

Relevance Criteria for Medical Images Applied by Health Care Professionals: a Grounded Theory Study

Shahram Sedghi Ilkhanlar

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

Doctor of Philosophy



Department of Information Studies

April 2008

Abstract

This thesis studies relevance criteria for medical images' applied by health care professionals. The study also looks at the image information needs and image resources used by health care professionals, together with the image seeking behaviour of health care professionals from different disciplines.

The work is a qualitative study that uses the Straussian version of grounded theory. The population of the study included health care professionals from different health and biomedical departments who worked in Sheffield Teaching Hospitals NHS Foundation Trust. In total twenty-nine health care professionals participated in this study and fifteen relevance criteria were identified from the data collected using semi-structured interviews and think-aloud protocols. The work forms part of the medical image retrieval track of ImageCLEF (ImageCLEFMed), and investigated the use of relevance criteria applied to search statements. Analysis indicates that some of the criteria identified by participants could be included in new topics used for future versions of the track.

The findings of the study showed that health care professionals paid more attention to the visual attributes of medical images when selecting images and that they applied topical relevancy as the most frequent and most important criterion. The study found that health care professionals looked for medical images mainly for educational and research purposes and judged the relevancy of medical images based on their pictorial information needs and the image resources they used. We identified the difficulties that health care professionals faced when searching medical images in different image resources. Other findings also highlighted the need for, and the value of, looking at narrower subject communities within health and biomedical sciences for better understanding of relevance judgment and image seeking behaviour of the health care professionals.

Dedication

To my wife, Jila, and my daughter, Sayna

To my mother and my father, and to my sisters and brothers

For all their support and encouragement throughout my life and study

Without which I could never have come this far

Acknowledgement

I would like to begin by expressing my deepest gratitude to all Iranian people, whose precious resources were used through the Ministry of Health and Medical Education and Iran University of Medical Sciences to fund my PhD. I humbly wish that the experience and the knowledge I have gained during my studies in the UK will make a small contribution to the progress of my beloved country, Iran.

Completing doctoral research is a difficult process with so many ups and downs. I am deeply indebted to my erudite supervisors Dr Mark Sanderson and Dr Paul Clough for their kind support and their admirable supervision.

In addition, I must acknowledge the help of the staff of the Research Department of Sheffield Teaching Hospitals NHS Foundation Trust, who willingly cooperated with me during the data collection stage. In particular, I would like to express my appreciation to Dr Nigel Hoggard, Dr Rob Ireland and those members of the health and biomedical departments in Sheffield Teaching Hospitals NHS Foundation Trust who gave up their valuable time and assisted in this study.

My bottomless appreciation also goes friends and scholars who assisted me with their knowledge and advice during my studies, in particular to Professor Nigel Ford, Dr Alireza Howaida, Dr Hamid Reza Jamali Mahmoei, Dr Abulfazl Fateh, Dr Hesam Seyedain and Dr Yazdan Mansourian.

There are many other people who have had an impact on my studies. There are too many to name individually, but I thank them all.

 Table of contents

Abstract	ii
Dedication.....	iii
Acknowledgement	iv
Table of contents	v
List of Figures	ix
List of Tables	x
CHAPTER 1- INTRODUCTION	1
1.1. Problem statement.....	1
1.2. Aims and objectives.....	2
1.3. Research questions.....	3
1.4. Significance of the study.....	3
1.5. Structure of the thesis	5
CHAPTER 2- LITERATURE REVIEW.....	7
Introduction	7
2.1. The concept of relevance.....	7
2.2. Classes of relevance.....	8
2.2.1 Objective or system-oriented.....	8
2.2.2 Subjective or user-oriented.....	9
2.3. Types of relevance	10
2.3.1 System or algorithmic relevance	10
2.3.2 Topical or subject relevance.....	11
2.3.3 Cognitive relevance or pertinence	11
2.3.4 Situational relevance or utility	11
2.3.5 Motivational or affective relevance	11
2.4. Situational relevance	13
2.5. Relevance criteria	14

2.6. Relevance studies	17
2.6.1 Image relevance studies.....	27
2.7. Information needs of health care professionals.....	32
2.7.1 Image information needs of health care professionals.....	38
Summary.....	40
CHAPTER 3- METHODOLOGY	41
Introduction	41
3.1. Quantitative vs. qualitative	41
3.2. Research method adopted in this study	46
3.3. Grounded theory	49
3.4. Qualitative methods and grounded theory in LIS.....	51
3.5. Straussian or Glaserian versions of grounded theory.....	54
3.6. Components of grounded theory.....	56
3.6.1 Data collection in grounded theory: Theoretical sampling.....	57
3.6.2 Data analysis in grounded theory: Comparative analysis	58
3.6.3 Open coding	60
3.6.4 Axial coding	61
3.6.5 Selective coding.....	62
3.6.6 Memos	63
3.6.7 Example coding	64
3.6.8 Schematic of the study.....	67
3.7. Data collection methods.....	69
3.7.1 Interviews.....	70
3.7.2 Think-aloud protocol.....	72
3.8. Data collection procedures.....	74
3.8.1 Prior to the data collection session	74
3.8.2 Research population	74
3.8.3 Sampling and recruiting the participants	75
3.8.4 Ethical issues	78
3.8.5 Time and place	79
3.8.6 During the data collection session	80
3.8.7 After data collection.....	80
3.8.8 Presenting the interview data	81
3.8.9 Data collection protocols	81
3.8.9.1 Interview protocol	82

3.8.9.2 Medical image search with think-aloud.....	82
3.9. Analysing data	83
3.10. Preliminary study	84
3.10.1 The findings of the preliminary study.....	85
3.11. Trustfulness and replicability.....	87
3.12. Limitations.....	88
Summary.....	89
CHAPTER 4- DEMOGRAPHICS OF THE SAMPLE	91
Introduction	91
4.1. Sheffield Teaching Hospitals NHS Foundation Trust.....	91
4.2. Participants' profile	94
CHAPTER 5- RESULTS	97
Introduction	97
5.1. How health care professionals apply relevance criteria	97
5.2. Relevance criteria	114
5.2.1 Visual criteria.....	116
5.2.2 Textual criteria	133
5.2.3 Other criteria.....	140
5.3. Quantitative analysis of relevance criteria.....	153
5.4. An experiment with ImageCLEFMed topics.....	156
5.5. Image information needs.....	159
5.6. Motivation for image searching.....	165
5.7. Medical image resources	168
5.7.1 Web-based resources.....	169
5.7.2 Articles.....	175
5.7.3 Personal collections.....	177
5.7.4 Books	178
5.7.5 Other resources.....	179

Summary.....	179
CHAPTER 6- DISCUSSION AND CONCLUSIONS.....	182
Introduction.....	182
6.1. Main research question: What criteria do health care professionals use to make relevance judgments when searching medical images?.....	182
6.2. Other research questions	185
6.2.1 Research question 1: Are the relevance criteria we identified different from those criteria suggested in the literature?	185
6.2.2 Research question 2: What are the core relevance criteria used for judging the relevance of medical images?	188
6.2.3 Research question 3: Where do health care professionals look for medical images? 190	
6.2.4 Research question 4: What difficulties do health care professionals face when searching medical images?	191
6.3. Conclusions	193
6.4. Further research	196
6.5. Final words	197
Summary.....	198
BIBLIOGRAPHY.....	199
Appendix 1: Ethics approval from the NHS.....	216
Appendix 2: The invitation letter for participation in the study.....	222
Appendix 3: The invitation email for participation in the study.....	223
Appendix 4: The information sheet of the project.....	224
Appendix 5: Reply slip	230
Appendix 6: Consent form.....	231
Appendix 7: Interview protocol.....	232
Appendix 8: Publications.....	235
Appendix 9: A summary of image relevance criteria	236

List of Figures

Figure 2-1: Illustration of types of relevance involved in an information retrieval session for a given sample.....	12
Figure 2-2: Synthesis of common concepts for relevance criteria in literature (Maglaughlin and Sonnenwald, 2002).....	24
Figure 3-1: Quantitative method as illustrated by Bryman (2001: p.63).	42
Figure 3-2: The stages of the research process in the current study.	69
Figure 3-3: The relevance criteria identified in the preliminary study.	86
Figure 4-1: Distribution of roles among health care professionals interviewed	96
Figure 5-1: Google image search retrieved 786 images for the initial query 'myocardial perfusion hibernating' submitted by P16.	99
Figure 5-2: Google image search retrieved 547 images for the second query, 'myocardial perfusion hibernating PET', submitted by P16.....	100
Figure 5-3: Google image search found 119 images for the third query, 'myocardial perfusion hibernating PET F18', submitted by P16.	101
Figure 5-4: Google image search results with the query 'Tool-like receptor 3'	103
Figure 5-5: An image search with the query 'TLR3', Google returned 2,430 images.	104
Figure 5-6: An image search with the query 'TLR-3', Google returned 6,450 images.	104
Figure 5-7: Images obtained for the query 'MIBG scan pheochromocytoma'.....	106
Figure 5-8: Candidate image-1 selected by P18.....	107
Figure 5-9: Candidate image-2 selected by P18.....	108
Figure 5-10: Candidate image-3 selected by P18	109
Figure 5-11: Results obtained for diabetes and retinopathy in Google image search.	111
Figure 5-12: Image selected by P12 for the topic 'diabetes and retinopathy'.....	112
Figure 5-13: General architecture of CBIR systems.....	113
Figure 5-14: Groups and subgroups of relevance criteria we identified.	116
Figure 5-15: IRMA search results	127
Figure 5-16: QBIC query based on colour	131
Figure 5-17: A summary of the criteria mentioned by all of the participants.	154
Figure 5-18: Top 5 most frequently mentioned criteria by participants.....	155
Figure 5-19: Coverage of relevance criteria by the topics of ImageCLEFMed.....	156
Figure 5-20: A sample image from ImageCLEFMed image collection.....	158
Figure 5-21: A comparison between the roles and image seeking motivators of the participants.	168

List of Tables

Table 2-1: Studies of Users' Relevance Criteria.....	16
Table 2-2: Types of document evaluated by participants in the studies investigated by Maglaughlin and Sonnenwald (2002).....	25
Table 2-3: The frequency of the relevance criteria identified by Hung et al. (2005).....	30
Table 2-4: Information sources used by Physicians in Covell et al. (1985)'s study.....	33
Table 2-5: Surgeons' purposes for medical information from Shelstad and Clevenger (1996: p.492).....	36
Table 2-6: Surgeons' source for their information needs (Shelstad and Clevenger, 1996: p.492).....	36
Table 2-7: Ten most common generic questions asked by 103 family doctors. Adapted from Ely et al. (1999: p.360).....	37
Table 2-8: Purposes of medical image retrieval by biomedical professionals (Hersh et al., 2005).....	39
Table 3-1: Some of the differences between qualitative and quantitative methods.....	45
Table 3-2: Differences between Glaserian and Straussian grounded theory.....	55
Table 3-3: Profiles of participants in preliminary study.....	84
Table 3-4: Image resources used by the participants of preliminary study.....	87
Table 4-1: Average number of persons employed in Sheffield Teaching Hospitals NHS Foundation Trust in 2007.....	93
Table 4-2: Profiles of participants.....	95
Table 5-1: Examples of queries used by the health care professionals.....	98
Table 5-2: Relevance criteria employed by participants.....	115
Table 5-3: Dublin Core metadata element set and definitions.....	148
Table 5-4: Visual Resources Association elements set and definitions.....	149
Table 5-5: Addressing medical image relevance criteria by Dublin and Visual Resources Association (VRA) Core Categories.....	150
Table 5-7: Summary of Data Pole and Objects Pole (Fidel, 1997).....	160
Table 5-8: What motivates health care professionals to look for medical images.....	165
Table 5-9: Resources used by health care professionals for medical image retrieval.....	169
Table 6-1: The appearance of criteria in the literature.....	187

CHAPTER 1-INTRODUCTION

Medical images contain useful information and these images are utilised by a variety of users with different levels of subject knowledge in medical schools, departments and research centres. Health care professionals search for medical images expecting to find images relevant to their needs.

The research utilizes a qualitative research method to study the relevance criteria applied by health care professionals to select relevant medical images in their work situation, and to derive their relevance judgment patterns in the context of their real medical image searching. Interviews and think-aloud protocols were used to explore the relevance judgment and the criteria that health care professionals apply when searching for medical images. The techniques for data analysis were based on using 'grounded theory', specifically the Straussian version of grounded theory. The population studied in this research was taken from various medical departments of the Sheffield Teaching Hospitals NHS Foundation Trust.

This chapter gives an overview of the study, defines its aims and objectives, illustrates the research problems encountered, and explains why the study is significant.

1.1. Problem statement

How do end users search documents such as text-based resources, video and photographs? What criteria do end users use to judge whether a document is relevant to their information needs? These questions have been the central concern and topic of much research in information science and a considerable number of criteria of relevance have been identified since 1965 (Schamber, 1994). However, to the best of our knowledge, no previous study among the large body of user-oriented relevance studies focuses on how health care professionals perform real medical image searching when faced with real information needs, and how health care professionals choose medical images that they need.

In several studies, researchers have investigated the relevance criteria applied by different user groups (see chapter 2). The findings of those studies highlighted the importance of the users' information needs in their relevance judgment of the retrieved documents. Thus, as discussed by Hersh (1994), we cannot assume that health care professionals use the same set of relevance criteria that other users apply, for selecting medical images.

The idea for this study came from the importance of relevance studies and understanding the criteria and concept of relevance in the information retrieval field. Any relevance judgment study that focuses on relevance criteria has been described as a step forward by Barry (1994). She stated that the findings of a study focusing on the relevance criteria used by a specific group of users could be valid only for that group, who evaluated specific types of documents relevant to their information needs. She mentioned that if a study focuses on users who evaluated text documents, the results of this study could not be generalized for other groups of users who might have different information needs, or look for different types of documents such as images. She also emphasised that it is impossible to investigate how all imaginable users evaluate various types of documents in different situations in a single study. Therefore, with a number of studies investigating the relevance judgment process of different users in a variety of situations, it will be possible to have an overview of the relevance judgment process (Barry, 1994).

Similarly, Tang and Solomon (1998) supported Barry and suggested that empirical studies of the particular criteria of relevance used by real users for specific situations or information needs is required to find out how users evaluate retrieved documents, and which criteria of relevance were used to mark documents as relevant to their information needs and tasks.

1.2. Aims and objectives

The primary aim of this research is to elicit and document the relevance criteria that health care professionals apply when searching for medical images, and to provide recommendations and strategies that could assist image retrieval research groups improve their systems. More specifically, the objectives of this study are as follows.

Relevance criteria for medical images applied by health care professionals

To identify the relevance criteria that health care professionals apply for selecting medical images relevant to their information needs.

To enrich our understanding of the medical image seeking behaviour of health care professionals, and to investigate the techniques and sources they use to find relevant images.

To identify difficulties that health care professionals might face when they look for images to satisfy their information needs.

1.3. Research questions

The central question of this study is to find out how health care professionals determine which medical images are relevant to their information needs when searching for medical images. Therefore the main research question for this study is:

What relevance criteria do health care professionals use to judge whether a medical image is relevant to their medical image information needs?

Further research questions include the following.

- 1) Are the relevance criteria we identified different from those criteria suggested in the literature?
- 2) What are the core relevance criteria used for judging the relevance of medical images?
- 3) Where do health care professionals look for medical images?
- 4) What difficulties do health care professionals face when searching medical images?

1.4. Significance of the study

Enser et al. (2006) reported that different types of medical images such as x-rays and ultrasound/MRI/CRT scans are widely used by health care professionals in health and biomedical departments. A comprehensive literature review (see chapter 2) revealed that despite of the importance of medical images, too little is known about how health care

professionals assess the relevancy of images and select images according to their information needs in a real work situation. Moreover, it was noted there is currently no known study that focuses on how health care professionals perform real medical image seeking with real information needs. Thus, it can be considered as an urgent need to know more about health care professionals and their information needs, their requirements for medical image retrieval systems, and the criteria of relevance for medical images used by them.

There have been a number of studies (see section 2.7) that discuss how health care professionals currently search medical text-based resources and examine the medical retrieval systems currently used. According to Crystal and Greenberg (2006) health care professionals, especially those working in academic environments, frequently make relevance judgments for their written information needs based on their knowledge of specific authors, the reputation of the information source (journals, books, and databases) and titles. However, no attempts have been made to study the criteria of relevance used by health care professionals when seeking in a medical image database. In addition, due to the considerable number of web-based medical image collections now available, it seems appropriate to begin to explore how these collections are being (or could be) used in medical departments.

As stated earlier, relevance studies to date have explored relevance criteria elicited from the users of traditional information retrieval systems; however, our understanding of relevance criteria for images including medical images has been limited by lack of research in this area. Therefore, it is hoped that the results of this study will be relevant and useful for the people working within the area of study and also expand our academic knowledge of health care professionals' image information needs, their criteria for selection of medical images, and the nature of the relevance judgment process for medical images.

From another point of view, the users of information retrieval systems assess the relevancy of retrieved items in terms of how far they meet their information needs as Borlund (2003a) stated. The information needs of health care professionals in general

have been investigated in several studies; however, to our knowledge no prior study in the literature specifically addresses their visual information needs and potential sources that meet their visual information needs. In this study, we asked health care professionals to describe their information needs for medical image in their own words as Barry (1994) suggested.

Finally our study differs from previous work. Firstly, we did not focus on a particular image collection or image retrieval system. For example, Markkula and Sormunen (2000) studied the relevance criteria applied by journalists using the Aamulehti digital image archive. However, we asked participants to look for images in any sources they preferred. Secondly, we did not ask participants to look for images from any predefined image retrieval task. Hung et al. (2005) asked ten students from the Department of Journalism and Media Studies to look for images in the AccuNet/Associated Press Photo Archive database system for three predefined image searches. In contrast, we asked our participants to look for images they really needed. Thirdly, we did not provide a list of criteria for our participants. Choi and Rasmussen (2002) offered a list of nine widely-used criteria of relevance identified in prior researches to the end users of the American Memory online image collection and asked their participants to assess the relevancy of images to their real information needs using these criteria. By contrast, we asked our participants to describe the criteria and attributes of images which were important for them for the relevance judgment of images.

1.5. Structure of the thesis

The remainder of this thesis comprises five chapters. Chapter 2 provides a literature review, commencing with a discussion of the concept of relevance in information retrieval and then focuses on previous relevance studies.

Chapter 3 presents the epistemological position and the subsequent choice of methodologies and research method selected in this study. The applied methods include the Straussian version of grounded theory with research data collected using

interviewing and think-aloud techniques. In addition, the chapter discusses issues such as the recruiting of participants, ethical issues and limitations of the study.

Chapter 4 presents the demographics of health care professionals who voluntarily participated in this study. It presents the characteristics of the health care professionals such as their speciality and affiliated departments.

Results of this study are discussed in chapter 5. It reports the findings of the study on health care professionals' medical image seeking behaviour, image sources used, criteria for relevant medical images, and attitudes regarding medical image retrieval systems.

Chapter 6 is devoted to the interpretation and discussion of findings from this study. Furthermore, it discusses how far the research meets its aims and its objectives. Chapter 6 also presents the conclusions of the study and summarizes its findings. Some implications for the development of medical image retrieval systems, and suggestions for future studies on relevance judgment, are addressed. Finally, the appendices and a bibliography are included.

CHAPTER 2-LITERATURE REVIEW

Information retrieval is the name for the process or method whereby a prospective user of information is able to convert his need for information into an actual list of citations to documents in storage containing information useful to him. Information retrieval embraces the intellectual aspects of the description of information and its specification for search, and also whatever systems, techniques or machines that are employed to carry out the operation. Quoted from Calvin Mooers (Saracevic, 1995)

Introduction

This chapter reviews the literature related to the scope of the current study: the relevance criteria applied by health care professionals to image retrieval. By focusing on studies that could be helpful for understanding our work, the chapter seeks to put the current study into context, making it possible to compare and relate the results of this study to past work.

The chapter includes two sections. The first section starts with a brief overview of the concept of relevance and then briefly discusses different views of relevance. The section concludes by reviewing the literature relevant to the concept of relevance: relevance judgments and relevance criteria applied by the end users of information retrieval systems. In the second section of the chapter, the information needs of health care professionals in general, and their medical image information needs in particular, are discussed. This section also mentions some related studies on the information seeking behaviour of health care professionals.

