Information Visibility on the Web and Conceptions of Success and Failure in Web Searching

Yazdan Mansourian

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University Of Sheffield.

Department of Information Studies

September 2006

Abstract

This thesis reports the procedure and findings of an empirical study about end users' interaction with web-based search tools. The first part is dedicated to address early research questions to discover web user's conceptions of the invisible web. The second part addresses primary research questions to explore web users' conceptualizations of the causes of their search success/failure and their awareness of and reaction to missed information while searching the web. The third part is devoted to a number of emergent research questions to re-examine the dataset in the light of a number of theoretical frameworks including Locus of Control, Self-efficacy, Attribution Theory and Bounded Rationality and Satisficing theory.

The data collection was carried out in three phases based on in-depth, open-ended and semi-structured interviews with a sample of academic staff, research staff and research students from three biology-related departments at the University of Sheffield. A combination of inductive and deductive approaches was employed to address three sets of research questions. The first part of analysis which was based on Grounded Theory led to discovery of a new concept called 'information visibility' which does make a distinction between technical objective conceptions of the invisible web that commonly appear in the literature, and a cognitive subjective conception based on searchers' perceptions of search failure. Accordingly, the study introduced a 'model of information visibility on the web' which suggests a complementary definition for the invisible web. Inductive exploration of the data to address the primary research questions culminated in identification of different kinds of success (i.e. anticipated, serendipitous, and unexpected success) and failure (i.e. unexpected, unexplained and inevitable failure). The results also showed that the participants in the study were aware of the possibility of missing some relevant information in their searches and the risk of missing potentially important information is a matter of concern to them. However, regarding the context of each search they have different perceptions of the importance and the volume of missed information and accordingly they react to it differently. In view of that, two matrices including the "matrix of search impact" and the "matrix of search depth" were developed to address users' search behaviours regarding their awareness of and reaction to missed information. The matrix of search impact suggests that there are different perceptions of the risk of missing information including "inconsequential", "tolerable", "damaging" and "disastrous". The matrix of search depth illustrates different search strategies including "minimalist", "opportunistic", "nervous" and "extensive".

The third part of the study indicated that Locus of Control and Attribution Theory are useful theoretical frameworks for helping us to better understand web-based information seeking. Furthermore, interpretation of the data with regards to Bounded Rationality and Satisficing theory supported the inductive findings and showed that web users' estimations of the likely volume and importance of missed information affect their decision to persist in searching. At the final stage of the study, an integrative model of information seeking behaviour on the web was developed. This six-layer model incorporates the results of both inductive and deductive stages of the study.

Dedication

I dedicate this thesis to the memory of my father, who sadly passed away when I was away from home. I would also like to dedicate this work to my mother for all her support and encouragement throughout my life, and to my brother and sister, Ehsan and Maryam, who have always encouraged me during the years of my education.

Finally, I would like to dedicate the thesis to two people who have been a great source of strength to me during my time in Sheffield, and will continue to be so when I return to Iran: i.e., my wife, Leila, whose company and patience during my PhD studies in Sheffield were a great help for me; and my son, Shayan, whose love keeps my life full of joy and happiness.

Acknowledgements

This research was funded through a doctoral bursary awarded by the Ministry of Science, Research and Technology of the Islamic Republic of Iran. My deepest appreciation goes to the Iranian people who generously funded this research.

I express my sincere appreciation to my knowledgeable supervisors, Mrs. Sheila Webber and Professor Nigel Ford for their admirable supervision. Without their sage comments on my work I would not have been able to complete this research.

In addition, I must acknowledge the help of the biology community at the University of Sheffield, who willingly cooperated with me during the data collection stage. In particular, I would like to express my appreciation to those members of the departments of Animal and Plant Science, Biomedical Science, and Molecular Biology and Biotechnology who gave up their valuable time and assisted in this study. Their enthusiastic participation in the research led to the development of a very rich dataset.

I would also like to thank everyone in the department of Information Studies for their kindness to me during my study in Sheffield: in particular Professor Peter Willett, Professor Steve Whittaker, Professor Micheline Beaulieu, Professor Sheila Corrall, Dr. Helen Grindley, Dr. Mark Sanderson, Dr. Barry Eaglestone, Dr. Paul Clough, Dr. Peter Bath, Dr. Daniela Petrelli, Ms. Briony Train, Dr. Angela Lin and Dr. Miguel Baptista Nunes.

Many thanks go to my friends Dr. Andrew Madden, Mr. Peter Stordy, Dr. Ali Shiri, and Mr. Hamid Reza Jamali who had many useful comments on my work, and to my first teachers in LIS Dr. Rahmatollah Fattahi, Dr. Mohammad Hossein Dayani, Dr. Asadollah Azad, Dr. Mehri Parirokh and Mr. Mohammad Naimabadi who have always been a source of encouragement for me.

Thanks too, to my friends in the department: Abdulhadi, Andy, Christine, Fenio, Hideo, Maryam, Nora, Phussadee, Peter, Paul, Robert, Salim, Theocharis, Woody, and Zapopan.

I would also like to express my gratitude to my friends Dr. Reza Fadayevatan, Dr. Reza Aflatoonian, Mr. Behrouz Aflatoonian, Dr. Ahmad Reza Daraei, Dr. Saeed Valizadeh, Mr. Poorang Pirouzfar, and Dr. Mohsen Aarabi whose friendship during my stay in Sheffield made life easier and more enjoyable for me.

Table of Contents

List of figures

List of tables

CHAPTER ONE:

INTRODUCTION

CHAPTER1: INTRODUCTION

"In order to succeed, your desire for success should be greater than your fear of failure." Bill Cosby

Introduction

This thesis presents a detailed report on the background, motivations, methodological procedure and findings of a qualitative study on a specific group of academic end users' information seeking behaviour on the World Wide Web (hereafter the web) in general and their feeling, thoughts and reactions after receiving unsatisfactory search results in particular. Since the study had an exploratory approach, the research's main focus has been changed over different stages due to emergence of new directions. However, the original idea which was enhancing the existing knowledge about searchers' interaction with web-based search tools remained unchanged during the whole course of the research.

This chapter seeks to illustrate the main components of the research and elucidate the exploratory nature of it. Moreover, the chapter introduces the main features of the study and its aims and objectives and also explains the main motivations for carrying out this investigation. This study sought to uncover a number of hidden aspects of end users' search procedure on the web. In order to achieve this goal only one aspect which was the user-oriented side was the key viewpoint of the research. Therefore, the research did not intend to investigate the system factors and focused entirely on the human side of information seeking on the web. The main reason of limiting the study to the user aspects comes back to the fact that it would be too broad to address both aspects in an individual project. Being more specific, it is not possible to thoroughly address these two aspects during one single study and it requires different investigations with different data collection and analysis methods.

Statement of the problem

Information seeking on the web has become a routine for daily life. It is one of the most widely used media for access to a variety of information resources for a huge range of users. The web as a global information resource provides different groups of society with different information resources. At the present time millions of users around the world seek their desired information through this media. The increasing number of submitted queries to web-based search tools is one of the indications of the astonishing popularity of the web in providing people with wide ranges of materials. Although there are many web-based applications, web searching is one of the most popular ones. According to previous studies searching the web is the second most popular web-based activity (Hsieh-Yee, 2001; Laycock, 2004).

However, information seeking on the web is not always a successful procedure and may culminate in unsatisfactory results for the users. Moreover, despite the recent developments and achievements in designing new web search tools, our knowledge about the human side of web searching is still limited.

For whatever reason users search the web they expect quick and easy access to information. Although web searching seems an efficient way of accessing a huge range of information resources, it is not always as quick as end users might expect and can be very time-consuming and frustrating. Ceaparu et al. (2004) investigated aspects of end-user frustration including the frequency, cause and level of severity of frustrating experiences. According to their study the three most frequent activities causing frustration for end users were web browsing, email and word processing. What the authors termed "web browsing" included different forms of information seeking on the web including searching as well as browsing. Figure 1.1 summarises the findings, which indicate that internet-related applications are one of the major sources of computer frustration for end users. Other examples of problematic situations noted by Ceaparu et al. (2004: 334) include:

> "The annoyances of losing work when a crash occurs, struggling to understand an error message or spending too much time to clear away spam and viruses have become symbolic of the struggles associated with modern technologies. Computers can be the cause of many problems, usually at the worst time possible. Some problems stem from the users' lack of knowledge, poor training, or unwillingness to read instructions or take tutorials."

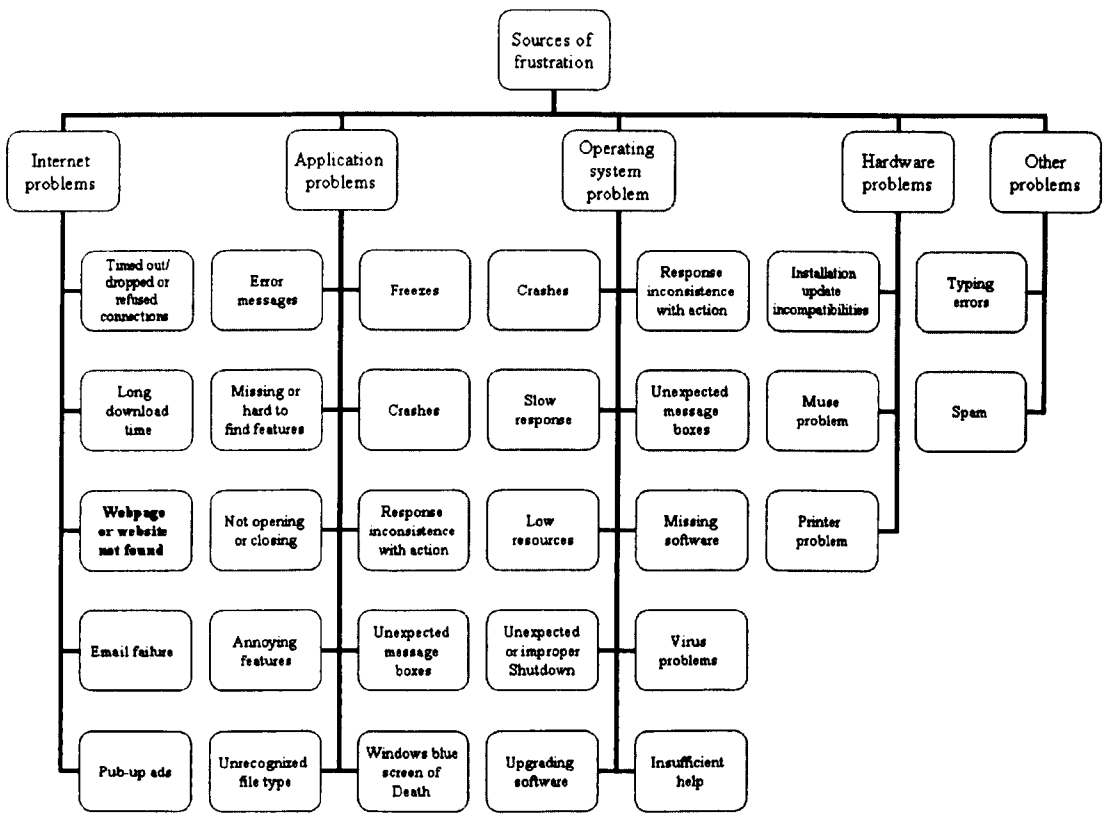

Figure 1.1: Sources of end users' frustration (based on Ceaparu et al. 2004)

On the other hand, as Huberman et al. (1998) stated that slow access to required resources and inability of finding relevant information on the web are the two most frequently reported problems of web users. They explained more about these two common categories of web users and argued that as slow access is the result of "congestion problems" and difficulty in finding relevant information is a result of "balkanization of the web structure" and solving these two problems are very difficult, therefore, research on web users' behaviour is one of the ways that will be useful to decrease the severity of users' difficulties of web searching. Huberman et al. (1998) reported:

> "The slow access has to do at least partly with congestion problems whereas the difficulty in finding useful information is related to the balkanization of the Web structure. Since it is hard to solve this fragmentation problem by designing an effective and efficient classification scheme, an alternative approach is to seek regularities in user patterns that can then be used to develop technologies for increasing the density of relevant data for users".

Therefore, this study focused on the users' search behaviour rather than technical issues of web searching. Moreover, this study concentrated on users' feeling and perceptions to find out their real feelings, thoughts and actions while web searching in general and when they encounter problematic situation in particular.

Ceaparu et al. (2004:334) believe that three steps should be considered in order to deal with user frustration namely:

> "A first step is to gain a better understanding of what frustrates users of computers. Then taxonomies of frustrating experiences can be developed, and means to measure their severity and frequency can be identified. These three steps should lead to solutions with enough supporting evidence so that requests for improvements will be well received by all parties involved."

The study reported here relates mainly to the first and second steps referred to above, in that it seeks to illuminate frustrations encountered during web searching. It is one of an increasing number of research studies devoted to the study of web-based information seeking – an increase noted by a number of authors (e.g. Jansen and Pooch, 2001; Hsieh-Yee, 2001; Spink and Jansen, 2004 and Martzoukou, 2004).

However, although existing models of information seeking and searching behaviour explain people's feelings, thoughts and actions while attempting to satisfy their information needs (e.g. Kuhlthau, 1991; Ellis et al., 1993; Ingwersen, 1996; Choo et al. 2000), there has been relatively little research into people's perceptions of search success and failure and of what factors influence them. These questions form the focus of the present study.

Indeed, in the field of information behaviour several models have been developed during the past three decades which include (1) information behaviour models, (2) information seeking models and (3) information searching models. These models mainly address people's behaviour to satisfy their information needs. The main focus of these models is on people's feelings,

thoughts, and actions to locate information resources. Nevertheless, these models do not address people behaviour when they are not able to satisfy their information needs.

In fact, there is a gap among the current research body about the information seeking behaviour. The existing models attempt to clarify how people feel, think and act towards satisfying their information needs. It is not clear how they feel, think and act when they are not able to locate their desired information and how they cope with the failure of information seeking.

In terms of research background, specific characteristics of this study make it distinctive and it is difficult to identify an explicit research background for this study. Nevertheless, it does not mean that the current study has no link with the previous literature body. In contrary, there are many strong link points between this study and some related research body. The point is because of the multidimensionality of this study it is not possible to ascribe it to one particular research field.

This research is multi-dimensional because it relates to different aspects of web searching including information visibility on the web in one hand and users' perceptions about their search experiences in another hand. In fact, there are a number of research areas that are associated to this study. The details of these links have been explained in the chapter two[1].

Nevertheless, as an introductory comment on the background of this study it can be mentioned here that understanding and interpretation of people's interaction with electronic resources and particularly online resources has been the topic of many studies over the past two decades. This interactive procedure can be investigated from different perspectives and within different contexts. In terms of the perspective, these studies can be carried out with system-oriented or user-oriented approaches.

In terms of the online resources this study is limited only to the web and regarding the main focal point of the research it solely focuses on the user-oriented aspects and excludes the technical aspects of information retrieval on the web.

Research aims & objectives

This study sought to achieve the following aims:

- To explore web users' information seeking behaviour on the web in general and how they perceive information seeking failure in particular.
- To identify the influential factors affecting users' perception about information seeking failure on the web and then explaining the role of each factor.

[1] Literature review

- To discover what people do and the strategies that they take into consideration when they encounter unsatisfactory results.

This research also concerns with users' awareness of missed information. Sometimes users may omit some relevant information resources but they might not be aware of such a problem. Consequently, users are not using the information retrieval system to full capacity. Therefore, users can not make a realistic judgment about what they have found.

Research rationale

The rationale of this research is based on three aspects including why the web has been selected as the context of the research, secondly, why among all functionalities of the web applications only web searching formed the main focus of the study and thirdly why the user-oriented aspect of web searching received the key attention of the researcher.

First of all, the web is a global medium of information resources and human communication and any investigation to make the web more efficient would be valuable for consideration. On the other hand, in parallel of popularity of the web as a source of information and a means of communication it has attracted researchers' attention over the last ten years. As Hsieh-Yee (2001:181) noted that:

> "Web search behaviour is a fertile ground for research... Researchers can pursue nearly any aspect of searching they wish. But ... they need to place users in the centre of investigation, understand users' needs and find out how the current web environment support or hinder their information seeking."

Regarding the reason of the specificity of the study on web searching, it seems sufficient if we consider the fact that after email web searching is the second most popular web-based application (Hsieh-Yee, 2001). Therefore, web searching is an important area of study and the growth of the number of investigations on this area in the recent years is an indication of researchers' attention to the area.

The third principle of the rationale behind this research is the reason of focusing on the user-oriented aspect of web searching. There are two issues that can be taken into consideration to clarify the importance of the third aspect of the research.

The study was focused on the user aspect of web searching because we still need to know more about the user's aspect of web searching. Moreover, this study explored qualitative aspect of the user side which is related to perceptions and conceptualizations rather than the quantitative measures of the number of queries and statistical tests. Reviewing the literature showed since commence of web search research in 1995 most of the studies (e.g. Jansen and Spink, 2000; Spink et al. 2001) have tended to the quantitative aspects of web searching and the need for covering the qualitative aspects of web searching to uncover users' perceptions and their real feelings of and reactions to web searching seems necessary. As Martzoukou, (2004) cited Newby (1998:4) and declared as follows:

> "What is lacking is the realisation that maximum effectiveness of an information retrieval system in producing information to address information needs could only be obtained with a detailed understanding of the conscious state of the information seeker' (Newby 1998: 4). Research on Web information searching is useful for examining behaviour and actions but is not adequate for explaining the factors and processes that have led to that behaviour."

Moreover, Ondrusek (2004:254) states:

> "The migration of vast amounts of information to online storage and retrieval systems, worldwide, provides strong justification to expand efforts to increase the understanding of end-user psychology. Because past research focused upon populations concentrated in academic centres, a return to those centres to conduct follow-up studies could be a viable next step in the end-user research continuum."

Regarding all above debates among all online resources the current study only focused on the web and among all web-based applications solely concentrated on web searching and among all aspects of web searching merely narrowed down to the qualitative characteristics of the user side of this process. Therefore, we can define a specific approach and framework for the research which is explained in the next immediate section.

Research approach & framework

As it was declared this study employed a user-oriented, exploratory, and qualitative approach to carry out an empirical study focusing on the user side of human computer interaction while information seeking procedure on the web.

A qualitative approach was selected to provide the researcher with more opportunities to develop new understandings and knowledge about users' interaction with the web resources. As Wilson (2000) states in information science we need to develop concepts rather than using pre-exist concepts:

> "Qualitative research ... is concerned with developing concepts rather than applying pre-existing concepts,' (Halfpenny, 1979) and, given the state of theory in information science (that is, its undeveloped state) it can well be argued that 'developing concepts' is what is needed".

In terms of the context of information seeking, this investigation is limited only to the web environment. In fact, the current research did not seek people's information seeking behaviour in general. There are two main reasons for making such a focused framework.

First of all, web has become an indispensable source to provide people with their desired information. Therefore, the results of the research can be useful for a wide range of end users. Secondly, apart from a few exclusive studies on information seeking on the web (e.g. Choo et al. 2000) the number of studies which have been focused on people's information seeking behaviour just on the web is still limited and most studied in this area cover different media of

information sources. As a result, limiting this study to the web makes it distinctive in the research area and it can enhance the originality, depth and specificity of the research topic.

In terms of the research population, the research was limited to a specific group of web users to provide the researcher with an opportunity to focus on a more homogenous group. However, the web-based information resources are not limited to a specific domain or subject and may include any web-based information. Nonetheless, because of the homogeneity of the research population as it had been anticipated there were more similarities among the web-based resources that the targeted group of the study have been using.

Research journey

One of the main features of the current research is its exploratory nature and it seems necessary to highlight this aspect in the introduction chapter. The research has naturally evolved through seven main stages including origination, orientation, exploration, elucidation, consolidation, reflection and culmination. Although there might be similar categorization in research methodology textbooks, what it is mentioned here is based on the actual experience of the researcher.

Origination

At the outset, the definition of the term "invisible web"[2] was the main core theme of the research. However, the level of uncertainty about the main focus of the study was quite high at that time and it was not clear for the researcher which is the best aspect to follow. The only clear point was the necessity of conducting this study to find out how serious this problem is and how this research might help to improve the situation.

The researcher knew some parts of the web are not easily accessible for end users and it is called the invisible web and this hidden part of the web makes the web searching process less efficient. In this stage a comprehensive review of the literature about the invisible web phenomenon was conducted and it became clear that research about this issue was mainly technical oriented and the user aspects of this phenomenon has not received enough attention up to that time, year 2003.

Nevertheless, further review during the next years also indicates that the existing publication about the invisible web has remained descriptive and no serious study has been conducted up to now, year 2006. In that time the researcher noticed the nearest area to this research is web search research[3].

[2] A detailed explanation about the definition of this term is presented in chapter two.

[3] A review of web search research is presented in chapter two.

Orientation

In orientation stage the researcher's viewpoint moved into two directions of breadth and depth of the area of web search studies. In terms of breadth it was decided to focus only on the user side of web searching and in terms of the depth it was decided to not go beyond than users' perceptions and conceptualizations of their genuine search experiences.

The orientation stage played an important role to form the main focus of the research. Although new directions appeared during the next stages, what has been formed in the orientation stage remained unchanged for the rest of the study.

Exploration

One of the most important and extensive stages of this study was the period of data collection and analysis. In this stage, the researcher sought to find the answer of the primary research questions[4] through a systematic data collection and inductive analysis which ended up with reasonable answers to the primary questions.

Grounded theory [5] was the main tool to carry out this step. The details about the methodological basis of the research which formed the exploration stage have been illustrated in the chapter three (methodology) and the results have been presented in five chapters from chapter four to nine.

Elucidation

After having some findings based on the results of the data analysis there were enough explanations to address the research questions reasonably well. However, there were two more issues remaining to address.

First of all, the findings were quite fragmented and it was necessary to integrate the results together. Secondly, it was necessary to validate the findings in any plausible way. The details have been reported in chapter ten. It was the time for the researcher to elucidate the findings of the exploration stage.

Consolidation

Consolidation stage was one of the most sensitive steps in the research. It was sensitive because the findings required more consistent links together to develop the final outcome of the research. The findings from the exploration stage and the further works in elucidation phase required more consistency to build up a coherent body of the results which would be easily comprehensible for the audience of the research.

[4] There is a detailed explanation about the research questions in the next section in this chapter.

[5] In the methodology chapter there is detailed illumination about the basics of Grounded Theory.

To cover the first remaining aspect of the elucidation stage it was useful to synthesise the fragmented results to foster the links between them. Therefore, the researcher called this stage as consolidation step. This stage also has been explained in chapter ten.

Reflection

In the reflection stage the researcher sought to reflect on the results from a broader viewpoint. In order to address the second part of the fourth phase, elucidation stage, it was necessary to map the findings into a bigger picture. Accordingly, the researcher has carried out another review of the literature to find out possible linkages, if any, with what he has identified inductively based on analysis of the dataset[6].

The second review revealed the fact that there were a number of suitable theoretical frameworks to support the results. In fact, in reflection stage the inductive results of the study were deductively analysed once more to foster the validity of the findings. In other words, in consolidation stage a new theory has been developed and it was a good chance to embed the newborn theory into a bigger picture of human behaviour. The details of this stage have been elucidated in chapter ten and eleven.

Culmination

The culmination step was the climax and maturity phase of the research. The culmination stage was the direct result of success in the reflection phase. The reflection phase was very productive because, interestingly, two theoretical frameworks were identified in the related literature bodies.

These theories were Attribution Theory (Heider, 1958) and Simon's Bounded Rationality and Satisficing (Simon, 1955; 1956). Moreover, suggesting further research was another outcome of this step. In fact, in the culmination stage not only the research addressed the research questions but also many fruitful questions have emerged which can be researched in future studies.[7]

Therefore, the current research began with a definition of the invisible web and ended up with merging the results into two theoretical frameworks of human behaviour. Although the beginning and the end of the study seem somehow different, the explanations in the above stages show how it happened and it just confirm the exploratory nature of the study. All these seven ordered stages have been visualised in a pyramid form and is presented in figure 1.2.

[6] The details of this stage are presented in chapter ten and eleven.

[7] The detail of this stage is presented in chapter twelve.

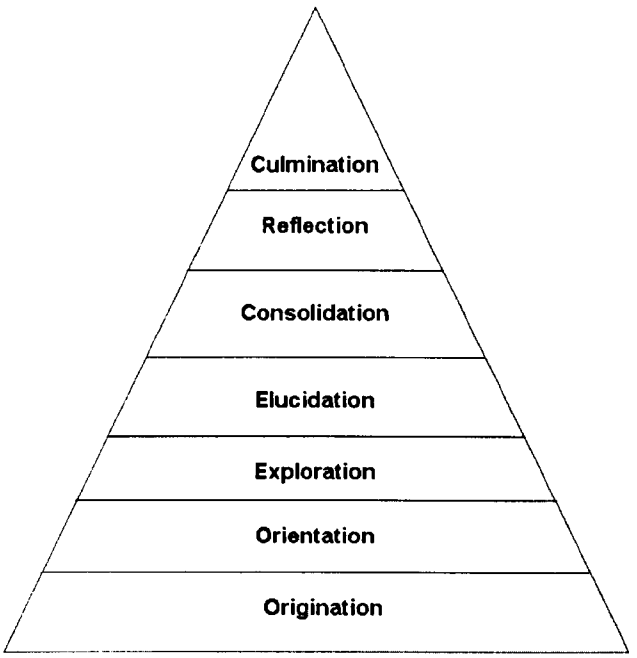

Figure 1.2: Seven stages of the research

Research questions

The research questions in this study differ from the conventional studies which usually begin with specific research questions at the outset and seek to answer the same questions over the course of the study. Developing research questions in this study happened in a long period of exploration and reflection. As a result, there are three sets of research questions in this study which emerged during various stages.

Citing to the figure 1.2 three sets of research question including early, primary and emergent questions were developed during three stages of origination, orientation and reflection. Nevertheless, all these questions are related together and there is a strong connection between them. In other words, emergence of the new directions in the research did not stop the researcher to pursue the early questions and shift from one perspective to another. Therefore, though there are three groups of research questions they share the same theme and all are related to users' problematic situation in web searching.

Indeed, raising new research questions did not distract the researcher from the main focus of the research and these new directions did not stop him to be consistent in the course of the research. In fact, new research questions in the later stages of the research helped the researcher to expand the research scope and follow up the research in a reasonable way. At the early stages of the research there were two Early Research Questions (ERQs) that formed the main motivation of the research at the beginning. These two key points included whether end users are aware of the existence of the invisible web and also how they perceive it in general.

However, conducting a preliminary study, which has been explained in chapter three, raised six new questions that formed the Primary Research Questions (PRQs). These PRQs formed the main focal point of the research in the main phase of data collection and analysis.

Eventually, in the final stage of the data analysis and regarding the findings three more related questions raised which formed the Emergent Research Questions (ERQs). These questions were called emergent because they literally emerged after inductive analysis of the dataset and reflection with the literature body. Figure 1.3 shows an overview of the development procedure of the research questions. There are more explanations about each category of questions and the role that each one played in the progress of the study.

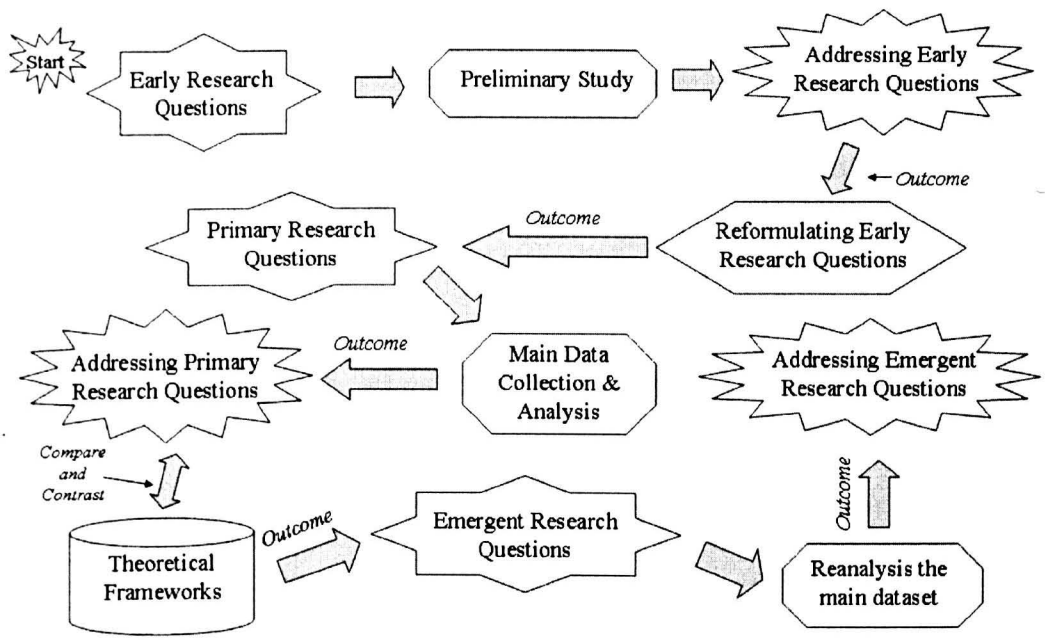

Figure 1.3: An overall image of the development of three sets of research questions

Early research questions

In the early stages of the study there were two ERQs that were formed during the origination stage of the study. These research questions were inspired by the definition of the invisible web and lack of any empirical research about this phenomenon.

In other words, the researcher's initial review of the literature showed though this term has appeared in a fairly notable number of publications there was no research on this issue. In that time there were no explanations in the literature to illustrate whether this term is just an idea of

librarians about the web resources or this phenomenon is really an important aspect of real web searching that makes a difference for end users. Therefore, the following research questions were developed to find out more about how much end user's real search experiences match with the definition of the invisible web and also whether there are a substantive ground to carry out a large-scale research on this issue or not. ERQs are as follows:

- ERQ1: How do a particular group of end users conceive the "invisible web"?
- ERQ2: To what extent do such conceptions and experiences map onto the technical models of the invisible web proposed in the literature?

These questions were explored initially in the preliminary study to attain three purposes. First of all, the researcher was looking for a way to formulate the final research questions before embarking on any major data collection and ERQs were the initial step to achieve this goal. Secondly, beginning with ERQs and carrying out the preliminary study was a good confirmation of the feasibility of the research in terms of suitability of the research theme and the effectiveness of interview as the main data collection method. Thirdly, exploring ERQs was a practical way to identify the most appropriate methodology for the research which was Grounded Theory[8].

Primary research questions

PRQs were developed after exploring ERQs during the preliminary study. PRQs are direct results from the evolution of ERQs. Having a closer look at ERQs and PRQs shows there are strong links between them. However, PRQs are more established and cover broader grounds.

For example, ERQ1 and PRQ1 are actually related together but PRQ1 possesses a much broader perspective than what ERQ1 does because the preliminary study showed that end users' conceptualization of failure in web searching covers a fertile ground for research in which the invisible web is only one aspect of this multidimensional issue. In other words, the researcher's initiative for this research was mainly formed by the concept of the invisible web but the preliminary studies indicated that there is a bigger ground for research behind the definition of the invisible web.

Two concepts of "information seeking failure" and "information missing" formed the axis of development of PRQ and both of them were developed as a result of exploring the definition of the invisible web. After carrying out the preliminary study the researcher discovered that the invisible web can be a subject for a user-oriented research if we consider its effect on the real web users' information seeking failure on the web. In fact, failure of locating the required information and the possibility of missing relevant information were identified as two aspects of the invisible web which was an original topic and worth investigating. Therefore, the following six research questions were developed.

[8] There are detailed explanations about Grounded Theory in chapter three.

- PRQ1. How do web users conceptualize their information seeking failure on the web?
- PRQ2. To what extent are web users aware of missing some information resources on the web?
- PRQ3. If web users are aware of missing relevant information, how important is it for them?
- PRQ4. How do web users seek to cope with information seeking failure on the web?
- PRQ5. What strategies do web users use to cope with information seeking failure on the web?
- PRQ6. What factors might influence different conceptualizations of information seeking failure, and different coping mechanisms?

Emergent research questions

The exploratory nature of this research was a great source of dynamism and vitality in the whole process of the study. Even after addressing the PRQs there were new fruitful grounds to cover through emergence of new directions.

In addition of the PRQs there are three EmRQs which emerged during the reflection stage of the research. Inductive analysis of a rich qualitative dataset yielded the main body of findings in the study which was able to address the PRQs. However, it was not the end of this research journey and a deductive exploration of the findings ended up to some powerful links between what was discovered in this study and what has already been developed in another areas including behavioral psychology and social science theories. In fact, at the beginning and in the middle stages of the research when the researcher was addressing ERQs and PRQs there was no link between the aforesaid areas and this study. The findings after addressing PRQs resulted to make a bridge between this study and the above subjects. In fact, these frameworks were unrelated to this research at the beginning and then became relevant in the final stage as a result of addressing PRQs. To explore the connections of the findings of this research and the new identified literature body the following research questions were developed as EmRQs:

- EmRQ1: What are the perceived causes of search success and failure of a particular group of academic users?
- EmRQ2: To what extent are Locus of Control and Attribution Theory useful in interpreting users' conceptualizations of success and failure in web searching?
- EmRQ3: To what extent is the theory of Bounded Rationality and Satisficing useful in interpreting users' awareness of and reaction to the missed information while web searching?

The three above EmRQs were formed the final focal point of the researcher in this study. However, it is very important to declare that emerging new research questions does not mean that the direction of the research has been constantly changing and after developing new sets of questions the researcher stopped paying attention to the previous aspects. In fact, during the course of the research all three sets of early, primary and emergent research questions have been addressed. As these three sets possess equal importance in the study this thesis covers all of them. Otherwise, it would be easier to choose one set as the final research questions and address them only. In fact, the researcher believes that the focal point of this study consists of the combination of these three sets and removing any part of them will make a deleterious effect on the whole picture that this thesis seeks to depict.

Thesis's summary

Chapter one describes aims and objectives of the study and elucidates three sets of research questions which have been addressed through the study. Moreover, the first chapter gives an overall view of the research approach and introduces the main components of the study. This chapter shows why the study is an exploratory research and how it evolved during different stages.

Chapter two presents a review of the literature body of information seeking on the web and with a specific attention to the invisible web. Moreover, chapter two reports a brief review of the related conceptual frameworks including Attribution Theory and the Theory of Bounded Rationality and Satisfying.

Chapter three is entirely about methodological aspects of the research and illustrates how Grounded Theory has been employed to address the research questions, why the study adopted a qualitative approach and why the biology community at the University of Sheffield has been selected as the target group.

Chapter four presents the demographic information of the participants who voluntarily participated in this study. Chapter five presents some descriptive results to report the findings about participants' search experiences, their motivations and feelings of information seeking on the web, and their positive and negative comments about web searching.

Chapter six focuses on the concept of information visibility on the web and reports results of the inductive analysis. This chapter demonstrates how inductive exploration of the data culminated in development of a new model of information visibility on the web.

Findings about users' conceptualizations of and reaction to unsatisfactory search results are presented in chapter seven. This chapter shows how a qualitative and in-depth investigation about users' interaction with web-based search tools in general and their perceptions of their search performance helped the research to identify different aspects of users' feelings, thoughts and actions whilst searching the web. Findings of the study about the participants' awareness of and reaction to missed information are presented in chapter eight.

Chapter nine reports results of the study about participants' coping strategies to overcome search failure. Chapter ten reports the result of a deductive analysis in the light of Locus of Control and Attribution Theory. Chapter eleven demonstrates the result of another phase of deductive analysis based on Bounded Rationality and Satisficing theory. The overall picture of the study is presented in figure 1.4.

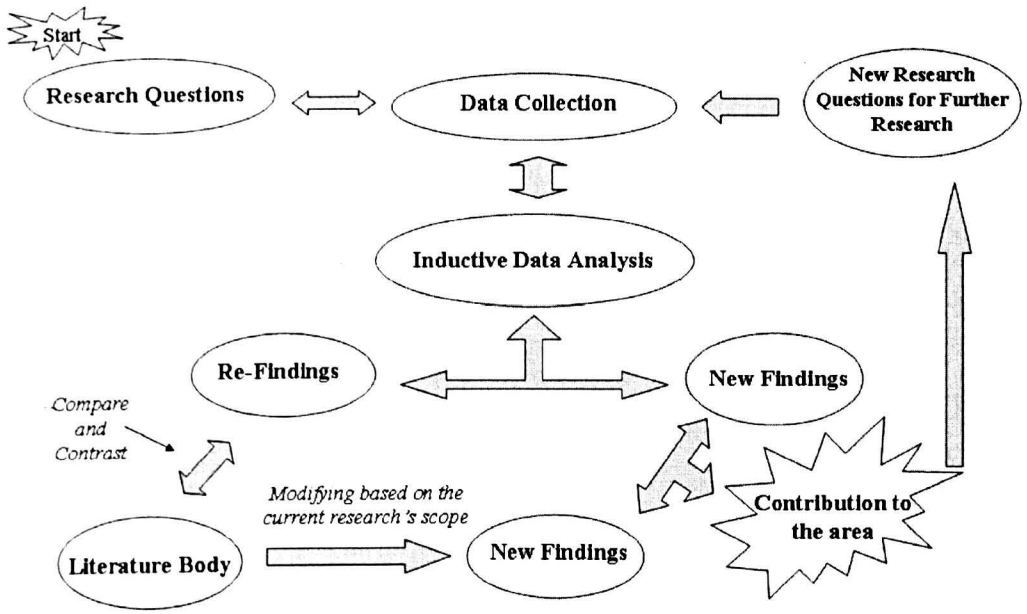

Figure 1.4: An overall diagram of some of the key components of the current research

Chapter summary

The introduction chapter presents an overall picture of the major characteristics of the study. The key features of the research forms its main theme which is about end users' conceptions of the invisible web, their information seeking behaviour on the web, their conceptualization of information seeking failure on the web, their awareness of and reaction to missed information whilst web searching, and their coping mechanisms to overcome search failure. This chapter addresses key aims and objectives of this study and spells out the foremost motivations and the anticipated outcomes of the research. As it has been stated in the chapter this is a user-oriented, exploratory, qualitative and empirical research.

CHAPTER TWO:

LITERATURE REVIEW

CHAPTER 2: LITERATURE REVIEW

"Most people think of success and failure as opposites, but they both are products of the same process." Roger von Oech

Introduction

This chapter presents a detailed and critical review of the related literature to the current study. Because of the multidimensionality nature of the study the literature review inevitably consists of different sections which are related together. Moreover, because of the exploratory approach of the research and its links to different fields during the course of the investigation there are different bodies of the literature that are relevant to this research and the researcher had to consider all of them in the literature review. In fact, this chapter reflects the multi-dimensionality nature of the study.

The order of the presented material in the chapter is based on the actual order of different stages of the research. Therefore, the chapter firstly includes the literature associated to the invisible web and reasons of information visibility/invisibility on the web. The review continues with considering the current issues and trends in information seeking on the web. Finally, the chapter ends up with a brief review of those theoretical frameworks which are associated to this study. Figure 2.1 shows the position of this study with the related areas.

Regarding the three sets of research questions which were explained in chapter one, each part of the reviewed literature is related to one set of the research questions. To be more precise, the literature on the invisible web mainly relates to the ERQs, the literature on the information seeking on the web is relevant to PRQs and finally theoretical framework on human behaviour associates to the EmRQs. This pattern exactly matches with the role of literature review in Grounded Theory. Glaser (1992:31-37) suggests about the role of the literature in developing Ground Theory as follows:

> "With regard to grounded theory methodology there are three types of literature: (1) non-professional, popular, and pure ethnographic descriptions, (2) professional literature related to the substantive area under research, and (3) professional literature that is unrelated to the substantive area."

In this categorization the first part of the literature is related to the phenomenon under the study at the beginning, the second part is related during the study and the third part which is unrelated at the outset of the research becomes relevant at the final stages.

Although these three proportions of the literature possess almost equal importance for the research, the researcher usually involve with the second part within a longer period of time and need to delve into this section more deeply. Because of the importance of the second part of the literature, Glaser divides this part of the literature into two areas. In the first areas the phenomenon under the study has not been researched in great deal and Glaser includes the subjects with "a modicum of literature" and subjects with "an abundance of literature".

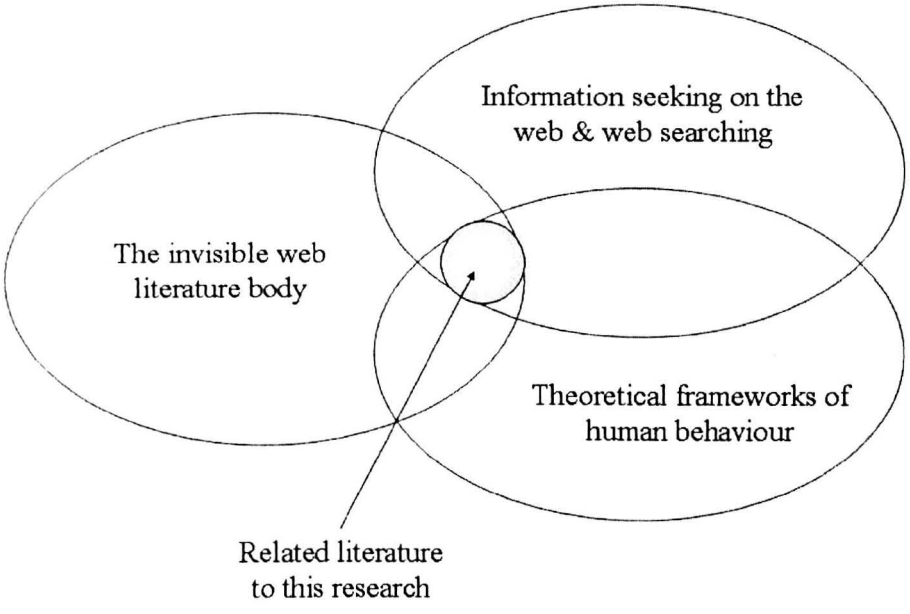

Figure 2.1: Position of the current research among the related areas

This categorization of the literature fits very well with this research. In fact, literature on the issue of success and failure represents the "modicum of literature" and web search research reflects "an abundance of literature". Therefore, this study relates to a large body of the literature but the main topic of the study is related to only a small portion it. Glaser (1992: 31-37) continued with describing the role of each part of the literature in a Grounded Theory methodology.

He highlighted that despite the existence of a rather common misunderstanding about the role of literature review in Grounded Theory that some people suppose in this methodology the researcher should not review the literature, s/he should review the literature but based on the above categorization. In general, Glaser's categorization of the literature fits with this research thoroughly.

Existing publications on the invisible web forms the first part of the literature which is mainly non-professional and descriptive, literature on information seeking on the web and web searching is the professional literature and finally theoretical frameworks forms that part of the literature which was unrelated at the beginning and then it became related in the concluding stages.

Figure 2.2 depicts the links between these sections of the literature and this research and the research questions.

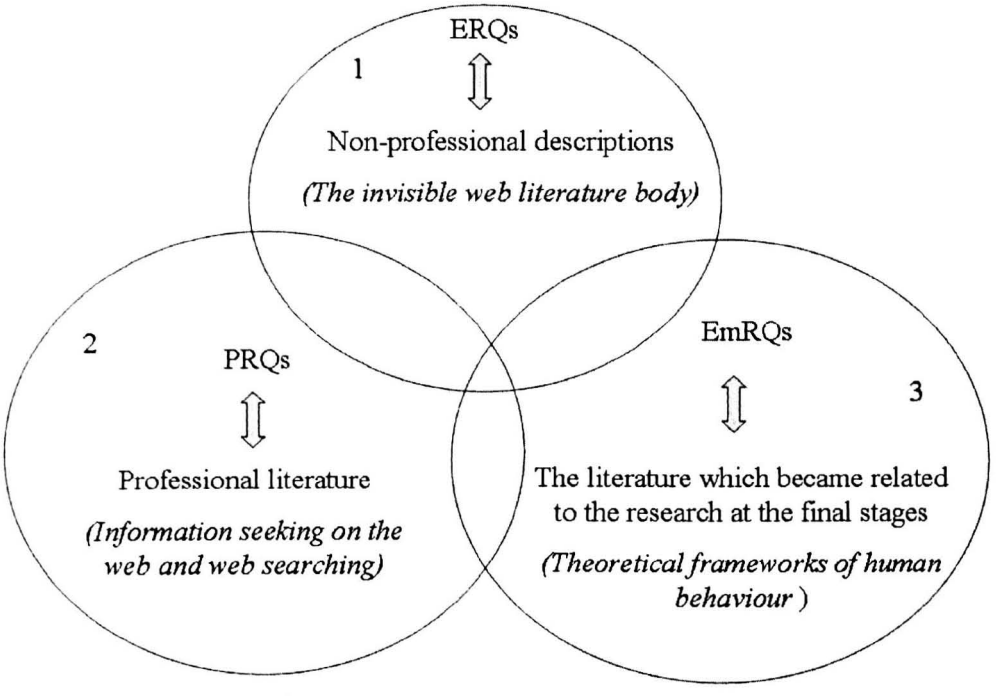

Figure 2.2: An overall image of the links between different parts of the literature and the research questions

Search strategy to identify the relevant literature

In order to identify and review the relevant literature the researcher took a long term search strategy. This long interaction with the literature body began at the origination stage of the research and continued over the remaining parts of the study.

At the first phase of the review the researcher carried out thorough searches using different databases including Library and Information Science Abstract (LISA) and Web of Knowledge and Science Direct. However, this introductory search only formed the initial part of the whole literature review and it continued through a long term browsing and surfing the area through different routes which are explained as follows.

The early search in the aforesaid databases was just a start for a long term interaction with the literature body. The researcher has been browsing different journals and was subscribed for alert services of the core journals including JASIST[9], JDOC[10], IPM[11], LISR[12] and JIS[13] to receive newly published issues. However, in order to show the status of the related research

[9] Journal of the American Society for Information Science & Technology

[10] Journal of Documentation

[11] Information Processing and Management

[12] Library and Information Science Research

[13] Journal of Information Science

there is brief explanations for the search strategies that have been employed to identify the related literature.

Moreover, the researcher gained access to the related literature through other routes including serendipity and during a long interaction with the research community in different events. For example, he published a number of papers and delivered some presentations to have other fellow researchers' comments which were very helpful for different aspects of the study including access to a comprehensive coverage on the literature. For example, during three years of the project the research submitted six papers to the related journals, presented two work-in-progress presentations in the department, three presentations in different doctoral forums in the UK and Europe, four posters, three conference presentations and constant discussions with fellow researchers in the department. The details of these publications and presentations are presented in Appendix 8.

Nonetheless, all these useful routes did not stop the researcher of performing classic search in the relevant databases. In the early stages of the research to find the current trends and issues on the invisible web the researcher has carried out extensive search in databases. For example, an advanced search in LISA for the term "invisible web" as descriptor from the earliest date in the database to year 2003, which was the start date of the research, produced 20 items and after limiting it to English documents it reduced to 18 records. The same search strategy with a broader time range from the earliest date to year 2006 which is the final year of this research produced 24 items. A broader search for the invisible web as keyword for all languages from the earliest date to year 2003 produces 50 documents and after limiting it to English documents it reduces to 44 items.

In the most general search to cover all existing publications which is based on "invisible web" as keyword for all years in all languages produced 57 items. Regarding the hugeness of LISA this number of retrieval shows that this issue has not been received sufficient attention. However, even among these limited number of documents there is no serious and empirical research on the issue. Almost all of the existing publications on this issue are mainly descriptive without profound analytical view. To increase the coverage the search and to minimise the possibility of missing relevant information on the topic similar searches have been carried out with extra keywords including "hidden web", "hidden internet", and "deep web". However, changing the keywords did not show considerable difference in the result.

Literature on the invisible web

Regarding the above categorization of the literature, the first part of the literature body which is related to this research is publication about the invisible web. The main theme of these publications is formed of introducing this phenomenon in a rather descriptive way and highlighting its importance in hiding information resources on the web. The first issue in this part of the literature is the definition of the invisible web which will be described as follows. However, before explaining what the invisible web is it would be useful to describe what the visible web is. Having a clear idea about the meaning of the visible web is helpful to understand the invisible web.

The visible web

The visible web consists of the indexable parts of the web which is easily accessible for general-purpose search engines including Google, AltaVista and Alltheweb. The main part of the visible web consists of the static HTML texts which are freely accessible on the web. Figure 2.3 which is a static HTML webpage is an example of millions of web pages which are visible for general purpose search engines. A simple search in almost any general-purpose search engine with an appropriate keyword like "University of Sheffield" will lead to locating this website.

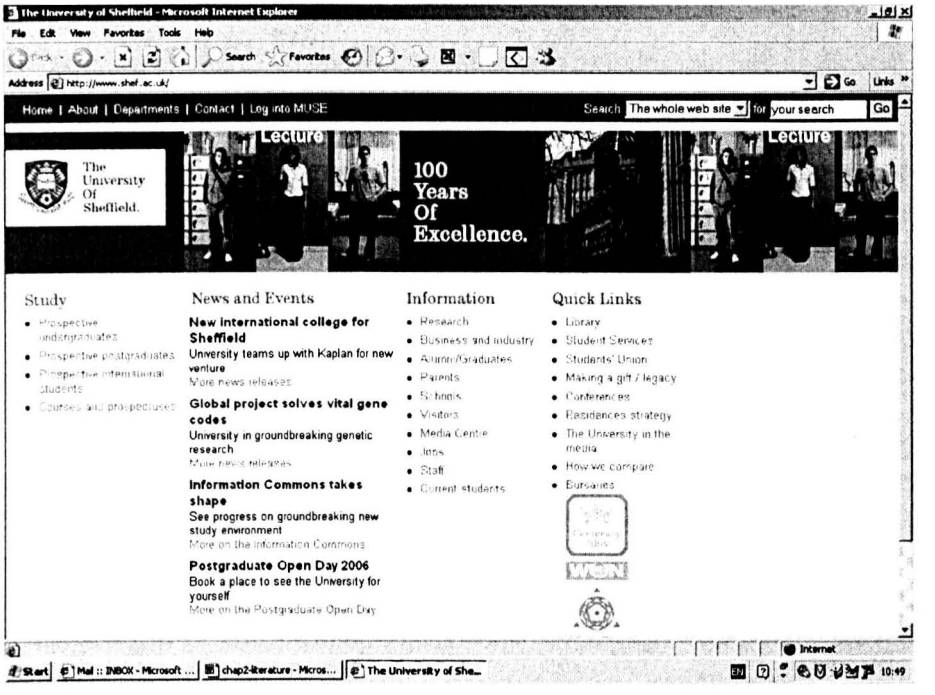

Figure 2.3: The homepage of the University of Sheffield's website as an example of visible web resources

How big is the visible web?

We are not able to answer this question unless we can find out how big the web is. Nevertheless, because of the specific features of the web environment including its unbelievable growing rate, universality, dynamism, distribution and so on if it is not impossible it is extremely difficult to find out how big the web is.

Accordingly, any calculation about the size of the web should be extremely approximate. On the other hand, even if we would be able to have a fairly correct estimation about the size of the web it will be out of date so quickly. Moreover, there is no credible way to measure the size of the web without facing many problematic questions about the validity of this measurement. In fact, if we measure the size of the web in terms of the number of websites, the size of websites are extremely different and varying from websites with a few pages to websites containing thousands web pages. For example, the website of BBC (http://www.bbc.co.uk) is a

site that includes perhaps hundred thousands of pages while a website of a small company or a personal website might consist of only a few pages.

The other possibility is measuring the size of the web regarding the number of web pages. The diversity of web pages is also enormously different. Some pages are very short with a little information/data while some of them are very long and include huge amount of data. The third possibility is measuring the size of the web by measuring the size of websites in terms of byte/megabyte. This method is also problematic because some websites includes images or other big size files that make them unnecessarily heavy. Therefore, any calculation based on this criterion can not be convincingly valid.

Nevertheless, Sherman (2001) reported the visible part or indexable part of the web should be between 2.5 and 4 billion pages.

> "Officials from Google and the other major engines readily admit they have not indexed the entire web. Aggregating estimates from several reputable sources puts the size of the visible web at somewhere between 2.5 and 4bn pages, growing at the rate of about 7m a day."

After four years another report estimates that the indexable part of the web is about 11.5 billion pages (Gulli and Signorini, 2005).

As the visible web is formed by the part of the web which can be indexed by search engines it is referred as indexable web as well. However, in this research the term "indexed web" is preferred because there are some resources on the web which are indexable but search engines ignore them either because of the financial limitations or because of the information owner's willing to hide the information. Besides, there are resources which are indexable but still have not indexed and include a part of the invisible web, opaque web, which are illustrated in the next sections.

The invisible web

A significant portion of the web cannot be accessed via conventional search engines because they do not know it is there (Smith, 2001). Consequently, different web-based sources have different levels of accessibility. This phenomenon has led to the emergence of the term "invisible web".

Although the term "invisible web" has appeared in many publications (e.g. Diaz, 2000; Sullivan, 2000; Sherman and Price, 2001; Sherman and Price, 2003; Pedley, 2001; Pedley, 2002; Walter, 2002 and Clyde, 2002, Crowther et al. 2004), little if any user-based empirical research has been published. Rather, the invisible web has generally been considered from a predominantly technical perspective in terms of the reasons for different types of invisibility.

Devine and Egger-Sider (2004) state that the earliest occurrence goes back to the papers published by Lawrence and Giles in 1998 and 1999. However, the origin is undoubtedly earlier, with Sherman (2003) attributing the term to Matthew Koll – others (e.g. Garcia, 1996; Bergman, 2001; Hricko, 2002; Hobbins, 2001; Crowther et. al. 2004) to Ellsworth, who used it in 1994 in an interview with the Canadian *Computer Paper* magazine when referring to:

> "... a site that's possibly reasonably designed, but they didn't bother to register it with any of the search engines. So, no one can find them. You're hidden. I call that the invisible web".

Sherman and Price (2001) commented that the invisible web is huge, and in all likelihood is growing more rapidly than the visible web. Pedley (2002) also notes the size and quality of information that "resides" in this area of the web and which may thus be missed. "Invisible" resources consist of material that general-purpose search engines either cannot, or are not intended to index.

Walter (2002) called such invisible resources the "buried treasure of the web". He noted that this portion of the web is not totally invisible – but rather, just invisible to users of conventional web-based search engines. As Devine and Egger-Sider (2004) state:

> ".. the invisible web comprises all the information sources available on the web that are overlooked by conventional search engines."

Block (2004) classified web resources into seven major groups according to possible reasons for invisibility. These include resources that are: too recent to have yet been indexed; non-text; dynamically generated; password protected; not linked to; too deep to be indexed; and simply not viewed (e.g. due to limitations in the number of search results viewed by users or the search engine retrieved the site but is screening it out).

However, a more extensive classification, which includes and extends Block's categories, is provided by Sherman and Price (2001 & 2003), who group the invisible web into four main categories: the opaque, the private, the proprietary, and the truly invisible web. These are shown in figure 2.4. The explanation of each part of the invisible web is as follows:

The opaque web

The opaque web comprises resources that could be included in search engine indexes but are not. There may be several reasons for this, including depth and frequency of crawling. It is expensive for a search engine to crawl and index every single page on a website. Therefore, some providers adopt a policy of "good enough" rather than deep crawling, the result of which is that information is missed beneath certain hierarchical website levels. Search engines also differ in the way they display search results, usually limiting them in some way.

Google, for example, only displays two results from any one site. Also, there is inevitably some delay between information being added or changed and its appearance in a search engine's index. This is particularly problematic with current, real-time or constantly changing information – for example, company stock price, news and weather forecasts. Pedley (2001) termed this area the *vanishing* web.

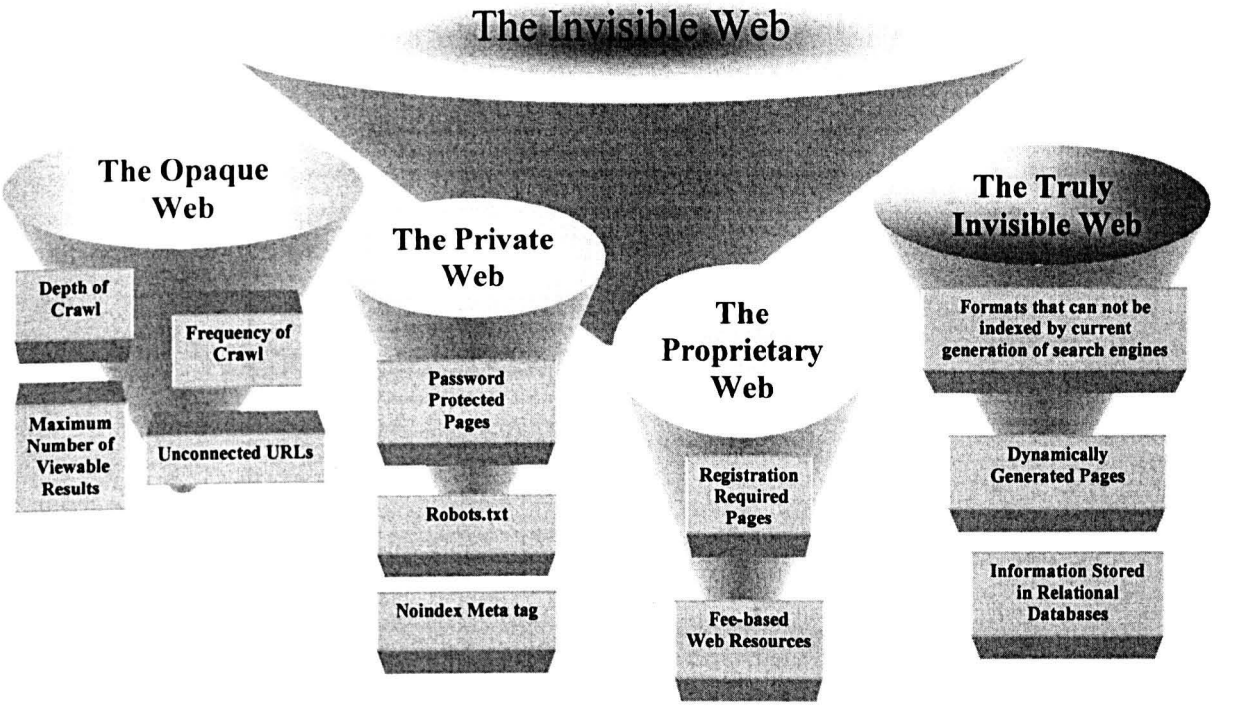

Figure 2.4: An illustration of different parts of the invisible web, based on Sherman and Price (2001)

Another part of the opaque web in Sherman and Price's categorization relates to search engine policies for displaying search results. For example, the query "biological research in the UK" (without quotation mark in the search box) at the time of searching in Google produced 23,500,000 results, of which only 623 were displayed with an option on the final page to display a further 370. Obviously, this is an unsophisticated search and in real situation it is very unlikely that a user would wish to visit the whole 23,500,000 results of this search session. The point is although search engines show a huge number of search results to the user, they would never display all identified documents. The figure 2.5 and 2.6 shows this search procedure.

Another example makes the importance of this issue more tangible. For instance, if a user conducts a more precise search and wants to have a collection of all web resources with the two phrase of "University of Sheffield" and "Department of Information Studies" the search will reveal 36,400 results (figure 2.7 and 2.8). Nevertheless, only 997 documents out of 36,400 hits will be displayed in the last possible search result which even includes omitted results in ordinary searches.

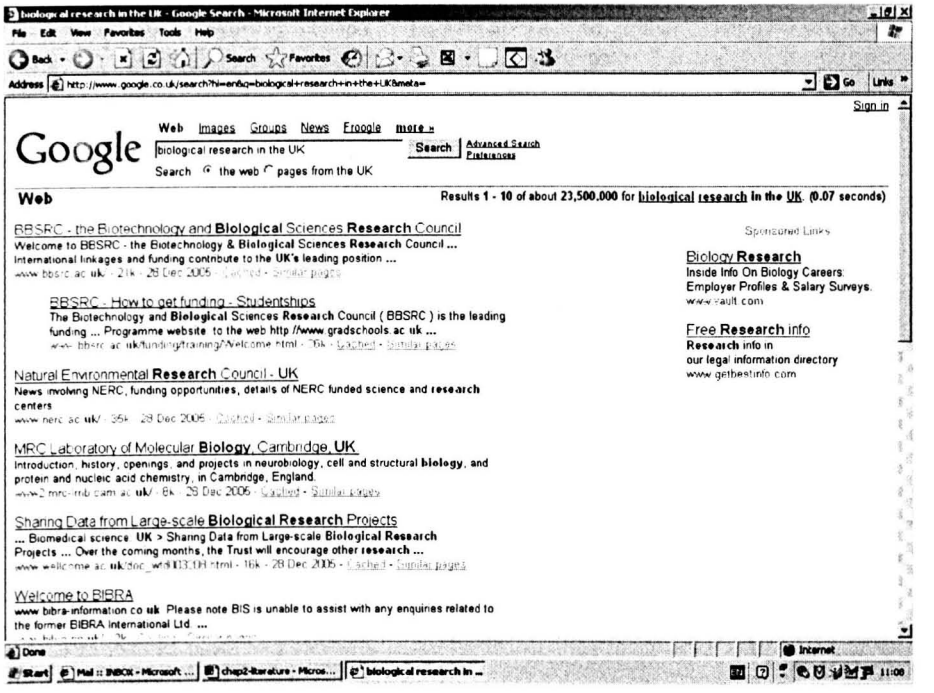

Figure 2.5: An example of a simple search in Google and the number of results

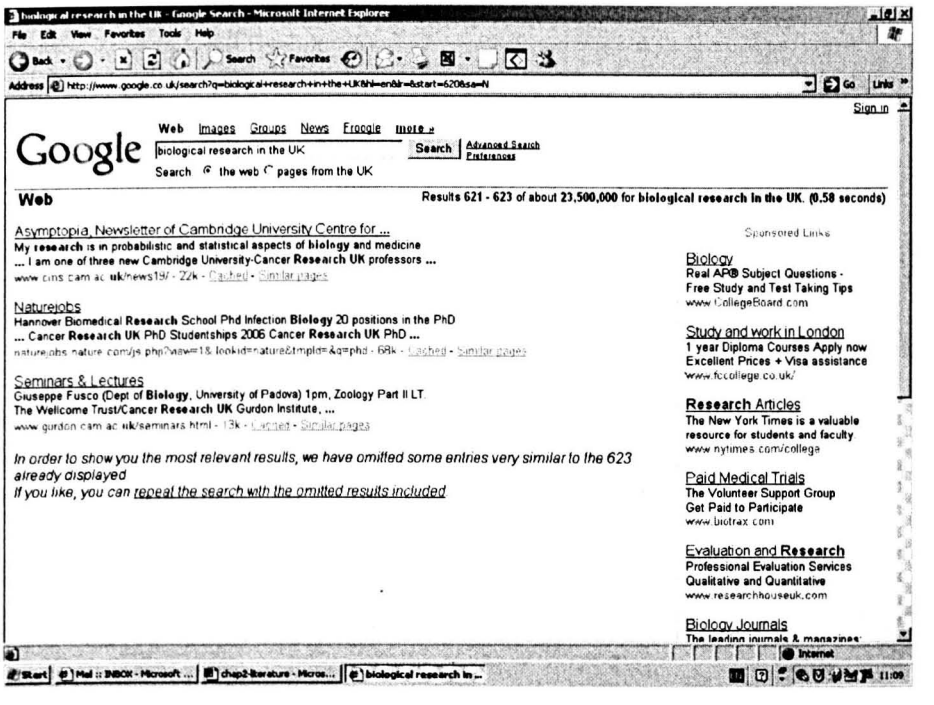

Figure 2.6: Search engines' inability of displaying all retrieved documents

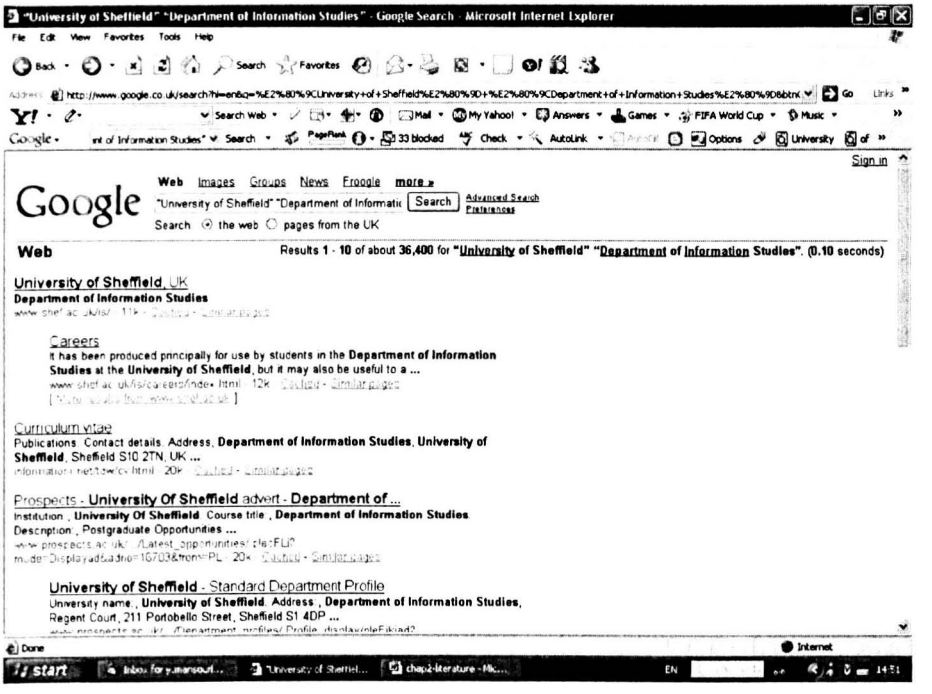

Figure 2.7: Search the web for all documents containing the "University of Sheffield" and "Department of Information Studies"

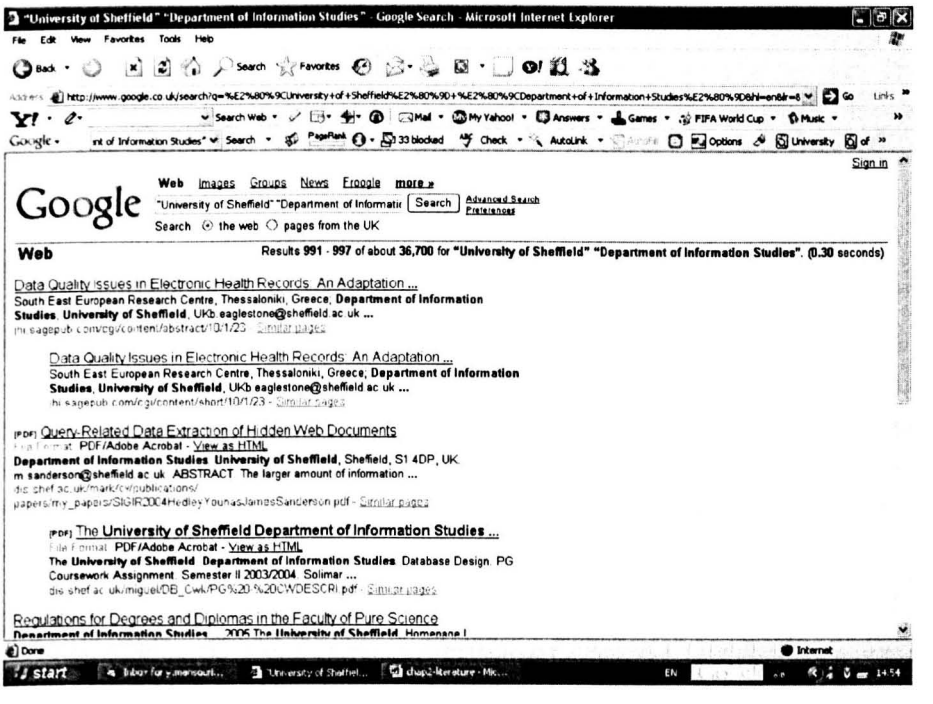

Figure 2.8: Google's last possible display search results.

Previous research (e.g. Jansen et al., 2000) indicated that the majority of search engine queries are short, advanced search facilities including Boolean operators – which might reduce potentially the very large hit lists that might result from such short queries – being seldom used. Another type of invisibility not mentioned by Sherman and Price relates to the fact that the majority of web users appear to view a very limited number of result pages (e.g. Jansen and Pooch, 2001; Spink, 2003; Spink and Jansen 2004). Spink (2003) estimates that searchers view on average only some 2.35 pages; whilst Jansen and Pooch (2001) report that web searchers typically view no more than ten results.

What have been termed disconnected URLs (Sherman and Price, 2003) represent another aspect of the invisible web. These include web pages that have not been submitted to a search engine and do not have any hyperlink pointing to them. Broder *et al.* (2000) estimated that disconnected URLs made up about 20% of the potentially indexable web.

Broken links are another problem, and can result from web pages being deleted or moved. There is usually a delay between moving a page and the updating of search engine indexes. Therefore, even if pages still exist elsewhere on the web they remain in the invisible realm. Block (2004) terms this area of the web the *dead* web.

Private web

The private web represents Sherman and Price's (2001) second type of invisibility, and it refers to resources that technically can be indexed but which are deliberately excluded from search engine crawlers.

These resources include the information which is part of people's private properties or some organizational resources which are accessible only for specific groups of users. There are different techniques to stop search engines of access to these kinds of information. Password protection, the robots.txt file and the noindex Meta-tag are three common methods of preventing access by crawlers.

This part differs from other parts of the invisible web in that these resources are not intended by their owners to be visible to general searchers. Therefore, invisibility of information in the private web is a very different type of invisibility because this part of the web is invisible because it should be for particular reason/reasons. There are many kinds of resources including people's private, personal and financial resources that must be protected on the web and should not be publicly available. Therefore, this part of the invisible web is totally out of this study's scope and it has been introduced here only to show this aspect of the invisibility of information.

Proprietary web

The proprietary web refers to resources that require registration for access. Some of these sites are fee-based, but many are free of charge. However, in both cases, search engine crawlers are not able to satisfy the registration requirements and cannot access these resources. Figure 2.9 shows an example of the proprietary web resources on the web that requires fee to have access to it and figure 2.10 is an example of the proprietary web resources that is free of

charge but requires registration before having access to it. This part of the invisible web also remains unseen for general-purpose search engines and access to its contents requires satisfying the access requirements; either through registration or paying the fees, and then performing separate searches inside them.

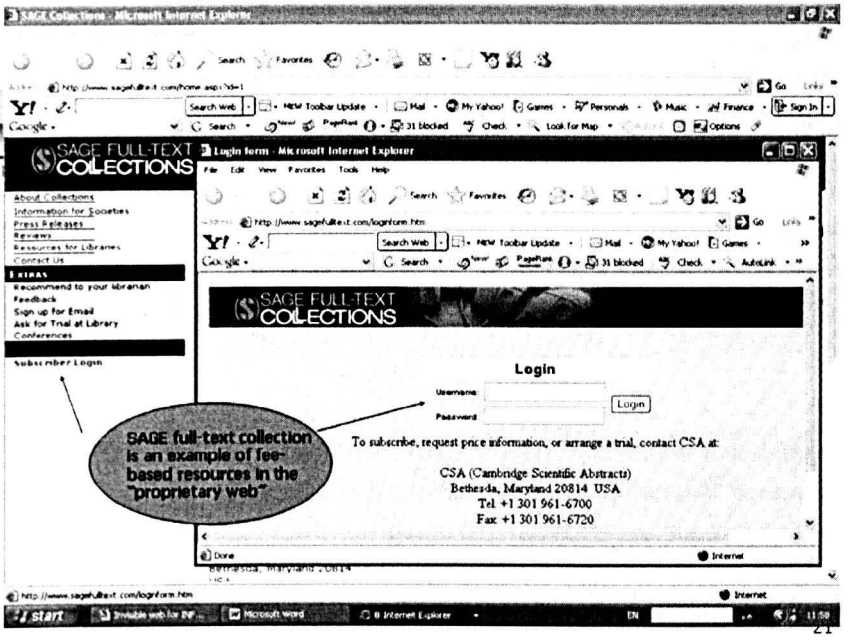

Figure 2.9: Sage full text collection (an example of the fee-based proprietary web)

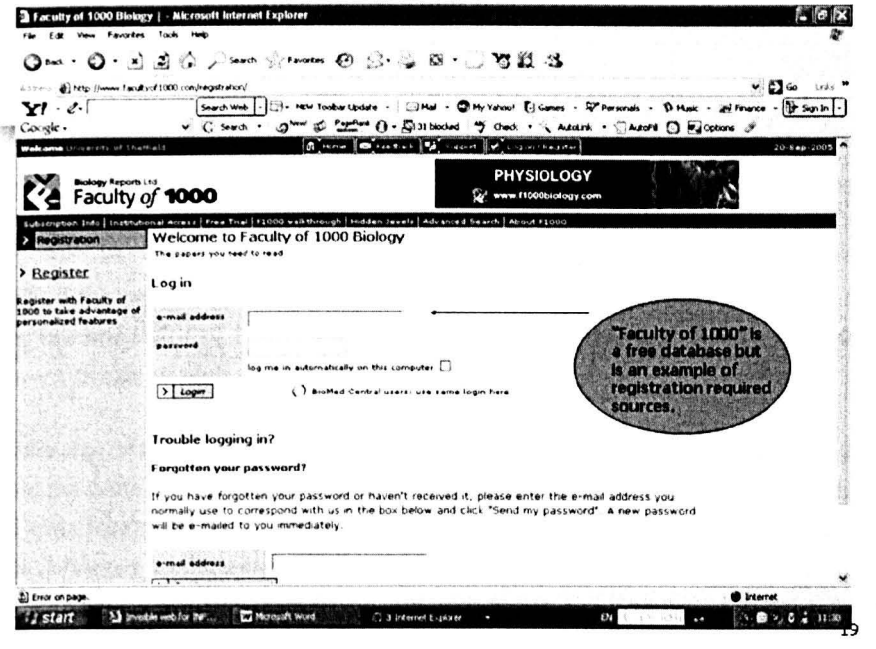

Figure 2.10: Faculty of 1000 (an example free of charge proprietary web resources)

Truly invisible web

Another type of information invisibility is due to the technical limitations of search engines. As Hsieh-Yee (2001) notes, most search engines are mainly geared to indexing static HTML pages. However, such pages represent just a small fraction of the information available on the web. The part of the web which is not in static format and cannot be indexed by general purpose search engines has been termed the truly invisible web. This includes resources that are in formats that can not be indexed by certain general search engines, dynamically generated web pages, and information stored in relational databases. A number of search engine crawlers cannot index resources such as postscript files, flash, audio, and streaming video.

The amount of such resources is substantial, and this can represent a serious limitation of search engines. Increasingly, however, mainstream search engines – including AltaVista, Google and Yahoo – are able to index a variety of other formats.

Deep web

Dynamically generated information is produced on the fly in response to a database query, or via a script. Such dynamically generated information has been called the "deep web" (Bergman, 2001). In terms of quantity the information resources on the deep part of the web should be much bigger than the "surface" content. Bergman's work in 2001 is one of the early attempts to quantify the immeasurable contents of the deep web. Any estimation of the size of the web is rapidly out of date. However, the key findings of his study showed that at the time:

> "Public information on the deep Web is currently 400 to 550 times larger than the commonly defined the World Wide Web. The deep Web, a part of the invisible Web, contains nearly 550 billion individual documents compared to the one billion of the surface Web. More than 200,000 deep Web sites presently exist. Sixty of the largest deep-Web sites collectively contain about 750 terabytes of information sufficient by themselves to exceed the size of the surface Web forty times. ...The deep web is the largest growing category of new information on the Internet. Deep Web sites tend to be narrower, with deeper content, than conventional surface sites. ...More than half of the deep Web content resides in topic-specific databases. A full ninety-five per cent of the deep Web is publicly accessible information, not subject to fees or subscriptions."

Nevertheless, four years after the publication of Bergman's study Lewandowski (2005) without denying the huge size of the deep web criticized the validity of Bergman's method to measure the size of the deep web and reported:

> "Regarding the size of the Invisible Web, it is surely quite smaller than proposed by Bergman in 2001. He said the Invisible Web was 400 to 500 times larger than the Surface Web, but his calculations were based on some of the largest Invisible Web databases, which included sites such as the National Climate Data Center (NOAA) and NASA EOSDIS, both of which are databases of satellite images of the earth. For each picture included, its size in kilobytes was added. As a result, Bergman concludes that the NOAA database contains 30 times more data than Lexis-Nexis, which is a mere textual database. But this says nothing about the amount of information. In conclusion, Bergman's figures seem highly overestimated. Other authors assume that the Invisible Web is 20 to 50 times larger than the Surface Web."

Lewandowski's (2005) criticism of measuring the size of the invisible web is an example of challenges to assess the quantity of the web and the invisible web. Sherman (2001) Nevertheless, normalising for both data type and format yields a more reasonable estimate that puts the size of the invisible web at between two and 50 times larger than the visible web. No matter how you measure it, the invisible web is much larger than the visible web.

Information seeking on the web

Regarding Glaser's (1992:31-37) categorization of the literature in Grounded Theory studies which has been illustrated in the early parts of this chapter, information seeking on the web forms the second parts of the literature review in this study. Information seeking on the web is part of a wider area of information seeking in online resources and information seeking in online resources is part of a general area of information seeking. Figure 2.11 shows three areas of research about information seeking.

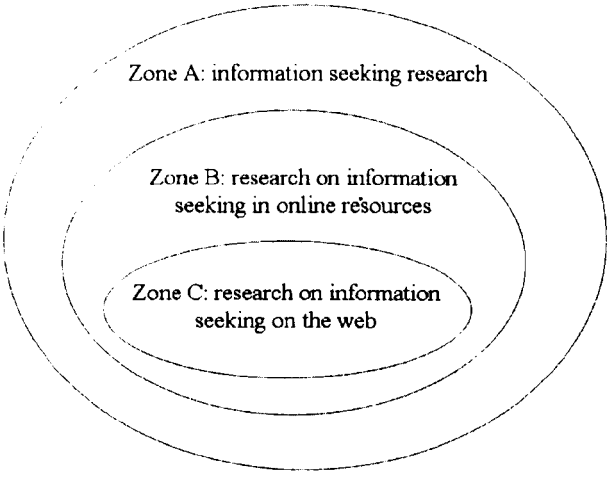

Figure 2.11: Information seeking on the web and wider areas

This chapter does not cover zone A and B and exclusively focus on zone C because not only it is virtually impossible to review the whole area of information seeking in this chapter but also it is beyond the scope of this study. Therefore, zone A and B has been totally excluded in this thesis. However, two useful and comprehensive review books can be recommended here if the reader wants to learn more about zone A and B. The first book is written by Donald Case in 2002. Hultgren (2003) reviewed the book (Case, 2002) and declared:

> "The book opens by briefly examining the concepts, issues and research that are pertinent to the topic of information behaviour and that have been developing over the last decades. The introduction contains a series of descriptive examples that put the student in the research picture of information seeking, needs and uses. ... On the whole, this well-written and perceptively analytical reference book cannot, in my opinion be overlooked".

The other recommended sources is "Theories of Information Behaviour" (Fisher, et al., 2005) which works as an encyclopaedia of theories and models in information behaviour research. Jamali (2005) reviewed this book and noted as follows:

> "This enlightening book presents an authoritative overview of many conceptual frameworks that help understand and interpret people's information behaviour including their information need, information seeking, and information use activities and so forth".

This book (Fisher et al., 2005) consists of 72 brief chapters about existing theories in human information behaviour. Among these 72 chapters only two chapters are related to information seeking on the web. Detlor (2005) contributed to the book by presenting a brief description about web information behaviours of organizational workers. Moreover, Turnbull (2005) contributed to the book by developing a brief description about information seeking on the web which is mainly based on his previous joint works on this issue (e.g. Choo et al., 2000).

Research on information seeking on the web has commenced in 1995. Catledge and Pitkow (1995) did the first study of in this new area (Choo et al., 2000; Kalbach, 2003). For example, Kalbach (2003) reported that:

> "Catledge and Pitkow (1995) were the first to publish a scientific study of Web browsing behaviour. A chief finding of this research is that Web users rely on a limited number of pages within a site. The study clearly shows a recurring pattern of frequently returning to given page as a sort of "home base." The researchers describe this as a "hub and spoke" style of navigation through a web space".

Tauscher and Greenberg (1997b) did the other early study about users' behaviour on the web. They concluded that there is a 58 percent probability that the next page visited was previously visited. Thus, they suggested web browsing happens in a recurrent way. In addition, they found out that users visit very few web pages frequently and many web pages are only visited once. The few frequently accessed pages tend to fall into certain categories. Besides, they reported that although many pages are revisited, users constantly add new pages to their repertoire. As Choo et al. (2000) stated the study conducted by Tauscher and Greenberg (1997b) was about users' behaviour on the web and they had identified seven web browsing patterns as follows:

> "First-time visits to a cluster of pages; revisits to pages; page authoring (where the subject used Reload to view the newly modified page); use of Web-based applications; hub-and-spoke visits (navigating to each new page from around a central page); a guided tour where links guide navigation through the Web pages; and a depth-first search where link paths are followed without returning to the first page in some cases."

Huberman et al. (1998) explored users' web surfing behaviour through mathematical laws. They identified some kinds of similarities among web user' surfing patterns, and developed a mathematical "law of surfing". The law of surfing calculates the probability distribution of the depth of web surfing which means the number of page that a user visits within each single web site. Huberman et al. (1998) identified different variables in their study including "cost of continuing surfing". They concluded that when the cost of moving to the next page is more than its expected value, the user stops surfing.

Choo et al. (1998, 1999, and 2000) investigated the information seeking behaviour of users on the web and developed the first model of information seeking on the web. Choo et al. (2000) concluded that there are a variety modes of information seeking behaviour on the web which varies from undirected viewing that does not pursue a specific information need, to formal searching that retrieves focused information for action or decision making. They identified that each mode of information seeking on the web is distinguished by the nature of information needs, information seeking tactics, and the purpose of information use. Moreover, they made a distinction between motivations and moves in behavioural frameworks in information seeking on the web. Motivations relate to the strategies and reasons for viewing and searching but moves link to the tactics used to find and use information.

One of the outcomes of research in the area of information seeking on the web was identifying new perspectives in human-computer interaction on the web. Researcher in this area noticed that web searching is a complicated process and possesses many aspects. For instance, Kalbach (2000) stated:

> "Traditional paradigms of information retrieval tend to over-simplify the information seeking process. Robertson (1977) presents models in which the seeker simply enters a query and is given matching results. Although this accurately describes part of the process, such models fail to take into consideration actual user needs, behaviours, situations and gaps in knowledge."

Furthermore, Kalbach (2003) sated information seeking process on the web can be viewed from a very narrow perspective to very broader situations:

> "An ISP [information searching process] can be viewed on different scales; from a very narrow perspective, for a single search session, for example, or for broader search situations over time".

In general, searching the web is an interactive procedure between end users and web-based search tools to achieve a goal. A successful search only happens when the required information exists on the web, the employed search tool is able to locate it and the user employs the search tool properly. As a result, all these three abovementioned elements should work properly to conduct a search successfully. When all elements work appropriately access to information can be very easy and straightforward. Nevertheless, if one or more elements do not work properly users are not able to locate what they are looking for.

We can focus on any of these elements to explore what is actually happening in the web search process. Some studies are concerned about the user's aspects (e.g. Spink et al., 1998; Choo et al. 1999; Jansen et al., 2000; Ross & Wolfram, 2000; Kim, 2000; Kim, 2001; Cockburn & McKenzie, 2001; Ford et al. 2002; Spink, 2002; Ford et al. 2003). The second group focuses on the search tools (e.g. Peterson, 1997; Gordon and Pathak, 1999; Vaughan and Thelwall, 2004) and the third group of studies are concerned about the web-based resources and presentation types (e.g. Zhang and Dimitroff, 2005a; Zhang, Dimitroff, 2005b). Figure 2.12 shows a very general perspective of these three major strands of research in this area.

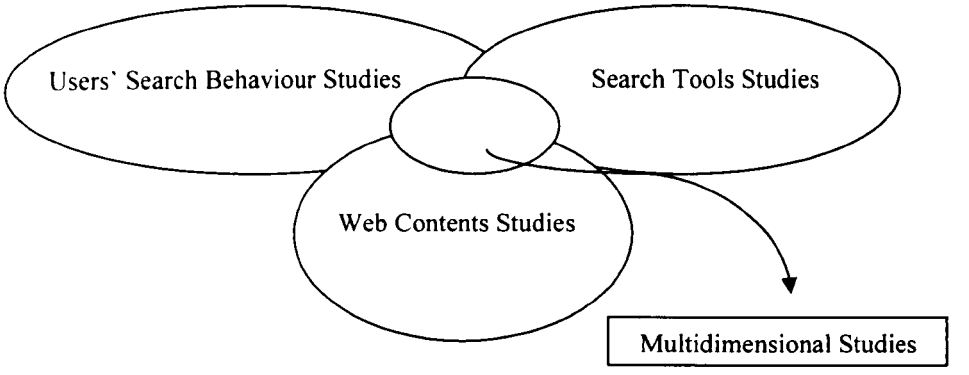

Figure 2.12: Three major part of the web search research body

The current study locates in the first group and just focuses on the user aspect of this procedure. User-oriented web search research is an area of investigation which includes studies that are concerned about users' web search process. This type of research has a relatively short history, but has covered many aspects of information seeking on the web and has exhibited impressive vigour in research design (Hsieh-Yee, 2001). Probably the reason of this progress in web search research is owing to the unique features of web searching which makes it distinctive from the pre-web online searching.

Unique features of web searching

Web searching is becoming an indispensable part of the daily life of various groups of people. Web usage is not limited to a few specific purposes and its functionalities are spreading across different aspects of life. Among all web-based applications, web searching is one of the most common ones. Information retrieval through the web is a multifaceted process and many side elements may affect on this procedure. Nevertheless, our current knowledge about the complexity of this course of actions is limited.

Although developing web-based search tools has been inspired by pre-web online information retrieval systems, some distinctive features of the web environment and the wide range of web users' characteristics impose new condition for searching and retrieval situation. In general, not only resources but also users in the pre-web systems were fairly homogenous and predictable. In contrast, both web users and web resources are so heterogeneous and diverse. Pre-web electronic resources including online and offline databases, Online Public Access Catalogues (OPACs) contain mainly structured data which has been stored in very well-organised systems. Each document has a specific structure with a number of fields and subfields. This structured environment make the storage and retrieval procedure much easier and more predictable. Accordingly, information resources in the classic pre-web online information systems were well organized and homogenous. On the other hand, the users were

limited to some specific groups of the society mostly the academics and researchers, librarians or subject expert people.

In summary, four main reasons in the recent years are involved in the creating new search environment. The first one is the uncontrolled expansion of electronic information in a very uneven, heterogeneous and unstructured environment. Because of the easiness and publicly accessibility of web-based publication facilities every moment new pages appear on the web without any control on the contents, structure and quality of the information. Consequently, the web is a highly dynamic and uncontrolled environment. This situation considerably impacts on the procedure of storage, searching and retrieval of information on the web.

The second event is the lack of a logical concurrence between the advance of computer technology in terms of technical sophistication and human computer interaction. Previous investigations about the information seeking and retrieval were mainly focused on the technical aspects and the users' side has been neglected in some ways.

The third issue is the rising number of inexperienced users. If previously computer-based equipment was mainly employed by professional groups, now the majority of the public need to have interaction with different computer-based implements. Previous research indicates that unfamiliarity with search tactics creates difficulties for many users of online information retrieval systems (Haverkamp & Gauch, 1998). As White & Iivonen (2001) concluded in their report the majority of web users are not very sophisticated searchers: preferring known sites, browsing, using only simple searches if they use a search engine, substituting words but not necessarily using other tactics to modify a search, being very trusting about information on the web, and sometimes being relatively non-discriminating in recognizing relevant information.

The fourth issue is related to the level of understanding about the users' features and needs while they are searching the web. In fact, one of the most important problems goes back to this reality that in spite of the importance of the issue of web searching, with a user-centred viewpoint, our existing knowledge about the information seeking on the web is still inadequate. Accordingly, search tool designers need to learn more about the real user's information seeking behaviours to develop their products based on their needs. As Cothey (2002) has mentioned:

> "We need to understand the phenomenon of web searching more fully to make information provision more effective."

Categorizing the research body

In order to achieve an overall picture about the publications in the area of "information seeking on the web" the researcher carried out some searches in related databases including LISA. As the required keywords in this stage were common words including "searching", "browsing", "the web" and "information seeking" the researcher had to define a specific search strategy to minimise the possibility of retrieving non-relevant documents and at the same time cover the most relevant part of the literature body.

Therefore, he initially carried out a simple but precise search in LISA which was followed by more sophisticated searches. The first search query was designed to look for the exact search

topic on the title. Thus, a query with TI = ("information seeking on the web") from the earliest time to 2003[14] produced four articles (Choo and Marton, 2003; Kari and Savolainen, 2003; Kim, 2001; Choo et al., 2000) which were all relevant[15]. In the next stage, the researcher identified those LISA descriptors which were common in these four highly relevant documents to use for an advanced search. In fact, by identifying the descriptors that have been allocated to these first four relevant articles he carried out another broad search to retrieve resources that are in the same subject but did not use the exact words "information seeking on the web" in their titles. Therefore, in the second search a combination of two main descriptor DE = ("world wide web") and DE = ("information seeking behaviour") were used to conduct the search which yielded 46 results.

Other descriptors of these four highly relevant documents (e.g. online information retrieval, user surveys, women, and information professionals) were not used in the advanced search because employing these descriptors produced less relevant and irrelevant results.

After reviewing the 46 documents produced in the advanced search the researcher noticed that the research body in this field can be categorised according to different perspectives. Six major kinds of criteria have been considered to categorize the studies. According to this viewpoint the studies in the arena of the web searching can be classified based on the following criteria:

- **Research approach: technical or user-oriented**
- **Research methodology: quantitative or qualitative studies**
- **Research population: public or specific targeted populations**
- **Research situation: experimental research or naturalistic studies**
- **Research extent/scope: specific websites or the whole web**

It is essential to emphasize that the above classification does not imply clear boundaries between different studies in this field. In fact, each study may possess a combination of these categories. At the end of this categorization the features of the current study is illustrated according to these criteria to illuminate the main characteristics of the current study and its linkage to the literature body.

Based on research approach

The literature review indicated that the studies in web search are dominated by two major strands which are user-oriented and technical-oriented. The importance of user-centred approach in IR and Information seeking studies has been declared in previous literature even earlier than emergence of web. Saracevic et al., (1988) explicitly anticipated that the main

[14] The beginning year of the study

[15] The same search has been carried out at the end of the study with a broader time span (from the earliest time to 2006) and the result was the same.

concern and future efforts of information retrieval research will be devoted to enhance our knowledge about human involvement with information rather than developing more sophisticated facilities. Due to universality of the web the importance of the user-oriented approach seems more needed. User-oriented research attempts to illustrate a clear picture of current situation and based on the findings, offer new solutions to help people in their search process to find what they need more quickly, accurately, and with less effort. This demand is becoming crucial, particularly by the extraordinary growth of data on the web and increasing number of users (Pollock and Hockley, 1997). The literature review also demonstrated that the number of user-oriented studies is increasing steadily (e.g. Spink et al., 1998; Ross & Wolfram, 2000; Cockburn & McKenzie, 2001; Spink, 2002 and Jansen et al., 2000).

Based on employed methodology

Regarding the various scopes and purposes in different web-based studies, a variety of research methodologies have been implemented in these studies up to now. Lacking of consolidated methodologies which can most effectively be used to study on the web environment was one of the main challenges in the area. As Martzoukou (2004) clearly states:

> "The existing limitations in method and the plethora of different approaches allow little progress and fewer comparisons across studies. There is urgent need for establishing a theoretical framework on which future studies can be based so that information seeking behaviour can be more holistically understood, and results can be generalised."

In a general viewpoint, the methodology approaches in this field, similar to other fields, can be divided into two main groups of quantitative and qualitative investigations.

Quantitative studies

Quantitative studies depend on quantitative data, usually in a large scale and utilising statistical analysis with a positivist paradigm. Due to the existence of huge amount of data, web searching is a suitable opportunity to carry out quantitative studies. In terms of the interaction between the web and web users, web transaction logs analysis is a common way in such studies. In users studies a wide range of users provide researchers with an apposite possibility to collect a huge quantity of data ready to process and investigation. The significant size and diversity of web users and web-based resources is a valuable prospect in quantitative research. There are several quantitative studies in this field (e.g. Spink et al., 1998; Silverstein et al., 1999; Jansen et al., 2000; Cockburn & McKenzie, 2001). As Martzoukou (2004) reported:

> "To date the most popular method of examining users' web-interactions has been through quantitative analyses of large sets of user-system transactional data taken from real online searches on the web."

Search sessions, queries, and terms are three key variables in quantitative studies about web searching. A session is the entire sequence of queries entered by a searcher. Sessions are composed of queries and each query is one or a string of search terms entered into a web search system. A term is defined as a string of characters separated by some delimiter such as a space, a colon, or a period (Jansen & Pooch, 2001). Reviewing recent research indicates that query

analysis is one of the major approaches in many studies. In fact query analysis is an important method to comprehend web users' behaviour (Ozmutlu et al., 2002). Everyday enormous number of queries is submitted to search engines and analyzing web search engine query logs is problematic due to the large amount of the available data (Ozmutlu et al. 2002). However, still web transaction log analysis is one of the major methods in current web-based information research (e.g. Silverstein et al., 1999; and Ozmutlu et al. 2002).

Although a quantitative study increases our understanding about the users search behaviors, it is not sufficient. For example, as Spink et al. (1998) mentioned user relevance judgments are central to both the systems and user-oriented approaches. Therefore a reliable study about search tools on the web should consider this important point. However, quantitative data can not satisfactorily support our deductions concerning issues like relevance judgment. Martzoukou (2004) argued:

> "Although web information searching studies can offer an informative insight into web searching activity they do not allow for a more in-depth investigation of individual users. The full extent of the user's role is not elaborated because they are either based on anonymous samples of web users or the involvement of users is minimal. Thus this type of research is not able to offer comprehensive interpretations of elements that are significant to the study of information seeking such as different information needs cognitive and affective characteristics and experience of individual users."

Therefore, we simply need to ask people about their feelings and ideas regarding search facilities performance. Gathering these kinds of data is possible by other methodologies such as qualitative research approaches which are discussed in following section.

Qualitative studies

As mentioned earlier although quantitative studies provide us with valuable understanding about different aspects of web-based search, the growing body of research in this field seeks something more than statistical results. Despite the widely accepted importance of the user in all kinds of information systems, our understanding about human involvement with web-based search systems is still limited. Filling this gap in research body is possible by carrying out more qualitative investigations. Qualitative studies which mostly are carried out in the survey format, attempt to clarify the observations and collected data within interpretive analysis with a holistic approach. Martzoukou (2004) identified the movement towards to qualitative mode of inquiry in the area of information seeking on the web and reported:

> "Research in Web information seeking has turned to the application of qualitative enquiry methods which are considered as more appropriate 'where in-depth understanding of human actions is the primary focus' (Mellon 1990: 20) concentrating in more focused groups of individuals (e.g., students, children and IT professionals) and within a wide variety of situations (e.g., health science, education, everyday life and work-settings). Case studies, ethnography, grounded theory and Dervin's (1992) sense-making theory have been the most commonly used qualitative approaches (Hill 1997; Nahl 1998; Pejtersen and Fidel 1998; Fidel *et al.* 1999; Brown 1999; Choo and Marton 2003; Rieh 2003)."

As a summary, the availability of large scale collections of transaction log data has led to the predominance of quantitative research in the area. Not only it is possible to analyse these data with less effort than would be required for qualitative data, the behaviour of large numbers of searchers can be studied. Qualitative studies, by contrast, require a great deal of effort to analyse the behaviour of just a few users. However, such a situation has led to a perception of web searching behaviour that is based on a superficial analysis. Quantitative studies alone can give us little insight into a user's interaction with the web. While undoubtedly useful, such studies are not sufficient. Martzoukou (2004) is one of a growing number of commentators to call for more in-depth research on cognitive and affective aspects of searching to enhance the findings from quantitative studies.

Some studies combine both qualitative and quantitative methods. Bilal (2000, 2001, 2002b), for example, in studying the search behaviour of school children, used recorded screen activity as a source of quantitative data, and supplemented it with interview data. Similarly, interviews are often supplemented by observation of the searchers' behaviour and actions (e.g. Fidel et al, 1999, Slone, 2002), or by 'talk-through' protocols, in which searchers are asked to explain their actions (e.g., Madden et al, 2006).

Based on data collection method

In the web-based information seeking investigations data collection process is associated with the web environment and its users. However, one of the major difficulties in studying web users is the lack of methods for collecting real-world data. Wang et al. (2000:236) stated that:

> "...searching Web resources is a highly interactive cognitive process that cannot be understood simply by comparing search outcomes with users' questions. Interview or survey data can only provide a partial picture of the interaction because users may not be aware of or able to recollect what they did during the process."

Data gathering tools in technical oriented studies are different with user-oriented projects. Additionally, within user-oriented studies data gathering tools are slightly different in quantitative and qualitative studies. Reviewing the previous research indicates that several techniques have been utilized to collect required data. The major methods include questionnaire, interview, observation, web logs transaction, and experiments.

In the user-oriented studies users, their needs, beliefs, and behaviours play an important role and it is necessary that data gathering tool suites to the study's purposes. Reviewing the literature indicates that there is a variety in data collection ways (Hsieh-Yee, 2001). Common data gathering tools in the web-based information seeking research are as follows:

Transaction Log Analysis

Transaction logs are a common method of capturing characteristics of user interactions with IR systems. Quantitative studies in web search research are relied heavily on transaction log analysis, which is defined by Griffiths et al., (2002) with reference to Peters et al, (1993) as:

> "...the study of electronically recorded interactions between on-line information retrieval systems and the persons who search for the information found in those systems."

Given the current nature of the web, transaction logs appear to be the most reasonable means of collecting user-searching information from a large number of users. Transaction log analysis uses transaction logs to distinguish attributes of the search process, such as the searcher's actions, the interaction between the user and the system, and the evaluation of results by the searcher (Jansen & Pooch, 2001). However, there are some advantages and disadvantages for assessing web transaction log methods.

On the positive side, this method provides the researcher with a large amount of data from the user population. As web transaction log is anonymous and can be gathered easily, studies based on this method is supported by a reasonable sample of whole population. On the negative side, the researcher can not communicate with the user to ask them about their ideas and judgments of their searches. Indeed, by this method there is no possibility to trace the users' attitudes. For example, when the web transaction log shows the majority of users did not follow the search results pages of search engines except the first or second page, the researcher does not know whether they could satisfy their information needs in the first or second pages or there was no relevant information and they stopped the search frustratingly. Therefore, while web transaction log is an efficient way to collect a large amount of data, it is not very reliable for further judgments. In summary, transaction log is an efficient tool for gathering quantitative data but it should be supported by qualitative data. Particularly, due to this widely accepted opinion that user relevance judgments are central to both system and user-oriented studies (e.g. Spink et al. 1998), it is important to explore users' attitude by some other methods.

The other problem of log analysis is its inefficiency to provide researchers with a comprehensive picture of using the web. For example, we do not know whether end users find the retrieved pages useful, print the visited pages, carry out any search inside the pages or not. Similarly, Choo et al. (2000) reported that difficulty in collecting complete sets of data to describe web browsing sessions is a source of difficulties for researchers in this area.

Questionnaire, interview & observation

Questionnaire is being widely used in the various web-based information seeking investigations (e.g. Lazonder et al., 2000; Spink, 2002). The questionnaire may be in printed or electronic format. The electronic questionnaire may be distributed by e-mail or locating in an interactive web-site (e.g. Spink et al., 1998). The electronic questionnaire is cheaper, easier to handle, time saving and arguably with more respond rate and respond speed (Roselle & Neufeld, 1998). However, in the cases that researchers need deeper understanding from their target group, interview is a more appropriate instrument. As in some user-oriented studies deep exploration of users' behaviours and their beliefs is important, researchers use interview to collect the data (e.g. Fidel et al, 1999). However, transcribing of interviews is a time consuming task and interpretations of users' expressions needs specific skills. Therefore, this method is used usually in qualitative and interpretative studies which focus on limited target populations.

Another method of data collection in web search studies is observation (e.g. Fidel et al, 1999). In this method the researcher seeks to follow the search behaviour through looking at a real search process performing by the user.

Based on research population

In terms of research population there are two types of studies. Sometimes, the target group is a specific population of web users, for example, children (e.g. Schacter et al, 1998; Bilal, 2005, 2004, 2002a, 2002b, 2001 and 2000), experts of an academic field, a group of professionals, or students (e.g. Fidel et al, 1999; Kim, 2000; Hoelscher & Strube, 2000; Lazonder et al., 2000; Ford et al., 2002; Ford, et al., 2003).

In contrast, in some studies, research population is not confined to a specific category and the samples are selected among the public of the web users (e.g. Ozmutlu et al., 2002 and Spink & Xu, 2000). In these studies researchers have to be optimistic that the sample reflects the diversity and enormity of the whole population of web users reasonably well.

Based on research situation

In terms of the reality or artificiality of research situation the web-based research can be divided into two groups of naturalistic and experimental studies. Naturalistic studies are carried out in a real situation and based on genuine needs of searchers (e.g. Fidel et al, 1999; Spink & Xu, 2000; Ross & Wolfram, 2000; Jansen et al., 2000; Ozmutlu et al., 2002). In contrast, experimental studies adopt a controlled laboratory-based design (e.g. Kim, 2000; Lazonder et al., 2000; Kim, 2001, Ford et al., 2005a, 2005b, 2003). In the first group, researchers seek to explore how end users approach to the web and carry out their searches. Exploring the real search process provides the researcher with an opportunity to investigate the user's information seeking behaviour, the quality and quantities of search results and users' attitude of search facilities. In the second group, searchers are asked to search the web and locate an appropriate answer to one or a number of specific search tasks defined by the researcher. Martzoukou, (2004) highlights some examples of previous studies in this category:

> "Another popular method has been the design of Web surveys of information searching practices and preferences of users. These however typically deal with hypothetical information searching situations rather than empirical investigations (Spink 1998; Hsieh-Yee 1998; White and Iivonen 1999; Vaughan 1999; Griffiths and Brothy 2002)."

Each type of the above categories is valuable in its place and they can be considered as complementary. Exploring users' behaviour in real situation is a chance to gain access to a genuine picture of what is really happening on the web everyday. Nevertheless, research in controlled environment is useful to investigate the details of end users' interaction with the web more thoroughly and focus on only one or a few aspects of their search process.

Based on research extent/scope

The review of the literature indicated that the scope of some studies have been limited to given websites and their users. Particularly, search engines form a common group in this category and many studies have been carried out about these search tools (e.g. Spink et al., 1998; Gordon & Pathak, 1999; Spink & Xu, 2000). However, in some other studies instead of considering a given website, the entire web is being considered as the source of information (e.g. Ford et al. 2002; Ford et al. 2003).

Similar to the search situation category, here in research scope each type has its benefits. Focusing on a specific website is useful to gain a profound insight on the users' interaction with that given website and the results will be valuable to improve the usability of the website. On the other hand, in the studies which are not limited to specific websites the research enjoys a much broader context which enables the researcher to generalize the findings to a broader perspective. Regarding the categories that are introduced in this chapter the current study has the following features which are summarized in table 2.1.

Table 2.1: The summary of the research main features

Research aspect	This study's feature
Research approach	User-oriented
Research method	Qualitative
Research population	Specific targeted group
Research situation	Real search situation
Research extent/scope	Whole of the web

Influential factors on web search procedure

As mentioned earlier many factors potentially affect web searching process. In fact, success or faultier of a search session stems from the overall effect of many variables. Through reviewing the literature these effective elements can be categorised into three groups:

Factors related to the users

In any information retrieval system users are the central part of the system. This pivotal point has been declared in almost all information retrieval studies. For instance, Chowdhury (1999:179) states that:

> "The user is the focal point of all information systems because the sole objective of any information storage and retrieval system is to transfer information from the source to the user".

Therefore, the user is not only a major part of any information system, but also is the most important one. Consequently, his/her features, needs and abilities have direct effect on the success or failure of search. Recently, the user side of web-based information seeking has received considerable attention and the number of user-centred investigation which are focused on users' characteristics is increasing (e.g. Jansen et al. 2000; Jansen and Spink, 2000; Wang et al. 2000; Palmquist & Kim, 2000; Palmquist, 2001; Kim, 2001, Cothey, 2002; Hargittai, 2002, Ford et al., 2005a, 2005b, 2003, 2002; Spink, 2002; Ford et al. 2003).

One of the most important elements which can affect on the search process and search results of web searchers is their searching experiences rate. Previous investigations revealed that people with more experience on web searching usually use a range of tools and facilities and would be more successful on their search (Hsieh-Yee, 1993; Palmquist and Kim, 2000; Slone, 2002). One of the controversial issues in the studies concerning the effect of experience on web search behaviour is how to measure people's experience and distinguish between a novice and expert searcher. Cothey (2002) stated that in previous studies generally experience has been substituted for expertise and it has been assumed that more experienced users are more expert in their field which is not always the case.

Subject familiarity is another issue which might make a difference in the search results. Based on the rate of subject familiarity the search process and results would be different for web searchers. Presumably, it should be easier for web searcher to locate an information item related to their area of interest or their profession because they can select more specific keywords and they are more familiar with special resources.

The experience level of people for working with computer can affect on the search process on the web. There are many facilities embedded in web browsers that users can facilitate their searching process by using them. However, we do not know what percentage of web users might utilise these facilities. Moreover, according to the aim of a search people might adopt diverse search approaches. Previous investigations suggested that people who seek information for job-related or educational purposes are highly motivated and they are more persistent on their information seeking procedure. However, those who are looking for recreational information are not very enthused and usually rely on serendipity (e.g. Slone, 2002).

Factors related to the search tool used

The second group of influential factors on web search process and its results goes back to the features and capabilities of the employed search tools. Undoubtedly, looking for information in the heterogeneous and huge environment of the web require capable search facilities. There is a considerable diversity in existing web-based search engines. General-purpose search engines including Google, AltaVista, and Yahoo are some examples of the most popular ones. However, there are some other kinds of search facilities including meta-search engines, directories, and subject gateways. Nevertheless, general purpose search engines are typical tools for information retrieval through the web.

Since the emergence and popularity of search engines many investigations have been conducted to evaluate the different features of these information retrieval tools (e.g. Schwartz, 1998; Gordon and Pathak, 1999; Jansen, 2000). These studies typically attempted to uncover the effectiveness of search engines in providing the web users with their information needs. Each search engines has its ranking result policy. This is very important that more relevant items appear on the top of retrieval results. Although the issue of relevance is a complex topic, search engine designers attempt to rank the most relevant items on the top of the list. The previous studies indicate that web users usually look at the first and second results page and simply do not visit the remaining (Spink, 2003). Therefore, ranking policy of a search engine can affect on search process. Obviously, the issue of relevance is very complex but out of the

scope of this research. There are many useful resources that discuss different aspects of relevance (e.g. Saracevic, 1975; Schamber, 1994; Anderson, 2005; and Harter, 1992).

Generally, search engines provide web users with different search facilities usually known as simple or basic search and advanced search. The prior studies indicate that the majority of searchers usually just use the basic or simple search facilities and their search queries contain two or three words which are similar to natural language (Keily, 1997; Silverstein et al. 1999; Jansen, 2000; Jansen et al. 2000). In the investigation conducted by Keily (1997) 12%, in Hoelscher's study (1998) just 3% and in Jansen et al. (2000) only 8.5% of posted queries contained Boolean operators.

Factors related to the search query

The third group of elements which is important on the web search procedure is related to characteristics of the query which is submitted to the search systems. As White and Iivonen (2001: 738) stated:

> "...questions are factors too important to be excluded from the research design in Web search research", the type of search query plays a vital role in information seeking procedure and search results"

White and Iivonen (2001: 721) examined web users' choice of initial search strategy and the role of questions on their decision making and reported:

> "...the participants not only indicated a fairly high degree of familiarity with the initial search options and used different search strategies but also were influenced in their choice of an initial search strategy by question-related characteristics. Of the two question characteristics in the study, the most influential is the predictable/unpredictable source of the answer".

Closeness or openness of search query may be influential in selecting search procedure and potentially will affect on search process. In definition of closed and opened questions White and Iivonen (2001: 723) stated:

> "When questions are closed, exact answers are wanted. Often these questions elicit brief, factual information. Searchers have little discretion in judging correct answers or choosing alternatives. Instead they may face many options about where to find correct answers. When questions are open, there is no one exact answer and searchers must develop acceptable responses."

Some other features of search query including clarity, complexity and familiarity, fact-based questions and subject questions, predictable sources and unpredictable sources, have been investigated in the prior studies (e.g. Jansen, 2000). Complexity refers to multiple aspects and number of concepts, familiarity, popularity and currency of topic indicates that topic or question is well-known or not. Therefore, characteristics of search query can potentially be influential on search process and consequently on search results.

Success and failure of search

Regarding the above explanations it can be concluded that a web search procedure will only be successful if the following essential factors are in place.

- Documents relevant to the needs of the searcher must exist on the web;
- The search tool employed must be able to locate and retrieve them;
- The user must be able to employ the search tool competently;
- The user must recognize the relevance of the retrieved documents with his/her need.

Only the second of these factors relates exclusively to technology. All the others relate either to the user, or to the user's interaction with technology. As mentioned earlier, the most important component of any web search system is the user. S/he provides the motivation for initiating and ending a search. Perhaps the most significant negative effect of the dominance of quantitative research in the area has been an overemphasis on the technological component of search systems. However, the gradual increase in qualitative research is helping to shift the focus of the researcher in the area to the interaction between user and technology.

Users' interaction with web-based search tools has different aspects including their conceptualization of success and failure in search. The issue of success and failure of search has not received sufficient attention in the literature. We still do not know enough about when and how web users feel successful and when they feel failed and what factors might affect their feelings. Besides, this issue has not been investigated in pre-web information retrieval systems. Ondrusek (2004: 243) in her exhaustive review identified 34 studies for "failure analysis" and 31 studies for success over the 33 years of research from 1977 to 2000. Nevertheless, Ondrusek (2004: 223) declared that she excluded the web-based studies in her review. She mentioned:

> "Research on Internet user behaviour, as it relates to searching web sites via publicly available search engines, was not included in this study. Rather, the selection process identified articles representing online searching behaviour as it applies to information retrieval from systems using indexed records and/or controlled vocabulary. Imposing this restriction provided continuity to the analysis, especially in classifying research variables."

Although Ondrusek (2004) totally excluded the web-based search research in her review, it is still worthwhile to mention it here because web-based studies about web users' information seeking behaviour were derived from the pre-web studies. In fact, these two strands of research in many ways are intertwined. Moreover, her review confirms that even before the emergence of the web the number of studies which specifically focused on the issue of success and failure in search procedure can be attributed to only twenty percent of the literature body[16]. In terms of

[16] What this thesis reports about the meaning of success/failure in search is different with what had been perceived in the literature about this issue. This study possesses a unique and holistic viewpoint on this issue which seeks users' views on success and failure which had not been considered in this way in the past.

the portion of studies which were about search failure in the pre-web online searches Ondrusek (2004:243) reported:

> "Thirty-four studies (20.9%) included examinations that singled out end-user search problems, and two forms of analysis were prevalent, identifying errors and isolating failed searches. Of these two methods, the more straightforward one was error analysis, or counting distinct errors in every search produced ... A broader term used to refer to problematic searches was search failure, a term indicating situations where the search did not lead the end-user to the desired materials. To measure search failure, analysts applied a technique sometimes called "failure analysis". Indicators of search failure included zero hits, multiple errors, incorrect or inadequate strategies, wrong database submissions, exceedingly low recall, and excessive postings."

Regarding the studies that considered success Ondrusek (2004: 244) classified different situations that imply success for end users:

> "In contrast, 31 studies (19.0%) examined end-user performances for signs of success, a performance measure encompassing many more facets than those embodied in the failure analysis techniques. In some studies, researchers used the same indicators employed in failure analysis but looked for "desirable" outcomes such as searches that produced hits or found items as opposed to analyzing no-hit searches. A number of investigators gauged success by scoring search outcomes such as speed, relevant document retrieval, precision, recall and combined recall, relevance, and precision."

Ondrusek's (2004) review can support the originality of the current study in terms of the way it addressed users' conceptualizations of success and failure in search. Her review shows during thirty three years of research, from 1977 to 2000, concepts of success and failure have been addressed mainly from an objective and system-oriented perspective, but the current study addressed these issues from a user-oriented and subjective perspective.

Theoretical frameworks

The last part of the literature review covers the conceptual frameworks which became relevant to the study after having the results of inductive analysis of the dataset. In fact, as this research was exploratory the researcher did not know where the research might end up and what sorts of theoretical frameworks might be useful to fortify the emergent results. Therefore, he had to follow the procedure of theory development to address the research questions first of all based on inductive analysis of the data before adopting any pre-conceived theories.

However, after emergence of the theory, presented in chapters six, seven, eight and nine, it was a suitable time to compare the results with some well-established theories to strengthen the emergent theory. These theories were identified mainly based on serendipity, long time interaction with the literature, reflections on the emergent results and comments from fellow researchers[17]. In fact, identification and implementation of these theories was an important part

[17] For example, the researcher learned about Simon's Bounded Rationality and Satisficing theory through a comment by an anonymous referee on one of his work-in-progress papers.

of the exploratory nature of the study. There are four theoretical frameworks which have been used in this study including Rotter's locus of control, Heider's attribution theory, Bandura's self-efficacy and Simons's Bounded Rationality & Satisficing. A brief review of these theories is presented here and the detailed explanations about the applicability of these theoretical frameworks have been illustrated in tenth and eleventh chapters.

Locus of Control and Attribution Theory

The attribution of causes to perceived success and failure has for long been studied in psychology. The long-established construct "Locus of Control" (Rotter, 1966), for example, maps people's perceptions of the extent to which they feel "in control" of events that surround them, individuals with a relatively external locus of control tending to feel that events are largely outside their control. There is research evidence (Mamlin, Harris, & Case, 2001) that:

- Internal locus of control is more associated with males than females;
- locus of control becomes more internal with age; and
- more senior individuals within organizations tend to be more internal.

Locus of control is an important element in Attribution Theory. Initially purposed by Heider (1958), and subsequently modified by other theorists including Kelley (1973) and Weiner (1979, 1986), this theory provides a conceptual framework for explaining how people ascribe causes to events. According to the theory, when people succeed or fail in performing a task there are four possible types of attributions to which they might ascribe their success or failure. These can be classified into two basic categories: internal and external. "Internal" or "dispositional" attributions include factors related to the person's ability to perform a task, or the amount of effort that s/he makes. "External" or "situational" attributions consist of factors that are out of the person's control such as luck, or the level of task difficulty. According to Attribution Theory, confident high achievers tend to attribute their success to internal factors (e.g. ability or effort) and failure to external factors (e.g. bad luck), whereas less confident under-achievers tend to attribute their success to external factors and failure to internal factors. The high achiever will tend to work hard since success is believed to be caused by effort, the low achiever tending to work with little effort since success is perceived to be related to external causes (Chinn, 2002; Weiner, 1986).

Attribution Theory has not been widely studied in relation to information seeking, but there is a considerable body of work in the related area of "self-efficacy". Self-efficacy relates closely to key elements of Attribution Theory, in that high levels of self-efficacy entail self-confidence in being able to produce success in tasks via one's own ability and effort.

Self-efficacy

As noted by Kurbanoglu, (2003), the concept of self-efficacy is an important component of "social learning theory" and has been developed in the social psychology mainly by Bandura (Bandura, 1977, 1986, 1994, 1995, and 1997). Low self-efficacy has been linked to anxiety and low levels of experience in relation to poor performance generally (Bandura, 1977 and 1986). These findings are supported by studies into computer anxiety (e.g. Meier, 1985) and

experience (e.g. McInerney et al., 1994). A useful overview is provided by Pajares (2002). Within an information seeking context, Savolainen (2002: 211) defines self-efficacy as:

> "A person's judgement of his or her ability to organize and execute action, such as finding information on the Web"

Indeed, self-efficacy has been investigated in a number of studies of information seeking (e.g. Ford et al, 2001, Nahl, 1996, Eastin and LaRose, 2000, Ren, 2000, Thompson et al. 2002, Kurbanoglu, 2003), and is conceived as one of the activating mechanisms of information-seeking behaviour in Wilson's (1997) model.

Nahl (1993; 1995; 1996) found evidence that individuals displaying high levels of self-efficacy outperformed those with lower levels in relation to a range of factors including search efficiency, success and satisfaction. A study by Ford and Miller (1996) found links between gender and a range of perceptions of the Internet relating to self-efficacy. Females reported being unable to find their way effectively around the Internet, getting lost, not being in control, and only looking at things suggested to them. All of these features were linked in a further study by Ford, Miller and Moss (2001) with retrieval failure. In the latter study, links were also found between poor retrieval effectiveness and fear of failure and poor time management.

Ren (2000) reports a study in which links were found between self-efficacy in information seeking and search performance. It was also found that self-efficacy could be improved through training. Jawahar and Elango (2001) found links between both self-efficacy and the setting of specific task goals and more general end user performance measured in terms of specificity and relevance. Rimal (2001) reported a significant relationship between self-efficacy, risk assessment and use of health information. Thompson et al. (2002) found links between Internet self-efficacy and the number of items correctly retrieved in an experiment in which participants were required to search to find names of industrial organizational psychologists. They also found that persistence in searching was greater when searchers were given specific task instructions as opposed to more general instructions.

Bounded Rationality and Satisficing

Simon's behavioural decision-making theory of bounded rationality and satisficing has been originated in human psychology and has been employed in many fields including web searching. Agosto (2002) reported the results of a study, apparently a PhD research, about applicability of this theory to gain a better understanding of a group of web users. Bounded rationality theory basically suggests that in real situations people are only "rational enough" or "partly rational" rather than being "absolutely rational". Agosto (2002:16), citing Simon (1955, 1956) and notes that:

> "Simon ... rejected the idea of optimal choice, proposing the theory of bounded rationality. He argued that due to time constraints and cognitive limitations, it is not possible for humans to consider all existing decision outcomes and then make fully reasoned, purely rational, choices. He suggested that humans operate rationally within practical boundaries or within the limits of bounded rationality".

Agosto (2002:17) notes that the concept of simplification forms the fundamental principle of Simon's theory:

> "The idea of simplification methods underlies all of Simon's work in human cognition. The most important simplification mechanism is "satisficing", or choosing decision outcomes that are good enough to suit decision makers' purposes, but that are not necessarily optimal outcomes. Satisficing involves "setting an acceptable level or aspiration level as a final criterion and simply taking the first acceptable [option]" (Newell & Simon, [1972], p. 681)."

Gigerenzer and Goldstein (1996:651) describe the meaning of "satisficing as a concept which covers two concepts of "sufficiency" and "satisfaction":

> "Satisficing, a blend of sufficing and satisfying, is a word of Scottish origin, which Simon uses to characterize algorithms that successfully deal with conditions of limited time, knowledge, or computational capacities".

About the implementation of bounded rationality theory in library and information science research Agosto (2002:17) reported that there are a few studies that have been employed this theory as a theoretical framework:

> "A few library and information science studies have considered the theories of bounded rationality and/or satisficing. Schwartz (1989) suggested that collection developers act within the limits of bounded rationality when selecting library materials. He proposed a descriptive model of collection development that combined bounded rationality, tacit knowledge, and symbolic content. In a related application of bounded rationality to library practice, Chu (1994) suggested that reference librarians' and academic library patrons' time and cognitive limitations often result in ambiguous reference questions and answers."

Summary

This chapter consists of three parts. The first part reviews the literature about the invisible web and describes the existing definitions of this phenomenon which refers to part of the web which is not accessible by general-purpose search engines. Moreover, different parts of the invisible web including the opaque web, the private web, the proprietary web and the truly invisible web are explained in detail. The second part reviews the area of information seeking on the web which is an area of research to explore web users' behaviour while searching the web and their interaction with web-based search tools. The chapter categorises the literature body in information seeking behaviour based on different criteria including research approach, methodology, research population, research situation and research scope.

The third part of the chapter reviews four theoretical frameworks which became related to the study at the final stages of the study and were employed by the researcher to reanalyse the dataset deductively. These frameworks are Rotter's locus of control, Heider's attribution theory, Bandura's self-efficacy and Simons's Bounded Rationality & Satisficing. This chapter briefly introduces these theoretical frameworks and the way in which the searcher has employed them to consolidate the results of the study are reported in chapters ten and eleven.

CHAPTER THREE:

RESEARCH METHODS

CHAPTER 3: RESEARCH METHODS

"Remember the two benefits of failure. First, if you do fail, you learn what doesn't work; and second, the failure gives you the opportunity to try a new approach." Roger von Oech

Introduction

This chapter describes the methodological issues of the current research including the research design, research approach, research population, data collection, data analysis, and detailed explanation on motivations and reasons of employing all these procedures and approaches. The chapter begins with a brief description about quantitative and qualitative research methods and the differences between these two approaches. Moreover, further explanation is presented about features of qualitative research in general and Grounded Theory in particular. The chapter illustrates why Grounded Theory has been employed as the research method in this project and how the researcher put this methodology in practice.

Moreover, the chapter reports the preliminary study which has been carried out before the main study. Performing the preliminary study was very useful to formulate the primary research questions and to develop the research methods. In addition, the chapter addresses the credibility issue of research methods including the validity and reliability of data gathering and interpretation methods. The aim of the chapter is to explain why this study is an exploratory research with a user-oriented and holistic approach and how the researcher conducted the research regarding Grounded Theory to investigate the phenomenon under the investigation systematically and address the research question thoroughly.

Qualitative vs. quantitative research methods

Success in conducting any research project requires employing a plausible research method to make sure that the outcome of the project is credible and trustworthy. In fact, the appropriateness of the employed research method provides the researcher with a dependable tool to carry out the research plausibly. Myers (1997) defines a research method as follows:

"A research method is a strategy of inquiry which moves from the underlying philosophical assumptions to research design and data collection. The choice of research method influences the way in which the researcher collects data. Specific research methods also imply different skills, assumptions and research practices."

Qualitative and quantitative research methods are two major approaches of research in different fields of studies including different areas in social science, management, marketing and library and information science. Although both approaches have much in common, there are a number of fundamental differences that make them distinguishable of each other. In fact, studies with a quantitative approach deal with numbers and figures (i.e., quantities) while studies with a qualitative approach deal with descriptions of concepts and perceptions. The results obtained from the two approaches are very different. Quantitative studies seek findings that can be used to make generalisations across the field of research. Although qualitative research can lead to generalisations, its prime function is to help interpret phenomena within

'real-life' contexts. Bauer et al. (2000:7) considered four criteria including data, analysis, prototype and quality to differentiate between qualitative and quantitative data. These criteria have been summarized in table 3.1.

Table 3.1: Differences between quantitative and qualitative research based on Bauer et al. (2000)

	Strategy	
	Quantitative	**Qualitative**
Data	Numbers	Texts
Analysis	Statistics	Interpretation
Prototype	Opinion polling	Depth interviewing
Quality	Hard	Soft

There is no superiority between these two approaches and each one is useful in its place. Moreover, qualitative and quantitative research can be complementary for each other. For example, a single research project might have different qualitative and quantitative phases and the overall outcome is a result of combining of both qualitative and quantitative data. To conduct a valid quantitative study we require a sample which should be statistically the representative of the whole research population.

However, in qualitative research the sample does not need to satisfy this requirement because in qualitative research we focus on a small number of subjects with more profound analysis. Sample size and the use of statistics are not the only difference between qualitative and quantitative research.

A review of methodological textbooks shows that there are more fundamental dissimilarities between these two types of research. Cook and Reichardt (1979) discussed the differences and their analysis results were summarized by Ratcliff (2004) in table 3.2. As table 3.2 indicates, qualitative studies usually focus on less well-defined phenomena. They are often exploratory in nature; and seek to describe rather than to explain. Many of the differences between qualitative and quantitative research can be understood by thinking in terms of the relationship between the researcher and those being researched. Qualitative research aims to identify the chief concerns of them, and to record what they consider to be the chief issues.

Therefore, quantitative research attempts to identify patterns and then look to see how well those being researched fit those patterns. This leads to a complementarity in which qualitative research generates hypotheses which can be subsequently tested through quantitative studies.

Table 3.2: Differences between qualitative and quantitative research based on Cook and Reichardt (1979)

Qualitative research	Quantitative research
Phenomenological	Positivist
Inductive	Hypothetical/deductive
Holistic	Particularistic
Subjective/insider centred	Objective/outsider centred
Process oriented	Outcome oriented
Anthropological worldview	Natural science worldview
Relative lack of control	Attempt to control variables
Goal: understand actor's view	Goal: find facts and causes
Dynamic reality assumed;	Static reality assumed;
"slice of life"	relative constancy in life
Discovery oriented	Verification oriented
Explanatory	Confirmatory

Each approach is therefore useful in its place, and a case can frequently be made for using just qualitative or just quantitative methods. Ford et al (2005a:744) for example, in their study of the impact of individual differences and web search strategies, justified the use of solely quantitative research:

> "This was a quantitative study in which all variables were reduced to numbers and analyzed statistically. The study adopted a controlled laboratory-based experimental design. The limitations of adopting such an approach in isolation, i.e., not complementing it with qualitative data and analysis are acknowledged."

Often however, a single research project will have both qualitative and quantitative phases, with the results generated being a combination of both qualitative and quantitative data. Alternatively, quantitative studies may be followed up with qualitative studies, or vice versa. There are some examples of this kind of combination in the literature. The case for a combination of approaches was made by Ford et al. (2005b: 762), who concluded their quantitative study with a recommendation that qualitative and quantitative approaches should be used in tandem:

> "The research adopted a positivist quantitative methodology, the limitations of which are well known. It would be profitable for future studies collectively to provide, as well as further statistical investigation of the relationships reported here, complementary interpretative viewpoints based on interviews and think aloud protocols."

In an ongoing research programme, Ford has managed to achieve this with studies of the web search behaviour of school children (Madden et al, 2006).

Since this study had a qualitative approach, the following section describes the qualitative approach in more detail and describes the applicability of this approach in library and information science studies.

Qualitative research in more detail

There are many descriptions about the qualitative research which are more similar to each other rather than different. Therefore, a few definitions are presented here just as some examples of existing definitions. Glaser (1992:11), one of the two originators of Grounded Theory, defines the qualitative approach as follows:

> "Qualitative analysis means any kind of analysis that produces findings or concepts and hypotheses, as in grounded theory, that are not arrived at by statistical methods".

Jones (1995) also put emphasis on the discovery aspect of the qualitative research and describes this line of enquire as follows:

> "Qualitative research begins by accepting that there is a range of different ways of making sense of the world and is concerned with discovering the meanings seen by those who are being researched and with understanding their view of the world rather than that of the researchers."

Myers (1997) highlighted the main differences of qualitative research with quantitative studies and defined the qualitative research as follows:

> "The motivation for doing qualitative research, as opposed to quantitative research, comes from the observation that, if there is one thing which distinguishes humans from the natural world, it is our ability to talk! Qualitative research methods are designed to help researchers understand people and the social and cultural contexts within which they live ... Qualitative research involves the use of qualitative data, such as interviews, documents, and participant observation data, to understand and explain social phenomena."

Reviewing the literature shows that qualitative research techniques have been successfully employed in many disciplines including library and information science (LIS). Wilson (2000) declares four main reasons why qualitative research methods are suitable LIS research in general and for information seeking studies in particular:

> "Qualitative research seems particularly appropriate to the study of the needs underlying information-seeking behaviour because: our concern is with uncovering the facts of the everyday life of the people we are studying; by uncovering those facts we aim to understand the needs that exist which press the individual towards information-seeking behaviour; by better understanding of those needs we are able to understand what meaning information has in the everyday life of the people; and by all of the foregoing we should have a better understanding of the user, be better able to design more effective information services, and be better able to create useful theory of information-seeking behaviour and information use."

There are different types of qualitative research methods including action research, case study research, ethnography and Grounded Theory. This study employed Ground Theory. There are specific reasons for this decision which are elucidated in the following section. However, before explaining the reasons it seems useful to describe what Ground Theory is.

Grounded Theory

In this section a brief description of the research method used in this study, Grounded Theory, is presented. This description includes the history of this method, its main features which make it different with other methodologies, two versions of Ground Theory and details of the practical procedure in this method.

Grounded Theory is a general, inductive and interpretive research method which was originated in 1967 by Barney Glaser and Anselm Strauss (Glaser and Strauss, 1967). However, it does not mean that the Grounded Theory is not connected to the previous research methods and has been emerged out of the blue. Heath and Cowley (2004: 142) cited Hammersley, (1989) and reported:

> "Grounded theory's roots lie in symbolic interactionism, which itself stems from pragmatist ideas of James, Dewey, Cooley and Mead (Hammersley, 1989), most notably the concept of the looking glass self (Cooley, 1922) ... the term 'symbolic interactionism' was invented by (Blumer (1937) and his development of the interactionist approach together with naturalistic inquiry is a key influence on grounded theory."

Glaser (1992:2) defined Grounded Theory as follows:

> "Grounded theory is based on the systematic generating of theory from data, that itself is systematically obtained from social research."

Similarly, Strauss and Corbin (1994: 273) described Grounded Theory as follows:

> "Grounded theory is a general methodology for developing theory that is grounded in data systematically gathered and analyzed. Theory evolves during actual research, and it does this through continuous interplay between analysis and data collection".

In fact, in Grounded Theory researchers do not test or verify any preconceived hypothesis. In contrary, they try to develop new theory based on the systematically collected evidence. Instead of having hypotheses to test, researchers in Grounded Theory studies have research questions to address. Just like the current study the research begins with some unanswered questions and researcher attempts to find out the answer of the questions without any bias in his/her mind. In fact, in Grounded Theory researcher should keep his/her mind open to any possible evidence that might exist in the dataset. The theory will emerge through the data analysis sooner or later. However, emergence of the theory is not a quick process and requires time and patience. As Glaser (1978:18) declared:

> "Generating grounded theory takes time. It is above all a delayed action phenomenon. Little increment in coding, analysing and collecting data cook and mature then to blossom later into theoretical memos".

Again in Glaser words, Glaser (1992:8), Grounded Theory is inductive generation of theory through a qualitative analysis of qualitative and/or quantitative data. The point is he makes a difference between qualitative analysis and qualitative data. In fact, qualitative analysis can be carried out by quantitative data as well. To define the meaning of the qualitative analysis Glaser (1992:11) believed:

> "Qualitative analysis means any kind of analysis that produces findings or concepts and hypotheses, as in grounded theory, that are not arrived at by statistical methods".

Although Grounded Theory has been developed and principally used within the field of sociology, it can be, and has been successfully employed by people in a variety of different disciplines including information science. Glaser and Strauss do not regard the procedures of Grounded Theory as discipline specific, and they encourage researchers to use the procedures for their own disciplinary purposes.The general goal of Grounded Theory research is to construct theories in order to understand phenomena. Haig (1995) believes that a good Grounded Theory should possess three main features that is inductively derived from data, subjected to theoretical elaboration, and judged adequate to its domain with respect to a number of evaluative criteria. The unique feature of Grounded Theory is simultaneity of data collection and data analysis because the collected data leads the researcher to make his/her decision for the next step. Although Grounded Theory is an inductive process it can be deductive at some stages as well as Case (2002: 165) reports:

> "Few investigators or studies stick solely to induction or deduction. Rather, they tend to move back and forth between those modes: collecting information that allows them to state a principle or tendency, then testing that generalization through further research, in an endless chain of logic."

Powell (1999: 97) explains about the nature of studies which employ Grounded Theory as the methodology:

> "Studies that seek to inductively and systematically develop taxonomies and theories through intensive analysis and coding descriptive data collected about the phenomenon under investigation theories emerge through iterative, constant comparison of concepts and categories against data and said to be grounded in given naturalistic setting being investigated."

Glaserian or Straussian approach

Over the last 39 years, from 1967 to 2006, since Glaser and Strauss originated the Grounded Theory this methodology has been constantly developed through enormous number of publications and discussions about the real nature of the Grounded Theory. Nevertheless, not everyone has perceived the main meaning of the Grounded Theory in the same way. Differences about the real nature of Grounded Theory began first of all among the originators of the Grounded Theory, Glaser and Strauss.

Glaser (1978, 1992) accused Strauss that he has not fully understood the meaning of Ground Theory and what he has written in his books (Strauss, 1987; Strauss and Corbin, 1990) is not Grounded Theory but full conceptual description. However, Glaser's criticisms about Strauss's new approach in Grounded Theory do not imply that Glaser nullify and dismiss Strauss's approach. Glaser in his writings (Glaser, 1978, 1992) always mentions that Strauss's approach is valuable in its place but it is not Grounded Theory anymore.

In fact, the later divergence in Grounded Theory and disagreements between the originators is not about the ontological and epistemological aspects of the Grounded Theory but it mainly comes from their differences about the details of methodological procedures like how to code the data and how develop the categories. Differences between Glaserian and Straussian approaches were summarized by Heath and Cowley's (2004) in figure 3.1 and 3.2.

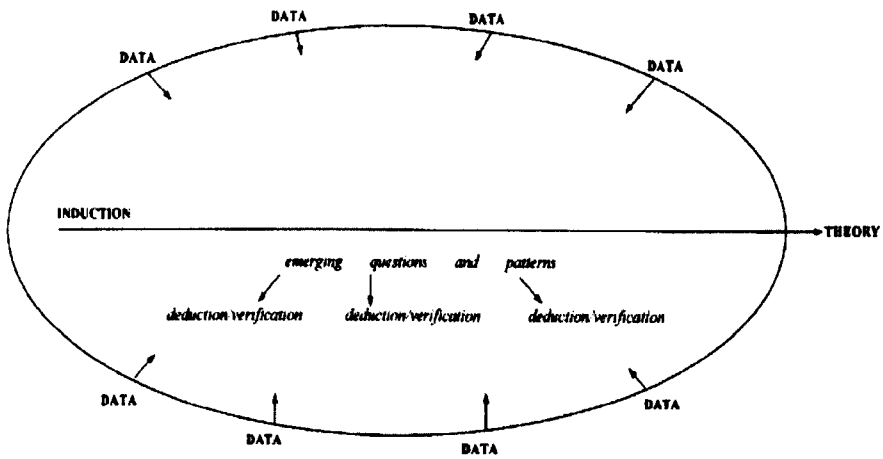

Figure 3.1: Summary of Glaser's approach of induction, deduction and verification in Grounded Theory adopted from Heath and Cowley's (2004)

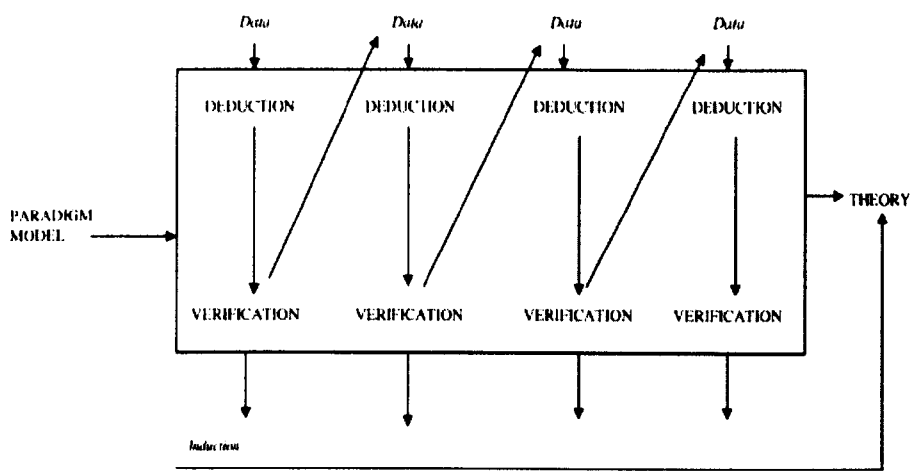

Figure 3.2: Summary of Strauss's approach of induction, deduction and verification in Grounded Theory adopted from Heath and Cowley's (2004)

Heath and Cowley (2004: 141) reviewed two approaches of Glaser and Strauss in Grounded Theory and concluded that:

> "A conclusion is drawn that, rather than debate relative merits of the two approaches, suggests that novice researchers need to select the method that best suits their cognitive style and develop analytic skills through doing research."

In fact, Straussian version of Grounded Theory claims that Glaserian version put too much emphasis on the induction nature of Grounded Theory. Nevertheless, Heath and Cowley (2004) cited Glaser (1998) and sated that:

> "Researchers should stop talking about grounded theory and get on with doing it, which seems like good advice. The novice researcher should set aside 'doing it right' anxiety, adhere to the principle of constant comparison, theoretical sampling and emergence and discover which approach helps them best to achieve the balance between interpretation and data that produces a grounded theory ... it is wise to remember, too, that the aim is not to discover the theory, but a theory that aids understanding and action in the area under investigation".

However, these debates have been constructive to develop this research method over the years. Probably, one of the main reasons of development of Grounded Theory comes back to all these arguments. On the other hands, these arguments did not stop researchers in many research areas of employing Grounded Theory and a considerable number of research projects have successfully accomplished by utilizing this method.

Although the researcher in this study reviewed most of the main resources about Grounded Theory (e.g. Glaser and Strauss, 1967; Glaser, 1978, 1992, 1998, Strauss, 1987; and Strauss and Corbin, 1990, 1994, 1998) he decided to use the first original book (Glaser and Strauss, 1967) as the main reference of Grounded Theory in the study. This decision helped him to minimise further ambiguity and confusion during the course of data collection and analysis.

Grounded Theory in information science

As has been declared earlier, Grounded Theory is a general research methodology which can be applied in different areas of study including library and information science (LIS). Glaser (1992: 18) states:

> "Grounded theory can be used successfully by people in many disciplines, it is a general methodology. What counts are that grounded theory methods are not bound by either discipline or data collection."

Powell's (1999) seminal paper about methodologies in LIS research supports the employment of Grounded Theory in LIS research. Powell (1999: 103) believes:

> "The fields of library and information science have no shortage of research questions and phenomena needing thorough exploration and continue to need more well-founded theories, so there is certainly a need for more grounded theory research."

There is more support in the literature about employing Grounded Theory in LIS. For example, Allan (2003: 9) advocates using Grounded Theory in LIS research and declares:

> "In conclusion, the Grounded Theory method is recommended as a powerful way to collect and analyse data and draw meaningful conclusions. This recommendation applies to any researcher in the hard sciences as well as the social sciences."

The history of employing Grounded Theory in LIS comes back to the early 80s. Since that time some seminal works in the information seeking studied have used Grounded Theory. Reviewing the literature shows that LIS researchers in Sheffield were among the first research groups who employed Grounded Theory in their studies (Selden, 2005).

Soto (1992:2) reported that Ellis (1987) was the first researcher who adopted Grounded Theory in information studies. Certainly, Ellis' works in Grounded Theory is among the first studies that adopted this methodology in LIS research. However, it seems the history of Grounded Theory in LIS goes back to a few years before Ellis' (1987) study. Selden (2005:120) cited Wilson (1980:7) and reported:

> "There have been advocates of GT within LIS from various corners ... it seems to have had its greatest stronghold in Sheffield ... the field for GT is prepared by the introduction in 1980 of the journal Social Science Information Studies. In its first issue studies using qualitative methods and having a phenomenological foundation were invited."

Therefore, to be more precise and regarding the literature review it can be said that Ellis (1987) was the first research in information seeking behaviour studies which successfully employed Grounded Theory and has been followed up by many further research including Ellis (1993). Similarly, Beaulieu (2003: 242) refers Ellis (1993) and reports:

> "Ellis turned to Glaser and Strauss's grounded theory approach, a well established inductive method for generating theories and models in the social sciences. The aim was to derive a more accurate model of information-seeking patterns from empirical data. The application of the approach is clearly described by Ellis in a paper on the application of grounded theory."

After successful use of Grounded Theory in LIS research it has received a great attention form researchers in different places. Selden (2005: 120) cites Mellon (1986) and Weingand (1993) as two early examples of Grounded Theory employment on another side of the Atlantic. Up to now many research projects have employed Grounded Theory (e.g. Pickard, 1998; Mellon, 1986; Soto, 1992; Ellis, 1993; Correia & Wilson, 1997; and Lucas, 1999). Beaulieu (2003: 242) declares:

> "The method has proven to be a reliable and fruitful way of handling data analysis and has contributed to establishing the validity of qualitative methods for user-based studies".

In particular many PhD studies around the world have been used Grounded Theory (e.g. Pace, 2003; Lehmann, 2001). As Selden (2005) reported Grounded Theory has received considerable attention from the Department of Information Studies at the University of

Sheffield and notable number of PhD studies adopted this method (e.g. Vedi, 1986; Brown, 1990; Soto, 1990; Musoke, 2001; Zafeiriou, 2001; Lucas, 2001; Zakaria, 2002; Kim, 2004).

Although Grounded Theory has been employed in mainly qualitative and user-oriented studies in LIS domain, there are other fields in LIS that have used Grounded Theory. For example, Star (1998: 222) compared Grounded Theory and Facet classification and found out some common grounds between these two and declares:

> "Both grounded theorists and designers of faceted classifications struggle with a common core problem. This is the question of how to represent vernacular words and processes. In both cases, the categories are empirically discovered in an almost self-contradictory fashion. The contradiction comes with the attempt simultaneously to represent, on the one hand, the local, specific, and empirical and on the other, abstractions and generalizations.