2.1. The concept of relevance

Relevance has been a central concept in the information retrieval field since the 1950s. This is hardly surprising, since the purpose of an information retrieval system is to retrieve documents in response to the users' information needs. Thus the concept of relevance has received much attention and has been a core issue for the evaluation of information retrieval systems (Schamber, 1994):

Relevance criteria for medical images applied by health care professionals

The foundations for the use of relevance as a criterion for IR effectiveness can be traced to the beginning of information science in the 1950s and 1960s. Researchers at that time were exploring the potential of computerized information retrieval, even before actual computerized systems were available. (Schamber, 1994: p.9)

Although research into the effectiveness of information retrieval systems is extensive, much of the research was based on a particular notion of relevance: topicality.

Despite its importance, there is no common agreement between information retrieval researchers on a single definition for the concept of relevance. Saracevic (1996: p.13); Anderson (2006: p.6) noted this lack of consensus, and stated that despite its significance, relevance remains one of the least understood concepts in the field of information retrieval.

2.2. Classes of relevance

Despite disagreements over the definition of relevance and the criteria actually used in judging the relevancy of retrieved documents, researchers such as Anderson (2006); Barry (1994); Borlund (2003a); Choi and Rasmussen (2002); Hersh (1994); Ingwersen and Järvelin (2005); Kim (2006); Maglaughlin and Sonnenwald (2002); Mizzaro (1997); Müller et al. (2006); Park (1993) are in general agreement that two classes of relevance exist and can be distinguished: objective or system-oriented relevance; and subjective or user-oriented relevance. The two classes correspond to an understanding of relevance as employed by objective evaluation of information retrieval systems, and by cognitive user-oriented relevance studies, respectively.

2.2.1 Objective or system-oriented

The earlier presentations of relevance can be defined as objective or system-oriented. This class has been derived from information retrieval systems evaluation (such as the Cranfield tests conducted by Cyril Cleverdon in artificial test situations, which began in the 1950s (Cleverdon , 1967)). The system-oriented approach considers relevance as a static and objective process, as opposed to the user-oriented approach that treats

relevance as a subjective individualized mental experience that involves cognitive restructuring (Swanson,1986).

In this view, a document is relevant to a query if some or all of its theme overlaps the query. Green (1995); Barry (1994) believed that this approach focuses on the evaluation of internal mechanisms that information retrieval systems employ to match the terms of query (topics) with the terms that were allocated to documents. With regard to the definition of topical relevancy, information retrieval systems retrieve documents which are topically relevant to the queries. Therefore all documents retrieved by the system could, by definition, be considered as potential documents relevant to the query. Barry (1994: pp. 149-150); Borlund (2003a); Green (1995) reported that this class of relevance has been used for the evaluation of traditional textual information such as bibliographic retrieval systems, and noted it refers to the relationship between the (textual) content of retrieved documents and query representation. Therefore in this class, relevance is a context-free concept, associated with the 'aboutness' of a document.

Barry (1994) also raised another point. She believed this class ignores the end users of information retrieval systems. Barry (1994) explained that in real-life situations users judge the relevancy of retrieved documents in relation to various criteria such as helpfulness and how effectively the documents meet their information needs. Park (1994) supported Barry's ideas and stated that this class of relevance is context-free and is based on fixed assumptions that ignore an individual's particular context and state of need. Saracevic (1996) notes however that dissatisfaction with the inappropriateness, inadequacy and ambiguousness of the system-oriented approach has generated much criticism. More recent research on relevance has thus focused on finding out how users describe relevance: a user-oriented approach to relevance.

2.2.2 Subjective or user-oriented

In recent years, the concept of relevance has been regarded as subjective in nature rather than as being objective (see Borlund, 2003a; Schamber, 1994; Mizzaro, 1997; Saracevic, 1996). This view of relevance promotes the importance of human subjectivity and cognition in making relevance judgments. In this approach, relevance is not considered

as associated solely with the topicality of a document, but is seen as a relationship between users' individual information needs and the documents. This approach pays attention to the relevance assessment of the users of information retrieval systems. Tang and Solomon (1998) stated that in the user-oriented approach, the relevance judgment could be described as an outcome of personal perception related directly to contextual factors such as subject knowledge. Therefore, in this approach users apply a non-binary, subjective and dynamic type of relevance to evaluate the outcome of the retrieval process. Consequently, this type of relevance assessment is multidimensional and does not refer to a single relevance criterion as Borlund (2000) stated, but rather refers to criteria such as quality of information, helpfulness, and utility of retrieved documents in relation to covering the information needs, interests, tasks and situations of users.

2.3. Types of relevance

Based on the two main classes, different types of relevance have been distinguished in the information retrieval literature. For example, Saracevic (1996) identified five types of relevance, which have been widely accepted and quoted (Ingwersen and Järvelin, 2005; Hersh, 2003; Borlund, 2003a; Maglaughlin and Sonnenwald, 2002; Tang and Solomon, 1998; Mizzaro, 1997). The next section discusses in detail the five types of relevance widely accepted in the literature.

2.3.1 System or algorithmic relevance

This type of relevance refers to the relationship between a query and a document in the file of a system as retrieved or missed, by a given procedure or algorithm (Saracevic, 1996). Information retrieval systems use methods to organize and present documents that the system inferred are relevant to the query. For example, search engines rank documents with regard to the relationship between query and retrieved items. According to Ingwersen and Järvelin (2005) the algorithmic type of relevance is the most widely-used type of relevance for traditional evaluation of information retrieval systems, and is the clearest type of relevance in terms of definition.

2.3.2 Topical or subject relevance

This type of relevance refers to the relationship between the 'aboutness' of retrieved documents and the subject of interest as perceived by the user (Saracevic, 1996). Hersh (1994) argued that topical relevance judgments are based on an assumption that both queries and documents are about the same topic. Therefore, this approach ignores the user's state of knowledge of during the judgment process.

2.3.3 Cognitive relevance or pertinence

This type of relevance refers to how documents affect the current knowledge of users. Saracevic (1996) described cognitive relevance as the cognitive correspondence between the 'informativeness' of retrieved documents and the background knowledge of users. Ingwersen and Järvelin (2005) described this type of relevance as the relationship between the nature of documents and the information need as perceived by the user at a given point in time. For example, the user might consider the novelty of a document to his/her information needs at a given point in time.

2.3.4 Situational relevance or utility

This type of relevance refers to the relationship between the user's work task, daily life situation or problem at hand underlying the user's information needs on the one hand, and retrieved documents on the other (Borlund, 2003b; Ingwersen and Järvelin, 2005; Salton, 1992). Borlund (2003b) notes a key difference between situational and cognitive relevance: situational relevance is based on the pragmatic utility of retrieved documents to solve users' immediate real-world problems, whereas in cognitive relevance users may search for information just to fulfil their curiosity.

2.3.5 Motivational or affective relevance

Motivational relevance refers to the user's emotional reaction to documents, and whether they fulfil users' intents, goals, and motivation. In fact interests, goals and motivation prompt users to use an information retrieval system and evaluate the relevance of retrieved items (Saracevic, 1996; Borlund, 2003a). Saracevic (1996)

believed that criteria such as satisfaction, success and accomplishment reflect this type of relevance.

Among the five types of relevance, topical relevance is perhaps the most widely used when judging the output or success of information systems. For example, using the topical type of relevance, a document is relevant to an information need (expressed as a query) if some or all of its theme overlaps the query. Cleverdon (1967) and Cleverdon and Mills (1963) noted that measuring the effectiveness of information retrieval systems depends on two factors: the first is the number of relevant documents retrieved as against the total number of relevant items in the collection (recall), and the second is the percentage of relevant documents among those actually retrieved (precision). Although precision and recall are the preferred pair of evaluation measures for performance of information retrieval systems, and both of them are based on the relevance judgment of the output of information retrieval systems (Cleverdon, 1967 and Cleverdon and Mills, 1963), these measures have also been severely criticised (Saracevic, 1996).

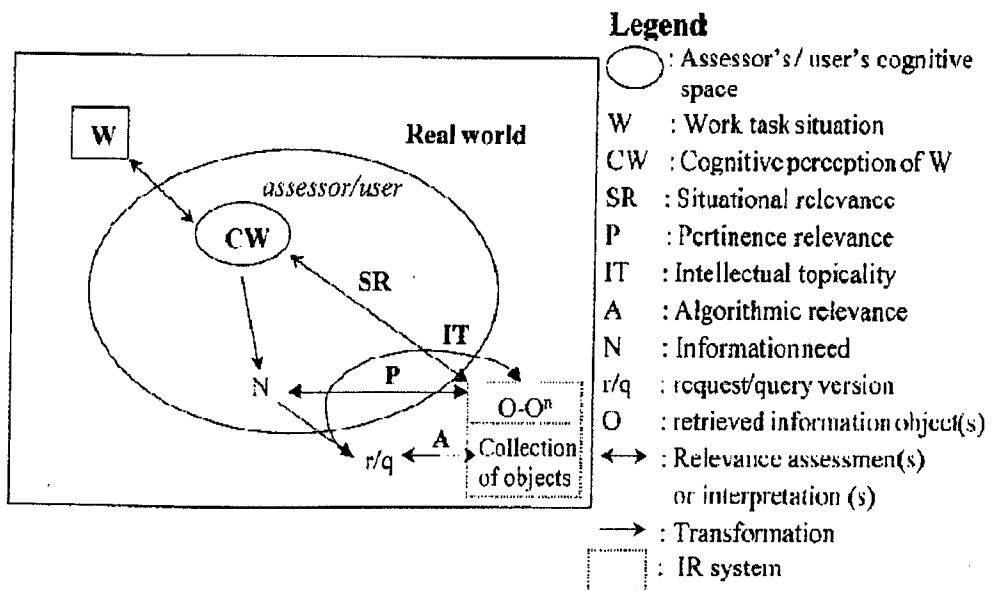


Figure 2-1: Illustration of types of relevance involved in an information retrieval session for a given sample
(Adapted from Borlund (2003a: p.915)).

Borlund (2003a) illustrates the main types of relevance, including situational relevance, involved in information retrieval during a session of relevance judgment for a given case (Figure 2-1).

In Figure 2-1, query (r) operates as an alteration and indication of the user's current information needs (N). However, situational relevance (SR) represents the subjective relevance (IT) and pertinence relevance (P). Situational relevance (SR) does not relate directly to topical (algorithmic) relevance (A). Borlund (2003a) also emphasized that the subjective relevance refers to the relationship between the query (Q), the information needs (N), or work task situation (W), as explained by the user (CW), and the retrieved items.

2.4. Situational relevance

Topical relevance is an important type of relevance, which helps system designers examine and measure the performance of information retrieval systems, but it cannot be used to investigate how systems respond to users' situational information needs in users' real-life situations. For example, topical relevance disregards the users' knowledge and their real-life information needs as previously discussed; in practice, user's knowledge plays an important role in the judgment process.

As mentioned earlier, Schamber et al. (1990) re-introduced the concept of situational relevance and stressed the importance of context and situation in information retrieval. Schamber et al. (1990: p.763) suggested that information retrieval researchers must focus on the perceptions of end users in real information-need situations, rather than on subject experts or search intermediaries, or on any judges in artificial test situations. Borlund (2003a) stated that in response to the call by Schamber et al. (1990), special attention has been paid to the situational type of relevance, and this type of relevance has been investigated in a number of studies (e.g. Park, 1993; Barry, 1994; Hersh, 1994; Saracevic,1996; Borlund and Ingwersen, 1998; Mizzaro,1998; Tang and Solomon, 1998; Xu and Chen, 2005).

The findings of the relevance studies mentioned above support the ideas of Schamber et al. (1990) with respect to the importance of context and situation in information retrieval, and the dynamic nature of relevance. For example, based on the findings of an empirical study, Park (1994) reported that user-based relevance is situational and involves an individual's interpretation and complex mental processes that go beyond simple topical relevance. Furthermore, Barry (1994) explained that the topical relevance of retrieved objects does not automatically mean that these documents provide the information that users need; users look for documents that have properties beyond topicality. Barry (1994) argued that situational relevance is a dynamic concept and is dependent on the perceptions and information needs of a user. Thus, this type of relevance is inferred by criteria such as the helpfulness of information in decision-making, and the usefulness of information for solving a problem. Barry (1994) described situational relevance as the relationship between an individual and information objects.

Similarly, Hersh (1994) stated that documents topically relevant to the query of the user cannot be described as relevant documents, and that there might be factors such as the information needs and knowledge of the user that affect the relevance judgment process. He added that a document can be marked relevant if it provides information that the user really needs.

What can be inferred from more recent relevance studies is that users' relevance judgment does not relate to the topical relevance between the query and document, but as Borlund (2003a) illustrated relates to judgment of retrieved items (presented as O-Oⁿ in Figure 2-1) according to the cognitive perception of the users' work-task situation.

2.5. Relevance criteria

Research has shown that relevance is subjective and depends on many different factors, several of which have been identified in studies by Schamber (1991); Park (1993); Barry (1994); Tang and Solomon (1998); Markkula and Sormunen (2000); Hirsh (1999). For example, Schamber (1994: p.11) reported that relevance must be studied beyond topicality. Schamber (1994) suggested that information retrieval researchers should

extend relevance studies into the realm of what is described as 'mental states of the user'.

The notion of IR interaction processes (including relevance assessment) should not be bounded at points of requests and documents, but rather should extend into the realm of mental states of the user. One concern that has been implicit throughout the discussion of relevance is IR system design. Various authors have indicated that a surprising number of subjective user criteria can be implemented in IR systems and services. This includes tangible document elements that users have said are important or that provide users with clues to relevance. (Schamber, 1994: p.35)

In a comprehensive overview of relevance, Schamber (1994) analysed related literature on relevance dating from 1960, and produced a list of eighty criteria which are likely to influence relevance judgements. She classified these into six groups: attributes of the person making the relevance assessment (e.g. knowledge and experience), queries or topics, documents, the information retrieval system, judgment conditions and choice of scale. She also believed the process of judging relevance to be a dynamic phenomenon and based on several criteria such as informativeness of documents, personal knowledge of the end users of information retrieval systems, and credibility of the source of documents.

Mizzaro (1997) analysed 157 papers published since 1959 and classified them within three periods: 'before 1958', '1959–1976' and '1977–present'. He analysed papers within each time period with regard to the following aspects: methodological foundation, kinds of relevance, beyond-topical criteria adopted by users, modes for expressing the relevance judgment, dynamic nature of relevance, type of document representation and agreement among different judges. Mizzaro (1997) concluded that the focus of papers published in the '1959–1976' period had been on relevance inherent in the document and query (topical relevance). However in the '1977–present' period,

researchers have attempted to understand, formalize, and measure a more subjective, dynamic and multidimensional relevance judgment (situational relevance).

Table 2-1: Studies of Users' Relevance Criteria.

Research By	Criteria or group of criteria
Schamber (1991)	Accuracy, Currency, Specificity, Geographic Proximity, Reliability, Accessibility, Verifiability, Clarity, Dynamism, Presentation quality
Barry (1994)	Depth (scope), Objective accuracy (validity), Tangibility, Effectiveness, Clarity, Recency, Background/experience, Ability to understand, Content novelty, Source novelty, Stimulus document novelty, Subjective accuracy (validity), Effectiveness, Consensus, External verification, Availability within the environment, Personal availability, Source quality, Source reputation (visibility), Obtainability, Cost, Time constraints, Relationship with author
Park (1994)	Internal context refer to user's background and experiences External context includes factors that stem from current search Problem (content) context, includes the motivations underlying the intended uses of a document
Tang and Solomon (1998)	Topical relatedness, Types of articles, Similar topical focus, Duplicates, Recency, Length, Depth (Breadth), Language, Geographical focus, Version of article (repetitiveness)
Wang and White (1999)	Topicality, Orientation/level, Discipline, Novelty, Expected quality, Recency, Reading time, Availability, Special requisite, Authority, Relation/origin, Cognitive requisite, Actual quality, Depth, Classic/founder, Publicity, Reputation, Prolific author, Journal spectrum, Peer review, Standard reference, Judge, Norm, Target journal, Credential
Hirsh (1999)	Authority, Convenience (accessibility), Interesting, Language, Novelty, Peer interest, Quality, Recency (temporal issues), Topicality
Markkula and Sormunen (2000)	Topicality, Technical, Contextual attributes and Visual attributes
Choi and Rasmussen (2002)	Topicality, Accuracy, Time frame, Suggestiveness, Novelty, Accessibility, Completeness, Appeal of information, Technical attributes of images
Maglaughlin and Sonnenwald (2002)	Citability, Informativeness, Author novelty, Discipline, Institutional affiliation, Perceived status, Accuracy-validity, Background, Content novelty, Contrast, Depth-scope, Domain, Citations, Links to other information, Relevant to other interests, Rarity, Subject matter, Thought catalyst, Audience, Document Novelty, Type, Possible content, Utility, Recency, Journal novelty, Main Focus, Perceived quality, Competition, Time requirements
Xu and Chen (2005)	Topicality, Reliability, Scope, Understand ability, Novelty, Relevance, Prior knowledge
Hung, Zoeller & Lyon 2005	Typicality, Emotion, Action, Aesthetic, Text, Familiarity, Context, Impression, Preference, Posture, Facial feature, and Appearance.

The aim of studies by Hirsh (1999); Maglaughlin and Sonnenwald (2002); Tang and Solomon (1998); Choi and Rasmussen (2002); Park (1994); Xu and Chen (2005); Barry (1994); Schamber (1991) was to identify criteria of relevance used by information retrieval systems users to judge documents against their information need: (Table 2-1 outlines the findings of some major empirical works). However, each study yielded criteria that were similar to criteria in other studies as Maglaughlin and Sonnenwald

(2002) stated. Maglaughlin and Sonnenwald summarized the findings of relevance studies in information retrieval and concluded that for textual documents, such as books and articles, there was a significant overlap between criteria identified by various studies, and therefore the information retrieval field seems to be reaching a consensus about which criteria are used in making relevance judgments of textual documents. Xu and Chen (2005) supported this and suggested that some criteria have overlapping meaning (e.g. novelty and recentness).

Barry (1994) suggested that the first step for researchers who want to study relevance criteria applied by different user groups within a range of situations would be to combine of the results of previous studies which have gathered relevance criteria, and variables which affect the relevance judgment process directly from users. The next sections now discuss the findings of some experimental studies on the concept of relevance.

2.6. Relevance studies

Borlund (2003a) reported that two rounds of relevance studies can be distinguished in information retrieval literature: the first round of studies can be traced back to the 1950s and 1960s. Borlund (2003a) reported that since the ASTIA¹ and Cranfield Uniterm tests carried out by Gull (1956) and Cleverdon (1960), the debate concerning the concept of relevance has become an important part of discussions in the information retrieval field. For instance, papers published by Cuadra and Katter (1967) , Rees and Schultz (1967) and Saracevic (1975) improved understanding of relevance and described how relevance was employed in the past.

The interest in relevance declined after the 1960s until the second round of studies began in the 1990s. Borlund (2003a) believed that the paper published by Schamber et al. (1990) revived and intensified the discussions on the concept of relevance in the information retrieval field. (e.g. Mizzaro, 1997; Ingwersen, 1992; Barry, 1994; Hersh,

¹ Armed Services Technical Information Agency

1994; Schamber, 1994; Saracevic, 1996; Park, 1994). We discuss related studies in this section.

Few studies have actually addressed the factors that influence relevance judgment (Hersh, 2003: p.106). Among them, two empirical studies begun in 1965 are regarded as pioneering studies which attempted to investigate the factors which influence the judgment process: Rees and Schultz, 1967; Cuadra and Katter, 1967.

Cuadra and Katter (1967) asked users to evaluate the relevancy of retrieved abstracts to the presented request. Hersh (2003) reported on Cuadra and Katter (1967)'s work, and stated that they found thirty-eight criteria such as style, specificity, and level of difficulty of documents that affect the human relevance judgment. They classified those criteria into five groups: (1) Type of document being judged, including its subject matter and level of difficulty; (2) Query or topic which expressed the information needs; (3) Judge (subject knowledge of judge and his or her familiarity with the subject of document); (4) Judgment conditions such as time available, order of presentation, number of documents in the document set; and (5) Judgment mode.

The findings of Cuadra and Katter (1967) indicated that criteria which relate to 'judge' and 'judgment mode' strongly influenced the relevance judgment process. It has also been suggested by Cuadra and Katter (1967) that a major priority for future research was the development of models of users and situations.