Star (1998:222) describes the common challenges and main features that make Grounded Theory and facet classification similar to each other:

> "The difficulty lies in making this representation both ethnographically faithful (faithful to the needs of users and particular populations), yet simultaneously powerful beyond the single instance or case study. Both grounded theory and faceted classification began as reform movements against powerfully entrenched a priori schemes with claims on universality."

Grounded Theory's process

The question that needs to be answered here is how we can employ Grounded Theory in practice. Charmaz (2000: 515) explains:

> "How do we do grounded theory? Analysis begins early. We grounded theorists code our emerging data as we collect it. Through coding we start to define and categorise our data. In grounded theory coding, we creates codes as we study our data ... we should interact with our data and pose questions to them while coding them ... coding starts the chain of theory development."

Similarly, Pace (2004: 337) describes four main stages of developing Grounded Theory and reports:

> "As with much qualitative research, data collection and data analysis occurred simultaneously in this study. A theory was derived from the data using a constant comparative method of analysis with four stages: generating categories and their properties; integrating categories and their properties; delimiting the theory; and writing the theory"

Constant comparative analysis

Constant comparative analysis is the heart of the process of emergence of the theory in Grounded Theory research. The importance of constant comparative analysis has been reported in many publications about Grounded Theory. For example, Cutcliffe (2000: 1477) cited Glaser and Strauss (1967) and Strauss and Corbin (1994) to highlight the important role of this issue and reported as follows:

> "A central feature of grounded theory is its method of constant comparative analysis (Glaser & Strauss 1967), in that data collection and analysis occur simultaneously and each item of data is compared with every other item of data. The theory induced is conceptually dense (Strauss & Corbin 1994), that is theory with many conceptual relationships, and these relationships are embedded in a context of descriptive and conceptual writing."

Moreover, Glaser (1992:14) reported constant comparison, saturation and core relevance are three important aspects of Grounded Theory which help researchers to minimise the possibility of bias in data analysis. Constant comparative analysis described in Glaser and Strauss (1967: 101-115) has been introduced as one of the main approaches to scrutinize the dataset in Grounded Theory research method. As Glaser and Strauss (1967:105) explained the Constant comparative method consists of four stages including (1) comparing incidents applicable to each category (2) integrating categories and their properties (3) delimiting the theory (4) writing the theory.

After a while of constant comparative analysis gradually the initial form of the theory will emerge from the collected data. In further stages of data collection and data interpretations the researcher might identify new categories or categorize the previous ones into some sub-categories. Then the researcher will identify specific properties of each category. Glaser and Strauss, 1967:28) mentioned:

> "Generation of theory through comparative analysis both subsumes and assumes verification and accurate description, but only to the extent that the latter are in the services of generation.

The second issue that Glaser (1992:14) recommends to minimize the bias is saturation. By saturation he means a stage in data collection when the researcher notices continuation of data collection will not lead to any new finding in the research. Cutcliffe (2000: 1477) described saturation stage in the data collection procedure as follows:

> "In order for concepts and categories to emerge during the data analysis, the need for sampling of specific data sources continues until each category is saturated. Therefore, at the beginning of the study, there are no limits set on the number of the participants, interviewees or data sources. The researcher continues selecting interviewees until they are saying nothing new about the concepts being explored."

Core relevance is the third issue that Glaser (1992) suggests to decrease the possibility of bias in data collection and analysis. Core relevance becomes clear for the researcher after coding the data which is explained in the next section.

Coding the data

Coding the collected data is one of the main stages of exploring the dataset by the researcher. By coding the raw data the researcher allocate specific meanings to the specific parts of the data to make the dataset manageable for the further stages of the research. Coding the data enable the analyst to identify categories and subcategories. In Grounded Theory text books (e.g. Glaser, 1978, 1992, 1998, Strauss and Corbin, 1990, 1994, 1998) there are detailed

explanations about coding procedure. However, as Heath and Cowley (2004: 146) mentioned Glaser and Strauss use different terminology for coding the data:

> "Glaser and Strauss (1967) originally described two levels of coding, first into as many categories as possible and then integration of categories. Neither in the original publication nor in later separate contributions from the two researchers are coding stages meant to be distinct and linear in their use. However, for Strauss and Corbin (1990), two levels become three. Strauss and Corbin (1990) describe the first level procedures as open coding whilst Glaser (1978) refers to substantive coding."

Heath and Cowley continues (2004: 146)

> "Kendall (1999) also suggests that Strauss and Corbin (1990) last coding procedure of selective coding is similar to Glaser (1978) theoretical coding, but paradoxically also says that they are used differently to generate different types of theory. It is the differences inherent in the terms 'selective' and 'theoretical' coding that are of key importance. The construction seen in axial coding continues in selective coding, with coding focusing on one category at a time until the researcher feels ready to *choose* the core and thus focus analysis on integration."

Substantive codes

The idea of substantive codes is a result of the first phase of coding which is called substantive coding and is very similar and almost the same with Strauss's suggestion for open coding. Glaser (1992:38) defines this stage as follows:

> "...the initial stage of constant comparative analysis, before delimiting the coding to a core category and its properties or selective coding. The analyst starts with no preconceived codes; he remains entirely open".

Theoretical codes

The same two stages of substantive and theoretical stages of coding based on Glaser (1978) happen in three stages of open, axial and selective coding in Strauss's version. Therefore, theoretical coding includes axial and selecting coding. As Babchuk (1996) believes:

> "At the heart of grounded theory analysis is the coding process which consists of three types: open, axial, and selective. Open coding is the initial process in grounded theory which involves breaking down, analysis, comparison, and categorization of data. In open coding, incidents or events are labelled and grouped together via constant comparison to form categories and properties. Axial coding, on the other hand, represents the delineation of hypothetical relationships between categories and subcategories, while selective coding can be described as the process by which categories are related to the core category ultimately becoming the basis for the grounded theory".

Memo writing and sorting

Glaser (1992:108) cited his previous definition of memo in *Theoretical Sensitivity* (Glaser, 1987) and defines it as follows:

> "...the theorizing write-up of ideas as they emerge, while coding for categories, their properties and their theoretical codes, ...they are written up as they strike the analyst when constantly comparing, coding and analysing".

Miles and Huberman (1984:69) support Glaser's (1978) definition of the memo and illuminate it in further details:

> "Memos are always conceptual in intent. They do not just report data, but they tie different pieces of data together in a cluster, or they show that a particular piece of data is an instance of a general concept."

Glaser (1992:111) reported that in writing memos and even in writing the first draft of the theory, researchers should not be worried about the perfect English and not let the concerns about the style of writing block their creativity. Therefore, memos are usually written in a very rough style to record the evolving procedure of theory emergence through a long time interplay with data. The manner of memo writing reassures the emergent nature of Grounded Theory because a useful memo is based on the promptness and spontaneity of researcher's mind about the collected data. Glaser (1992:108) declares that Strauss's definition is very similar with his definition. In terms of Strauss's definition, (Strauss & Corbin, 1998: 217), memos are:

> "...specialised types of written records, those that contain the products of analysis or directions for the analyst, they are meant to be analytical and conceptual rather than descriptive."

Strauss & Corbin, (1998:218) also stated that:

> "Memos and diagrams evolve. Perhaps the most important point to keep in mind is that there are no wrong or poorly written memos. Rather, they grow in complexity, density, clarity, and accuracy as the research progresses".

Strauss & Corbin (1998) believe that memos are an important part of the research process and the conceptual density and integration of the final theory is highly depends on the researcher's attention in making memos and diagrams constantly and meticulously. Working on memos is helpful for the analyst to move from working with data to conceptualizing and gaining analytical perspective about the data. In describing the features of memos and diagrams Strauss & Corbin, (1998) goes on and state that memos and diagrams vary in content, degree of conceptualization, and length depending on research phase, intent, and type of coding.

In order to facilitate and systematize the memo making procedure a couple of practical policies can be taken into consideration. Strauss & Corbin (1998: 221-223) suggested fourteen recommendations to apply while developing memos. These practical guidelines are useful to facilitate the memo-writing procedure and have been employed in developing the memos in the current research. It is recommended by Strauss & Corbin (1998: 221-223) each memo and diagram should be dated and contain references including the date, short quotes or phrase of raw data, type of memo, and the code number of the interview to identify the document from which ideas were derived. Moreover, they continues and highlight the points that because of the

dynamism of the research process the analyst should not be afraid to modify the content of memos over the time as the analysis progress.

Writing the theory

Glaser (1992:111) defines writing stage of Grounded Theory as "write-up of piles of ideas from theoretical sorting". Writing the theory is the final stage of Grounded Theory which leads to announcing the results of the study to a wide community of audiences. In this stage the researcher has to summarise and restructure the findings into a coherent body of text which would be easily understandable and informative for the reader. This is a rather delicate task that should be carried out meticulously.

This is also a difficult task because the researcher who has done the research knows more than anyone else about the phenomenon under study. S/he knows every single detail of the research and it is not easy to share all of this knowledge with others. In fact, at the end of the research there is a very clear picture of the phenomenon under study in the researcher's mind which is only clear for him/her because of a long term interaction with the research context. Successful transfer of this clear picture from the researcher's mind on the paper is a challenging task. However, it is essential to carry it out to achieve one of the main goals of the research which is contribution in knowledge. Glaser (1978:128) declares the importance of writing and publishing the results of studies based on Grounded Theory and states:

> "The goal of grounded theory methodology, above all, is to offer the results to the public, usually through one or more publications ... both feedback on and use of publications will be the best evaluation of the analysts' grounded theory. It will be his main source of criticism, constructive critique, and frequently of career rewards. In any case he has to write to expand his audience beyond the limited number of close colleagues and students ... the rigor and value of grounded theory work deserves publication."

The researcher in this study also considered Glaser's advice seriously and has published some papers and delivered some presentations based on the study. The details are presented in the appendix 8.

Preliminary study

As it was an exploratory research there were many ambiguous points at the outset of the study that needed to become clear. For example, it was not clear for the researcher how feasible this study is, what is the best aspect/aspects to focus on, how to collect data, how to analyse the data and so on. In order to reduce the level of initial uncertainty the researcher decided to conduct a preliminary study.

Therefore, between July and September 2003 an early study was carried out to prepare the researcher for the main study and reduce the ambiguities. In fact, the rationale behind this initial can be summarized by in a quotation from Heath and Cowley (2004:142):

"Unlike quantitative research, where time is spent reviewing the literature and planning details of all stages of the research process, there is a need to start gathering data in order to formulate ongoing plans and, perhaps, to discover the nature of the research questions."

The preliminary study's questions and procedure

The researcher carried out the preliminary study to prepare himself for the main study and formulate the primary research questions. In addition, the preliminary study helped him to reduce the initial uncertainty about the research topic. In that stage, the researcher wanted to illuminate the current awareness of the web users of the existence of the invisible web and how compatible is this awareness with the definition of the invisible web.

In fact, the researcher was looking to discover whether end users are aware of the existence the invisible web, have they already felt the presence of the invisible web in their real search experiences, does it matter for them that part of the web is invisible, what are their perceptions of the invisible web, and also identification some new aspects in the research context.

Ten volunteer interviewees participated in the preliminary study. The participants included six PhD students and four research assistant in the department of Information Studies at the University of Sheffield. There are two reasons for selecting a sample of people in this department. First of all, because it was only a preliminary study the researcher wanted to conduct it in a familiar area with peers and colleagues and preferably with people who were experienced in information seeking on the web. The result showed that they had long experiences searching in online environments including the web and even in information systems before emergence of the web.

They search the web almost everyday or sometimes more than once a day. Interview was the data collection method and it took in average 26 minutes for each interview session. Details of the participants' characteristics are presented in table 3.3. In the table there are some abbreviations that are defined as follows.

PAS stands for "participant's academic statues", PS for PhD student, RS for "research staff", ID for "interview date", IL for "interview length" in minute, and PSA "participants' subject area".

Table 3.3: Profile of the participants in the preliminary study

No	Code	PAS	Gender	PSA	ID	IL
1	FP01	PS	Female	Health Informatics	04/08/03	35
2	FP02	RS	Male	Information Science	05/08/03	40
3	FP03	PS	Female	Health Informatics	21/08/03	19
4	FP04	PS	Male	Chemoinformatics	26/08/03	17
5	FP05	PS	Male	Chemoinformatics	28/08/03	25
6	FP06	RS	Female	Librarianship	28/08/03	19
7	FP07	PS	Female	Health informatics	29/08/03	25
8	FP08	PS	Male	Chemoinformatics	29/08/03	34
9	FP09	RS	Male	Information retrieval	02/09/03	29
10	FP10	PS	Male	Information retrieval	11/09/03	16

The preliminary study's outcomes

Analysing the transcribed interviews indicated that the interviewees usually categorise their search experiences into different categories including work-related and personal-related searches. All of them use search engines for information seeking through the web. However, search engines were not their only search tools and they also use other search facilities including academic databases. There was a common feeling among interviewees that there should be further information resources on the web and they can not manage to access to all resources.

They were relatively confident about their search results and they believed that they usually locate what they need but the search results could be better. They usually use Google at the first stage of their information seeking procedure in the web. All of them had at least one failure incident in their search experiences.

The interviewees believed that in these cases their desired information should be somewhere on the web but for whatever reason they were not able to retrieve it. They had different conceptions about the meaning of the invisible web. An interesting aspect of this preliminary study was people's perceptions of the meaning of the invisible web. They was hesitated about their definition and just declared it as their first assumption after hearing the term of the invisible web. Here some of them are presented as follows:

- The hardware and servers behind the user's sight which they can not see them.
- Irrelevant information which is not as relevant as information in the visible web.

- The information on the search engines which you can not find or search engines have not got the information.
- The information that people can not find because they do not know the right search term.

Testing the efficiency of interview method

Conducting the preliminary study was very helpful to find out how useful the interview is for the data collection method in the study. After conducting the preliminary study the researcher found out that interview is an appropriate technique to explore people's perceptions of their search experiences. In qualitative studies the researcher's skills play an important role in success of the study. Conducting informative and in-depth interviews is a crucial skill for qualitative research. The preliminary study provides the researcher with an opportunity to develop his interview skills.

Meanwhile he attended a useful course in the department of Information Studies to learn interview skills. The course was a module for postgraduate student called 'people skills for information work'. Different interview techniques including active listening, eye contact, clarification, paraphrasing, summarizing and body language were discussed in the module and it was very useful for the researcher to practice these skills in the preliminary study and fully employ them in the main study.

Moreover, during the preliminary study the researcher learned how to effectively set up the interview sessions and prepare the requirements for each interview session. He learned that paying attention to small things might improve the whole procedure of the interview. For example, for the first two interviews the researcher used an ordinary tape recorder to tape the interviews. However, he noticed the quality of voice is not good enough for transcription and he decided to use a digital tape recorder for the next interviews. Nevertheless, because it was the first time that he used the digital tape recorder he missed a small point in operating it and at the end of the interview he noticed that the recorder did not record anything and he missed one interview which is not included in the study. Consequently, he learned how to operate the digital tape recorder properly. Nonetheless, to prevent any further problem next interviews have been taped by both digital and ordinary tape recorders.

Through the preliminary study the researcher learned that transcribing interviews is an important task to make a better use of the collected data and conducting a thorough analysis depends on having the details of all interview sessions. At the beginning, transcription was as a very time-consuming and tedious task for him. However, carrying out the preliminary study was an opportunity for the research to practice this task and develop different skills to perform it thoroughly. The interview questions in the preliminary study are presented in appendix 5.

Formulating primary research questions

Perhaps the main outcome of the preliminary study was formulating the final research questions and identification of new directions in the research which led to fixing the main primary research questions (PRQs).

After completion the preliminary study it became clear for the researcher that the direction of the research is original and there are good opportunities to carry out a large scale and in depth study in the area of research.

In addition, successful completion of the preliminary study was a confirmation for the researcher that interview is a suitable technique for data collection and he is able to conduct fruitful interviews to build up an informative and rich dataset. Moreover, he learned that Grounded Theory can be an appropriate methodology to employ for a thorough data analysis in the main phase of the research.

Target group

The target group in this study was the biology community at the University of Sheffield. The detailed description about the research population is presented in below.

Biology community

There are certain reasons that the current study focused on the biology community. In fact, the main reasons of selecting this research population rise from four main motivations as follows:

- The literature review suggested that the information seeking behaviour of biologists – in particular web-based information seeking – has received little attention to date.
- There is a large community of biologists in Sheffield, the three selected departments comprising one of the largest in the UK.
- The researcher used to be a reference librarian for almost three years before conducting this research (from 1998 to 2001) and he used to help web users to locate their information needs in an academic library[18]. It also was very helpful to have more constructive discussions with the interviewees. Accordingly, he prefers to carry out the research in an academic environment rather than the other areas.
- The researcher's subject background is in agricultural science[19] and he is familiar with the relevant subject areas including biology in general and plant science in particular.

The contribution of the fourth reason became evident after the first interview session. The researcher noticed his background knowledge about biology considerably helps him to engage to discussion with the interviewees about their search experiences. On the other hand, because of this background knowledge he has been familiar with many web-based resources that the interviewees kept mentioning in their talks including PubMed and Web of Knowledge.

[18] The researcher holds a degree of Master of Arts in LIS.

[19] The researcher holds a BSc degree in Agricultural Engineering and Plant Breeding.

His understanding about the search topics encouraged the interviewees to discuss about their search experiences in more details because what they were talking about was mainly highly specialized searches about their research topics and having knowledge about the area was very helpful to have a more constructive discussions.

There are three departments which cover the biology science at the University of Sheffield. Departments of Animal and Plant Sciences (APS), Molecular Biology and Biotechnology (MBB) and Biomedical Science (BMS) all work in biology. According to the information presented at the university website these three departments together comprise one of the largest groupings of biologists in the United Kingdom. Having a closer look at the number of people in each departments confirm this claim. For example, only the department of Animal & Plant Sciences at the time of this research comprised 34 academic staff and 46 research assistants/ research fellows, and 50 PhD students. Additionally, there are some centres which work under supervision of these departments. Department of biomedical science contains two internationally recognised centres of research including the Centre for Developmental Genetics and the Institute of Molecular Physiology and centre for Stem Cell Biology.

Further details including facts and figures about the research population and web-based resources that they mainly have access to them are explained in the chapter four.

Data collection procedure

Because the study adopted a qualitative approach the data collection procedure had to be compatible with the main approach of research in qualitative studies. As Ondrusek (2004: 228) declared:

> "Researchers who used this method generally hand-picked individuals or small groups of end-users and then recorded, coded, and analyzed aspects of their behaviour. The other distinctive hallmark of qualitative data collection exhibited in the descriptions of the end-user studies that follow is that the researcher(s) often interacted with subjects".

The details of the data collection stage including the method of recruiting the participants, and other considerations to collect the data are presented in below.

Interviews

In terms of the data collection, this research used one of the most common ways of data collection methods associated the Grounded Theory which is intensive and semi-structured interview. However, it dose not mean that interview is the only valid data collection method in Grounded Theory. As Charmaz (2000: 514) states researchers who adopt Grounded Theory use a variety of data gathering tools:

> "Grounded theory methods specify analytic strategies, nor can data collection methods ... researchers use grounded theory techniques with varied forms of data collection.

Semi-structured, open ended and face-to-face interview has been selected as the most appropriate data gathering tool for the study. There are a number of reasons that justify why the

interview was as a proper method of data collection in this investigation. First of all, interview is a common technique for data collection in qualitative research. Interview provides the researcher with an opportunity to have a face-to-face interaction with research population and let them talk freely about their feelings, thoughts and actions through their words.

Secondly, in terms of the research approach, interviewing was a suitable data collection method because this research required a large and rich set of data and interview provided the researcher with a tool to achieve this purpose. Thirdly, as this investigation was concerned about what people in the research population feel, think and do, the researcher needed a tool to allow him having access to people's thoughts, intentions, and feelings. Accordingly, achieving the above goals was feasible by employing interview method. Patton (1980: 196) on advocating the necessity and value of interview comparing observation or other methods in qualitative research stated:

> "We interview people to find out from them those things we cannot directly observe... we cannot observe feelings, thoughts, and intentions. We cannot observe behaviours that took place at some previous point in time. We cannot observe situation that preclude the presence of an observer. We cannot observe how people have organised the world and the meanings they attach to what goes on in the world. We have to ask people questions about those things. The purpose of interview, then, is to allow us to enter into the other person's perspective.

There are different types of interview. Patton (1980: 196) classified them into five main categories as follows:

> "There are a number of interview forms which might be used in an interview sessions including unstructured, partially structured, semi-structured, structured, and totally structure interviews."

Davidson and Rincones (2001) have summarized the meaning of these five common ways of interview. In the first type, unstructured, which is based on the instinct reaction of interviewees to the questions there is no control on the direction of interview. In the second type, partially structured, the questions are defined without any order. In the third type, semi-structured, although questions are in order any changes are allowed in the order. In the structured interviews questions are formulated in order and coded and finally in totally structured way not only questions are formulated with order, and are coded but also there are alternative answer for each question.

In this study the open-ended and semi-structured interview has been selected because this method enabled researcher to probe the phenomenon under investigation in more depth. In this method although the questions had already been formulated and were supposed to be asked in order, both the interviewee and the interviewer were allowed to talk about anything else which they might find useful or related to the topic. Consequently, all necessary aspects of the research questions could be covered during an interview session and also always there were chances to explore any other unexpected aspects.

In total forty seven interviews have been carried out over the whole course of this research. Ten pilot interviews in preliminary study and thirty seven interviews in the main stage of the

data collection. However, only the data collected through the main data collection, thirty seven interviews, was used for the analysis. All interviews have been conducted by the researcher. The details about the interviews are presented in the next section in three sub-sections including before, while and after interviews.

Before interviews

Before each interview session the researcher prepared all necessary requirements including sending invitation letter[20] for potential participants, communicate with those who agreed to take part in the research, confirming the date and place of the interview with the interviewee, printing the consent form with the full name of each participants, booking an appropriate room for those interviewees who did not have personal office, checking the tape recorders, and so on.

Ethical issues of data collection

The anonymity of interviewees has been taking into consideration and the interviewees have been made ensured about this issue. The interview sessions audio files have been saved in secure files that have been protected by passwords. Additionally, the transcribed interviews have been stored in sealed envelopes in a secure place.

A consent letter has been developed and signed by all participants indicating the complete anonymity of data collection and insuring them that the collected data will remain anonymous in the thesis, as it is, and any publication from this research. They also have been informed that they may decline to answer any question they wish and if they decide to withdraw from this project at any time they can do it simply by informing the researcher. The consent letter for participation in the research has been presented in appendix 2 and the consent letter for taping the interview is presented in appendix 3.

An introduction letter was written by the researcher then revised and approved by his supervisors. The letter was signed by the head of the department and has been sent to the heads of three departments of Animal and Plant Sciences, Molecular Biology and Biotechnology and Biomedical Science to introduce the researcher and explain the purpose of interviews. Luckily all three heads of the biology departments accepted that the researcher can contact their staff and students to conduct the interview with them.

Recruiting participants

Recruiting participants in qualitative research studies is somewhat different with quantitative studies. As Mason (2002: 134) stated if the research adopts a theoretical or purposive sampling strategy, then whether or not the sample is big enough to be statistically representative of the total population is not the researcher's main concern. Nonetheless, the qualitative research textbooks suggest that the participants should consists of all categories of

[20] A copy of the invitation letter is presented in appendix 1.

the research population. Accordingly, the participants of this study were from all levels in the target groups including academics, research staff and research students.

Undergraduate students in the biology science have not been included in the recruiting procedure because the researcher wanted to limit the study to a group of volunteers who probably have more experience in web searching with more well-established search topics. Although nowadays undergraduate students also use the web frequently, the researcher sought the participants who could probe the interview questions in more thoroughly. This approach has been supported by the previous publications about Grounded Theory studies. For example, Cutcliffe (2000:1478) cited Morse (1998) and reported:

> "Since the researcher in grounded theory is concerned with uncovering the situated, contextual, core and subsidiary social processes, the social processes need to be shared and experienced by the individuals who make up the researched group. Otherwise, if an individual has no experience of the social or psychosocial process, how can they comment on it? Consequently, grounded theorists using a more narrow or focused sample seek out participants who have experience, the most experience, in the topic of interest (Morse 1998)."

Therefore, the target group of the study was limited to the research population in the biology community. However, it can be enlightening if in future another study investigate similar research questions in the student population and compare the results with the findings of this study. More details about this issue are presented in chapter twelve as part of the recommendations of this study for further research.

Theoretical sampling is used in a considerable number of qualitative studies and particularly in the investigations which adopt Grounded Theory as their research methodology. Flexibility in data collection procedure and adjusting this procedure with the new horizons of the research is one of the main benefits of theoretical sampling. Glaser and Strauss (1967: 45) stated:

> "Theoretical sampling is the process of data collection for generating theory whereby the analyst jointly collects, codes, and analyses his data and decides what data to collect next and where to find them, in order to develop his theory as it emerges."

The researcher recruited a self selected sample for two reasons. Firstly, the procedure of data collection did not suggest looking for a specific category of participants to follow theoretical sampling perhaps because the research population was fairly homogenous. Secondly, the researcher had no authority to select people and he had to wait for volunteers. Therefore, he had to interview those biologists who were willing to participate in the research. The only inclusion criterion for each participant that made him/her eligible to be a participant in the study was being a biologist at the University of Sheffield.

During interviews

At the beginning the of each interview session, the researcher explained the main purpose of the interview to the participant once more because they had already been informed about the details of the interview. However, to make sure that the participants are fully aware about

everything related to the research, it was useful to remind them once more. In particular, to provide the interviewees with a friendly and comfortable atmosphere, the researcher reassured them that whatever they would say in the course of the interview will remain anonymous and their name will not appear in any publication whatsoever.

While interview sessions the researcher employed the main skills of interviews including clarification, active listening, paraphrasing, and body language and so on to encourage the participants to talk about the questions in more details.

Interview protocol

Twenty questions were designed as the basic questions to ask during each interview. However, the possibility of expanding each question to as many sub-questions as required was open for the researcher. The basic questions are presented in appendix 6.

Interview technique

The study employed the Critical Incident Technique (CIT) which is an interview method to help interviewees to remind important incidents related to the interview topic and explain what happened in those incidents in detail.

Meanwhile the interviewer can trace the data that s/he is looking to extract from the interview sessions. As mentioned earlier, a copy of the interview questions has been included in appendix 6. However, this is very important to mention that these questions are not the only questions that have been asked in the interview sessions. As the researcher used a semi-structured interview each questions has been followed by many more sub-questions.

For example, when the researcher asked the interviewee to describe a recent successful or failed search it was just a beginning of a series of more detailed questions. The researcher tried to cover different aspects of the search scenario including which search facilities had been used, which steps the interviewee took to conduct the search, what was the search topic[21], why the search was important for the interviewee, how long did it take to perform the search, and so on. Moreover, in each interview many unexpected but highly relevant questions have emerged through the discussion that the researcher followed them and attempted to pursue in the next interviews as well. For example, after first few interviews the research noticed the interviewees usually make a difference between work-related and non-work related searches and it was a new aspect that enabled him to explore more search scenarios (almost double) which was useful to enrich the dataset and opened a new direction in the research.

In addition, while transcribing the interviews the researcher has simulated on the web the search scenarios that the interviewees explained to him to see how detailed they could remember these searches. Interestingly, because interviewees had described their important and

[21] The researcher's background knowledge about biology was really helpful to discuss about search topics with interviewees.

recent search experiences the researcher has always become astonished about the level of accuracy and details that they could remember. It means web users can recall the details of their recent and vital searches in detail. Accordingly, the appendix just shows the main theme of the interview sessions and not the only questions that were discussed with the participants. Therefore, the researcher did not limit himself to specific questions and attempt to cover different aspects of the search scenarios.

As Gorman and Clayton (2005: 130) declared "the interviewer is a principal determinant of the value of any interview". Accordingly, it is believed that three features of the interviewer in this research can be considered as rather helpful characteristics to improve the quality of the interviews and enhancing the data collection procedure. Firstly, his background knowledge about the biology science helped him to have more effective discussion with the interviewees and encouraged them to explain their searches in more detail. Secondly, his three years work experiences as a reference librarian in an academic library was helpful to interact with the interviewees more effectively.

Thirdly, he had been trained to conduct effective interviews [22] and attempted to employ useful interview techniques including active listening, paraphrasing, body language, eye contact, summarising, and fully transcribing the interviews.

Interviews' locations & durations

Interviews were held in the offices of the participants and were recorded digitally. For those participants including the research students who did not have personal rooms a suitable room was booked by the researcher in the department of the participant. The average duration of the interviews was forty five minutes, ranging from a maximum of one hour and fourteen minutes to a minimum of twenty eight minutes.

Recording interviews

All interviews were fully taped and then were transcribed, each interviewee being sent her/his transcript to enable confirmation of its accuracy, and to provide an opportunity for any changes and possible further comments.

Digital tape recorder was used to record all interview sessions from beginning to the end. The advantage of employing this kind of voice recorder in comparing the ordinary recorders goes back to the quality of recorded voice and the easiness of storage, manage, and transcription. Digital tape recorders produce computerised audio files which can be transferred to the personal computer. These files may be played several times without any damage to the voice quality.

[22] Through taking a course in the department of Information Studies called "People Skills for Information Work" in 2003.

After interviews

After each interview session the researcher immediately started making notes to document the main ideas emerged through the interview sessions and then he transcribed the interviews. Because the researcher was the only person involved in whole procedure of the interviews he was able to transcribe and later on analyse the interviews more effectively.

In fact, the researcher had interaction with the dataset in three stages including while the interview, while transcribing the interviews, and after transcription through reading the transcriptions and coding them.

Transcribing

The researcher transcribed the interviews by himself. He occasionally sought help from his native English speaking friends to figure out some unclear words or local expressions. Although the researcher could employ somebody to transcribe the interviews, he decided to carry out this job by himself because it was a chance to interact with the data before further analysis. In fact, transcribing the interviews was very useful to become more familiar with the dataset. Therefore, it is recommended to the future researchers that although transcribing is a very time-consuming and tedious job, it is useful to become prepared for analysis stage.

Interviews were transcribed immediately after each interview session. Approximately, for each hour interview the researcher had to spend about fifteen to twenty hours to transcribe. After completion each transcription the transcribed text has been sent to the participants to provide them with an opportunity to make any changes and confirm the accuracy of the transcription. It was very helpful because some of the participants made some comments on the text and provided the researcher with further explanations.

In general, it was useful for three reasons. First of all, it was a good way to have confirmation of the accuracy of transcriptions. Secondly, it was an opportunity to follow up any points that the interviewees wanted to explain more. Thirdly, in terms of research ethics the participants had a second chance to change what they had said or remove any part of it or even stand down of the research.

Presenting interview excerpts in the thesis

The volume of the transcribed interviews was much more than what the researcher had expected at the beginning of data collection. Therefore, it was necessary to select only a small amount of the data to be presented in the thesis. As Oliver (2004: 23) states:

> "As a general rule, qualitative data is much more voluminous than quantitative data. Whereas quantitative data may be condensed to several short tables, qualitative data can often seem to be so extensive that the major problem is deciding on the sections to omit...one of the first issues which can arise in the use of qualitative data, is the process whereby some data is selected to be included in the thesis, and other data is omitted..."

For that reason, the researcher selected only small fractions of the transcribed texts to be included in the thesis. A number of excerpts from the interviews are presented in chapters five to eleven to support the discussion. Indeed, in qualitative studies this is a common way of presenting the data in reports. However, the researcher attempted to make a reasonable balance between the amount of the data and the writings in the thesis. As Oliver (2004:24) also states:

> "The usual strategy with say interview data, is to include verbatim extracts from the interviews in the text of the thesis. These extracts are employed as the evidential basis of the arguments developed in the thesis ... an approximate guide might be that the total number of words of analysis should be double the total number of words of quotations. People may differ in opinion, but this would be a reasonable guide".

Luckily, the interviewees were very precise in answering the questions and most of the time it was easy to select a brief and relevant quote to support discussions in the thesis. Therefore, most of the interview quotes are exact transcription of the interviews. Nevertheless, in some cases it was difficult to extract a short quote form a long conversion in the interview to support a specific argument. As spoken language is somehow different from writing language, a direct quote from a dialogue might seem slightly disjointed because of existence of some repetitive and unfinished sentences or unnecessary catch phrases such as "you know", "I mean" etc. In order to present interview quotes in a brief, concise and fluent way, and for the sake of succinctness in the thesis, in some quotations the researcher had to remove some of unnecessary or repetitive words.

The removed parts of the quotes are shown by three dots (...) and also sometimes the researchers added a few words to some quotes in bracket to show that these words are for more explanations. For example:

> AS05: "I was hoping you wouldn't ask me this question because I don't really know [laugh] ... I don't really know how it works. I just know there is amazing stash of information and information transfer ... I use it all the time but about the mechanics how it works I have no idea. ... I suppose I would say this is a huge resource that you can tap into to get information about anything and a very useful resource. ...You can do so many things much more conveniently ... I shop on the web, order holidays ... do a lot of things and that saves me a lot of time."

In addition, sometime the researcher had to remove some parts of a quote for the sake of absolute anonymity of people in the thesis. For example:

> AS01: "It is hard to say really because it has got various levels in one level. It is a great mine of accessible information on a huge huge range of subjects. As another level, it is a very useful means of communication. For instance, I am president of the international ...[23]society and we try to move all our operation to the web-based"

[23] The name of the society was removed from the quote to keep the interviewee as anonymous as possible.

Validity and reliability

Validity and reliability are two important criteria which were taken into consideration both in the data collection procedure and data analysis process in the study. Reviewing the literature about the validity and reliability revealed two important issues. First of all, there is no unanimous agreement on the real meaning of "validity" in qualitative research literature body.

Secondly, the meaning and the way of measuring the research validity is different in quantitative with qualitative research. In fact, definition and method of measurement of validity and reliability are defined differently in each one. In actual fact, these two notions mainly have been developed and employed in quantitative research with the positivist approach and it is difficult to measure them in the same ways in qualitative research with naturalistic approaches.

Regarding the conceptual differences among quantitative and qualitative research we need more appropriate term referring to these concepts. Consequently, researchers who carry out qualitative investigations prefer to use some similar terms including credibility, truthfulness, and dependability. Winter (2000) categorized the validity and reliability in qualitative research from others definitions as follows in table 3.4.

Winter (2000) in his review of the previous researchers' definitions of validity in qualitative research stated that:

> "Some qualitative researchers have argued that the term validity is not applicable to qualitative research and have at the same time realised the need for some kind of qualifying check or measure for their research. As a result many researchers have espoused their own theories of 'validity' and have often generated or adopted what they consider to be more appropriate terms, such as 'trustworthiness', 'worthy', 'relevant', 'plausible', 'confirmable', 'credible' or 'representative' (Denzin & Lincoln, 1998; Guba & Lincoln, 1989; Hammersley, 1987; Mishler, 1990; Wolcott, 1990)."

In qualitative research usually two concepts of "validity" and "reliability" is considered in one single but three-dimensional concept of "credibility". In fact, credibility possesses three aspects which are neutrality (objectivity), truthfulness (validity), and replicability (reliability).

Table 3.4: Different definitions of validity and reliability in qualitative research based on Winter (2000)

Validity	"An agreement between two efforts to measure the same thing with different methods" Campbell and Fisk (as cited in Hammersley, 1987). "The measure that an instrument measures what it is supposed to" Black and Champion (1976: 232-234). "Accuracy" Lehner (1979: 130). "Degree of approximation of 'reality'" Johnston and Pennypacker (1980:190-191). "Are we measuring what we think we are?" Kerlinger (1964: 430, 444-445). "To the extent that differences in scores yielded...reflect actual differences" Medley and Mitzel (as cited in Hammersley, 1987: 150).
Reliability	"An agreement between two efforts to measure the same thing with the same methods" Campbell and Fisk (as cited in Hammersley, 1987). "Ability to measure consistently" Black and Champion (1976: 232-234). "Reproductibility of the measurements...stability" Lehner (1979: 130). "Capacity to yield the same measurement...stability" Johnston and Pennypacker (1980: 190-191). "Accuracy or precision of a measuring instrument?" Kerlinger (1964: 430, 444-445). "To the extent that the average difference between two measures obtained in the same classroom is smaller than...in different classrooms" Medley and Mitzel (as cited in Hammersley, 1987).

Credibility in data collection & analysis

Regarding the above-mentioned illustrations about the meaning of credibility in qualitative research the following sections explain how these aspects of credibility have been addressed in this research.

Neutrality (objectivity)

Objectivity implies that how unbiased the researcher has been in collecting and exploring the data. Regarding the nature of Grounded Theory that requires full reliance and connection to the collected data, implementation of objectivity is a pivotal point to success in developing a robust theory. To ensure that this research enjoyed a high level of objectivity and the researcher took different strategies during data collection and analysis to achieve this goal.

To enhance the objectivity in research Strauss and Corbin (1998: 43-46) recommend grounded theorists to think comparatively and looking for similarities and differences in emerging concepts, collect more evidence, questioning the reality of the collected data, consider the emerging concepts and categories as provisional at the beginning and validate them in the next stages of data collection and keep making comparison.

In this research all of the above strategies were employed. Constant comparative analysis for a long time was the main job of the researcher to enhance the objectivity. Concurrence of data collection and analysis was very helpful to validate findings based on new data. During the course of the study the ability of the researcher for higher level of objectivity was improved too. In terms of neutrality in data collection, the preliminary study was very useful to enhance the researcher's abilities to interact with participants in the main study in a more efficient way.

The first series of interviews took place between August 2003 and September 2003 during the preliminary study and the main data collection began in November 2004 and ended in September 2005. This long period of interplay between the researcher and the dataset which was built up gradually and constantly was an important point to ensure the objectivity of the research. Moreover, the whole courses of interview sessions have been tape recorded and then have been transcribed. Through analysing the preliminary study interviews new perspectives appeared in the research and also conducting this stage provided the researcher with new experiences to carry out the main data collection phase with more practical skills.

For example, after carrying out the preliminary study the initial hesitation of the researcher about the success of interview as a suitable method for data collection was diminished because he found out that conducting interview and then transcribing is a manageable task for him. Furthermore, there are other advantages of carrying out the preliminary study which are summarized as follows:

- The researcher learned that some of the participants required more encouragement to answer the questions and he should provide them with enough time during the interview session.
- The researcher learned how to avoid steering the interviewees' minds to any direction and just let them talk freely about the main topic.
- The researcher learned how to develop a mutual understanding about the meaning of used terms with the interviewees to make sure that they understand the questions accurately.
- The researcher also learned that transcribing is a meticulous task and requires extra attention and carefulness to make sure that the transcribed texts are as accurate as possible.

Interview questions have been formulated thoroughly to cover the research questions. However, all questions were open-ended and the interviewees were allowed to express their views and ideas. As a semi-structured interview technique has been selected for carrying out the interview while interview sessions, the pre-defined questions were asked in order but the data collection was not limited to these questions and interviewees were encouraged to talk

about their views for as long as they wish and they could refer to any aspects which they thought as relevant to the research topic. Interviews were transcribed to provide the researcher with details of what had been discussed during interview sessions. Although transcribing was a very time-consuming and tedious job, the researcher preferred to carry it out in order to develop a rich dataset. Transcribed interviews have been sent back to the participants to confirm the accuracy of the transcription and also to provide them with an opportunity to make any changes including adding more explanations or remove any part if they wish.

Trustfulness (validity)

The researcher attempted to follow the Grounded Theory method as thoroughly as possible and considered the constant comparative analysis method as the key issue to enhance the trustfulness of the findings.

Accordingly, the data collection took place in an unrushed manner over a lengthy period of time. The advantage of this was that it allowed the researcher to develop sufficient insights and explanation about the emergent concepts through constant interaction with the data and extensive memo development. Thus, the process conformed to the essence of Grounded Theory. Another way to evaluate the authenticity and veracity of the study is comparing its outcome with features of a good Grounded Theory. Glaser (1992:13) declares:

> "Virtually every major category in the theory can be a subject for a paper or chapter. The researcher will soon realize that while a core category will be excellent to write a book on, near core categories themselves are occasions for good papers".

Up to now a number of papers and presentations have been developed based on the study which can be considered as an indicator of the productivity of the research. The details of these presentations and publications are presented in the appendix 8.

Replicability (reliability)

There are different techniques to enhance the reliability of a research project. For example, Miles and Huberman (1984: 60-63) suggested double-coding as a measurement tool to assess the reliability of qualitative research. Double-coding means two researchers coded the same dataset separately and then compare the results to find out how similar the codes are.

Although people's perspectives are different and consequently they would code the same dataset in a very different way, double-coding might be a feasible idea to gain a fairly accurate idea about the reliability of a qualitative study's analysis. Miles and Huberman (1984:60-63) recommended the following formula to evaluate the reliability:

Reliability = number of agreements ÷ (total number of agreements + disagreements)

The problem with implementation of this formula in practice comes back to the fact that Miles and Huberman (1984: 54) mentions:

> "A chronic problem of qualitative research is that it is done chiefly with words, not with numbers. Words are fatter than numbers and usually have multiple meanings"

To reduce the harmful affect of this problem on the abovementioned formula Pace (2003, 2004) expanded the formula by developing the meaning of agreement by dividing the agreed codes into two groups of "matches" and "agreements". Pace's (2003: 103) expanded formula as follows:

$$\text{Reliability} = (M + A) \div (M + A + D)$$

However, the researcher in this study believes Miles and Huberman's approach to reliability is very quantitative and does not seem appropriate in the context of this qualitative study. Therefore, this procedure has not been employed in this study. Instead of that, another method which sounds more appropriate has been used to address the reliability (replicability) of the study.

In fact, the next section which is about the "audit trails" and "FOCUS weekly reports" relates to reliability of this research. Over the long period of data collection and analysis the researcher has been documenting all of his ideas and thoughts in different formats. The first attempt was called "audit trails" and then in a new format called "FOCUS Weekly reports". During 51 weeks the researcher developed these reports constantly which includes all thoughts and activities about this research in detail and also includes the memos as well.

In summary, as mentioned earlier, the concept of credibility can be divided into three sub-concepts of neutrality (objectivity), replicability (reliability) and truthfulness (validity) and the following considerations in the study are the techniques that the researcher employed to enhance the credibility of the research:

- Long time reflection on the topic in the first year of the study, year 2003, and before embarking on the main data collection stage.
- A preliminary study to evaluate different aspects of the research to consolidate the foundations of research questions before carrying out main data collection.
- Long time data collection without any rush to push the research without having sufficient insight and explanations about the emergent concepts.
- Constant interaction with the data, extensive memo development through FOCUS weekly reports, following the essence of Grounded Theory, thoroughness in data collection.
- Developing publications and delivering presentations about the research and having other researchers' comments about the progress of the study.

What has been carried out in the study to ensure the credibility of research can be employed by future researchers. For example, the researcher's initiative to develop weekly reports and send it to supervisors to have their comments was an effective way for constant interplay with the dataset.

Summary

This chapter describes different aspects of the adopted research method in the study. The chapter begins with an explanation about differences between quantitative and qualitative research and continued with more illustration about qualitative research.

In addition, the chapter reports the procedure and outcomes of a preliminary study which helped the researcher to formulate the primary research questions and to make the final decision for adoption of Grounded Theory as the research method in the study.

Moreover, the chapter depicts a history of Grounded Theory and main features of this methodology and reports how the researcher followed the essence of Glaser and Strauss's (1967) methodology to generate the theory after a long time constant and active interplay with the dataset.

Finally, the chapter reports that in qualitative research usually "validity" and "reliability" are considered in one single but three-dimensional concept of "credibility". The chapter reports that credibility possesses three aspects including neutrality (objectivity), truthfulness (validity), and replicability (reliability).

CHAPTER FOUR:

DEMOGRAPHICS INFORMATION

CHAPTER 4: DEMOGRAPHICS OF THE SAMPLE

"Success is not final, failure is not fatal: it is the courage to continue that counts."

Winston Churchill

Introduction

This chapter reports the demographic information of the research population and the characteristics of the participants in the study. In fact, the chapter reports those results which are mainly related to the facts and figures rather than analytical outcomes of the research. The purpose of this chapter is to provide the reader with an overall picture of the main characteristics of the target group in this study.

Obviously, it is important to know who participated in this investigation because the findings of the study which will be presented in the following chapters may reflect the characteristics of the participants in terms of the ways in which they interact with the web.

Demographic information

As has been stated in the earlier chapters, the biology community at the University of Sheffield has been selected as the research population for this study. This community consists of three departments including Animal and Plant Science (APS), Biomedical Science (BMS), and Molecular Biology and Biotechnology (MBB). In the following section some basic statistics about the biology community and the participants in the study are presented.

Biology community profile

Table 4.1 shows three main categories of people who are involved in the research activities in the biology community. The table demonstrates the number of academic staff, research staff, and research students (Mphil and PhD) in three departments of the biology science at the time of the data collection (from November 2004 to September 2005).

In fact, as the table indicates this investigation only includes the research-based part of the biology community and excludes the teaching-based part. The reason that the researcher has excluded the undergraduate students from the study comes back to the fact that the researcher intended to focus on those member of the biology community who have more interaction with web-based resources, use the search facilities more often, have more search experiences and have more established search topics.

However, it would be valuable and enlightening if other researchers in future can investigate the same or similar research questions on the undergraduate students in a separate study and compare the results with the findings of the study. In fact, this is one of the recommendations for the further research which are suggested in chapter twelve in more detail.

Table 4.1: Number of research-based people is three department of the biology community at the University of Sheffield

Category \ Department	APS	BMS	MBB	Total
Academic staff	41	38	41	120
Research staff (postdoctoral associates & research assistants)	46	47	40	133
Research students (Mphil & PhD students)	50	38	48	136
Total			389	

Participants' profile

The main data collection stage was carried out in three phases. In the first phase fifteen members of the academic staff, in the second phase eleven members of research staff, and in the third phase eleven research students participated in the study. In total thirty seven participants voluntarily took part in the investigation.

Regarding the total number of the research population the sample of the study includes 9.52% of the whole population which is quite substantial for an entirely qualitative research. Nevertheless, as it has been declared earlier in chapter three the sample size is not a concern for researchers in qualitative research. However, more number of participants in the study might arguably have a positive affect on the validity of the research and can cover different aspects of the phenomenon under study.

As figure 4.1 shows, forty percent of the participants were from academic staff category. Research staff and research students have the equal percentage of thirty percent. As figure 4.2 indicates the department of Animal and Plant Science with 44 percentage of the participants in the study stands on the top of the list.

Seventy percent of the interviewees were male and thirty percent were female. This difference between the percentages of the two groups reflects the fact that the biology community at the University of Sheffield is a male-dominant area. For example, amongst 120 members of academic staff in the community only 22 of them were female which formed 18.3 percent of the population at the time of the study.[24]

[24] Characteristics of the research population reported in the thesis are based on the time of the data collection, November 2004 to September 2005, and might be different now.

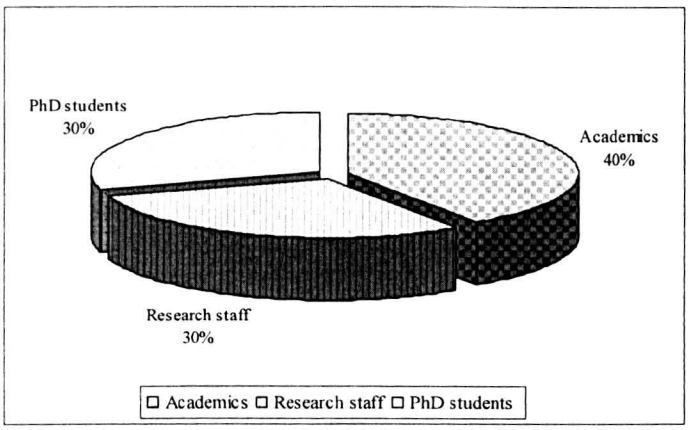

Figure 4.1: Distributions of participants in the whole course of the main data collection based on their level of work/study

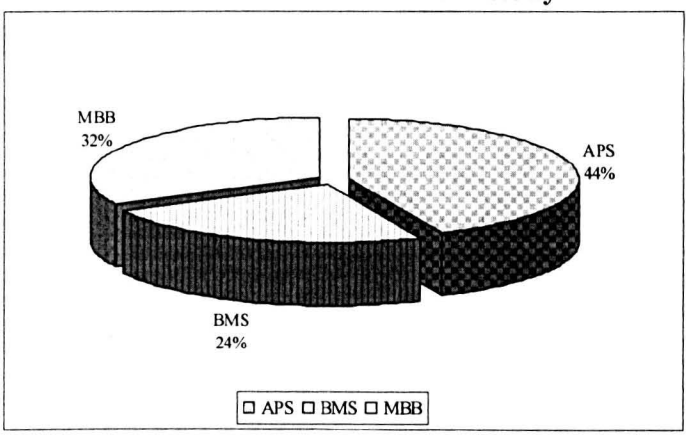

Figure 4.2: Distributions of participants in the whole course of the main data collection based on their department

The details of each phase are presented in the following sections.

Phase 1: Academic staff

In the first stage of the main data collection fifteen members of academic staff accepted to participate in the study. This was a self-selected sample because the researcher had invited all of them through individual invitation letters. Among one hundred and twenty invited people fifteen volunteers accepted to take part in the study.

Table 4.2 shows the details of the participants' basic characteristics including their age, gender (G) and Participant's Academic Status (PAS). Moreover, the table presents the details of the interview sessions including the Interview Date (ID) and Interview Length (IL) of each session. PAS in the first phase of data collection included different categories which are Professor (Prof.), Senior Lecturer (SL), Lecturer (L) and Research Fellow (RF).

Table 4.2: The profile of the participants and the interview sessions

No	Code	PAS	G	Age	ID	IL	Dept.
1	AS01	AS/Prof.	M	62	05/11/04	32	APS
2	AS02	AS/Prof.	M	54	16/11/04	33	APS
3	AS03	AS/Prof.	M	55	25/11/04	38	APS
4	AS04	AS/L	M	31	07/12/04	38	APS
5	AS05	AS/Prof.	F	43	08/12/04	34	BMS
6	AS06	AS/Prof.	M	47	13/12/04	41	APS
7	AS07	AS/SL	M	46	13/12/04	32	BMS
8	AS08	AS/SL	M	45	16/12/04	49	BMS
9	AS09	AS/SL	M	62	16/12/04	52	BMS
10	AS10	AS/Prof.	M	59	16/12/04	38	MBB
11	AS11	AS/SL	M	62	17/12/04	35	MBB
12	AS12	AS/L	M	44	20/12/04	64	MBB
13	AS13	AS/Prof.	M	42	21/12/04	31	MBB
14	AS14	AS/L	M	36	21/12/04	30	MBB
15	AS15	AS/RF	M	39	12/01/05	28	MBB

As table 4.2 shows all except one of the participants in this stage of the study were male staff and it reflects the fact that the biology community at the university in general and in the academic staff in particular is mainly male-dominant. The average age of participants was 47 years old. The average length of interviews in this phase is forty minutes. The longest interview in the first phase was one hour and four minutes (AS12) and the shortest interview was twenty eight minutes (AS15).

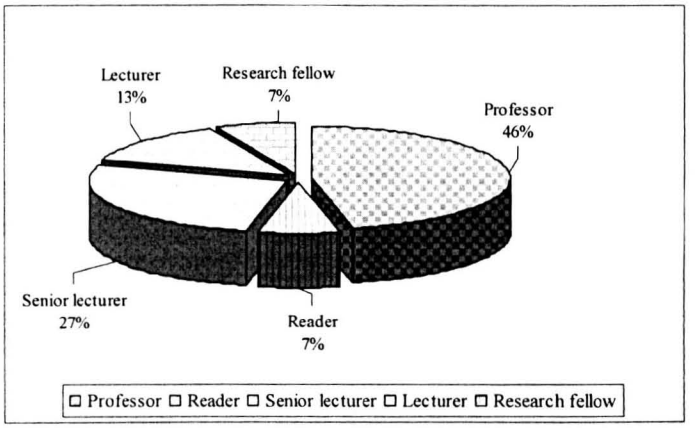

Figure 4.3: Distributions of participants in the first phase of the main data collection based on their academic statues

Phase 2: Research staff

In the second phase of the main data collection the same procedure in recruiting of the participants were employed and all research staff in the three departments have been invited through individual invitation letter to take part in the study.

However, despite the number of people in the second category were greater than the first category the number of volunteers were less. In fact, out of one hundred and thirty three members of research staff eleven of them accepted the invitation and participated in the study.

The other difference between these two groups was the greater number of female participants in the second group. Out of eleven participants four of them were female which was much higher than the number of female participants in the first group.

The average age of participants was thirty six years old (36.4). The last participant in this stage, RS26, refused to declare his age and therefore his age is not stated in table 4.3.

The average of interview length was forty five minutes. The longest interview in this phase took one hour and fourteen minutes and the shortest interview was thirty one minutes. The details of the second phase of data collection are presented in table 4.3.

Table 4.3: The profile of the participants and interview sessions in the second phase of the data collection

No	Code	PAS	G	Age	ID	IL	Dept.
16	RS16	RS (Post-Doc)	M	32	19/05/05	31	APS
17	RS17	RS (Post-Doc)	F	45	01/06/05	60	APS
18	RS18	RS (Post-Doc)	M	37	02/06/05	43	APS
19	RS19	RS (Post-Doc)	M	31	03/06/05	41	APS
20	RS20	RS (Post-Doc)	M	34	03/06/05	34	APS
21	RS21	RS (Post-Doc)	F	26	09/06/05	49	BMS
22	RS22	RS (Post-Doc)	F	31	17/06/05	36	BMS
23	RS23	RS (Post-Doc)	F	29	24/06/05	39	BMS
24	RS24	RS (Post-Doc)	M	49	11/07/05	36	MBB
25	RS25	RS (Post-Doc)	M	50	12/07/05	74	APS
26	RS26	RS (Post-Doc)	M	?	14/07/05	45	MBB

Phase 3: Research students

In the final phase of the study eleven research students participated in the research. The variety of the research students' year of study was from first year students to the final year ones. It was a useful to provide the research with some level of diversity among the third category. Table 4.4 demonstrate the details of this phase of study.

As table 4.4 shows five interviewees in this stage were male and six of them were female students. The average age of the participants in this phase was 25.1 and the average length of interviews was 47.1 minutes. The longest interview in the third phase was one hour and twelve minutes and the shortest interview took thirty six minutes.

Table 4.4: Profile of the participants in the third phase of the data collection

No	Code	PAS	G	Age	ID	IL	Dept.
27	PS27	PhD Student	F	25	12/07/05	36	APS
28	PS28	PhD Student	F	29	14/07/05	47	APS
29	PS29	PhD Student	F	25	19/07/05	54	APS
30	PS30	PhD Student	F	25	20/07/05	38	BMS
31	PS31	PhD Student	F	23	20/07/05	42	BMS
32	PS32	PhD Student	M	26	22/07/05	45	APS
33	PS33	PhD Student	M	25	09/08/05	55	APS
34	PS34	PhD Student	F	24	25/08/05	39	MBB
35	PS35	PhD Student	M	24	25/08/05	46	MBB
36	PS36	PhD Student	M	25	01/09/05	72	MBB
37	PS37	PhD Student	M	26	26/09/05	38	MBB

Summary

This chapter depicts an overall picture of the main features of the participants in this study. The main purpose of the chapter is reporting two main issues. First of all, the features of the research population and then more detail about those members of the population who participated in this investigation. The biology community at the University of Sheffield has been selected as the research population for this study.

The community consists of three departments including Animal and Plant Science, Biomedical Science, and Molecular Biology and Biotechnology. The department of Animal and Plant Science with forty four percent of the participants contains the largest part of the sample. In each department there are three categories of people who are directly involved in research based activities including academics, research staff and PhD students. Forty percent of the participants were from academic staff category, thirty percent from research staff and the rest which is thirty percent were PhD students. The majority of the participants were male because seventy percent of the interviewees were male and thirty percent were female. Probably, this difference between the percentages of the two groups reflects the fact that the biology science at the University of Sheffield is a fairly male-dominant area of study.

CHAPTER FIVE:

SEARCH EXPERIENCE, MOTIVATIONS & FEELINGS

CHAPTER 5: SEARCH EXPERIENCE, MOTIVATIONS & FEELINGS

"You cannot learn anything from success, you only learn from failure." Jim Dale

Introduction

This chapter reports the first part of the findings of the study. The chapter seeks to answer a number of questions about participants' feelings, thoughts and actions while searching the web. The chapter describes how participants in this study perceive web searching in general, how experienced they are in web searching, how often they search the web and what kinds of resources they usually look for on the web.

The chapter begins with the results about the participants' level of search experience and it continues with their positive and negative feelings of web searching. Next part of the chapter devotes to the findings on how participants in the study define the web in their own words and based on their real experiences of interaction with the web. The results which are presented in this chapter indicate that there are two main types of web searching including work-related and everyday life searches. The findings of this study explain how these two types are similar and different from each other.

The last part of the chapter gives an overall picture about the participants' motivations of web searching. The findings suggest there are a number of categories that explain what motivations drive end users to carry out search on the web.

Participants' search experience

One of the main features of any web searcher which plays an important role in success of search is his/her search experience. Almost all of the participants in this research were experienced searchers with long time interaction with the web. Most of them have been engaged in web searching more or less since the web has become accessible for the academia. The one exception started searching three years ago. All participants typically conducted multiple searches daily.

There are some quotations in reply to the question that "how long have you been searching the Web?" as follows. For example, two of the interviewees, AS08 [25] and AS07, reported they had been searching the web since it has become available at the university.

[25] Each participant's code consists of two parts. First part of the code is formed of two letters which shows the category that the participant belongs to. There are three categories including AS, RS and PS. AS stands for "Academic Staff", RS for "Research Staff", and PS for "PhD Student". The second part of the code consists of a number which shows the number of the participants in order. Therefore, AS08 means the eighth participant which was a member of academic staff.

AS08: "Really since it became available. I have been here for 15 years or so. So, my use of internet databases is parallel with the development of it within this university and for as long as we have had web based search engines I have been using them in this university."

AS07: "Hmm, gosh, probably since the early 1990s I think. When I got my first computer which was accessible to the web, I think it was early 90s."

Interestingly, some of the participants were not able to remember exactly when they began searching the web. However, they knew over the past years their usage of the web has constantly increased. For example, one of the interviewees believed he had been using other applications of the internet including email before searching the web.

AS03: "That's a good question, hmm, certainly since I have been in Sheffield. I came to Sheffield in 1992. I mean before that I always was using e-mails and searching the web has been more recent. So, within the last ten years I suppose and it increases all the time and it happens everyday, most of the day."

AS13: "Hmm, no idea [laugh]. How long has it been available in the University of Sheffield? So, that's how long it is"

There was only one exception among the participants who has been using the web only within the last three years. He was one of the oldest interviewees and despite most of the participants he still did not consider the web as an absolutely necessary source of information. He expressed his view about web searching as follows:

AS01: "I treat the web as a sort of luxury item. I don't think as an essential tool, but of course if I was in your age I would think it as an essential tool and not a luxury."

The reason that this interviewee has not been using the web since it has become available at the university is related to the specific features of his area of study – plant classification - and the types of information resources that he needs in his research. In fact, the majority of information resources that he requires are still in the printed format rather than online form. He explained his history of web searching and his motivations of using the web in the recent years as follows:

AS01: "Well, I would think probably only the last few years, last two or three years probably. I have been using it because when I recognized that students were using it as their primary source, undergraduate students were using it as a primary source. At that point, I decided, you know, at least [I] had to find out what they're up to."

Web search frequency

In addition of long time experience in web searching, the participants in this study search the web very often. They mostly search the web almost everyday. Some of them search the web more than once a day.

In reply to the question "how often do you search the web?" The following quotes are two examples of the answers:

AS02: "Hmm, good question, perhaps ten times a day."

AS04: "Multiple times the day. So, ten, twenty, I have never counted."

It seems web searching has become an integrated part of the participants' daily activities. Some comments show that web searching is an included part of the participants' everyday work which begins in the morning with other daily activities and continues during the day as a normal action.

AS03: "Well, it starts at half past seven in the morning and carries on until half past four in the afternoon, but obviously I am doing other things in between. So, all of those times during the day..."

Although they search the web everyday, it does not mean they search the web in the same way or for the same things everyday. In fact, there are different motivations for different searches. One of the interviewees said:

AS08: "I search the web everyday whether I am doing that necessarily for biological research or teaching or my own research ... so, most of search I do is in relation to content I need for teaching. Yes, most days I search the web and most days that would be searching for teaching."

However, as one of the interviewees declared the frequency of search depends on other daily activities that they are involved with.

AS12: "It very much depends on what I am involved in. Sometimes it would be, [pause] ...sometimes it may be only once or twice a day but sometimes I would do many searches in a day."

The above quotes about web search frequency shows how web searching has become a part of the everyday life for the participants in this study. In fact, it is similar to ask them, for example, how frequent you make a phone call or how frequent you talk to your colleagues during a day which are routine and normal actions along other daily activities.

General feelings of web searching

Exploring the collected data in the light of Grounded Theory method culminated in identification of a number of categories about the participants' general feelings of web searching. The results showed that the participants' general feelings on web searching are relatively similar in some ways. However, there are some differences among their responses as well. To answer the first question of interview sessions which was about their general feeling on web searching they have both similar and different views.

They mentioned that they usually find it "reasonably straightforward" but have some difficulties to filter the retrieved items. For example, one of the interviewees found is fairly easy to search but he also had the problem of information overload which inevitably affected the convenience of search for him.

AS02: "Reasonably straightforward. You can frequently get much more information than you actually require. So, I mean by doing a search or something like this, I type in a few keywords that I think are relevant, I'll get much more information than what I ever wanted, but not always be relevant. It is not always well-sifted."

As the above quotation shows this interviewee is generally pleased with web searching because of the abundance of resources on the web. However, he has some problems with irrelevant results. Nevertheless, he still finds it reasonably straightforward. Another quotation also shows the general view of one of the participants about web searching which is similar to AS02's comment.

AS04: "In general, I find it reasonably straightforward to find what I want on the web. ... I think, perhaps that is because I am usually quite specific about what I am looking for."

Another interviewee highlighted two aspects of web searching which are information accessibility and convenience to access. He compared web searching with traditional way of looking for information through libraries and concluded web searching is more convenient.

AS07: "In general I find it very convenient. It is good to have information accessible from my office because although the library isn't very far away, it is easer to have five minutes to go on the web than just walk to the library."

However, the same participant continued his comments on web searching and immediately referred to two common problems that he had with web searching. Limitations to access to all journals that he needs and slowness of downloading PDF files were two issues that he pointed out. Nevertheless, these two issues are not directly related to searchability of information on the web. The first problem is an example of "proprietary web" that has been explained in chapter two as a part of the invisible web and the second problem is related to the technical deficiency of the hardware that he uses.

AS07: "It is frustrating sometimes, two things particularly [are] frustrating. One is that our library, I understand why, but our library doesn't have permission to access every journal I want. The second thing that is frustrating is that sometimes it takes a long time to download PDF files. I understand the importance of PDF files but my computer is quite slow and my printer is quite slow and it just takes quite a long time to download things."

Another interviewee also highlighted the problem of filtering irrelevant information whilst searching the web. However, he was generally pleased with web searching because of the abundance of available web-based resources and also because of high speed of access to information.

AS08: "Generally good but there is always a problem of filtering the information that you get. Obviously you get so much information that it can be difficult to pick up those things that are really of use. Having said that, it is quicker than the old fashioned paper based ways of searching through literature, and it is broader. So, from that point is good but you have to take downside of need to filter the information that you get."

In general, the overall viewpoints about web searching were mainly positive and the participants in the study declared that they are primarily pleased with the advantages of web searching and regardless of the problems that exist they still find it useful. Another interviewee referred to the advantages of web searching and like many other participants compared it with traditional ways of information seeking through the library and declared:

> PS27: "I find it extremely useful especially for something that I am not too sure about. If I come across a new field or something that I want to explore, maybe some experiment I want to set up, I find it really useful to be able to just to sit there and have a general overview about things and not have to go through acres and acres of technical papers and things. Often easy to find summaries about things...it is like an immediate answer to a question. Isn't it?"

As a summary, the participants' feelings about web searching were more positive rather than negative. Although there were some common problems in web searching including information overload or inaccessibility of some sources, they were positive about web searching. Even some of them explicitly mentioned they would not be able to carry out their daily duties without the web. For example, one of the participants said:

> RS22: "Very useful. I think I would probably struggle without it ...I use it for all sorts of things. On a day to day basis searching for online journals, papers, searching for information about companies and the products they sell, prices, contact numbers, often there is some equipment that we use at the hospital, and things like booking forms and things like that then you can fill in online. Sometimes if I want to use a reagent and I have no idea where to get it from just searching, using Google, can be useful because it can bring up the company name as well."

Participants' comments about web searching were classified into two main groups of positive and negative comments which are presented in table 5.1 and 5.2. The categories presented in these tables and table 5.3 all emerged based on an inductive data analysis in the light of Grounded Theory. The details about this methodology are presented in chapter three.

Table 5.1: Participants' positive comments about web searching

1. Ease of use and convenience

- In general, I find it very convenient. [AS07]
- I think it is very convenient. [PS30]
- I find it quite convenient, quite easy. [AS09]
- In general, I find it good. Yes, it is quite good for finding information. [AS10]
- It seems to be remarkably convenient nowadays. [RS25]
- I find it quite easy to find what I need. [PS37]
- I general, I find it really easy even if sometimes the information I am looking for is a little bit cryptic. [PS28]
- It is quite convenient. It is a tool for me to search for anything. [PS29]

2. Straightforwardness (simplicity)

- It is straightforward to find things on the internet. [RS21]
- Often easy to find summaries of things...it is like an immediate answer to a question. Isn't it? [PS27]
- I find it reasonably straightforward to find what I want. [AS04]
- In general, I find it to be reasonably straightforward. [AS12]
- It is relatively straightforward. [AS13]
- It is very simple to search the web now. [RS24]
- That was very simple to bring those up and print those off. [RS25]

3. Usefulness

- It is often useful. [RS16]
- I find it really useful to be able to just to sit there and have a general overview about things and not have to go through acres and acres of technical papers and things. [PS27]
- I recognize the value of it as a fabulous resource. [RS17]
- Very good, very useful tool whether it is through databases or just general searching for information, it is an extremely useful tool. [PS27]
- I am confident that I can get to enough information that allows me to do my job. [RS20]
- It is obviously a brilliant resource. [RS17]

4. Necessity

- Without the internet I can't imagine how I am going to get my PhD. [PS30]
- It is just helping and you can just find anything on the web. So, it is just really amazing. [PS30]
- I find it having the internet is extremely useful and hard to imagine being able to find information without having it. [PS37]
- I think I would probably struggle without it ... I use it for all sorts of things. [RS22]
- It took a while to get that but without the web we wouldn't have got it [AS03]

5. Accessibility to information

- It is good to have information accessible from my office. [AS07]
- I am fairly happy with them in terms of accessibility. [AS12]
- I have a very positive feeling to the searching the web and I use it a lot and I enjoy using it and I find it very successful. [RS20]

Continue in the next page ...

6. Abundance of information

- They have a lot of material readily accessible on the web. [AS01]
- There is a lot out there for biologists if they're prepared to look obviously. [AS09]
- There are quite a lot of resources out there. [RS21]
- When you search the web you get so much information about so many things that are happening. I have a quite a positive feeling being part of community. [RS20]
- It is like a library with a lot of information. [PS29]

7. Quickness

- This is the fastest way of getting information. [AS02]
- It is quicker than the old fashioned paper based ways of searching through literature. [AS08]
- It is quite quick and most of the time you can find what you want. [PS29]
- It might not give you the best thing but it is quick. So, at least you know quickly whether you failed or not. [AS13]
- Yeah, I think most of us only use the web because we want it to be a quick and easy way of finding information. [AS15]
- Google is fast and also it gives you something else that Medline will never give, like personal homepages and published articles ... I found what I was looking for and I told you that was fast. [RS24]

8. Make Communication to obtain information

- I search for the authors and go to their websites and sometimes they have those papers on the web and or they have their e-mail address and I'll e-mail them to ask a copy of the paper. [AS02]
- This email included a reference to this conference at the royal society. So, I thought that is really interesting and clicked on the link that they helpfully provided, went to the royal society website ... and found details of this meeting. [RS25]
- I got an email when it was up and then I was able to find it and print it off ... so that was another way of keeping up-to-date. [RS25]
- When I don't find the picture I have to email this person and see if they have got an image. [PS28]
- Often easer to drop somebody an e-mail than phone them. [AS05]
- I get e-mailed the content list of about three or four journals but not of 30 journals and of course the information, the papers that I would want could be published in one of thirty journals. So, this is the only way that I can do it. [AS13]

9. Gratification of searching

- I enjoy searching the web. [RS20]
- I am satisfied. I often find enough information that answers the question that I have. [RS20]
- It is like a challenge, if there is a particular piece of information which I know must exist ...it is like a puzzle to try to navigate and find out the piece of information. So, I enjoy that challenge. [RS20]

10. Sufficiency of relevant information

- I can often find something that is related. I can always find enough to keep me with enough reading or more than enough reading. [PS37]

Sometimes it is quite quick and most of the time you can find what you want. [PS29]

Table 5.2: Participants' negative comments about web searching

1. Information Overload

- Just far too much information is out there. [AS02]
- Generally you get too much information when you do a search and you want to refine it. [AS05]
- You can frequently get much more information than you actually require. [AS02]
- The problem with Google is that you end up with a load of rubbish as well as what you want ... you end up with hundreds and hundreds hits ... most of the websites you get form Google are useless. [AS13]
- The internet is a great place to find information but you find a lot of junk that you have to treat carefully. [AS08]
- I found a lot of other information but I didn't find that specific thing. [PS37]

2. Irrelevant information

- There was a lot of information for strictly agricultural commercial viewpoint and there was not so much about biological things.[AS01]
- It is easy to find websites. It is not easy to find the website you want or a useful website. [AS10]
- It can be a bit annoying because sometimes you get a long list of hits and maybe a lot of them are kind of irrelevant to what you need. [RS22]
- There is a lot of ...I would say irrelevant information out there. [RS23]
- You often get stuff that shouldn't come up with your keywords but it does. [AS02]
- It gives you a lot of irrelevant information sometimes or repeated information. [PS29]

3. Necessity of filtering the results

- I do spend time to filter out all this stuff. [RS18]
- I have to filter out a lot of low quality information. [AS12]
- There is always a problem of filtering the information that you get. [AS08]
- It is just quite labour-intensive simply because of filtering out the inappropriate information. [AS12]
- You get so much information that it can be difficult to pick up those things that are really of use. [AS08]
- It is a fairly blunt instrument to do things with and requires a bit filtering and careful thought about what you ask for. [AS15]

4. Lack of access to the required details

- I can't always get the full copy of the article. [AS07]
- Our library doesn't take all our journals. [AS02]
- I understand why; but our library doesn't have permission to access every journal I want.
- Could be the most significant paper of this person not be available as full text online. So, your focus can be directed by what is accessible. [AS12]

5. Search tools' lack of Comprehensiveness

- I don't have independent tools to compare. I might think that I am correct. I might trust the tool. But it might be wrong and I know these tools make mistakes. [AS02]
- I know the web is not perfect. [RS24]
- It is good but is not perfect. [RS17]

Continue in the next page ...

6. Systems' Deficiencies

- The only limitation I have found is really to do with the facilities that I have in the university in term of computers, the computing facilities provided are out of date ... when I work with my laptop at home I have much more positive experience than when I work in my office. [AS12]
- It is a fairly blunt instrument to do things with and requires a bit filtering and careful thought about what you ask for. [AS15]
- I don't think the internet is as optimized as it could be. [RS26]

7. Problem with credibility

- I'm slightly concerned about things you can pull up from the web and how validate and how accurate the information is. [AS05]
- It is important to make sure that what we use from the web has been validated through peer review. [AS07]
- You can't find an independent source to know how reliable the information you got is. [AS02]

8. Download Delay

- I can't cope with a site takes a long time to load, 30 second maximum, and then I just leave it. [AS12]
- The second thing that is frustrating is that sometimes it takes a long time to download PDF files. I understand the importance of PDF files but my computer is quite slow and my printer is quite slow and it just takes quite a long time to download things. [AS07]
- I have difficulty in accessing simply because of it is slow basically but generally. [AS12]
- Some days it is slow. I think it still takes a fair amount of time to get to what you want and there are times there would be quicker for me to go to the library. [RS17]
- It just goes slow. Whether that is our network here or the wider inside the community. I don't know. [RS17]
- In certain times of the day Web of Knowledge is quite slow. [AS06]

9. Inefficient search tools

- I don't know how to tell Google not to do that. [AS06]
- The machine itself will find things for you but you're still the biological individual at the end of it. You've got to filter it through again. [AS08]
- When I am searching I find things that the search machines correctly found but it is not what I want and then it is no use to me. [AS10]
- The system is just a machine and if you kick the machine but it won't make it work any better. [AS13]
- If I search for a specific brand, say, Paul Smith handbags or Paul Smith website it gives me all the e-bay Paul Smith offers before I find the official Paul Smith website. [PS29]

10. Time-consuming

- It is really time consuming with all the irrelevant information coming along. [PS29]
- Sometimes it takes a lot of looking for something. [PS27]
- I have to look in different ways to get the thing that I am looking for, instead of just typing into the browser what I am looking and get an immediate reply. [PS28]
- It is very convenient but also very time-consuming. [PS29]
- With current resources it would take me too much time to ensure that I have found everything and on a cost-benefit it is not good enough. [AS02]

Continues in the next page ...

11. Possibility of missing relevant information

- I am equally sure you would miss a lot of things. [AS08]
- I would say I probably find 80 percent which are relevant. So, I am sure I miss something [AS06]
- Undoubtedly I miss things. I am sure I miss things. [AS15]
- I always think it is likely that I miss something. [RS20]
- I always miss something. I don't think you can ever find everything. [PS37]
- Sometimes I will miss some. I would say maybe only 60 to 70 of things or maybe 70 to 80 percent. It depends but definitely 20 percent of things I will miss out. *[sic]* [PS30]
- On some searches they are completely open-ended. So, you can never really have a satisfactory search because you never know what else there might be around the corner. [AS08]

12. Frustration and burden

- It is not an interesting part of my work. It is rather boring typing something in and then getting something in return for information that you want. [RS18]
- I would say I am not totally satisfied with them. Sometimes it is frustrating. [PS27]
- It is necessary, but I do not feel particularly comfortable [with it]. I do not like it...I spend so much time on screen already. I do it when it is necessary, but I am not enjoying it particularly. [RS18]
- That is very frustrating because we work very quickly. [AS03]
- Frustrated I suppose, but frustrated with myself rather than frustrated with the web, because I know the information must be there somewhere and I know if I did it properly, I'd be able to find it but I just haven't been able to find it. [AS06]
- Kicking the computer you mean [laughing]. You know it is out there but you can't find it, I suppose you could be very frustrated. I mean you could leave very frustrated. [AS01]
- If I really can't get it and I just know I have seen it somewhere that is very frustrating. [AS07]

13. Repetitive information & duplications

- There is so much duplication on the web. So, it would be great if you could find some sort of way to get rid of duplication. [AS02]
- Sometimes when I type in a term I get fifteen sites all the same. [AS07]
- Anything you type in Google, there is always going to be ten references to the same site or in your top twenty or ten things all tell you the same answers. [AS02]
- I don't want anymore about it which might be the same or similar or repetitive. [AS07]
- You have repetitive information in different sites. [RS26]
- I pulled out lots of pages describing the same thing. [AS14]

14. Failure in finding information

- Sometimes I don't find what I am looking for. [AS12]
- That is the problem with the negative results. You never know with a negative result whether you've missed something or whether there wasn't anything there to miss. [AS08]
- Occasionally, I don't find what I am looking for. [RS16]
- It does happen if you look for something and you don't find any information on the web. [RS18]
- Yes, I do fail. I do fail to find things on the web, whether that's because they aren't there or whether because of I don't have time to find things. [AS08]

Participants' negative comments on web searching are associated with some emotions that are involved in the search process. As the interview quotes in table 5.2 shows in the negative comments the emotional aspects (e.g. frustration and feeling inadequate) are stronger while some positive comments in table 5.1 seem more neutral in terms of the users' feelings. End users usually experience these feelings after failure in satisfying their information needs. These feelings in unsatisfactory situations include frustration, disappointment, impatience, uncertainty, annoyance, blame, and feeling of being inadequate. These feelings with examples of related quotations are presented in table 5.3. As the quotations in table 5.3 shows the participants in this research expressed different kind of uncomfortable feelings which are related to web searching. However, the ways that they conceptualize it for themselves were very different. The findings of the study about their participants' conceptualization of failure are presented in chapter six and seven.

Table 5.3 participants' uncomfortable feelings about web searching

	Emotions	Example Quotes
Uncomfortable Feelings (Impeding Emotions)	**Frustration**	It is a little bit frustrating. ... It is buried in hundreds of papers and you don't want to really plough through that ... so, that is frustrating. I think and again it maybe that I am not searching in the right way and the reviews that I am looking for exist but I just don't know how to find them. [AS05]
	Disappointment	I am quite often disappointed. As I said ... I knew how to get to the web pages and I did but when I have got there it didn't give me what I wanted. So, it was that university there haven't bothered to put on the information that I expected. So, I felt to be a bit let down. [AS07]
	Inadequacy	I feel inadequate and then I do think those things are so specific they may not be there. [RS17]
	Impatience	I don't have the patience to search that exhaustively. I mean if it is not there then immediately I'll look at something in a different way. [AS14]
	Uncertainty	So, frustrated and maybe wonder if there is a better way to do it that I don't know about. [RS22]
	Annoyance	When you typed in and you hit a lot of articles or pieces of information and then you come up with nothing concrete at the end of it, it is annoying. You do feel like you have wasted your time and also maybe ... it is your own fault for that, that is annoying. [AS08]
	Blame	I always blame myself ...yeah, I haven't put in right questions ... I understand [laughing] yeah, I expected and it was a silly thing to ask. It was too general. So, I just I've put it wrong ... I have not defined my question well enough ... I don't blame Google; you know it does a pretty good job bearing in mind. [AS07]

Perceptions of the Web

As the study was mainly concerned with the less accessible parts of the web - which is invisible web - it is necessary to find out how the participants perceive the web in general before exploring their perceptions of the invisible web. In fact, it is useful to find out how they envisage the web first to examine their perceptions of less accessible parts of the web. However, as Kari and Savolainen (2001) stated:

> "Defining the internet in one unambiguous sentence is a mission impossible, because it is such a complex phenomenon"

Nevertheless, based on the previous works Kari and Savolainen (2001) categorised the approaches of internet definition into four categories including technical, practical, cultural and metaphorical definitions.

In Kari and Savolainen's (2001) categorization technical definition of the internet/web defines it regarding its physical aspects including the connections of computer networks. Definition of the internet/web with a practical view highlights its functionality as a source of information supplier. Nonetheless, definition of the internet/web with a cultural approach emphasizes on the human aspect of it and perceives it as a network of people. Finally, a metaphor perspective of the internet/web includes any definition of the web/internet which envisages it with another perception to make it more understandable or tangible including an "ocean of information", "information mall", "mine of information" or etc.

Although, these four categories are based on experts' views on this issue, interestingly, users' perceptions sought in this study reasonably match with these categories. However, sometimes it seems difficult to attribute subjects' perceptions of the web to just one of these categorises. For instance, one of the interviewees envisaged the web with a mixture of technical and practical approach. He said:

> AS01: "It is hard to say really because it has got various levels in one level. It is a great mine of accessible information on a huge huge range of subjects. As another level, it is a very useful means of communication. For instance, I am president of the international ...[26]society and we try to move all our operation to the web-based so we are communicating with our members through the web all the time. That is very very helpful. Hmm, and I suppose that is more or less it really."

Some of the participants had more technical definition about the web. One of the interviewees said:

> AS06: "What do I mean by the web? Hmm, well ... I mean pages hosted on peoples' computers. I guess the most useful thing generally is the search tools run on the web, things like Google. So, as I said the single most useful thing to me probably is Web of Knowledge which is again a tool for

[26] The name of the society was removed from the quote to keep the interviewee as anonymous as possible.

getting to the papers that I am looking for. I don't want to restrict it too much because as I said at the beginning it includes things like intranet we have in the lab. So, it isn't public ... only are readable on the lab and almost any information on the computers but I am not including my files which are only on my computer."

Similarly, another interviewee had a rather technical viewpoint towards the web and referred to the connections of databases around the world that makes the web work to provide end users with their desired information.

AS08: "Anything with WWW dot comes up ... I think to be more specific ... it is a set of shared databases, a global set of databases together with search engines associated with it ... a useful tool and like many tools sometimes doesn't work probably."

In contrast, another interviewee had a practical approach to the web. Moreover, he made a distinction between the communication aspect of the web and its functionality for search.

AS07: "What I mean by the web is the World Wide Web or the internet and I include any searching such as Google, PubMed, [and] the web pages of the university, anything to do with the web. I do make a slight distinction with e-mail because I do think with e-mail [which is] something different from the web. So, e-mail as a communication device, I think, it is different from the web. But the web is everything else that I can access through my computer [like] library sources, time tables, music, [and] anything."

Another interviewee's perception of the web also was matched with the practical category and highlighted the functionality of the web in providing the information for its users.

AS02: "The web is the means of accessing information from around the world. So, it just gives you access to locations that you would never have had access to before. So, I do not have to catch a plane to go to America to go to their library I can go to their library by the web. ... So, it is a means of accessing information from different locations to and it is also a means of communication."

Similarly, another interviewee had a practical perspective about the web and considered it as a source of information to look for information:

AS10: "I usually mean a computer screen with the Microsoft navigator, with the Microsoft browser on there, or the Netscape browser. That is what I think of the web just a picture in my mind of it as a starting place to start searching. I also think of having links to anywhere and everywhere in the world that I can access almost everywhere.

Some other participants had simpler views about the web and only perceive it from the point of the view which helps them to solve their problems in locating information and satisfying their information needs without paying so much attention to the mechanism that makes this system to work. For example:

AS05: "I was hoping you wouldn't ask me this question because I don't really know [laugh] ... I don't really know how it works. I just know there is amazing stash of information and information

transfer ... I use it all the time but about the mechanics how it works I have no idea. ... I suppose I would say this is a huge resource that you can tap into to get information about anything and a very useful resource. ...You can do so many things much more conveniently ... I shop on the web, order holidays ... do a lot of things and that saves me a lot of time."

Some of the participants also had a practical approach about the web. For example, the following quotes show how they perceived the web in terms of its functionality in providing access to information resources:

AS13: "[By] the web I just mean anything that I access via Netscape and that is the web to me and it is the way of finding, the way of getting to sites. So, the web to me is a way of getting from one place to another, from one compute to another one if you like."

AS11: "What do I mean by the web? An interesting question, isn't it? Hmm, I suppose it is what comes up when I start a browser and go to a search engine or start a browser and follow the links through the topics. So, that is one side of the web. The other side of the web for me is the databases that I access and those clearly aren't just web pages. Those are organized information and in many ways those are the most useful things for research."

Another comment on the definition of the web is a good example of the practical viewpoints about the web. He declared that the web for him is just a means to access the information and he does not care about the other aspects of this network.

AS14: "Hmm, well, for me it is the computer interface with a huge amount of information spread throughout the world, mainly in the US but everywhere. It means to me the interface I have within access to it. You know it doesn't matter to me there're servers and things like that. It is just pages and texts that come out when you call ... It is the connection between my computer and a server with a website on it. That's it. ...It goes down a cable and somebody else's cable and all that stuff and the other thing is I don't care. It is a tool ... it is like most machines, I don't care how the machine works. It is just that it does work and it is the most important thing. I don't have time to worry about philosophical things."

Another interviewee had a cultural perspective about the web and highlighted the importance of the web in making links between different communities through the virtual environment. He said:

RS20: "Just when you think of the community of people that you are involved. You know, say, if it is research where in one university and within this one university there are three thousands ten thousands web pages and then there're hundred universities in the UK and there is hundred of countries. So, that is just the academic information and never minds all the other potential sources of information."

Another interviewee also mentioned the cultural aspect of the web and explained how it affects the communication between people in a community like biologists around the word. She said:

RS17: "I think there is a growing culture of that in order to be more efficient in a community, particularly in biology I think. You get a lot of "I found this, this is useful for you and it doesn't take you long time to have a quick look" and it get from all over the world people send stuff and that is very useful because I don't have to search [laugh]. And I think there is a reason behind that to be more efficient as a community searching for things helping each other and this way is quite useful."

Some of the participants had interesting metaphorical perceptions of the web. For example, one of the interviewees envisaged the web as a huge network which is consisted of thousands hubs and little globes scattered around the web with a high level of dynamism which makes it different all the time. He highlighted three features of the web including its dynamism, connections and network-ness and said:

AS03: "To me it means a network of information and if every set of information which is put on this network is like a little globe, there are lots of these globes all over the place wherever they are situated in computer spaces if you like. For me what is critical is each glob is connected to another globe and to some hub somewhere. So, that is how I see it. But it is not static. So, you look at it this morning and this web look like this and then look at it tonight the web has grown, the hub's grown and some more things and more connections but equally some have gone. I mean you do find of course a website which is absent which is no longer active but that is how I see it so that is sort of connectedness, network-ness, dynamic-ness of information."

Some of the participants' metaphorical perceptions have different aspects including biological and political bases. For example, one of them described the web as follows:

RS17: "This [web] is like my signalling [her research topic]. It is very much like the way the internet works. So, I can view it in that way and in a sense I then use it to see 'can I turn that back the other way?' 'If the web works like that would the plant signal system work like that?' So, that sounds very useful as a concept and the way it grows, but there is no overall control of the web and how it behaves which I find fascinating. Nobody controls this thing. It is growing by itself or all users control it which again I mean you can get political as well about it. It is about the only democratic thing that I have ever come across and where it takes us I don't know. It is just fascinating to see what has happens to be there, witness it even if you don't use it but everybody does and you know when more and more people gain access and you can see eventually everybody is hooked up to everybody else. And I don't think anything else could have achieved that. It is absolutely phenomenal and ... so amazing and self-perpetuating. I don't even know a word to describe it. I haven't really got a word to describe what it is because we haven't seen anything like it before. It just keeps growing without control and from a biologist's point of view that is fascinating itself how it grows. Whether it is mechanical or not doesn't matter."

In general, participants in this study had a variety of perceptions about the web which can be classified into the four above-mentioned categories. They usually included communication as part of the web as well.

Two major search types

In response of the question "what kind of information do you usually look for on the web?" there were a variety of answers from academic information to shopping. However, all these searches can be divided into work-related and everyday life search categories. For example, the quote below gives a fairly clear description of the diversity of his searches.

> AS06: "Well, as I said virtually everything really. Hmm, so, I quite commonly look for people's phone numbers or e-mail addresses, hmm, I look to see what research people do, I look for train time-table and the way I am going to and the other big things is looking for papers for information and scientific subjects."

This finding, two types of search, is supported by previous research. According to Kari and Savolainen (2001) two main domains of information seeking are work and leisure. Research by Klobas and Clyde (2000) showed though people use the internet/web for variety of reasons, work-related searches just includes a small portion of their searches.

Work-related searches consist of searches that are carried out to perform those tasks which are related to the searcher's job, career or profession. This category of searches includes both study-related and job-related searches. For example, when students are looking for information to carry out their coursework it is part of this category. In contrast, those searches which are just for the personal interest of the user are considered as leisure. In fact, leisure or personal searches are related to the personal and private matters of life. Searches for online shopping, accommodation, holidays and etc. are examples of leisure searches.

These two categories of search are studied in a broad and well researched area in information seeking research which is called Everyday Life Information Seeking (ELIS) (e.g. Savolainen, 1995, 1999; Kari and Savolainen, 2001, 2002, 2003; Spink and Cole, 2001; Savolainen and Kari, 2004; Bruce et al. 2004; Agosto and Hughes-Hassell, 2005; and Venkatesh, 2006). The importance of studying ELIS comes back to the key role of information availability in the era that we live.

On the other hand, since the early stages of data collection in this study almost all participants made such a distinction between their searches. Accordingly, the researcher became interested to explore this aspect in further depth. However, because of keeping the main focus of the study, it was decided to leave this interesting direction for further research. In the following section there is only a brief explanation on this issue.

Work-related searches

Analysis the data indicated there is a fairly common set of web-based resources that are used by the participants including scientific databases, online version of printed academic journals, online journals, subject gateways, research institutes' websites, other scientists' homepages, suppliers of materials, and etc. For example, one of the participants gave a detailed picture of work-related resources that he uses in his work.

AS14: "Journal articles, technical resources for lab techniques, and buy chemicals and things like that ...price information and availability of things to buy for our lab, DNA resources, that sorts of information quite a lot and then pictures for putting into presentations and then background research information."

In addition, some of the interviewees had their own categorizations about their work-related searches which are mainly based on the search purposes. For example, one of them divided his work-related searches as follows:

AS03: "There're three sources of information. When I am writing a paper then I would maybe looking for relevant references. So, in that case I might go for a targeted search using Web of Science. ...When I am writing lectures, and PowerPoint presentation these days, and then I am looking for good pictures to go with the data. So, I have data and I want some nice pictures of forests or a process or whatever it is ... Finally, if I'm breaking into a new area I would have web general public availability or paper availability of journals of these areas looking at. So, every of three sort of areas really and the fourth one I am also editor-in-chief of a journal and have a lot of papers, there are about one hundred papers a year that I actually edit and we need to find referees for these papers."

Similarly, some other interviewees had their own categorization for work-related searches. For example:

AS09: "Well several categories, a lot of images because my teaching is very much into producing illustrations of human anatomy, [and] human biology. So, a fair amount of time is spent looking for and finding links to image databases."

The participants were more confident to search work-related information and the search results for this kind of topics are more satisfactory. There are different explanations for the difference between these two categories. For example, searching work-related topics are more successful because users know how to formulate more appropriate search queries. Particularly, selecting more well-defined keywords play an important role to develop efficient search queries. However, there is always a feeling that there must be more relevant information about the search topic that might have been missed.

The level of certainty about success or failure is greater for work-related searches because they just make use of a limited number of web resources for work-related purposes and they know very much about these resources and their contents. Accordingly, they know how comprehensive the results are. They have clear idea about the usefulness of each frequently used source and have a clear idea about the boundaries of their search topic.

In a nutshell, the data also indicated in the work-related category biologists search the web for academic published material in general and the publications in their research interest in particular. One of the interviewees explained his experiences in more detail.

AS12: "I look for technical information about research methodology. I look for published literature primarily within my own research field but also beyond that in some extent. I view, as I said, the

table of contents of journals that sort of cover the field that I work in, and to see if there is anything new and interesting to me ... I spend quite a lot of time using web-based resources to identify products that I might want to purchase for my research in my laboratory. And occasionally I would use web-based resources in response to something that I have read just happened to stimulate my interest in an area or something scientific that is completely beyond what I normally study."

Web-based academic resources including Web of Science have priority for them. However, searching these kinds of resources does not stop them to search the publicly available materials. Particularly, homepages of other scientists or research groups in the same or similar areas are sought frequently. One of the motivations that encourage them to visit other scientists' web pages is scholarly communication and the possibility of gaining access to those information resources that they do not normally have access including some fee-based journals that the university does not subscribe. Informal communication between scientists is a way to share information through the web.

AS02: "Because our library doesn't take all our journals. I might know a paper by a group of authors and I search for the authors and go to their websites and sometimes they have those papers on the web and or they have their e-mail addresses and I'll e-mail them to ask a copy of the paper. This is the fastest way of getting information."

Biologists' work-related searches mainly include more factual information rather than subjective topics. Perhaps the reason relates to the nature of their work which is more practical rather than theoretical.

AS07: "It tends to be factual information, times, dates, numeric data rather than opinion. I think most of biology departments you'll find people more interested in looking at papers, articles or data, level of hormone in blood stream or length or times of a particular biological event rather than opinion. ...My research is about fertility and infertility. So, I am very interested in times and levels of hormones in the blood."

The participants also use different alert services for work-related topics. For example, they use alert search services every week to complete the coverage of other searches that they usually carry out through well-known databases.

AS13: "There are few main things. One is literature. But I link up to one of the searches with my keywords that delivers me the new papers every Monday. So, I get those every Monday. I get 150 papers to look through the titles. The other types of things are mainly sort of molecular, biological genetic things. So, I look up a lots of genome information particularly within the organisms that I am interested in and there is whole genome websites that I look at a lot. And also the other main thing that I would do is doing PubMed searches for papers on specific topics... So, these are main things that I do."

The participants also search the web for work-related topics to keep themselves up-to-date in terms of being aware of the current trends and issues in their research topic. For instance, one of the participants described how web searching is a useful way to remain aware of the research activities in his area.

AS02: "The other thing that I use the web a lot is for molecular biology. There are so many molecular and genetic resources out there now. But actually unless you search the web first you are not informed about what you are doing because a lot of this information is not in the library. So, you can't go and read it anywhere. It is only available on the web. So, for example sequencing of human genome, if you're working on a gene you must go and look at the human genome of your organism ... So, without doing a web searching you would never know about this and there are whole series of web tools available just allow you to go to these websites and to access the information."

In addition of all above-mentioned types of work-related searches, the participants also search the web for acquisition of chemicals and laboratory equipment that they use in their experiments. In order to acquire appropriate equipment in terms of the quality and price they have to do sophisticated searches. For example, one of the participants reported that besides all work-related topics, she has to check different suppliers to find the best deal for laboratory equipment.

RS23: "New journal articles, relevant information, gene sequences, protein sequences and also reagents for experiments that kind of things like Oligonucleotides, because we work on Oligonucleotides, just finding out different suppliers for chemicals and that kind of things. We would search for supplier of Oligonucleotides because we buy them and we just want to see who is the cheapest, who is doing good deals, that sort of things."

Everyday life searches

The other major type of search for the participants was everyday life searches. In these searches their search topics are as common as everybody else on the web. Everyday life searches consist of a huge variety aspects of life and include online shopping, train time table, flight time, accommodation, online banking, phone directories, and so on. The data showed there is a considerable diversity in the search topics in this category and in some cases there are considerable difference in the search results comparing the biology topics. For example:

AS14: "You know it is easy to look at phone numbers [on the web], I use it [the web] to find maps, to book hotels, to book almost all of our travels and things like that. You need to research things you are going to buy and quite a lot of different things for home as well. I do a lot of shopping on the web."

The data shows sometimes they have some sort of difficulties to locate non-work related search topics. For example, the following quote shows an example of difficult searches in everyday life searches:

AS05: "I had a real problem. I was looking for ... full spectrum lights. I'd tried to find them in the UK and couldn't, and I could only find a USA supplier and I tried quite hard and I failed. ... I couldn't find it and eventually a friend found it for me through another source – and they had a website. ... She didn't find it on the internet; she actually found it fortuitously in a leaflet she has got. So, she found it that way."

The participants believed that the search type plays an important role to achieve the search goal. Feeling of frustration might happen more often in everyday life searches because everyday life searches cover a huge range of possibilities. Therefore, users are not confident enough to select appropriate search tools and develop efficient search strategies for every single search and in particular for less familiar topics. For example, another participant explained one of his searches in which he had experienced some problems to carry out the search and he had not been able to find a convincing answer for his question.

> AS07: "To some extent it depends on the information that I am looking for ... If I am looking for a railway time and the time isn't the time I want I can get annoyed about it. Let me give an example. I used the web to find a railway time in a station in France. I was going to France for a private thing as a holiday and I was trying to find the time of a train that stops in a particular station and it was a station near Tours in France and I couldn't find any train that stops there. So, I went to the travel agent and they found some train stop there and then I phoned the French embassy and they gave me a number and they found the same thing but the problem was that the railway station has a little side station called St. Pierre de Cours and my train went on to that station and then you had to take a small community train five minutes to the main station. I didn't know about this other station. The travel agent did and I didn't and I tried to look at the information and I couldn't find the information. So, I just got terribly frustrated because there were no train stopping where I wanted to and yes I was told, I had a ticket said that there was a station and I just didn't believe anything then because I felt it is wrong ... I didn't know who to believe. I was very anxious because it was important. It was in a different country and I didn't know where to go. So, I got very very frustrated and I didn't know who to believe."

The above quote also shows the importance of users' ability to evaluate the credibility of information on the web. As the quote indicates this interviewee had a problem to assess the authenticity of the information that he had obtained and he did not know which source can provide him with the correct information. In particular, participants in the study had more problems for evaluation of the authenticity of search results in ELIS. They were more component in assessing work-related items. There are a number of reasons for their better performance in work-related searches including their ability to assess the accuracy of the retrieved information by some criteria (e.g. reputation of journals or authors and etc) but for everyday life they do not have the same level of competence. Moreover, their work-related searches are more well-defined:

> AS08: "In the case of doing something for work where I need a review of the toxicity of Mescalin that's it. That's a closed-ended search. You go, you find it. That's fine. But if you are doing something more general then one search engine is never going to be enough because you never know what else might be there. This is like foraging behaviour in birds. I don't know if you've ever discussed things like mating strategies or foraging behaviours in animals with the zoologists down the road [in Animal and Plant Sciences Department] but some of the strategies that animals use is ... relevant to search strategies that we use for obtaining information from the library."

AS04: "When I search the web for academic reasons, my work, I have 99 percents of the time a well-defined piece of information or image that I am looking for because there is a well-defined goal."

They usually have a set of limited number of frequently used web sources for the work-related topics and usually find the results more satisfactory. Nevertheless, for everyday life searches there are not such a well selected set and they have to try a variety of search facilities. One of the participants had an explanation about this issue and he said:

AS10: "I use it [the web] in work for looking for biological science based material and I use it at home sometimes for anything for interest. So, probably for finding scientific based material on the web I use a limited number of sites. For genome sequences, for journals, publications, for electronic journals, and for all those of kind of things I find it very easy. At home using it to find maybe hotels, flights and things like that I find much more difficult and unsatisfactory ...probably because in searching for non-work related information the scale isn't limited. There are millions of websites out there and the possibility to hit the right one that you want that you're looking for is quite difficult. Whereas, in work it is probably are much more circumscribed set of information I am looking for. So, there are limited numbers of places that I am looking at."

Sometimes non-work related search might lead to failed results. Although failure may happen for the work-related searches, the possibility of failed non-work search seems higher. In addition, for the work-related failed searches they have more coping strategies including asking colleagues, reformulating search query etc. but for the second type they are not able to locate a coping strategy easily.

AS02: "Sometimes I don't find what I am looking for. I can give you an example. Two weeks ago I wanted to buy a new memory stick for my computer ...I decided if I am going to buy a new one I buy the largest amount of memory that I can find. So, I decided to have a look on the web for where I can find this. And there was one site offering a memory stick with four gigabytes of memory which is huge but eventually I couldn't find this site at all. You can find information about it but whenever I tried to go there, there was a faulty link or whatever. So, I knew it was available and I tried for about a week to find high memory capacity memory stick and didn't get anywhere. So, in the end I abandoned there and I just bought a one gigabyte memory stick but I do know the four exist but I couldn't find it. ... No way to buy it and no way to get to the website that sold it because the links have gone. You can try yourself and if you find it let me know [Laughing]"

As mentioned earlier, the issue of information seeking in everyday life is an independent area in information seeking behaviour research. Therefore, discussion about the differences between work-related searches with everyday life searches can continue to more detailed level. However, as this issue was not the main concern of the study it stops here. Nevertheless, it can be enlightening if in future another study can explore this issue in more depth. In particular, it can be useful if future researchers investigate how information visibility on the web and success and failure in search, which form the main them of this study, differ for work-related and everyday life searches. This idea is also highlighted in chapter twelve as a suggestion for further research.

Search purposes

Five categories were identified in the data set about the participants' search motivations. Each category has its own dimensions. In fact, participants search the web to fulfil one or some of the following purposes.

Table 5.4: The identified categories of the participant's search purposes

	Search Purpose	Dimension
1	Question Answering	From fact-based to subject-based
2	Decision Making	From trivial to vital decisions
3	Increase of Awareness	From familiar to unfamiliar fields
4	Problem Solving	From simple to severe problems
5	Making Communication	From informal to formal communications

Question answering

Answering a question is a common motivation to conduct a search. When people have a question in their mind it means there is a gap in their knowledge about a specific issue and they need to fill this gap. As Belkin (1980) and Belkin et al. (1982) stated the Anomalous State of Knowledge in people's mind plays an important role to guide them to seek new information to rectify the existing anomaly.

In a simple word, people search the web because they have a question in their mind which they want to answer. This question might be a fact-based or a subject-based one. If the question is fact-based participants expect that the answer might be a fixed piece of information. Usually answering this kind of questions is more straightforward. Nonetheless, for subject based questions there is no fixed answer and people need to spend more time to find the best possible option. Answering a question through employing any search system requires formulating a search query to negotiate with the system. This is an important part of the process of answering a question. The question maybe presented in a well-defined search query or maybe in an ill-defined query. There are some examples in the data that shows the participants search the web to answer their questions. For example, one of them described the way that he uses the web to find answers of questions and compares it with the pre-web time.

> AS02: "I basically use the web frequently to answer simple questions. So, in the past you would go to your reference book ... now it is quicker for me to go to the web and type in the question and use Google to find the answer."

Another example is a simple search by one of the interviewees to find out the scientific name of a plant through a simple search in Google. She said:

> PS28: "In general, I find it really easy. I usually use Google ... for example, yesterday I was looking for the scientific name of a plant, Euphorbia antisyphilitica, and I knew but I wasn't sure. So, I just typed the common name, Candelilla and Euphorbia, and it just looked up a lot of resources."

RS25: "I think for answering idle queries it [web searching] is often very useful. My daughter had a word in her spelling test in school which wasn't in our dictionary. We have got quite a big dictionary. So, I typed that into Google and it came up with a website which is about words and it happened to have an entry for this particular word which explained how it arose. It was a sort of unusual word and sort of half slang word which obviously hadn't a main entry in a dictionary. It was 1990s edition. So, when I work like that it is extremely useful in a way that would be either hard or very time-consuming to pursue that query in any other way."

Decision making

Sometimes searching is a device to make a decision. Any decision is made based on the available information. Accordingly, when people need to make a decision they need to collect certain amount of information. People may need to make a decision but there are some gaps in their knowledge and this gap does not let them to make a sensible assessment. For example, one of the participants wanted to buy laboratory equipment and he needed to carry out a web search before making any decision.

AS12: "I spend quite a lot of time using web-based resources to identify products that I might want to purchase for my research in my laboratory."

This is the same for many other decisions including travelling and so on:

AS12: "Also I use web-based resources for pragmatic reasons like for example, if I am travelling to conferences trying to find out cheap flights, hotels, and things like that basically to cut down the cost of the university because the cost all over come form the university but it comes from the department and allocation of my own personal grant resources. So, I use it to save money for myself and the department I suppose as well."

Task performance

Although search is a task, we might need to search to conduct another task. The task might be a simple or a very complicated one. In fact, in this category there is a dimension of complexity which is a continuous spectrum from very simple to very complicated tasks. However, in any cases performing the task require collecting a certain amount of relevant information. For example, online shopping is a popular task these days which requires intensive searches to make a sensible decision to buy different things.

AS10: "I mean everyone searches for, you know, airline tickets or information on this and that. We do a reasonable amount of buying across the web and this is for use in our laboratory. So, we use things like e-buyers [http://www.e-buyers.co.uk/] which is actually based in Sheffield for buying our electronic equipment."

The other example of task performance is developing teaching materials for teaching or supervision a research project. The following quotation is an example of task performance searches.

AS11: "On lecture based material it [browsing the web] can be quite useful because in a way this is what I am looking for; I am looking for sites which have a lot of information about other sites. So, browsing works well in that context. And very often it does give the information I want because I am looking for something that is not absolutely cutting edge but more likely review and for that browsing seems to work quite well."

Increase of awareness

Awareness of current issues in any area is a vital point for researchers. Increasing this awareness requires gathering new information through web searching.

AS06: "I have a student, a PhD student, just started a new research project who is working on Alzheimer disease. He has a particular area to look at it which is we think maybe the protein that is responsible for Alzheimer disease bind to a particular kind of lipid on a certain number of nerves. So, we have tried to find more information about those lipids just to look what they are and it was straightaway."

In this category sometimes participants have a certain interest and they just want to keep their knowledge up-to-date. Accordingly, their searches are specific and usually well-defined.

AS07: "I tend to search [the web] for specific things. I tend not to type in broad terms and troll the web. I don't find that as a satisfactory way of doing it. I tend to use the web in a very fixed way. I tend to look specifically for details. I know some people look at it in a different way. They type in the word "ethics" and try to look at ethics of In Vitro Fertilization [IVF] or cloning or genetic engineering. I tend not to do that."

Satisfaction of inquisitiveness

Sometimes participants search the web just as a matter of curiosity. They would like to know what is going on. Searching for news is a common example. This category of search is not very directed and well-defined. For example:

RS25: "Other idle queries like you type in names of people who used to know at school or university ... if you can identify them amongst all the others who have the same name. I had a friend that was in a remote part of Canada fifteen years ago and I wondered whether she is still there she is now in another remote part of Canada but still is practicing as a midwife. So, I am confident this is her. I have lost touch with her ... it was just interesting to find out where she was. So, those sorts of things are fun."

AS07: "A friend of mine has recently have been appointed to a new post at a university outside this country as a chair in a particular subject. I wanted to find out more about his department and so I did a Google search."

Searching for news on the web was also mentioned by some of the participants. In fact, sometimes they search the news websites just to respond their curiosity or to spend their short

leisure times. For example, one of the participants said that she search the web for the news related to her country everyday to just see what is going on there:

> PS30: "Sometimes I search news. Especially I am a foreign student. So, I would like to know what is happening in my country. ...so, I would like to know the news or anything is happening in my country. So, I just do the news search sometimes and then yeah everyday. I don't know how many times really because ... we do experiment and then annoyingly there are just like three minutes, five minuets waiting period during our experiment many many these kind of waiting period. So, you can't really use it [these five minutes waiting periods] for something because your mind is still on your experiment. All you can do is just find something that is not important to do. So, I would normally use this time to go wondering on the internet."

Clarification

In this category participants have basic knowledge about an issue and they require clarifying it by acquiring more information through web searching.

> AS06: "I usually search for different things in Web of Knowledge and Google. For a trivial example, I had a tutorial yesterday and there was a student to write an essay about Thalassemia, blood disease, and I wanted to find out more about Thalassemia. So, I can say something intelligent about the essay. So, I went to Google and then looked up Thalassemia and the pages answered what I was looking for."

> AS08: "I needed to find some information on a drug called Mescaline and very simply I searched at two different searches one through PubMed and one through Google. The reason why I've used those two was because I wanted some hard good quality scientific information about the drug toxicity, profiles, overdoses, etcetera information of pharmacological actions."

Problem solving

Problem solving is another important aspect of information seeking behaviour. This is also an important aspect of web searching and a number of participants referred to this category in their searches. For instance, two examples which are related to this category are presented as follows:

> AS09: "I was helping a student with trying to find some information on what we thought was a part of a condition part of Cystic Fibrosis and she was trying to confirm that a particular problem arose from Cystic Fibrosis as a disease and she knew I could resolve the issue."

> RS26: "Just an example, for instance a friend of mine, she is not a biomedical scientist, but she had a problem with her son. She said my son has been diagnosed with this particular disease and because you are closer to this field you can maybe give me some information about it. So, I spent a few days, maybe five and six days looking at the internet just to gather information to give to her ... or for instance, someone how I know who actually had a postpartum depression, a woman, and then after birth she was actually very depressed I had not any enough knowledge to look for that ...so I search for any type of health issues.

Making communication

Searching might lead to make communication with other people. When web users need to make communication they need more information and the required information can be accessed through searching.

> AS14: "You know, I must have done ten Google search this morning already. As I do it all the time looking for things. So, this morning I had to phone a colleague in Chichester in a hospital and looking on the web for the hospital phone number and that was successful."

> PS30: "I would say that the web is very important communicating thing for nowadays. People use it to communicate to each other and people put their information there. So, other people can get it ... it is just millions of computers are connected and then people can use it to improve their lives.

Summary

This chapter encapsulates the findings of the current research about different aspects of web searching including the search experience, search frequency, search motivations and the participants' feelings on web searching. The findings presented in this chapter showed that the participants in this study are experienced web searches and they search the web quite frequently. They usually look for both work-related and everyday life searches. Work-related searches are more well-defined. In terms of their feelings towards web searching there are both positive and negative feelings.

Ease of access, quick access, availability of a wide range of information resources on the web, and sufficiency of relevant information are some examples of positive comments of web searching. In contrast, information overload, inefficiency of search tools, delay in response and high possibility of missing relevant information are some examples of negative comments on web searching.

CHAPTER SIX:

INFORMATION VISIBILITY ON THE WEB

CHAPTER 6: INFORMATION VISIBILITY ON THE WEB

"Failure is instructive. The person who really thinks learns quite as much from his failures as from his successes." John Dewey

Introduction

This chapter reports this study's findings about the participants' conceptions of the invisible web and how they perceive this phenomenon in their real web search experiences. Moreover, the chapter presents the study's first model that has emerged based on addressing ERQs. To remind the reader about ERQs which were presented in chapter one these research questions are revisited here once more as follows:

- ERQ1: How do a particular group of end users conceive the "invisible web"?

- ERQ2: To what extent do such conceptions and experiences map onto the technical models of the invisible Web proposed in the literature?

This chapter addresses the above questions and presents a model which enables us to have a comprehensive view about the issue of information visibility on the web. The model suggests a new definition for the invisible web which reflects web users' perceptions of the invisible web. The model contributes to the area of information seeking on the web by suggesting a dynamic and comprehensive description of the concept of information visibility on the web.

Conceptions of the "invisible web"

End users' conceptions of the invisible web consist of two aspects including users' knowledge of this term and their experiences of this phenomenon during their real web searching. Obviously, users' conception of this phenomenon was more important for the purpose of the study than their knowledge about the terminology of the invisible web because the employed terminology has been developed by librarians and information scientists and is not necessarily the best way of describing the phenomenon. In fact, in this part of the study the researcher was primarily concerned about end users' conceptions of the phenomenon of the invisible web and not users' understanding of its terminology. However, it is interesting to know how participants in this research expressed their first impression of hearing this term which in most cases did not match with the definition of the term.

All participants experienced some parts of the invisible web in their actual searches. Unsurprisingly, the term "invisible web" was new to all but one of the interviewees. The following excerpts from the dataset indicate how the participants perceive the term. Some of the participants expressed absolute ignorance about the meaning of the term.

AS06: "I don't know what that means at all. It doesn't mean anything to me. I don't know."

RS16: "I have never heard this term used before. So, it doesn't mean anything to me really. I am not sure what it means".

RS25: "The invisible web? I am not quite sure what you mean by that ... I haven't heard the term before."

RS26: "[The invisible web] means nothing to me because I have no idea what it means."

However, some of participants who also had no idea about the definition of the term took one step forward and tried to guess something about it based on their first impression of hearing the term. Although their speculations did not match the definition of the invisible web in the literature, it was interesting to see how this term might be perceived by those who have not heard it before. One of the participants guessed the invisible web is part of the web which can not be accessed by users because of selecting inappropriate search terms.

AS10: "What the invisible web might mean? Nothing, it doesn't mean anything to me. I don't know. I have never heard of the term. It doesn't mean anything to me. ... I don't know unless this term is used by bioinformatics people for pages that one can't access because you can't find the correct search word."

Another interviewee assumed the invisible web might be part of the web that people do not use and the information resources in this fraction of the web is not subject of any use.

AS07: "Hmm, I don't know. It could it be information that is there [on the web] that people don't use – perhaps things that are just dumped there for all sorts of purposes, but nobody ever accesses, such as minutes of committee meetings and that sort of thing. I have to say I am struggling with it and I don't know what it means. What is it?"

Another participant expressed his feeling about this term as a term which describes the technical infrastructure of the internet that is not visible for the user.

RS19: "What the invisible web is! Never have come across the phrase, hmm, the invisible web! It sounds sinister doesn't it? It means nothing to me because I have never heard it before. Do you want I say what it feels like? ...the invisible web to me means the communication systems that aren't actually visible. So, like, you know, electronic signals, satellite communication that is all stuff rather than telephone lines. That is my initial feeling what the invisible web is."

Nevertheless, some of these speculations about the meaning of the invisible web were slightly close to the actual definition of the term but mainly refer to part of it. For example, two of them who have not heard the term before articulated their first impression in a way which covers just one part of it which is the "private web[27]" in these cases.

RS16: "I have never heard this term used before. So, it doesn't mean anything to me really. I am not sure what it means ... I guess it's mainly websites which need some kinds of authorization to access. So, it is not generally accessible to everybody. Possibly that! I am not sure."

[27] The private web is part of the invisible web in Sherman and Prices' (2001) categorization which has been illustrated in chapter two.

RS24: "Well, everything which is not public ... well, because you say the invisible web and that is something that I can not see. So, if I can not see it because it is not public."

Another interviewee who had not heard the term before defined the invisible web somehow close to its definition. She ascribed the existence of the invisible web to the inefficacy of search engines and described her feeling about the term as follows:

RS22: "I don't really know exactly what it means. I can just imagine. I think it might mean ... well maybe it means the information that isn't accessible maybe because the search engines have some problems that they can not retrieve everything properly."

The one exception who had a background in computing managed to define it fairly well. He firstly explained about his background as follows:

AS11: "I came into biology from ... the computing side so this might well explain the bias I have in this regard".

Then, he defined the invisible web in some detail:

AS11: "I mean it is sort of nebulous concept. I think it is clearly the parts of the web that are not reached by search engines, partly because they are dynamic. There are pages just created ... and then disappear. And also I think there are obviously parts of the web that are deliberately hidden from search engines by choice of the authors, either by being locked on campus, the departmental teaching material, or simply by being marked as not to be searched by the robots that generate the web. And I think apart from that, there are pages that do not really appear on search engines because they are not referenced by other pages".

Therefore, it became clear that the participants are not familiar with the term and their first impressions about it did not usually match with the existing terminology. However, when at the end of each interview the researcher explained the definition of the term they declared that they have already experienced some parts of this phenomenon including the deep web, opaque web and so on without being aware of existing such definition or knowing about the severity of the problem. In fact, some of the interviewees said they have already felt the existence of the invisible web without being aware of the term. For example, one of the participants had been aware of the fact that if a webpage has not any link to other resources on the web it remains inaccessible for search engines. These web pages are part of disconnected URLs which has been described in chapter two.

RS18: "Actually I knew this fact but I didn't know the word the invisible web. I mean I know that, for example, sometimes you are not allowed to put reprints of your own papers on your websites because you breach the copyright. But if you generate an extra website with no connection to other websites you can not find it but you can access it if you have the precise address."

Also another participant had experienced the deep web in his previous searchers without knowing this term. He said:

RS25: "Yes, I was conscious of that [the deep web] ... I came to that conclusion that was the case. There were some searches that I have done fairly idly I think and thinking right! I didn't find that probably because it is in a database which the search engine doesn't have access to or wouldn't want to because it doesn't pick up everything. It maybe some kinds of data like air temperature or something like that where I realized that I would need to specifically [find] where the database was and find some entries through that way. ... well, I mean I was conscious that there are bits that search engines couldn't cover like database but I did not hear the phrase the invisible web or know in this extent."

Similarly, another interviewee reported that he had been aware that search engines do not cover all parts of the web and there are resources that remain inaccessible for them.

PS37: "I guess I knew that there're things that engines like Google couldn't access. So, I guess that was kind of in my head already that Google wouldn't search every single thing on the internet. So, I guess I have a kind of confirm in my mind now after you described it. Sometimes when you can't find something you got to know where to go."

Some of the interviewees after learning about the invisible web started comparing what they usually do in their daily searches with the definition of the invisible web. For instance, one of the participants noticed that some of information resources that she uses for subject searches including databases are part of the invisible web and said:

RS21: "So, I guess some of the stuff we would use is probably private to most people. So, people couldn't get access to because either there is a subscription or you have to know to go to the certain database for the information."

In addition, some of the participants attempted to explain different reasons of invisibility of some resources. For instance, one of the participants explained how in different areas of research in science the value of newly discovered data causes invisibility on the web for end users and she referred to it as part of the invisible web which is in fact the "private web".

RS23: "Yeah, I can see that actually... and then I can imagine that and particularly for my research. If it is on the invisible web we wouldn't be ... you know just because of the way people are to an extent quite secretive about their research. I know it is different in maths and physics, because as far as I am aware, people put stuff out there as soon as they get the data. Whereas, we are very 'oh no!' you know this is my baby. It has taken me years to get this data. They won't put out what they don't want people to see. So, I am not going to find anything from the research ... on the invisible web anyway because people just won't publish preliminary data out there. Because then you risk your publication, because somebody else might use your work."

Learning about the invisible web seemed useful for the participants and they gained more realistic viewpoints about the accessibility of information on the web including the information that they publish on the web. For example, one of the interviewees evaluated the visibility of her personal homepage to see whether it is visible or not and said:

PS27: "I am in the visible web and I know that because a friend of mine in Ireland searched for my website and it came up straightaway and also I think I don't have a very common surname ... not many people in my surname. So, I come up in number one or number two I think which is nice."

In general, awareness of the invisible web seems useful for the users to increase the efficiency of their searches in different ways. After explaining about the meaning of the invisible web the researcher asked the participants whether knowing about this issue makes any difference in their future searches or not. The answers were mainly positive. One of the participants declared that this awareness will help her to be more persistent in her future searches and before giving up the search she will try different routes to gain access to her desired information. She said:

RS21: "It [awareness of the invisible web] might make me not give up quite so easily perhaps. Especially...I mean I might not use a lot a general web search like Google straightaway. I might be more likely perhaps now to use something more technical database ... not worrying that if it doesn't come up in my general search it doesn't mean it is not there. But I am not sure how I would try to penetrate the invisible web."

The other useful point of RS21's idea about the invisible web is the necessity of providing web users with search instruction because she mentioned she is not sure "how to penetrate the invisible web". In contrast, another participant believed knowing about the invisible web might have a double-edged effect which may make her more uncertain about the comprehensiveness of the information that she retrieve through web searching or might encourage her to be more persistent in the future searches.

PS28: "I don't know if I am just going to get more the feeling that there is something there that I am not getting to it or maybe it is going to encourage me to search in a different way and looking more instead of just expecting to get the answer in the first place."

One of the participants had a clear statement that shows how the awareness of the existence of the invisible web might affect users' search behaviour. At the beginning of the interview with he declared that he trusts Google:

RS18: "I trust Google, if you want. So, if I type something in and Google doesn't find it I trust it is not out there. So, if I can not find it so it is not my fault then. It is not there. So, in a way it seems that I completely trust Google maybe I shouldn't."

However, at the end of the interview and after learning about the invisible web he changed his attitude and said:

RS18: "I usually do not feel that I miss information but when you mentioned all this I might think actually I do miss something."

Nevertheless, some of the interviewees believed that because of the type of resources that they use regularly knowing about the invisible web would not make any significant difference in their regular searches. For instance:

> RS25: "My first thought is that the kinds of information I get like publications and sort of current developments in research groups are not likely to be invisible because you wouldn't want to be invisible. Yes, I think you would expect them visible. The bit that I was vaguely aware of that, databases, okay I was vaguely aware of that you've made me more definitely aware and I would think okay we need different strategies to find the relevant databases rather than just a random search on Google. I have heard that some of the other search engines are better at doing different kind of searches rather than just putting in terms. So, yes, knowing the invisible web, as I said my first thought is material that I am looking for probably won't be made invisible and if it is like databases [deep web] then I am aware I would need different strategy to find them."

> RS23: "I don't think so actually. It is interesting to know that it is there but I think about the stuff that I search for a day to day I can't see it makes a lot difference really".

In addition to the benefits of knowing about the invisible web for end users, discussion about this issue with the participants was very informative for the researcher to provide him with two important issues. Firstly, some of the interviewees added new categories to the existing definition of the invisible web. For example:

> RS24: "There are more and more about what you call the invisible web. Now Google and many other dot coms [search engines] have maps but in their maps there is not everything. You can't find everything in the maps. There is some information you never find in maps, never, never in your country, in my county or here in the UK. For example, if you look at maps for the military areas, it is not there. You can pass everyday or maybe driving near the military base but when you look at the map it is not there. Maps never have military information ... so, this is information but you can not find on the web. You have the map but you can not find this information."

The same person continued this issue and made a difference between the invisible and hidden information and declared:

> RS24: "Well, I call it hidden. This is not invisible. This is hidden. Maybe the information is there but because of filter or something that stops you to see the information. Here at this university this is very interesting because if you go to the public IT centres and then if you look some information that can affect people you can not do it. For example, you can not see pornographic information. So, the problem is what is the pornography you know because if it is a naked woman or a man in the web maybe this is anatomy but maybe somebody says 'no!' this is not. So, they also cut the possibility of freedom to search the web... so, it is not completely free."

One of the participants managed to summarize the different reasons of information invisibility after hearing about the meaning of the invisible web very well.

RS20: "It would be the information that is not available to me whether because I am not allowed to get that information, or it might be I haven't got access officially to get that information, or it might be that my method of searching or the tools that I use to look at the web aren't sufficient.

The above discussions produced new directions in the research. This new direction was the role of the user in the visibility or invisibility of information. This was the "missing link" in previous definition of this phenomenon. Participants' reactions to the definition of the invisible web and exploring their successful and failed searches guided the researcher to a new idea of cognitive invisibility and eventually ended up the development of the model of information visibility on the web presents below.

Cognitive invisibility

Although the term "invisible web" was new to all but one of the interviewees, all could recall experiences of search failure. The notion of search failure is potentially much wider than that of the "invisible web" as defined in the literature review (chapter two), search failure potentially being due to a range of factors of which technical invisibility is but one. Nevertheless, it was decided to explore search failure more generally, in order to illuminate the relationship between perceived search failure and the invisible web. This line of enquiry led to the emergence of a tentative model of "cognitive invisibility" and to consideration of its relationship to the sort of "technical invisibility" discussed in the literature review chapter.

In relation to search failure, a number of different searching "scenarios" were described. These are presented below, along with short quotations from the interview transcripts to illustrate each.

1. To set the notion of failure, and degrees and types of failure in context, it is worth noting that all interviewees could recall search experiences that lacked any perceived level of failure – and indeed difficulty – entailing the straightforward retrieval of relevant information via a search engine without problems. For example, one of the participants narrated a successful search as follows and explains he managed to retrieve what he has been looking for very quickly and without facing any difficulties.

AS12: "That [search] was very straightforward and I had a very efficient and rapid answer. Basically, I got from where I was to where I wanted to be very very quickly and I guess the reason it worked well was because I knew what to do and there was no learning curve there. It was just very straightforward."

Similarly, one of the participants reported one of his straightforward searches that he had done recently by using Google as a general purpose search engine and he answered the question that was his motivation for the search.

AS07: "A friend of mine has recently been appointed to a new post at a university outside this country as a sort of chair in a particular subject. I wanted to find out more about his department and so I did a Google search in that case and typed in the name of the university department and

went straight to it. The searching process was very good, very quick it was literally straight there. I typed in the words in inverted commas and I went straight to that page and clicked on that page and went straight to the department and it was very successful."

Another example of this type of search was reported by one of the participants. He describes his successful search as follows:

AS04: "I was invited to give a presentation at a university in Canada and I wanted to find out about the person who had invited me. So, I searched [Google] for that person and found their website first or second – I think it was the first returned, first element of that list"

2. The following quotations, however, all entail a degree of deviation from a straightforward, problem-free search using a search engine. Indeed, interviewees noted that relevant information could be retrieved, of course, without recourse to a search engine, via bookmarked websites, and/or URLs given by friends and colleagues.

Although not representing search failure, these experiences are considered relevant here insofar as (a) they represent the bypassing of search engines, and (b) search engine use is central to the concept of both technical and cognitive conceptions of "invisibility" that form the focus of the study.

This category of scenario thus represents a gradation in the move from total visibility to total invisibility shown in the emerging model presented below. Two examples of this kind of search have been reported by two of the interviewees. One of them reported a category of search in which he can gain access to his desired information through different routes which are not necessarily provided by general search tools. For instance, he referred to his colleagues that can guide him to an information source by providing him the URL of the required sources.

AS10: "Well, there is one example, yes. If ... it was a colleague who told me that it is there somewhere I would then e-mail the colleague and say 'give me the website'. That is probably what I would do and I might say that for other things even scientific things if I couldn't find them ... Because someone told me that it exists ... I would e-mail my colleague and say 'Can you give the URL?' "

Similarly, another interviewee reported one of her successful searches which entailed a degree of difficulty but eventually she has managed to proceed the search through other routes which in this case again was not search engines.

AS05: "I had a real problem. I was looking for ... full spectrum lights. I'd tried to find them in the UK and couldn't, and I could only find a USA supplier and I tried quite hard and I failed. ... I couldn't find it and eventually a friend found it for me through another source – and they had a website. ... She didn't find it on the internet; she actually found it fortuitously in a leaflet she has got. So, she found it that way."

The question of whether the information retrieved could have been retrieved via a search engine (indeed, bookmarked sites could have been so retrieved originally) may simply not be known by – or indeed relevant to – the user, and thus not referred to in the interviews.

3. Ultimately successful searches could entail a degree of difficulty, relevant information being retrieved but, compared with the examples above, entailing a more circuitous route. One of the participants reported an example of this category through narrating one of his challenging search experiences in which he had spent a considerable amount of time and energy to locate an information item on the web.

> AS03: "We have been trying to get a video tape of that and it has taken ages actually. ... Google didn't help initially and Amazon didn't help, and there was another site which said, well, it would be a great cost to find it for you and I thought 'Oh! Forget it', you know, [laughing]. 'I am not doing that admission of failure', and then I found a firm which is now based in Somerset and it was delivered the other day. So, I was very impressed with that. But it took a while because that is not my familiar area of work ... So, it took a while to get that anyway but without the web we wouldn't have got it..."

Likewise the above example, another interviewee reported a search experience is which he had been fairly certain that he would not be able to locate the required information on the web. However, after spending some time on the search he had managed to fulfil the search successfully which had been very encouraging for him.

> AS09: "Well it was successful partly because I didn't think I'd be able to find the information, and because I was rather heartened and surprised that I did find it. I mean it took three or four attempts refining the search to find it."

Similarly, another participant reported one of his demanding searches that he had carried out by using different search tools including Google Scholar and Google Image. Although his search in that case eventually ended up with successful results, he considered it as a result of a 'rather roundabout route' as follows:

> AS11: "I was searching on stem cells and initially I thought I'll use Google Scholar and put in 'stem cells' and of course you get hundreds of thousands of replies. This was just too many. You know they were accurate, they were specific, there were papers on stem cells but not really what I wanted. And then I actually used Google image search for pictures of stem cells and that was much more useful because it got me to lectures that people give on stem cells, and also it got me a link to the NIH information site on stem cells which was just what I wanted. So I didn't get the first time. I got the second time by a rather roundabout route."

4. Inevitably, however, there were perceived searching failures. In some cases, the interviewee did not only fail to retrieve the information hoped for, but also felt confident that it was there to be retrieved. For instance, one of the participants believed most of the time the desired information is somewhere on the web and access to it only depends on following the right direction towards it which might require spending more time or using more appropriate search tools.

AS14: "I'm sure the information is out there, and I'm sure this is the case for most things – that information is there you just have to spend a bit longer looking around, or maybe using more directed search engines or something like that."

Similarly, another participant ascribed failure in access to his required information to his inability to carry out the search procedure properly because in his opinion the required information must be somewhere on the web.

AS06: "Frustrated I suppose, but frustrated with myself rather than frustrated with the web, because I know the information must be there somewhere and I know if I did it properly, I'd be able to find it but I just haven't been able to find it."

Indeed, there are more examples of this type of feeling among the participants. For instance, another participant attributed his failure to the inefficiency of his search strategy and states as follows:

AS10: "I think sometimes when I am frustrated and have not found what I want, I know that somewhere there must be a page with this on but I just can't find it. I would not necessarily blame the search engine, because I don't particularly know why it is not finding it. ... Maybe I am not putting the correct words in to enable the search engine to find it. And I would try to think of different words to put in, or different things to type in there. So, I will experiment and try a little while."

5. Sometimes, however, interviewees were very unsure whether or not the information was "out there" and retrievable with greater effort or skill on their part. In some cases, the balance was more towards the feeling that the information probably was there. For example, although one of the participants considered himself as an experienced searcher, he was not confident enough to ascribe his failure in search to the lack of information and said:

AS09: "Well, although I do a lot of searching on the web, I don't necessarily think that I am the world's best at it, the world's best searcher. I spend a lot of time searching the web and in some respects I am quite professional at it, but if I fail I am more inclined to think that is because I haven't adopted the best search strategy rather than it not being there – that is my first thought in most cases. And if I do it intensively and I'm still not getting anywhere then I progressively begin to think that obviously it is not there. But if you spend 20 minutes or something half an hour looking for something and you don't find it my first reaction is not to blame the web. It is to blame myself."

Alternatively, the balance may be more equal between both possibilities. For example, two participants expressed their uncertainty about the reason of failure in search as follows:

AS07: "But the problem is I don't know if it is there and it is the search engine which fails, or if I am putting the wrong information or the information isn't there."

AS12: "I don't always find what I am looking for and one of the biggest challenges is ... it's hard to know, you never know, whether you can't find it because it is not there or if you can't find it

because your search method is inappropriate. ... You don't know when to stop, when to give up, because you don't know if you are wasting your time and there is literally no information out there online or whether you just not searched appropriately."

6. Towards the other extreme, whilst not actually knowing the reason for certain failed searches, interviewees reported occasions when they felt confident that information was not "out there" indexed and ready to be retrieved. For instance, one of the participants believed that failure in search means the required information does not exist on the web and he confidently ascribes the failure to lack of the required information. The following quotes show this type of attitude of failure in web searching.

AS13: "Well, it was never a failure because there was not anything to find. So, if there is nothing to find, that is just as useful because it tells you there isn't any information. ... So, that is a failure because probably the specific information that I am asking for just is not there. I mean often times when we are using genetic information you do a search with genetic information and you don't find what you are hoping for, but that is not a failure it just tells you that you were hoping for something that didn't exist".

According to this participant's belief if the required information exists on the web there is no reason to be hidden from his view. In particular, he highlights the fact that he usually looks for specific information on the web which is usually stored in reputable resources and if a search ends up with failure it just means the required information does not exist on the web rather than being hidden.

AS13: "If I am doing searches for information on a certain topic and I can't find, I mean, whatever keyword you put in Google you always get something but you might not find the information you want. So, yes it does fail on occasion but that is probably because the information isn't there rather than being hidden."

As AS13 has this attitude he does not seem persistent to keep looking for the required information after facing initial failure in search because in his opinion failure in search is the result of lack of the required information. Therefore, comparing with other participants that are uncertain about the existence of the required information he might give up easily:

AS13: "Failure is probably mostly generated by the fact that the information just isn't there. Then so what? We just carry on. So, it is not a big deal."

Figure 6.1 shows a categorisation of these different scenarios in the form of a "matrix of cognitive invisibility", beginning in the top left corner with total visibility and going on to include gradations of invisibility. It is important to note that the notion of "invisibility" depicted here is different from what was discussed in chapter two about the literature-based definition of the invisible web. The matrix shown in figure 6.1 represents the subjective views of the interviewees. That is, their perceptions of whether information they were searching for was in reality indexed and available for retrieval are independent of this reality, and thus independent of the technically-oriented models of the invisible web previously discussed. This

introduces a cognitive subjective, as opposed to a technical objective definition of "invisibility".

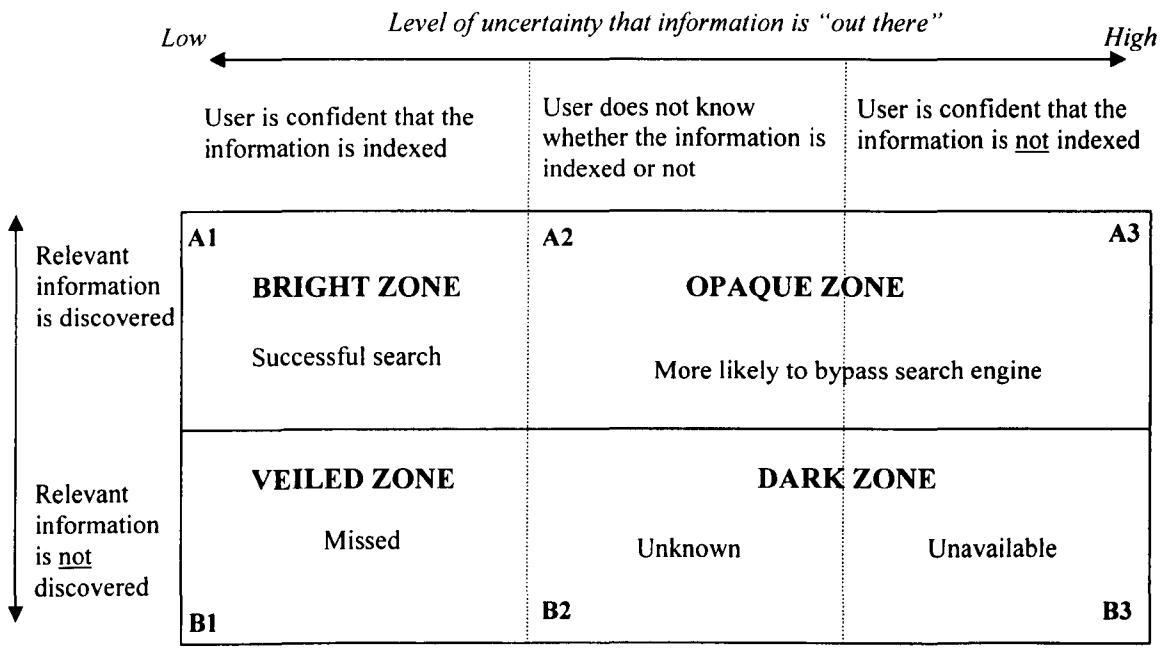

Figure 6.1: Model of information visibility on the web

Cognitive model of information visibility

Cognitive model of information visibility is the first product of this research that contributes to information seeking area on the web through offering a conceptual framework which enables us to understand users' interaction with the web regarding the visibility of information resources. This model suggests a comprehensive view of the issue of information visibility/invisibility by dividing the whole web into four zones.

In fact, regarding the literature-based definition of the invisible web the whole web is divided only into two parts of the visible and the invisible which is not adequate to address the real meaning of this phenomenon. This research offers a more sophisticated perspective about this issue through the model of information visibility on the web which consists of bright, veiled, opaque and dark zones.

Bright zone

In zone A1, the searcher searches and finds relevant information using a search engine without problems. This zone is termed bright since the search is fully illuminated with no hindrance. This zone shows the most straightforward kind of search on the web which entails the least difficulty to locate the required information. For example:

AS05: "Usually I am looking for papers, specific papers relevant to the research I am doing. So, I want to check a paper or see if something has been done. That is generally what I am doing and also I am retrieving the same piece of information multiple times because I have found it before and I know it is there. I check it again."

There are a number of real search scenarios about the participants' past search experiences that show some examples of searches that can be considered in the bright zone. For example, one of the interviewees had carried out a simple search in Google to find out the scientific name of a specific plant and she managed to fulfil the task without facing any difficulty.

PS28: "In general, I find it really easy. I usually use Google ... for example, yesterday I was looking for the scientific name of a plant, Euphorbia Antisyphilitica, and I knew but I wasn't sure. So, I just typed the common name, Candelilla and Euphorbia, and it just looked up a lot of resources."

Similarly, the same person had carried out another search which can be mapped into the bright zone when she has been looking for images of another plant and succeeded the search without any difficulties.

PS28: "For my research I needed some pictures for that presentation. What I just typed was picture of that species, Mammillaria Lenta that is a cactus, and I had many different options form different pages or addresses to go ...and I found good pictures and used for my presentation."

Sometimes a search session might consist of a few stages but the whole procedure goes smoothly and though the user may rephrase the original search strategy it is still very straightforward and can be considered in the bright zone. For example, one of the participants explained one of her successful searches as follows:

RS21: "I need to present a journal article tomorrow in our lab meeting ... I used various search terms for topics that our research group are interested in. So, I could find a recent paper that I think they would be interested in and I have successfully found something quite quickly that I think it would be useful. Initially I went to PubMed which is a literature searching website and I typed in various keywords for the topic. One of the words that I typed in was Dystroglycan because we work on this protein and I came up with a selection of papers and obviously they were ordered in terms of the most recently published. I looked at a couple of the abstracts by clicking on the links and decided I didn't quite fancy doing neither of those. So, I went back for a different search term. I think I searched for Filopodia and I came up with the list of papers. One which I looked to the abstract ... I thought it might be interesting. I decided to do that one and I used the link from the abstract to the journal which actually brought up the actual paper which I could then download as PDF. So, you know that was in a matter of ten to fifteen minutes. So, you can go from searching to actually getting the article you want because they are very well-linked between them."

The same person narrated one of her everyday life searches as well. As she explained in the following quote, she easily had managed to book a holiday accommodation through a simple web searching without experiencing any particular difficulty.

RS21: "Yeah, last night we were looking for a camp site on the Isle of Skye. So, we started out using the Visit Scotland website and we used their search engine for accommodation. You type in where you want to go. So, we could choose. We want to go to Skye and we just chose that. We wanted camping accommodation and it brought up whole list of different camp sites which if you click on the link it tells you more information about the actual campsite. Then some of them had web addresses that you could click at links which was the actual campsites' own web address and [by] doing that we managed to find somewhere that we quite like so."

The above examples shows most straightforward searches which can happen successfully. However, web searching is not always as straightforward as these examples and sometimes it entails more steps to take.

Opaque zone

In zones A2 and A3, the searcher finds relevant information perhaps by means other than a search engine. The question of whether or not the information could have been found via a search engine may be unknown and/or irrelevant to the searcher (A2). Alternatively, the searcher bypasses a search engine because s/he feels that the information is unlikely to be indexed (A3). These zones are termed opaque since illumination takes place in a roundabout way in relation to a search engine, which is bypassed. In opaque zone (A2) the user is not sure whether the required information is indexed by search tools or not. However, s/he manages to fulfil the search through the web successfully either via search engines or other means.

One of the participants narrated one of her search experiences in which shows an example of A2 in which she has not been sure whether her required information is available on the web or not and eventually she succeeded the search task through using different search tools.

RS23: "Recently ... I had to give a talk for the department and I was making slides for my presentation and I knew that I needed certain, or it would be helpful to have certain, diagrams. We work on the development of DNA in the fish and I wanted to see if I could find diagrams of the DNA in mammals in also in the fish. Because a small problem that I have, it sounds silly, but I can't draw, even using things like Carol Draw or Paint. I can't even using the mouse. I find it very difficult to do. So, it is just easier for me to actually download pictures and the other problem is in the lab we don't have a scanner. So, we can't scan anything from textbook. So, I went to Google and Yahoo and particularly Google Image search which is very useful tool and it is becoming more useful as time goes on to find the actual diagram that I wanted. It took me probably about an hour to find everything that I wanted, in the exact format that I wanted,...but I did actually find the diagrams that I was looking for and it was just easy to copy and paste some into PowerPoint. Then that was it, I was done, rather than me spending hours to actually attempt to draw these things myself."

Access to information in opaque zone usually may require passing through a multi-stage search. For example:

PS28: "Yeah sometimes I actually have to search like that. You type the word. It takes you somewhere and from there you start making a more like a deep search in the things that didn't come up first."

Access to information through colleagues is an example of opaque zone which provides the user with the required information without using search engines. For example, one of the participants said:

RS17: "There is enough information and just because of a sheer volume of information you know you are going to miss most of it ... and I may rely on colleagues to say "did you see this?" word of mouth still also I think is important. When using the net I mean people send sites to each other and say this maybe useful to you which is a way of cutting down the time that take to find something. So, I think there is a growing culture of that in order to be more efficient in a community particularly in biology I think you get a lot of 'I found this, this is useful for you and it doesn't take you long time to have a quick look' and it get from all over the world people send stuff and that is very useful because I don't have to search [laugh]. And I think there is a reason behind that to be more efficient as a community searching for things helping each other and this way is quite useful."

As has been mentioned earlier, in A3 the user is confident that the required information is not indexed. However, s/he carries out the search to either confirm his/her belief or to find out whether s/he might end up with a different result. In fact, when users search the web they are looking to discover information which is not necessarily entail retrieval of some documents. In contrary, sometimes users search the web hoping they will not locate any document in their search topic.

For example, one of the interviewee narrated the story in which he had expected that the search will fail but he had done it to confirm his belief and he concluded that his initial belief was right. He said:

AS15: "I went to a conference couple of months ago. We work on a certain subject and it is totally unexpected the results that we've had which tells something that the organism that we were working on is unable to do something that we didn't expect to. So, I've looked at the background literature and there isn't any because I've looked around the subject and found other organisms can do so. I have looked through in some details and I went to a conference and this guy said to me: 'Oh! Yes I know about so and so and [I] just published papers about that'. So, I wrote down that person's name and I put them through PubMed yesterday and there wasn't any information at all [laugh]. So, therefore the question is have I written the spell of the name wrong or did this guy gets it wrong? My conclusion is probably this guy got it wrong and he was confused in what I was saying to him."