Schamber et al. (1990: p.763) quoted Rees and Schultz (1967) and stated that the researchers studied the qualities of 'judges' engaged in various stages of biomedical research. Rees and Schultz (1967) identified more than forty variables including primary variables (e.g. research stage, judgment group, document set, document representations), and secondary variables such as education, professional experience and research experience. The results of the study by Rees and Schultz (1967) indicated that relevance judgments had depended on the subject knowledge of judges, the ranking of documents, judgmental variations and the research stage of judges. Although, the criteria that Rees and Schultz (1967) examined were similar to the criteria identified by Cuadra

and Katter (1967), they examined a new factor which they defined as 'research stage of users'.

The studies by Cuadra and Katter (1967) and Rees and Schultz (1967) share a number of qualities that make them significant to our current knowledge of relevance. Both studies produced findings about the relevance criteria that have become widely accepted. They showed that relevance judgement depends on many non-topical criteria including the type of document, the way the information needs are expressed, characteristics of the judges, the mode for expressing the judgment and the situation in which the judgment is carried out. Unfortunately, while the two studies stand unreplicated, both studies examined judgments by experts rather than actual users.

Schamber et al. (1990: p.774) examined the papers published since the 1960s and made three conclusions about the nature of relevance and its role in information behaviour:

- 1) Relevance is a multidimensional cognitive concept and its meaning is largely dependent on searchers' perceptions of information and their information needs. Relevance judgment has multidimensional characteristics.
- 2) Relevance is a dynamic concept, because judgment of documents may change over time. It depends on users' decisions about the relationship between the documents and their information needs at a certain point in time.
- 3) Relevance is a complex but systematic and measurable concept if approached conceptually and operationally from the users' perspective.

Borlund (2003a: p.922) believed that Chamber et al. (1990)'s conclusions stressed the importance of context and situation in information retrieval and brought dimensions and dynamism into relevance. 'Context' may come from the documents or knowledge sources in systems, but may also be part of the user's real information-seeking situation. 'Situation' also involves a set of dynamic cognitive conditions in the mind of the user while she or he is searching for information. Essentially, it implies that if relevance is dynamic, the corresponding information need is dynamic as well and vice versa. A

document may be perceived to be topically relevant, but may not be judged useful for a particular situation the user is dealing with at that time. It is obvious that only the user can conduct this type of 'subjective' relevance. In a cognitive sense, it is linked to a work-related or daily-life task or event placed in a particular context.

A year later, and in her empirical research, Schamber (1991) used an open-ended questionnaire and interviewed thirty professional users of weather information employed in aviation, electric power utilities, and construction fields (ten in each field). She asked users to describe one recent task situation in which they needed information about the weather and discuss how they evaluated weather information that they obtained from seven different sources including themselves, other people, weather information systems, television, radio, newspapers, and weather instruments. Schamber (1991)'s study was different from other relevance studies, since she investigated the users' criteria of relevance for evaluation of obtained information and not retrieved documents. She identified twenty-two criteria, in ten groups including accuracy, currency, specificity, geographic proximity, reliability, accessibility, verifiability, clarity, dynamism and presentation quality. Schamber (1991) elicited two new criteria that had not appeared in the relevance literature. These new criteria were geographic proximity and dynamism. Schamber defined geographic proximity as information which covered geographic area, location or altitude. She also defined dynamism as the presentation of information (e.g. a display with tracking or zoom capabilities). Therefore, it can be said that users of specific types of information may apply certain criteria of relevance such as geographic proximity that were generated by them. Based on the results of her study, Schamber (1991) noted that four out of the ten categories had been applied more than others. These categories of relevance criteria were: presentation quality; verifiability; geographic proximity; and dynamism. Schamber (1991) also stated that users applied different criteria for evaluating the source of obtained weather information. For example, reliability was applied more than any other criteria in evaluating the information obtained from another person. Users also had stated that currency is the most important criteria for evaluating the results of retrieval using a weather information system. Schamber et al. (1990: p.763) criticized Cuadra and Katter (1967); Rees and

Schultz (1967) and declared that they failed to settle on a definition of relevance; however, Schamber (1991)'s efforts also did not result in a definition of the concept of relevance.

A study by Cool et al. (1993) explored the factors which affect the judgments of the relevance or usefulness of documents to particular information problems. They studied two different groups of participants using two different techniques. In the first group, they asked 300 college freshmen to fill in a questionnaire and write brief explanations concerning their decision to use or not use documents for a research paper. For the second group of participants, they used interviews and asked eleven humanities scholars about their information needs and the documents they used to meet those needs. Their results indicated that six categories of relevance criteria were used by these populations: topic (how a document relates to a person's interests), content/information (characteristics of what was in the document itself), format (formal characteristics of the document), presentation (how a document was written/ presented), values (dimensions of judgment), and oneself (relationship between a person's situation and other categories). Cool et al. (1993: p.81) compared their findings from two groups and stated that: firstly, for both groups of people, factors beyond topical relevance were important to decide whether to use a document or not. Indeed, there was a great deal of overlap between the two groups in the types of criteria that were important. Secondly, there were also differences between the two groups in the type of decisions that were made, and the criteria that were applied to support them. This revealed that the nature of the user's situation was significant in using the criteria of relevance. Thirdly, the relationship between a person's situation and relevance criteria may be related to the goals of the user and the problems that users are trying to address.

In an empirical study, Park (1994) investigated how university students and faculty members judge the relevancy of scholar citations to their information needs. Park reported that three major categories of factors that affect the relevance judgments emerged from data. The first category, internal context, reflects the participants' interpretation of a citation based on their own prior experiences or perceptions. The second category, external context, represents criteria that originated in the individual's

search and current information needs. The third category, problem (content) context, examines the motivations underlying the potential uses of a citation. Her findings also showed that some of the criteria mentioned by participants included the subject matter indicated by the title, readability, author's status, quality of publication and type of document. Park (1994: p.140) concluded that the psychological, situational, and interpretative nature of information needs and relevance is not controllable, nor can it be isolated from the individual and social contexts in which human beings operate.

Barry (1994) provided additional insight into relevance and conducted an exploratory study to identify criteria of relevance using eighteen participants within an academic environment. This study yielded twenty-three criteria: depth/scope, objective accuracy/validity, tangibility, effectiveness, clarity, recency, background/experience, ability to understand, content novelty, source novelty, stimulus document novelty, subjective accuracy/validity, effectiveness, consensus, external verification, availability within the environment, personal availability, source quality, source reputation/visibility, obtainability, cost, time constraints and relationship with author. Barry (1994)'s findings supported the argument that situational factors beyond the inherent topical content of documents influence the relevance judgement process. She reported that the situation included any factors that users bring to the situation (such as skill, education, knowledge level, views and individual preferences) and these assessments of retrieved documents occur within the larger context of the information environment. Barry (1994) also noted that each participant did not possess a unique set of criteria by which information is judged. However, some relevance criteria were shared across participants and situations. She also stated that there had been a great deal of overlap in the criteria identified by her study and studies by Schamber (1991); Park (1994). Though Barry (1994) stated that results of a study such as hers could not be generalized to other groups of users, examinations of different types of information, or other information need situations, she stressed the importance of the findings of previous studies on relevance judgment. She believed that the results of previous studies have some implications: firstly, there are factors beyond topical relevance that affect the relevance judgment process; secondly, there is an obvious overlap in the criteria of relevance that had been recognized in

previous studies; and thirdly, users' of information retrieval systems can recognize and describe non-topical factors that influence their relevance judgment.

Maglaughlin and Sonnenwald (2002) examined empirically the relevance criteria applied by twelve graduate students with real information needs to assess the twenty most recent documents retrieved in response to their information needs using a three-partite scale (relevant, partially relevant, and not-relevant). They identified twenty-nine criteria including: citability, informativeness, author novelty, discipline, institutional affiliation, perceived status, accuracy-validity, background, content novelty, contrast, depth-scope, domain, citations, links to other information, relevant to other interests, rarity, subject matter, thought catalyst, audience, document novelty, type, possible content, utility, recency, journal novelty, main focus, perceived quality, competition and time requirements. They classified the criteria in six categories: abstract, author, content, document, journal or publisher and participant. Among them, the criteria in the content category (content novelty, contrast, depth-scope, domain, citations, links to other information, relevant to other interests and rarity) were the most frequently-used criteria and were mentioned more than the combination of all other criteria.

In addition to these criteria, Maglaughlin and Sonnenwald (2002: p.334) studied the positive and negative value of each criterion for the relevance judgment process. Maglaughlin and Sonnenwald reported that criterion such as citability was always considered as a positive criterion for the judgment of documents; however, most criteria were discussed both negatively and positively with several exceptions. The authors also suggested that users may spend more time examining useful documents carefully, or may find it easier to discuss positive associations between their information needs and documents. The authors also compared the criteria they identified with those suggested in the studies by Schamber (1991); Park (1993); Cool et al. (1993); Tang and Solomon (1998); Barry (1994); Wang (1994); Schamber and Bateman (1996); Bateman (1998a); Tang (1999). They also reported that there has been substantial overlap of criteria (the average number of studies in which a particular criterion is identified was 6.8 (see Figure 2-2). Maglaughlin and Sonnenwald (2002) believe that the more frequently a

criterion is identified, the more likely the criterion is to be applicable across document domains and situations.

Category	Criteria concepts	Scharber and Tang and Baerman Spink Tang Maglaughlin and Sonnenwald (1998)(1998)(2002)				Total times identified	Average per category		
		Scharber (1991)	Part (1992)	Cool et al. Barry (1993) 1994	Wang (1994)				
Author Content/ topics/ aboutness	Credibility/status	N/A	✓	✓	✓	✓	7	7.0	
	Subject matter/topics*	✓	✓	✓	✓	✓	11		
	Breadth/completeness/depth/breadth/ scope/specificity	✓	✓	✓	✓	✓	10		
	Accuracy/credibility/quality/ validity/verifiability	✓	✓	✓	✓	✓	9		
	Clarity/presentation quality/readability/ understandability	✓	✓	✓	✓	✓	7		
	Novelness/ information connections/links to other information	✓	✓	✓	✓	✓	6		6.8
	Background information	✓	✓	✓	✓	✓	6		
	Methodological information	✓	✓	✓	✓	✓	6		
	Stimulus/thought catalyst/ Geography focus/proximity	✓	✓	✓	✓	✓	4		
	Full text	Current/recent/timeliness	✓	✓	✓	✓	✓		10
Document/article type		✓	✓	✓	✓	✓	7		
Availability/accessibility/ Obtainability		✓	✓	✓	✓	✓	6		
Novelty		✓	✓	✓	✓	✓	5		
Utility		✓	✓	✓	✓	✓	4		
Authority/quality/reliability/ reputation/value/visibility		✓	✓	✓	✓	✓	10	7.5	
Journal/ publisher/ source		✓	✓	✓	✓	✓	5		
Novelty		✓	✓	✓	✓	✓	9		
Affectiveness/appeal/competition		✓	✓	✓	✓	✓	8		
Oneself/ participant/ user		Bedside/performance/understanding	✓	✓	✓	✓	✓	9	7.0
	Time constraints/requirements	✓	✓	✓	✓	✓	4		

* Category often assumed but not studied directly.

Figure 2-2: Synthesis of common concepts for relevance criteria in literature (Maglaughlin and Sonnenwald, 2002).

As shown on the previous page, Figure 2-2 gives a synthesis of the common concepts for relevance criteria in the literature.

Table 2-2: Types of document evaluated by participants in the studies investigated by Maglaughlin and Sonnenwald (2002).

	Schamber (1991)	Park (1992)	Cool et al. (1993)	Barry (1993), Barry (1994)	Wang (1994)	Schamber and Bateman (1996)
Type of document users evaluated	Varied	Textual documents (bibliographic information)	Textual documents (full text)	Textual documents (full text and bibliographic information)	Textual documents (full text and bibliographic information)	Textual documents (full text and bibliographic information)

	Tang and Solomon (1998)	Bateman (1998a); Bateman (1998b)	Spink et al. (1998)	Tang (1999)	Maglaughlin and Sonnenwald (2002)
Type of document users evaluated	Textual documents (bibliographic information)	Textual documents (full text and bibliographic information)	Textual documents (bibliographic information)	Textual documents (bibliographic information)	Textual documents (bibliographic information)

Although Maglaughlin and Sonnenwald (2002) did not explain why there has been a great deal of overlap in terms of relevance criteria in those empirical studies, they noted that in the studies presented in Figure 2-2, researchers investigated the relevance criteria applied for traditional textual documents (see Table 2-2). Textual documents share a number of textual features such as title, author, and publication date. These textual features have generally been considered as relevance criteria for judgments, or criteria have originated from these features, as Barry (1994: p.151) stated.

Based on the results of the study by Maglaughlin and Sonnenwald (2002), and the types of document judged in the studies they investigated Figure 2-2, we believe that the more shared features there are across documents investigated in relevance studies, the more likely it is that the criteria are applicable across situations. This assumption supports the idea that the document itself is a central variable in the judgment process (Barry, 1994: p. 152), and supports the feasibility of research into the shared criteria employed by

users as a possible means of incorporating such criteria into the retrieval process. Further investigation is required, however, to examine whether there is a relationship between the type of document and the relevance criteria identified in relevance studies.

We can draw the following conclusions about relevance studies in information retrieval literature. In the earlier studies, relevance was regarded as a measure for the evaluation of the information retrieval process for traditional textual documents, regardless of the person who made the judgment. However, since the 1990s relevance studies have become user-oriented. The findings of the earlier studies suggested that the degree of document relevance changes across documents and users. For example, two end users of an information retrieval system do not produce two similar search queries, yet both judge the different sets of documents as relevant. In the most recent studies, the concept of relevance has been regarded as having moved beyond topicality in relation to information retrieval and the information seeking behaviour of the user.

It is thus very evident from the literature that relevance judgment is based on criteria beyond topicality. In real-life situations, users are concerned with the usefulness of sought objects for their information needs situation. However, we may not know exactly which criteria, or which specific combination of criteria, determine the situational relevance of a document for a user's information needs in a particular situation. Studies such as Maglaughlin and Sonnenwald (2002) found evidence of a great deal of overlap between the criteria applied by users in several relevance studies (e.g. studies listed in Figure 2-2). This indicates the centrality of users in relevance studies, which might be explained by the fact that users have the ability to determine whether or not particular criteria exist for a given document. On the other hand, the document (and its features) plays an important role in determining its relevance. Traditional textual documents share a range of tangible textual features such as title, publisher, publication date and author. As mentioned earlier, situational relevance is different and is based on criteria beyond topicality. Through studying the features that indicate to users whether those criteria are present or not, we may be able to take information retrieval beyond the topical approach. That means it may be possible to improve information retrieval systems by attempting to

incorporate features of documents that users employ to detect criteria beyond topical relevancy.

2.6.1 Image relevance studies

There are few studies which investigate relevance judgments for visually-orientated documents.

Markkula and Sormunen (2000) investigated the relevance criteria typically applied by journalists when selecting images for tasks in realistic work situations. The authors interviewed eight journalists who were given twenty illustration tasks based on searching for images in the Aamulehti digital image archive, a collection containing over 83,000 photographs.

Based on the results of this preliminary study, Markkula and Sormunen (2000) classified relevance criteria that journalists applied for relevance judgment of images into four groups. These were: topicality, technical, contextual attributes and visual attributes (unfortunately the authors did not make clear the total number of relevance criteria they identified, nor how many criteria existed in each group, though they did mention criteria such as cost of images, recency of images and layout of pages). The first group of criteria the journalists employed was 'topical' and they used captions to assess the topical relevancy of images to obtain information about a relevant image and its background. 'Technical' and 'contextual attributes' were the second and third group of relevance criteria used, with most journalists preferring to find images which were technically good, rather than recently published and current (Markkula and Sormunen, 2000: p.277). Further important factors included the financial cost of an image, recency or freshness of images.

The journalists also had paid much attention to the fourth group of criteria, 'visual attributes'. For example, sometimes journalists required images of a particular type such as a passport photograph of a specific person. In addition, journalists would often use the message they wanted to convey through an image as relevance criteria (e.g. dramatic, surprising, effective, shocking, funny, expressive or threatening).

Markkula and Sormunen (2000: p.277) mentioned that topicality was the first group of criteria applied at the beginning of the image search; criteria related to the visual attributes of images were always used at the final stage of image selection. At this point candidate images were topically, technically and contextually relevant. However, the selection of images for publication in the newspaper was based solely on their visual attributes and aesthetic attributes such as colour and composition. In addition to these four groups of relevance criteria, some other factors affected the decision of journalists. Examples of these criteria were the article, the layout of the page, the section and its illustrative style, its editorial policy and the ethical rules journalists follow. Markkula and Sormunen (2000) believed that some factors such as the layout of pages restricted journalists' options for illustration. In addition, images were rich in content, and they could be used in various contexts and in different ways. Moreover, journalists might use selected parts of images in the article or establish the associative links between the article and images in the captions. Sometimes journalists used neutral images to fill an empty space on a page. Markkula and Sormunen (2000) concluded that journalists apply a range of relevance criteria regarding their work-task situation, and that the importance of relevance criteria varied in different situations. Nevertheless, when the researchers asked journalists to specify the most important criterion, they stated that 'the technical quality of the photo' was the most important criterion for them.

Choi and Rasmussen (2002) conducted a quantitative study to examine relevance criteria before and after searching for images of American history. Thirty-eight graduate students of American history and faculty members from departments of history at Carnegie Mellon University, Duquesne University and the University of Pittsburgh participated in the study. All of them looked for images in the American Memory¹ online image collection, and they were asked to discuss how they evaluated relevant images. Since previous studies indicated a significant overlap between the criteria applied by end users, Choi and Rasmussen (2002) had offered a list of nine common criteria from those studies to participants and asked them to rate the criteria regarding the importance of each for their information needs: topicality, accuracy, time frame,

¹ See <http://memory.loc.gov/ammem/index.html>

suggestiveness, novelty, completeness, accessibility, appeal of information and technical attributes of images. Participants were then asked to search for images and evaluate retrieved images using these criteria. They were also asked to list other relevance criteria which they might apply. Before starting the search, participants were asked to rank the nine criteria: topicality, accuracy and completeness were ranked as the top three. Once participants had seen the retrieved photographs, they were found to apply criteria relating to aspects such as 'time frame' and 'accessibility of the photos'. The authors stated there had been a significant difference in the ratings of each criterion before and after users saw the images.

In a preliminary study, Hung et al. (2005) investigated the relevance criteria applied to images by ten students from the Department of Journalism and Media Studies at Rutgers University. The aim of this study was to find out what criteria searchers employ to select relevant images. They asked participants to look for images in the AccuNet / Associated Press Photo Archive¹ based on three pre-defined image search tasks, including those deemed as specific, general and subjective. Hung et al. (2005) define specific, general, and subjective image search tasks as follows:

Task 1(specific): You are photo editing a story on Tiger Woods for a sports magazine. For this story, you need to find some photos of Tiger Woods as illustrations. Task 2 (general): You are photo editing a report on the crisis in the Middle East for a newspaper. For this report, you need to find some photos regarding this topic to be used as illustrations. Task 3 (subjective): You are photo editing a special report on the topic of "Peace" and you need to find some photos to illustrate the meaning of "peace". (Hung et al., 2005)

Based on the finding of their study, Hung et al. (2005) identify several relevance criteria applied during three image search tasks. These criteria were typicality, emotion, action,

¹ The AccuNet database is a selected collection of the Associated Press Photo Archive and contains 400,000 photographs that have been published by the Associated Press since the 1840s. The Associated Press Photo Archive is a library of more than 50 million images (See <http://accuweather.ap.org/>).

aesthetic, text, familiarity, context, impression, preference, posture, facial feature and appearance. They found that typicality, emotion and aesthetic appearance were the three most important criteria, applied across all three tasks, where typicality was deemed the most frequent and most important criterion for all three tasks (according to the authors, typicality was a criterion that can exhibit universal representation of an object in a photograph).

Emotion was the second frequently applied criterion by participants across these three tasks. Hung et al. (2005) report that 'emotion' refers to situations in which an image contains the emotional context telling what is happening in the photograph. Thus they believe that images containing emotional context were more likely to be marked and be selected as relevant images. The third criterion, 'aesthetic', was not applied as frequently as emotion and typicality, and only female participants used this criterion. Hung et al. (2005) state that this might be due to the gender difference in relevance judgments in image retrieval; however, they stated that this assumption requires further investigation to find out whether such a gender pattern exists.

Table 2-3: The frequency of the relevance criteria identified by Hung et al. (2005).