Moreover, another interviewee reported an interesting example of his successful searches in which there are more explanations about this kind of search. Interestingly, the quote below is the interviewee's immediate answer to the question 'can you remember and describe one of your recent search experiences in which you were successful in web searching?' He answered:

AS13: "I did a PubMed search yesterday looking for a couple of keywords and I didn't pull out any papers, but it was successful in the fact that I didn't pull out any paper because there wasn't any information about the subject that I was interested in finding. So, a negative [result] in that respect is actually quite useful because when you are working in science and you're working on things and you're broaching into a new area that you don't know very much about and you want to find out information and then there might not be any information. So, a negative result is sometimes illuminating as a positive result."

Similarly, another participant explained a rather unusual but interesting kind of search success in which the user hopes to retrieve nothing. In fact, this kind of search which leads to no result always is considered as failure in terms of information retrieval (IR) point of view. Nevertheless, in terms of the context of search and users' viewpoint this is a success. Indeed, this is not "retrieval of information" that determines users' conception of success/failure in search. This is "discovery of information" which enables the user to discover something to enhance his/her knowledge and then feel successful.

In fact, in the above case the relevant information that the user had been looking for and was discovered by him was a confirming proof that no paper exists about the search topic. Therefore, in this case retrieval of "no result" is perceived as success for the user. However, if the user was confident that the required information is not indexed but s/he carried out the search hoping to find some information, in that case it could be considered as failure. Nevertheless, as this study was concerned about users' perceptions of success and failure, the above case is considered as successful search. Similarly, the quote below shows another example of users' perceptions about this kind of success in search.

RS19: "I am so confident with the Web of Knowledge. It is good. If I only get back six papers on a certain topic one of my responses is either ...that has not been researched very much and then it is either Oh! Dam! I wish I had the information that I really want, and if I am writing a grand Oh! Great! No one has done it yet, brilliant! ... I mean obviously for papers you write for grants you have to be absolutely certain [about] what you've said."

Veiled zone

In B1, the user searches expecting to find information via a search engine but realises that s/he has failed to locate it. This zone is termed veiled since information is felt to be there but somehow concealed from view. For example, one of the interviewees referred to one of her search experiences in which she had failed to buy her desirable flight ticket because of failure in locating the required information. However, she was sure that what she had been looking for is accessible somewhere on the web:

PS29: "It is a failure because I know something is available but I can't get it. Also I expected that I am able to translate the page that is one of the functions of search engines but I can't. Obviously, I didn't get the ticket at the end and this is a failure. It hasn't fulfilled my purpose of this search."

As participants in the study stated this is fairly common to receive a number of search results pages in most search sessions but for whatever reason including time limitation they do not follow the further pages and usually look at the first or second page. This is also supported by the previous research that usually users just look at the first few pages of the results (Spink, 2003). One of the participants referred to this fact and mentioned that sometime search engines locate what she is looking for but it might be out of her access because it is at the bottom of a very long list. However, the search engine has found it and she makes a difference between this kind of failure and when the required information has not been found by search engine at all.

> PS29: "Well, if it is in the twentieth page of your result you have got the result but you don't access to it. It is different from when it is not in your twenty pages of your search result."

On the other hand, this zone includes the search scenarios that the user fails to employ the full capacity of a search engine through selecting inappropriate keywords and s/he misses the relevant documents. What one of the interviewees said in the following quote is an example of veiled zone when she is confident that her required information is indexed but she does not have enough patience to carry on searching to find it.

> RS17: "Oh, I think it [information] is there. I think it is there. I just give up on the search. I think everything you need is out there. I mean even now cutting edge research is automatically put out there ... I can't think many things that aren't there except what I am looking for in science [laugher]."

Similarly, another interviewee said:

> AS02: "You've gone to the first ten pages of Google [and] you still haven't got what you wanted, but it doesn't mean it is not there it could be there. You just need to be a bit more patient. That failure is either my fault because of the way I have tried to search, using the keywords that I have chosen, or it could be the other person using words that ...Google recognition sequences ... give them priority ... and that happens quite a lot I think."

In the veiled zone the user is confident his/her required information is indexed by search engines and what s/he needs to do is employ an effective search strategy to locate the desired information. Therefore, sometimes users might continue the search to move from veiled zone to the bright zone and sometimes they might give up the search. For example, one of the participants explained how being aware of the existence of the required information encourages her to continue the search.

> PS29: "If I search for your name on the web, just as an example, and it doesn't appear in the search engines, but I know for sure that you are in this university and you are a registered student, because I know for some reasons that you explain about the invisible web, I'll actually go on to this university address and I'll search through the departments and then search for you."

In contrast, sometimes though the user is confident the required information is available on the web for different reasons s/he might cease the search in the veiled zone. For example, one

of the interviewees reported a search experience that he had been looking for an information item that in his opinion could be found by spending more time.

> PS37: "Yeah I was looking for information about a specific strain of E.coli that I was interested in and probably the main reason that I wasn't successful because I didn't have much time. So, it wasn't a very thorough search."

Similarly, sometimes the user knows that the required information is indexed by search tool but s/he does not have enough patience to carry on the search to access the required information and it remains veiled for him/her. One of the participants stated an example of this situation.

> PS28: "Maybe I don't have the patience to keep searching until I get all the things that I could get from there".

The importance of search strategy was obvious for most of the participants in this study and they were aware that the type of search terms that they use makes a big difference in the search results. For example:

> RS26: "The most important reason why failure occurs in web search engines is when the keywords used for the search are generic terms that are applied in either to many different fields or within a given field can have a lot of applications or even if it has only one field and one application it is so widely used and you have one thousand for the same thing and you never know exactly what you are looking for."

Another possibility of missing the required information comes back to some elements that are out of users' control. For example, access to full-text of a considerable number of academic journals requires fee-based subscription which regarding Sherman and Price's (2001) definition of the invisible web is part of the proprietary web[28]. For example, one of the participants mentioned this issue as one of the reasons of his failure in many search experiences which has stopped him from access to the required information. However, he is confident that the information is accessible on the web.

> PS37: "Sometimes I can't access a PDF [full-text] of an article which I want because the university doesn't subscribe to it which can be frustrating. I think most often times that I don't find is when the university doesn't subscribe."

There are more examples in the data for the veiled zone. For examples, one of the participants had failed to fulfil a search task and she explains the search scenario as follows:

> RS21: "I remember trying to find a protocol for a particular technique which I wanted to do [for] a particular experiment. I tried to find a protocol online and although there was lots of links

[28] Detailed explanation about the proprietary web and other parts of the invisible web based on the existing literature is available in chapter two (literature review).

...they weren't very detailed or I couldn't access to the page where I tried to click on the link because it was on a restricted server or we didn't have access to that particular thing through the library."

In terms of conceptualization of this failure she continued in the following quote and declared that in her opinion there are two reasons for this failure. Firstly, lack of a specialised search tool which can enable her to search specifically in the area which she requires. Secondly, the limitation of the university subscription for online journal restricts accessibility of information for her. Therefore, as she knew her require information exists on the web but she was not able to locate what she wanted. This example also can be located in veiled zone.

RS21: "There hasn't been such good central and searchable database of methods, like literature we have got searchable databases or DNA or protein sequence as a searchable database. But there is only recently people started to put together searchable databases of methods and sometimes these are subject to subscription and if the university isn't subscribed with it you can't access to that particular database. So, you have to look elsewhere."

Sometimes the user knows that the required information is indexed because s/he has already seen it somewhere and now wants to revisit it. However, another search does not lead to it. For example:

RS27: "To be honest now you talked about it I have come across that through maybe when I have searched for something and I have not come across it whereas I know I might have already known this person has got a website about that but it hasn't come up on the search engine. I didn't realize it is a common thing."

Dark zone

In zones B2 and B3, the user fails to find relevant information but either simply does not know whether or not it was there to be found (B2), or feels that in all likelihood the information was not there to be found (B3). The latter situation could pertain to a search specifically aiming to confirm that no relevant information is "out there" – for example, on a new research topic. Alternatively, the conclusion that there is no information to be found may be drawn after the search. Presumably in the first case, there was at least a small suspicion that there may have been information, otherwise the search would be unlikely to have been conducted in the first place.

One of the interviewee gave an example of the dark zone (B2) in which she failed to find her required information. During the search procedure she has encountered two down websites which she supposed possibly possess her required information. Therefore, she was not sure whether the information is available on the web or not.

RS17: "I was looking for mass spectra, a mass spectrum from mass spectrometry, spectrum of a particular compound. I do have some chemistry websites that I looked on for this and I found it but I couldn't find the breakdown products which are very specific thing ... and I don't know if it is out there yet. But I was not successful to find the breakdown products or information relating

to the breakdown products. So, that was an unsuccessful search even with the specific keywords you know the "compounds" itself and "mass spectrometry" [which are] quite specific keywords but I was unsuccessful. Now whether that there were two sites which were down ... but they may give me the information but they were not available at that time."

Sometimes users are confident that a failed search is the result of lack of the required information. These cases are located in the dark zone (B3). For example, the following quote shows an example of this area of the model.

AS13: "When I am looking for things, if I don't get anything then I am pretty certain that it is not there ..., because most of the information on the web is a load of rubbish for my purposes because it is not specific, it is not peer reviewed, or whatever. So, if there is really useful information for me it is going to be on a reputable website where you can expect journals and these kinds of things. So, it is not going to be one of these things which are difficult to find. So, I think if you can't find stuff then generally it is because it is not there."

Discussion on the model

The matrix of perceived invisibility is arguably a useful complement to more objective technical conceptions of invisibility in that the latter may not reflect users' experiences of information visibility/invisibility in their daily web searching. This derives from the fact that what constitutes the objective invisible web is a moving target since it is subject to constant change. What is invisible today may not be so tomorrow. As one interviewee put it:

AS03: "You look at it this morning and this web looks like this and then look at it tonight the web has grown, the hub's grown and some more things and more connections but equally some have gone."

The capability of search engines is also constantly being improved, affecting objective visibility. For example, file formats that were not crawled some years ago are now visible to many general search engines. Also, each user's search skills are likely to change with experience over time. For this reason, it is important to distinguish, and to be clear about the relationship between, what is termed in this thesis the objective technical invisible web and the cognitive subjective invisible web. A simple example illustrates the distinction. A user searches for but fails to find information on a topic. In reality, there is no information "out there", but the user feels confident that there is, and that s/he has failed to find it (category B1 in figure 6.1). This example does not relate to the objective invisible web, since in reality there is no web-based information "out there" rendered invisible by, for example, search engine indexing practice or policy. However, it does relate to the perceived invisible web, since the searcher clearly feels that s/he has to some extent failed to find what was there to be retrieved.

This is arguably important in that there is interplay between (a) searchers' perceptions of search failure, and (b) the objective reality of search failure. This interplay may affect searchers' confidence, and has implications for training. (a) may not relate to (b), and vice versa. (a) may help to determine a user's confidence, and her or his motivation and perceived need for self improvement thus impacting on training. (b) may affect the nature and extent of

training that may be required to render the user a more effective searcher in that training should reference the "reality" of a searcher's skill as well as her/his perceptions.

Therefore both objective and subjective conceptions of the invisible web, and interaction between them, may be useful. It is likely also that the perceived importance of a failure is a factor in such processes – failure to find information in relation to a relatively trivial enquiry having less impact than failure in a search of significant importance to the user.

Drawing such a distinction between the objective and perceived web enables some mapping of a searcher's skills onto "reality". This mapping may be useful in terms of helping the searcher to develop appropriate expectations, confidence, and improved search skills. It is acknowledged that an objective assessment of any individual's "actual" invisible web may be unknowable with any accuracy, the nearest approximation being the assessment of an experienced web searcher. Nevertheless, such an approximation may still be useful. For example, where a searcher to perceive that search failure was not due to any lack of skill on is/her part, confidence may be beneficially affected – or, at least, detrimental effects may be avoided. However, if we are to help people develop skills that are likely to result in more effective searching, we need to map such perceptions onto as realistic a view as possible of these skills. Thus more objective assessments may be useful in the process of helping people develop more effective search skills. Such a mapping may be instrumental in helping to ameliorate both excessively negative misperceptions on the part of a searcher in relation to her/his abilities (which may lead to a lack of confidence and motivation), and excessively positive misperceptions (by suggesting the need for appropriate self-critical reflection).

Further remarks on the model

The categories shown in the figure 6.1 matrix (a) were derived inductively from the interviews, and (b) account for all the relevant interview data, in the sense that no contradicting or competing data has been ignored. Whilst it is argued here that the matrix validly represents the range of search experiences reported by the academic staff, research staff and PhD students interviewed, the research did not seek to investigate the frequency with which any of the six types of search experience depicted in the matrix occurred, or are likely to occur in any wider population. Rather, it aimed to chart the possibilities, or identify what Olaisen (1991) has termed "sensitising concepts". The function of this type of research is to "map the territory" and provide at least a tentative scheme that could form the basis for more systematic study.

Such subsequent study could, for example, investigate the validity of the categories in different and/or larger samples, and seek further to illuminate, challenge, extend or refute the categories. Once the model was thus further developed, the frequency of occurrence of different types of experience in different populations and issues of generalisability could also be investigated using statistical methods. The particular group of users studied here exhibited differences in the extent to which they attributed search failure to "internal" or "external" causes – i.e. to their own lack of ability to find relevant information or to the unavailability of information to be retrieved. Although the notion of "invisibility" applied to the web meant nothing to the great majority of the academics interviewed here, it is nevertheless arguably

useful as a metaphor for describing differences in users' search failure perceptions, as in the matrix presented above.

Whilst as a metaphoric term, "invisibility" bridges the objective technical search engine realities that may bring about search failure, and subjective perceptions of such failure by users, like any analogy it exhibits differences as well as similarities as used in relation to these two areas of study. However, potential confusion is arguably outweighed by the power of such a common denominator to emphasise the close links and inter-dependencies of technical and cognitive aspects of search failure, and the importance of mapping such links and inter-dependencies to support education and training to help people become more effective in their information seeking.

Summary

This chapter reports the findings of the research that address the early research questions which were about end-users' perceptions of the invisible web and how literature-based definition of this term map onto users' perceptions. Moreover, the chapter reports how the research moved from the technical definition of the invisible web toward a deeper and broader explanation of this phenomenon.

The chapter shows that existing definitions of the invisible web are not sufficiently comprehensive to describe this phenomenon and we need to develop a new definition for this issue to cover its different aspects. Accordingly, the study offers the concept of "information visibility" to make a distinction between technical objective conceptions of the "invisible web" that commonly appear in the literature, and a cognitive and subjective conception based on searchers' perceptions of search failure.

This model shows how the term "invisible web" can be developed to a new concept called "level of information visibility" or "cognitive invisibility" which was developed based on an inductive analysis of the qualitative data in this study.

CHAPTER SEVEN:

CONCEPTUALIZATION OF SEARCH FAILURE

CHAPTER 7: CONCEPTUALIZATION OF SEARCH FAILURE

"Failure happens all the time. It happens every day in practice. What makes you better is how you react to it." Mia Hamm

Introduction

Emergence of the information visibility model - which has been illustrated in chapter six - provided the researcher with a new conceptual framework to address the PRQs[29]. This chapter presents the results of addressing PRQ1 in the light of the model of information visibility. Therefore, there is a strong link between what is presented in this chapter and what has been reported in chapter six. To remind the reader, PRQ1 is:

- How do web users conceptualize their information seeking failure on the web?

Over the course of the study this question played an important role to keep the researcher focused on the main theme of the research. In fact, end users' conceptualization of search failure had a key role in all stages of the data collection and analysis. Therefore, the whole contents of this chapter are devoted to the findings about this issue. However, it is important to mention that although the main focus of the current research began with a focus on users' conceptions [30] of failure in search, since the early stages of the data collection the researcher realized that exploring users' conceptualization of failure depends on their perception of success as well.

In other words, success and failure are two sides of the same coin and it is not possible to consider either without paying attention to the other. Therefore, the chapter not only focuses on users' perception of failure, but also it reports the findings about their conception of success as well. However, still the main focus of this study was on users' conceptualization of failure.

Two aspects of the conceptualization process

Data analysis indicated conceptualization of success/failure consists of two main aspects which are:

- What is success/failure?
- Why does it happen?

[29] Primary Research Questions

[30] Two words "conception" and "conceptualization" have been used in this thesis interchangeably. However, conceptualization implies the process of reasoning and conception is the product of conceptualization.

These two aspects emerged since the early stages of data collection when the researcher asked the participants to narrate one of their recent successful and failed search experiences and then asked them to explain why they think these specific cases are successful or failed. This question was perceived in two ways by the participants both of which were illuminating for the researcher. Some of the participants supposed they should explain their opinion about the differentiation of success/failure in search and some of them attempted to explain the causation of success/failure in search. However, the researcher let both groups to express their views freely and then he covered the other aspect through some follow-up questions. Therefore, now the research findings enable us to address both aspects of the conceptualization of success/failure.

The first aspect is differentiation of success from failure in web searching and the second aspect is causation of success/failure in search. Further exploration of the second aspect led to emergence of EmRQ1[31] which was explained in chapter one and the results of addressing this question are reported in chapter ten.

Differentiation of success from failure

Participants in this study expressed their ideas about the difference between success and failure in web searching and it was a rich ground of exploration which provided the researcher with a deep insight into a key theme of the research.

Achieving the search goal was a common definition of success. Most of the participants frequently declared that if they achieve the goal which they had in search they will consider it as a successful search. One of the participants explained his definition of success in a very simple but informative way as follows:

> AS01: "If I find the information that I need for a particular purpose that is fine. That's success and if I don't that's failure."

Another participant also illustrated his definition of success in a very similar way. He said:

> AS10: "Well, success is where whether I find what I am looking for. If I am looking for a definition of a word, if I am looking for the structure of a compound for instance or something like that then that is the success, because I found exactly what I wanted to find.

The same participant continued his illustration about the meaning of success and failure and said if he does not retrieve what he exactly wanted then it will be failure.

> AS10: "If when I am searching I find things that the search machines correctly found but it is not what I want and then it is no use to me then of course that is what I call a failure. So, it is not a failure of the search engine. It is a failure of my search approach to find what I wanted."

[31] The first Emergent Research Question

As the above quotes show discovering the desired information determines whether a search session should be considered as successful or failed. If a search session ends up with satisfying the searcher's information need, it will be considered as success and if it does not lead to this stage it will be considered as a failure. However, further defining features emerged form other interviewees' responses. For example, in AS15's definition of success he considers the amount of time that is spent on a search as a leading factor to differentiate success from failure.

> AS15: "If I get the piece of information that I'm after [is success]. Just as simple as that, I mean success would be finding the answer within fifteen minutes. I would think that is successful. I have got the information that I wanted, I've done it in fifteen minutes preferably within five or even two but within fifteen I'll accept. If it takes me more than half an hour to find what I am after then I would suggest even if I found it, it's not very successful use of my time."

Moreover, in the early days of data collection the researcher identified that there is not a binary picture about the success and failure and users consider their conceptions of success and failure on a continuous scale.

> AS02: "I don't use success and failure. There is a middle road. So, successful is I've achieved my aims, failure is there is nothing to come up and more frequently so much information comes up and it takes too long to go through it all. So, you are not able to come to a satisfactory conclusion. You actually just take the first or a few that they are on offer....So, success is definitely where I get exactly what I want and yeah and there is a way between the grounds."

> AS14: "If you are looking for a small piece of information and you can find it very quickly that is very easy success to find a phone number and that is a sort of routine success, isn't it? I mean, I think it is not an absolute failure, it is very difficult, you do one search and it is not anything and it is not there and it doesn't... does it? It is a time-related failure, it is sort of spectrum of failure you can't find it instantly, you can't find it easily, you can't find it with lots of looking, you can't find it after a week exhaustive searching. It is very different kinds of failure, isn't it? And I don't have the patience to search that exhaustively. I mean if it is not there then immediately I'll look at something in a different way."

As the above quote shows this interviewee envisaged a scale in which one extreme is absolute success and another extreme is complete failure. However, in his opinion and based on his experience the frequency of partly successful searches is more often than absolute success or complete failure.

Causation of success and failure

Exploring the users' perceptions of the causation of success and failure in web searching was a rich ground of exploration in this research. Therefore, there is a detailed explanation about users' viewpoints of the reasons of success and failure in their web searching. The next section, which presents a matrix of success and failure in web searching, illustrates the findings of this study about this issue. A conceptual framework has been developed in the matrix format to address users' perceptions of the causation of success and failure in search. The matrix is an extension of the model of information visibility which was illustrated in chapter six.

Matrix of success and failure

The matrix shown in figure 7.1 illustrates how the first model presented in chapter six about information visibility can be extended to develop the model on success and failure in search. In fact, the matrix integrates the concept of information visibility on the web with the concepts of success and failure in search.

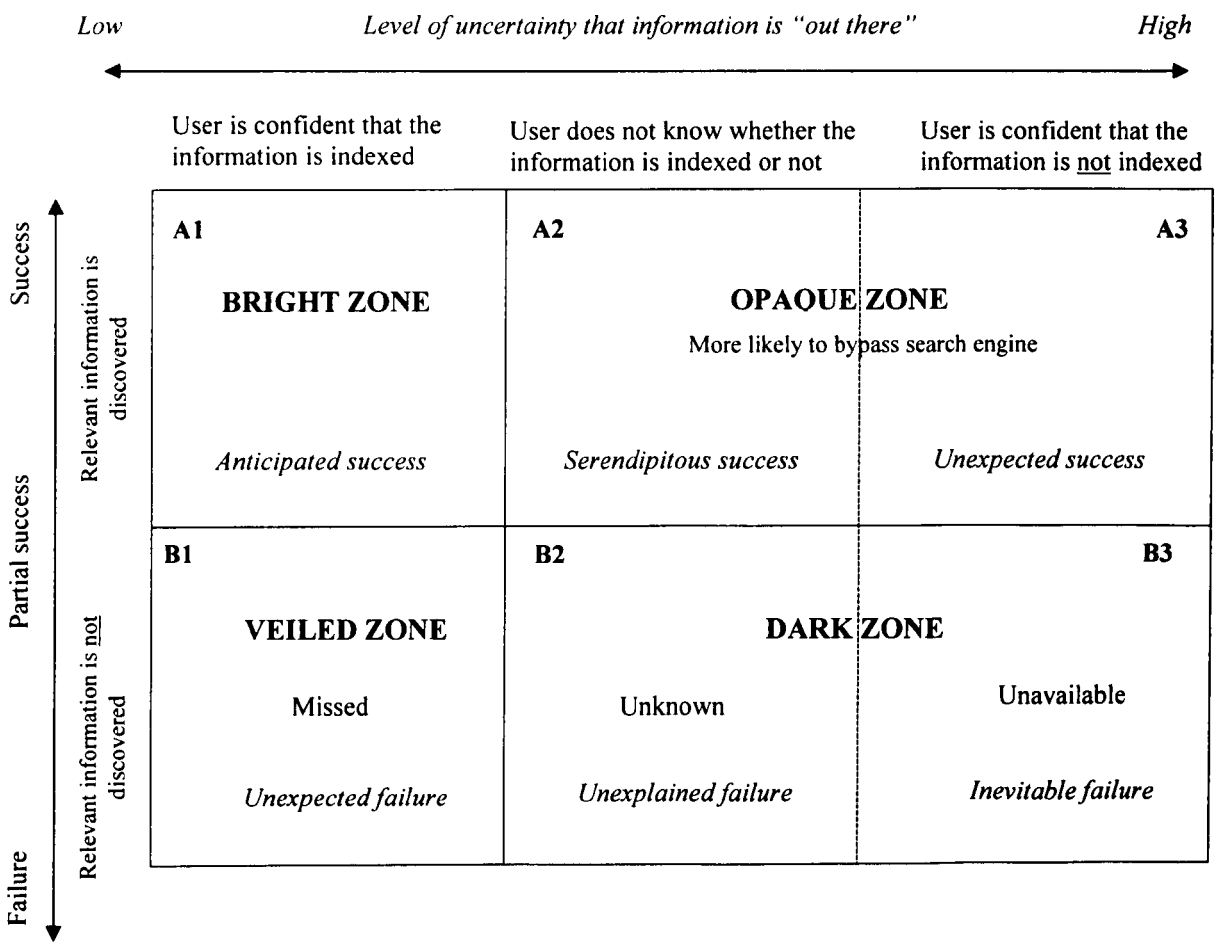

Figure 7.1: The extended model of information visibility integrating the concepts of success and failure in web searching with the concept of information visibility

Three types of successful searches

There are six cells in the matrix. All these six cells can be divided into two major sections including upper and lower areas. Three cells on the upper section have an important characteristic in common. They include search scenarios whereby users eventually discover what they look for and based on the data when users retrieve their required information they consider it as success. Even if they spend a lot of time and energy to fulfil the search task successfully they still consider it as success. One of the interviewee believed:

RS21: "I would say success in general is just being able to find the information that you want eventually. So, whether it takes you ten minutes or three hours, as long as you find it I would say successful. Whereas, failure is after a long period of time getting to the point that you actually give up because you just have spent far too much time doing it and being unable to actually find something."

Therefore, the upper section of the matrix is the success area. Moreover, based on the matrix there are three kinds of success which differ from each other as follows:

Anticipated success

In the "bright zone" users are confident that the required information is indexed by search tools and should be available, and they achieve their search goal. Therefore, they have already anticipated that they will be successful in the search task. This type of success is termed as "anticipated success" in this study. For example, one of the participants mentioned:

RS17: "I went to the Sheffield [university] homepage, to the library, to the e-journal table, to "N" for Nature, pull up Nature online, use their browser with an author name, put in the author name, get the journal, get PDF, download it and print. And that was successful because I knew where I am going there."

Another example of this kind of success is explained by another participant. In the quote below the interviewee explicitly declared that she had already been aware that her required information exists on the web and she expected to be successful in her search. She said:

RS23: "[I was successful] because I was looking for something which I know that other people are working on ... I know that it is covered in a lot of, say, American medical school courses and they tend to have a lot of their course information online for students. So, it is available to anyone to just go in and pull things out of there. So, I think because what I was actually accessing was a probably a fairly large pool of information and fairly well-studied."

Users' previous experiences in search help them to anticipate success in search because they have already carried out similar searches and they are aware that what kinds of information are and what kinds are not available on the web:

RS26: "I like formula one and I am interested in Spanish driver Fernando Alonso and I am interested the last modification they made to the car that he drove in the last racing Silverstone you can find the information. I can find information about refrigerators, I can find information about computers, I can find information about biology, I can find information about throat cancer, [and] I can find information about anything you want ...Also there is an important limitation is that not all countries have equally open web pages to be able to access to for instance companies in the US they are very open, apart from some, you can get into the company, all the jobs that are available, the profile of the company, the products that they have, you have the all price list of all the chemical products that this people are selling.

Serendipitous success

In zone A2 the user is simply not sure whether the required information is available or not. Nevertheless, s/he finds what is required. This kind of success can be called as "serendipitous success". For instance, one of the interviewees reported that for a specific search he had not been sure whether his required information is available on the web or not. However, he had been aware that for that specific search he requires a general search engine rather than a specialized database and he managed to succeed the search through a general search. He reported as follows:

> AS10: "The point is NCBI[32] is good if you know what you want. If you want a gene or medical publication or something about genetic disease it is fine. But if you don't really know where to look then it won't work. You got to look for specific things within NCBI whereas AltaVista you can just do it a random with anything. I am a writing a review on fungal metabolize at the moments and the genes involving making them. Sometimes I need a structure and not sure what the chemical structure is for this compound. Well, again I wouldn't know where to find the chemical structure for the compound. I could look for a compound in a journal, find a journal, look through several articles about the compound, [and] look to see if the authors put a figure in for that compound. You could do that which takes time but the other day when I just wanted the structure of some chemicals. I simply typed the name on there with AltaVista again and with the top few hits it had catalogues ... and I just clicked on there and getting the diagram describing the compound and where it is from and a picture of the structure on there."

Unexpected success

In zone A3 despite the user's certainty that the required information is not indexed by search tools he/she finds it. This kind of search is very interesting and can be called "unexpected success".

Obviously this kind of success is different from serendipity in information seeking because in this kind of access to information the user locates something that is not supposed to be found.

> AS09: "I was looking for some detailed information on the blood supplier to kidney just a few days ago mainly to confirm what I've already read in elsewhere. I am pleased to say that I found from an American site quite a detailed description of it together with some nice pictures ... yes, well it was successful partly because I didn't think I'd be able to find the information and because I was rather heartened and surprised that I did find it. I mean it took three or four attempts refining the search to find it."

[32] National Centre for Biotechnology Information http://www.ncbi.nih.gov/

Three types of failed searches

Similarly zones B1, B2 and B3 have an important feature in common. In all of these zones the user does not find the relevant information and he/she fails to conduct the search successfully. Therefore the lower section of the matrix is the failure area. The failure area of the matrix is divided into three zones compromising "Initial Failure", "Unexplained Failure" and "Inevitable Failure".

Unexpected failure

In zone B1 failure happens mainly because the user had supposed that his/her required information must exist on the web. Nevertheless, s/he did not manage to succeed the search unexpectedly. For example, as the quotes below show, unpredictably the search did not lead to success.

> AS07: "I looked up her department and found the department but couldn't find her. I looked up her name in the search directory but although I knew how to spell her name, I knew her initials she wasn't in there. Because although she works there, she works on a contract, and she is not a permanent member of staff and I couldn't find her."

> RS25: "I suppose in that case I had a very definite aim. I wanted this equation which I thought somebody must have created in the past ... that was a failure and I didn't find anything."

This kind of failure is "unexpected" rather than anything else because the user does not expect his/her search fails but it does. For example, one of the participants in the quote below believed that most of the time when he fails to carry out a search successfully, it happens because of his fault in adopting an inefficient search strategy. He said:

> AS09: "If I fail I am more inclined to think that is because I haven't adopted the best search strategy rather than it not being there. That is my first thought in most cases."

Similarly, another interviewee also blamed the inefficiency of his selected search terms and she said:

> RS17: "I think that is my problem. I don't use specific enough search terms."

In the same way, the other interviewee believed that what he required in a specific search must be available somewhere on the web. Therefore, if he fails to find it the reason is because she did not choose an appropriate search term. She said:

> RS22: "Possibly it is hidden and I am not managing to put in the correct search terms or not looking in the right place."

Likewise, another participant thought that her search failed because she had used an inappropriate search tool and her required information is accessible if she uses an appropriate search tool. She said:

RS23: I just think, I maybe have just looked in a wrong place. It was quite a surprise that I didn't find it. I just need to look in a couple of other places.

Unexplained failure

In zone B2 the user fails in search but it is not clear for him/her what the reason of failure is. This is an "unexplained failure". As the bellow quote shows, in this case the user was not confident whether what he had been looking for was accessible on the web or not.

RS25: "Maybe that doesn't exist but I suspected hardly half belief that I think it probably does exist. Somebody must have done this before but may not be available through the web. It might be buried in a research paper somewhere."

In another example which cab be mapped into zone B2 as well, the user is uncertain whether the required information does not exist on the web and his persistence in web searching will not make a difference in the result or he just need to modify his search strategy to succeed the search. He said:

AS12: "I don't always find what I am looking for and one of the biggest challenges is ... it's hard to know, you never know, whether you can't find it because it is not there or if you can't find it because your search method is inappropriate. ... You don't know when to stop, when to give up, because you don't know if you are wasting your time and there is literally no information out there online or whether you just not searched appropriately."

Inevitable failure

In zone B3 which is the exact opposite of zone A1, the user is certain that the required information is not available on the web and this kind of failure is called "inevitable failure". The below example shows that the user knew that because he had been looking for something which was too specialised the search failed and this failure had been expected by him. He said:

AS15: "I think for this particular search example yes. I think it was most likely that actually nobody has produced a website or a webpage with that information on ... I think it was highly specialized. I was looking for an aspect of the theory. You know, it is a very specific area of the theory and probably a limited number of people in the world are studying this aspect ... it wouldn't be something for instance that somebody would perhaps put in a webpage for undergraduate teaching. It was more specialized than that. So, it is limiting the number of people who would be interested in ... I would believe nobody bother to construct a webpage to explain this aspect of theory."

Another interviewee was confident in some other occasions that when he fails to find what he is looking for it just means that the required information does not exist on the web. He said:

AS13: "Well, it was never a failure because there was not anything to find. So, if there is nothing to find, that is just as useful because it tells you there isn't any information. ... So, that is a failure because probably the specific information that I am asking for just is not there. I mean often times

when we are using genetic information you do a search with genetic information and you don't find what you are hoping for, but that is not a failure it just tells you that you were hoping for something that didn't exist".

In a similar way, another interviewee knew that what he had been looking for on the web falls in the category of personal and private information and he will not be able to gain access to it. Although in this specific case which was about telephone number of people he believed that this kind of information should be publicly accessible but he knew this is not the case for everyone. He said:

RS24: "It was a failure well because it is private information and people consider it as private and they don't want to put it on the web. The address and telephone number or e-mail address of scientists has to be public but some people keep it private."

Equally, another interviewee knew that his required information, about a particular species of Orchids, should not be easily accessible on the public web because of the specific feature and importance of this kind of information in protecting rare plant species including Orchids. He declared:

RS16: "I think the most logical reason for that [failure] was people are quite protective about giving information about where to find specific plants particularly Orchids because it is a rare plant. So, they are worried about the collectors."

Different dimensions of search success and failure

The above matrix illustrates the main categories of users' conceptualization of success and failure in web searching. However, web searching like any kind of human computer interaction happens in a "context" which needs to be taken into consideration to understand users' behaviour.

Exploring the context of search in terms of users' reaction to search failure led to identification of different properties and dimensions of search failure. Table 7.1 (at the end of this chapter) summarizes a number of properties and dimensions of search failure. The identified properties are severity, time allocation, frequency, uncertainty, predictability, inevitability, complexity, measurability, subjectivity, adjustability, tolerability, functionality, feelings, attribution, and experimentation.

The above properties of search failure can be considered for more exploration in further research. However, regarding the limitations of this thesis it was not possible to probe into all of them in any more detail. Nevertheless, the first two properties including the severity of failure and immediacy of search have been selected for further exploration which is presented in the matrix of importance and immediacy of search in the following section.

Matrix of importance and immediacy of search

Identification of different aspects of failure in context enabled the researcher to explore the phenomenon of information seeking failure on the web through new perspectives. For example, considering two aspects of search including the immediacy and importance of search led to development of another matrix shown in figure 7.2. Consequently, four types of searches emerged through the further analysis.

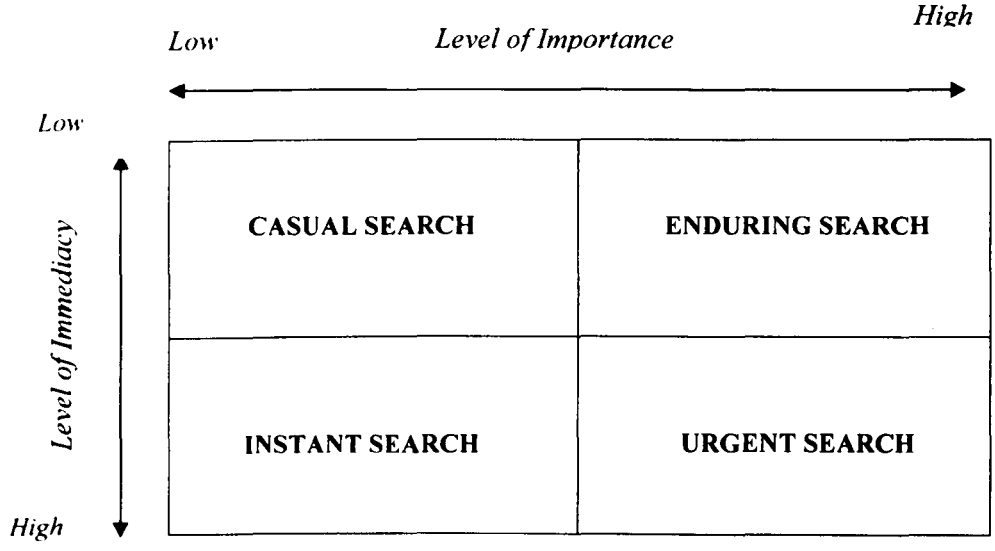

Figure 7.2: The matrix of time and importance in each search scenario

Casual searches

In casual searches the search is neither particularly urgent nor important. This kind of search is carried out to answer trivial questions which usually arise through users' simple inquisition as a matter of curiosity. For example, one of the interviewee declared:

> RS25: "I think for answering idle queries it [web searching] is often very useful. My daughter had a word in her spelling test in school which wasn't in our dictionary. We have got quite a big dictionary. So, I typed that into Google and it came up with a website which is about words and it happened to have an entry for this particular word which explained how it arose. It was a sort of unusual word and sort of half slang word which obviously hadn't a main entry in a dictionary. It was 1990s edition. So, when I work like that it is extremely useful in a way that would be either hard or very time-consuming to pursue that query in any other way."

In the same way, another interviewee reported another example of casual searches:

> AS12: "Occasionally, I would use web-based resources in response to something that I have read, just happened to stimulate my interest in an area or something scientific that is completely beyond what I normally study."

Enduring searches

In "enduring searches" which usually includes a series of successive searches the level of search importance is quite high but it is not very urgent and user has enough time to deal with the search. For instance, one of the participants said:

> RS23: "It depends on how important I perceive that piece of information to be. If it is something that would have been interesting, for instance to do with my project, ... not crucial I might just forget about it ... I'll probe again through a few weeks time or couple of months time and it might be there next time because things, you know, appear and reappear. If it is something that I really need to know I will try different avenues."

Similarly, another participant reported that when he fails to conduct a search successfully he will search again in sometime later:

> AS03: "I mean if I really can't get it and I just know I have seen it somewhere that is very frustrating. ...so I don't like it if I am not successful and I know something is there. So, I certainly swap between my search engines and then leave it and come back another day and then think about another approach."

Sometimes a specific reason stops the user to fulfil the search successfully and s/he has to postpone the search for a while. For example, the quote below shows that this participant try to avoid using Web of Knowledge because of the heavy traffic in the network and postpone his search for a suitable time:

> AS06: "In certain times of the day Web of Knowledge is quite slow, end of afternoon it is not ever good, but it I try not to use it in the afternoon, and then actually when it is really busy there is a limit to the number of people using it at the same time from Sheffield University. So, occasionally it didn't just let me in so really too many people using it ... yeah, [I] just go away and come back half an hour later again."

Instant searches

"Instant searches" are not very important but the user wants a quick answer to them. These searches are usually part of a bigger line of enquiry and they are important because a more important task depends to them. For example, one of the interviewee declared as follows:

> RS26: "Well, it depends on how important is what you are looking for. So, if it is something very important that I can not find it that would be a problem. For instance, I wanted to buy plane ticket to go with my mum to Rome. The thing is quite likely I would be able to find good prices searching the net. I have already done some. It is possible you may not find good prices because it is high season and it is going to be expensive. So, it is rarely to find a cheaper [ticket] because the time of the year which is not any cheaper but I still look at the failure it is just the nature of the beast."

The same person also explained another example that shows sometimes the user has to stop a search because of limitation in time. However, the search is not also very important and the immediacy of result seems more important but it might become a challenge for the user:

> RS26: "Let's say I am looking for the plane ticket to go to Prague for a cheap price. Then I realize I can not find the plane ticket because I can not find the right pages because I do realize that there might be some other companies there are some smaller plane companies from Prague to somewhere else. So, the most important thing ... is not just the fact that I was not able to find the plane ticket in that case because that is the nature of my mind ...that is the intellectual challenge that I think I can do it and then you don't have the time or not able to spend time on it and then you have to drop it ... people who ... spend five hours on the net to look for a plane ticket and they have plane tickets that costs 55 pounds, 50 pounds, 45 pounds, 47 pounds ... they spend five hours on the net to find the cheapest ticket but the difference between them is actually less than ten pounds."

Another example of instant search is presented in the following quote. In this search the user requires an immediate answer to a question which does not seem very crucial:

> AS04: "I was invited to give a presentation at a university in Canada and I wanted to find out about the person how had invited me so I searched for that person and found their website".

Urgent searches

Urgent searches are the most crucial type of search. The level of both immediacy and importance are high in this kind of search and users require immediate answer to very important questions. Of course, the level of importance is perceived by the user which depends on many elements in the context of search. For example:

> AS07: "Last week I was trying to find a telephone number of a colleague who works at the Sheffield hospital. I knew her name, I knew her department, I knew her initials but I wanted to get either a telephone number or an e-mail address because I needed to contact her quite urgently."

> AS08: "Yes, I do fail. I do fail to find things on the web, whether that's because they aren't there or whether because of I don't have time to find things. Well, actually probably both of these things occurred. You know, there are sometimes when you are looking for a particular piece of information and the time pressure like review a lecture, for example, or writing something that needs to put in by the deadline."

Summary

The chapter encapsulates the main findings about users' conceptualizations of successful and failed searches. What is presented here is a summary of a long time interaction with the dataset to explore users' conceptualizations and perceptions of success/failure in web searching.

The chapter shows firstly that success and failure in search are closely interconnected and it is not possible to address them in a totally separate way.

Users usually have a perception of the results that they have received. They might perceive a search procedure in a spectrum varying from absolute success to total failure. Data analysis shows any search procedure which does not lead to satisfaction of users' information needs is considered as an unsuccessful search. This might happen after receiving no result, ambiguous results or too many results. When the number of search results is too many to identify the relevant documents among them it is still a failure in terms of end users' viewpoint.

Some of the interviewees thought they fail in some cases because their required information for that specific topic does not exist on the web. They neither blame the search engine nor themselves. They believe the information does not exist at all, or if exists it has not been uploaded on the web yet. In contrast, some of the participants believed the search failure happens because of their searching incompetence.

The chapter also shows that conceptualization of success/failure happens is multi-dimensional context which needs to be taken into consideration to understand users' behavior. Exploring different aspects of failure led to identifying different dimensions of failure in web searching which is presented in table 7.1.

Table 7.1: Dimensions of failure in web searching

Property	Explanation	Dimension	Example quotations
Severity	Severity shows how important a failure might be for the user and differs from very sever to insignificant.	From vital to trivial	AS07: "I was very anxious because it was important. It was in a different country and I didn't know where to go. So, I got very very frustrated."
Time Allocation	Time allocation shows how much time the user can spend in order to fix a failed search. The user might have sufficient time to carry on the search or the search might be urgent.	From urgent to negligible	RS17: "But it is a crucial point these days there aren't enough hours in a day anyway. How much time can you afford to spend searching the web? You have to limit it."
Frequency	Frequency shows how often failure happens regarding the total number of search sessions.	From very often to rarely	AS03: "I suppose if I am getting 90 percent success everyday. So, failures do come up reasonably regularly even then."
Uncertainty	Users are not certain whether they are experiencing a failure or not.	From well convinced to highly uncertain	AS12: "... you don't know when to stop when to give up because you don't know if you are wasting your time and there is literally no information out there online or whether you just not searched appropriately."
Predictability	Failure may or may not be predictable for many reasons. Sometimes users have already known this is very unlikely to succeed a specific search task and sometimes they can not predict whether failure might happen or not.	From highly predicted to absolutely surprising	AS10: "I think sometimes when I am frustrated and have not found what I want, I know that somewhere there must be a page with this on but I just can't find it. I would not necessarily blame the search engine, because I don't particularly know why it is not finding it. ... Maybe I am not putting the correct words in to enable the search engine to find it.

Property	Explanation	Dimension	Example quotations
Inevitability	Sometimes the user is convinced that the required information does not exist. In contrast, the user might hope to find what s/he is looking for.	From unavoidable to highly preventable	AS11: "If none of them (search strategies or search engines) worked I don't know what to do. I mean because the only other resource really is looking on the literature. And if it is not in the literature well, you know, I think it comes [to] the time when you have to say really I am sure whether the information is available or not."
Complexity	Failure in search can be very complicated regarding many factors that play direct or indirect role in the search context.	From highly complicated to very simple	AS03: "I never understood that really because I thought the question and details I have put in were appropriate and it just failed ... Google had trouble to find what I wanted and I didn't define it obviously correctly. So, that was once quite recently because I have to do a lot of searches ... but that was a quite a strange example. I suppose in the end you don't really look at it very much as a failure. I must have done something wrong and move on but it was a kind of strange thing."
Measurability	Although it is not easy to measure the degree of failure, users consider failure on a scale and try to assess the level of failure.	From highly measurable to non-measurable	AS02: "I don't use success and failure. There is a middle road. So, successful is I've achieved my aims, failure is there is nothing to come up and more frequently so much information comes up and it takes too long to go through it all. So, you are not able to come to a satisfactory conclusion. You actually just take the first or a few that they are on offer....So, success is definitely where I get exactly what I want and yeah and there is a way between the ground."
Experimentation	If users fail to achieve their search goal, they usually seek for alternative ways.	From various alternative ways to no experimentation.	AS10: "I would try to think to different words to put in or different things to type in there. So, I will experiment and try a little while."
Attributions	Users ascribe their failure to either internal factors or external factors.	From only internal attributions to only external factors.	RS17: "Possibly because those two sites were down. And I think I was doing it for about twenty minutes and I just think that was enough time to not get anywhere because there're sort of things to do. So, probably I gave up on a time limit so I think the time was and the fact that the couple of sites were down contributed to that unsuccessful search."

Property	Explanation	Dimension	Example quotations
Subjectivity	What determine a search is a failure or not is users' perception about it rather than an objective measure.	From highly subjective to highly objective.	RS17: "To him it [web searching] is a hobby, whereas to me it is a part of the job. So, the way that he approaches that is going to be different. You know computing is his big hobby ... anything to do with this machine ... to him is like a pleasure. So, he learns the search as part of his learning process. But I just use that to get my stuff. To me it is a chore, a task, a job that has to be done. So, it is not so pleasurable."
Adjustability	A failed search can be fixed by the user by making some modification in the initial search strategy.	From highly adjustable to the least adjustable	AS03: "I have three major search engines which are quite different really, the three I mentioned earlier on. So, I would use the other two. If all of those failed and then I will sit back and think about the question again and there are two ways for a question usually. One is the general topic area or the other one is identifying a person and that would be particular relevant when I am doing my journal work looking for referees."
Tolerability	Users might tolerate and accept a failure or for any reason can not accept it.	From highly tolerated failures to unbearable failures.	AS03: "That is very frustrating because we work very quickly. I mean there is a time constraint there, but I need five people to say referee a paper if after three or four minutes I have not really found a person, we'll try somebody else. So, there are alternative people. There aren't just five referees in the world for one paper and only five there are clearly expectation of possible people in that case. I'll move on. I need to do it quickly and I can't hang about."
Functionality	Failure in search may be perceived by the user as a useful event because it helps him/her to enhance his/her search abilities and learn new techniques or learn more about the search topic.	From very useful to absolutely useless.	AS04: "I mean it is a failure in the sense that one search didn't [work], the first search for example, didn't return the desired information, but if you like it is loosing the battle but the war is a bigger thing ... it is only part of the process ...in terms of the interim goal unsuccessful, but it terms of the process of reaching your ultimate goal was useful sometimes."
Feelings	The affective side of failure plays an important role in users' reaction to the failed searches.	From highly frustrated or highly concerned to absolutely unconcerned and indifferent.	AS10: I'll get quite annoyed. Yes, get annoyed and frustrated. Because usually you don't have one go you try one way and it doesn't work and you try another way and it doesn't work and you try the third way and that doesn't work and you say ah! And you just give up and use some other methods and give up on that, so very frustrated.

CHAPTER EIGHT:

OVERLOOKED INFORMATION

CHAPTER 8: OVERLOOKED INFORMATION

"You miss a lot of opportunities by making mistakes, but that's part of it: knowing that you're not shut out forever, and that there's a goal you still can reach." Johnny Cash

Introduction

This chapter presents this study's findings about users' awareness of and reaction to missed information while searching the web. The issue of missed information relates to both the invisible web which was discussed in chapter six and users' conceptualization of search failure which was presented in chapter seven.

Firstly, it relates to the invisible web because in terms of the definition, the hidden parts of the web are the resources that are overlooked by search engines and consequently will be missed by the end users who employ these search tools. Therefore, the phenomenon of the invisible web exists as a result of overlooking some information resources that could be used by end users. Of course, in real search situations there is always a possibility of missing information in the visible web as well. Therefore, any information item – either in the invisible or the visible part of the web - which is missed by a user, or which a user assume that s/he has missed, is related to this issue.

Secondly, it relates to search failure because sometimes users fail in satisfying their information needs as a result of overlooking their required information that they could retrieve. Nevertheless, in a broader sense if overlooking information does not lead to failure in search it will reduce the search performance in different degrees. Accordingly, in any case if users miss some information it negatively affects final search performance.

The chapter at the outset reports the results of inductive analysis which enable us to addresses PRQ2 and PRQ3[33]. To remind the reader, these questions are as follows:

- PRQ2. To what extent are web users aware of missing some information resources on the web?
- PRQ3. If they are aware of missing relevant information, how important is it for them?

Four identified categories

Exploring the collected data to identify categories of the participants' awareness of and reaction to missed information led to identification of four categories, namely (a) perceived volume of missed information, (b) perceived importance of the missed information (c) awareness of information likely to be missed (d) reaction to missed information.

These four categories suggest two matrixes. The first matrix which is presented in figure 8.2 is called in this study the "matrix of search impact" and the second one, presented in figure

[33] The second and third primary research questions of the study which are discussed in chapter one

8.3, is termed the "matrix of search depth". Before presenting the matrixes there are brief explanations for each category below:

Perceived volume of the missed information

Data analysis indicated that the participants believed that while they search the web they miss some information anyway. However, key aspects of this reality were the amount of missed information and its degree of perceived relevance. What they miss on the web locates on a scale which varies from missing a large amount of highly relevant information to a small amount of less relevant information. In fact, web users are concerned about two aspects of missing information which are its quantity and quality. The quantity refers to the number of missed information items and the quality implies the degree of relevance of the missed information to the user's information need.

As perceived relevance of a document to the user's need determines how important it is for the user therefore the relevance of information can be considered as equal of its importance for the user. Therefore, the more relevant document contains the more important information. In fact, importance of information is a subjective concept which is determined by its degree of relevance to the user's need.

Perceived importance of the missed information

The second category is the perceived importance of missed information. Users want to know how important/relevant the missed information is. They need to know how important missed information is to take an appropriate action afterwards. If they perceive missed information as too important to be missed then they might spend more time to chase it up.

Awareness of missed information is highly dependent on the searcher's knowledge about the search topic. There is a level of certainty about the missed information. For a less familiar area they are more uncertain if they missed something or not. When the user is uncertain about the importance of missed information his/her reaction is different. S/he might be worried that what s/he missed is a "really seminal work" in an "obscure journal" or a "sleeping paper" [AS01] which for some reason has been missed.

Awareness of information likely to be missed

The third category is users' awareness of the likelihood of missing relevant information. This category shows how confident users are in terms of the comprehensiveness of their search results. As the straightest answer to this question, data analysis showed the participants in this study were aware of missing some information:

> AS05: "You will miss things for sure, yeah. I don't think it is a 100 percent ...I don't think it is an absolute retrieval. We do miss some information."

However, there is a tricky aspect about users' awareness of missed information. Indeed, if something is missed then it will be very difficult to have an exact idea about it. Nonetheless,

this study did not intend to evaluate the recall and precision of users' web searching. In contrast, this research dealt with users' perceptions about missed information to explore their search behaviour on the web. Therefore, users' perception of missed information might be matched with the reality or not which was out of this research scope. Nevertheless, it will be enlightening if in future another study explore how compatible is users' perception of what they miss on the web with what they actually miss in reality. This issue is recommended as one of the suggestions for further research in chapter twelve.

In fact, in real situations people rarely can be absolutely certain about the information that they overlook unless they have already learned about the existence of it and when they search for it, it does not appear in their search results again and then they will be certain that the information must be missed. Accordingly, participants in this study expressed their uncertainty of missed information. For instance, one of the interviewees expressed his uncertainty about the comprehensiveness of search results and said:

> AS02: "Generally the search tools I use are very good, but I don't have any way of verifying them. So, say for example, I ask a question when I want to know how many of these there are in the world and it comes back and says 333. Is '333' correct? I don't have independent tools to compare. So, that can be a problem. So, I might think that I am correct. I might trust the tool. But it might be wrong and I know these tools make mistakes."

As another example, one of the participants used the interesting phrase of "lurking doubt" which shows users' feeling of the possibility of missing some information. He said:

> RS25: "When I was looking for the manufactures of air pollution measuring equipment I did always have a lurking doubt for a while that I didn't look at some prominent and important companies."

In general, the study suggested that there are three kinds of awareness of missed information. The first type can be called "certain or firm awareness" in which users not only are sure that they have missed something, but also they know what they have missed. For example, they might have already found a particular document on the web and know it exists but they search for it again it is not in the search results and they know what exactly has been missed. For example, one of the interviewee said:

> AS02: "Today for example, I went to Web of Science. Head of the department asked me about the papers that a PhD student had written and I could not remember them all. So, I said I'll phone you back. Go to the Web of Science, do a search and comes back with two papers. Now I know that person has done seven papers and I know what the answers are but today it did not come up with seven papers. So, that is an error. I only know that because I could verify."

The second type is "fairly certain awareness". In this situation users know they might have missed something but they do not know what has been missed.

However, because of their knowledge about the search topic or because of their previous experiences, they know there should be more information on the topic and they missed some relevant information.

> PS37: "I think it is likely. I always miss something. I don't think you can ever find everything. ...no database is completely exhaustive. Is it? ... I reckon that always there is a high possibility of missing something and even not missing very much I am sure you always miss something."

Finally, in the third scenario the user is "uncertain of missed information" in which the user has no idea whether s/he has missed something or not. In fact, s/he has no criterion to evaluate the comprehensiveness of the search result. These three kinds of awareness can be shown in figure 8.1.

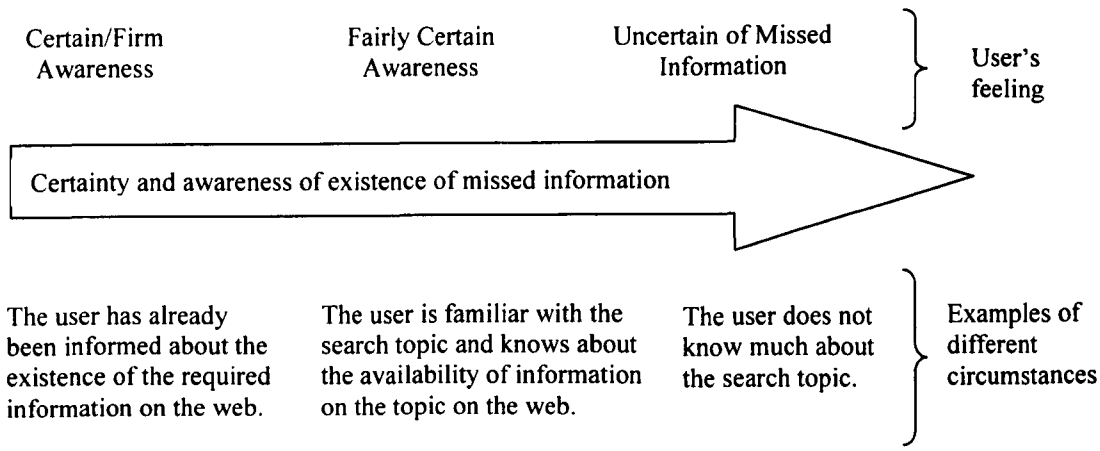

Figure 8.1: Users' certainty about missed information

Regarding the above debate information missing has two aspects. One is information that users know exists but has been missed and the other is information that s/he does not know about its existence. When users are certain that their required information exists it helps them to delve into the web until they find it through whatever route. For example, one of the participants said:

> AS05: "There are two sorts of different issues. One is there is the information you don't know you've missed and there is the information that you are looking for but you can't find. And so the latter category doesn't happen very often. I usually can find in some way what I need. But the information that I don't know I missed, there is nothing to do about that ... I have missed because it didn't come up in the search and I just don't know it exists and the information that I need to know but I can't find."

The same participant expressed her idea about the second type of missed information and declared:

AS05: "You keep creeping around with search with keywords until you pull up the paper. So, it is kind of trial and error because sometimes I would use those keywords to find a paper that I know it exists but I can't remember who the authors are ... and occasionally when I have a lot of time I do a very broad search and scroll down through pages and pages of references and that is really helpful because you see things that you normally would miss."

Reaction to missed information

The fourth category which was identified in this research about missing information was users' reaction to missed information. The study sough to discover how people react to the phenomenon of missed information and it led to developing the fourth category. This category shows how persistent users might be in chasing up missed information. The data showed users' persistence to pursue missed information highly depends on their perception of the importance of missed information.

PS28: "It depends on the case because ... sometimes I really find what I am looking for ... So, if there is more information out there but I don't need, it I think it is all right. But if I don't find what I want, then I'll get the feeling that I am doing something wrong ... in the Chihuahuan desert image example, if I could have got it from the internet it would have saved a lot of time but I didn't. So, in that case it matters. So, if there is more pictures of other species that I was looking for ... I don't mind. I've found what I like. In the case of the paper [referring to the earlier parts of the interview] that was really annoying, that was annoying that I couldn't find the paper of the ecological niche one ... I really needed that and I thought maybe I am not looking correctly, but I thought I was looking in the correct way and I spent a lot of time looking for the paper."

Exploring the participants' awareness of and reaction to missed information ended up with emergence of the following matrix which is presented in figure 8.2.

Matrix of search impact

The matrix of search impact has two dimensions including (1) users' perceived volume of missed information and (2) their perceived importance of missed information.

In each search session each user has an idea of the volume of the missed information. In one extreme s/he might be very confident that s/he has not missed any relevant information and in another extreme s/he might be very worried that s/he has missed a huge amount of information.

Indeed, when we talk about missed information we need to know how much and how important information has been missed. Regarding the effect of information missing on the search performance, we can envisage a scale to show the four possibilities.

There are four zones in the following matrix, figure 8.2, including "inconsequential zone", "functional zone", "damaging zone" and "disastrous zone".

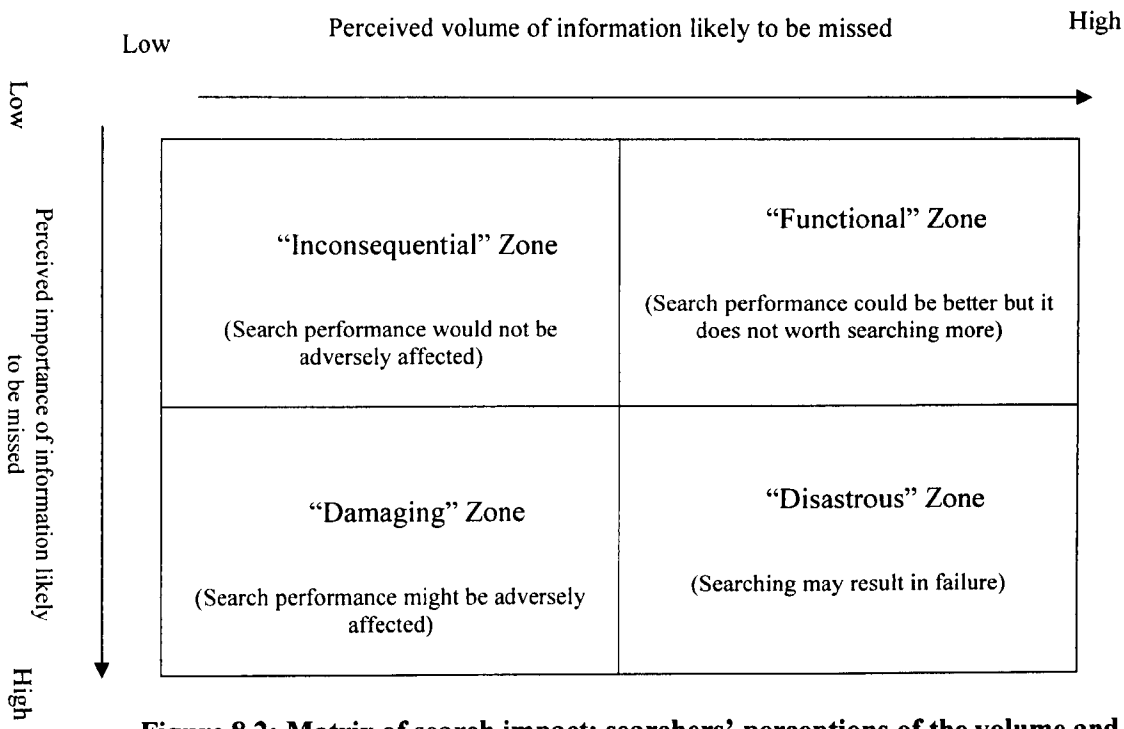

Figure 8.2: Matrix of search impact: searchers' perceptions of the volume and importance of possibly missed information

Inconsequential zone

In the best situation, only a small amount of less relevant information is missed which does not affect the search performance considerably. An example quotation of interviews is presented as follows:

> AS13: "There are probably things that it [the search tool] misses but not too many. I don't think mainly because we work on certain organisms and it pulls up everything about them. So, it pulls up a lot of stuff which is completely spurious and tedious and in fact the numbers that I select to go into EndNote each week are probably about four or five out of the 150, but in fact that is a good hit rate I think".

Another interviewee pointed out the role of users' competence of using search tools and constant interplay with the literature to minimise the amount of missed information. She commented:

> RS21: "I think if you search the literature and the web regularly then there are very few things that you would miss on your research topic. But I think if you get out of the habit of checking regularly then you will miss various publications or whatever in your search topic, but if you're good at using the internet, which is part of your training as a scientist to learn how to be able to use the web. ... You have to learn to how to use the internet and it is very important. So, I think if you learn how to use it effectively I think most of the time you will get most of the information that you want and you miss very little."

Functional zone

In the functional zone overlooking some relevant information does not lead to search failure but the result is not as satisfactory as it could be if the user had not missed those resources. In this zone missing the relevant information does not affect the search results because either there are alternative resources which are able to satisfy the user's information needs or the amount of the relevant found items is sufficient to achieve the search's goal. One of the interviewees described this kind of search as follows:

> RS20: "Basically I can usually get enough information to do my job. I might be able to get more information which might be enable me to do my job better but at the moment I'm still able to do my job well enough with the information that I get. So, it is sufficient. The word would be 'sufficient'. I think that describes my experience of using the web. It could be better but it does the job."

Similarly, another interviewee said:

> AS04: "I would say it is probably like 70 percent coverage if you like. So, 30 percent chance that I would miss something, but I am not sure how important the thing that I missed would be."

Sometimes users do know that they overlook some relevant information but they ignore it for a variety of reasons including the lack of sufficient time. Simply they are not able to spend more time to look for it or do not afford to pay the required fee for access to fee-based resources. Similar to the conceptualization of failure users' awareness and apprehension about missed information is affected by the ultimate goal of each search. For example, one of the participants said:

> RS20: "I often find enough information that answers the question that I have. So, if I have a particular task which might be to find out about funding opportunities I can usually find some funding opportunities. Sometimes I would like to find more but I suspect that they don't necessarily exist ... I am aware that I don't have all the answers and I don't get all the possibilities. I don't get all, but I get enough for me to carry out my job ...I know there is some stuff that I can't find but I don't worry about it. If I feel like that I have got eighty percent of what I need then I move on and just do the next thing. So, I don't try to get every last drop of information."

The same interviewee used an interesting metaphor of 'low hanging fruit' to describe his reaction to missed information. He explained that when he has sufficient amount of information to fulfil the search's purpose, he does not look for more information which access to them require extra time and energy. Although he is aware that his search could culminate in more results, he would stop searching after having enough information. He said:

> RS20: It is like, have you heard the metaphor 'low hanging fruit'? It is like there is lot of low hanging fruit. There might be a lot of apples on the top of the tree but I don't need to get them because there are lots just hanging. I don't need to get my ladder."

Based on the search aim users envisage a saturation point for each search. Obviously the saturation point is not the same for every search and it considerably differs in different circumstances. For example, a saturation point should be much higher in a literature review search. One of the interviewees used the phrase 'reaching the plateau' which seems an appropriate expression for this context.

Accordingly users might be aware of missing some information but they do not continue searching anymore because the amount of collected information is sufficient to accomplish the search satisfactorily.

> RS25: ... in the end I found I think six or seven companies and I seem to have reached the plateau in finding new ones. I didn't find any more. So, I thought well, that is probably the representative selection. I would not get bothered looking for anymore."

Damaging zone

In the "damaging" zone users are concerned mainly about the importance of the missed information rather than its quantity. If the number of relevant documents is high then failure to retrieve a few highly relevant items might not affect the search performance drastically but if there are only a few highly relevant resources to fulfil the search successfully and the user misses the few possibilities, it might lead to failure in search. Accordingly, in the damaging zone search performance will be affected regarding the abundance of the relevant information. The quote below shows an example of this situation. The interviewee mentioned because of the sheer volume of information on the web, even about her research topic which is very specific, she might miss a lot of information while searching the web but it does not worth searching for all possible relevant information. She said:

> RS17: "It is obvious that I miss lots. Again because although I am saying there is not a huge amount [on her search topic], there are still enough to keep you busy for half a day every day. There is enough information and just because of a sheer volume of information you know you are going to miss most of it ... so, I know I miss lots."

Similarly, another interviewee referred to the vital importance of search comprehensiveness in some specific situations in which he should make sure that he has not missed any information which might be very important. For example, he mentioned when he applies for research grants he needs to cover the related area as comprehensive as he can. In particular, when the numbers of retrieved items are not considerable he requires searching again to find out how comprehensive his search is to minimise the possibility of missing any relevant item. He said:

> RS19: "...there is a feeling that Oh! There are only six papers. Oh! I must have missed something. I mean obviously for papers you write for grants you have to be absolutely certain [about] what you've said. ... then I will do some more related searches on related words just to check it doesn't bring up any other stuff ... I mean there is only a feeling that I have missed a little bit not a big feeling that there is only six papers on this, I am really missing a lot. It is more like, well there're only six papers better to do some check to see if there is only six."

Disastrous zone

In the disastrous zone the user might miss a large portion of highly relevant information. This is the most severe situation of overlooking relevant information which might lead to search failure. One of the participants compared two search scenarios that in the second of which she was concerned that she might miss a very seminal item of information.

> PS29: "I would say it depends on what I am searching for. Like say for the price of a ticket if I think this is good enough even though there is a cheaper ticket okay I missed the cheaper ticket but that is not as such as a matter as you're missing a really important dataset that everybody knows about it except you."

Another interviewee mentioned even nowadays despite the availability of new methods of access to information, sometimes gaining access to some information resources might be very difficult. As he was working in a very specialized field of biology he knew that missing some information items might has a deleterious effect on his search because of the high importance of the missed information. He said:

> AS11: "On the research side, I think it can be a problem if you miss things that you know are important to your protein you work with. Part of the problem is that some of these things, even now there is lots of journals, older articles in journals existing on paper that is very very difficult to get that, you know if it is missed in a review then there is a chance of just disappear forever."

The importance of missed information also depends on the abundance of required information. If there is little information on the topic users can not afford to overlook anything but when there is plenty of similar information overlooking some part is not a real problem. However, as the quote below shows in specific cases missing relevant information can be disastrous:

> AS13: "Missing the information is extremely important because if we miss, as we are moving to different scientific areas you have to read up all the background literature. If we miss an extremely important paper and that is very important to us because it might have that key piece of information which would have led us to make a new hypothesis or something. Particularly, when you are going to publish a paper or you are going to write a grant application ... into a scientific area that you have not been before, you can not afford to miss any information."

Matrix of search's depth

The matrix presented in figure 8.3 has two dimensions. One dimension shows the users' confidence that the information has been missed. In the best situation users are highly confident that they have not missed any relevant information and what they have found is what exactly exists about their search topic on the web. In the worst situation users feel so suspicious and doubtful about the comprehensiveness of whatever they have found. The other dimension of the matrix shows the degree of effort that users make to chase up the perceived missed information. Accordingly, four types of search might happen which occupy the four zones of the matrix

which are termed here: "perfunctory searching", "minimalist searching", "nervous searching" and "extensive searching".

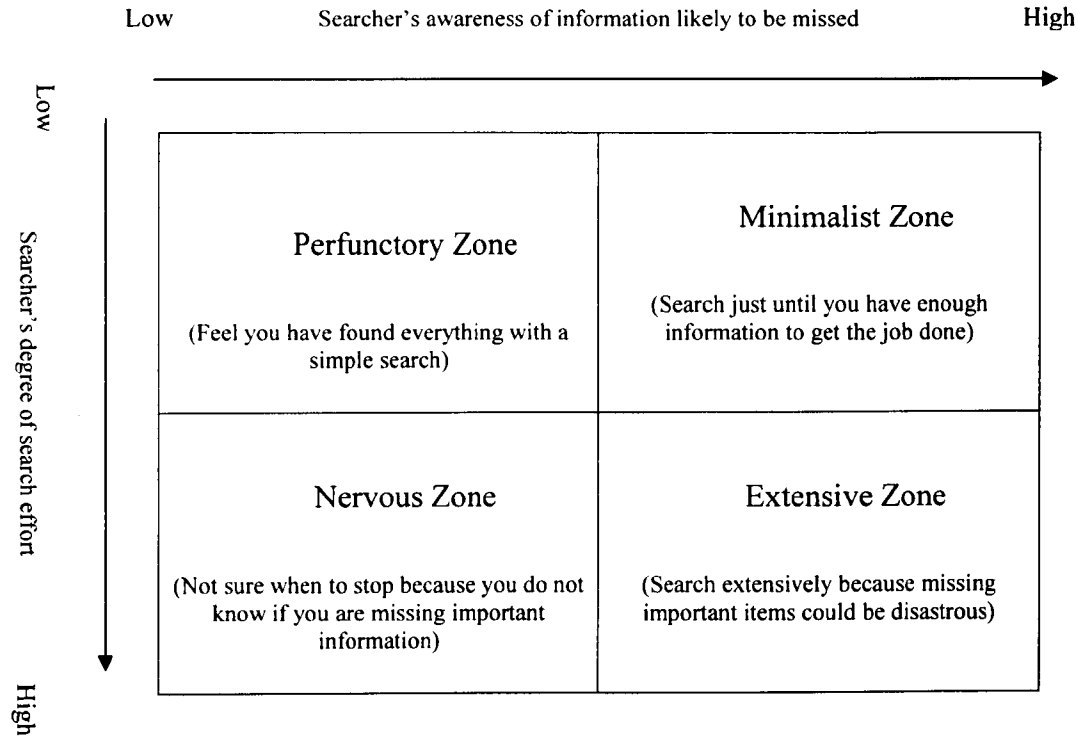

Figure 8.3: Matrix of search depth: Searchers' awareness of the possibility of missing relevant information and their determination to retrieve it.

Perfunctory searching

In the perfunctory zone users do not attempt to find any missed information because they are confident that they do not miss any important information. In this zone users trust the results of their straightforward searches without worrying about the need to carry out any more complex searches. Users might have different reasons to believe in this way.