Relevance criteria for specific task	Relevance criteria for general task	Relevance criteria for subjective task
Typicality	Typicality	Typicality
Emotion	Emotion	Emotion
Facial feature	Aesthetic	Aesthetic
Aesthetic	Action	Impression
Action	Impression	Familiarity
Posture	Familiarity	Context
Appearance	Preference	Text
Affection	Text	

In addition to these three core criteria, Hung et al. (2005) also point out that relevance criteria were used differently in these three image search tasks. For example, in the specific image search task participants considered the outward properties of objects in an image. This might be due to the large size of retrieval sets and a high level of similarity between the retrieved photographs. Hung et al. (2005) believe that in a specific image search task participants chose an image (photograph) that can display the characteristics or features of the objects and has an artistic quality. Thus, participants applied the aesthetic criterion more than other criteria in the specific image search. However, in the general image search, relevance judgment of retrieved images was based on criteria such as text, familiarity, impression or preference. They state this might be because of the unfamiliarity of the participants with the topic. Hung et al. (2005) conclude that participants' search tasks affect the relevance criteria that participants applied for the relevance judgments of images. They also mention that image users applied different criteria than those suggested in the literature for textual documents.

From the example studies mentioned above, we can conclude that topicality has been the most important criterion in relevance literature. While the assessment of topicality through the application of textual features of textual documents seems quite easy, topicality judgment is not easy for other types of media. For instance, Ingwersen and Järvelin (2005: p.239) state that topicality assessment in music is meaningless. Additionally there is no agreement in the relevance literature as to whether or not there are tangible and shared features for topicality assessment of other types of documents such as images.

A review of the image relevance literature also reveals that the process of relevance judgment for images is multidimensional and the judgment, no matter how the classes and types of relevance are described, is based on criteria extending beyond topical match between a query and the documents. Moreover, image retrieval poses difficulties compared to text document retrieval. Because the relevance judgment of images depends heavily on different levels of situational interpretation by the users, there are no agreed features to support the retrieval, interpretation and judgment process (Ingwersen and Järvelin, 2005: p.179). For example, textual documents (i.e. articles and books) have

common textual features such as title, author and publication date. By contrast image relevance studies, as discussed in this section, address a range of relevance criteria that might be applied by image users in various disciplines such as journalism and history. Nevertheless, researchers are unable to generalize about which image relevance criteria might be shared across users of information retrieval systems in various situations. Additionally, researchers of previous relevance studies could not specify which criteria could be the most important criteria for the users and in which circumstance this significance could happen. We found that our understanding of the judgement process for images is incomplete; as yet there are no distinct explanations for the way that relevance judgment is made for images.

Studies by Schamber et al. (1990); Ingwersen and Järvelin (2005); Borlund (2003a); Saracevic (2007); Hersh (2003) highlight the importance of the information needs of users in relevance judgment of retrieved documents in specific domain. Therefore, the next sections will focus on the information needs of health care professionals.

2.7. Information needs of health care professionals

This section describes the findings of a number of studies on the information needs and information seeking behaviour of health care professionals in general, and in particular on the medical image information needs and medical image seeking behaviour of health care professionals.

Lancaster (1979); Yang (2005); Tang (1999); Mizzaro (1998) reported that the information needs of users of information retrieval systems influence the criteria applied by users for the judgment of retrieved documents. Moreover, Revere et al. (2007) reported that the information seeking process is situational, contextual and unique to the information seeker; knowledge of the information needs of users can help design information retrieval systems that support those information needs. In the study of relevance judgment, it is crucial to know about the information needs of users. A clear understanding of the medical image information needs of health care professionals is also vital to the design process and development of medical image retrieval systems.

The information needs of health care professionals has been investigated in a number of studies. A comprehensive review of the literature would be impractical for the purposes of this study as our aim is to investigate the relevance criteria for the judgment of medical images. We thus limited the scope of our review to a few studies. Perhaps one of the most widely-quoted studies about the information needs of health care professionals is by Covell et al. (1985). The authors investigate the information needs of 47 physicians practising internal medicine during a half day and used a questionnaire and interviews to identify the physicians' self-perceived information needs. The physicians saw one to sixteen patients during the half day. The questionnaire was completed before physicians saw patients and an interview was performed after each patient was seen. Physicians mentioned that they had two questions for every three patients seen. Of these questions, 40% were factual questions (e.g. 'What are the side effects of bromocriptine?'), 45% were questions of medical opinion ('How do you manage a patient with labile hypertension?'), and 16% were questions of non-medical information ('How do you arrange home care for a patient?'). About a third of the questions were about treatment of specific conditions, a quarter about diagnosis, and 14% about drugs. Covell et al. (1985) report that the physicians looked both for information in print and for human resources to respond quickly after the patients' visits.

Table 2-4: Information sources used by Physicians in Covell et al. (1985)'s study.

Information source	Percentage use	
	Reported (n = 182)	Observed (n = 80)
Print sources:	62	27
General and specialty textbooks	25	3
Pharmaceutical textbooks	14	9
Journals	18	7
Drug company information	1	
Self made compendia	4	7
Human sources:	33	53
Specialist doctors	18	24
Generalist doctors	1	1
Office partner	3	4
Pharmacist	6	3
Other	5	21

The physicians stated that they used print sources such as general and specialty textbooks, pharmaceutical textbooks and journals, but in practice they were most likely to consult other health care professional (see Table 2-4). Print sources were not often used because of reasons such as age (publication year), poor organization, inadequate indexing, lack of knowledge of an appropriate source, and the time required to find the desired information. Most of the physicians stated that they could find information from books rather than from journals to fit their information needs. In average, four questions (from one half day) of each physician remained unanswered and the physicians reported barriers such as lack of time, cost, poor organization and non-availability of information sources for finding information they needed.

Based on the findings of the interviews with primary care physicians and the findings of published studies of health care professionals' information needs, Gorman (1995: p.730-731) defines information used by health care professionals in five types:

- 1) Patient data: information about an individual patient that includes information about past medical history, observations from physical examination, and results of diagnostic testing. Health care professionals usually obtain this type of information from the patient, his or her family and friends, and a patient's medical record.
- 2) Population statistics: aggregated information about groups or populations of patients. Health care professionals usually use their personal knowledge of recent illness patterns in the local population as a form of informal epidemiologic information, modifying their practices according to recent experience.
- 3) Medical knowledge: generalized information pertaining to the care of all patients. Basically medical knowledge may exist in the form of original research and systematic overviews published in the literature, or it may exist in the form textbooks.
- 4) Logistic information: refers to local knowledge about how to get the job done, often specific to a practice setting or payment mechanism. Gorman (1995) explains that logistic information may be as important in day-to-day medical practice as other types of

information. For example, health care professionals frequently ask questions such as: 'Which medicines (for treating a particular condition) are included in the hospital formulary?'

5) Social influences: refers to knowledge about the expectations and beliefs of others, especially peers such as colleagues and consultants, but also including patients, families, and others in the local community. Local practice patterns and expectations regarding prescribing of medication or performance of surgical procedures are examples of this type of information needs.

Gorman (1995) declares that not all health care professionals such as physicians have similar information needs. The different work situations of health care professionals create varying levels of information need. He explains that a physician in a small clinic may have less need for patient information than for medical knowledge; whereas a physician in a large hospital or other institution may experience the reverse. Gorman also adds that health care professionals rely heavily on human sources of information to respond their information needs. This reliance may result from a need for higher-order information than descriptive medical knowledge. For some questions, purely descriptive medical knowledge (such as that found in medical textbooks) may be sufficient. In other situations, however, health professionals may require higher-order information, such as confirmation, explanation, analysis, synthesis, and ultimately assessment; assessment that takes into account the complexity and patient's specificity.

Shelstad and Clevenger (1996) examine the information needs of surgeons in general medicine and information seeking patterns. Ninety-nine surgeons participated in the survey and they were asked to describe the purpose for which they required information. The results of the study showed that 'patient care', 'continuing medical education', and 'casual curiosity' were the most common purposes for which they needed information.

Table 2-5: Surgeons' purposes for medical information from Shelstad and Clevenger (1996: p.492).

Purpose	Number
Patient care	98
Continuing medical education	83
Curiosity	53
Patient education	48
Medico-legal purposes	43
Teaching	41
Research for publication	28

Surgeons also declared that 'professional meetings', 'the medical literature', 'colleagues', and 'continuing education courses' were the main sources for their information needs.

In an observational study, Ely et al. (1999) analyse the information needs and information-seeking behaviour of family doctors to determine the number, type, and urgency of patient care questions encountered. One hundred and three family physicians participated in this study. Ely et al. (1999) state that the most frequent questions were: 'What is the cause of symptom X?', 'What is the dose of drug X?' and 'How should I manage disease or finding X?'(see Table 2-7)

Table 2-6: Surgeons' source for their information needs (Shelstad and Clevenger, 1996: p.492).

Source of information	Number	Source of information	Number
Professional meeting	96	Audiotapes	53
Medical literature	95	Self literature search	50
M.D. colleagues	92	Computers	37
CE courses	82	State medical school	35
Personal library	79	Community library	33
Hospital/medical library	74	Television	25
Drug representatives	70	UNM Medical Library outreach	17
Librarian literature search	65	Compact discs	10
Videotapes	64	Motion pictures	9
Professional organizations	61	Computer information networks	6

It can be said that the nature of the questions asked revealed that the information needs of health care professionals had a direct relationship to their work task. The results of this study also indicated that textbooks and colleagues were the primary source of answers to patient care questions; formal literature searches in medical databases and the internet being rarely performed. This actually corresponds with the findings of previous research (Covell et al., 1985; Gorman, 1995; Shelstad and Clevenger, 1996) that health care professionals' information needs were usually met by medical literature, such as textbooks, and their colleagues. The authors suggested that the information needs of doctors had a direct relation to their work task and an effect on their relevance judgments. For example, Ely et al. (1999) report that when doctors were faced with a clinical problem, often they tended to ask questions such as 'How should I manage disease or finding X?' The authors added that doctors needed quick and up-to-date answers to their questions.

Table 2-7: Ten most common generic questions asked by 103 family doctors. Adapted from Ely et al. (1999: p.360).

Generic question	Questions asked [*]	Questions pursued ^{**}	Questions answered ^{***}
What is the cause of symptom X?	94 (9%)	8 (9%)	4 (50%)
What is the dose of drug X?	88 (8%)	75 (85%)	73 (97%)
How should I manage disease or finding X?	78 (7%)	23 (29%)	19 (83%)
How should I treat finding or disease X?	75 (7%)	25 (33%)	18 (72%)
What is the cause of physical finding X?	72 (7%)	13 (18%)	6 (46%)
What is the cause of test finding X?	45 (4%)	18 (40%)	13 (72%)
Could this patient have disease or condition X?	42 (4%)	6 (14%)	4 (67%)
Is test X indicated in situation Y?	41 (4%)	12 (29%)	10 (83%)
What is the drug of choice for condition X?	36 (3%)	17 (47%)	13 (76%)
Is drug X indicated in situation Y?	36 (3%)	9 (25%)	7 (78%)

^{*}Percentage is proportion of total questions asked (n=1101). ^{**}Percentage is proportion of questions asked.

^{***}Percentage is proportion of questions pursued. [†]Not specifying diagnostic management versus treatment.

The main conclusion reached from studies to date about the information needs of health care professionals is that to learn more about their needs requires an examination of the questions, situation, and context within which their needs arise. As Gorman (1995: p.733) states, some of the questions are fairly simple and direct; others are complex, multidimensional questions embedded in context. Additionally, the meaning and value of information provided by an information retrieval system depends not on the information itself but on the context in which it is received¹. As is mentioned earlier, Schamber et al. (1990) puts emphasis on the dynamic and multidimensional nature of relevance. Based on these studies, we suggest that although the user's background knowledge influences the dynamic and multidimensional process of the relevance judgment, the dynamism and multidimensionality of the information needs of users affect the relevance judgment and the relevance criteria applied by them.

2.7.1 Image information needs of health care professionals

To the best of our knowledge, there have been few previous studies specifically addressing the relevance criteria employed by health care professionals when searching for medical images, although existing studies have investigated their information needs. In a qualitative study Hersh et al. (2005)² examine the pictorial information needs of thirteen biomedical professionals with various roles including clinician, researcher, educator, librarian and student.

The results of a study by Hersh et al. (2005) show that the medical image needs of biomedical professionals can be categorized into four groups: research-related; patient care-related; education-related; and other (see Table 2-8).

Hersh et al. (2005) investigate some real tasks of biomedical professionals, which were supported by medical images; however, this was a small-scale study. Moreover, they did not study the source of medical images that biomedical professionals use. No prior

¹ This supports the recent flurry of work in information retrieval on context (e.g. Information Retrieval in Context (IRiX) conference).

² Using similar methodology and questionnaire Muller et al., (2006) conducted a repeat study.

research has examined the source of medical image information needs of health care professionals while they are using these images. Furthermore, once an image information need and its context are identified, context-specific image collections can be created and tailored to meet the image information needs of health care professionals during interactions with medical image retrieval systems.

Table 2-8: Purposes of medical image retrieval by biomedical professionals (Hersh et al., 2005).

Research-related	
Images as data: analysis and interpretation of images (photomicrographs, magnetic resonance imaging, etc.)	
Images for presentation and publication of research findings to research audience	
Patient care-related	
Check image test result in electronic health record	
Diagnosis of uncommon or unrecognized condition	
Illustration and explanation to patient	
Education-related	
Educational presentations to students, etc. (listed for all roles)	
Learning – librarian support of others' learning; clinician self-education on new technique	
Other	
General Information	Educators
Expert witness testimony	Clinician
Developing collections	Educators
Marketing (before and after)	Clinician (plastic surgery)

Paling and Miskiewicz (2005) investigated the image information needs of 34 dental faculty members and clinicians. They reported that participants looked for images in a variety of sources, such as search engines, personal collections, digital textbooks, digital journal articles, database and CD/DVDs. The authors reported that a substantial number of the participants preferred to find and use digital images, and that none of the participants indicated an overall preference for physical slides. Paling and Miskiewicz (2005) suggest that online dental image collections could be a good match for the participants' dental image information needs. Participants also preferred to access higher quality, manipulable images and metadata schemes for describing the content of images such as the name of a disease or injury.

The importance of digital images in domains such as medicine is great; digital imaging has become a vital component of a large number of applications within current clinical settings (Eakins and Graham, 1990; Glatard et al., 2004; Müller et al., 2004). According to Eakins and Graham (1999), medical images are utilized by a variety of users such as medical students, lecturers in medical departments, and clinicians, each with different levels of subject knowledge. Access to images is commonly mediated through an electronic patient record system such as DICOM¹ or PACS². Although research into the effectiveness of such systems is extensive, much of the research is based on a particular notion of relevance (i.e. topicality), and there is relatively little research into the criteria used by professional users who search for medical images as part of their daily work.

Researchers such as Markkula and Sormunen (2000); Greisdorf and O'Connor (2002); Cunningham et al. (2004); Tsai (2007) have studied the needs and information seeking behaviour of users searching for images. However, such studies have not addressed the search for images in clinical settings. Similarly, although relevance studies have explored criteria elicited from the users of document retrieval systems (see e.g. Saracevic, 1996; Mizzaro, 1997), the understanding of such criteria, particularly for medical images, is limited. Previous work within the large body of user-oriented relevance has not addressed how health care professionals judge the relevancy of medical images for their information needs. In addition, there have been few studies concerning the medical image information needs of health care professionals. We were therefore motivated to conduct such a study based on interviews with such professionals, to explore relevance criteria commonly employed for medical images.

Summary

This chapter has provided an overview of the concept of relevance and has given a description of different types and classes of relevance. It includes a review of the related literature and some closely related relevance studies have been discussed. In addition, the information needs of health care professionals have been discussed in detail.

¹ Digital Imaging and Communications in Medicine

² Picture Archiving and Communications System

CHAPTER 3-METHODOLOGY

Introduction

This chapter describes the methodology used to study relevance criteria for medical images as applied by health care professionals. It justifies the reasons for selecting a qualitative research approach in general and in particular our motivation for using grounded theory. The Straussian version of grounded theory was selected for data analysis and the methodological procedures adopted in this study are based on the framework developed by Strauss and Corbin (1998). This chapter also reports a pilot study that we performed before the main study. In addition, we discuss ethical issues, research participants, and data collection protocols, together with strategies applied to enhance the trustfulness and replicability of the findings.

3.1. Quantitative vs. qualitative

In order to conduct successful research and ensure the findings are valid, it is important for all researchers to know which research methods are appropriate. Myers (1997) defines the term 'research method' and highlights its importance 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. (Myers, 1997: p.241)

Research methods are traditionally classified into quantitative versus qualitative. Although there are many definitions of qualitative and quantitative research methods, some examples are given here to demonstrate the variety in the existing definitions.

Bryman (2001) describes quantitative research as a method that focuses on the collection of numerical data and quantification in the data analysis stage and outlines the main steps of quantitative research as illustrated in Figure 3-1:

Relevance criteria for medical images applied by health care professionals

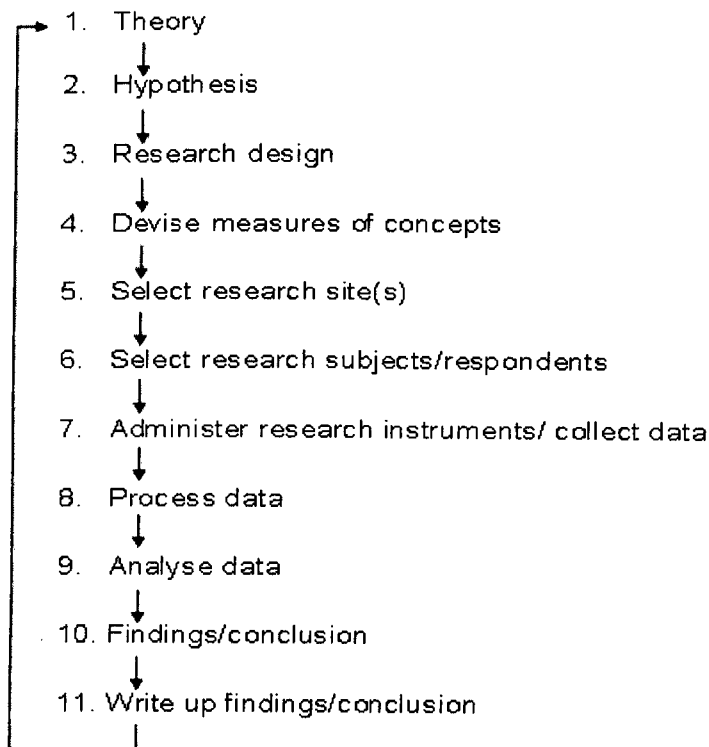


Figure 3-1: Quantitative method as illustrated by Bryman (2001: p.63).

In contrast, he defines qualitative research as an interpretive methodology which deals with verbal data:

Qualitative research is a research strategy that usually emphasizes words rather than quantification in the collection and analysis of data. As a research strategy it is inductive, constructionist, and interpretivist, but qualitative researchers do not always subscribe to all three of these features. (Bryman, 2001: p. 266)

Strauss and Corbin (1998) define a qualitative approach as follows:

By the term qualitative we mean any kind of research that produces findings not arrived at by means of statistical procedures or other means of quantification. It can refer to research about persons' lives, lived experiences, behaviours, emotions, and feelings as well as about organizational functioning, social movements, cultural phenomena, and interactions between nations. Some of the data may be quantified ...but the bulk of the analysis is interpretative. (Strauss and Corbin, 1998: p.11)

Myers (1997) puts emphasis on the origin of data in qualitative research and described this method as follows:

Qualitative research involves the use of qualitative data, such as interviews, documents, and participant observation data, to understand and explain social phenomena. (Myers, 1997)

There has been a prolonged and ongoing debate known as quantitative vs. qualitative in social science research since the 1960s. Some believe that only quantitative studies can be used to study human behaviour; others think that only qualitative studies are appropriate (Punch, 2005). The debate, sometimes described as the 'paradigm wars', challenged the traditional dominance of quantitative methods in social science and accompanied a major growth of interest in using qualitative methods which in turn led to a split in the field between quantitative and qualitative researchers.

The quantitative and qualitative approaches to research have important differences, and the distinction between these approaches is based on a range of considerations. The main difference between the two methods lays in the nature of their data, and in methods for collecting and analysing data. Studies using quantitative research deal with numerical data (i.e. quantities): qualitative research emphasizes non-numerical data such as terms, concepts, meanings and knowledge of participants. Thus the findings of the two methods are very different. According to Punch (2005), findings of quantitative studies can be generalized from a sample to some larger population. The intention of

qualitative studies is not to generalize the findings, but rather to provide an in-depth understanding of the phenomena under study within its real-life context.

Another difference between quantitative and qualitative methods relates to the way in which concepts or variables are employed by researchers in the two methods. Bryman (2001) describes concepts as categories for organisation of ideas and observations. He states that concepts are the building blocks of theory in both quantitative and qualitative research, and represent the points around which study is conducted. In fact quantitative research is based on measurement and the researcher measures preconceived concepts. For example, IQ (intelligence quotient) is a measure of a concept known as intelligence. Once concepts are measured, they can be classified as dependent or independent variables. However, emergence is the foundation of qualitative research and concepts emerge from the data. Once emerged concepts have been validated and revised in relation to the data, the researcher might turn to quantitative analysis if this will enhance the research process.

Moreover, qualitative research method can be distinguished from quantitative research methods in terms of the instruments of data collection and the type of data collected. Myers (1997); Glazier and Powell (1992); Strauss and Corbin (1990) state that the data collection tools used in qualitative research include observations, interviews, content analysis of documents, and audio-video recordings. Additionally, Glazier and Powell (1992) state that the type of data collected in qualitative research methods include communications between people, groups and organizations, together with explanations of certain phenomena. Thus, qualitative research methods are rich in description. Strauss and Corbin (1990) also emphasize that while researchers might utilize qualitative data collection tools such as interviews and observation to collect data, they might then analyse the qualitative data and produce their findings using mathematical and statistical procedures. Qualitative research methods refer to an informal analytic approach that allows researchers to derive findings from data collected using tools such as observation, interviews, documents and videotapes.

Table 3-1 presents some characteristics of quantitative and qualitative research suggested in the literature by Bryman (2001); Glaser (2003) ; Berg (2001); Slater (1990); Pickard (2007); Glazier and Powell (1992); Myers (1997).

Table 3-1: Some of the differences between qualitative and quantitative methods.

Quantitative research	Qualitative research
Objective	Subjective
Deals with numeric data	Deals with verbal data
Uses statistical sampling: typically has large samples	Uses theoretical sampling: typically has smaller samples
Data analysis is based on mathematical techniques	Data analysis is based on interpretation
Creates explanatory rules	Provides in-depth description
Context free	Context dependent
Researcher examines theory (deduction)	Researcher develops theory (induction)
Researcher does not participate in the research	Researcher participates in the research
Researcher measures the concepts	Researcher develops the concepts
Researcher uses instruments to collect data	Researcher collects data via communication or observation

Both quantitative and qualitative approaches are needed in social research and both have roles to play in theorizing. It is not a case of trying to determine whether one approach is superior to the other, but rather to see how the two approaches might work together to foster the development of theory. In fact each of these methods has its own strengths and weaknesses; therefore these methods can be complementary to each other. For example, quantitative methods enable standardized, objective comparisons to be made. On the other hand, qualitative methods are flexible and can be easily modified as study progresses (Punch, 2005). What can be concluded from the definitions and differences defined for the two approaches is that we need to be clear in our response to the following question: What exactly are we trying to find out?

3.2. Research method adopted in this study

Strauss and Corbin (1998) note that once researchers have specified their research aims and objectives and research questions, they should select appropriate research methods and data collection tools in order to reach those aims and objectives and to respond their research questions:

The original research question and the manner in which it is phrased lead the researcher to examine data from a specific perspective and to use certain data-gathering techniques and modes of data analysis. (Strauss and Corbin, 1998: p. 53)

Additionally, Strauss and Corbin (1998) believe that research questions not only help researchers to stay focused, but also affect selection of the research method and data collection tools. In some cases, it is easy to say whether a research question could be answered using a quantitative or a qualitative approach. However there are situations in which a question is more specific. Punch (2005) argues that when researchers make research questions more specific, they can often see the interaction between the research question on the one hand, and the design and method of the study on the other. This interaction raises the point that some questions are quantitative and a quantitative method is required to answer them, while other questions are qualitative and researchers can only answer them using a qualitative method. The method chosen can also affect the

research questions asked: the important thing, however, is the matching of research questions with the research method.

As stated earlier, the goal of this study is to elicit and document the relevance criteria that health care professionals apply when searching for medical images and to provide recommendations and strategies that could assist medical image retrieval research groups improve their systems. The central question in this study is to find out how health care professionals determine which medical images are relevant to their information needs when searching for medical images. The question now raised is: Should we take a quantitative approach or a qualitative approach in our study?

Sonnenwald et al. (2001) claim that quantitative research methods are unable to provide data on the dynamic nature or complexity of many information seeking situations and contexts such as relevance judgment process and relevance criteria. However, Maglaughlin and Sonnenwald (2002); Ingwersen and Järvelin (2005); Hirsh (1999); Park (1994); Myers (1997) maintain that qualitative research methods can capture an in-depth understanding of the complex and dynamic concept of relevance. Additionally, in studies such as Hirsh (1999); Maglaughlin and Sonnenwald (2002); Tang and Solomon (1998); Choi and Rasmussen (2002); Park (1994); Barry and Schamber (1998); Xu and Chen (2005); Barry (1994) researchers applied qualitative, exploratory, and descriptive research methods to study situations of information retrieval users and the relevance judgment process.

Schamber (1994) believes that researchers apply qualitative research methods in order to document users' feedback using techniques such as open-ended questioning. The researcher then analyses participants' answers to elicit criteria of relevance applied by users during the relevance judgment process. There are a number of reasons that a qualitative method seemed most appropriate for the study of the relevance criteria for medical images used by health care professionals.

Firstly, the research questions of the current study are qualitative by nature. As we stated earlier, the central question in this study is to find out how health care

professionals determine which medical images are relevant to their information needs when searching for medical images. To answer this question we need to access the health care professionals' perception of the concept of relevance, the criteria of relevance for medical images they use, and their image information needs.

Secondly, there is no known research exploring the relevance judgment process and relevance criteria for medical images used by health care professionals. Qualitative methods are especially suitable when there is very little knowledge about the subject of the study. Moreover, the qualitative method is most appropriate when the researcher does not know what is likely to be found during the research (Morse and Richards, 2002; Pickard, 2007).

Thirdly, the purpose of this study is to establish a substantive theory or "*a theoretical framework that reflects reality*". As Morse and Richards (2002) state:

If the purpose is to construct a theory or theoretical framework that reflects reality rather than your own perspective or prior research results, you may need qualitative methods that assist the discovery of theory in data. (Morse and Richards, 2002)

Fourthly we are interested in studying and understanding the relevance judgment process and relevance criteria for medical images in depth and in its real-life context. According to Morse and Richards (2002) and Pickard (2007), this cannot be easily done unless a qualitative method is used.

Myers (1997) notes that a qualitative research method is often employed by researchers to investigate phenomena relating to society and culture. Similarly, Strauss and Corbin (1998) state that the qualitative method includes any methodology that refers to investigation about "*persons' lives, lived experiences, behaviours, emotions, and*

feelings as well as organizational functioning, social movements, cultural phenomena, and interactional relationship.”

Strauss and Corbin (1998) suggest that qualitative research contains three basic elements: data, analytic or interpretive procedures, and written and verbal reports. A researcher gathers data using tools such as interviews, observations, documents, records, and films. The analytic or interpretative procedures refer to procedures that researchers apply to reach results or develop theories. In qualitative research the bulk of the analysis is interpretative. Written and verbal reports might be published in scientific journals or presented verbally in conferences.

In the light of the reasoning above, we adopted a qualitative method. We expected that the qualitative study we carried out would enrich our understanding of the concept of relevance and the criteria of relevance used by health care professionals for their medical image information needs.

There are many types of qualitative research methods including action research, ethnography, case studies, and grounded theory. Suitable tools and techniques for data collection and data analysis were based on using grounded theory, specifically the ‘Straussian version’ of grounded theory. The reasons for selecting grounded theory is explained in the next sections, however before that we describe this strategy.

3.3. Grounded theory

Bryman (2001) notes that “*grounded theory has become by far the most widely used framework for analyzing qualitative data*”. As a qualitative research strategy, grounded theory is specific and different as it is both a methodology and a set of methods for data analysis.

In the 1960s Anselm Strauss and Barney Glaser began collaborative work in medical sociology. They published a book under the title of ‘The Discovery of Grounded Theory’ in 1967 and introduced the grounded theory method. Strauss and Corbin (1990) published their book ‘Basics of Qualitative Research’ in 1990. Following this Glaser

(1992) stated that Strauss had changed the meaning of grounded theory, and that what he has described in his book could not be called grounded theory.

After this split, grounded theory was divided into two versions: Glaserian and Straussian. Glaser and Holton (2004) define grounded theory as follows:

GT [grounded theory] is simply a set of integrated conceptual hypotheses systematically generated to produce an inductive theory about a substantive area. (Glaser and Holton, 2004: p.7)

Strauss and Corbin (1990) emphasize that grounded theory is a qualitative research method and define it as follows:

The grounded theory approach is a qualitative research method that uses a systematic set of procedures to develop an inductively derived grounded theory about a phenomenon. The research findings constitute a theoretical formulation of the reality under investigation, rather than consisting of a set of numbers or a group of loosely related themes. Through this methodology, the concepts and relationships among them are not only generated but they are also provisionally tested. (Strauss and Corbin, 1990: p. 24)

Grounded theory is not actually a theory at all, it is a strategy. Grounded means that the theory will emerge from the data, rather than from examining a previous theory; the theory therefore will be grounded in data. Theory means that the objective of data collection and analysis is to generate theory. The essential idea in grounded theory is that theory will be developed inductively from data. Once the theory has been developed on the basis of data, it can be validated by comparing it with other theories suggested in the relevant literature.

There are a number of reasons that justify the use of grounded theory in this study. Firstly, we have not generated any hypotheses with respect to the area of study. In other words, we needed to apply an inductive research method in this study to construct the

theory from an experiential perspective rather than testing a preconceived hypothesis. Grounded theory was created for this purpose.

Secondly, there is no literature available concerning the relevance criteria for medical images used by health care professionals. According to Goulding (1998) grounded theory is a strategy that has been used to develop theory where little is already known about a research topic.

Thirdly, we were interested in enriching our understanding of the concept of relevance, the relevance judgment process, and the relevance criteria used for medical images in its real-life context. According to Strauss and Corbin (1998), in grounded theory the theories emerging from the data are more likely to resemble the reality, and to offer insight and enhance understanding of phenomena under investigation.

3.4. Qualitative methods and grounded theory in LIS

Although research methods and data analysis strategies in Library and Information Science (LIS) are predominantly quantitative, there are more modest increases in the use of qualitative approaches in contemporary library and information science research as Powell (1999) reports. Similarly Wilson (1999) believes that the adoption of qualitative methods has risen since the early 1970s. He also comments that the use of quantitative research methods for studying human behaviour within the library and information science context is inappropriate (e.g. the information seeking behaviour of library users). He explains why the number of qualitative studies in LIS has increased and reports:

First, in the positivist tradition, quantitative research methods were adopted that were inappropriate to the study of human behaviour: many things were counted, from the number of visits to libraries, to the number of personal subscriptions to journals and the number of items cited in papers. Very little of this counting revealed insights of value for the development of theory or, indeed, of practice. Secondly, researchers in the field of information science seem generally to have ignored allied work in related areas that might offer more robust theoretical models of human behaviour... Thirdly, general models of information behaviour have only begun to emerge, and attract much attention, in the past ten to fifteen years. (Wilson, 1999: p.250)

As mentioned earlier, the current research was undertaken to investigate the image seeking behaviour of, and the relevance criteria for medical images applied by, health care professionals. Wilson (1999: p.249) defines information seeking behaviour as a set of “activities a person may engage in when identifying his or her own needs for information, searching for such information in any way, and using or transferring that information” and states that quantitative methods are unable to study information seeking behaviour in LIS. Wilson (2006: p.666) advocates the use of qualitative methods in user studies in LIS and lists a number of reasons for the appropriateness of qualitative methods for information seeking behaviour studies:

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. (Wilson, 2006: p.666)

According to Yang (2005: p.35) in the earlier empirical relevance studies in LIS, methods of data gathering and analysis have predominantly been quantitative. This situation was related to the centrality of topical relevance in relevance studies: the relation between the search query and content of retrieved documents. As discussed previously, this class of relevance, referred to as an objective approach, ignores the end users of information systems. In recent years, researchers have focused on the subjective approach to relevance in their studies and have attempted to discover what end users think about relevance. The shifting from an objective approach to relevance to a subjective approach also resulted in a shift in research methods employed in relevance studies. Instead of applying quantitative methods such as surveys, researchers used qualitative methods such as grounded theory (the research method employed in this study) to investigate the concept of relevance in its real-life context.

Myers (1997) reports that grounded theory has become a popular research method in LIS since the 1990s. Powell (1999) reports on qualitative methods such as ethnography, grounded theory, and phenomenology applied by LIS researchers in detail, and expresses the need for further studies based on grounded theory as follows:

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. (Powell, 1999: p.103)

3.5. Straussian or Glaserian versions of grounded theory

As mentioned earlier, there are two different versions of grounded theory. In both versions of grounded theory, the researcher does not begin the research with a theory, and then look for proof. Instead the theory is derived from the data. Furthermore, during the data analysis stage in both versions, the researcher constructs concepts from the obtained data. Then, the researcher carries out new interviews/observations to verify and amplify those concepts and group them to form the categories (sometimes a concept might become a category). However, notwithstanding the similarities, there are a number of differences between Glaserian and Straussian versions of grounded theory. For example, in a Glaserian approach the researcher plays a neutral role during the study; however in a Straussian approach the researcher is an active participant of the study. Table 3-2 shows some of the differences between data analysis stages in Glaserian and Straussian versions of grounded theory.

The differences between the Straussian and Glaserian versions of grounded theory could be summarised by stating that the Glaserian version of grounded theory is a 'purist' approach that emphasises an 'open' attitude to the research where the researcher is professionally naïve. In this version, theory does not come from the researchers' preconceptions, but comes straight from the data. By contrast, Straussian grounded theory is as a 'pragmatic' method that emphasises a 'structured' attitude in grounding the theory. In this approach the researcher must apply a set of tools and procedures. Moreover, Strauss and Corbin (1998) suggest researchers have an active role in the research in order to apply existing insights and experience during the research.

According to Hekkala (2007), for the majority of studies based on grounded theory in the field of information science, researchers used the Straussian version of grounded theory. She also cites Urquhart (2001) and reports that most people in information

science research believe that Strauss and Corbin's book entitled 'Basics of qualitative research: grounded theory procedures and techniques' is the definitive book on grounded theory.

Table 3-2: Differences between Glaserian and Straussian grounded theory.

Glaserian	Straussian
The grounded theory is independent of the researcher's ideas	The grounded theory is influenced by the researcher's ideas
Theory emerges directly from the data	Researchers use the data to shape the theory
Glaserian approach emphasizes an open attitude to the research	Straussian approach relies on a systematic attitude to grounding theory
Researcher does not need to review the literature in the area under study, either at the beginning of the study or during the data analysis	To define the research questions for the study, the researcher needs to review the literature in the area under study
All data are important, and the researcher should avoid data selection	Researcher selects data that relate to the identified concept and categories
Researcher must verify concepts by all data and constantly refit the categories	Concepts and categories will be verified regarding their appearance in the data
Researcher plays a neutral role during the study	Researcher is an active participant in the study

In addition, Pickard (2007: p.242) notes that selecting the Straussian or Glaserian versions of grounded theory is up to the individual researcher. However, he mentions that Straussian grounded theory offers researchers something to hang on to in what can

often be a turbulent sea of excessive amounts of unbounded descriptive data. As mentioned earlier, this study proposes to investigate participants' perspectives on the relevance criteria and relevance judgment process for medical images. Fernández (2004) cites Lehmann (2001) and reports that the Straussian approach appears to be more useful for studies of individuals than studies involving organizational, political, and technical issues. Therefore, the Straussian approach seems most appropriate for this study.

Heath and Cowley (2004) compare Straussian and Glaserian approaches and suggest that *“researchers need to select the method that best suits their cognitive style and develop analytic skills through doing research”*. Thus, we conducted a preliminary study to examine the appropriateness of Glaserian and Straussian versions of grounded theory (see section 3.10). Initially we started with a Glaserian approach, but found it troublesome to use its coding method for our data. Based on our experiences during the preliminary study and with regard to the differences identified between the Glaserian and Straussian versions of grounded theory, we believe that the Straussian version of grounded theory best fits the aims and objectives of this study.

Additionally, we found that the Straussian approach suits this study better due to its structured pragmatic approach to data collection and data analysis.

Though acknowledging and recognising the spirit of the Glaserian version of grounded theory, this study employs the Straussian version of grounded theory developed and tested by Strauss and Corbin (1998) as a research strategy and data analysis model. Choosing the most appropriate methodology helped us to decrease the risk of methodology mistakes and uncertainty – see Heath and Cowley (2004).

3.6. Components of grounded theory

Grounded theory incorporates a number of steps to ensure good theory development. We discuss its main steps and components in the following sections.

3.6.1 Data collection in grounded theory: Theoretical sampling

According to Strauss and Corbin (1998), in grounded theory data collection, data analysis and building the theory are regarded as reciprocally fused. Guided by some initial research questions, we selected the research population, the type of data and the data collection method. Then we collected the first piece of data. At that point, we could start the data analysis process using a constant comparison method (its procedures are described in the next section). After the first set of data was analysed, the second set was collected using the directions that emerged from the first data analysis. This is the principle of theoretical sampling. Strauss and Corbin (1998) define theoretical sampling as follows:

Theoretical sampling enables the researcher to capture all potentially relevant aspects of the topic as soon as they were perceived. (Strauss and Corbin, 1998)

According to Corbin and Strauss (1990) this is the characteristic feature of grounded theory. Corbin and Strauss (1990) also believe that theoretical sampling decreases the risk of researcher bias. This is because researchers do not begin their studies with a predefined theory in mind, but begin to develop and verify a theory from the data.

According to Strauss and Corbin (1990), the researcher continues the cycle of alternation between data collection and data analysis until theoretical saturation is achieved. Strauss and Corbin) state that saturation is achieved when no new concept or category emerges, and the researcher has identified the main category and established the relationship between the main category and others.

The general rule in grounded theory research is to sample until theoretical saturation of each category is reached. This means, until: (1) no new or relevant data seem to emerge regarding a category; (2) the category development is dense, insofar as all of the paradigm elements are accounted for, along with variation and process; (3) the relationships between categories are well established and validated.. Theoretical saturation is of great importance. Unless you strive for this saturation, your theory will be conceptually inadequate. (Strauss and Corbin, 1990: p. 188)

3.6.2 Data analysis in grounded theory: Comparative analysis

As stated in section 3.3, grounded theory is both an approach to research and a way of data analysis. The grounded theory approach was described in previous sections. This section deals with the basic ideas of grounded theory. Grounded theory has a set of procedures for developing theory through the analysis of data. Strauss and Corbin (1990) recommended that in grounded theory researchers need to use its rules and procedures for data collection, data analysis and theory generating.

Grounded theory analysis aims directly at generating theory to explain what is central in the data. Glaser and Holton (2004); Corbin and Strauss (1990); Glaser and Strauss (1967) maintain that a grounded theorist must identify the elements of theory in order to establish it. These are concepts (or variables), categories, and hypotheses (propositions).

Concepts are the basic units of data analysis in grounded theory, thus conceptualization of the data is the first and the most important stage as Corbin and Strauss (1990) state:

Theories cannot be built with actual incidents or activities as observed or reported; that is, from "raw data." The incidents, events, and happenings are taken as, or analyzed as, potential indicators of phenomena, which are thereby given conceptual labels. ... Only by comparing incidents and naming like phenomena with the same term can a theorist accumulate the basic units for theory. In the grounded theory approach such concepts become more numerous and more abstract as the analysis continues. (Corbin and Strauss, 1990: p. 7)

In grounded theory, any concept involved or discovered in the study is considered as provisional. Then the importance of each concept in the induction of the theory relate to its repeated presence in the data.

The next step is to find relationship between these concepts in order to form the categories, the second elements of grounded theory. Corbin and Strauss (1990) define categories as follows:

Categories are higher in level and more abstract than the concepts they represent. They are generated through the same analytic process of making comparisons to highlight similarities and differences that is used to produce lower level concepts. Categories are the "cornerstones" of a developing theory. They provide the means by which a theory can be integrated. (Corbin and Strauss, 1990: p. 9)

There is a systematic relationship between concepts and categories in grounded theory. Each category contains a group of interconnected concepts. Grouping of concepts and forming categories are based on the similarities and differences between concepts.