For example, one of the participants believed that useful information for him exist only in a limited number of reputable and high profile resources that he is aware of and has access to them. Therefore, he was not worried about missing information that exists in other resources.

> AS13: "When I am looking for things if I don't get anything then I am pretty certain that it is not there ...if there is really useful information for me it is going to be on a reputable website where you can expect journals and these kinds of things. So, it is not going to be one of these things which are difficult to find. So, I think if you can't find stuff then generally it is because it is not there."

Another interviewee had full confidence in the comprehensiveness of Google search result. Therefore, he supposed he does not miss anything and if Google does not find something it just means it does not exist.

> RS18: "I trust Google, if you want. So, if I type something in and Google doesn't find it I trust it is not out there. So, if I can not find it so it is not my fault then. It is not there. So, in a way it seems that I completely trust Google maybe I shouldn't. If I would not trust Google then I maybe more open to blame myself. That is true, yeah."

In contrast, another interviewee did not seem very passionate about the comprehensiveness of information resources on the web. Although she ultimately perceived her searches similar to RS18, she conceptualized it in a different way. She said:

> RS23: "It [missing some information] doesn't really bother me because I know it is only the internet. I think people do get too obsessed with the internet and how amazing it is and all these but you know it is only as good as the people who put the information out there in the first place. It is not magic. So, if something is not there, well it is not there."

Minimalist searching

In the minimalist zone users are aware that they have probably missed some information but for various reasons they do not expend much effort to fill the gap. For example the following quote shows a minimalist search:

> AS01: "Well, it is possible to miss, but it always was. I mean the all historical things one could possibly miss. Things that are in rather obscure journals, which are not in back issues, which are not archive electronically yet you could miss some of that ... Yeah it is possible you could miss things. Well, I don't think whether you use the web or whether you use paper that is a similar problem in my experiences with the web ...you know it is out there but you can't find it, I suppose you could be very frustrated ... but by and large I don't think is a problem. I don't see it is a real problem at all."

This kind of search is 'minimalist' in that easily retrieved information fulfils the searcher's goal. For example one of the participants said:

> RS20: "The criteria that I have are quite specific. So, there is a limited number of places that could actually fulfil those criteria. I am aware that I don't have all the answers and I don't get all the possibilities. I don't get all, but I get enough for me to carry out my job ... so, I don't try to get every last drop of information."

It was believed by the participants in this study that there is no absoluteness in web searching and the possibility of missing some information always exist anyway.

> AS02: "I mean there is so much information out there ... I mean sometimes, I revisit subjects one month later and come back again another month and if I do a search each time I always go to different places. I always get different groups of information comes to me. So, that suggests to

me that I have never really looked at the entire of the data because every time I go I get something else. So, that means there is a whole load of information out there and I just get a part of it every time."

Another participant had a similar viewpoint. He said:

AS02: "If I did want every paper that has ever been written on my gene would I'd been annoyed if I'd miss one. Hmm, [pause] I have to say no, it doesn't matter ... and the reason I say this is we don't have at the university access to all of the scientific literature and the Web of Science doesn't have access to all scientific literature and so consequently if it is not on those two things then I have no idea that paper exist. So, there could be papers out there but I do not worry about it.

Moreover, he mentioned the issue of cost-benefit and reported that not only he is not worried about the information which might be missed in his searches, but also he deliberately excludes some less useful resources from his search domain to save his time. He said:

AS02: "There are also journals that I ignore and I do not include in my search ... because they are very unlikely to contain a paper that I am interested in and it would cost me too much time to look through their information to see they had information ...to be honest ideally I don't want to miss anything but it is the cost of time frankly so with current resources it would take me too much time to ensure that I have found everything and on a cost-benefit it is not good enough. So, I have to accept that I loose something."

Therefore, it is important to mention that when users ignore some information resources it is a different issue from information missing. This is a strategy which can be called as 'information exclusion' policy. The same interviewee gave more explanations about the problem of duplication on the web and his policy of excluding repetitive information. He said:

AS02: "I was going to write a lecture. I could type the topic of the lecture and there might be 2000 lecture already written out there, all on the same topic. They are all the same. So, what you are getting is 2000 versions of the same thing because there is so much duplication on the web. So, it would be great if you could find some sort of way to get rid of duplication. I mean you can do it yourself. Anything you type in Google, there is always going to be ten references to the same site or in your top twenty; ten things all tell you the same answers."

The importance of missed information also depends on the abundance of the required information. If there is little information on the topic users can not afford to miss anything but when there is plenty of similar information overlooking some parts is not a real problem.

AS03: "Probably I always miss something. I think because we don't have, we never have to get everything. As long as I am satisfied with the products that I have found, that is it. I don't want anymore about it which might be same or similar or repetitive."

Nervous searching

Sometimes searchers may feel distinctly nervous that they may be missing important information. They do not know for sure whether they are, but fear doing so. Consequently, they are unsure of when they can safely stop:

> AS12: "... you don't know when to stop, when to give up, because you don't know if you are wasting your time and there is literally no information out there online or whether you just not searched appropriately."

Similarly, another participant expressed his concerns about the possibility of missing information while searching the web and he compared it with the pre-web time and reported that the possibility of missing information has been always a problem in academia:

> AS01: "well, I suppose you always worry you might miss something. But it is not new, is it? I think, it is more likely when you are moving into a field of research which you are not too familiar with it. You know, if you are trying to take a new line in research it is easy to miss ... particularly the important paper which is not widely cited what I call 'sleeping papers' in our line of work. I suppose in biology there are Mendel's papers [which] slept for many years didn't them? You are always worried one of those is out of there, some sorts of really seminal works, but for some reasons have been overlooked and I don't know what you do about that, whether the web is helpful or not."

Extensive searching

In the extensive zone searchers consider it highly likely that they will miss important information if they fail to make extensive searching efforts. This kind of search is particularly appropriate in the case of, for example, literature reviews forming the basis for theses or research proposals. An example for this zone is as follows:

> AS08: "I can see how people would think that missing something is a failure. If I am thinking say of a PhD student. In the first year of their studies when typically they are doing their major literature survey I am pretty sure they would feel they have failed if they missed getting 250 papers on a particular area they are working in because they're terrified they might be asked a question on it in their viva."

Another interviewee also mentioned that he would continue searching unless he feels that he has covered all necessary information. He said:

> AS04: "I don't stop until I get all of the information I want which must be the right way around to do it rather than rely on the Internet to supply information to rather demand information. So, I think most of the time I get all of the information I want but who can say if I've missed something I wouldn't know.

The above matrixes have been developed based on the inductive analysis of data. However, in a further analysis the same dataset was explored in the light of Simon's Bounded

Rationality and Satisficing Theory deductively and the results are presented in the eleventh chapter. It was also interesting that the scheme developed by Agosto (2002) for coding interviews with a very different sample (ninth- and tenth-grade schoolchildren) and in relation to different web-based information behaviour (evaluating websites rather than specific searching) mapped well onto the data here. It would appear that there is at least a prima facie case for Bounded Rationality and Satisficing being considered as potentially useful concepts in our quest better to understand aspects of end users' information seeking on the web. More explanations about this issue are presented in chapter eleven.

Summary

This chapter reports the findings of the study about end users awareness of and reaction to missed information during web searching. The results shows when a user carries out a search there is always a possibility of missing some relevant information for a variety of reasons and participants in the study were mainly aware of this issue.

In fact, they were mainly aware of missing information while web searching and they considered it as an inevitable reality in web searching because of many reasons including immensity and diversity of the resources on the web. However, based on their perceptions of the volume and importance of the missed information they react to this issue in different ways. In fact, the risk of missing potentially important information was a matter of concern to the interviewees in the study reported here.

It is interesting that they had varying views on the extent and potential value of information that they thought they might be missing – as opposed to expressing any universal ignorance. The ways in which they went about searching, and the ways in which they went about deciding when to stop searching, showed considerable variation. The chapter suggests two model including matrix of search impact and matrix of search depth. These two matrixes address different reactions that users might take into consideration about missed information.

CHAPTER NINE:

COPING STRATEGIES

CHAPTER 9: COPING STRATEGIES

"Success is to be measured not so much by the position that one has reached in life, as by the obstacles which he has overcome while trying to succeed." Booker T. Washington

Introduction

This chapter addresses the fourth, fifth and sixth PRQs [34]of this study which are:

- PRQ4. How do web users seek to cope with information seeking failure on the web?
- PRQ5. What strategies do web users use to cope with information seeking failure on the web?
- PRQ6. What factors might influence different conceptualizations of information seeking failure, and different coping mechanisms?

To provide the reader with a clear description of the terminology employed in the chapter, what is meant by coping strategies is defined at the beginning of the chapter. In the next parts two main categories of active and passive coping strategies are presented to explain the identified methods that the participants in the study adopt to overcome their information seeking failure on the web.

Moreover, the chapter presents a matrix to illustrate how different concepts including the perceived importance of a search by users and their determination to overcome search failure forms four types of searches. Categorization of different search sessions based on this matrix enables us to address users' behaviour to react information seeking failure.

What is a coping strategy?

Taylor (1998) defines coping strategies as follows:

> "Coping strategies refer to the specific efforts, both behavioural and psychological, that people employ to master, tolerate, reduce, or minimize stressful events."

Regarding the above definition, coping strategies are employed by people when they find themselves in uncomfortable or unsatisfactory situations, when they do not possess the desirable level of comfort or satisfaction. In the web search context a feeling of failure to satisfy a specific information need is a circumstance that might motivate end users to employ coping strategies. Regarding the major types of coping strategies Taylor (1998), citing Folkman and Lazarus (1980), reported that in general there are two types of these strategies:

> "Two general coping strategies have been distinguished: problem-solving strategies are efforts to do something active to alleviate stressful circumstances, whereas emotion-focused coping strategies involve efforts to regulate the emotional consequences of stressful or potentially stressful events. Research indicates that people use both types of strategies to combat most stressful events."

[34] Primary Research Questions

The findings of the current study suggests that web users select and employ coping strategies based on the context of their search. As has been declared in the previous chapters the context of search plays an important role in web users' behaviour including their preference of selection and implementation of coping strategies. Data analysis suggested, for example, there is a difference between coping strategies for work-related searches and everyday life searches. For work-related searches the possibility of finding the most credible information is an important factor to decide on selecting appropriate coping strategies. However, for everyday life searches, convenience of use is one of the main criteria for searchers to decide on adopting any coping strategy. For instance, the following quotes show examples of the participants' coping strategies for both cases. For work-related information seeking failure one of the participants said:

> AS07: "[when I fail] then I go to text and I'll order on interlibrary loan or occasionally I will write to people or telephone them ... I mean for scientific information I prefer the journal, I prefer the article."

Similarly, another participant referred to work-related searches in general and looking for the literature in particular and said:

> RS19: "If it is paper searching you just go to the related paper and just look at the reference list at the back. I have no idea how ... you know abstracting journals ... the way people used to search the literature I have no idea how to do that. So, if I want to do literature search on my subject area I probably get three or four most key papers and look at the reference lists."

Interestingly, the same people have a different order of priority in coping strategies for everyday life searches. In these cases ,when they fail to satisfy their information needs on the web they firstly start thinking about the easiest way to deal with the failure.

> AS07: "Very often if I fail on the web, I'll go to the telephone. And if I fail on telephone, I'll go to text, to paper, or visit the place. So, if it is travel information ... I look on the web, if I don't get it there, that is the most convenient, I prefer the web to telephone. But if I can not find it there, I'll go to telephone and if I can't find there, I'll go to the travel agent or something like that in person."

> RS19: "Actually if, say, I can't find someone's telephone number. I mean the response ought to be to pick up the phone and ring up the reception and say can you give me the staff number but actually I usually e-mail another colleague somewhere else and say: 'do you have the telephone number?'..."

Interestingly, the above results about users' attitudes about coping strategies resonate with the "Human Behaviour and the Principle of Least Effort" (Zipf, 1949). Based on this theory basically people tend to adopt the easiest way to achieve a goal.

All identified coping strategies in this study can be categorised into two groups of "passive strategies" and "active strategies". When users employ passive coping strategies they usually change their initial search goals and try to avoid any further action. This type of strategy consists of indirect approaches including avoidance, and giving up the search. In contrast, active coping strategies entail users following new directions to gain more satisfactory results

and include refining the initial search techniques or changing databases. The details of these two main types of strategies are presented in the following sections.

Passive strategies

Passive strategies include the strategies that involve less action toward modifying the situation and mainly relate to accepting the situation as it is rather than as it can be. For instance, giving up after facing an initial failure can be considered an example of a passive coping strategy.

Giving up

When users feel their search leads to nowhere they may bring it to a halt and just accept the result whatever it is. One of the interviewees said:

> AS10: "If you think about it too much you waste time ... that is bad if you get frustrated because you can't find something. Sometimes it becomes like a challenge, doesn't it? You have to try to look for it and you think the system is defeating you, you know, and you get angry at it. So, that is probably when I say that is enough. Stop it. I am not going to look at it anymore."

Of course, there are different reasons for stopping the search after an initial failure. For example, the quote below shows the interviewee stopped the search because of time shortage:

> RS22: "I sometimes just give up because you haven't got time to keep going round in circles."

Goal modification

One of the coping strategies is making changes in the information need or modifying the goal of search to make it suitable for the existing situation. In fact, when users encounter unsatisfactory results they may adopt a passive reaction to it and instead of making any further effort to overcome the initial failure they may change their search goal. This strategy in the study is called "goal modification". For instance, one of the participants declared that if he looks for information and not able to find it he will look for something else that does the job in a slightly different way. For example, he said if he looks for the cheapest flight ticket and not able to find it, he will change his criteria and modify his goal and will book a more expensive ticket which is available.

> RS26: "If I fail to find a cheap flight to Rome and I have the money, so I'll buy a more expensive flight to Rome."

Active strategies

Active strategies refer to those methods that require further activities to change the search results from unsatisfactory to satisfactory. There are two levels of active coping strategies. At the first level, users may try to find a solution to overcome their search failure on their own. At the second level, they may look for help from other people.

In this research the first level is called *revising strategies* and the second level as *help-seeking* strategies.

Revising strategies

Revising strategies include modifying search queries, shifting databases/search tools, narrowing search strategy/search domain, and employing printed materials including personal archives.

Modifying search queries

Revising search query covers many activities including amendment of search terms, using alternative Boolean operators, and using broader or narrower search terms. Data analysis shows users employ revising mechanisms when they believe the failure of search is the result of their inappropriate search strategy rather than the lack of the required information. For example, the quote below shows an example of revising strategies:

> AS08: "Well, two strategies. Trying to think over different keywords, different combinations of keywords, to look at or look at different database or abstracting or search engines."

Similarly, another quote shows sometimes users might think their failure in search is due to using too specific search terms and they need to employ some broader terms to increase the possibility of retrieving more information. One of the interviewees who used this technique said:

> RS22: "Sometimes I wonder whether the specific term isn't there and a more general term might bring out more information which then you can look through."

Another interviewee mentioned that a failed search might provide him with some new ideas through a number of retrieved materials which will be helpful to carry out a more specific search. He said:

> AS11: "I think on a science side I would hope that my failed searches have thrown up enough references to the literature that I can work from there and then do a direct search on the literature rather than on the web search. If you are searching only across the web and as I said different search strategies I might try. So, either Google scholar or Google images or the directories of groups [are] three different strategies that I think is the way that I would try to fill in the holes."

Shifting database/search tool

Another coping strategy is adjusting the search procedure by shifting from one search tool to the other one or trying new databases. The quote below indicates an example of this strategy:

> AS08: "Chop from Google to DogPile. Change from PubMed to the one that we are supposed to use from the library licence and I don't remember what was it called or maybe change from PubMed to Medline."

When users attribute their search failure to selecting inappropriate search tool they may use this technique and try different search facilities:

> PS30: "I went to the Sheffield Forum as well. Maybe I just didn't go to the right place and I went to e-Bay as well but they are either too far away or too expensive that I could not afford it because I really want a second hand one because I only stay here for a year and I want to just buy a second hand one and throw it out when I leave."

Another participant also pointed out his strategy in changing the search tool hoping to overcome the search failure. He reported:

> AS03: "I have three major search engines which are quite different really, the three I mentioned earlier on. So, I would use the other two ... so, you got to be flexible you can't say it's got to be happen otherwise I am not moving. We haven't got time for that."

Narrowing down the search strategy/search domain

Participants in this study were aware of the immensity and diversity of web-based resources. Accordingly, they knew that looking for specific information in such a huge and heterogeneous source of information might be very challenging sometimes. For example, one of the interviewee used the metaphor of "*needle in haystack*" to show the difficulties of looking for information on the web. He said:

> AS15: "I would say on the whole, given the number of websites that are out there, tens of thousands of websites, therefore you try to find your 'little needle' in your 'very very big haystack' it is probably not too bad but I do spend further time getting frustrated probably like everybody else does."

Accordingly, they knew that a failed search might be a consequence of the broadness of the search domain. Therefore, one of the identified coping strategies was narrowing down the search domain to a more specialised set of information resources. For example, the same interviewee who used the metaphor of "*needle in haystack*" used the phrase of "*hone down*" the search procedure to retrieve better results. He said:

> AS15: "Generally, they [search results] are quite good but that is because I have gained some experience now in knowing that I have to put usually a series of keywords and I have to be quite selective. But I still sometimes spend half an hour or more searching for one single topic because I just don't have a right combination of keywords to hone down the search to something sensible."

Employing printed materials

This study showed that despite the undeniable role of the web in providing people with different sources of information still printed materials are useful to overcome failed searches on the web. Printed materials that the participants in this study use exist either in the library or in their personal collections. Printed resources were used for four different purposes including:

- Identifying new keywords to carry out the search on the web with a new search strategy.
- Gaining access to the full text of documents that are not available on the web.
- Searching for the required information in the printed collections which are not available on the web.
- When all revised search strategies fail on the web and users have no choice for search apart from printed resources.
- Users' personal preference for using printed materials instead of the online resources.

The following quotations show examples of the usefulness of printed materials in helping users to overcome their failed web searches. For example, as the quote below shows this interviewee reported that he uses textbooks to identify more appropriate keywords for those broad search areas which are new to him:

> AS07: "If I am looking at something very very broad, an area that I don't know or if I want to look at something such as genetic engineering. I don't know really what that I want to find out. So, my preliminary searches would probably be done using a text book and then when I feel I know what the terms are, what specific terms are, and then I'll use the web to look at the detail."

If searching on the web culminates in failure as a result of inaccessibility of the full-text of the required information, then using printed materials can be an effective coping strategy:

> RS17: "Sometimes you go back and again library will contain archives and old records, information that probably less people are aware of. I don't know how much of that part is there [on the web] or is not there yet. So, still I need further archive over there. If you stumble across a paper say 1927 ... and you want that paper that would be hard to get on there but over there it's there [library] in printed format."

Besides, the participants in this study knew that despite the hugeness of the web, it does not contain everything. Therefore, when they fail to satisfy their information needs on the web they back to the "*Gutenberg Galaxy*" [35] (McLuhan, 1962). For example, one of the participants said:

> PS30: "I always feel that not all the information is on the internet. For example, a very simple example, I wanted to use this technique ... I saw it in some paper that people did the techniques and I wanted to do the same thing. But when I went to that exact paper they said reference to their previous paper and then I wanted to find the previous paper but that was published in 1989 which means that is not really on the web. Even though it is a very famous journal, they just have their collection started from 1992. So, I think I just couldn't get it on the web. So, I had to go to the library and saying that I need that and they finally managed to find it in Manchester University and send it to me a photocopy of the paper."

[35] The Gutenberg Galaxy is a metaphor to define the whole printed materials in the world since the invention of printing machine by Johannes Gutenberg in fifteenth century. This term was coined by Marshall McLuhan in 1962.

Some of the participants mentioned that they usually search the web first to satisfy their information needs and then will revise their search strategy a few times. However, if they do not manage to succeed the search on the web they will turn to the printed materials. For example:

> RS16: "I'll probably try two or three times, type in a slight variance on my search terms and if it doesn't work, I'll forget about using the web perhaps. I'll look through a couple of general text books or ask colleagues."

Finally, some interviewee said that they prefer using the printed materials as a matter of personal choice. The following quotes are two examples of users' preference to use printed materials:

> RS17: "I am still a bit old fashioned if you like and impatient with it. I still don't mind going to the library and looking in books ... you know the young students here they will say oh, why you don't do it this way why you don't do it that way. So, I mean that is a difference in age and in the whole primary education things which we had no computers, when I was child in primary normal education."

> AS01: "As I said at the outset you are talking to a dinosaur here [laughing]. So, I don't look at it [the web] as a primary tool. I'm sure it is a primary tool now for many people, but because I am at the end of my career, so it is not a primary tool for my hobbies and retirement [laughing] ... I suppose, I treat the web as a sort of luxury item. I don't think as an essential tool but of course if I was in your age I would think it as an essential tool and not a luxury."

One of the participants mentioned that sometimes to gain access to some highly-specialised information personal archives are very useful. She declared that she had found some information resources in her supervisor's personal archive which were unavailable both on the web and at the university library. She said:

> PS28: "I thought who can have this paper? And then I remembered oh my supervisor! He has a small room for the papers. So, I am going to look there and it was there ... and I thought how stupid of me because I should have gone to that option first than the internet. You know it is just like that is easier you sit ... just type in and print the paper instead of going get the key, going outside, getting in the room, looking different files ... yeah, that was more interesting because I saw another paper that I was interested in. So, you know at the end it gave me more information than just found it on the internet."

Help seeking

In a broad sense, it seems that coping strategies are part of natural strategies of asking for help. One of the interviewee said:

> RS17: "I think it is something that comes naturally to me anyway to ask for help. And I am also a bit bossy. So, when I see my students, my colleagues, my peers. [I say] do me a favour. I know they'll

say yes. It is very lazy I think. A lazy strategy because it works, I use it more. So, it was developed. I don't think I did it deliberately but because I think it works for me I use it more and more."

As an example of help-seeking strategy another interviewee mentioned:

AS10: "Well, there is one example, yes, if there is something on the university website if it was a colleague who told me that it is there somewhere, I would then e-mail a colleague and say give me the website. That is the probably what I would do and I might say that for other things even scientific things if I couldn't find them and it is because someone told me that it exist there, I would e-mail my colleague and say can you give the URL."

Some of the interviewees mentioned that they will ask help after they do their best. In fact, they prefer to adopt the coping strategies that they know and if it does not solve their problem they will ask other people to help them:

RS25: "I think I am a kind of person too who likes to work on his own and quite happy to work on his own not very good in networking in conferences for example. So, I am quite content to find things out through libraries, through web searches which don't disturb other people. But then when I do need some difficult bits or want to have some assurance of being comprehensive and not overlooked anything then I would go to people to sort of making small use of their time for critical points."

Asking others for advice

Help seeking has two sub-categories. In the first category which is asking help to carry out the search in a new way, users seek advice to develop a more efficient search strategy which might lead to success. The participants in the study mentioned that they ask help from their colleagues, peers and friends. One of the interviewees said:

PS30: "I will seek for help maybe from my supervisor or other colleagues, yeah I'll do that. I'll ask people ... I think you should put your effort first if you still can not get it. Then you ask for help. If you haven't tried hard enough, why should you go for somebody else?"

One of the help seeking strategies is asking somebody else to carry out the search. For example, one of the participants said if she fails on a search she will ask somebody to search it for her. She also mentioned that she will ask somebody who is skilful in searching the web to carry out a search for her. She said:

RS17: "I'll get someone else to look for me [laugh]. Those people who are better in searching. I will ask somebody else to do it for me ... my husband is very good at this. Then I can leave my list and then he will send me."

Remote collaborative search

One of the participants highlighted a specific coping strategy which in this thesis is called "*remote collaborative search*". In remote collaborative searches more than one person who are

in different locations at the same time are involved in a search procedure. The following quote shows an example of this search:

> PS29: "I am not sure [whether] you are aware of this or not, but it happens a lot nowadays, especially within Chinese or my friends. When I am in the department, I am in front of computer all the time. So, sometimes they are in the train station and call me and say: 'do you know when the next train is to so and so?' and say: 'I am in Cambridge now' and you know from Cambridge to Sheffield you have to change your train. There is no direct train usually from Cambridge. So, they might call and say: 'look I am in Cambridge now when is the next train to come back to Sheffield and where [do] they change?'..."

The same person reported another example of her "*remote collaborative searches*" in which she has been travelling and looking for a hotel's address. In that case, she telephoned a friend in another city and asked him/her to search the address on the web. In fact, it was easier for her to call a friend to have access to the web to search for the address rather than looking for the information through other conventional sources including local tourists' centres, telephone directory, or map. She said:

> PS29: "Also it is not my search again but it is a really successful search. I booked a hotel in Manchester but forgot to print out the address. I had the reservation page but they hadn't the address on it for some reason [laugher]. So, I called my friend last week and said: 'are you in front of computer? So, can you search for this particular hotel and tell me the address?' We always use the internet which is quite odd."

Postponing the search

One of the coping strategies is leaving the failed search for a while and coming back to it again to overcome the initial failure. This method is different with giving up the search and can not be considered as an entire passive method because searchers know they will come back to the search in near future and they just leave the search temporarily.

In fact, users believe excessive effort would not be useful to overcome the initial failure. There is a time in each search that the user feels extra time and energy does not make any difference in the search result and it is better to leave the search task for a while and being away from it to find a new approach. Accordingly, there is a difference between search session and whole search procedure. Each search procedure might include several search sessions until users feel their information needs have been satisfied or regardless the result they feel that they have searched sufficiently and stop it. The quote below shows an example of this strategy:

> AS03: "Oh, sometimes it [information searched for but not found] may come up again through another route and I'll try it and it might be successful and it does happen. You know, it is kind of strange how it happens. It is not purely scientific. It is very much, you know, the dices roll correctly today for you whereas it didn't do yesterday ... and I always assume this is part of the fact that none of the search engines really address whole of the web or maybe for some reasons a new link has happened or maybe someday has made a new link since I have last look at it."

The same interviewee, AS03, declared that when he resumes searching he keeps searching until the search succeeds or eventually he might leave it, but he will not give up easily. He said:

> AS03: "So, I certainly swap between my search engines and then leave it and come back another day and then think about another approach. ... So, if I can't make it so it would nag me until I get it or alternatively I'll just leave it. I won't put it at that lecture or won't put it in that paper. So, I spend a week probably and back and forward with that until I got it or didn't get it. So, it does matter."

Another participant also mentioned that sometimes she might postpones a failed search for a while and will come back to it sometime later to carry out the search again which might lead to more satisfactory results. She said:

> RS23: "It depends on how important I perceive that piece of information to be. If it is something that would have been interesting, for instance to do with my project, ... not crucial I might just forget about it ... I'll probe again through a few weeks time or couple of months time and it might be there next time because things, you know, appear and reappear. If it is something that I really need to know I will try different avenues."

Moreover, sometimes users might postpone a search because of their personal mood. For example, one of the interviewee said that he is a morning person and he carry out every task including web searching more efficiently in the morning:

> RS18: I would think it depends a little bit on the time of day. ... I am a morning person. So, I am good in the morning. If I do a search in the morning it often means it is an important issue. If it is not important, I'll do it sometime later. Because it is important I more likely to search say first two hundred hits and then say in the afternoon I look at maybe fifty and then I am done. So, it means if I haven't found the information after fifty in the morning I would go on and thereafter I would someway call it success because I might have found what I wanted whereas in the afternoon I would be unsuccessful although I didn't look for as many as in the morning. So, one element is certainly my patience."

Matrix of coping strategies

Data analysis showed that the "importance of the search" and "searchers' determination to use coping strategies to overcome failure" are two elements which play important role in users behaviour after receiving unsatisfactory search results. Accordingly, the current study suggests another matrix which enables us to address users coping strategies. This matrix consists of four zones including "trivial searches", "alternative searches", "fascinating searches" and "crucial searches".

In the following section there are some explanations about each zone of the matrix. However, it is important to mention that this matrix could be considered for further analysis to find out whether there are links between different aspects of the search and zones of this matrix. Nevertheless, any further exploration in this issue was beyond the scope of the study and additional analysis can be carried out in future research.

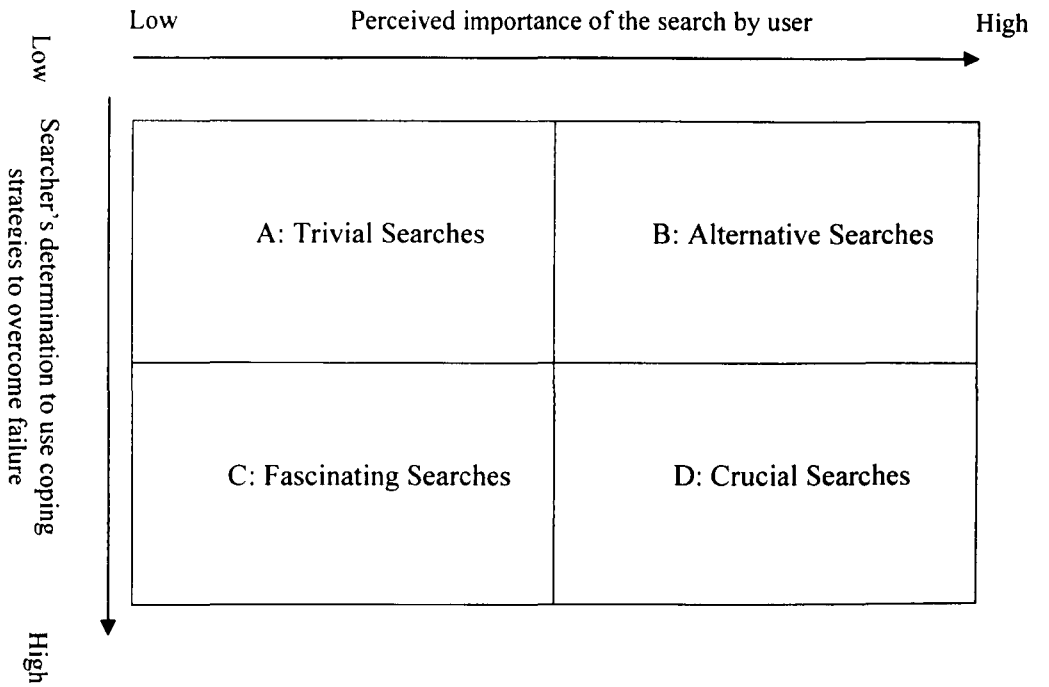

Figure 9.1: Matrix of searchers' determination to employ coping strategies and the importance of the search

Trivial searches

In zone A the search is not perceived as an important issue and if the search fails searchers also are not determined to employ different coping strategies, or if they adopt any they are not so persistent to complete the search successfully. This kind of search is called "trivial searches" in the study.

This kind of searches is usually stimulated by not very powerful motivations. As earlier reported in chapter five sometimes users search the web just to respond to their inquisitiveness to find out more information about a specific issue. For example, one of the participants mentioned that sometimes he might search the web just as a matter of curiosity to answer idle questions.

> RS25: "I think for answering idle queries it [web searching] is often very useful. ...other idle queries like you type in names of people who used to know at school or university ... if you can identify them amongst all the others who have the same name. I had a friend that was in a remote part of Canada fifteen years ago and I wondered whether she is still there she is now in another remote part of Canada but still is practicing as a midwife. So, I am confident this is her. I have lost touch with her ... it was just interesting to find out where she was. So, those sorts of things are fun."

Therefore, if a trivial search fails, the user will not be so persistent to adopt copying strategies and leave it as it is. For instance one of the interviewee said:

RS21: "It depends I suppose how much the information means to you to get it. I mean if it is something that appears quite trivial and you can't find it in an hour you might consider that a failure whereas if it is something that you know is very important to get that information you maybe willing to spend more time on it and even if it takes you three hours if you get the answer that is a success. Do you see what I mean? I think it depends on how important the information you want is to you whether you would consider it as a success or failure within a certain length of time."

Alternative searches

In zone B although the search is important, searchers are not determined to carry on the search by employing coping strategies. This zone can be termed as "*alternative search*". In this kind of search web searching is not the one and only way to find the required information and there are other alternative routes. For example, one of the participants said:

RS25: "Then I would look for alternative roots and these might often involve actual people. I think writing or these days e-mailing to people, for example these people in Washington state university [I would ask] I have found your publication up to about 2002 have you published anything more recently which is relevant to what I am doing. Again with the temperature curve [another search scenario that he reported earlier in the interview] it maybe a case of asking around we have people we know in met office or as I said I would look at the meteorology sites ... I guess it is probably more by asking people."

Fascinating searches

In zone C although the search is not intrinsically important, there is an element in the search topic that fascinates searchers and consequently they are persistent to find the required information. In fact, in zone C the search becomes like a puzzle for users which they just want to carry out successfully regardless of the importance of the search topic for them. For example, in the following quote the interviewee says that he has been very interested to find the answer of question and he spent long time looking for it but it was just a "fairly idle" question and not very important for him:

RS25: "It was something I searched for, sort of fairly idly, and I could not narrow it down ... I was interested, can't of the moment think why, in the diurnal change of temperature. ...I couldn't find it immediately on the books that I had to hand. So, I thought okay I'll do a web search on this and put in "diurnal air temperature" and various other search terms. I tried and I could not narrow it down sufficiently to find what I wanted. I just used Google as I generally do and it will come up with hundreds of sites where the various combinations of words I had put in occurred and some of them you felt you are getting quite close but nowhere did I find a site. And I did spend awful long time on this ... but I was really interested in. So I could not find a site which said oh yes we modelled the diurnal change in the air temperature using the following equation in which minimum temperature and maximum temperature are here as variables. I was a bit surprised by that I though I was going to run that down quite simply because it is sort of obvious problem and I am sure lot of people have solved it and I had a little play around myself ... I didn't find the complete answer."

Similarly, in the quote below there is an example for this kind of search:

RS26: "I would say that I want to find information and that is the important thing. But that is not strictly true because there is an intellectual satisfaction of searching in the net. So, the main thing that ... I should admit when I don't find something is not the failure of not having it, it is the failure to be able to attain the intellectual challenge that entails being successful at that task."

Crucial searches

Finally in zone D because of the importance of the search topic for searchers they employ all possible coping strategies to overcome failure. This study suggests the term *"crucial searches"* for this kind of search. The quote below shows an example of this category of search:

RS21: "It depends I suppose how much the information means to you to get it. I mean if it is something that appears quite trivial and you can't find it in an hour you might consider that a failure. Whereas, if it is something that you know is very important to get that information you maybe willing to spend more time on it and even if it takes you three hours if you get the answer that is a success. Do you see what I mean? I think it depends on how important the information you want is to you whether you would consider it as a success or failure within a certain length of time."

Similarly, two other interviewees said:

RS26: "Well, it depends on how important is what you are looking for. So, if it is something very important that I can not find it that would be a problem."

PS37: "I think there is only one occasion that I failed to find something that I really needed. In that case I went to the library may happen to have it there ... I found the hard copy and photocopied the article ... It was a journal called "Methods in Enzymology" ... and there was an article that I couldn't find that on the databases. Then I went to the library website and found that they have it in the library. So, I went to the library and found it in the library and photocopied it which took me back to what I had to do in Lancaster [when he was an undergraduate student] all the time."

Developing coping strategies

Another issue that has been considered in this study was how web users develop coping strategies. In fact, this research not only identified a number of coping strategies but also explored the ways that the participants adopt and enhance their strategies.

This study identified four methods of developing coping strategies including attending training courses, using trial and error, knowledge sharing with colleagues and adoption of the pre-web coping strategies on the web.

Training courses

Some of the participants mentioned that they have attended training courses to learn how search the web and it had been useful for them to develop their coping strategies. For example, the following quote shows how one of the interviewees developed his coping strategies.

AS08: "Mostly experience I would think. But also from a training course a long time ago, but mostly by experience, and shared experience, talking to colleagues ... CICS did do a training course in the days when the World Wide Web was something new."

Trial and error

Some of the participants mentioned that they use trial and error to improve their search results and interestingly they were quite happy about the efficiency of this method.

AS10: "Trial and error. I never attended any course or training or anything like ... It is just by learning by trying things which I guess my children told me how to do it. Because many years ago when we used to have computer at home and my children were quite young and I would be wondering how to work something on windows or Word and I would get the book out to look and eventually my children said to me this is crazy looking in the book to work out how to use a computer. You use it by using it. They said: 'just do it, just try it ...you can't break anything just keep having a go until it works.'

In fact, developing these strategies is part of a learning process. Web searching is a learning curve for end users. As users perform more searches they will learn more techniques because of their errors. Indeed, over real searches they are learning how to search more efficiently.

AS03: "The learning curve was what is the appropriate search engine for me and then after that is to target the way I ask the questions in a minimum number of words I get the right answer and you learn while you go on."

As has been mentioned earlier, the result of this study about users' behaviour in adoption of coping strategies is compatible with "*the principle of least effort*" (Zipf, 1949). In general, users adopt the mechanisms that require less effort. For example, one of the participants in answering the question that "how did you develop your coping strategies" said:

RS19: "Laziness. Yeah, using e-mail is so much easier. I don't know. People in general or perhaps me became less comfortable to actually picking up the phone speaking to people when it can be done by e-mail. It is likes school kids just text each other rather than have a conversation. ... The coping strategy was developed by what is the easiest and least hassle way to get this."

Knowledge sharing

This study showed that users who work in similar areas tend to share knowledge of searching with each other through different ways including sharing search terms, exchanging databases' addresses and so on. One of the interviewee said:

RS17: "There is enough information and just because of a sheer volume of information you know you are going to miss most of it ... and I may rely on colleagues to say "did you see this?" word of mouth still also I think is important. When using the net I mean people send sites to each other and say this maybe useful to you which is a way of cutting down the time that take to find

something. So, I think there is a growing culture of that in order to be more efficient in a community particularly in biology I think you get a lot of 'I found this, this is useful for you and it doesn't take you long time to have a quick look' and it get from all over the world people send stuff and that is very useful because I don't have to search [laugh]. And I think there is a reason behind that to be more efficient as a community searching for things helping each other and this way is quite useful."

Another interviewee mentioned that when he fails to find something he will go to the library or asks his colleagues:

AS02: "There are two routes really. One is library and the second is colleagues. So, it could be I know colleagues that might know more than I and undertaking different approach. I would probably first contact colleagues [who] have a similar need and see if they had anymore success and then the second would be access to whatever I could get from the library."

Similarly, three other participants said:

PS27: "Through experience really I think ... probably I'd use the internet first immediately and then if that is not satisfactory then go to papers and ask somebody else. I think it is just experience and also talk to postdocs about what they would do."

RS18: "I don't know it is quite natural. If you don't know something you ask a colleague whether he or she can remember. I think it is most natural thing. I probably grew up with it."

AS03: "In the lab in our research group I might ask them [colleagues] ... on lunch activities they do crossword in one of the newspapers. So, always looking for words on there, so that is amusing, so there are quite different sets of questions. But I say look I just not getting anywhere so it might be that occasion different views. And that is great I think different ways to come up with few words and it is helpful if somebody else has a different attitude."

Adoption of pre-web coping strategies

Some of the participants pointed out that coping strategies that they use to overcome failure in web searching have been developed before the emergence of the web. In fact, they use the old techniques in the new environment:

RS26: "Actually this is the other way around. I think I didn't develop coping strategies because of I fail using the internet no. I had these coping strategies before the internet was here. Because before the internet was here you just get your phone and call companies or you just take a walk and go to John Lewis downtown and ask for the qualities of blenders and then they gave you the direction of appliances and then you asked for a leaflet. So, what I am trying to say the coping mechanisms of internet failure are born or were there before you got the internet ... you don't develop the coping strategy because of the net failure ... the young generation that start with the net then they will have to develop on the coping. It depends what particular sector of the population you are talking to."

Summary

The participants in this study reported a series of coping strategies that they usually implement to overcome their failure in satisfying their information needs. They have some sets of coping strategies. These strategies are divided into two categories of active and passive. Active strategies are those methods that require further activities to obtain more satisfactory. The study identified two levels of active strategies. At the first level, users may try to find a solution to overcome their search failure on their own. At the second level, they may look for help from other people. In this research the first level is called revising strategies and the second level as help-seeking strategies.

Passive strategies include the strategies that involve less action toward modifying the situation and mainly relate to accepting the situation as it is rather than as it can be. For instance, giving up after facing an initial failure can be considered an example of a passive coping strategy.

The matrix presented in the chapter shows that web users usually prioritize their coping strategies according to the importance of the search and their determination to change a failed search to a successful one. This study identified four methods of developing coping strategies including attending training courses, using trial and error, knowledge sharing with colleagues and adoption of the pre-web coping strategies on the web.

CHAPTER TEN:

CAUSATIONS OF SUCCESS AND FAILURE

CHAPTER 10: CAUSATIONS OF SUCCESS AND FAILURE

"Success and failure are greatly overrated. But failure gives you a whole lot more to talk about."
Hildegard Knef

Introduction

This chapter presents findings of this study about the participants' perceptions of the causes of their search failure and success in the light of the Attribution Theory[36]. The chapter which presents the first part of the deductive analysis illustrates to what extent the constructs Locus of Control and Attribution Theory can provide useful conceptual frameworks for understanding searchers' perceptions of success and failure in web searching. This chapter addresses EmRQ1 and EmRQ2 which are:

- EmRQ1: What are the perceived causes of search success and failure of a particular group of academic users?
- EmRQ2: To what extent are Locus of Control and Attribution Theory useful in interpreting users' conceptualizations of success and failure in web searching?

To be more precise, initially perceptions of failed and successful searches were derived from the inductive analysis of the dataset, which was discussed in chapter seven. In further exploration these perceptions were classified into "internal" and "external" attributions, and the relationships between these categories and "successful" and "failed" searches were analysed deductively.

Factors perceived to determine success/failure

Analysing the dataset uncovered a number of factors that participants perceived to find out why some searches lead to failure while some other searches culminate in successful outcome. The results are categorised into two groups of perceived factors related to success and perceived factors related to failure. The identified factors are summarized in table 10.1.

Successful searches

Factors perceived to bring about success in searching included domain knowledge, availability of information, search experience, linguistic and conceptual ability, confidence, persistence, search tools efficiency, website structure, and good fortune – which are not necessarily mutually exclusive. There are some explanations and a number of quotes form the dataset about each factor in the following section.

[36] There are detailed explanations about the Attribution Theory in chapter two.

Table 10.1: The perceived factors related to success and failure in search

Perceived factors related to success	Perceived factors related to failure
Domain knowledge	Lack of domain knowledge
Availability of information	Unavailability of information
Search experience	Lack of search skills
Linguistic and conceptual ability	Lack of linguistic and conceptual ability
Confidence	User or system errors
Persistence	Lack of sufficient effort
Search tools efficiency	Search tools deficiency
Website structure	Deficiency of website design
Good fortune	Bad Luck
Combinations of factors	Combination of factors
Uncertainty about the causes	Uncertainty about the causes

Domain knowledge

Having domain knowledge helps the user to select apposite search terms which enable him/her to develop efficient search strategies. Therefore, familiarity with an area helps users to carry out their search more efficiently:

> AS06: "Very often actually, if I am looking for information in an area that I know reasonably well I can actually, for example, search for the author's name on Web of Knowledge because I know whose work might be most likely to be interested in. So, I can quite often get to the thing that I want more quickly."

As the quote below shows this interviewee highlighted that when he carries out a search in an unfamiliar area he firstly needs to find out some appropriate search terms before conducting the search. However, if the search is related to his area of research he has enough knowledge about the area and is able to develop efficient search strategies. He reported:

> AS07: "If I am looking at something very very broad, an area that I don't know or if I want to look at something such as genetic engineering. I don't know really what that I want to find out. So, my preliminary searches would probably be done using a text book and then when I feel I know what the terms are, what specific terms are, and then I'll use the web to look at the detail."

However, when a search locates in the users' domain of study s/he is able to target appropriate resources and conduct more effective searches:

RS21: "If I am looking for something specific like if I want to find a paper because I know where to look, it is probably easier than perhaps trying to find something at home."

Some of the interviewees attributed their success in search to their familiarity with the search topic and their knowledge about the domain of search. For example, one of the participants reported that she usually feels more comfortable with familiar websites in her area of study that she visits frequently. She believed that she is more successful in searches that are related to her area of research:

RS17: "The majority of it [searches] will be e-journals for my work. That would be the majority. Then there would be other information, say, I use what we call the Nottingham centre for Arabidopsis seed stock [Nottingham Arabidopsis Stock Centre] where it is an information centre then I look for various mutants. There is an American one. So, these are sites that I regularly look to for information. So, it is going to be eighty percent work in that sense ... obviously if I am going to the sites that I know where how to navigate and where to get the information that I want, then it is very easy and I am hundred percent successful."

Moreover, another participant reported that she usually carries out searches in her area of study and she retrieves the same piece of information several times. Therefore, she feels more comfortable with searches in her area of research. She said:

AS05: "Usually I am looking for papers, specific papers relevant to the research I am doing. So, I want to check a paper or see if something has been done. That is generally what I am doing and also I am retrieving the same piece of information multiple times, because I have found it before and I know it is there. I check it again."

Availability of information

Availability and abundance of information one the web which was mentioned in chapter five was one of the positive comments of the participants about web searching and also was pointed out by some of them as the reason of success in search:

AS01: "Because they have a lot of material readily accessible on the web and sometimes they may do this because they are looking for test out the keys and whatever. So, when more people use it they get more feedback perhaps. That maybe one of the reasons [of success], I don't know, but certainly there are many very good keys [related to his search topic] out there [on the web].

Search experience

Experience was mentioned as a key factor leading to successful searching. Some interviewees spoke of experience in a general sense of "knowing what to do". For example one of them expressed his opinion about a successful search as follows:

AS12: "That was very straightforward and I had a very efficient and rapid answer. Basically, I got from where I was to where I wanted to be very quickly and I guess the reason it worked well was because I knew what to do and there was no learning curve there."

Another participant believed that there is a direct link between search experience and success in search. He said:

> RS26: "Searching through the net is like playing chess and at the beginning of playing chess ... playing chess stimulate your mind and your mind is sharper because of the chess playing, you learn to play more chess and every time you play better. So, what happens doing different many searches, in the biology or outside the biology, you develop a certain mind to establish a collection of links which are the keywords that actually going to lead you to the place where you are going to be."

Linguistic and conceptual ability

Linguistic ability was also mentioned as an important determiner of search success. For instance, the quote below shows how one of the participants ascribed one of his successful searches to his ability to develop a search query with appropriate keywords:

> AS03: "[The search] was successful because I could target from a number of items not just one word. Three separate areas which I felt sort of triangulated on the answer."

Similarity, another interviewee attributed his success to his ability to develop efficient search query:

> AS04: "I think the success was the function of using the extra information and being more specific."

Some interviewees linked experience to the development of such linguistic ability. In fact, experience could foster awareness of the way in which the same topic may be described and indexed using a variety of different terms. For example, the quote below shows how one of the participants thought about this issue. He said:

> RS26: "If you are in Spain, for instance, and you are looking for a blender for your kitchen you use, for instance, the word would be "blender" and "comparison", and then it [search tool] will give you different places then you can actually compare different blenders. If you use the same "blender and "comparison" in the UK you don't get that information. You just get to shops that sell you the blender. When you have a lot of experience or some experience you realize that you need to put "blender" and "review" and that is why you get to the places that they tell you how good the blender is. So, the word "review" is a key. ...Language is a key issue of web searching. So, quite likely in Farsi people use different keywords to look for Iranian web pages than the words that they use for English web pages and so on. The structure of the language is the structure of your mind and then leads to different structure when you search the net."

Confidence

Experience could also engender confidence – itself a factor contributing to success. For example, one of the participants attributed his success to his confidence of using the web which was the result of several years experience of web searching. He said:

RS20: "At least two things [reasons of success]. One of them is my experience in using the internet in which I would say I have strong experience, a high level of experience. I have been doing it for a long time for a lot of different reasons and I have successfully managed to retrieve information which has been useful. So, I am quite confident in my abilities. The other thing is I am confident that I can get to enough information that allows me to do my job."

Persistence

Persistence could also result in achieving the specificity required for a successful search. In the following example, the searcher hit upon a more specific search term after a variety of unsuccessful attempts to narrow his search:

RS20: "I was looking for an example of a successful proposal for a European funding opportunity. You're not meant to post your successful proposals on the internet. The EU says you shouldn't do that, but I found some that had been put out there and I've found one and maybe there are others. I didn't find them to start with. I did a lot searches and spent quite a long time – half an hour to an hour – trying to find out these things. I did different sorts of searches, combinations of keywords that sort of thing. And then I eventually narrowed down my search to look for the document. I realized there was a particular spelling, a grammatical mistake really that put the word "start" and "page" together with no space so you have got "startpage" with no space and in PDF and I searched for all PDF with "startpage" in them and it just came up with the whole long list of the completed proposals for this particular call. So, I was very pleased with myself."

Search tools efficiency

Some of the participants found the performance of web-based search tools (e.g. search engines and databases) in a satisfactory level and sometimes they attributed their success to their efficiency in information retrieval. For example, one of them explicitly ascribed his success in a search session to the efficiency of the search tool which he had used:

RS19: "I think it thanks to Web of Knowledge. I think it is superbly set up. It is very fast. Yeah I think Web of Knowledge is extremely well-designed. Only when you fail on it when you are too broad in your search term and you come back with ten thousands possible articles and then you have to learn how to be more specific. But I think it is a quality of Web of Knowledge. I mean it works very well for me."

A number of the participants said that they only use one favourite search facility because they feel satisfied with its performance. The following quotes are examples of users' views:

RS18: "You could use different search engines, but I never do that. I just use Google. I never delve with other ones. I think they pretty much covered by Google as well."

AS07: "I think Google is probably the best of the general ones for me but things like Ask Jeeves and all those, I don't like these at all and they never give me what I want."

Some of the interviewees reported that since they initially have started searching the web they have been trying different search tools to find out which one could provide them with more satisfactory results. Moreover, progress in search facilities on the web encouraged them to select new search tools which offered new services (e.g. progress in Google's services). The next excerpts are examples of the participants' attitudes of search tools that they use:

> AS06: "If you had asked me five years ago I would have said Web of Science, Web of Knowledge would be almost the only thing that I used. Now I use Google a lot actually because Google is getting better and coming up with more relevant hits. So, particularly if there are bits of information that I am looking for, I tend to do it on Google it is usually quicker actually."

> AS10: "I think Google is more effective. I think over the search engines that I have tried. Google is the most effective. And the other way of course you can search across via the groups ... whether they have directory of web pages that on Google and I guess Yahoo as well. Those can also be very useful."

> AS14: "I have used Lycos and AskJeeves and Microsoft MSN search engine but I prefer Google and stick with that."

Website structure

In some cases searchers ascribed their success to website structure which provided them with an effective way to retrieve the required information. For example, the quote below shows how one of the participants attributed his success to the effectiveness of the website which he had used. He had found Amazon as a very well-designed website which enabled him to retrieve what he wanted. Furthermore, he was satisfied with the easy ways to access to this site which is possible either through searching the Google or just remembering the URL address. He said:

> AS15: "I was searching through Amazon yesterday. I was trying to find a book for my brother.... So, that was a nice and easy one and I found the book for him. ... Amazon again it is an obvious place to go with a nice clear brand name. You can search in Google or you can just remember it. It is very easy to remember it and when you get there it is a very nicely set up website and it is very easy to find information you want. So, that was so easy and straightforward."

Good fortune

Occasionally, where the participants were not able to identify other factors to conceptualize their success, they attributed it to good fortune. The following two excerpts from the dataset demonstrate this type of attribution:

> AS03: "Oh, sometimes it [information searched for but not found] may come up again through another route and I'll try it and it might be successful and it does happen. You know, it is kind of strange how it happens. It is not purely scientific. It is very much, you know, the dices roll correctly today for you whereas it didn't do yesterday."

Similarly, another interviewee ascribed his success in a search session to luck, probably because he was not able to attribute it to any specific reason:

> AS01: "It was only successful because there happened to be this particular family [of plants] I was interested in a particular genus and there was a key of a large genus of a large fungi... There was a key and it was a Norwegian key and I was in Sweden and it was very very relevant and probably it was just luck."

Uncertainty about the causes

Sometimes the user might be uncertain about the actual reason behind the success of his/her search. When the user is uncertain about the reason of success thus s/he is not able to attribute it to a specific element. In particular, if the user had not expected to be successful and the search ended up with success, the possibility of being uncertain can be higher.

> AS09: "I was looking for some detailed information on the blood supplier to kidney just a few days ago mainly to confirm what I've already read in elsewhere. I am pleased to say that I found from an American site quite a detailed description of it together with some nice pictures ... well it was successful partly because I didn't think I'd be able to find the information and because I was rather heartened and surprised that I did find it. I mean it took three or four attempts refining the search to find it. ...well, I went to, I suspected it would have been PubMed and found some recent articles."

Combination of factors

Data analysis indicated that in some cases users attributed their success to a number of reasons. For example, in the quote below the user mentioned a few reasons for his success:

> AS07: "Because I knew the name of the department and I knew the location of the department and I knew if I put it in inverted commas it would search specifically. So, I knew exactly what I wanted to get. I knew the department name, I knew that it was the research section of that department and I knew where I wanted to go to. So I knew all of those things and because I have put all the information in the search engine took me straight there which is good, which is exactly what I wanted."

Unsuccessful searches

Similarly, unsuccessful searches were attributed to a range of factors, including lack of domain knowledge, unavailability of information, lack of search skills, lack of linguistic and conceptual ability, user or system errors, lack of sufficient effort, search tools deficiency, deficiency of website design, bad luck, or combination of some of these factors. Besides, sometimes searchers are not able to figure out why they failed to fulfil the search successfully and they are uncertain about the reason/reasons behind their failure which forms another category that is uncertainty about the causes. More explanation about each kind of attribution for unsuccessful searches with a number of examples from the dataset is presented as follows:

Lack of Domain Knowledge

When the user is not an expert of the area of search, therefore, s/he does not know much about the area and it might be somehow difficult for him/her to develop an efficient search strategy. In particular, selecting appropriate search terms in an unfamiliar area might be challenging for the user. For example the following quote shows that the interviewee needed to carry out an initial search before the main search to find out apposite search terms:

> AS10: "I give some lectures on mitochondrial genetics and genetics is my area, but some of diseases with a lot of medical terminology and I don't know the definition of some of these medical words. So, you think where I am going to find what this is. One of the diseases is called Myoclonic Epilepsies, I know what epilepsy is, but Myoclonic I did not know. So, 'what do I do?' I always just go to the web, I don't look for a dictionary or book or try to find a medical book. I go to the web. I probably go to AltaVista just type Myoclonic ... when I don't know where to look I tend to use AltaVista. So, I am fairly conservative with what I use."

Similarly, another interviewee reported that he failed in a search because of lack of domain knowledge. He said:

> AS06: "I guess probably because I didn't know the right places to search. So, I just used Google and I'm sure there are better ways to find it. But I don't know what they are. And I guess it is not the kind of thing that I search very often. The kind of information [which] I am normally looking for is published papers and I know how to do that and it works fine. But it was just a strange request that I don't normally make."

Unavailability of information

Unavailability of information was cited by a number of interviewees as reasons for a failed search. In some cases, this could be due to the specialised nature of what was being sought. For instance one of them said:

> AS15: "I think for this particular search example yes. I think it was most likely that actually nobody has produced a website or a webpage with that information on ... I think it was highly specialized. I was looking for an aspect of the theory. You know, it is a very specific area of the theory and probably a limited number of people in the world are studying this aspect ... it wouldn't be something for instance that somebody would perhaps put in a webpage for undergraduate teaching. It was more specialized than that. So, it is limiting the number of people who would be interested in ... I would believe nobody bother to construct a webpage to explain this aspect of theory."

The same interviewee also reported the recency of the topic as the reason of unavailability of his required information. He said:

> AS15: "It is quite possible that nobody has yet got a website to explain that area very well or at least the aspect that I was interested in or it is possible that my search was not specific enough,

but I did try for quite a while and I did try quite a lot of different keywords. So, it is most likely that there isn't actually a nice website or an explanation about the theory."

Faced with an apparent unavailability of information, some searchers maintained a high level of self-confidence that search failure was not due to any limitations in their own abilities. For instance, the quote below shows this participant had enough confidence in his ability to conduct the search sufficiently well and if failure happens he will blame the web. He said:

RS19: "If I am being honest, if I fail to find something I guess on the line I blame on the web and not my ability to search. Because if you are looking for something very specific I guess I have been honest I feel you should be able with a correct search term to bring it up in a good search engine."

To the extent that some searchers felt able to conclude that what they were searching for simply did not exist. For example, one of them said:

AS13: "I mean often times when we are using genetic information you do a search with genetic information and you don't find what you are hoping for, but that is not a failure it just tells you that you were hoping for something that didn't exist."

Lack of search skills

Others, however, felt more inclined to blame their own lack of search skill – which by no means implied a lack of experience. For example, one of the participants reported:

AS09: "I spend a lot of time searching the web and in some respects I am quite professional at it, but if I fail I am more incline to think that is because I haven't adopted the best search strategy rather than it not being there. That is my first thought in most cases ... my first reaction is not to blame the web it is to blame myself."

The sheer volume of material available on the web was perceived by some to shift the balance of probability that a lack of skill, rather than information unavailability, was to blame for a failed search. The quote below is an example:

AS09: "I have a blind faith that there is more out there than I have found but I can't categorically say that there is everything there. Realistically there can't be, but as I said several times I think it is fairly often my fault in not searching adequately".

Indeed, such apparent ready availability of information on the web coupled with search failure can lead to feelings of inadequacy and frustration. For example, the quote below shows how one of the interviewee described her feeling of frustration after failing to succeed a search:

RS21: "Frustration probably more than anything and disappointment I think. It is frustrating because you think you should be able to find anything you need on the internet because it is huge. So, you think it is bound to be somewhere, but it is frustrating when it is not or you can't actually [find it]. You tend to blame yourself because you searched inadequately maybe. You are not searching in the right way. Yeah, it is frustration more than anything."

Lack of linguistic and conceptual ability

Perceived lack of linguistic ability has been reported as one of the reasons for search failure. For instance, the quote below shows how one of the interviewees was disappointed of his inability to select appropriate keywords. He said:

> AS11: "I have never been successful in refining words sufficiently to pull out exactly what I wanted."

Another interviewee had similar perception and she supposed that she is not sufficiently competent to narrow down a search adequately. She said:

> RS17: "But I think that is my problem. I don't use specific enough search terms."

As a consequence of inability to select apposite search terms she considered herself as an unsuccessful searcher. She reported:

> RS17: "I am not a very good searcher. My keywords are useless. I don't really use keywords properly, I don't think. So, it takes me longer to narrow down my searches and then I will get frustrated... So, I am not that successful."

Similarly, another interviewee ascribed his failure to inappropriateness of his search strategy. He believed that failure happens as a result of inadequacy of his own search strategy and not because of the deficiency of search engines. He said:

> AS10: "So, it is not a failure of the search engine. It is a failure of my search approach to find what I wanted ... I am not putting the correct words in to enable the search engine to find it".

Indeed, some interviewees displayed a strong confidence in web search tools – even to some extent to compensate for limitations in their own searching abilities. For example, one of the interviewee reported:

> RS18: "I trust Google, if you want. So, if I type something in and Google doesn't find it I trust it is not out there. So, if I can not find it. So, it is not my fault then. It is not there. So, in a way it seems that I completely trust Google maybe I shouldn't. If I would not trust Google, then I maybe more open to blame myself. That is true, yeah."

User or system errors

User or system errors represented another factor to which search failure was attributed. Searchers occasionally acknowledged being to blame through, for example, misspelling. The quote below shows an example of users' perceptions of attribution of failure to an error which in this case is a user's error:

> AS07: "I've had problems when I haven't been able to find something because I misspelled the name of somebody ... there are two vowels on the name and I've typed in one."

System errors were also reported, whether relating to an assumed temporary unavailability of particular websites:

> RS17: "I was looking for ... a mass spectrum from mass spectrometry, spectrum of a particular compound. I do have some chemistry websites that I looked for on this and I found it. But I couldn't find the breakdown products which are a very specific thing to look for ... I was not successful in finding the breakdown products or information relating to the breakdown products. So, that was an unsuccessful search even with the specific keywords, you know, the "compounds" itself and "mass spectrometry" [which are] quite specific keywords, but I was unsuccessful. Now whether there were two sites that were down, but I tried to access but they were down... they may give me the information but they were not available at that time."

Or to assumed system behavioural errors such as PubMed's "funny half hours" experienced by the following searcher:

> RS23: "I also know there is a small problem with PubMed that it does have funny half hours when it won't find what you are looking for, you know you can put the name of someone that you know and you can find a reference on there one day and it would be there and occasionally things disappear. It just has these mad minutes. Things disappear from PubMed and come back again."

The same interviewee went on to surmise a possible cause for such inconsistent behaviour in the system. She said:

> RS23: "Maybe like I've said like PubMed can lose data every so often or systems go down, you know it happens because ultimately they all run by people and people get sick and people have days off. So, the internet sometimes has days off".

Occasionally what were perceived as system inconsistencies arguably resulted more from misunderstandings of the nature of search facilities being used – though the following interviewee did acknowledge this possibility:

> RS23: "I also found that things would actually come up as hits in Google web search that didn't come up in Google image search even though there were images within the web hits ... then I guess that is to do with the way that they categories things and I don't really know how it works."

Lack of sufficient effort

Lack of sufficient effort was another cause to which search failure was attributed. This could be a result of being pressed for time. For example, the following quote shows that one of the participant believed that failure had happened because of making inadequate effort to carry out the search as a result of time shortage. He said:

> RS16: "Because I was doing the search in work time. I wasn't putting in a lot of effort and if I'd put in more effort into the search and then I would have found the information elsewhere. It was a quite a quick general search."

Similarly, another interviewee reported an example of his failed searches in which he attributed the failure to the lack of effort. He said:

> PS37: "I wasn't successful because I didn't have much time. So, it wasn't a very thorough search. I didn't look on PubMed. I just looked on Google, but it just came up with articles that had used this strain rather than what the strain actually was itself."

Moreover, the need to spend time and effort filtering information retrieved was acknowledge by another interviewee. He said:

> AS08: "...there is always a problem of filtering the information that you get. Obviously you get so much information that it can be difficult to pick up those things that are really of use. Having said that, it is quicker than the old fashioned paper based ways of searching through literature, and it is broader. So, from that point is good but you have to take downside of need to filter the information that you get."

However, some interviewees felt they lacked sufficient patience to do this. For example, the quote below shows that the interviewee was aware of this problem but she is not patient to carry on the search in anticipation of gaining satisfactory results. She said:

> RS17: "Sometimes I will lose my patience and not continue. If I am not getting there in so many stages, then I'll give up. So, that is a complete unsuccessful search. And that is the patience thing. I don't have the patience to go through all the links that has been given to me... I think probably if you put in a search and you get to a site that looks promising, say, a chemistry site and I am looking for a particular compound but there is too many links to go and find one is better. So, I can probably waste a lot of time saying 'no!' I want a better site."

The same interviewee felt that impatience and unwillingness to spend the time and effort narrowing excessively broad searches derived from a basic inability to select appropriately specific keywords – an ability, she felt, that comes more easily to the young:

> RS17: "I think that [selecting effective keywords] is something that you have to learn. If I go and will put in say 'Arabidopsis' then I'll get thirty billion hits whereas I am just not being specific but that is just the way that my brain works. Again I think the younger people have more aptitude for this because they have been taught you know "be more specific" in your searching and it won't take you so long. I just can't seem to grasp that concept which is why I think I am not terribly successful and then I don't have the patience to narrow it down and still you have three billion hits. And I'm thinking 'no, no, no!' it is just too much, and sometimes it can be overwhelming really."

Search tools deficiency

Some of the participants attributed their failure in search to the inefficiency of the search facilities that they had used. The following two quotes are examples of this kind of attribution.

AS02: "Generally the search tools I use are very good, but I don't have any way of verifying them. So, say for example, I ask a question when I want to know how many of these there are in the world and it comes back and says 333. Is '333' correct? I don't have independent tools to compare. So, that can be a problem. So, I might think that I am correct. I might trust the tool. But it might be wrong and I know these tools make mistakes."

RS21: "There hasn't been such good central and searchable database of methods, like literature we have got searchable databases or DNA or protein sequence as a searchable database. But there is only recently people started to put together searchable databases of methods and sometimes these are subject to subscription and if the university isn't subscribed with it you can't access to that particular database. So, you have to look elsewhere ... I mean obviously people have books in the lab but if you want something a bit out of ordinary or whatever then you have to find that elsewhere and most of the time you tend to go to the internet because you will find it and in the end you probably end up searching the literature to find where someone did it instead of actually having specific place you can search."

Deficiency of website design

Search tools are not the only device which enables users to carry out the search and there are other elements form the system side which need to be efficient in order to bring about success including the efficiency of website design. Sometimes users are able to find their required website but because of the poor design they are not able to gain access to their required information.

AS12: "I do sometimes find it incredibly difficult to find contact information of people both for companies and for other academics. I find for lots of university websites it takes a long time to find a person. If all you have is a name and a university then it can be very frustrating. Some of the websites are really good but in others it is practically impossible. For example, to get a telephone number that may be because they have some security issues or something that they are concerned about. ... I actually think that the information is just not there in some cases or probably deliberately not there ...so, that is a kind of scenario that I found quite frustrating ... there is incredible variation and ease of navigation is not always good."

Similarly, another interviewee attributed her failure in search to the poor design of the website which reflected to its searchability as well:

AS05: "I think the web site where we can actually get the lights was poorly set up. So, possibly if I had gone to the page twenty of Google I would have found it, but I always give up after about second or third page. I lose interest."

Bad luck

Being unlucky is another element that might be considered by the user as the reason of failure:

AS01: "Well, it is not a problem with the web. It is just a question of the community, I mean the community, taxonomy community, taxonomists are very specialized, there are old texts and there are old keys which are in paper copies, but the keys are not always robust there are relatively few people in this field so some families are treated very very thoroughly and some others not at all. So, this is a case where there hasn't be probably a modern treatment and people wanted to test out and I was just been unlucky."

Uncertainty about causes

A number of interviewees reported that they were uncertain why particular searches were unsuccessful. They were unsure whether it was their fault or not. For example:

AS12: "I don't always find what I am looking for and one of the biggest challenges is ... it's hard to know, you never know, whether you can't find it because it is not there or if you can't find it because your search method is inappropriate. ... You don't know when to stop, when to give up, because you don't know if you are wasting your time and there is literally no information out there online or whether you just not searched appropriately."

In a similar way, another participant expressed his uncertainty about the reason of his search failure. He was doubtful between whether he should attribute the failure to an internal element such (e.g. the user's error) or to an external factor (e.g. the search tools' deficiency):

AS07: "The problem is with searching is that if I don't find it I don't know whether it is my fault because I have misspelled something or whether it is the search engine's fault that it is there but I haven't been able to find it."

Combination of factors

In many cases the participants attributed their search failure to a combination of factors. For example, failure could be due to a combination of lack of enough information and not using apposite search terms:

RS16: "Yeah occasionally I do [fail on the web] which is probably a combination of lack of the research actually have been done and lack of appropriate keywords."

Internal and external attributions for success/failure

Regarding Attribution Theory[37], the above factors for success and failure can be categorised into two groups of internal and external factors. Examples of interview extracts describing search success and failure, classified in terms of internal and external attributions which are present in table 10.2.

[37] Explanation about internal and external factors in Attribution Theory is presented in chapter two.

Table 10.2: Internal and external factors related to success and failure in search

	Internal Factors	External Factors
Success	Domain knowledge Search experience Linguistic and conceptual ability Confidence Persistence	Availability of information Search tools efficiency Website structure Good fortune
Failure	Lack of domain knowledge Lack of search skills Lack of sufficient effort User errors Lack of linguistic and conceptual ability	Unavailability of information System error Lack of linguistic and conceptual ability Search tools deficiency Deficiency of website design Bad Luck

Success attributed to internal factors

Data analysis showed sometimes users attribute their success to different internal elements, those elements which are under the control of the user. For example, the next two quotations show that one of the participants ascribed her success to her ability to select suitable search terms as a result of linguistic and conceptual ability and the other one to his search experience:

> RS17: "Because my keywords were not too bad. My keywords were quite good ... So, I think that was successful."

> RS20: "At least two things [reasons of success]. One of them is my experience in using the internet in which I would say I have strong experience, a high level of experience. I have been doing it for a long time for a lot of different reasons and I have successfully managed to retrieve information which has been useful ...."

Success attributed to external factors

In contrast, data analysis showed sometimes users attribute their success to different external factors, those factors which are beyond the control of the user. For instance, one of the participants ascribed her success to the efficiency of the employed search tool which had provided her with advanced search facilities to carry out the search based on different criteria:

> RS22: "Well, PubMed is a facility specifically designed to search for papers and it gives you various criteria you can use and you can really narrow down the results. So, you can put in

author, journal, year of publication and any keywords as well and so often you can end up with one paper from thousands."

Similarly, some other interviewees who were very satisfied with the search tool that they had used explicitly attributed their success to the effectiveness of the search system and said:

RS16: "Well, Web of Knowledge is always successful because it is a ready to search database and it is structured in a very definite way."

RS19: "I think it thanks to Web of Knowledge. I think it is superbly set up. It is very fast. Yeah I think Web of Knowledge is extremely well-designed. Only when you fail on it when you are too broad in your search term and you come back with ten thousands possible articles and then you have to learn how to be more specific. But I think it is a quality of Web of Knowledge. I mean it works very well for me."

The same interviewee generalised his satisfaction with Web of Knowledge to other web-based search tools. In his opinion, web-based search tools are generally efficient enough to satisfy users' information needs. Moreover, he believed that because it is very easy to use these facilities, there is no need to be an expert or learn sophisticated search techniques to employ them effectively. He said:

RS19: "Usually I am able to find it because like what I said because search engines and web based databases are so good and you don't have to be particularly great to be able to use them. You don't have to be an expert to use them. Basically I mean that is why I don't blame myself when I can't find something, because I don't think you need to be an expert to use it. Therefore I can't say to myself oh! ...[38] You need to be lot better at this go away and teach yourself more about this. Because my feeling is that you don't have to know a lot to be able to use the web very well. Therefore, if I can't find it, is not my fault because even someone with basic skills is already good enough to use the web effectively."

Success attributed to both internal and external factors

Sometimes users attribute their success to both internal and external factors. For example, the following quote shows that the interview ascribed his success to having sufficient details about the search topic as a result of domain knowledge, familiarity with the employed search tool because of using it before which are internal factors and also to the availability of the required information, existence of well-maintained information sources about the search topic with up-to-date information which are external factors.