The third element of grounded theory, hypotheses, clarifies the relationship between a category and its concepts and separate categories. Glaser and Strauss (1967) emphasize that the hypotheses have at first the status of suggested, not tested, relations between categories and their concepts. At the beginning, the researcher needs evidence only to construct hypotheses, not to collect evidence to prove the hypotheses. The initial

hypotheses may seem unrelated. However, as the researcher develops categories and concepts, hypotheses become verified and related.

The essential idea in grounded theory analysis is to select a core category which is grounded in the data. The basic operation in grounded theory analysis is coding. Strauss and Corbin (1990) distinguish between three types of coding in the Straussian version of grounded theory. These are: open coding, axial coding and selective coding. Strauss and Corbin (1990) emphasize that these stages are not clearly separated. Thus, researchers may move from one stage of coding to another stage. The different phases of coding sequence in Straussian grounded theory could be described as follows.

3.6.3 Open coding

This is the first phase of data analysis in grounded theory and the intention of this stage is to reveal fundamental characteristics of the phenomenon under investigation. Strauss and Corbin (1998: p.102) describe open coding as the most important analytical step in grounded theory and report that the rest of the data analysis and communication follows from it. They also explain why the term 'open' is used: researchers must open up the text line-by-line or paragraph-by-paragraph and expose the thoughts, ideas, and meanings contained therein. Open coding as defined by Strauss and Corbin (1998) is:

The analytic process through which concepts are identified and their properties and dimensions are discovered in the data. (Strauss and Corbin, 1998: p.101)

Open coding is the process of conceptualizing and it aims to identify and name concepts from the written data.

The researcher begins by breaking down and conceptualizing the data in order to recognize phenomena (concepts) such as events, object, happenings, or actions/interactions that seem to be significant in the data. Conceptualizing is directed by two procedures: theoretical sampling and questioning of the data. Once phenomena are identified, researchers must give each phenomenon a conceptual label. In fact, a

concept is a labelled phenomenon. The names or labels chosen for a concept can be the name of an object, or the words of respondents themselves, referred to as '*in vivo* codes'.

The purpose behind labelling a concept is to enable researchers to compare each concept with the others in order to find similarities and differences between them. Conceptualizing the data also requires asking questions of the data. Questions such as: 'What is this data referring to?' 'Do other participants hold similar ideas?' 'Is there a specific theme or concept to which this issue relates?' Through the comparison researchers can examine the concepts, group similar concepts, and construct categories. While a category stands for a phenomenon such as an event or object, a concept answers questions such as when, where, why, and how a phenomenon is likely to occur. In fact concepts describe the properties and dimensions of categories. According to Strauss and Corbin (1998) concepts become clearer and more precise during axial coding.

3.6.4 Axial coding

Axial coding is the next stage of coding in grounded theory, where the main concepts which have been generated in the open coding stage become interconnected with each other. Strauss and Corbin (1998) explain this process as follows:

The process of relating categories to their subcategories, termed 'axial' because coding occurs around the axis of a category, linking categories at the level of properties and dimensions. (Strauss and Corbin, 1998: p. 123)

The purpose of this stage is to identify causal relations between concepts which open coding has developed, and generate a model which illustrates the relationship between concepts. If open coding breaks the data apart, in order to identify concepts, axial coding puts the concepts back together. Now the question that might be raised here is that of how concepts become interrelated to form categories?

To do interrelating, we need some concepts developed during open coding.

There are three basic groups of concepts in grounded theory: conditions, actions/interactions, and consequences. Conditions are responses to the questions why, where, how come, and when. They form the structures and sets of circumstances or situations in which categories are embedded. Actions/interactions are answers to the questions whom and how. They are "*strategic or routine responses made by individuals or groups to issues, problems, happenings, or events that arise under those conditions*" (Strauss and Corbin, 1998: p.128). Consequences are outcomes of actions/interactions and they respond to questions such as what happens as a result of those actions/interactions. By the end of the axial coding stage researchers must produce hypotheses or propositions which explain how concepts are related to each other.

3.6.5 Selective coding

Selective coding is the third stage of analysis in grounded theory. Strauss and Corbin (1998) define selective coding as "*the process of integration and refining categories*" (Strauss and Corbin, 1990). In this stage the researcher integrates categories that have been developed to identify the core category. The core category systematically relates to other categories, and represents the main phenomenon around which all other categories are based. This is the final stage in grounded theory analysis, where the researcher has reached theoretical saturation. Saturation is a point at which no new concept or category emerges from the data. There are some criteria provided by Strauss and Corbin (1998) to decide whether a category is the core category or not. We used these recommendations in our study.

[1] It must be central; that is, all other major categories can be related to it. [2] It must appear frequently in the data. This means that within all or almost all cases, there are indicators pointing to that concept. [3] The explanation that evolves by relating the categories is logical and consistent. There is no forcing of data.. [4] The name or phrase used to describe the central category should be sufficiently abstract that it can be used to do research in other substantive areas, leading to the development of a more general theory. [5] As the concept is refined analytically through the integration with other concepts, the theory grows in depth and explanatory power. [6] The concept is able to explain variation as well as the main point made by the data; that is, when conditions vary, the explanations still holds, although the way in which a phenomenon is expressed might look somewhat different. One also should be able to explain contradictory or alternative cases in terms of that central idea. (Strauss and Corbin, 1998: p.147)

What we can conclude from open, axial, and selective coding is that grounded theory analysis is based on abstraction. The essential idea of grounded theory is to select a core category at the highest level of abstraction, but grounded in the data. In fact, grounded theory and its components are the outcome of different levels of abstraction during the data analysis process. Categories are identified in the first level of abstraction (open coding). They are at a more abstract level than the data themselves. The aim of second level abstraction (axial coding) is to bring categories together and to interrelate them in a higher level of abstraction than first level. The objective of the third level of abstraction (selective coding) is to find a higher-order, more abstract construct, and identify the core category which integrates the other categories.

3.6.6 Memos

Strauss and Corbin (1998: p.111) describe 'memos' as "*the researcher's record of analysis, thought, interpretation, questions, and directions for further data collection.*" Typically, memoing helps researchers to abstract and to record ideas during the

research.. As researchers begin the analysis of data and develop the codes, they should start memoing and continue throughout the research. Strauss and Corbin (1990) believe that without memoing, grounding the theory could not occur. Due to the importance of memoing Strauss and Corbin (1998: pp.221-223) suggest recommendations to be applied by researchers while memoing. We applied these recommendations in this study. According to Strauss and Corbin (1998) memos should contain headings, short quotations of raw data, and type of memo; all of the memos should be dated. They also noted that researchers should not be afraid of style of memoing. Each researcher can develop his or her own style of memoing.

The aim of memoing in grounded theory is to construct a store house of ideas or memo fund. Memos facilitate identification of the core category and the integration of categories. Although researchers continue memoing until data saturation occurs, they must sort and resort memos as categories become better elaborated and integrated. Furthermore, sorted memos help researchers to generate a framework for integration of categories and grounding the theory; thus, sorting of memos in the fund is an important stage in grounded theory as Strauss and Corbin (1998: p.240) state.

3.6.7 Example coding¹

Using the Straussian version of grounded theory, we analysed the interview transcripts to allocate every line or paragraph a concept label. The following examples are illustrations only of the coding process and do not represent findings of the current study:

¹ The researcher has put this section in as an example of the coding process.

P3¹: I prefer to use images of my previous patients who gave me their consent. I mean I prefer real images. I tend to look at those images more than searching for images in books. If I could not find them I go through Medline articles or medical organizations' websites like Sport Medicine Association, Cartilage Association and Joint Association websites. For example, if I am looking for images about meniscus tear in the knee I go to a relevant site or a journal site and I will definitely find what I need.

The previous paragraph (and associated part of the interview) illustrates a range of emerging concepts. By referring to “*I prefer to use images of my previous patients who gave me their consent*” the participant states that he uses his personal image collection. The participant then explained he prefers “*to look at those images more than searching for images in books.*” In addition he stated that he would use image sources other than books to find images: “*If I could not find them I go through Medline articles or medical organizations' websites like Sport Medicine Association, Cartilage Association and Joint Association websites.*” Finally the statement “*For example, if I am looking for images about meniscus tear in the knee I go to a relevant site or journal site and I will definitely find what I need*” implies that the participant looks for medical images in sources such as specialized journals or websites. Therefore, the following memos were recorded in the open coding stage:

The participant has a personal image collection containing images from his previous patients, and he prefers to use images from his own collection.

The participant uses images from published articles.

The participant looks for images in related websites.

¹ These codes refer to the interviewees. A list of interviewees and their corresponding codes is given in chapter 4 to which the reader can refer for details.

Following the coding paradigm of Strauss and Corbin (1998), we then compared emerged concepts to establish any relationship between them. At the next stage, as the following example illustrates, we put related concepts labels together to form concepts:

P3: I prefer to use images of my previous patients who gave me their consent.

Memo: The participant has a personal image collection containing images from his previous patients

Concept label: Personal collections

P2: I also use books to find images. For example, if you want to know and see the mechanisms that already have been discovered, you can use the books.

Concept label: Books

P1: But for the special images I go to find them in papers.

Concept label: Papers

These example concept labels indicate that participants looked for images in personal collections, books and papers. We therefore classified them under the category of 'image resources' during axial coding. Then we continued with the analysis to identify the properties of each concept label classified under the category 'image resources'. The following examples illustrate this process:

Axial Code: Image resources

Concept label: Web

Properties of the concept level:

Property 1- Using the internet helps participants to find images quickly

P2: Using the Internet saves time as you can find images quickly.

Property 2 - Participants may find many irrelevant images in the internet.

P3: The Internet is available everywhere, but you have to change or modify the query term and see many pages of images to find the ones that you need. As you know, there are lots of irrelevant images that do not fit your needs.

Property 3 - Participants believe that they may find required images on the internet.

P4: I also believe that through the Internet and image search engines, it is more likely that I find the images I need.

Using grounded theory and its coding paradigm, we could identify other criteria and the core relevance category to address our initial aims and objectives. In this study, the core category was identified as 'topical relevancy' and we concluded the selective coding stage of the grounded theory approach and made a statement as follows: 'health care professionals looked for medical images in different image resources and they used a variety of relevance criteria to judge the relevancy of images to their information needs'. Although, there was no agreement between them on the most important criterion, they judged the relevancy of images based on their information need and the image resources they used.

3.6.8 Schematic of the study

Although we discussed the methods and procedures in grounded theory separately, Strauss and Corbin (1998) state that the steps and procedures in grounded theory are not taken in a linear sequence nor they are separate in practice. Researchers can move back and forth between them. They stated:

We emphasize strongly that techniques and procedures, however necessary, are only a means to an end. They are not meant to be used rigidly in a step-by-step fashion. Rather, their intent is to provide researchers with a set of tools. (Strauss and Corbin, 1998: p.14)

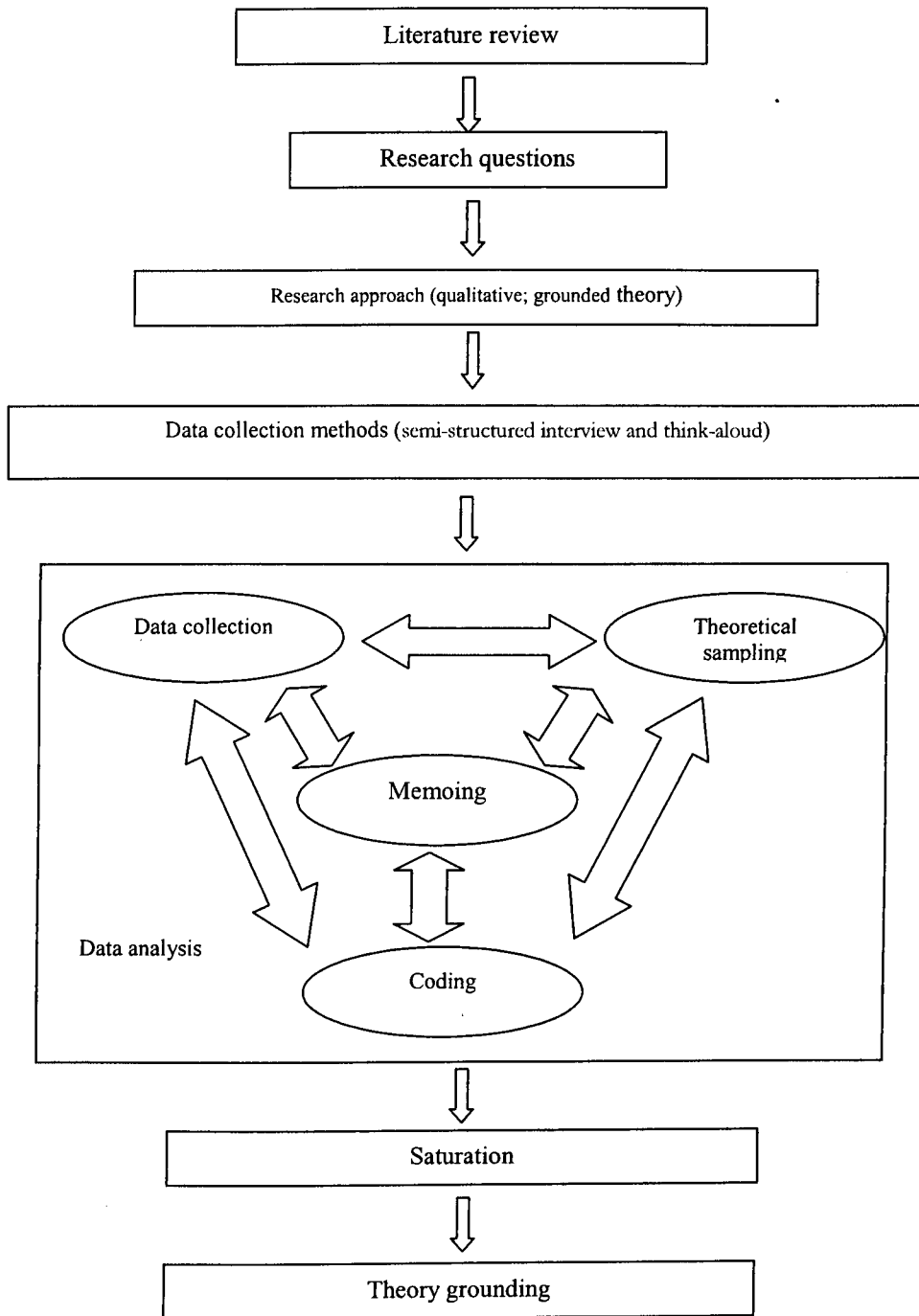


Figure 3-2: The stages of the research process in the current study.

We have discussed the Straussian version of grounded theory and the methodology applied in this study. However, it seems appropriate to summarize the main procedures of the current study. Figure 3-2 depicts the stages and process of the current study. It should be mentioned that we stopped data analysis after we reached saturation. This means no further new concept or category was developed. Additionally in grounded theory, data collection is guided by theoretical sampling.

3.7. Data collection methods

As mentioned earlier, it was important to see the relevance criteria and the judgment situation explained in the health care professionals' own terms. Maglaughlin and Sonnenwald (2002); Ingwersen and Järvelin (2005); Hirsh (1999); Park (1994) state that methods typically used to investigate relevance criteria and related issues include interview, think-aloud, questionnaire and direct observation techniques. Strauss and Corbin (1998) state that they regard interviews and observation as the main tools for data collection in grounded theory. Thus they suggest that once the researcher decides on the research participants, the site of study, and the type of data to be collected, she or he must develop a list of interview questions or areas for observation. Strauss and Corbin (1998: p.204) report that the choice of data collection method depends on the type of data to be collected:

A decision must be made about the types of data to be used. Does the investigator want to use observations, interviews, documents, biographies or combination of these? The choice should be made on the basis of which data have the greatest potential to capture the types of information desired.
(Strauss and Corbin, 1998: p.204)

The types of data we collected from health care professionals were verbal; therefore we selected two common ways of gathering verbal data: the semi-structured interview and the think-aloud approach. We used these methods to identify and describe the relevance criteria applied by health care professionals and their perceptions of the concept of

Relevance criteria for medical images applied by health care professionals

relevance. In the first, and the main, part of this study we used interviews. We also asked the participants to conduct real medical image searches and explain exactly how they went about looking for images, and how they evaluated the retrieved images. In the following sections, each of research methods used in this study is explained in detail.

3.7.1 Interviews

There are two main types of interviews in qualitative forms of research such as grounded theory: the unstructured interview and the semi-structured interview. Sometimes researchers use the term qualitative interview to encapsulate these two types of interview. Although flexibility and variation are the main characteristics of qualitative research, using the qualitative interview maximizes flexibility and variation. (Bryman, 2001: p.319).

In the semi-structured interview the researcher has a list of questions, without optional response categories, which is referred to as an interview guide. Semi-structured interviewing allows the researcher to open up the responses of the participants and direct the interview according to the research objectives. The researcher asks all participants the same questions; however the researcher may not follow the protocol exactly. For example, sometimes the researcher may ask questions that are not in the protocol.

Unstructured interviews include open-ended questions that participants are allowed to respond to freely. In fact the researcher does not control or direct the interview. Unstructured interviews are employed when the researcher wanted to obtain a holistic understanding of the thoughts and feeling of participants.

Regarding the nature of structured and semi-structured interviews, we believed the need was to collect data according to the aims and objectives of the study. Thus, we used a semi-structured interview protocol for data gathering in our study. After we started collecting the data, it became evident that in grounded theory the researcher needs to use semi-structured interviews as described by Park (1994). Park (1993) also emphasizes that in grounded theory the researcher must control and direct the data collection, thus we could not use unstructured interviews in a grounded theory study.

Since grounded theorists prefer to construct (or discover) theory from the data, data collection and analysis proceed simultaneously. The data collection activities are guided by analytic interpretations and guided to focus further observation or interview procedures. Glaser and Strauss called this a constant comparative method to generate grounded theory. (Park, 1994: p.139)

Reviewing the literature showed that the face-to-face semi-structured interview is probably the most widely used method for data collection in grounded theory and other forms of qualitative research. Additionally semi-structured interviews are a common data collection tool in relevance studies. For example, Schamber (1991) uses semi-structured interviews to investigate how participants make relevance judgments. She interviewed thirty professional users of weather information obtained from different resources such as radio and newspapers and asked all participants to talk about one recent job-related situation in which they had needed information about the weather to make a decision or perform a task. The aim (Schamber, 1991) was to understand the underlying meaning of users' evaluation criteria in the context of contextual and multimedia information retrieval.

Using semi-structured interviews, Park (1994) conducts a study and interviewed ten participants in order to understand participants' evaluation criteria of information. She asked each participant their information seeking context, their information needs, and the criteria they applied for relevance judgment of scholarly citations. Park states that semi-structured interviews were used in order to collect users' verbal descriptions about their own thoughts and reasoning behind their perceptions and relevance judgments, for each citation.

Although we believe that a grounded theory study requires the use of face-to face semi-structured interviews, using face-to-face semi-structured interviews in this study had some further advantages:

The researcher could help respondents if they had difficulty answering questions.

There was an opportunity to probe respondents to elaborate on their response.

The researcher was free to ask questions in a flexible way, with respondents answering them in the order the researcher asked them, which made each question independent.

The researcher could collect additional data.

There was little risk of missing data, making partially answered questions less likely.

3.7.2 Think-aloud protocol

We asked all participants to verbalize their thoughts while searching and selecting images. This is known as the think-aloud technique. Using this technique we recorded verbal data about relevance judgment, which is a cognitive process as Ingwersen and Järvelin (2005: p.92) state. Kagolovsky and Moehr (2004) specifically put emphasis on employing the think-aloud protocol in user studies in information retrieval, and stated that the results of think-aloud protocol analysis can be helpful for designing information retrieval systems:

When a "gap" in the IR process is identified, other methods can be used to investigate it. If a "gap" is related to users' involvement in IR, methods of cognitive psychology can be used. For example, although all users have to formulate an information needs statement, choose a strategy, or evaluate retrieved documents, these tasks can be done differently by different users. Therefore, if a "gap" is related to one of these steps, evaluators have to understand how users perform these tasks. One possible approach is a "think-aloud" protocol analysis that is used in cognitive psychology. The results of this type of analysis would allow the creation of information systems that can support commonalities and accommodate differences between users. (Kagolovsky and Moehr, 2004: p.108)

The think-aloud protocol was introduced by Lewis and Rieman (1994). The think-aloud protocol allows researchers to capture a person's cognitive processes while she or he is performing some task of interest. They described this protocol as follows:

The basic idea of thinking aloud is very simple. You ask your users to perform a test task, but you also ask them to talk to you while they work on it. Ask them to tell you what they are thinking: what they are trying to do, questions that arise as they work, things they read. You can make a recording of their comments or you can just take notes. You will do this in such a way that you can tell what they were doing and where their comments fit into the sequence. You'll find the comments are a rich lode of information. (Lewis and Rieman, 1994: p.83)

A review of the literature shows that the think-aloud protocol has been successfully employed in relevance studies. In studies by Barry (1994); Tang and Solomon (1998); Maglaughlin and Sonnenwald (2002); Yang (2005); Tombros et al. (2005) the participants were asked to discuss their reasons for judging the relevancy of document. Using the think-aloud protocol was a crucial part of data collection in these studies, since users described the relevance criteria and the judgment process in their own words. Kagolovsky and Moehr (2004: p.110) note the appropriateness of the think-aloud protocol for relevance studies in the information retrieval field. Kagolovsky and Moehr (2004) declare that if the aim of the study is to investigate how participants judge the relevancy of retrieved documents to their information needs, the relevance judgment process and relevance criteria can be captured using 'think-aloud' protocol analysis:

The user's task is to make a decision about the relevance of retrieved documents to the query. The process of decision making is profiled using methods of cognitive psychology: "think-aloud" protocol analysis. (Kagolovsky and Moehr, 2004: p.110)

3.8. Data collection procedures

3.8.1 Prior to the data collection session

A local research contact from Sheffield Teaching Hospitals NHS Foundation Trust facilitated our access to interviewees. To recruit suitable participants we sent potential participants an invitation letter (see Appendix 2) by traditional postal services. Potential participants also received an information sheet (see Appendix 4), a reply slip indicating their interest in participating (see Appendix 5), and to return these to us, they received a pre-stamped and pre-addressed envelope. The letters of invitation also were distributed via email (see Appendix 3) to subscribers of a Sheffield-based health and biomedical mailing list. The recruitment process began in September 2007 and interested respondents were then selected based on their suitability for this study.

3.8.2 Research population

The research population of this study is the health care professionals at Sheffield Teaching Hospitals NHS Foundation Trust (see chapter 4 for detailed information about the research population of the study). There were certain reasons that the current study focused on health care professionals:

Our understanding of relevance judgment and the relevance criteria for medical images is limited.

The medical image seeking behaviour of health care professionals has received little attention to date.

There is a huge community of health care professionals in Sheffield; the research site selected for this study is one of the largest sites in the UK.

The main researcher was a lecturer at Tabriz University of Medical Science from 2001 to 2005 prior to starting this project. In addition he has done his first and second degrees in medical information and librarianship and worked as a

medical librarian for almost four years before becoming a lecturer. He is thus experienced in helping health care professionals locate information in an academic library and in medical databases.

Since the main researcher's subject background is medical librarianship and he has worked as a lecturer in a Medical Information and Librarianship department, he is familiar with the relevant subject areas.

The reasons for studying health care professionals and the selection of Sheffield Teaching Hospitals NHS Foundation Trust as the site for this study are explained in this chapter and in chapters 1 and 4. In the next section, we will discuss the sampling methods used in this study.

3.8.3 Sampling and recruiting the participants

Although Strauss and Corbin (1998: p.214) believe that in grounded theory concepts "*are indicatives of phenomena and are not counting individuals or site per se*", they suggest that the site of study or the research population must be selected according to the research questions. In the light of the research questions for this study, we therefore selected Sheffield Teaching Hospitals NHS Foundation Trust as the site of our study. In particular we decided to study health care professionals who met the following criteria:

- 1) They worked in the site of our study. Since the research was carried out in Sheffield, we selected the Sheffield Teaching Hospitals NHS Foundation Trust as the site of our study. The Trust includes five hospitals: a) Northern General Hospital; b) Royal Hallamshire Hospital; c) Jessop Wing; d) Weston Park Hospital and e) Charles Clifford Dental Hospital.
- 2) They used any type of medical image.
- 3) They were skilled and knowledgeable internet and computer users.
- 4) They held a degree in health or bio-medical sciences.
- 5) They had access to the internet.

The following health care professionals are excluded from the study:

Relevance criteria for medical images applied by health care professionals

- 1) Health care professionals who did not have an academic degree since we were looking for participants with an academic medical background.

Sheffield Teaching Hospitals NHS Foundation Trust consists of about 11,689 staff (see Table 4-1). The staff includes medical, dental, scientific, therapeutic, technical, nursing, midwifery, and health visiting professionals, health care assistants and other support staff. The sampling method for the interviewing stage of the current study was a purposive non-random sampling and was a theoretical sampling as suggested by Strauss and Corbin (1998). In describing this method, Strauss and Corbin (1998) state that in grounded theory the researchers must collect the data in a non-random purposive, convenient manner and they must continue data collection until they reach data saturation. Strauss and Corbin (1998) declare that theoretical sampling is sampling on the basis of "*concepts that emerged from analysis and that appear to have relevance to the evolving theory*" [Theoretical sampling]. (Strauss and Corbin, 1998: p.202).

Moreover in the previous studies on relevance criteria, the number of interviewees varied between eighteen and forty. For example, Barry (1994) had interviewed eighteen participants. Similar qualitative studies of relevance judgment used thirty-eight (Choi and Rasmussen, 2002), thirty (Schamber, 1991), twenty-six (Yang, 2005) and twelve (Maglaughlin and Sonnenwald, 2002). Based on the results of her study, Barry (1994) believed that the full range of criteria of relevance can be collected through interviewing less than ten interviewees. She also states that a limited number of criteria of relevance are shared across users and situations. This means each user does not apply a unique set of criteria for the relevance judgment process. Barry (1994) emphasizes that the aim of her study was to provide a comprehensive list of criteria of relevance that users applied during the relevance judgment process. Thus the only way to identify these criteria is to inspect the superfluity of responses, and the redundancy will be achieved when no new criteria was mentioned by participants. Barry (1994) adds that in any possible ordering of participants, redundancy was achieved after the ninth participant had been interviewed. In other words, without regard for the order of interviewing the participants, no new criterion was identified once the ninth had been interviewed. This

supports the findings of previous studies in which redundancy of criteria of relevance was reached throughout interviews with less than ten participants.

Taking into account the number of participants in previous studies, and in view of participants' time limitation and availability, it seemed logical to interview a minimum of twenty and maximum of forty health care professionals in this study. However, this is a qualitative study based on grounded theory and in grounded theory the size of research sample is based on theoretical sampling; thus we started the recruitment process and data collection as Strauss and Corbin (1998) suggest.

In order to facilitate and systematize the sampling in grounded theory a number of practical guidelines can be taken into consideration. Strauss and Corbin (1998) recommend three suggestions to apply while sampling. These practical guidelines are useful to help decide how long a researcher must continue with data collection and sampling. We applied them during the data collection and sampling in this study. In response to the question ' *how long a researcher must continue to sample*', Strauss and Corbin (1998) responded as follows:

The general rule when building theory is to gather data until each category is saturated. This means until (a) no new or relevant data seems to emerge regarding a category, (b) the category is well developed in terms of its properties and dimensions demonstrating variation, and (c) the relationships among categories are well established and validated.
(Strauss and Corbin, 1998: p.212)

While interviewing participants, we used a method known as 'Snowball Sampling' for the recruitment of further participants. In this method interviewees are identified based on the recommendations of previous interviewees (Bryman, 2001).

Using the theoretical sampling approach, we continued data collection until we reached saturation in data gathering. After interviewing fourteen participants, we reached data

saturation: however, we continued the recruitment and data collection processes to ensure that we reached data saturation. In total we interviewed twenty-nine participants.

3.8.4 Ethical issues

Punch (2005: p.276) notes that all social research involves ethical issues. This is because the research involves data from people, and about people. Thus researchers should anticipate the ethical issues that might arise, and consider how they will deal with these ethical issues in their studies. He reported that the substantial literature on ethical issues is of two main types. The first type offers researchers guidelines for ethical conduct and a checklist of points to consider. The second type of literature describes what issues have arisen for social science researchers in previous studies, and how they have been dealt with.

Ethics approval to carry out the study was obtained from the NHS¹ National Research Ethics Service and Sheffield Teaching Hospitals NHS Foundation Trust in August of 2007 in order to approach, recruit and interview participants (see Appendix 1). It took about six months to get the approval from the South Sheffield Research Ethics Committee acting on behalf of the NHS² National Research Ethics Service. All participants were aware of the legal issues involved and the need for protection of patients' privacy. Independent research monitoring officers from The University of Sheffield and Sheffield Teaching Hospitals NHS Foundation Trust monitored the study to ensure that neither the researchers, nor the participants, breached the rules. We noticed that participants could access anonymous images through the health information systems they used, and in addition used images in different ways for a variety of reasons (e.g. viewing images from medical websites or electronic journals for clinical purposes). Some participants used patient images only if they had written consent from the patient.

The anonymity and confidentiality of all data was maintained and participants were informed about these issues. There was no disclosure of information and no reference

¹ National Health Service in the United Kingdom

² <http://www.nres.npsa.nhs.uk/contacts/find-your-local-rec/?EntryId11=10435&p=14>

was made in verbal or written form, which could link participants' names to the study: participants are referred to by code number only. All data were collected anonymously respecting participants' confidentiality at all times. In addition, all documents and all recorded data and contact information of participants were stored in a locked drawer at the Department of Information Studies of the University of Sheffield. The audio files of data collection sessions have been saved in secure files and password protected folders. Data were available only to the main researcher and his research supervisors. After the main researcher completes his PhD, all data including participants' contact information, recorded tapes and transcriptions, will be destroyed.

All participants signed the consent letter presented in Appendix 6. This letter informed all participants about the anonymity and confidentiality of data collection and assured them that the data we collected would be used anonymously in the thesis, and in any publication resulting from this study. The participants have also been informed that the data collection session does not include any sensitive issues, and that they were free not to respond to any question if they did not wish to. The participants could also request the researcher to stop the interview at any time, and could withdraw from the study without having to give any reason. Before starting the interview, participants were asked if they felt upset or distressed.

3.8.5 Time and place

We interviewed participants during October 2007 – March 2008. We contacted potential participants to arrange for the venue, date and time of the interviews. The interviews with health care professionals who consented to take part in the study were carried out in mutually agreed venues, at mutually agreed dates and times. Most of the interviews were carried out in the offices of participants, and with participants' permission we recorded the interviews using a digital voice recorder. For two participants who did not have a suitable room, a room was booked in the Department of Information Studies. The interviews were carried out face-to-face by the main researcher and the mean interview duration was approximately 42 minutes (range 28 to 92 minutes).

3.8.6 During the data collection session

At the beginning of data collection, we again outlined the aims and objectives of the current study to the participants, and the voluntary nature of their participation. Doing this was also found to make the atmosphere friendlier and give the participants time to ask questions before signing the written consent form (see Appendix 6).

We reassured participants that their anonymity would be protected at all times and access to their data would be managed carefully. They were informed that we are the only people who could access the data. No identification was recorded on the audiotape or transcribed interviews. All participants received a copy of the signed consent form, and with their permission we recorded the whole interviews including the medical image search sessions with a digital voice recorder. Using a digital voice recorder facilitated the follow-on transcription work for the main researcher.

Interviews were conducted in a range of locations and the format of interviews was semi-structured (see Appendix 7). The questions focused on how participants search for and select medical images they need. All interviews started with the following question: 'Tell me about your profession and work experience?' This question was followed by a question about the frequency of their medical image searching. After the participants had said what they wanted to say, we asked them additional questions as planned.

3.8.7 After data collection

A hallmark of the ground theory is the simultaneous data collection and data analysis as Strauss and Corbin (1998) suggest. Thus, we transcribed each interview immediately after the interview session. As we stated earlier, in grounded theory data collection and data analysis are based on theoretical sampling, meaning we must commence data analysis as soon as we collect the first bits of data. Theoretical sampling could not occur unless the researcher transcribes each interview immediately after the interview session.

Strauss and Corbin (1990) and Glaser (2002) state that in qualitative research, the researcher should transcribe all tapes recorded during the interviews. However,

according to Glaser (2002); Strauss and Corbin (1990) in grounded theory the researcher may also need to transcribe earlier interviews.

Although Glaser (2002); Strauss and Corbin (1990) suggest that the transcription in grounded theory is selective, the main researcher transcribed the audiotapes verbatim. Sometimes he consulted native English-speaking postgraduate students in the Department of Information Studies and School of Medicine and asked them to explain medical and local expressions used. To examine the accuracy of transcriptions, we sent the transcribed text interviews to participants. This helped us to get participants' confirmation of the accuracy of the transcription. Additionally, we received further explanation and recommendations from health care professionals. This was also ethically helpful, because the health care professionals had another opportunity to amend any part of their interviews as desired, or to clarify what they had said previously.

3.8.8 Presenting the interview data

A number of quotations from the participants are included in chapters 3, 5 and 6 to substantiate the discussions. Although the transcription of each interview was verbatim, the direct quotations from the interviewees might seem a little disjointed because of use of some repetitive words, incomplete sentences, or redundant catch phrases such as 'well', 'wow', 'you know', 'I mean' and 'you see'. In order to quote from interviewees concisely and for the sake of brevity, we removed some repetitive or unnecessary words or phrases.

3.8.9 Data collection protocols

The data collection took place over approximately five months. We used interview and think-aloud protocols in order to investigate participants' perspectives on the relevance criteria and relevance judgment process for seeking medical images. Most of the time, participants conducted medical image searches during the interview to explain their image search behaviour, relevance judgment processes and other activities. In other words, interviews and think-aloud sessions were not separated. As participants responded to our questions and conducted searches for medical images, we were able to

get additional insights and information from them regarding research aims and objectives. Therefore, we considered the data transcribed from the medical image search sessions as part the interview sessions, and did not analyse the data obtained during the two sessions separately. We now discuss the protocols used.

3.8.9.1 Interview protocol

In the first stage of the study, we used the interview protocol presented in Appendix 7 and asked health care professionals to describe one of their recent medical image needs. We developed the initial questions of the interview protocol based on the aims and objectives of our study. Additionally, the experiments obtained from each interview helped us to revise the interview protocol. Thus the questions in the interview protocol became directed and more focused through theoretical sampling (Strauss and Corbin, 1998: p.203).

The interview protocol includes nine sections (see Appendix 7). At the beginning of the interview we asked health care professionals about their background, departmental affiliation, the nature of the activities they undertook, and the type of work they were normally involved in. Then the participants were asked about the frequency of medical image searching, and were asked to describe one recent job-related situation in which they needed to find some medical images. In the third section of the protocol we asked participants to talk about their favourite medical image sources. Then the interview was continued covering the health care professionals' specific medical image needs, image searching queries, medical image selection process, and their final medical images uses. Since the interview was a semi-structured interview, we constantly checked whether all the questions on the protocol had been responded to.

3.8.9.2 Medical image search with think-aloud

The goal of our study is to provide insight into the cognitive process of relevance judgments for medical images and the interactions of health care professionals during the judgment process. The focus of this study was not on the outcome of the relevance judgment process, but rather on the process itself: the relevance criteria that health care

professionals applied and the attributes of medical images on which those criteria were based.

Before starting the data collection, a brief training session about think-aloud was given to the participants. The participants were informed about our interest in the way they formulated queries and selected medical images, but that there were not specific guidelines for participants on what we wanted them to talk about or how they should express their thoughts. We asked each participant to perform a series of medical image searches based on information needs that were developed independently by participants and that they discussed in the interview session. Participants were allowed to search for medical images in any way they found useful or natural. We encouraged them to search as they would normally. Participants could use any available image search tools such as image search engines, personal knowledge of the image sources (e.g. useful medical image collections), or databases. Therefore, there was no restriction placed on the search strategy that health care professionals would apply as suggested by Tombros et al. (2005). While searching for images, we prompted participants to keep up the flow of comments and discuss why they chose a specific image and what attributes that choice was based on (Lewis and Rieman, 1994: p.84). We recorded their comments and believe that the inclusion of think-aloud in our data collection stage provided insights into the cognitive relevance judgment process that could not be achieved in any other way.

3.9. Analysing data

Using a grounded theory research approach, we tried to characterize the relevance criteria that health care professionals use when making relevance judgments. There are some software packages available for qualitative data analysis. Using such software enables the researcher to search the assigned codes for patterns; and to establish categories of codes that reflect testable models of the conceptual structure of the underlying data as reported by Lewis (2004). Based on his experiments during the preliminary study, the main researcher chose the NVivo software package.

3.10. Preliminary study

It was crucial for us to conduct a preliminary study for several reasons. Firstly, we carried out a preliminary study to ensure the feasibility of the main study, to examine the data collection tool, and to assess the data analysis strategy selected for our research. It was also an opportunity to develop interviewing skills and to prepare for the main study.

Moreover, we believed that the findings of such a study would help us to clarify the research topic, and its aims and objectives. At the preliminary stage, we wanted to know how health care professionals search for and select images they need. In addition, we wanted to know which is the main source of medical images for health care professionals. We were looking to discover whether health care professionals use specific medical image collections to find images, how they begin their search for medical images, and how they assess the relevancy of medical images for their information needs. Are the relevance judgments and the relevance criteria for medical images different from relevance judgments for non-medical images suggested in the literature?

Between May and June 2007 health care professionals in medical departments at the University of Sheffield were asked to participate in the preliminary study. Seven volunteer interviewees participated and the mean average of the work experience of the participants was 7 years 7 months (range 2 to 21 years). The results showed that they used a variety of resources to find the required medical images. They would first rely on articles and the internet (image search engines) for finding medical images. In addition, they look for medical images almost every day or sometimes more than once a day. Interview was the data collection method used and the mean interview duration was \approx 50 min (range 37–92 min). Table 3-3 shows profiles of participants interviewed and their characteristics.

Table 3-3: Profiles of participants in preliminary study
(PhD: Doctor of philosophy; DDS: Doctor of dental surgery; MD: Doctor of medicine; MSc: Master of science; AD: academic degree; IL: interview length; WE: work experience; and ID: interview date).

ID	Participants' Speciality	AD	IL	WE	ID
P1	Dentist, Dental Materials	DDS	50	4	29/05/07
P2	Molecular Genetics	PhD	57	6	13/06/07
P3	Orthopaedic Surgeon	PhD	92	12	06/06/07
P4	Immunologist	MSc	58	4	19/06/07
P5	General Surgeon	MD	47	5	20/06/07
P6	Consultant of Orthopaedic Surgery	PhD	45	21	26/06/07
P7	Stem Cell researcher	PhD	37	2	24/06/07

3.10.1 The findings of the preliminary study

From the interviews and think-aloud sessions and using grounded theory, twenty relevance criteria (concepts) were elicited from participants. We grouped them into two categories: *medical and photographic* (Figure 3-3). The findings of our preliminary study indicated that image quality, orientation and topical relevancy were the three most frequently-used relevance criteria, and these criteria were used by six out of the seven interviewees.

The participants applied different criteria as the most important and first relevance criterion. Topical relevancy of images was the most important criterion for three of the participants; modality was the most important criterion for two participants. One participant stated he would consider the photographic quality of medical images and another participant stated that the importance of a criterion depends on what is required.

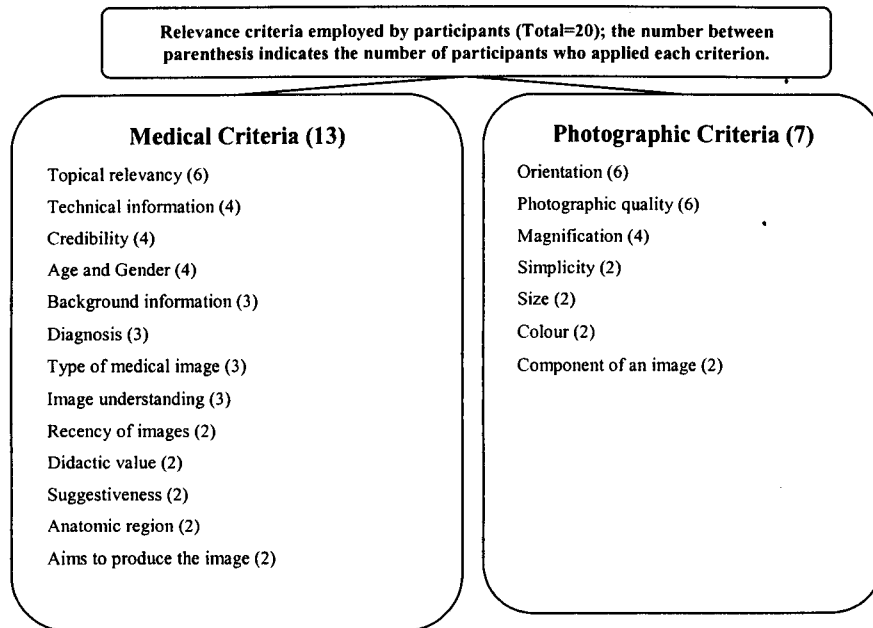


Figure 3-3: The relevance criteria identified in the preliminary study.