RS25: "Well, it was straightforward and easy because I knew where she worked. I have not met her but my colleague at IRRI has met her when he visited her in the UK. I used to work in the

[38] The interviewee referred to his first name in this part, which has been removed from the quote to keep the anonymity.

department of plant science of Oxford. So, I am familiar with the website. I have actually visited their website fairly recently to see whether my old friends are still there. So, it was easy to find the website, [and] easy to find her webpage. And I think the other critical thing in contrast with the person I mentioned earlier [who does not keep his webpage up-to-date] is she obviously keeps her webpage up-to-date."

Similarly, the next quote shows the interviewee reported a number of factors as the reasons for success. The interviewee attributed her success to her search experience which enabled her to begin the search with an appropriate search tool and also the efficiency of the employed search tool and effective design of the website:

RS21: "I think to start with I went to a good websites ... because I was looking for something that was in Scotland and tourism and then I went to Scottish tourism website and started there. I think in that case because these website are like more kind of more nationwide type things. I think they are very often very good at linking out to other things and I think that how I have been quite successful and quite quick at doing these things because they have very good links to other sites that can help you to get the right information."

Failure attributed to internal factors

Sometimes searches attribute their failure to internal factors. For example, the following quote shows that the interviewee ascribed her failure to not selecting an appropriate search tool, not adopting an efficient way of search and not spending enough time to fulfil the search:

RS22: "Because I don't know the best place to look or the best way to go about it. I mean there are things like Google, I don't know a lot about them, but they must be quite sophisticated to list the most relevant hits at the top of the list. So, frustrated and maybe wonder if there is a better way to do it that I don't know about. ... Possibly it is hidden and I am not managing to put in the correct search terms or not looking in the right place."

Similarly, another interviewee ascribed his failure to his lack of search skills to select an apposite search tool and also his unfamiliarity with the search topic due to lack of domain knowledge. He said:

AS06: "I guess probably because I didn't know the right places to search. So, I just used Google and I'm sure there are better ways to find it. But I don't know what they are. And I guess it is not the kind of thing that I search very often. The kind of information [which] I am normally looking for is published papers and I know how to do that and it works fine. But it was just a strange request that I don't normally make."

Failure attributed to external factors

In some cases failure is attributed to external elements which are out of the control of the user. For instance, one of the interviewee supposed he failed to succeed a search because of the unavailability of the required information on the web as a result of the deliberate action of the information owners to hide it. He said:

RS24: "It was a failure, well, because it is private information and people consider it as private and they don't want to put it on the web. The address and telephone number or e-mail address of scientists has to be public, but some people keep it private."

Another interviewee attributed her search failure to three factors which all are external including scarcity of information about her search topic, high level of specificity and obscurity of the search topic, unavailability of the required information on the web:

PS27: "Because I work with these megaspores and there might be only be two people in the world that have looked at these megaspores. Sometimes I want to search for these things and find papers and they just aren't there because they are so obscure. So, I'm looking for old papers that just aren't on the web. So, I am not frustrated by that I just accept it because they [papers] are very old things, very obscure things ... long Latin names that mean to me and possibly five other people in the world."

Failure attributed to both internal and external factors

Failure in search might be attributed to both internal and external elements. For instance, one of the participants supposed she failed in a search as a result of the combination of two internal factors and two external ones. She first blamed herself because of using an inappropriate search tool and also because of not putting sufficient effort to explore other search facilities. Moreover, she attributed her failure to the inefficiency of the employed search tool, and occurrence of some unexpected and unexplained technical errors. She said:

RS23: "Maybe I have just looked in a wrong place. It was quite a surprise that I didn't find it. I just need to look in a couple of other places. And also computers don't replace people just go and ask the person who told me about this piece of research it might be quicker to do that ... I also know there is a small problem with PubMed that it does have funny half hours when it won't find what you are looking for, you know you can put the name of someone that you know and you can find a reference on there one day and it would be there and occasionally things disappear. It just has these mad minutes. Things disappear from PubMed and come back again."

In a similar way, another participant believed there are a combination of user fault, search tool's deficiency and unavailability of the required information on the web that cause failure:

RS20: "Yeah, in general if search engine can't find any result, it is combination of me, the search engine and the data or the information. They are not just working on that day to answer the question. But I do look at it that it is me not the tool that I use of the community of information out there which I can't access."

The details of their reasoning for causes of successful and failed searches have been illustrated in appendix 7. The Table reports the result of exploring all participants' attributions of their successful and failed searches.

As table in appendix 7 shows, in a large majority of cases (82.4%), they clearly attributed each search to either internal (e.g. ability or effort) or external (e.g. luck or information not

being available) factors. The pattern of such relationships was analysed, and mapped onto those that would be predicted by Locus of Control and Attribution Theory. It is interesting to note that uncertainty about causes of failure was reported in 4 of the total of 20 reported "failed searches". Uncertainty relating to causes of success was found in only 1 of the 20 total reports of "successful searches".

Where interviewees did not report uncertainty (i.e. in 69 of the 74 reported searches), reasons can be categorised in terms of either (a) internal attributions of success or failure – i.e. relating to themselves (e.g. ability or lack of it, or effort on their part), or (b) external attributions – i.e. outside their control (e.g. luck, or information not being available to be found). Table 10.3 shows the frequency of "internal" and "external" causes for both successful and unsuccessful searches.

Table 10.3: Correspondence of internal/external attributions to successful and unsuccessful searches

	Internal factors	External factors	Both Internal and External	Uncertain
Successful searches	12	4	3	1
Unsuccessful searches	4	10	2	4
Both successful and unsuccessful searches	3 *[= 6 searches]*	11 *[22 searches]*	3 *[= 6 searches]*	0

Each of the 37 interviewees could recall a successful and an unsuccessful search, giving a total of 74 recalled searches. The table shows, for example, that 12 of the successful searches were attributed exclusively to internal factors, four exclusively to external factors. Ten of the unsuccessful searches were attributed exclusively to external factors, four to internal factors.

Three successful and two unsuccessful searches were attributed to a mixture of both internal and external factors. In three cases, the interviewees attributed both successful and unsuccessful searches (i.e. six searches) exclusively to internal factors, in 11 cases exclusively to external factors. In relation to 58 of the 74 searches (78.4%), interviewees made either exclusively internal or exclusively external attributions. Eleven (15.3%) reported mixed internal/external attributions, and five (6.9%) were uncertain.

Of those 58 searches where attributions were made exclusively to either internal or external factors, in 28 (48.3%) both success and failure were attributed to the same (internal or external) factors.

Twenty two (37.9%) attributed both success and failure to external factors, six (10.3%) attributing both to internal factors. In the case of the remaining 30 (51.7%) of the 58 searches, 22 (73.3%) conformed to a pattern whereby success was attributed exclusively to internal factors, and failure exclusively to external factors. In eight of the searches (26.6%) the

converse pattern was reported, in which success was attributed to external and failure to internal factors.

Participants' self-efficacy

Figure 10.1 shows the data classified into five groups, along a dimension ranging from extreme external to extreme internal Locus of Control.

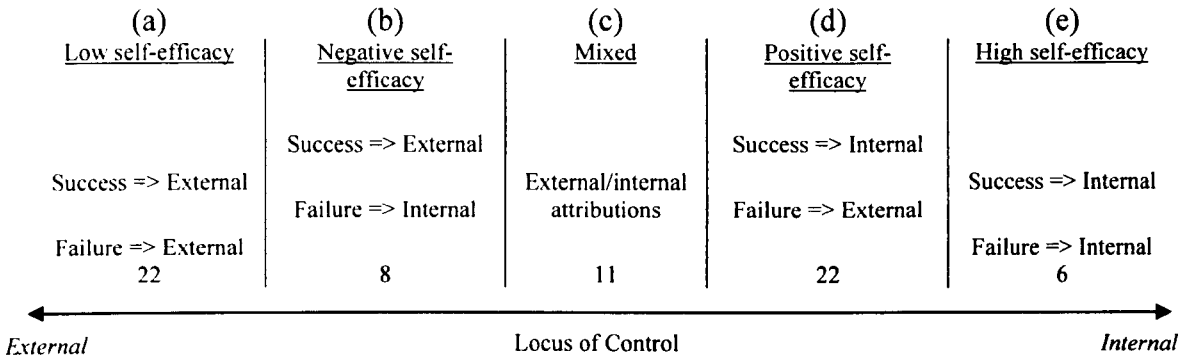

Figure 10.1: Searches grouped according to internal and external attribution and search success and failure.

The groups are as follows:

Low self-efficacy group

a) A "low self-efficacy" group (accounting for 22 searches) consisting of searchers with a high external locus of control in that they perceive not only failure – but also success – to be outside their control. The attribution behaviour of this group is not as self-deprecating as that characterising group (b) below, who not only fail to take any personal credit for success, but also blames themselves for failure.

Negative self-efficacy group

b) A smaller "negative self-efficacy" group (accounting for 8 searches) who attribute success to relatively external factors, and failure to relatively internal factors. Self-efficacy is negative in the sense that the searcher perceives himself/herself to have been effective only in bringing about failure – not success. This pattern of behaviour accords with that predicted by Attribution Theory for the typical unconfident low achiever.

Mixed group

c) A "mixed" group (11 searches) in which internal and external attributions are not linked to search success or failure. This group includes (i) successful searches attributed to mixed internal/external attributions (3 searches); (ii) unsuccessful searches attributed to mixed

internal/external attributions (2 searches); and (iii) both successful and unsuccessful searches attributed to mixed internal/external attributions (6 searches).

Positive self-efficacy group

d) A "positive self-efficacy" group (accounting for 22 searches) in which success is attributed by searchers to relatively internal factors – their own abilities and efforts, and failure is attributed to relatively external factors such as the unavailability of relevant information "out there" available to be found, or search engine inefficiency. The attribution behaviour of this group accords with that predicted by Attribution Theory in relation to confident high achievers.

High self-efficacy group

e) A "high self-efficacy" group (accounting for 6 searches) consisting of searchers with a high internal locus of control in that they attribute not only success – but also failure – to internal factors. The attribution behaviour of this group is not as self-flattering as that of group (d) who, whilst accepting the credit for success, attributes failure to external factors beyond their control.

Model of self-efficacy in web searching

According to the above explanations this study hypothesizes the model of self-efficacy in web searching which is presented in figure 10.2. The model consists of five parts including confidence, pessimism, optimism, uncertainty and mixed zones.

In the confidence zone which belongs to those users with high self-efficacy they attribute both success and failure to internal elements. In contrast, users with low self-efficacy will accommodate in the uncertainty zone and they ascribe both success and failure to external factors. In the pessimism zone which is occupied by users with negative self-efficacy only successful searches are attributed to external factors and failed searches are ascribed to internal elements. In contrary, optimistic users with positive self-efficacy accommodate in positivism zone because they attribute their success to internal factors and their failure to external elements. Finally, in the middle of the model there is "mixed group" which shows those searches who attributed both successful and failed searches to both internal and external elements.

Comparing the model of self-efficacy with previous two models (i.e. cognitive model of information visibility on the web in chapter six, figure 6.1, and the extended model of information visibility integrating the concepts of success and failure in web searching, figure 7.1) shows arguably there might be some links between the current model, figure 10.1, and the previous two models. In fact, it can be argued that users with high self-efficacy who are in confidence zone in this model are more likely to have "anticipated success" in their searches associated with figure 7.1, and will be in the "bright zone" in figure 6.1.

Similarly, users in optimism zone who attribute their success to internal factors and their failure to external elements can be associated to the "veiled zone" in figure 6.1 and have "unexpected failure" in figure 7.1 which encourage them to put more effort to carry on the search after having initial failure because they ascribe their success to internal factors and are ready to put more effort to achieve it. In contrast, users in uncertainty zone with low self-efficacy who attribute both success and failure to external elements are more likely to be in the "dark zone" in figure 6.1 and experienced "unexplained" or "inevitable" failure which is associated with figure7.1 and give up the search more easily.

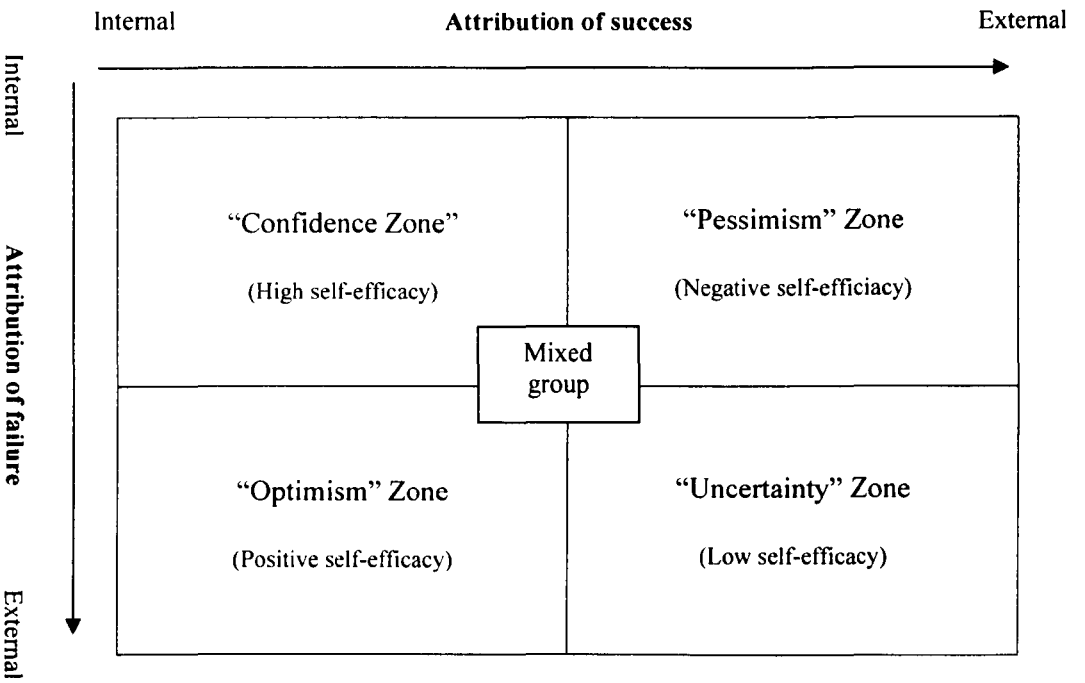

Figure 10.2: Model of self-efficacy in web searching

It is also important to acknowledge a more fundamental limitation. Critics adopting a social constructivist epistemological position (Burr, 1995; Gergen and Gergen, 2003) have emphasised, both generally and specifically in relation to Attribution Theory, the potential dangers of assuming that verbal accounts can necessarily be interpreted as undistorted "windows on the mind". Edwards and Potter (1993: 23), for example, note that:

> "[L]anguage is by no means a transparent, or neutral, system for conveying information but rather the words that people use to describe simple, everyday actions and states carry with them powerful implications for the causal explanation of those events."

Attributions are discursive actions, engaged in within particular contexts and situations, for particular purposes. Edwards and Potter (1993: 30) note that traditional experimental studies of attribution may be criticised to the extent that they deny subjects the essential opportunity to choose their own descriptions of events, and that compared to findings from data where subjects are allowed to express themselves more naturally, their findings represent "only a

narrow focus of explanatory concern". Within a social constructivist perspective, it is important that researchers consider the extent to which the research findings are essentially constructed by the research question and methodology adopted (Willig, 1999).

Such a perspective invites sanguine reflection on the methodology of the study. A danger of deductive research is that researchers may find what they are looking for not entirely because it is "there", but at least to some extent because it is what they are looking for. Untested contingencies (for example, possibly self-reinforcing research values and conceptions shared by researcher and interviewees) may mask the possible dependence of findings on hidden aspects of the research context.

However, militating against such dangers is the fact that the interview data analysed in the study were collected and inductively analysed without reference to the psychological theories subsequently adopted for their interpretation. The data collection entailed open-ended questions in which the participants were free to choose their own descriptions relating to web searching. In fact, it was inspection of the results of this inductive analysis that suggested to the researcher resonances with these theories.

Summary

This chapter encapsulates the findings of the current study about the participants' attribution of their successful and failed searches in the light of Self-efficacy and Attribution Theory. The chapter concludes these theories are useful conceptual frameworks to illuminate our understanding of web searching. Therefore, the result presented in the chapter can have implications particularly for better understanding of searchers' motivations, and for the design and development of more effective search training programmes. The chapter illustrates the factors perceived to determine success/failure.

For successful searches domain knowledge, availability of information, search experience, linguistic and conceptual ability, confidence, persistence, search tools efficiency, website structure, good fortune, or a combinations of some of these factors are the reasons that the participants mentioned. For unsuccessful searches, lack of domain knowledge, unavailability of information, lack of search skills, lack of linguistic and conceptual ability, user or system errors, lack of sufficient effort, search tools deficiency, deficiency of website design, bad luck, combination of factors, or uncertainty about the causes are the identified attributions.

Internal and external attributions for successful and failed searches are identified and classified in the chapter. Participants' self-efficacy are categorised into five groups of low self-efficacy, negative self-efficacy, mixed, positive self-efficacy and high self-efficacy groups and the model of self-efficacy in web searching is introduced and the links between this model and the previous models in chapter six and seven are illustrated.

CHAPTER ELEVEN:

BOUNDED RATIONALITY AND SATISFICING

CHAPTER 11: BOUNDED RATIONALITY AND SATISFICING

"Whether we like it or not, each of us is constrained by limits on what we can do and feel. To ignore these limits leads to denial and eventually to failure. To achieve excellence, we must first understand the reality of the everyday, with all its demands and potential frustrations." Mihaly Csikszentmihalyi

Introduction

Deductive analysis of the dataset with regards to Locus of Control, Attribution Theory and Self-efficacy, presented in chapter ten, strengthened the inductive results about users' conceptualizations of success/failure which has been presented in chapter seven. Therefore, the researcher adopted a similar strategy to fortify the inductive findings about users' awareness of and reaction to missed information presented in chapter eight.

To achieve this aim, he examined the dataset in the light of another conceptual framework to find out to what extent existing theoretical frameworks can be applicable to support the inductive findings of the study. Accordingly, this chapter presents the result of another phase of deductive analysis based on Simon's Bounded Rationality and Satisficing theory[39]. In fact, the chapter addresses EmRQ3 [40] which is:

- To what extent is the theory of Bounded Rationality and Satisficing useful in interpreting users' awareness of and reaction to missed information while web searching?

Although there are detailed explanations about this theory in chapter two, here a brief description is presented to remind the reader.

Bounded rationality

Simon's Bounded Rationality and Satisficing Theory is a well-established conceptual framework which is useful to address human behaviour related to decision-making in general. This theory suggests that in real situations human kind is only "rational enough" or "partly rational" rather than being "absolutely rational". Therefore, people usually make their decisions within specific boundaries which stop them from exploring all possible options. Time constraints, information overload and physical constraints are three categories in the Bounded Rationality theory which are influential elements on people's decision making behaviour. One of the interesting parts of the Simon's theory is the concept of "stop rule". Agosto (2002) stated:

"Satisficing acts as a "stop rule" (Simon, [1979], p. 4) - once an acceptable alternative is found, the decision maker concludes the decision process. Nonetheless, satisficing does not limit the decision

[39] There are some explanations about this theory in chapter two.

[40] Third Emergent Research Question

> maker to one deciding factor: When the criterion of problem solution or action has more than one dimension, there is the matter of calculating the relative merits of several alternatives, one of which may be preferred along one dimension, another along another The satisficing rule stipulates that search stops when a solution has been found that is good enough along *all* dimensions. (Simon, [1979], p. 3)

Agosto (2002) explored young people's web-based decision-making. She developed a framework for coding interview transcripts based on Simon's theory, and was able to map interview transcripts onto all main elements of the theory. This framework was developed to code interview data from ninth- and tenth-grade schoolchildren concerning their web-based decision-making in relation to evaluating websites, as opposed to searching.

This chapter utilizes Agosto's coding framework to assess the extent to which elements of Bounded Rationality theory were present in – and thus may illuminate – the dataset in the study. However, Agosto's additional "personal preference" codes have not been included in this study since they relate to website evaluation and are arguably not relevant to web searching. The framework is shown in table 11.1.

Table 11.1: Bounded Rationality and Satisficing coding scheme based on Agosto's (2002)

Bounded Rationality	Time constraints ▪ Imposed ▪ Self-generated Information overload ▪ Textual overload ▪ Outcome overload Physical constraints ▪ Discomfort ▪ Exertion
Satisficing	Reduction ▪ Known sites ▪ Synopsis ▪ Categorization Termination ▪ Acceptance ▪ Discomfort ▪ Boredom ▪ Time limits ▪ Snowballing

Time constraints

The data analysis revealed that shortage of time is one of the reasons for stopping searching. There are two types of time constraints: *imposed* and *self-generated.*

Imposed

Imposed time constraints happen as a result of external factors that are out of the searcher's control and limit the amount of the time that s/he can afford to spend on the search. The following two quotes are examples of the participants' ideas about imposed time limitations:

> RS17: "It is a crucial point these days. There aren't enough hours in a day anyway. How much time can you afford to spend searching the web? You have to limit it."

> RS24: "They want to make it open access to articles and this is fantastic and also there is huge amount of information available but there is no time to read everything."

Self-generated

Self-generated time limitations may result from searchers simply preferring not to spend time searching – as opposed to any specific limitation on the time they have available so to do if they wished. For example, one of the interviewee said:

> AS10: "I suppose I don't play with the web a lot. I don't experiment with it. I just, what I know works is okay and I don't want to waste time by doing it."

Similarly, another participant also reported that because of various tasks that she has to carry out during a day she has to limit the time that she spend on web searching:

> RS17: "There are a lot of things to do, always a lot of other things to do, I mean in the lab, down in the growth rooms. So, I mean always something to do. So, I limit probably more than most people how much time I would spend searching."

Another interviewee mentioned that although she knows there is much more information to find on the web, she prefers to not spend more than a specific amount of time on it:

> AS05: "I am not that interested in surfing the internet for pleasure. I don't always look up holidays and things. I booked a holiday about a month ago, but I just looked up one particular site and booked it immediately. I didn't go through the other ones. I went through a little bit but not that much. I mean some people I know spend a lot of time, but I don't."

Information overload

The enormous volume of information on the web can cause difficulties for searchers. The participants in this study indicated that they felt uncomfortable with having to deal with such huge amounts of information.

For example, the following excerpt is an example of users' views on this issue, in this information overload and time constraint is explained:

> RS17: "It is obvious that I miss lots. Again because although I am saying there is not a huge amount [on her search topic], there are still enough to keep you busy for half a day every day. There is enough information and just because of a sheer volume of information you know you are going to miss most of it ... so, I know I miss lots."

Information overload may be differentiated into textual and outcome overload. Textual overload is defined here as an inability or unwillingness to read information that one has identified as potentially relevant were it not for the sheer volume of it. Outcome overload represents a stage before textual overload, in that it is an inability or unwillingness to process large hit lists in order to establish what is and is not potentially relevant.

Textual overload

Although the web offers information in many different formats, still the major portion is in the form of textual materials. Sometimes web users feel overwhelmed by the substantial amount of textual resources that are potentially relevant to their needs. Examples from the dataset are presented below:

> AS12: "I would probably do a journal search rather than using the web-based resources. Because I would expect spending a lot of time going through thousands of hits which I'll have to do with, you know somebody's opinion in favour or somebody's opinion against it, legal debate is going on United States, and etcetera, etcetera. So, I suppose you could say that in fact I don't really want to spend as much time that I need to refine my searches on the web."

> AS08: "I am sure that I miss things every time, but that doesn't necessarily mean I fail. As I said, if you are looking for a specific piece of information then you get that, that's fine. A case would be looking for recent information on a particular bit of a cell. For example, we call it "cytoskeleton" and I know if I try to search for some recent general information on the "cytoskeleton" I can hit two or three reviews. I will not hit everything but I haven't got time to read everything anyway. So, as long as what I am getting is up-to-date and reliable ... that is a success. Unless you're looking for very specific items of information, but even then you are unlikely ever to get everything that is available and I don't think it is important you do get everything available".

In particular, some of the participants mentioned that it is more tiring for them to read a text on a computer screen than reading the same volume of text on paper and this causes more difficulties for them to deal with the huge volume of text on the web:

> RS24: "To read stuff ... is faster on paper than on screen. You can read two hundred pages everyday, but you can't read hundred pages on screen. It is impossible.

Outcome overload

The volume of retrieved items can often prevent searchers from scanning them all. Indeed, a number of studies have suggested that web users tend to look only at the first or first two pages of results (Spink, 2003; Spink and Jansen, 2004). The following quotes show some of the participants' opinions about this issue:

> RS17: "If I go and will put in say 'Arabidopsis' then I'll get thirty billion hits, whereas I am just not being specific. But that is just the way that my brain works."

> RS26: "You never make twenty five pages of search. You just do two, three or five pages and then you type some other words to see you can get it any better. You just don't go for twenty five pages one after the other".

> AS02: "... more frequently so much information comes up and it takes too long to go through it all. So, you are not able to come to a satisfactory conclusion. You actually just take the first or a few that they are on offer."

> AS05: "Possibly if I had gone to page twenty of Google, I would have found it. But I always give up after about second or third page. I lose interest."

Physical constraints

Discomfort

Using computers can be exhausting; particularly for those for whom such use is not a leisure activity (e.g. playing computer games). Resultant discomfort may be physical and/or mental, and may become a factor in decisions as to when to stop searching. For example, in the quote below the interviewee was complaining of the discomfort causing by searching the web:

> RS18: "It [web searching] is necessary, but I do not feel particularly comfortable [with it]. I do not like it...I spend so much time on screen already. I do it when it is necessary, but I am not enjoying it particularly."

One of the interviewees compared her attitude about web searching with somebody else's attitude and reported:

> RS17: "To him it [web searching] is a hobby, whereas to me it is a part of the job. So, the way that he approaches that is going to be different. You know computing is his big hobby ... anything to do with this machine ... to him is like a pleasure. So, he learns the search as part of his learning process. But I just use that to get my stuff. To me it is a chore, a task, a job that has to be done. So, it is not so pleasurable."

Another participant pointed out an important issue which is worthy of mention here. He highlighted the role of the user in carrying out the search procedure. He declared that although computers execute a large portion of the job of finding or displaying information, completion of a search session still requires effort on the part of the user to go through the retrieved

materials, and everyone has his/her physical constraints which inevitably affect the search procedure:

> AS08: "The machine itself will find things for you, but you're still the biological individual at the end of it. You've got to filter it through again. So, you also have the human element, and you could miss things [that] the computers have found for you."

Exertion

The level of required effort may also be an influential factor. Although usually the web provides end users with quick and easy access to information, sometimes web searching requires a considerable amount of effort. The following two quotes are examples of this issue:

> RS17: "Sometimes I will lose my patience and not continue. If I am not getting there in so many stages, then I'll give up... I mean there are some sites that I avoid because there are too many links within a site which really annoys me ... too many links jut to get where you want to get. If you bookmark everything you want to then you never get through your bookmark list. So, I don't bookmark everything. So, sites like that have too many links to get to ... it would take thirty clicks, you know, something crazy and that puts me off."

> AS14: "I think the difficulty of using Google is it generates lots of hits and quite a lot I spend a long time searching through for something that is relevant. Yeah I guess there are times that you don't find things ... It depends how long you are prepared to spend looking for it. So, if you have got a long time to spend and you think you'll find it at the end, but after ... the first few ... quickly I'll just give up."

However, it is worth noting that conversely, the same factor may militate in favour of web searching as opposed to other forms of information seeking:

> AS07: "In general I find it very convenient. It is good to have information accessible from my office because although the library isn't very far away, it is easer to have five minutes to go on the web than just walk to the library."

Satisficing

It became clear from the data analysis that a number of interviewees brought their searches to a halt before they felt they had found all relevant information. In particular, the concept of "compromised" as opposed to "ideal" searching – i.e. stopping a search when it was felt to be "good enough" or further searching was deemed impractical for various reasons – was a reason for bringing searches to a halt:

> RS20: "If there are thousand documents and I have found five and it allows me to do my job I don't really mind that I haven't read 995. As long as I have done the job and make the argument that needs to be made, then it is okay. I just move on".

AS03: "Probably I always miss something. I think because we don't have, we never have to get everything. As long as I am satisfied with the products that I have found, that is it. I don't want anymore about it which might be same or similar or repetitive."

In terms of satisficing strategies, it was possible to map the data onto all the categories identified by Agosto (2002), as illustrated in the sections below.

Reduction

One satisficing strategy is to reduce the search task, for example by restricting searching to known sites, making use of synopses or using categorized sites.

Known sites

A number of participants in this study mentioned that they usually have a specific set of frequently visited websites that they have found useful to satisfy their information needs. They usually keep a record of their frequently used websites in different formats including bookmarking them in a favourites list. The following three quotes are examples of known sites:

AS10: "I use a limited number of sites for genome sequences. For journals, publications, [and] for electronic journals, I use a very limited number of sites. We are talking about three or four sites."

RS19: "I go to news pages and catch up on news, but I don't search for them because they are on my favourites."

AS15: "I looked for the book on the Amazon website. I had that bookmarked anyway. Amazon is a quite regular source of information of books for here and for home. So, that was a nice and easy one."

One of the participants declared that having a list of bookmarked websites helps her to carry out better defined searches:

RS17: "I have sort of journals and specialized sets and I have bookmarked so and so and also websites that are specialized in plant science. So, I can go to them which narrows down my search ... and I use that."

Synopsis

One of the strategies that help users gain more satisfactory results through web searching is use of synopsis in different formats. For example one of the interviewees mentioned that he uses a web-based SDI [41] system which helps him reduce the amount of accessible information:

[41] Selective Dissemination of Information

AS13: "I have signed up to ... PubCrawler [http://pubcrawler.gen.tcd.ie/] ... and all it does is search for new publications with my keywords and this is this week [referring to the computer screen] and there are about 130 a week. So, I have about 10 or 15 keywords and so it just pulls up all the new papers every week ... this service is very useful because there is such huge amount of literature. You can't look through the titles in all the journals every time they come out. I mean I get e-mailed the contents list of about three or four journals but not of 30 journals and of course the information, the papers that I would want could be published in one of thirty journals. So, this is the only way that I can do it. Because ... it prompts me to actually look through this list each week and then I just download directly into EndNote. And if there are any interesting ones, I just print them out straightaway."

One of the participants used Current Contents weekly report as a synopsis of newly published materials. He had been using it for a long time even before the emergence of the web and he still does. Although he does use the web-based service, he believed the printed format of this service is faster:

RS24: "There was a big index book called Index Medicus and Current Contents. They were coming every week. So, instead of using computers it was manual and thinking a little bit about that it was faster to do it manually than computers. I could look at one issue of Current Contents in about eight to ten minutes. Now I have to switch on the computer, log into Current Contents, type the keywords, find information. So, manually was faster."

Categorization

Similar to the notion of known sites is that of categorized sites, these two being differentiated in Agosto's (2002) coding scheme. Likewise, the data showed that users usually categorize the resources that they use in a different ways. For example, the following quote shows how this interviewee categorized his work-related websites:

AS09: "Well several categories, a lot of images because my teaching is very much into producing illustrations of human anatomy, [and] human biology. So, a fair amount of time is spent looking for and finding links to image databases."

However, another participant had a more detailed categorization of different kinds of searches that she does. She firstly categorized her searches into main category of work-related and everyday life searches. Furthermore, she classified work related searches into different subcategories (e.g. scientific papers, experimental methods, and purchasing laboratory equipment). She also divided everyday life searches into several subcategories (e.g. house hunting, travel information, and shopping). She said:

RS22: "In the work environment online journals, general protocols, information about where different labs are and who works in them and maybe facilities and services available for experiment. For private reasons, I did use it [the web] quite a lot when I was buying a house searching estate agents' pages for properties. I use it for weather forecasts, sometimes TV guides or cinema guides, and yesterday I used it to find out if there is any problem with the trams because of G8 summits. And e-mail I have Yahoo account as well as the university. Sometimes I

go to the websites of shops, I don't know like B&Q and something like that. Sometimes I use Scoot [http://www.scoot.co.uk/] to find like a business, a local business in Sheffield ... Recently I wanted to hire a skip to put rubbish in it. So, [I put] "skip hire" or something like that and ... then you can put "Sheffield" and it lists the local businesses and you can select ... like a business finder which is quite useful or yell.com [http://www.yell.com] which is yellow pages online. The search engine that I usually use is Google."

Termination

There are a number of what Simon termed "stop rules" which in the present context enable people to decide when to terminate a search. The below quote shows an example of the participant's view about this issue:

AS10: "If you think about it too much you waste time ... that is bad if you get frustrated because you can't find something. Sometimes it becomes like a challenge, doesn't it? You have to try to look for it and you think the system is defeating you, you know, and you get angry at it. So, that is probably when I say that is enough. Stop it. I am not going to look at it anymore."

Termination has five aspects including acceptance, discomfort, boredom, time limits and snowballing. They are explained and illustrated below.

Acceptance

Searches may stop as soon as they find what the searcher considers to be an "acceptable" search result. The following quotes are examples of this category:

AS14: "I think it is quite likely [to miss information] and we have to accept this. If you look at specific thing you can find it but there may be more and more."

AS15: "I think it is just part of the life. It is part of the research searching for things and it doesn't always work. I mean it's disappointing – yeah – if I've spent half an hour or more trying to find something. Yes, you do get disappointed and think 'oh yeah – should I have been doing something else or am I wasting my time?' You know you've got to accept it. You've got to do these things sometimes. I get lucky if it works and I have more successes than failure. So, I generally accept that. Occasionally it just doesn't work ... yeah, I accept. Life isn't always going to work".

AS10: "Often you just keep looking until you've found as much as you want or as much as you need. But I know that often there're more things out there. But I stop when I find what I want."

Discomfort

Discomfort may also cause a searcher to stop searching. For example, the following quote shows that users may feel uncomfortable as a result of long time working with computers for different purposes including searching the web.

AS08: "There are only a certain number of articles you can look at when you go through a PubMed search before your brain just glazes over, your eyes glaze over and you just skip and miss things."

Consequently, some of the interviewees mentioned that they were happy to be away from searching the web from time to time to avoid the tiredness of using it everyday. For example:

RS17: "So, I do it [searching the web] at least once everyday except Sundays. That is when I garden. I turn it off and have a day without it, unless it is an emergency".

Boredom

Searching the web requires a certain degree of motivation and it can not be continued when the searcher feels bored:

RS17: "To him it [web searching] is a hobby, whereas to me it is a part of the job. So, the way that he approaches that is going to be different. You know computing is his big hobby ... anything to do with this machine ... to him is like a pleasure. So, he learns the search as part of his learning process. But I just use that to get my stuff. To me it is a chore, a task, a job that has to be done. So, it is not so pleasurable."

Another participant found web searching a boring task and he mentioned that he personally likes the old fashioned way of looking for information through printed materials, and it makes him less motivated to carry on web searching unless he has to do it. Besides, he believed that the traditional library search is a more sophisticated method of information seeking:

RS18: "It is rather boring. Typing something in, and then getting something in return for information that you want. Also it is quite an overload of information, depending on what you are searching. This is sometimes a bit constrained, and I am also a bit an old fashioned guy. I really like to go to libraries and dig into books and stuff. Maybe I have the feeling this is a bit more special. Because everybody can go to the web and just type things in and look for it. But not everybody has skills to extract like hidden interesting references from the library."

Similarly, another interviewee said:

RS26: "Sometimes you might be unsuccessful not necessarily because it is not there but because you don't have the patience or time to actually keep looking.

Time limits

As has been mentioned before, time is a leading element that makes a big difference to users' search behaviour in terms of when to stop the search. The following quotes are examples of the importance of time in users' decision making:

RS17: "I think I was doing it for about twenty minutes. I just think that was enough time to not get anywhere, because there're sort of things to do. So, probably I gave up on a time limit."

AS08: "Yes, I do fail. I do fail to find things on the web, whether that's because they aren't there or whether because of I don't have time to find things. Well, actually probably both of these things occurred. You know, there are sometimes when you are looking for a particular piece of information and the time pressure like review a lecture, for example, or writing something that needs to put in by the deadline."

One interviewee pointed out the issue of cost-effectiveness and resource allocation in relation to the comprehensiveness of the search result:

AS06: "Ultimately, I suppose it is a case of weighing up how much it matters against the fact if I really want to be hundred percent sure. It would take me ten times longer to be sure."

Snowballing

According to Agosto (2002), "snowballing" refers to a situation in which an individual begins to experience repetition or saturation in the information sources being found. The following two quotes show examples of snowballing in the dataset:

AS03: "Probably I always miss something. I think because we don't have, we never have to get everything. As long as I am satisfied with the products that I have found, that is it. I don't want anymore about it which might be same or similar or repetitive."

RS25: "I think when it comes to searching for literature, which I suppose it is something that I do quite a lot ... you feel you get to a plateau when you are not finding many more references as you read a paper and make a note about what they have cited. If you can start with something recent and you can go backwards ... once you start finding the network of citations and you are not finding many more and you think okay well more or less I covered that area."

Model of information seeking on the web based on Bounded Rationality & Satisficing theory

As has been discussed earlier, this chapter complements the results of the inductive analysis presented in chapter eight by adding findings based on a deductive analysis of the dataset. In fact, this chapter aims to complete the picture that has been illustrated in chapter eight about users' awareness of and reaction to missed information. Chapter eight enabled us to addresses PRQ2 and PRQ3 [42], and presented two matrices including the "matrix of search impact", presented in figure 8.2, and the "matrix of search depth" presented in figure 8.3.

The matrix of search impact consists of four zones including the "Inconsequential Zone", "Tolerable Zone", "Damaging Zone" and "Disastrous Zone". In contrast, the matrix of search depth is made of four other zones including "Perfunctory Searching", "Minimalist Searching",

[42] The second and third primary research questions, to remind the reader, these questions are:
PRQ2. To what extent are web users aware of missing some information resources on the web?
PRQ3. If they are aware of missing relevant information, how important is it for them?

"Nervous Searching" and "Extensive Searching". Therefore, through a closer look at the above matrices it can be hypothesized that there is some logical correspondences between:

- Perceived "Disastrous" impact and "Extensive" searching, since dire consequences are likely to result from missing any important source;
- Perceived "Damaging" impact and "Nervous" searching since, if an extensive search is not to be attempted, there may be uncertainty about when to stop searching coupled with worry about potentially damaging results of so doing;
- Perceived "Tolerable" impact and "Opportunistic" searching, since the retrieval of "low hanging fruit" is likely to be acceptable; and
- Perceived "Inconsequential" impact and "Minimal" searching since there is low perceived likelihood of any adverse consequences resulting from relatively shallow searching.

Therefore, as a result of the integration of the above matrices in the light of the theory of Bounded Rationality and Satisficing, a new model has emerged through this study which is called "a model of information searching behaviour in the light of Bounded Rationality and Satisficing theory. This is presented in figure 11.1.

This model not only incorporates the previous two matrices but also brings more contextual elements into the picture, providing us with a more comprehensive view of what is actually happening during the web search process regarding the issue of missed information and users' reactions to it. This model shows that search-specific bounding factors are likely to interact with generic ones, with certain ones over-riding others.

For example, physical and/or psychological discomfort or boredom may be over-ridden where the impact of missing information is perceived as potentially disastrous. In such circumstances, strenuous attempts may also be made to overcome time limitations. This would be much less so where the possible impact of missing information is less deleterious. Different search strategies may render different satisficing approaches more or less relevant. For example, in the context of an extensive search, snowballing (or saturation) becomes particularly appropriate since it may be an indicator of the completeness of a search.

Conversely in this context, factors such as boredom and physical discomfort should be less relevant. In the context of an opportunistic search, snowballing is likely to be less appropriate and physical and/or psychological factors may be perceived as more influential.

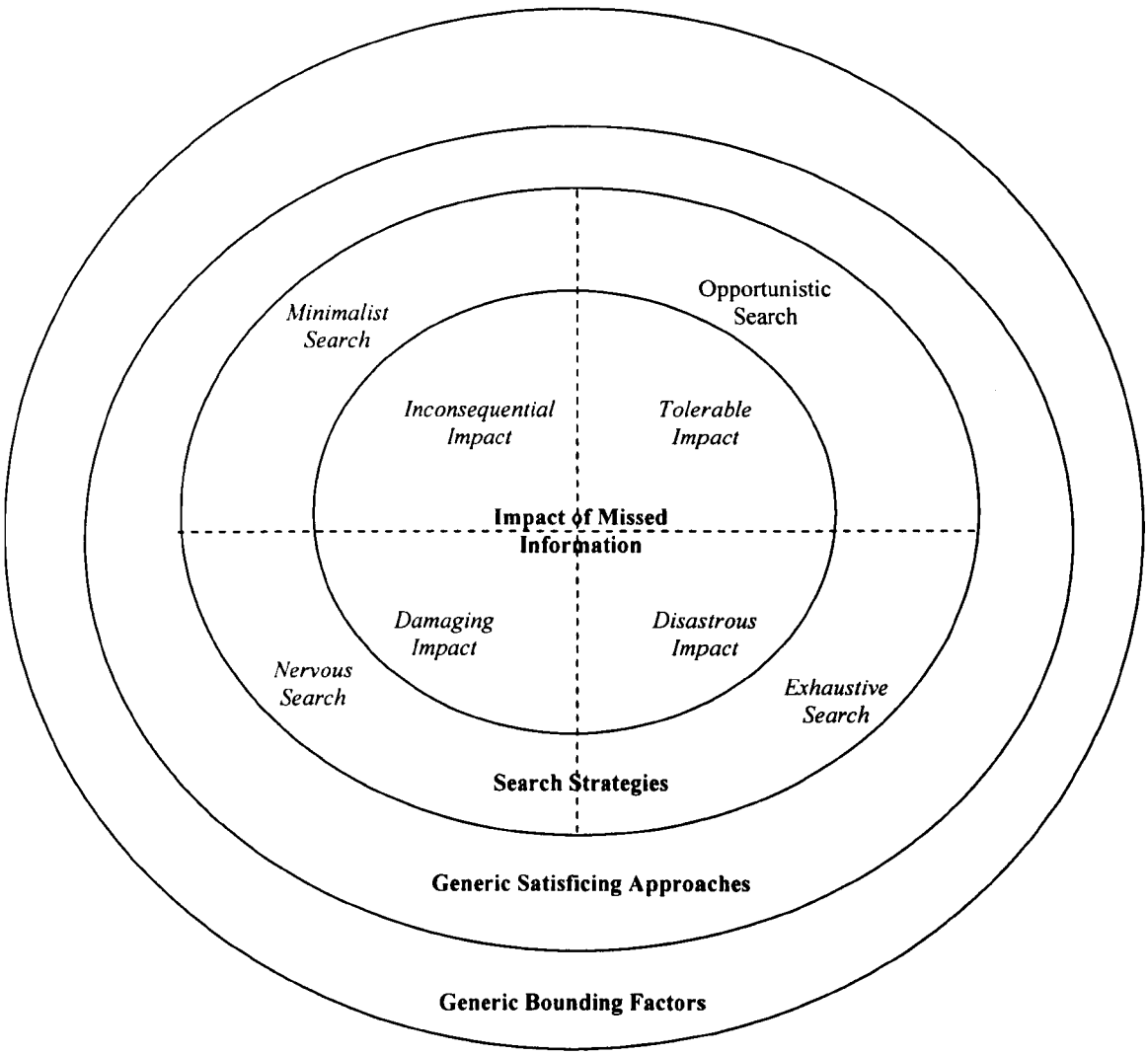

Figure 11.1: Model of information searching behaviour in the light of Bounded Rationality and Satisficing theory

Summary

This chapter presents the results of the second deductive phase of analysis of the dataset taking into consideration the theory of Bounded Rationality and Satisficing. The chapter aims to complete the picture presented in chapter eight which was developed based on the inductive phase of data analysis, and related to users' awareness of and reaction to missed information while searching the web.

The findings show that the data collected in this study can be successfully interpreted in the light of the theory of Bounded Rationality and Satisficing and, that there is sufficient evidence in the data to support key components of this theory including time-constraints, information overload, physical constraints, reduction and termination. For example, the data analysis showed that shortage of time is one of the reasons for stopping searching. Moreover, the enormous volume of information on the web can cause difficulties for searchers, and using computers can be tiring, particularly for those for whom such use is not a leisure activity. Resultant discomfort may be physical and/or mental, and may become a factor in decisions when to stop searching.

The level of required effort may also be an influential factor for searchers to decide when to stop. Although usually the web provides end users with quick and easy access to information, sometimes web searching might require a considerable amount of effort. The chapter also showed that the interviewees might bring their searches to a halt before they felt they had found all relevant information. In particular, the concept of "compromised" as opposed to "ideal" searching was identified in the dataset. Results presented in this chapter illustrate that the theory of Bounded Rationality and Satisficing can be considered as offering potentially useful concepts to understand different aspects of information seeking behaviour on the web. To support this claim the chapter presents a new model of information seeking behaviour on the web which has been developed as a result of the integration of two matrices including the "matrix of search impact", presented in figure 8.2, and the "matrix of search depth" presented in figure 8.3. This model enables us to address users' awareness of and reaction to missed information while searching the web.

CHAPTER TWELVE:

CONCLUSION & FURTHER RESEARCH

CHAPTER 12: CONCLUSION & FURTHER RESEARCH

"The most successful men in the end are those whose success is the result of steady accretion. It is the man who carefully advances step by step, with his mind becoming wider and wider - and progressively better able to grasp any theme or situation - persevering in what he knows to be practical, and concentrating his thought upon it, who is bound to succeed in the greatest degree." Alexander Graham Bell

Introduction

This chapter presents the overall picture of the main findings of the current study and encapsulates what has been achieved through the course of this research in terms of addressing the research questions and contributing to the knowledge of the related areas including information seeking on the web, web search research, information literacy and human computer interaction.

Furthermore, the chapter presents a six-layer integrative model of information seeking behaviour on the web as the overall result of this study. The model integrates the previous models which emerged through different stages of this study and are presented in chapter six, seven, eight and eleven of this thesis. The overall model aims to incorporate the findings of the study into a coherent body and address the research questions. Therefore, to remind the reader the three sets of research questions and a brief discussion about the results relating to each one are presented in this chapter.

Revisiting the research questions

As has been declared in chapter one, this study was an exploratory study and passed through different stages, in each stage of which the researcher addressed a specific set of research questions. At the outset, the 'invisible web' formed the main focus of the study and there were two main research questions as follows:

Early research questions

In the early stages of the study, from February 2003 to June 2004, there were two Early Research Questions (ERQs) that originated the research. These questions were about (1) end-users' perceptions of the invisible web and (2) the compatibility of their perceptions with the existing literature-based definition of this term. The early questions which formed the main focal concern of the early stages of the research were:

Perceptions and definition of the invisible web

- ERQ1: How do a particular group of end users conceive the 'invisible web'?
- ERQ2: To what extent do such conceptions and experiences map onto the technical models of the invisible web proposed in the literature?

The result presented in chapter six shows that although the term 'invisible web' was new to all but one of the participants, they have experienced different parts of the hidden territory of the web in their real searches. For example, when they were reporting details of their recent failed searches it was declared by them that their required information must be somewhere 'out there' on the web but they have failed to retrieve it and it remained hidden for them. However, the perceptions of the participants about the invisible web did not match the existing definition of the invisible web. This incompatibility led to developing the model of information visibility which has been explained in chapter six, in the model of information visibility on the web presented in figure 6.1. This model which is in a matrix form, consists of four zones including 'bright', 'opaque' for successful searches and 'veiled' and 'dark' zones for failed searches.

In the bright zone the web user searches and finds relevant information using a search engine without facing any impediment or problems and the search is fully illuminated for the user with no hindrance. However, in another part of the model which is called the 'opaque' zone the user eventually finds relevant information possibly by means other than a search engine (e.g. via URLs given by friends or colleagues or through bookmarked web-based resources in their favourites lists). Although in the opaque zone the user does not fail to fulfil the search, it represents a gradation in the move from full visibility in the bright zone to total invisibility in the dark zone.

As has been mentioned earlier, both 'veiled' and 'dark' zones represent search failure in the model presented in figure 6.1 in which users fail to succeed in the search due to different kinds of information invisibility, including perceived as hidden, unknown or unavailable. Therefore, based on the type of the perceived invisibility there is a difference between these two zones. In fact, what makes a big difference between 'veiled' and 'dark' zones is the level of users' certainty of the existence of their required information. When the user is confident the required information exists on the web and s/he fails to retrieve it this kind of search is located in the veiled zone. Nonetheless, if the user does not know whether the information is indexed or not or is confident that the information is not indexed then these kinds of searches are positioned in the dark zone.

This model enhances the existing definition of the invisible web by adding the user aspect to it and evolves the term 'invisible web' into a more comprehensive concept termed in this study 'information visibility'. In fact, this model does make a clear distinction between the invisible web which is static, technical and objective, and the information invisibility which is dynamic personal and subjective.

The definition of the invisible web is a fixed term which divides the whole web environment into two parts - the visible and the invisible. However, in reality we can not make a boundary between two parts of the web. For each individual web user in each single search session the portion of the visible part and the invisible fraction of the web are different. The existing definition of the invisible web appeared in the literature (e.g. Sherman and Price, 2001) offers a binary picture about this issue but this study considers it in an analogue image and on a continuous scale which varies in different contexts and it is not possible to consider concrete boundaries for it.

Moreover, there are no clear links between different parts of the invisible web suggested in the literature, summarized in figure 2.4, but the model of information visibility on the web suggested in this study presents an integrated picture which incorporates different kinds of the invisible web into an integrated and coherent image in figure 6.1.

In addition, there are some kinds of the information visibility on the web that the literature-based definition of the invisible web is not able to address. For example, there are many information resources that are used by millions of web users everyday but are totally invisible for general-purpose search engines. The existing definition of the invisible web cannot characterize these kinds of web-based resources because based on the definition these resources are invisible but in reality they are visible for end users and are used by them. However, the model of information visibility suggested in this study easily addresses this issue. Based on this model these kinds of resources are part of the 'opaque zone' in figure 6.1.

Similarly, there are many resources that can be found by search engines but for many reasons end users might overlook them. Based on the existing definition of the invisible web these resources are part of the visible web but in reality they remain invisible for end users. According to the model of information visibility in figure 6.1 these kinds of resources are part of the 'veiled zone'.

The findings of the study indicate that the existing definition of the invisible web is limited and does not take into account the user and we need to introduce a new concept to address the issue effectively. Therefore, this study suggests that the concept of 'information visibility' which is presented in the model of information visibility on the web, figure 6.1, as a complement to the literature-based definition of the invisible web by adding user dimension to the previous definition. The model of information visibility on the web demonstrates that developing any definition about the issue of information visibility/invisibility without considering the user's aspect is not sufficient to convey the meaning of the invisible web. In fact, the current study highlights the fact that the existing definition of the invisible web, that arguably is still useful even though partial, overlooks two important aspect of the issue including the user aspect and also the dynamism of the web environment.

Regarding the literature-based definition of the invisible web those resources which can not be retrieved by search engines form the hidden part of the web. This objective and static view provides only a limited perspective. It fails to address a key question: for whom are information resources invisible? This study showed there are resources that are invisible for search engines but are visible for some users, and there are resources that are visible for search engines but remain invisible for some users.

Furthermore, the capability of the search engines is improving everyday and what is invisible today may not be so tomorrow. Therefore, the improvement in the capability of search engines is also constantly affecting objective visibility. For example, file formats that were not crawled some years ago (e.g. PDF and PPT files) are now visible to many general search engines including Google and AltaVista.

Also, each user's search skills are likely to change with experience or training over time. For this reason, it is important to distinguish, and to be clear about the relationship between, the 'objective technical' invisible web in the literature and the 'cognitive subjective' invisible web in this thesis. Accordingly, this study demonstrated that although the phenomenon of the invisible web does exist, existing definitions only provide a somehow limited perspective which fails adequately to take into account the user which is addressed by the concept of information visibility developed in this study.

Primary research questions

The first primary research question (PRQ) was:

Conceptualizations of failure

- PRQ1: How do web users conceptualize their information seeking failure on the web?

The current study managed to address this question both inductively and deductively. The inductive analysis of the collected data showed that end users experience failure in web searching in different occasions. All of the participants in this study were able to recall at least one recent failed search. Some of them remembered more than one occasion, although the majority declared that failure does happen less frequently than success. Based on the inductive phase of data analysis, this study suggests an extended version of the model of information visibility which integrates search success and failure with the concept of information visibility. The model presented in figure 7.1 illustrates how the first model, presented in figure 6.1, is related to the model of success and failure in search.

There are six cells in figure 7.1 which are divided into two main upper and lower areas. Three cells on the upper area include search scenarios whereby users eventually discover what they are looking for. Therefore, the upper section of the matrix is the success area. Furthermore, based on the matrix there are three different kinds of success. The first kind is termed as 'anticipated success' in which the user is aware that their required information exists on the web, they expected to be successful and they are. Therefore, a successful result is anticipated by the user. In the second type which is termed in this thesis as 'serendipitous success', the user is simply not sure whether the required information is available or not. Nevertheless, s/he finds what is required. The third type is 'unexpected success' when despite the user's certainty that the required information is not indexed by search tools s/he finds it.

This study also identified three types of search failure. In failed searches the user does fails to conduct the search successfully. Therefore the lower section of the matrix is the failure area. The failure area is divided into three zones compromising 'unexpected failure', 'unexplained failure' and 'inevitable failure'.

In 'unexpected failure' the user has expected to be successful because of his/her certainty about the existence and availability of his/her required information. However, in spite of the users' expectation of success s/he fails to succeed the search. In 'unexplained failure' the user

fails to succeed the search without knowing whether the search could be successful or not. Nevertheless, at the end the user can not manage to accomplish the search successfully.

Finally, 'inevitable failure' is the exact opposite of 'anticipate success' in which the failure has already been expected by the user. In fact, 'inevitable failure' represents those failed searches in which the user is aware that his/her required information is not available on the web and this failure seems unavoidable to him/her. The data showed there are possible motivations behind those searches that are predicted to be failed. For example, the searcher carries out a search to confirm his/her certainty of the unavailability of the required information.

Awareness of and reaction to missed information

Awareness of and reaction to missed information while web searching formed two primary research questions as follows:

- PRQ2. To what extent are web users aware of missing some information resources on the web?
- PRQ3. If they are aware of missing relevant information, how important is it for them?

The results showed that the participants in this study were aware of the possibility of missing some relevant information in their web searches. However, regarding the context of each search they have different perceptions of the importance and the volume of missed information, and accordingly they react to it differently.

In general, data analysis regarding this issue led to identification of four categories, namely (a) perceived volume of missed information, (b) perceived importance of missed information (c) awareness of information likely to be missed and (d) reaction to missed information.

According to these four categories two matrices were developed which are presented in chapter eight. The first matrix, figure 8.2, is called the 'matrix of search impact' and the second one, figure 8.3, is termed the 'matrix of search depth'.

The matrix of search impact consists of four zones including 'Inconsequential Zone', 'Functional Zone', 'Damaging Zone' and 'Disastrous Zone'. In the 'Inconsequential Zone', only a small amount of less relevant information is missed which does not have a significant effect on the search performance. In the 'Functional' zone missing some relevant information does not bring about search failure but the retrieved result is not as satisfying as it could be if the user had not missed those resources. However, in this zone the amount of the relevant found items is usually sufficient to achieve the search's goal and therefore there is a 'tolerable impact' on the search results. Nevertheless, in the 'Damaging' zone users are concerned about the importance of missed information rather than its quantity. If the number of relevant documents is high then failure to retrieve a few highly relevant items might not influence the search outcome hugely, but if there are only a few highly relevant resources to accomplish the search successfully and the user misses the few possibilities it might bring about search failure. Finally, in the 'Disastrous' zone the user might fail to retrieve a large portion of highly relevant

information. Obviously, this is the worst situation of missing relevant information which probably cause search failure.

The matrix of search depth illustrates different types of search strategies that users take into consideration based on their awareness of the possibility of missing some information and their determination to retrieve missed information. This matrix is formed of four zones which are 'Perfunctory Searching', 'Minimalist Searching', 'Nervous Searching' and 'Extensive Searching'. In perfunctory zone the user is confident that s/he is not missing information and do not make effort to find any missed information. In this zone users trust the results of their straightforward searches without being worried enough to conduct any more searches.

The 'minimalist' zone shows those search strategies in which although users are aware that they probably miss some information, they do not attempt to fill the gap. However, in 'nervous searching' web users might noticeably feel nervous that they may be missing important information and consequently, they are unsure of when they will be able to safely stop searching without fear of missing seminal items.

Finally, in the 'extensive' zone of the matrix the user is very persistent in conducting the search as extensively as possible because missing any important item can be disastrous in their search context. In fact, in this zone users are fully aware of the possibility of missing important information if they do not conduct extensive searches. This type of search strategy is mostly taken into consideration in the case of, for example, literatures review searches or for writing research grant proposals when the user can not afford to miss any important information item.

Coping strategies

The issue of coping with information seeking failure on the web formed three primary research questions as follows:

- PRQ4. How do web users seek to cope with information seeking failure on the web?
- PRQ5. What strategies do web users use to cope with information seeking failure on the web?
- PRQ6. What factors might influence different conceptualizations of information seeking failure, and different coping mechanisms?

A series of coping strategies have been identified through the study. These strategies are adopted by users to overcome search failure. The identified coping strategies have been divided into two groups of passive and active strategies. Passive strategies entail strategies that involve the least action toward modifying the situation and mainly relate to accepting the situation as it is rather than as it can be. Giving up after facing a failure is an example of a passive coping strategy. Similarly, when users come upon unsatisfactory results they may perhaps adopt a passive reaction and instead of making any additional effort to conquer the failure they may perhaps change their search goal which in this study is called 'goal modification'.

However, employing active strategies requires further activity to overcome the failure. There are two levels of active coping strategies including 'revising' and 'help-seeking' strategies. At the first level, the user may try to find a solution to defeat his/her search failure on his/her own. The data analysis led to identifying a number of these revising strategies including verifying the initial search query or primary keywords, using another search facility, looking for new keywords in other resources, chaining authors' names, browsing as an alternative way to searching, and searching with more details. In 'help seeking' the user looks for help from other people, including colleagues, peers, supervisors, friends and family.

Moreover, data analysis suggested that the importance of the search and searchers' determination to use coping strategies to overcome failure are two important elements for users to decide how persistent they should be to overcome the search failure. Accordingly, another matrix was developed in this study, presented in figure 9.1, which shows different kinds of searches based on searchers' determination to employ coping strategies and the importance of the search for them. Therefore, the matrix consists of four zones including 'trivial searches', 'alternative searches', 'fascinating searches' and 'crucial searches'.

A trivial search is not perceived as an important issue for the user and s/he also is not determined to employ different coping strategies, or if s/he adopts any strategy s/he is not so persistent to finish the search successfully. In alternative searches though the search is perceived important by the user, s/he is not determined to carry on the search by employing coping strategies. In this kind of search web searching is not the only way to look for the required information and there are other substitute routes to achieve the goal.

Fascinating searches and crucial searches are common in the high level of users' persistence to overcome the failure. Nevertheless, there are different motivations behind the persistence in these two kinds of search. In crucial searches the user peruses the search persistently because of the importance of the search topic for him/her. However, in fascinating searches although the search is not essentially important, but the search topic for whatever reason fascinates the user and consequently s/he is very determined to find out the wanted information. In terms of how users develop different coping strategies the data analysis led to identification of four methods including attending training courses, using trial and error, knowledge sharing, and adoption of pre-web coping strategies on the web.

Data analysis showed there are possible links between adoption of coping strategies and how users conceptualize their search failure. When failure is ascribed to their own inability they might change their search strategy. Therefore, when they think their own ability is preventing them from conducting the search successfully they start seeking help. However, if they become convinced that their required information dose not exist on the web they look for it through other media including printed materials in the library. In fact, when failure is ascribed to any element which is out of the users' control they might act differently. For example, when it is believed the search failed because there is no information to retrieve they will stop searching because they are convinced that there is nothing to find, and logically there is no need to continue the search.

Emergent research questions

As has been discussed in chapter one after completion of the inductive analysis of the dataset a number of new research questions emerged which were addressed through a deductive examination of the same dataset. There were three emergent research questions which were addressed in the study as follows:

Causation of success/failure

- EmRQ1: What are the perceived causes of search success and failure of a particular group of academic users?

Apart form a number of the participants who were somehow uncertain about being successful or failed in some cases, the rest of searchers who participated in the study were able to clearly distinguish between 'successful' and 'unsuccessful' searches. Analysing the dataset led to identification of a number of elements that participants perceived to influence the fact that searches end in failure whilst others result in success. The findings were categorised into two groups of perceived factors related to success, and perceived factors related to failure in chapter ten.

For successful searches domain knowledge, availability of information, search experience, linguistic and conceptual ability, confidence, persistence, search tools efficiency, website structure, good fortune, or a combinations of some of these factors are the reasons that the participants mentioned.

For unsuccessful searches, lack of domain knowledge, unavailability of information, lack of search skills, lack of linguistic and conceptual ability, user or system errors, lack of sufficient effort, search tools deficiency, deficiency of website design, bad luck, combination of factors, or uncertainty about the causes are the identified attributions.

In the next stage of the analysis the identified factors were divided into two groups of internal and external factors. The internal factors are those elements that are in the control of the users (e.g. linguistic abilities) and they are able to change them. For example, the user can improve his/her linguistic abilities by training or gaining more experience in searching. However, external elements are beyond the users' control and they are not able to make any changes in them. For instance, efficiency/deficiency of search tools is not in the control of the user and s/he has to accept the search facilities that are provided by search tool designers.

Exploring the dataset in regard to internal and external factors showed that a certain percentage of the sample attributed their successful search to internal, and their unsuccessful search to external factors. A percentage displayed converse attributions; and a further group reported a mixture of internal and external attributions for some searches. The results of this phase of analysis are reported in chapter ten.

Locus of Control and Attribution Theory

- EmRQ2: To what extent are Locus of Control and Attribution Theory useful in interpreting users' conceptualizations of success and failure in web searching?

As has been explained in chapter ten the findings of the current study demonstrated that Locus of Control and Attribution Theory are possible theoretical frameworks to analyse the interaction of end users with web-based search facilities. Attribution Theory provided the current study with a well-established conceptual framework to reanalyse the inductive results of the study in a deductive way. In fact, this study is arguably useful in that it presents some empirical evidence that Locus of Control and Attribution Theory are possible theoretical frameworks through which to interpret web-based information seeking behaviour. It supports the notion that they represent plausible frameworks insofar as they are coherent, account for much of the data, between them do not ignore negative or counter data, and are supported by a good evidence base within the domain of psychology.

Participants' self-efficacy was categorised into five groups of low self-efficacy, negative self-efficacy, mixed, positive self-efficacy and high self-efficacy, and a model of self-efficacy in web searching was introduced in chapter ten (page 202). The results suggested that the more extreme low self-efficacy individual– if tending to attribute both success and failure to external factors to a greater extent than is warranted by a more objective assessment of the reality of the situation – may underestimate the potential benefits, in terms of increasing success and reducing failure, of engaging in training/self-learning, s/he may perceive little to be gained by attempting to enhance internal factors such as knowledge and skills if both success and failure are due to external factors beyond his/her control. Conversely, the extreme high self-efficacy individual – again, assuming that his or her tendency distorts a more objective assessment of search attributions – may be excessively self-critical over search failures, and possibly expect too much in terms of the capacity of training to remedy them where in reality they are the result of external factors. On the other hand, it might be argued that people who have high confidence may not perceive the need to improve their skills, thus they are hard to train.

Bounded Rationality and Satisficing

- EmRQ3: To what extent is the theory of bounded rationality and satisficing useful in interpreting users' awareness of and reaction to missed information while web searching?

Simon's Bounded Rationality and Satisfying were used in a similar way to that in which Attribution Theory was used in this study to re-analyse the dataset deductively.

Simon's theory was utilized to interpret the users' behaviour in reaction to problematic search situations in general, and their attitudes to missed information in particular. In order to conduct the deductive analysis in the light of Simon's theory, the study adopted Agosto's

(2002) coding framework[43], apart from her additional 'personal preference', to evaluate the extent to which elements of Bounded Rationality theory can be useful to scrutinize the dataset.

Data analysis revealed that this theory is a very useful conceptual framework to examine the dataset. Different elements of the theory were well-matched with the dataset in this study. For example, it was found that two types of time constraints including imposed and self-generated were mentioned by the participants as reasons for bringing their searches to a halt before they feel they have found all relevant information. Similarly, other elements of Agosto's (2002) framework including 'information overload', 'outcome overload', and 'physical constraints' were visible enough in the dataset. The analysis showed the sheer volume of information on the web, the volume of retrieved items, and discomfort caused by excessive work with computers are some examples that cause difficulties for searchers and prevent them from scanning all possible options while searching the web.

Accordingly, they stop searching when the result is just 'good enough' and they do not spend more time and energy to achieve 'ideal' results. In fact, they usually bring their searches to an end whilst being aware of missing some information. Based on Simons' theory and Agosto (2002), these strategies are known as satisficing approaches. Data analysis showed that different satisficing approaches including 'reduction' and 'termination' are distinguishable in the dataset. 'Reduction' is a satisficing strategy entailing reducing the search task to make it more manageable. For example, restricting the search domain to 'known sites', making use of synopses or using categorized sites are examples of 'reduction' strategies.

'Termination' is another part of Simon's theory which entails 'stop rules' enabling people to decide when to terminate a search. Termination has five aspects including acceptance, discomfort, boredom, time limits and snowballing. Searchers might stop searching once they find what they consider to be an 'acceptable' result.

Moreover, discomfort may force a searcher to stop searching before gaining ideal results. In fact, web searchers, like other computer users may feel exhausted because of spending a long time working with computers and may terminate their searches as a result of discomfort. Furthermore, searching the web requires a certain degree of enthusiasm and if the searcher feels bored, even if s/he is aware of missing information, s/he may stop searching.

Integrative model

An overall model of information seeking on the web is suggested in figure 12.1. This model integrates different specific models presented in previous chapters.

[43] Agosto's (2002) framework is based on Simon's Bounded Rationality and Satisfying theory.

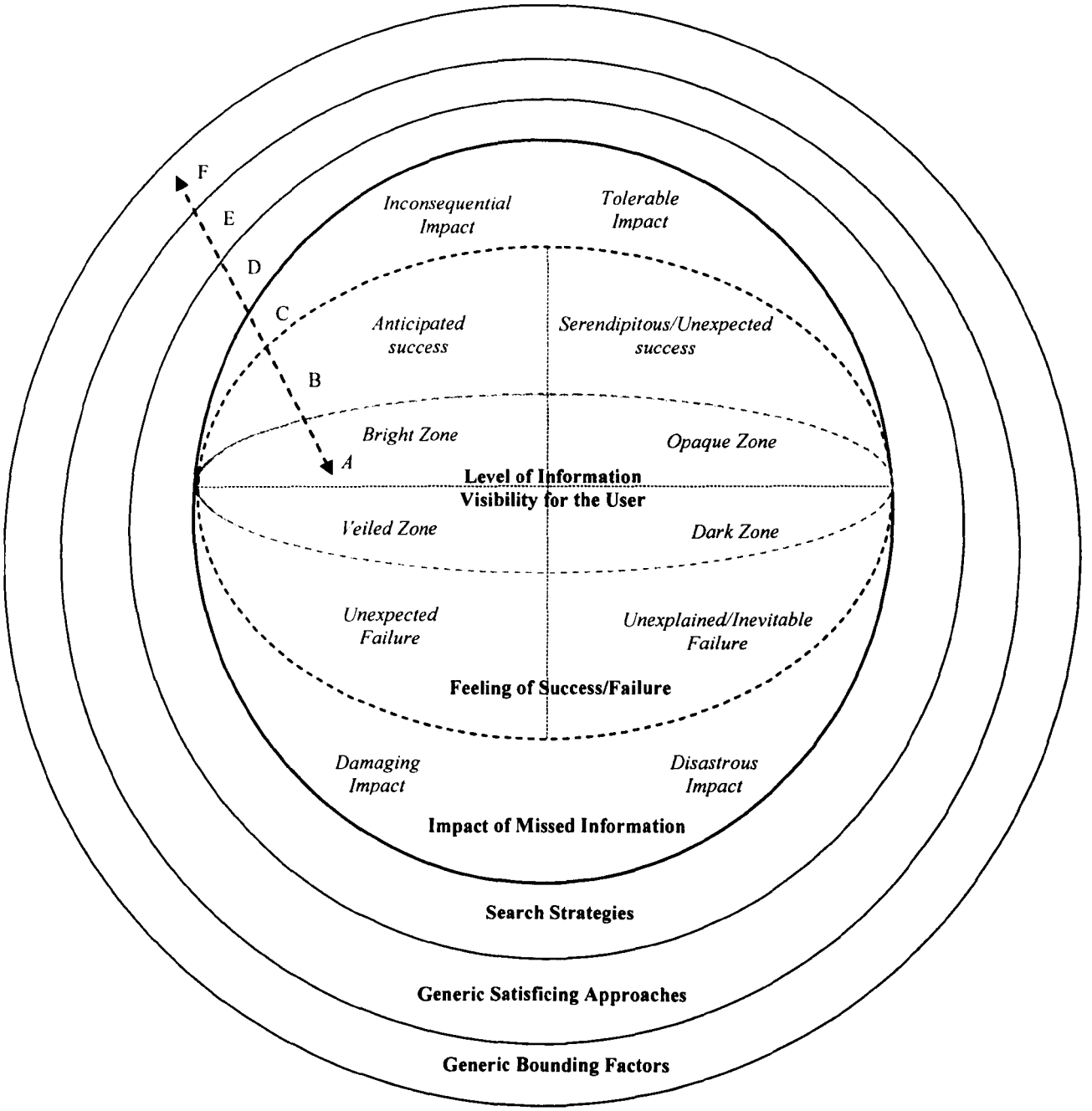

Figure 12.1: The six-layer integrative model of information seeking on the web

There are six interrelated levels in figure 12.1 which emerged one after another in different stages of the research. The first three inside layers, A; B and C, emerged through an inductive analysis of the dataset based on Grounded Theory, which was the employed research method in the study and there are detailed explanation about it in chapter three. The next three outer surface layers, D; E and F, were developed based on the deductive analysis in the light of three conceptual frameworks including Locus of Control, Attribution Theory and Bounded Rationality and Satisficing theory. There are more explanations about the model as follows:

A. The central part of the integrated model is occupied by the first model which emerged through the inductive stage of the data analysis. In fact, layer A is a simpler version of the model of information visibility presented in chapter six. This part of the integrated model explains what happens when users interact with web-based search tools. This level is directly deals with human and computer interaction (HCI). This domain is very dynamic and relates to the information visibility on the web.

B. The next layer relates to the domain of users' perceptions of success and failure. Users' feelings play the key role in this part of the model which is more stable than the HCI layer and less stable than the next layers. By stability it is meant that feelings of failure or success remain with users for a period not restricted to a particular, which may covers a series of search sessions.

C. The third layer shows search-specific bounding factors which have different influences on search, varying from inconsequential to disastrous impact.

D. The fourth layer illustrates different search strategies that users might take into consideration regarding different search contexts.

E. The fifth layer shows generic satisficing approaches based on Simons' theory of Bounded Rationality.

F. Finally, the outermost layer shows generic bounding factors including time constraints, information overload and physical constraints.

Contribution of the study and further research

The results of this study contribute to the related areas at two levels of theory and practice. The first level of contribution is to the theoretical base in the area of online information seeking behaviour in general, and information seeking on the web in particular. The models developed in the study arguably provide these areas with fresh understanding of how end users interact with web-based resources. In fact, this research presents a conceptual framework that is intended to be a useful reference point for researchers wishing to investigate user-based aspects of web searching.

The second level is contribution to practical aspects of different related areas including information literacy. The findings of the current study may be useful to trainers and those interested in developing information literacy in that it (a) differentiates 'technical objective'

and 'cognitive subjective' conceptions of 'invisibility', and (b) by implications for helping searchers to develop more effective searching capabilities.

Additionally, this study suggests a 'visibility learning diary' which can be developed based on the model of information visibility. This is a new kind of learning diary to monitor the progress of web users' information seeking skills over time. This diary can be used for different purposes. For example, it might be helpful for information literacy trainers to measure the progress of their trainees during a specific period of time. The diary might help to find out how the trainees move within different zones of the model during time and how the training course might help users to have a more realistic viewpoint about their search results and conceptuality become more successful in web searching.

Furthermore, designers of web-based information retrieval tools may benefit from the results of the study to enhance the efficiency and usability of their products. For example, the integrative model of information seeking on the web, presented in figure 12.1, is useful for search tool designers to have a more holistic and realistic view about the process of information seeking on the web. In fact, it reminds them that their users employ their products in a real situation which is surrounded by generic bounding factors and their users' may adopt very different search strategies based on the contextual elements of the search situations. Therefore, the designers of web-based search tools can consider users' satisficing approaches to provide them with more search facilities.

The lines of enquiry that the current study opened does not stop with the completion of this research. One of its notable results is producing more research questions to address in future research. The models which have been developed over different stages of this study can work as conceptual frameworks for further investigations. For example, the model of information visibility on the web can be employed for different purposes. One possibility is testing the model in other contexts. As this model is based on the web environment it can be useful to test its applicability in broader context. Although the model addresses the issue of information visibility/invisibility on the web, it would be useful to find out how it can be generalized to the world beyond the web. Then it will be possible to introduce a new concept of 'information visibility' which can be compared to other well-established concepts such as 'information accessibility' and 'information availability' to show how these concepts are different. In fact, information visibility implies the status of each information item in terms of the possibility of utilise by the person who is looking for it.

This is a new approach to human information interaction. In this approach an information item will be used by the user only when it is visible enough for him/her. In other words, availability and accessibility of information are essential requirements for using a piece of information but are not sufficient because that piece of information can be used only when it is available, accessible and visible. For example, information resources in the opaque web are available on the web but not accessible for general purpose search engines users and obviously not visible for them. However, information in the 'veiled zone' is available on the web and is accessible by search tools but still not visible for users. Accordingly, just those information items are usable which are at the same time available, accessible, and visible. Further research

is required to find out how the model of information visibility can be employed to map the level of information visibility for different groups of web users or within different subjects.

Whilst it is argued that models presented in this study validly represents the range of search experiences reported by the academic staff interviewed, the research did not seek to investigate the frequency with which any of the types of search experience occurred, or are likely to occur in any wider population.

In addition, it would be interesting to test the robustness of the constructs identified within the small scale qualitative study reported here by using larger-scale statistical survey techniques. If found to be robust, it would then be interesting to discover:

- What is the incidence within particular populations of different generic and search-specific bounding factors and satisficing strategies in relation to web searching;
- What is the incidence of inappropriate use of search satisficing approaches, and what might be the effects of this. For example, effectiveness would be compromised were a young person to consider a 'perfunctory' search as entirely appropriate (or possibly the only known) strategy in the context of schoolwork requiring a deeper analysis. Similarly compromised in terms of quality would be the undergraduate using a 'minimalist' approach to deliver some low hanging fruit when what is required is a more critical comparative analysis – for which a wider-ranging search may be more appropriate;
- Whether some people use multiple satisficing approaches and if so, how these interact, and whether certain approaches and interactions between approaches are more effective than others;
- What might be, if any, the effects on choice of satisficing strategies – and the effectiveness of this choice – of levels of prior knowledge of the search subject domain on the part of the searcher;
- What might be, if any, the effects of individual factors such as cognitive style. Are different search satisficing approaches favoured, for example, by searchers with a more impulsive as opposed to more reflective style, or a more dependent as opposed to autonomous thinking style, or a more sequential as opposed to holistic thinking style?
- What might be, if any, the effects of age and education on the selection and use of appropriate and effective approaches? Do people's knowledge and choice of strategies develop in range and/or effectiveness over time? Or do some people become entrenched in comfortable (if ineffective) approaches?
- To what extent different people and/or groups are aware of the range of possible approaches and their implications.

Perhaps the most interesting, yet at the same time most problematic question relates to the extent to which perceptions map onto 'reality'. Clearly, any attempt to assess how 'accurate' a searcher's perceptions are (of information that might be missed, and of how importance this information might be), and how 'appropriate' and 'effective' are their decisions when to stop searching, bring into relief the essentially subjective and constructivist nature of these concepts.

Nevertheless, it may be useful to attempt to obtain relatively objective assessments – for example, by pooling the ratings of a number of expert judges, against which to rate individual searchers' perceptions. Even if the differences between the two are far from accurate in any absolute sense (indeed such accuracy would arguably be impossible to attain given the subjective nature of relevance and the infeasibility of measuring recall within a web context) they may nevertheless be useful in terms of helping less experienced searchers move in a beneficial direction.

Indeed, the mapping of an individual searcher's perceptions against more objective assessments may be useful in the context of information literacy education and training. It could be instructive to know, at least to a degree, the extent and importance of information likely to be missed within the context of different satisficing approaches and related search strategies. The study reported here also raises a number of questions within the context of LIS practice. In particular, it could be beneficial to know:

- Whether, and if so how, we can help people develop more effective satisficing approaches;
- The extent to which an important component of information literacy is to know when to stop searching;
- The extent to which 'critical searching' is needed to provide the raw materials for critical thinking. In certain contexts, critical thinking nurtured on low hanging fruit may be stunted;
- Whether we can educate people to move away from what for them might be overly comfortable and possibly entrenched approaches where these are inappropriate;
- How can we help searchers, where appropriate, be aware of the ineffective use of particular approaches (e.g. the use of perfunctory and minimalist strategies where something more extensive is required);
- How can we help searchers – particularly 'nervous' and 'extensive' searchers – to know when it is safe to stop searching. It may be useful to attempt to develop quality filters such as those used in medical search facilities whereby the highest quality sources are listed first. Therefore a searcher could see where the quality had dropped to a certain level and could take the decision to stop searching with more confidence and less uncertainty. Cleary, the notion of 'quality' is arguably less problematic in areas such as medicine where 'gold standards' of quality can be agreed.

As well as search training, the theories employed in the study have potential implications for our understanding of search behaviour.

For instance, where a searcher attributes a failure to internal factors such as ability or effort, s/he is might increase effort and/or persistence, and/or take steps to learn more effective search techniques. Conversely, when an unsuccessful search is attributed to, for example, the inefficiency of the search tool, it may be more likely that the searcher will give up or move to another search tool.

The sample used in this research was made up of high achieving individuals – mainly male, and relatively high in the organisation in which they worked. According to the Locus of Control research, such individuals are likely to have a generally high internal locus of control – particularly, according to Attribution Theory, relating to success. Such biases were not supported by this study, which found an association of external attribution to a large percentage of the searches. It may be that we need to differentiate – in relation to the predictors in these theories – between levels of achievement and confidence in relation to the specific activity of searching, as opposed to more general academic seniority and achievement. Again, this is a question requiring further research.

Furthermore, since this study was focused on academia it would be valuable if another research project could investigate how different categories of the public perceive the invisible web or how they conceptualise their information seeking failure in their real information seeking experiences. In addition, other academic groups can be subject of the future studies. For instance, it could be investigated how users in different areas of study including social sciences differ with users from pure science or engineering in terms of their conceptualization of success/failure in web searching.

Moreover, it will be interesting to see how, for example, students at different levels of their study conceptualise their information seeking failure and whether there is any link between their educational progress and the way that they perceive the web. Some routes could include whether for example:

- There is a significant relation between users' information literacy abilities and the way that they conceptualize their information seeking success/failure.
- There is significant relation between users' subject of study and their conceptualization of success/failure in web searching.
- End users conceptualizations of success/failure in web searching change over time.
- Experienced users are more successful than inexperienced users to attribute their search success/failure to the factors that actually cause success/failure in their searches.
- The level of information visibility is higher for work-related than everyday life searches.
- The frequency of success in work-related searches is higher than frequency of success in everyday life searches.
- Users experience more anticipated success and less unexplained failure in work-related searches.
- Users experience more serendipitous success and more unexplained failure in everyday life searches.

As all models presented in the study are based on users' perceptions rather than what actually happens in web searching, it would be very enlightening to find out how compatible the perceptions of end user are with what really happens while they search the web.

For example, a group of users can be recruited to carry out a specific search task and then they can be asked to map their search results into the model of information visibility based on their opinion. In the next stage, search expert could examine the search procedure and map the same search into the model. The compatibility of these two versions, one based in the user' perceptions and another based on an expert's viewpoint, should be very illuminating to find out what is happening while end users search the web and if they fail in search what is the real reason of their failure.

Summary

The chapter summarizes the main findings of this study in terms of addressing three sets of research questions including early, primary and emergent. The chapter showed that the end users interviewed do experience the invisible web in their searches. However, the term 'invisible web' and the suggested definitions for it in the literature are limited in the extent to which to address the meaning of this phenomenon. Therefore, this study suggested a new concept termed 'information visibility' which is more comprehensive.

The chapter showed how web users conceptualize their information seeking failure on the web and how they attribute their success and failure to internal, external or both internal and external elements. Furthermore, the chapter showed how this study explored the dataset with some well-established conceptual frameworks including Attribution Theory, Locus of Control and Simons' theory of Bounded Rationality and Satisficing.

Moreover, the chapter presents a holistic model of information seeking on the web which integrates the models presented in the previous chapters of the thesis. Finally, the chapter discusses the contributions of this study to related areas and also the implications for practice and policy. Finally, the chapter suggests new lines of inquiry for further research.

BIBLIOGRAPHY

BIBLIOGRAPHY

Agosto, D. E. & Hughes-Hassell, S. (2005). "People, places, and questions: An investigation of the everyday life information-seeking behaviors of urban young adults". *Library and Information Science Research*, **27**(2), 141-163.

Agosto, D.E. (2002). "Bounded rationality and satisficing in young people's web-based decision making". *Journal of the American Society for Information Science and Technology*, **53**(1), 16-27.

Allan, G. (2003). "A critique of using grounded theory as a research method". *Electronic Journal of Business Research Methods*, **2**(1), 1-10.

Anderson, T.D. (2005). "Relevance as process: judgments in the context of scholarly research". *Information Research* [Online]. **10**(2) http://InformationR.net/ir/10-2/paper226.html [Accessed 15 April 2006].

Babchuk, W.A. (1996). "Glaser or Strauss?: Grounded theory and adult education" In: Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, University of Nebraska-Lincoln, Lincoln, Nebraska [Online]. http://www.canr.msu.edu/aee/research/gradpr96.htm [Accessed 15 April 2006].

Bandura, A. (1977). "Self-efficacy: Toward a unifying theory of behaviour change". *Psychological Review*, **84**, 191-215.

Bandura, A. (1986). *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1994), "Self-efficacy". In: Ramachaudran, V.S. (eds.), *Encyclopedia of Human Behaviour*, **4**, pp. 71-81. New York: Academic Press.

Bandura, A. (1995). "Exercise of personnel and collective efficacy in changing societies". In: Bandura, A. (eds.), *Self-efficacy in Changing Societies*, pp.1-45. New York: Cambridge University Press.

Bandura, A. (1997). *Self-efficacy: The Exercise of Control*. New York: W.H. Freeman and Company.

Bauer M. W., Gaskell, G., & Allum, N. C. (2000). "Quality, quantity and knowledge interests: avoiding confusions". In: Bauer, M.W. & Gaskell, G. (eds.) *Qualitative Researching with Text, Image and Sound*, pp. 3-17. London: Sage.

Beaulieu, M. (2003). "Approaches to user-based studies in information seeking and retrieval: Sheffield perspective". *Journal of Information Science*, **29**(4), 239-248.

Belkin, N.J. (1980). "Anomalous States of Knowledge as a basis for information retrieval". *Canadian Journal of Information Science,* **5,** 133-143.

Belkin, N.J., Oddy, R. & Brooks, H. (1982). "ASK for information retrieval: Part I. background and theory". *Journal of Documentation,* **38**(2), 61-71.

Bergman, M.K. (2001). "The Deep Web: Surfacing hidden value". *The Journal of Electronic Publishing* [On-line]. 7(1) http://www.press.umich.edu/jep/07-01/bergman.html [Accessed 17 November 2003].

Bilal, D. & Wang, P. (2005). "Children's conceptual structures of science categories and the design of Web directories". *Journal of the American Society for Information and Technology,* **56**(12), 1303-1313.

Bilal, D. (2000). "Children's use of the Yahooligans! search engine. I. Cognitive, physical, and affective behaviors on fact-based search tasks". *Journal of the American Society for Information Science,* **51**(7), 646-665.

Bilal, D. (2001). "Children's use of the Yahooligans! Web search engine. II. Cognitive and physical behaviors on research tasks". *Journal of the American Society for Information Science and Technology,* **52**(2), 118-137.

Bilal, D. (2002b). "Children's use of the Yahooligans! Web search engine. III. Cognitive and physical behaviors on fully self-generated tasks". *Journal of the American Society for Information Science and Technology,* **53**(13), 1170-1183.

Bilal, D. (2004). "Research on children's information seeking on the Web. In: Chelton, M.K. & Cool, C. (eds.) *Youth Information-Seeking: Theories, Models, and Approaches,* pp. 271-291, Lanham, MD: Scarecrow Press.

Bilal, D., & Kirby, J. (2002a). "Differences and similarities in information seeking: Children and adults as Web users". *Information Processing and Management,* **38**(5), 649-670.

Black, J.A. & Champion, D.J. (1976). *Methods and issues in social research.* New York: Wiley.

Block, M. (2004). "The Invisible Web: A presentation for LACONI-RASS on at Batavia Public Library" [Online]. http://marylaine.com/hidden.html [Accessed 2 February 2005].

Blumer, H. (1937), "Social psychology". In: Schmidt, E.P. (eds.), *Man and Society,* Englewood Cliffs, NJ: Prentice-Hall.

Botluk, D. (1999). "Exposing the Invisible Web" [Online]. *LLRX.com's ResearchWire.* http://www.llrx.com/columns/exposing.htm [Accessed 22 October 2003].

Broder, A., Kumar, R. Maghoul, F., Raghavan, P. Rajagopalan, S. Stata, R., Tomkins, A. Wiener J. (2000), "Graph structure in the web". *Ninth International World-Wide Web Conference* [Online]. http://www.almaden.ibm.com/cs/k53/www9.final/ [Accessed 22 October 2003].

Brown, A.D. (1990). *Information, Communication and Organizational Culture: A Grounded Theory Approach*. PhD, University of Sheffield.

Brown, C.M. (1999). "Information-seeking behavior of scientists in the electronic information age: astronomers, chemists, mathematicians, and physicists". *Journal of the American Society for Information Science*, **50**(10), 929-943.

Bruce, H., Jones, W. & Dumais, S. (2004). "Information behaviour that keeps found things found." *Information Research* [Online]. **10**(1). [http://InformationR.net/ir/10-1/paper207.html] [Accessed 15 October 2005].

Burr, V. (1995). *An Introduction to Social Constructionism*. London: Routledge.

Caregnato, S. (2000), "*Modelling the Information Skills Development Domain: A Grounded Theory Approach to Knowledge Elicitation*". PhD thesis, University of Sheffield.

Case, D.O. (2002). *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. New York: Academic Press.

Catledge, L.D. & Pitkow J.E. (1995). "Characterizing browsing strategies in the World-Wide Web". *Computer Networks and ISDN Systems*, **26**(6), 1065-1073.

Ceaparu, I., Lazar, J., Bessiere, K., Robinson, J., Shneiderman, B. (2004). "Determining causes and severity of end-user frustration". *International Journal of Human-Computer Interaction*, **17**(3), 333-356.

Charmaz, K. (2000). "Grounded theory: Objectivist and constructivist methods". In: Denzin N.K. & Lincoln Y.S. (eds.), *Handbook of Qualitative Research, 2nd edition*, pp. 509-535. Thousand Oaks, Ca: Sage.

Chinn, S.S. (2002). "Attribution theory applied to information technology" [Online]. http://www2.hawaii.edu/~sherlock/attribution.pdf [Accessed 22 October 2005]

Choo, C. W and Marton, C. (2003). "Information seeking on the Web by women in IT professions". *Internet Research: Electronic Networking Applications*, **13**(4), 267-280.

Choo, C.W., Detlor, B. & Turnbull, D. (1999). "Information seeking on the Web: An integrated model of browsing and searching". In: Woods, L. (eds.). *Proceedings of the 62nd ASIS Annual Meeting*, pp. 3-16. Medford: Information Today.

Choo, C.W., Detlor, B. and Turnbull, D. (2000). *Web work: information seeking and knowledge work on the World Wide Web*. Dordrecht: Kluwer Academic Publishers.

Choo, C.W., Detlor, B., & Turnbull, D. (1998). "A behavioral model of information seeking on the web: Preliminary results of a study of how managers and it specialists use the web". In: Preston, C. (eds.), *Proceedings of the 61st Annual Meeting of the American Society for Information Science*, pp.290-302. Pittsburgh, PA.

Choo, C.W., Detlor, B., and Turnbull, D. (2000), "Information seeking on the Web: An integrated model of browsing and searching". *First Monday* [Online] **5**(2). http://firstmonday.org/issues/issue5_2/choo/index.html [Accessed 12 August 2003].

Chowdhury, G.G. (1999). "The Internet and information retrieval research: a brief review". *Journal of Documentation*, **55**(2), 209-225.

Chu, F.T. (1994). "Reference service and bounded rationality: Helping students with research". *College & Research Libraries*, **55**, 457-461.

Clyde, A. (2002). "The invisible Web". *Teacher Librarian*, **29**, 47-49.

Cockburn, A. & McKenzie, B. (2001). "What do Web users do? An empirical analysis of Web use". *International Journal of Human Computer Studies*, **54**(6), 903-22.

Cook, T. & Reichardt, C. (1979). *Qualitative and Quantitative Methods in Evaluation Research.* Beverly Hills: Sage.

Cooley, C. H. (1922). *Human Nature and the Social Order*. New York: Scribner & Sons.

Correia, Z. & Wilson, T.D. (1997). "Scanning the business environment for information: a grounded theory approach". *Information Research* [Online]. **2**(4). http://informationr.net/ir/2-4/paper21.html [Accessed 2 December 2004].

Cothey, V. (2002). "A longitudinal study of World Wide Web users' information-searching behaviour". *Journal of the American Society for Information Science and Technology,* **53**(2), 67-78.

Crowther, P., Bechhofer, S., Horan, B. (2004). "Shedding light on the hidden Web" [Online]. http://www.cs.man.ac.uk/cspreprints/PrePrints/cspp28.pdf [Accessed 15 March 2005].

Cutcliffe, J.R. (2000). "Methodological issues in grounded theory". *Journal of Advanced Nursing,* **31**(6), 1476-1488.

Dahn, M. (2000). "Spotlight on the invisible Web". *Online*, **24**(4), 57-62.

Davidson, J. & Rincones, R. (2001). "Qualitative research methods" [Online]. http://homepages.wmich.edu/~jdavidso/resmeth/640wk5/640wk5.PDF [Accessed 20 August 2004].

Denzin, N.K., & Lincoln, Y.S. (1998). *Collecting and Interpreting Qualitative Materials.* London: Sage.

Dervin, B. (1992). "From the mind's eye of the user: the sense-making qualitative-quantitative methodology". In: Glazier, J. & Powell, R. (eds.), *Qualitative research in information management*, pp. 61-84. Englewood, CO: Libraries Unlimited.

Detlor, B. (2005). "Web information behaviors of organizational workers". In: Fisher, K. E., Erdelez, S., & McKechnie, E.F. (eds.), *Theories of Information Behavior*, pp. 377-381. Medford, NJ: Information Today.

Devine, J. and Egger-Sider, F. (2004). "Beyond Google: the invisible web in the academic library". *The Journal of Academic Librarianship*, **30**(4), 265-269.

Diaz, K. R. (2000). "The invisible Web: navigating the Web outside traditional search engines". *Reference & User Services Quarterly,* **40**(2), 131-134.

Eastin, M.S. & LaRose, R. (2000), "Internet self-efficacy and the psychology of the digital divide". *Journal of Computer-Mediated Communication* [Online]. **6**(1). http://jcmc.indiana.edu/vol6/issue1/eastin.html [Accessed 20 August 2004].

Edwards, D. & Potter, J. (1993). "Language and causation: a discursive action model of description and attribution". *Psychological Review*, **100**(1), 23-41.

Ellis, D. (1987). *The Derivation of Behavioural Model for Information Retrieval System Design, PhD,* University of Sheffield.

Ellis, D. (1993), "Modelling the information-seeking patterns of academic researchers: A Grounded Theory approach". *Library Quarterly*, **63**(4), 469-486.

Ellis, D., Cox, D., Hall, K.A. (1993), "A comparison of the information seeking patterns of researchers in the physical and social sciences". *Journal of Documentation*, **49**(4) 356-369.

Fidel, R., Davies R.K., Douglass M.H., Holder J.K., Hopkins C.J., Kushner, E.J. Miyagishima, B.K., & Toney, C.D. (1999). "A visit to the information mall: Web searching behaviour of high school students". *Journal of the American Society for Information Science,* **50**(1), 24-37.

Fisher, K.E., Erdelez, S., & McKechnie, E.F. (eds.) (2005). *Theories of Information Behavior.* Medford, NJ: Information Today.

Folkman, S., & Lazarus, R.S. (1980). "An analysis of coping in a middle-aged community sample". *Journal of Health and Social Behavior,* **21**, 219-239.

Ford, N., Miller, D. & Moss, N. (2003). "Web search strategies and approaches to studying". *Journal of the American Society for Information Science and Technology,* **54**(6), 473-489.

Ford, N., Miller, D. & Moss, N., (2005a). "Web search strategies and human individual differences: Cognitive and demographic factors, internet attitudes, and approaches". *Journal of the American Society for Information Science and Technology*, **56**(7), 741-756.

Ford, N., Miller, D. & Moss, N., (2005b). "Web search strategies and human individual differences: A combined analysis". *Journal of the American Society for Information Science and Technology*, **56**(7), 757-764.

Ford, N., Miller, D. (1996), "Gender differences in Internet perception and use". In: *Electronic Library and Visual Information Research: Papers from the third ELVIRA Conference*, 30 April-2 May, 1996, ASLIB, London, pp. 87-100.

Ford, N., Miller, D., & Moss, N. (2002), "Web search strategies and retrieval effectiveness: An empirical study". *Journal of Documentation*, **58**(1), 30-48.

Ford, N., Miller, D., Moss, N. (2001), "The role of individual differences in Internet searching: an empirical study". *Journal of the American Society for Information Science and Technology*, **52**(12), 1049-66.

Garcia, F. (1996). "Business and marketing on the Internet" [Online]. http://web.archive.org/web/19980113075450/http://www.tcp.ca/Jan96/BusandMark.html [Accessed 10 November, 2003].

Gergen, M. & Gergen, K.J. (eds.) (2003). *Social Constructionism: A Reader.* London: Sage.

Gigerenzer, G., & Goldstein, D.G. (1996). "Reasoning the fast and frugal way: models of bounded rationality". *Psychological Review* [Online]. **103**, 650-696 http://www.dangoldstein.com/papers/FastFrugalPsychReview.pdf [Accessed 20 December 2005].

Glaser, B. & Strauss, A. (1967). *The Discovery of the Grounded Theory: Strategies for Qualitative Research.* New York: Aldine de Gruyter.

Glaser, B. (1978). *Theoretical Sensitivity.* Mill Valle, CA: Sociology Press.

Glaser, B. (1992). *Emergence v Forcing Basics of Grounded Theory Analysis.* Mill Valle, CA: Sociology Press.

Glaser, B. (1998). *Doing Grounded Theory: Issues and Discussion.* Mill Valle, CA: Sociology Press.

Glaser, B. (1999). "Keynote address for the fourth annual qualitative health research conference". *Qualitative Health Research*, **9**(6), 836-845.

Goede, R. & De Villiers (2003). "The applicability of Grounded Theory as research methodology in studies on the use of methodologies in IS practice" [Online].

http://www.cs.vu.nl/~gordijn/QRM/lib/IS-GroundedTheory-Goede-Sep2003.pdf [Accessed 2 December 2005]

Gordon, M., & Pathak, P. (1999). "Finding information on the World Wide Web: The retrieval effectiveness of search engines". *Information Processing and Management*, **35**(2), 141-180.

Gorman, G. E. & Clayton, P. (2005). *Qualitative Research for the Information Professional: A Practical Handbook*. London: Facet Publishing.

Griffiths, J.R. & Brophy, P. (2002). "Student searching behaviour in the JISC information environment". *Ariadne* [Online]. (33). http://www.ariadne.ac.uk/issue33/edner [Accessed 20 March 2003].

Griffiths, J.R., Hartley, R.J., & Wilson, J.P. (2002) "An improved method of studying user-system interaction by combining transaction log analysis and protocol analysis" *Information Research* [Online]. 7(4). http://InformationR.net/ir/7-4/paper139.html [Accessed 20 March 2003].

Guba, E.G., & Lincoln, Y.S. (1989). *Fourth Generation Evaluation*. Newbury Park, CA: Sage.

Gulli, A. & Signorini, A. (2005). "The indexable web is more than 11.5 billion pages". In: *Proceedings of the Special interest tracks and posters of the 14th international conference on World Wide Web*, 10-14 May, 2005, Chiba, Japan. pp. 902-903.

Haig, B.D. (1995). "Grounded Theory as scientific method" [Online]. http://www.ed.uiuc.edu/EPS/PES-yearbook/95_docs/haig.html [Accessed 2 December 2004]

Halfpenny, P. (1979). "The analysis of qualitative data". *Sociological Review (New series)*, **27**, 799-825.

Hammersley, M. (1987). "Some notes on the terms 'validity' and 'reliability". *British Educational Research Journal*, **13**(1), 73-81.

Hammersley, M. (1989). *The Dilemma of Qualitative Method: Herbert Blumer and the Chicago Tradition*, London: Routledge.

Hargittai, E. (2002). "Beyond logs and surveys: In-depth measures of people's Web use skills". *Journal of the American Society for Information Science and Technology*, **53**(14), 1239-1244.

Haverkamp, D.S., & Gauch, S. (1998). "Intelligent information agents: review and challenges for distributed information sources". *Journal of the American society for information science*, **49**(4), 304-311.

Heath, H. & Cowley, S. (2004). "Developing Grounded Theory approach: a comparison of Glaser and Strauss". *International Journal of Nursing Studies* **41(2)**, 141-150.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.

Hill, J.R. (1997). "The World Wide Web as a tool for information retrieval: an exploratory study of users' strategies in an open-ended system". *School Library Media Quarterly*, **25**(4), 229-236.

Hobbins, K. (2001). "Surfing the Invisible Web" [Online]. http://www.oaltabo.on.ca/iw.ppt [Accessed 10 December 2004).

Hoelscher, C. & Strube, G., (2000). "Web search behaviour of Internet experts". In: *Proceedings of the Ninth International World Wide Web Conference*, 337-346.

Hoelscher, C. (1998). "How Internet experts search for information on the Web". In: Maurer, H. & Olson, R.G. (eds.) *Proceedings of WebNet98 - World Conference of the WWW, Internet & Intranet*. Charlottesville, VA: AACE.

Hricko, M. (2002). "Using the Invisible Web to teach information literacy". *Journal of Library Administration*, **37**(3-4), 379-386.

Hsieh-Yee, I. (1993). "Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers". *Journal of the American Society for Information Science*, **44**(3), 161-174.

Hsieh-Yee, I. (1998). "Search tactics of web users in searching for texts, graphics, known items and subjects: a search simulation study". *Reference Librarian.* **60**, 61-85.

Hsieh-Yee, I. (2001). "Research on web search behaviour". *Library and Information Science Research,* **23**(2), 167-185.

Huberman, B.A., Pirolli, P.L., Pitkow J.E., & Lukose, R.M. (1998). "Strong regularities in World Wide Web surfing". *Science,* **280**, (5360), 94-97.

Hultgren, F. (2003). "Review of: Case, D.O. *Looking for information: a survey of research on information seeking, needs and behavior.* San Diego: Academic Press, 2002. *Information Research* [Online]. **8**(3). http://informationr.net/ir/reviews/revs095.html [Accessed 12 March 2005].

Ingwersen, P. (1996). "Cognitive perspectives of information retrieval interaction". *Journal of Documentation*, **52**(1), 3-50.

Jamali, H.R. (2005). "Review of: Fisher, K.E., Erdelez, S., & McKechnie, E.F. (eds.). *Theories of information behavior.*" Medford, NJ: Information Today, 2005. *Webology* [Online]. **2**(4). http://www.webology.ir/2005/v2n4/bookreview2.html [Accessed 15 January 2006].

Jansen, B. J. & Pooch, U. (2001). "A review of Web searching studies and a framework for future research". *Journal of the American Society for Information Science and Technology*, **52**(3), 235-246.

Jansen, B.J. (2000). "The effect of query complexity on Web searching results". *Information Research* [Online]. **6**(1). http://InformationR.net/ir/6-1/paper87.html [Accessed 19 August 2003].

Jansen, B.J., & Spink, A. (2000). "The Excite research project: A study of searching characteristics by Web users". *Bulletin of the American Society for Information Science*, **27**(1), 15-17.

Jansen, B.J., Spink, A. & Saracevic, T. (2000). "Real life, real users and real needs: A study and analysis of users' queries on the Web". *Information Processing and Management.* **36**(2), 207-227.

Jawahar, I.M. & Elango, B. (2001). "The effects of attitudes, goal setting and self-efficacy on end user performance". *Journal of End User Computing*, **13**(2), 40-45.

Johnston, J. M. & Pennypacker, H.S. (1980). *Strategies and Tactics of Human Behavioural Research.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Jones, R. (1995). "Why do qualitative research". *British Medical Journal* [Online]. **311**(2). http://bmj.bmjjournals.com/cgi/content/full/311/6996/2 [Accessed 20 August 2003].

Kalbach, J (2000). "Designing for information foragers: a behavioural model for information seeking on the World Wide Web." *Internet Working* [Online]. **3.3** (9). http://www.internettg.org/newsletter/dec00/article_information_foragers.html [Accessed 18 September 2005].

Kalbach, J (2003). "I'm feeling lucky: The role of emotions in seeking information on the web" [Online]. http://home.earthlink.net/~searchworkshop/docs/JKalbach_Emotions-InformationSeeking-Web_short21.pdf [Accessed 20 April 2004].

Kari, J. & Savolainen, R. (2001). "Web searching in the context of information seeking in everyday life: The cases of civic and spiritual action" [Online]. http://www.uta.fi/~csjakar/kari-savolainen.pdf [Accessed 20 June 2005].

Kari, J. & Savolainen, R. (2002). "Web searching in the context of information seeking in everyday life: A synopsis of research proposal" [Online]. http://www.uta.fi/~csjakar/ka-sa.pdf [Accessed 20 June 2005].

Kari, J. & Savolainen, R. (2003). "Towards a contextual model of information seeking on the web". *The New Review of Information Behaviour Research,* **4,** 155-175.

Keily, L. (1997). "Improving resource discovery on the Internet: The user perspective". *21st International Online Information Meeting.* pp. 205-212. London: Learned Information Europe.

Kelley, H.H. (1973). "The processes of causal attribution". *American Psychologist*, **28**, 107-128.

Kendall, J. (1999). "Axial coding and grounded theory". *Western Journal of Nursing Research.* **21**(6), 743-757.

Kerlinger, F. (1964). *Foundations of Behavioural Research.* New York: Holt.

Kim, K. (2000). "Effects of cognitive style on web search and navigation". In: *World Conference on Educational Multimedia, Hypermedia and Telecommunications 2000* [Online]. (1), pp. 531-536. http://slisweb.lis.wisc.edu/~kskim/Publications/Ed-media00.pdf [Accessed 20 March 2003].

Kim, K. (2001). "Information-seeking on the web: effects of user and task variables". *Library and Information Science Research*, **23**(3), 233-255.

Kim, Y-S. (2004). *The Public Library and Its Support Groups: In What Way and to What Extent Do Friends Influence Public Library Policy.* PhD, University of Sheffield.

Klobas, J.E. & Clyde, L.A. (2000). "Adults learning to use the Internet: a longitudinal study of attitudes and other factors associated with intended Internet use". *Library and Information Science Research,* **22**(1), 5-34.

Kuhlthau, C.C. (1991). "Inside the search process: information-seeking from the user's perspective". *Journal of the American Society for Information Science*, **42**(5), 361-371.

Kurbanoglu, S.S. (2003). "Self-efficacy: A concept closely linked to information literacy and lifelong learning". *Journal of Documentation*, **59**(6), 635-646.

Lawrence, S. & Giles, C.L. (1998). "Searching the World Wide Web." *Science,* **280**(3), 98-100.

Lawrence, S. & Giles, C.L. (1999). "Accessibility of information on the web." *Nature*, **400**, 107-109.

Laycock, J. (2004). "Search ranks as second most popular online activity". *Search Engine Guide* [Online]. http://www.searchengineguide.com/laycock/002108.html [Accessed 10 December, 2005).

Lazonder, A.W., Biemans, H.J.A., Wopereis, I.G.J.H. (2000). "Differences between novice and experienced users in searching information on the World Wide Web". *Journal of the American Society for Information Science,* **51**(6), 576-581.

Lehmann, H. (2001). *A Grounded Theory of International Information Systems*. PhD, University of Auckland.

Lehner, P.N. (1979). *Handbook of Ethological Methods*. New York: STPM Press.

Lewandowski, D. (2005). "Web searching, search engines and information retrieval" [Online]. http://www.durchdenken.de/lewandowski/doc/isu_preprint.pdf [Accessed 10 March, 2006).

Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage Publications.

Lucas, H. (1999). "Representing WWW navigational data: A graphical methodology to support qualitative analysis". *Information Research* [Online]. **4**(3). http://informationr.net/ir/4-3/paper59.html [Accessed 2 December 2005].

Lucas, H.S. (2001). Patterns of Influence: A Qualitative Model of Users' Interactions with World Wide Web Medical Resources. PhD, University of Sheffield.

Lucas, W. and Topi, H. (2002). "Form and functions: the impact of query term and operators usage on web search results". *Journal of the American Society for Information Science and Technology*, **53**(2), 95-108.

Macluhan, H.M. (1962). *The Gutenberg Galaxy: the Making of Typographic Man*. London: Routledge & Kegan Paul.

Madden, A.D., Ford, N.J., Miller, D., Levy, P. (2006). "Children's use of the Internet for information-seeking: What strategies do they use, and what factors affect their performance?" *Journal of Documentation*, (in press)

Mamlin, N., Harris, K.R., Case, L.P. (2001). "A methodological analysis of research on Locus of Control and learning disabilities: Rethinking a common assumption". *Journal of Special Education*, **34**(4), 214-25.

Marchionini, G., Lin. X., & Dwiggins, S. (1990). "Effects of search and subject expertise on information seeking in a hypertext environment". In: *Proceedings of the 53rd Annual Meeting of the ASIS*, pp. 129-142.

Marchionini, G.M. (1995). *Information Seeking in Electronic Environments*. Cambridge: Cambridge University Press.

Martzoukou, K. (2004). "A review of Web information seeking research: Considerations of method and foci of interest". *Information Research* [Online]. **10**(2) http://InformationR.net/ir/10-2/paper215.html [Accessed 10 March 2005].

Mason, J. (2nd eds.) (2002). *Qualitative Researching*. London: Sage.

McInerney, V., McInerney, D.M., Sinclair, K.E. (1994), "Student teachers, computer anxiety and computer experience". *Journal of Educational Computing Research*, **11**(1), 177-89.

McLeod, R. (2002). "Information frustration". *Information World Review*, **178**, 28.

Mead, G.H. (1934). *Mind, Self and Society*. Chicago: Chicago University Press.

Meier, S. (1985), "Computer aversion". *Computers in Human Behaviour*, **1**(1), 171-179.

Mellon, C. (1990). *Naturalistic Inquiry for Library Science*. New York: Greenwood Press.

Mellon, C.A. (1986), "Library anxiety: A grounded theory and its development". *College & Research Libraries*, **47**, 160-165.

Miles, M. & Huberman, A. (1984). *Qualitative Data Analysis: A Sourcebook of New Methods*. London: Sage.

Mishler, E.G. (1990). "Validation in enquiry-guided research: The role of exemplars in narrative studies". *Harvard Educational Review*, **60**, 415-442.

Morse, J.M. (1998). "What's wrong with random selection?". *Qualitative Health Research*, **8**, 733-735.

Musoke, M. (2001). *Health Information Access and Use in Rural Uganda: An Interaction-Value Model*. PhD, University of Sheffield.

Myers, M.D. (1997). "Qualitative research in information systems," *MIS Quarterly* [Online]. **21** (2), 241-242. http://www.misq.org/discovery/MISQD_isworld/ [Accessed 10 March 2005].

Nahl, D. (1993). *CD-ROM Point of Use Instructions for Novice Searchers: A Comparison of User-centered Affectively Elaborated and System-centered Unelaborated Text*. PhD, University of Hawaii.

Nahl, D. (1995). "Affective elaborations in Boolean search instructions for novices: Effects on comprehension, self-confidence, and error type". In: Kinney, T. (eds.) *Proceedings of the 58th ASIS Annual Meeting*, 9-12 October 1995, Chicago, Illinois. pp. 69-76. Learned Information, NJ.

Nahl, D. (1996), "Affective monitoring of Internet learners: perceived self-efficacy and success". In: Hardin, S. (eds.) *Proceedings of the 59th Annual Meeting of the American Society for Information Science*. 21-24 October 1996, Baltimore, Maryland. pp. 100-109. Information Today, Medford, NJ.

Nahl, D. (1998). "Learning the Internet and the structure of information behaviour". *Journal of the American Society for Information Science*, **49**(11), 1017-1023.

Newby, G.B. (1998). "The strong cognitive stance as a conceptual basis for the role of information in informatics and information system design" [Online]. http://www.petascale.org/tenure/papers/cogspace-98.pdf [Accessed 12 September 2005].

Newell, A. & Simon, H.A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Olaisen, J. (1991). "Pluralism or positivistic trivialism: Important trends in contemporary philosophy of science". In: Nissen, H.E. Klein, H.K. & Hirschheim R. (eds.), *Information Systems Research: Contemporary Approaches and Emergent Traditions* pp. 235-265, Elsevier, Amsterdam.

Oliver, D., Whymark, G., & Romm, C. (2005), "Researching ERP adoption: An internet-based Grounded Theory approach". *Online Information Review*, **29**(6), 585-603.

Oliver, P. (2004). *Writing Your Thesis*. London: Sage.

Ondrusek, A.L. (2004). "The attributes of research on end-user online searching behaviour: A retrospective review and analysis". *Library and Information Science Research*, **26**(2), 221-265.

Ozmutlu, H.C., Spink, A. & Ozmultu, S. (2002). "Analysis of large data logs: an application of Poisson sampling on Excite Web queries". *Information Processing and Management*, **38** (4), 473-490.

Pace, S. (2003). *Understanding the Flow Experiences of Web Users*. PhD, Australian National University.

Pace, S. (2004). "A grounded theory of the flow experiences of Web users". *International Journal of Human-Computer Studies*, **60**(3), 327-363.

Pajares, F. (2002). "Overview of social cognitive theory and of self-efficacy" [Online]. http://www.emory.edu/EDUCATION/MFP/eff.html [Accessed 11 December 2005]

Palmquist, R.A. (2001). "Cognitive style and users' metaphors for the web: An exploratory study". *The Journal of Academic Librarianship*, **27**(1), 24-32.

Palmquist, R.A., & Kim, K.S. (2000). "The effect of cognitive style and online search experience on Web search performance". *Journal of the American Society of Information Science*, **51**(6), 558-567.

Patton, M.Q. (1980). *Qualitative Evaluation Methods*. Beverly Hills: Sage Publications.

Pedley, P. (2001). *The invisible Web: Searching the Hidden Parts of the Internet*. London: Aslib.

Pedley, P. (2002). "Why you can't afford to ignore the Invisible Web". *Business Information Review*, **19**(1), 23-31.

Pedraza, M. (1986). "*Learning as a Process of Gaining Control of the Sources of Information: A Grounded Theory Approach Based on Evidence from Columbia*". PhD, University of Sheffield.

Pejtersen, A.M. & Fidel, R. (1998). "A framework for work-centered evaluation and design: a case study of IR on the Web". In: *Multimedia Information Retrieval Applications (Mira) Workshop*, Grenoble, France.

Peters, T. (1993). "The history and development of transaction log analysis". *Library Hi Tech*, **11**(2), 41-66.

Peterson, R. (1997). "Eight Internet search engines compared". *First Monday* [Online]. **2**(2). http://www.firstmonday.dk/issues/issue2_2/peterson/ [22 July 2003].

Pickard, A. (1998). The impact of access to electronic and digital information resources on learning opportunities for young people: A grounded theory approach, *Information Research* [Online]. **4**(2). http://informationr.net/ir/4-2/isic/pickard.html [Accessed 2 December 2004].

Pollock, A. & Hockley, A. (1997). "What's wrong with Internet searching?" *D-Lib Mag* [Online]. **3**(3). http://www.dlib.org/dlib/march97/bt/03pollock.html [Accessed 15 March 2004].

Powell, R. (1999). "Recent trends in research: a methodological essay". *Library and Information Science Research*, **21**(1), 91-119.

Ratcliff, D. (2004). "Qualitative research resources" [Online]. http://don.ratcliff.net/qualresources/ [Accessed 5 January 2006].

Ren, W.H. (2000). "Library instruction and college student self-efficacy in electronic information searching". *Journal of Academic Librarianship*, **26**(5), 323-8.

Rieh, S.Y. (2003). "Investigating Web searching behaviour in home environments". In: Todd, R.J. (eds.), *Proceedings of the 66th ASIST annual meeting*, **40**, pp. 255-264. Information Today, Medford, NJ.

Rieh, S.Y. (2004). "On the web at home: information seeking and web searching in the home environment". *Journal of the American Society for Information Science and Technology*, **55**(8), 743-753.

Rimal, R.N. (2001), "Perceived risk and self-efficacy as motivators: Understanding individuals' long term use of health information". *Journal of Communication*, **51**(4) 633-54.

Roselle, A. & Neufeld, S. (1998). "The utility of electronic mail follow-ups for library research". *Library and Information Science Research,* **20**(2), 153-161.

Ross, N. & Wolfram, D. (2000). "End user searching on the Internet: An analysis of term pair topics submitted to the Excite search engine". *Journal of the American Society for Information Science*, **51**(10), 949-958.

Rotter, J.B. (1966). "Generalized expectancies for internal versus external control of reinforcement". *Psychological Monographs*, **80**(1), 1-28.

Saracevic, T. (1975). "Relevance: A review of and a framework for the thinking on the notion in information science". *Journal of the American Society for Information Science*, **26**(6), 321-343.

Saracevic, T., Kantor. P., Chamis, A.Y., & Trivison, D. (1988). "A study of information seeking and retrieving. I. Background and methodology". *Journal of the American Society for Information Science*, **39**(3), 161-176.

Savolainen, R. & Kari, J. (2004). "Conceptions of the Internet in everyday life information seeking". *Journal of Information Science*, **30**(3), 219-226.

Savolainen, R. (1995). "Everyday life information seeking: Approaching information seeking in the context of 'way of life' ". *Library and Information Science Research*, **17**, 259-294.

Savolainen, R. (1999). "The role of the Internet in information seeking: Putting the networked services in context". *Information Processing and Management*, **35**, 765-782.

Savolainen, R. (2002), "Network competence and information seeking on the Internet: From definitions towards a social cognitive model". *Journal of Documentation*, **58**(2), 211-226.

Schacter, J., Chung, G.K.W.K., & Dorr, A. (1998). "Children's internet searching on complex problems: Performance and process analyses". *Journal of the American Society for Information Science,* **49**(9), 840-849.

Schamber, L. (1994). "Relevance and information behaviour". *Annual Review of Information Science and Technology,* **29**, 3-48.

Schwartz, C. (1998). "Web search engines." *Journal of the American Society for Information Science*, **49**(11), 973-982.

Schwartz, C.A. (1989). "Book selection, collection development, and bounded rationality". *College & Research Libraries*, **50**, 328-343.

Selden, L. (2005), "On grounded theory - with some malice". *Journal of Documentation*, **61**(1), 114-129.

Sherman, C. & Price, G. (2001). "Exploring the Invisible Web". *Online*, **25**(4), 32-35.

Sherman, C. & Price, G. (2001). *The Invisible Web: Uncovering Information Sources Search Engines Can't See*. Medford, N.J: CyberAge Books.

Sherman, C. & Price, G. (2003). "The Invisible Web: Uncovering sources search engines can't see". *Library Trends*, **52**(2), 282-298.

Sherman, C. (2001). "Search for the Invisible Web" [Online]. *Guardian Unlimited* http://www.guardian.co.uk/online/story/0,3605,547140,00.html [Accessed 22 November 2005].

Sherman, C. (2003). "The Invisible Web". *The Simmons Librarian: Alumni Newsletter of Graduate School of Library and Information Science* [Online]. **35**(2) http://my.simmons.edu/gslis/alumni/Sherman_2003_Presentation.pdf [Accessed 22 November 2004].

Silverstein, C., Marais, H., Henzinger, M., & Moricz, M. (1999). "Analysis of a very large Web search engine query log". *ACM SIGIR Forum*, **33**(1), 6-12.

Simon, H.A. (1955). "A behavioral model of rational choice". *Quarterly Journal of Economics*, **69**, 99-118.

Simon, H.A. (1956). "Rational choice and the structure of the environment". *Psychological Review*, **63**, 129-138.

Simon, H.A. (1979). *Models of Thought*. New Haven, CT: Yale University Press.

Slone, D. (2002). "The influence of mental models and goals on search patterns during Web interaction." *Journal of the American Society for Information Science and Technology*, **53**(13), 1152-1169.

Smith, C.B. (2001). "Getting to know the Invisible Web". *Library Journal*, **126**(11), 16-19.

Snow, B. (2000). "The Internet's hidden content and how to find it" [Online]. http://www.findarticles.com/ [Accessed 15 February 2003].

Soto, S. (1990). *Information in Dentistry: Patterns of Communication and Use*. PhD, University of Sheffield.

Soto, S. (1992). "Using grounded theory analysis to study the information-seeking behaviour of dental professionals". *Information Research News*, **3**(1), 2-12.

Spink, A. & Cole, C. (2001). "Introduction to the special issues: Everyday life information-seeking research". *Library and Information Science Research*, **23**(4) 301-304.

Spink, A. & Jansen, B. (2004). *Web Search: Public Searching of the Web*. Netherlands: Kluwer Academic Publishers.

Spink, A. & Qin, J. (2000). "Introduction to special issue on Web research and information retrieval". *Information Processing and Management,* **36**(2), 205-206.

Spink, A. & Xu, J.L. (2000). "Selected results from a large study of Web searching: the Excite study". *Information Research* [Online]. **6**(1). http://InformationR.net/ir/6-1/paper90.html [Accessed 2 April 2003].

Spink, A. (2002). "A user-centred approach to evaluating human interaction with Web search engines: An exploratory study". *Information Processing and Management,* **38**(3), 401-426.

Spink, A. (2003), "Web search: Emerging patterns". *Library Trends*, **52**(4), 299-306.

Spink, A., Bateman, J. & Jansen, B. J. (1998). "Searching heterogeneous collections on the Web: behaviour of Excite users". *Information Research* [Online]. **4**(2). http://informationr.net/ir/4-2/paper53.html, [Accessed 2 April 2003].

Spink, A., Wolfram, D., Jansen, B. J., & Saracevic, T. (2001). "Searching the Web: The public and their queries". *Journal of the American Society for Information Science and Technology*, **52**(3), 226-234.

Star, S.L. (1998). "Grounded classification: Grounded theory and faceted classification". *Library Trends*, **47**(2), 218-232.

Strauss, A. & Corbin, J. (1990). "*Basics of Qualitative Research: Grounded Theory Procedure and Techniques*. Newbury Park, London: Sage.

Strauss, A. & Corbin, J. (1994). Grounded theory methodology: An overview. In: Denzin, N.K. and Lincoln, Y.S. (eds.) *Handbook of Qualitative Research*, pp. 273-285, Sage, Thousand Oaks.

Strauss, A. & Corbin, J. (1998). *Basics of Qualitative Research: Grounded Theory Procedures and Technique*. London: Sage.

Strauss, A. (1987). *Qualitative Analysis for Social Scientists*. New York: Cambridge University Press.

Sullivan, D. (2000). "Invisible Web gets deeper". *The Search Engine Report*, [On-line]. http://searchenginewatch.com/sereport/00/08-deepweb.html [Accessed 12 August 2003].

Tauscher, L. & Greenberg, S. (1997a). "How people revisit web pages: Empirical findings and implications for the design of history systems". *International Journal of Human-Computer Studies,* **47**, 97-137.

Tauscher, L. & Greenberg, S. (1997b). "Re-visitation patterns in World Wide Web navigation". In: Pemberton, S. (eds.) *Proceedings of CHI 97 Human Factors in Computing Systems,* pp. 399-406. Atlanta, Georgia.

Taylor, S. (1998). "Coping strategies" [Online]. http://www.macses.ucsf.edu/Research/Psychosocial/notebook/coping.html [Accessed 15 May 2006].

Thompson, L.F., Meriac, J.P., Cope, J.G. (2002), "Motivating online performance: the influences of goal setting and Internet self-efficacy". *Social Science Computer Review*, **20**(2), 149-60.

Turnbull, D. (2005). "World Wide Web information seeking". In: Fisher, K.E., Erdelez, S., & McKechnie, E.F. (eds.), *Theories of Information Behavior*, pp. 397-400. Medford, NJ: Information Today.

Vaughan, J. (1999). "Considerations in the choice of an Internet search tool". *Library Hi Tech*, **17**(1), 89-106.

Vaughan, L & Thelwall, M. (2004). "Search engine coverage bias: Evidence and possible causes". *Information Processing and Management*, **40**(4), 693-707.

Vedi, S. (1986), "Information and the Awareness of Leaching of Plant Nutrients: The Case of Lake Ringsjon". PhD, University of Sheffield.

Venkatesh , A. (2006). "Introduction to the special issue on ICT in everyday life: home and personal environments". *Information Society*, **22**(4), 191-194

Walter, M. (2002). "The Invisible Web". *School Library Journal*, **49**(12), 29-29.

Wang, P., Hawk, W.B. & Tenopir, C. (2000). "Users' interaction with World Wide Web resources: an exploratory study using a holistic approach". *Information Processing and Management*, **36**(2), 229-251.

Weiner, B. (1979). "A theory of motivation for some classroom experiences". *Journal of Educational Psychology*, **71**, 3-25.

Weiner, B. (1986). *An Attribution Theory of Motivation and Emotion*. New York: Springer-Verlag.

Weingand, D.E. (1993). Grounded theory and qualitative methodology. *IFLA Journal*, **19**(1), 17-26.

White, M.D. & Iivonen, M. (2001). Questions as a factor in web search strategy. *Information Processing and Management*, **37**, 721-740.

Willig, C. (eds.) (1999). *Applied Discourse Analysis: Social and Psychological Interventions*. Buckingham: Open University Press.

Wilson, T.D. (1980). "On information science and the social sciences". *Social Science Information Studies*, **1**(1), 5-12.

Wilson, T.D. (1981), "On user studies and information needs". *Journal of Documentation*, **37**(1), 3-15.

Wilson, T.D. (1997). "Information Behaviour: An interdisciplinary perspective". *Information Processing and Management*, **33**(4) 551-572.

Wilson, T.D. (2000). "Recent trends in user studies: Action research and qualitative methods". *Information Research* [Online]. **5**(3). http://informationr.net/ir/5-3/paper76.html [Accessed 4 January 2006].

Wilson, T.D., Streatfield, D. (1982), "Action research and users' needs". In: Friberg, I. (eds.), *The 4th International Research Forum in Information Science*, pp.51-70. Högskolan i Borås, Borås.

Winter, G. (2000). "A comparative discussion of the notion of 'validity' in qualitative and quantitative research". *The Qualitative Report* [Online]. **4**(3/4). http://www.nova.edu/ssss/QR/QR4-3/winter.html [Accessed: 18 August 2004].

Wolcott, H.F. (1990). "On seeking and rejecting validity in qualitative research. In: Eisner. E.W. & Peshkin A. (eds.), *Qualitative inquiry in education: The continuing debate*, pp. 121-152. Teachers College Press, New York.

Zafeiriou, G (2001), "*Students' Perceptions of Issues Arising from and Factors Influencing Group Interaction in Computer Conferencing: A Grounded Theory Approach*". PhD, University of Sheffield.

Zakaria, S. (2002), "*Information Processing Strategies and Styles in Relation to Instructional Design: A Grounded Theory Approach*". PhD, University of Sheffield.

Zhang, J. & Dimitroff, A. (2005a). "The impact of webpage content characteristics on webpage visibility in search engine results (Part I)". *Information Processing and Management*, **41**(3), 665-690.

Zhang, J. & Dimitroff, A. (2005b). "The impact of webpage content characteristics on webpage visibility in search engine results (Part I)". *Information Processing and Management*, **41**(3), 691-715.

Zipf, G.K. (1949). *Human Behaviour and the Principle of Least Effort*. Reading, MA: Addison-Wesley.

APPENDICES

Appendices

Appendix 1: Invitation letter for participants

Dear ....,

This letter is an invitation to consider participating in an investigation I am conducting as part of my PhD degree in the Department of Information Studies at the University of Sheffield under the supervision of Professor Nigel Ford and Ms Sheila Webber. I would like to provide you with more information about this project and what your involvement would entail if you decide to take part.

This study will focus on the web-based information seeking behavior of Biologists at the University of Sheffield both generally and in terms of their conceptualization of information seeking failure on the web in particular.

Participation in this study is of course voluntary and it would involve an interview of approximately thirty minutes to take place in a mutually agreed location. You may decline to answer any of the interview questions if you so wish. Further, you may decide to withdraw from this study at any time simply by informing me.

With your permission, the interview would be tape-recorded and later transcribed for analysis. Shortly after the interview has been completed, I would send you a copy of the transcript to give you an opportunity to confirm the accuracy of our conversation and to add or clarify any points that you wish.

All information you provide is completely confidential. Your name will not appear in any thesis or report resulting from this study. Data collected during this study will be in a locked and secure place. On completion of my research, I will be pleased to send you a summary report detailing the results.

If you have any questions regarding this study, or would like additional information to assist you in reaching a decision about participation, please contact me at 0114 2222 675 or by email at y.mansourian@shef.ac.uk . You can also contact my supervisors as well.

I hope that the results of my study will be of benefit to all internet searchers and search engine designers in general and the Biology community in particular. I very much look forward to speaking with you and thank you in advance for your assistance in this project.

Yours Sincerely,

Yazdan Mansourian

Research Students

Appendix 2: Consent form

I agree to participate in a study being conducted by Yazdan Mansourian of the Department of Information Studies, University of Sheffield under the supervision of Professor Nigel Ford and Ms Sheila Webber. I have made this decision based on the information I have read in the Information Letter. In addition I have had the opportunity to receive any further details I wanted about the study. I understand that I may withdraw from the study at any time I wish simply by informing the researcher.

Participant Name (Please print): _____

Participant Signature: _____

Date: _____

☐ Please tick if you would like to be sent a copy of the summary report.

Appendix 3: Consent form for tape recording interview

I understand that the interview will be audio taped to facilitate the collection of information on the understanding that all information I provide will be held in confidence, and I will not be identified in the thesis, summary report, or any publication. I understand that I may withdraw this consent at any time by informing the researcher.

Participant Name: _____

Participant Signature: _____

Date: _____

Appendix 4: Contact information for participants

Yazdan Mansourian

Research Student, Room 204 Tel: +44 (0) 114 22 22675

Department of Information Studies Fax: +44 (0) 114 27 80300

University of Sheffield, Regent Court,

211 Portobello St., Sheffield, S1 4DP

y.mansourian@sheffield.ac.uk

Nigel Ford

Professor, Room 210 Tel: +44 (0) 114 22 22637

Department of Information Studies Fax: +44 (0) 114 27 80300

University of Sheffield, Regent Court,

211 Portobello St., Sheffield, S1 4DP

n.ford@sheffield.ac.uk

Sheila Webber

Lecturer, Room 234 Tel: +44 (0) 114 22 22641

Department of Information Studies Fax: +44 (0) 114 27 80300

University of Sheffield, Regent Court,

211 Portobello St., Sheffield, S1 4DP

s.webber@sheffield.ac.uk

Appendix 5: Preliminary study's interview questions

Q1. Would you please let me know, what are your feelings about looking for information on the web?

Q2. How long have you been seeking your required information on the web?

Q3. How often do you search and what kinds of information do you usually looking for?

Q4. How easily do you usually find your required information through the web?

Q5. How satisfactory is the search result for you?

Q6. Do you usually use any search engine for information seeking on the web?

Q7. Could you please try to remember and describe one of your recent experience in which you have located an important information item on the web that helped you to carry out a significant action or making a key decision?

Q8. Do you remember which search facilities did help you to find this information item?

Q9. How do you usually begin your information seeking process?

Q10. Can you remember which site is your usual starting point for searching the web?

Q11. Why have you selected this site as starting point?

Q12. Is there any specific feature in this site encouraging you to start your search through it?

Q13. Can you remember and describe one of your recent experience in which you did not find your information on the web? What did you looking for and in your opinion why you did not find it?

Q14. Do you ever get the feeling that there is more information on the web that you are not getting to?

Q15. Could you please describe how you decide how successful you have been at web searching?

Q16. Do you search on the web in the same way every time?

Q17. Could you describe the 'invisible web' might mean for you?

Q18. If this term is new for you what is your first impression of it?

Q19. Is there anything else about your experiences of web searching that you would like to add?

Appendix 6: Main study's interview questions

Q1. What is your general feeling about searching the web? How do you find it in general?

Q2. How long have you been searching the web?

Q3. How often do you search the web?

Q4. What kind of information do you usually look for on the web?

Q5. How satisfactory are search results for you in general?

Q6. Can you remember and describe one of your recent experiences in which you were successful in web searching?

Q7. Why do you think in this specific case you have been successful?

Q8. Do you always manage to find what you want on the web or do you ever not find what you have been looking for?

Q9. Can you remember and describe one of your recent experiences in which you failed to find what you have been looking for? Any specific and recent event is useful.

Q10. Why do you think this specific case was a failure?

Q11. How do you feel about your information seeking failure on the web?

Q12. How do you differentiate between failure and success in searching the web?

Q13. Do you ever get the feeling that there should be more relevant information about your search topic on the web but you are not getting to it?

Q14. Do you think that you always find everything exist about your search topic on the Web? I mean, in your opinion how likely is it that you miss something about your search topic?

Q15. How much does it matter to you if you know you have missed something while searching the Web? How do you feel about it?

Q16. What do you do about it? I mean, how do you seek different solutions to cope with information seeking failure on the web?

Q17. How did you develop these coping strategies?

Before ending the interview, I need to clarify just two things else:

Q18. What does the web mean to you? I mean, what do you mean by the web?

Q19. Could you describe the 'invisible web' might mean to you?

Q20. Is there anything else that you would like to add?

Appendix 7: Participants' perceived causation of success and failure in search

There are a number of abbreviations in the table which are explained here. "SS" stands for "Successful Searches", US = Unsuccessful Searches, "I" for "Internal Factors", "X" for "External Factors" and "B" for "Both Internal and External Factors".

Code	SS	US	Causation of Success	Causation of Failure
AS01	X	X	Availability of information, and luck.	Unavailability of information and being unlucky.
AS02	X	X	Availability of information and search tools efficiency.	Search tools deficiency and unavailability information.
AS03	I	X	Linguistic and conceptual ability.	Unavailability of information and being unlucky.
AS04	I	U	Linguistic and conceptual ability and search experience.	Uncertainty about the causes.
AS05	I	X	Domain knowledge and search experience.	Deficiency of website design.
AS06	I	I	Linguistic and conceptual ability, domain knowledge, and search experience.	Lack of domain knowledge and lack of search skills.
AS07	I	X	Domain knowledge and search experience.	Unavailability of information.
AS08	I	X	Domain knowledge and search experience.	Unavailability of information.
AS09	U	X	Uncertainty about the causes.	Unavailability of information.
AS10	X	X	Search tools efficiency.	Unavailability of information.
AS11	X	I	Search tools efficiency.	Lack of conceptual and linguistic ability and lack of search skills.
AS12	I	I	Linguistic and conceptual ability and search experience.	Lack of domain knowledge and lack of search skills.
AS13	X	U	Search tools efficiency.	Uncertainty about the causes.
AS14	X	I	Availability of the required information on the web.	Lack of sufficient effort.
AS15	X	X	Search tools efficiency and website structure.	Unavailability of information.

Code	SS	US	Causation of Success	Causation of Failure
RS16	B	B	Combination of linguistic and conceptual ability, search experience and search tools efficiency.	Combination of unavailability of information and lack of linguistic and conceptual ability and lack sufficient effort.
RS17	I	B	Domain knowledge, linguistic and conceptual ability, and search experience.	Combination of system deficiency and system errors, lack of sufficient effort, lack of linguistic and conceptual ability, and time constraint.
RS18	X	X	Availability of information.	Unavailability of information.
RS19	X	X	Search tools efficiency.	Unavailability of information, deficiency of website design.
RS20	I	B	Confidence and persistence in search and search experience.	Combinations user error, search tools deficiency and unavailability of information.
RS21	B	X	Combination of search experience, linguistic and conceptual ability efficiency of search tool and website structure.	Unavailability of information and search tools deficiency.
RS22	X	I	Search tools efficiency.	Lack of search skills and lack of sufficient effort.
RS23	B	B	Combination of domain knowledge, search experience, and availability of information.	Combination of lack of search skills, lack of sufficient effort, and search tools deficiency and system error.
RS24	X	X	Search tools efficiency.	Unavailability of information.
RS25	B	U	Combination of domain knowledge, search experience, availability of information and website structure.	Uncertainty about the causes.
RS26	I	X	Search experience and linguistic and conceptual ability.	Unavailability of information.

Code	SS	US	Causation of Success	Causation of Failure
PS27	X	X	Availability of information.	Unavailability of information.
PS28	X	X	Availability information.	Unavailability of information.
PS29	X	X	Search tools efficiency.	Unavailability of information.
PS30	I	X	Search experience.	Unavailability of information.
PS31	I	I	Domain knowledge, linguistic and conceptual ability.	Lack of search skills.
PS32	I	U	Domain knowledge, linguistic and conceptual ability.	Uncertainty about the causes.
PS33	X	X	Availability of information and search tools efficiency.	Unavailability of information.
PS34	B	I	Combination of domain knowledge and search tools efficiency.	Lack of linguistic and conceptual ability
PS35	I	X	Linguistic and conceptual ability and search experience.	Unavailability of information.
PS36	B	B	Combination of availability of information, linguistic and conceptual ability, domain knowledge and search tools efficiency.	Combination of lack of sufficient effort and unavailability of information.
PS37	B	I	Combination of search tools efficiency and linguistic and conceptual ability.	Lack of sufficient effort.

Appendix 8: List of publications & presentations based on the current study

Peer-reviewed papers

- Mansourian, Y. & Ford, N. (2006). "The invisible web: An empirical study of 'cognitive invisibility'". *Journal of Documentation*, **62**(5), 584-596.
- Mansourian, Y. & Ford, N. (2006). "Web searchers' attributions of success and failure: An empirical study". *Journal of Documentation*, (in press).
- Mansourian, Y. (2006). "Adoption of Grounded Theory in LIS research". *New Library World*, **107**(9/10), 386-402.
- Mansourian, Y. & Madden A.D. (2006). "Methodological approaches in the web search research". *Electronic Library*, (in press).
- Mansourian, Y. (2006). "Biologists' search patterns on the web". *Online Information Review*, (in press).
- Mansourian, Y. (2006). "How do researchers explore end users' web search patterns?" *Informology*, (in press).
- Mansourian, Y. (2004). "Technical and non-technical aspects of the invisible web". *Informology*, **1**(2), 221-237.
- Mansourian, Y. (2004). "Similarities and differences between web search procedure and searching in the pre-web information retrieval systems". *Webology* [Online]. **1**(1) http://www.webology.ir/2004/v1n1/a3.html [Accessed 20 April 2005]

Conference proceedings and oral presentations

- Mansourian, Y. (2005). "Identifying effective factors on the level of information visibility for end users on the web environment". In: Ingwersen, P. and Larsen, B. (eds.) *Proceedings of 10th International Conference of the International Society for Scientometrics and Informetrics (ISSI 2005)* **2**, pp. 722. Stockholm: Karolinska University Press.
- Mansourian, Y. (2005), "Enhancement of information seeking efficiency in the electronic resources". In: *The Abstract Book of Computer and Information Technology Group, Proceedings of 13th Iranian Researchers Conference in Europe*, 2nd July 2005. The University of Leeds, Leeds. pp. 21-22.
- Mansourian, Y. (2004). "Searching the Invisible Web: An empirical study of users' interactive searching behaviour in the opaque side of the web environment". In: Dearden, A. & Watts, L. (eds.) *Design for Life: Proceeding of the 18th British HCI Group Annual Conference*, 6-10 September 2004, Leeds Metropolitan University, Leeds. pp. 231-232. Bristol: Research Press International.

Oral Presentation:

- Mansourian, Y. and Webber, S. (2006). "Context in web searching: implications for information literacy". *Librarians' Information Literacy Annual Conference (LILAC2006)*. 27th - 29th March 2006, Leeds University, Leeds.
- Mansourian, Y. (2006). "Exploring the concept of information visibility on the World Wide Web". *Quantitative Methods for Internet Research*. 6th - 8th April 2006, University of Wales, Aberystwyth.
- Mansourian, Y. (2005). "Cognitive and affective aspects of human information seeking behaviours." *The Annual Interdisciplinary Conference for Postgraduate Students in the Social Sciences and Humanities*, 7th April 2005, the University of Sheffield.
- Mansourian, Y. (2004). "End users' perceptions of the invisible web and their conceptualization of information retrieval failure on the web". *Information Studies Seminar Series (2004-2005)*. 11th October 2004, the University of Sheffield.
- Mansourian, Y. (2004). "What is the invisible web?" *The Annual Interdisciplinary Conference for Postgraduate Students in the Social Sciences and the Arts*. 22nd April 2004, the University of Sheffield.
- Mansourian, Y. (2003). "Towards a model of user-oriented aspects of the invisible web". *Information Studies Seminar Series (2003-2004)*. 8th December 2003, the University of Sheffield.

Poster presentations and book reviews

- Mansourian, Y., Ford, N. & Webber, S. (2006), "The hidden territory of the Internet". *Second Annual Poster Competition of the UK GRAD Programme*. 19th May 2006, University of Northumbria, Newcastle.
- Mansourian, Y., Ford N., & Webber, S. (2006). "Searching in nebulous, uncertain or unsatisfactory situation (SINUOUS)". *Research Advisory Panel of the Department of Information Studies*. 19th April 2006, the University of Sheffield.
- Mansourian, Y., Ford N., & Webber, S. (2005). "How people search the Internet: Success and failure on the online world". *First Annual Poster Competition of the UK GRAD Programme*. 26th May 2005, University of Leeds.
- Mansourian, Y., Ford N., & Webber, S. (2005). "Failure ontologies: Conceptualizing unsuccessful searches (FOCUS). *Promoting Information Literacy for Information and Computer Sciences*, 8th June 2005, University College Worcester.
- Mansourian, Y. (2004). "Review of: Spink, A., & Jansen, B. J. Web Search: Public Searching of the Web. Dordrecht, the Netherlands: Kluwer Academic Publishers, 2004". *Webology* [Online], **1**(2) http://www.webology.ir/2004/v1n2/bookreview1.html [Accessed 15 May 2005]
- Mansourian, Y. (2005). "The past, present and future of web search research: An interview with Dr. Amanda Spink". *Webology* [Online]. **2**(2). http://www.webology.ir/2005/v2n2/a15.html [Accessed 18 March 2006]

Persian papers & presentations

- Mansourian, Y. (2006). "Grounded Theory: Inductive theory building based on real data". *Library and Information Science: Quarterly Journal of Central Library and Documentation Centre of Astan Quds Razavi*, (in press).
- Mansourian, Y. (2005). "Development of information retrieval model in the hidden web environment". *Faslname-ye Ketab (Quarterly Journal of the National Library of Iran in Library and Informaiton Science)*, **16**(3), 75-86.
- Mansourian, Y. (2004). "The hidden internet and the buried information resources under the invisible depth of the World Wide Web". *Library and Information Science: Quarterly Journal of Central Library and Documentation Centre of Astan Quds Razavi.* 7(1), 25-42.
- Mansourian, Y. (2004). "What is the invisible web and why it is important". *Nama, the Online Journal of IRANDOC* [Online]. **2**(1). http://www.irandoc.ac.ir/data/E_J/vol2/Invisible_Web.pdf [Accessed 20 January 2005].
- Mansourian, Y. (2003), "A review of the recent user-oriented studies in the web-based information retrieval research". *Library and Information Science, Quarterly Journal of Central Library and Documentation Centre of Astan Quds Razavi.* **6**(1) 1-22.
- Mansourian, Y. (2003), "Effective factors on information seeking and information retrieval procedure in the World Wide Web environment. [Online]. *Ketabdar Electronic Journal* [Online]. **2**(2) http://www.ketabdar.org/magazine/detailarticle.asp?number=23 [Accessed 23 January 2004].
- Mansourian, Y. (2004). "Different aspects of the invisible web". *Iranian Library & Information Science Association Conference on Library Websites: Design and Evaluation*). 14th January 2004. The University of Tehran.
- Mansourian, Y. (2004). "The invisible web". *Seminar of Khorasan Branch of the Iranian Library & Information Science Association.* 8th January 2004, Ferdowsi University of Mashhad.