Other findings of our preliminary study indicated that health care professionals used a variety of resources including academic papers, the internet (image search engines), books, personal collections, friends and colleagues, departmental collections, CDs and DVDs to find medical images they needed (see Table 3-4). Amongst them academic papers and the internet were used by all participants to find medical image they needed. All of them were familiar with general image search engines such as Google.

Participants stated that the number of irrelevant images retrieved by search engines is often very large and that they were unable to check all retrieved images and select the images they need. Therefore, they stated that they usually first searched for relevant papers in medical databases such as PubMed¹ and then looked for medical images within papers. They added that the text of papers explain and clarify various aspects of the presented objects in the images.

¹ <http://www.ncbi.nlm.nih.gov/pubmed/>

Table 3-4: Image resources used by the participants of preliminary study.

Medical image resources	Number of participants
Papers	7
Web	7
Books	4
Personal collection	4
Friends and colleagues	3
Departmental collection	2
CDs and DVDs	2

We also went on to examine how images are used by health care professionals. The results of the preliminary study showed that health care professionals required images for the following purposes: patient care, education, research, publication and documentation.

3.11. Trustfulness and replicability

According to Punch (2005) the trustfulness and replicability of the data are important criteria for judging the quality of collected data, evaluating the outcomes of data analysis and concluding of empirical studies.

Glaser and Strauss (1967) state that results of a study based on grounded theory should meet four central criteria in order to be trustful and replicable: fit, understanding, generality and control. Fit entails that the theory should be composed of components that correspond to the 'everyday reality' of the research area that is studied and that has

been developed from data. In other words, developed categories must fit the data. Understanding entails that the theory should be comprehensible either to the researcher or participants involved in the area of study. Generality means that the theory can be applicable in similar contexts. Control entails that the theory should provide control of components that relate to the phenomenon studied. Although we applied the criteria suggested by Glaser and Strauss (1967), replicability data in this study has been ensured throughout by using the think-aloud data collection technique. Using think-aloud also decreased the risk of bias in the results as we asked all participants to explain their medical image needs in their own words, including their search experiences as well as the relevance criteria for medical images in a real context.

In addition the grounded theory method used in this research raised the trustfulness of the findings the study, because the findings emerged from the data and not from examination of preconceived ideas or theories predating the data analysis. Moreover, we adhered closely to all procedures and rules of grounded theory for the recruitment of participants, data collection, data analysis such as theoretical sampling, open coding, selective coding, axial coding, saturation, and memoing to ascertain that all concepts and categories emerged from the data. This enhanced the trustfulness and replicability of the data and the findings of this study. To validate the coding the main researcher asked a grounded theory researcher¹ to check the assigned codes to four interviews. This method is known as double-coding (Gilbert, 2001). Since both researchers were in agreement, the trustfulness and replicability of the data were further confirmed. Additionally, the data collection and analysis based on theoretical sampling allowed us to document any changes to the questions of interview protocol.

3.12. Limitations

The study was constrained by some methodological limitations. Generally, due to time and resource constraints of the research project, this research is limited to health care professionals in a particular geographic area. Differences exist among health care

¹ This person has graduated and he used grounded theory in his PhD project.

professionals in terms of speciality, roles, academic degree and information needs. The findings of this study cannot be entirely generalised to all health care professionals.

Additionally the concept of relevance is not a simple concept. It is affected by several factors. For example, the role of health care professionals, their image information needs and image resources used to find medical images might affect the relevance criteria applied by health care professionals. Thus, one needs to be cautious in drawing conclusion and generalisations from the findings of the current study even when the generalisations are restricted only to health care professionals.

There were also limitations and bureaucratic problems in recruiting health care professionals: it took six months to get the ethics approval and permission from the NHS¹ National Research Ethics Service and Sheffield Teaching Hospitals NHS Foundation Trust to approach health care professionals. In addition, it was not easy to find and access potential participants for the study. The number of participants did not limit this research since we reached data saturation after we interviewed fourteen participants. Nevertheless, findings from this study might be applicable to similar populations, their image information needs, and users of similar image resources and subject areas. Since we recruited health care professionals from different departments, a repeat of this study, with participants from specific medical department could enrich our understanding.

Summary

This chapter has provided a justification and description of the proposed research methods in this study. It outlined an explanation of two major research methods: quantitative and qualitative. We have concluded that grounded theory, one of the widely-used qualitative research techniques, could be used in this study. Moreover, the data collection method of semi-structured interviewing was justified. The method of data analysis, with special attention to the Straussian version of grounded theory, was described and how the findings will be analysed was documented. In addition, a

¹ National Health Service in the United Kingdom

rationale and description of how study participants were selected and recruited was discussed.

CHAPTER 4-DEMOGRAPHICS OF THE SAMPLE

Introduction

This chapter reports the demographic characteristics of the study population and the demographic characteristics of the participants in the current study. The data collection techniques and findings are presented in chapters 3 and 5 respectively. This chapter gives further details of the study population in the form of tables and figures.

The data presented here provides some contextual information that helps to define who participated in the study and aids our understanding of the results presented in the following chapters. Although the information covered in this chapter could have formed a section of the next chapter (chapter 5: Results), we thought that devoting a separate chapter to the demographics of the sample would make it easier for readers to understand the findings.

4.1. Sheffield Teaching Hospitals NHS Foundation Trust

Sheffield Teaching Hospitals NHS Foundation Trust was formed on 1st of April 2001 and manages the five NHS hospitals in Sheffield including Northern General, Royal Hallamshire, Jessop Wing, Weston Park and Charles Clifford hospitals. Each year, more than 175,000 operations and day case procedures are performed and over 825,000 outpatient appointments are provided across the hospitals which together offer almost every kind of treatment available through the NHS (Sheffield Teaching Hospitals NHS Foundation Trust, (2007)).

Although the majority of admitted patients are from Barnsley, Doncaster, Rotherham, Sheffield and parts of North Derbyshire, around five per cent of patients are from other parts of the United Kingdom. The patients come to Sheffield Teaching Hospitals for specialist treatments, many of which are offered in only a few NHS Trusts in the UK. Specialist services of the Trust include:

Weston Park Hospital: one of only three dedicated cancer hospitals in the UK.

Sheffield Cardiothoracic Centre: a regional centre of excellence for heart services.

Princess Royal Spinal Injuries Unit: a unit for the treatment of spinal cord injuries and associated illnesses that treats patients from around the country.

Sheffield NHS Immunology and Allergy Service: the department provides a National External Quality Assessment Service (UK NEQAS) and is responsible for monitoring the quality of laboratory services in Immunology and Clinical Chemistry in the UK and Europe.

Ophthalmic Services: the Trust hospitals run the largest specialist ophthalmology service (for the treatment of eye conditions) in the region. Specialist services include the treatment of patients with cancer of the eye and services for patients with low vision.

Neurosurgery: the Trust has one of the best neurosurgical teams in the UK who treat patients with complex diseases and injuries to the brain.

The Skull Base Group: a multi-disciplinary team made up of specialists from different fields such as maxillofacial surgery, neurosurgery, head and neck surgery and ophthalmic surgery, who work together to give patients with the most difficult to treat illnesses the best treatment and quality of life possible.

Sheffield Teaching Hospitals is called a 'Teaching Hospital' because of its close association with academic centres. The strong commitment to teaching and research, with close links to the University of Sheffield, Sheffield Hallam University and other learning establishments, has established the Trust as a national and international centre of excellence: new treatments and services pioneered in Sheffield have changed the face of medicine across the country and the world. In 2008 a major new Picture Archiving and Communication System (PACS) was opened which facilitates the storage of, and access to, diagnostic images by clinical staff. PACS enables all images to be stored in such a way that doctors, nurses and other health care professionals can instantly access a patient's X-rays, MRI scans or other images from computer terminals across the

Relevance criteria for medical images applied by health care professionals

hospitals. As Strauss and Corbin (1998) suggest it is important to obtain permission from appropriate authorities to use the site of study and recruit participants. Thus, permission was sought from the Research Department of Sheffield Teaching Hospitals NHS Foundation Trust to access the site and approach participants (see Appendix 1).

**Table 4-1: Average number of persons employed in Sheffield Teaching Hospitals NHS Foundation Trust in 2007
(from Sheffield Teaching Hospitals NHS Foundation Trust, (2007)).**

	Total	Permanently Employed	Other
Medical and dental	1,380	1,340	40
Administration and estates	2,418	2,306	112
Health care assistants and other support staff	1,269	1,269	0
Nursing, midwifery and health visiting staff	4,738	4,556	182
Scientific, therapeutic and technical staff	1,883	1,869	14
Total	11,689	11,340	349

Table 4-1 shows five main categories of people who worked in Sheffield Teaching Hospitals NHS Foundation Trust in 2007. The table details the number of Medical and dental, Administration and estates, Health care assistants and other support staff, Nursing, midwifery and health visiting staff and Scientific, therapeutic and technical staff in post in 2007. Table 4-1 also shows that the Trust employs nearly 12,000 staff in a wide range of occupations and professions. Amongst the employees of the Trust, 1,380 people were medical and dental staff. Medical and dental staff and scientific, therapeutic and technical staff were included in this investigation. The reason that we have excluded the other staff from the study relates to the aims and objectives of the study mentioned in section 3.8.2.

4.2. Participants' profile

Using a qualitative study, grounded theory was used to identify and describe relevance criteria applied by health care professionals and their perceptions for medical image retrieval systems. To recruit suitable participants, letters of invitation were distributed via email to subscribers of a Sheffield-based health and biomedical mailing list and by traditional postal services. Interested respondents were then selected based on their suitability for this study. A local research contact from Sheffield Teaching Hospitals NHS Foundation Trust facilitated our access to interviewees.

Unlike previous studies (i.e. Markkula and Sormunen, 2000; Choi and Rasmussen, 2002; Hung et al., 2005), we did not focus on a particular image collection or image retrieval system. By contrast, participants were asked to specify (and conduct) medical image searches as typically carried out in their day-to-day activities. Example topics chosen by the participants included: a pathologic image of a biopsy of bone marrow, MRI scans of the meniscus tear in the knee and microscopic images of cartilage injuries in children. Two participants did not have access to the internet at the time of the interview, and so were simply asked to describe their searches and the relevance criteria they had applied.

In grounded theory, sampling and recruitment of the research participants is based on the theoretical sampling: i.e. the selection of participants is based on the initial analysis of the findings. The recruitment process continues until the researcher reaches data saturation. In total, twenty-nine health care professionals participated in our study. Bearing in mind the difficulties in accessing and interviewing health care professionals and the lack of time mentioned by those who were contacted by the researchers, this is a reasonably good rate of participation by health care professionals and suitable sample size. (Data saturation in this study was achieved after fourteen people were interviewed, though we continued the data collection until we were sure that no new data would emerge. It was based on the personal recommendations of two senior researchers in the Department of Information Studies at the University of Sheffield). With the participants' permission, we recorded the whole interview during the medical image search sessions.

Table 4-2: Profiles of participants.
 (AD: academic degree; M: male; F: female; IL: interview length; WE: work experience years;
 FRCOG: Fellow of the Royal College of Obstetricians and Gynaecologists).

ID	Speciality of participants	AD	Gender	IL	WE	Participants' roles
P1	Dental materials	PhD	M	50	4	Research and Educational
P2	Molecular genetics (genetics and immunology)	PhD	M	57	6	Clinical and Research
P3	Orthopaedic surgeon	PhD	M	92	12	Clinical, Educational and Research
P4	Immunologist, Molecular immunology	MSc	M	58	4	Research
P5	General surgeon	MD	M	47	5	Clinical and Research
P6	Sport medicine/Consultant orthopaedic surgeon	PhD	M	45	21	Clinical, Educational, Research and Administrative
P7	Stem cell	PhD	M	37	2	Clinical, Educational and Research
P8	Molecular medicine and Female infertility	PhD	M	51	7	Clinical and Research
P9	Bone metabolism researcher	PhD	F	53	17	Research and Educational
P10	Superintendent radiographer	PhD	F	38	16	Clinical, Educational Research and Administrative
P11	Virologist	PhD	F	43	3	Research
P12	Epidemiologist, Non-clinical lecturer	MD	M	33	22	Research and Educational
P13	Neurologist	PhD	F	38	5	Clinical, Educational and Research
P14	Medical physicist	PhD	M	33	15	Clinical and Research
P15	Radiologist	PhD	M	42	16	Clinical, Educational, Research and Administrative
P16	Nuclear medicine	PhD	M	28	17	Clinical, Educational and Research
P17	Medical physicist	PhD	M	42	7	Research and Educational
P18	Nuclear medicine	PhD	F	38	11	Clinical, Educational and Research
P19	Medical physicist	PhD	M	35	35	Clinical, Educational, Research and Administrative
P20	Consultant haematologist	PhD	M	31	25	Clinical, Educational, Research and Administrative
P21	Obstetrician gynaecologist	FRCOG	F	34	10	Clinical and Research
P22	Gynaecologist	MSc	F	39	7	Clinical and Research
P23	Haematologist	PhD	F	28	18	Clinical, Educational, Research and Administrative
P24	Obstetrician gynaecologist	PhD	M	36	30	Clinical, Educational, Research and Administrative
P25	Neurologist	PhD	M	30	2	Clinical and Educational
P26	Human reproduction and development biology	PhD	M	29	19	Research, Educational and Administrative
P27	Molecular medicine and Female infertility	MD	M	55	15	Clinical and Research
P28	Medical physicist	PhD	M	32	20	Clinical, Research and Administrative
P29	Nephrologists	MD	M	53	15	Clinical and Research

Relevance criteria for medical images applied by health care professionals

As Table 4-2 shows, all participants, with the exception of two of them (P4 and P11), mentioned that they had several roles at the time we interviewed them. The duration of the interviews varied between 28 and 92 minutes (the mean average of interview duration was 42 minutes). The work experience of participants ranged between 2 and 35 years (with the mean average of 13 years and 3 months of work experience). Table 4-2 also indicates that only eight of the participants were female staff which reflects the make-up of health and biomedical departments in Sheffield Teaching Hospitals NHS Foundation Trust: they are mainly male-dominated.

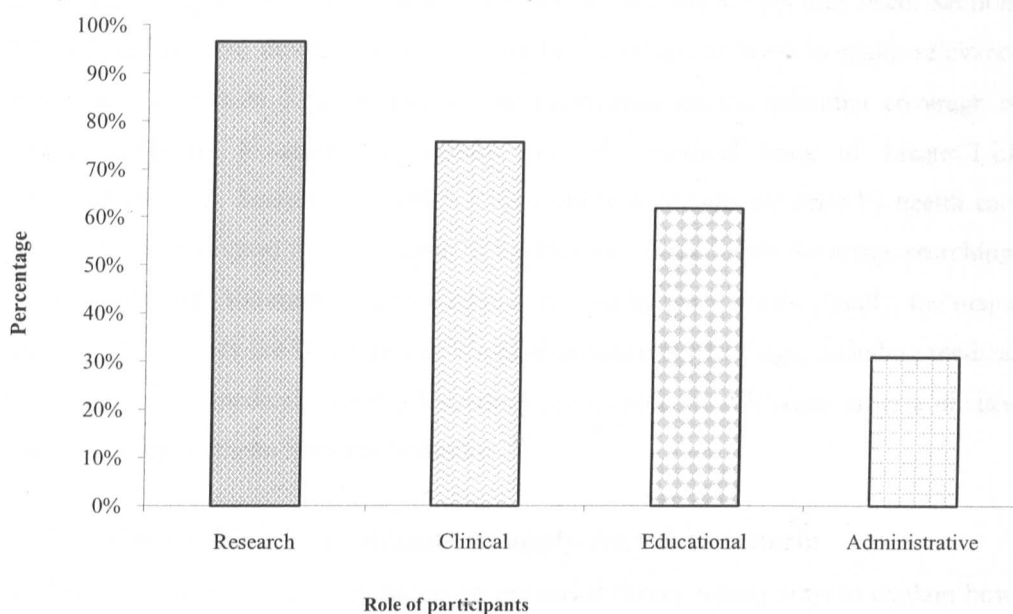


Figure 4-1: Distribution of roles among health care professionals interviewed

(Total=29): (the total is more than 100 per cent as some participant mentioned having more than one role).

Figure 4-1 shows the distribution of specific roles of health care professionals. Twenty-eight (96.55%) of the participants were health care professionals with research roles, and 31% of participants (nine out of twenty-nine) had administrative roles.

Chapter 5 now presents the detailed findings of the study.

CHAPTER 5-RESULTS

Introduction

Twenty-nine participants from various medical departments were studied using semi-structured face-to-face interviews, and with them conducting medical image searches using the think-aloud approach. We believe that the most logical way for presenting and structuring our results is to base this chapter on the study's research questions, its aims and objectives, and the issues that have been examined. Our findings are therefore presented in the eight sections below, as follows. Section 5.1 describes how health care professionals apply relevance criteria and select the medical images they need. Sections 5.2, 5.3 and 5.4 are devoted to the criteria that participants used to make relevance judgments. In section 5.5, we discuss our experiment on the potential coverage of relevance criteria to search statements from the medical track of ImageCLEF (ImageCLEFMed). Sections 5.6 and 5.7 illustrate how images are used by health care professionals and report the motivation of health care professionals for image searching. Section 5.8 details the medical image resources used by participants. Finally, the major findings of the study are summarized. We also compare our findings, including medical image resources, medical image information needs, and the relevance criteria elicited from our study, to those from the literature.

5.1. How health care professionals apply relevance criteria

As stated earlier, this is a study based on grounded theory which aims to explain how health care professionals search for and select images they need. However, before we describe relevance criteria in detail, we discuss how health care professionals search and apply medical image relevance criteria.

The findings of our study showed that medical image searches consisted of multiple sessions. Although participants conducted multiple interactions, in particular when they used Google image search, we found that each interaction was related to the same image information needs. We also noted that query reformulation and image browsing was an essential strategy used by health care professionals when they looked for images. According to our observations, health care professionals reformulated their queries after

they browsed through the retrieved images. The reformulation of the query, and interaction with the system, related to various factors. These were, for example, the number of retrieved images, the proportion of relevant and irrelevant images, and the purpose for the participants' image searches.

Table 5-1: Examples of queries used by the health care professionals.

Denture base materials	Amalgam restoration
Bone marrow and immunity deficiency	Bone marrow severe congenital neutropenia
Bone marrow biopsy immune deficiency	Cartilage injuries in children
Closure bone epiphysis	Foreman skull closure
MRI cartilage ligaments of the knee	Monteggia fracture MRI
Interleukin-4 electrophoresis	Interleukin proteins derived by mast cell
Allergy phenomena and mast cells	Mast cells and an allergic reaction
Anterior Cruciate Ligament of the Knee	A photo of Anencephalic embryo
TLR3 (Toll-like receptor 3) receptors	Bone metastasis
A diagram of how osteoclasts are formed	Meningioma MRI scans T3
MRI aneurysm	Structure of the enveloped proteins
Diabetes and Retinopathy	'Hot cross bun' MRI
Lung, head and neck PET scans	Cardiac 3D CT images
Selenium detector mammography	PET scans for myocardial perfusion hibernating
T-2 weighted MR images of brain	MIBG scan image for tumour in adrenal glands
Mentalis muscle interior view	Red blood cells
Epidemiology images and ultrasound scans of polycystic ovarian	PCO ultrasound images
Immunostaining on endometrial tissue CD-56	Recurrent miscarriage
Symptoms of bleeding or clotting disorders	Queen Victoria family tree hemophilia
Photograph of American meadow vole	A picture of a crocodile newly hatched
Multiple sclerosis	Microscopic image of a sperm approaching an egg

In each search session, the participants tried to find relevant images using particular query terms or a particular search strategy. We noted that health care professionals paid

much attention to selecting search terms, and they always changed or formulated the queries based on the search results.

As Table 5-1 shows, the participants used medical terms and used single word or single phrase topics. The following examples (Figure 5-1, Figure 5-2 and Figure 5-3) illustrate a search. A senior clinician and lecturer in the medical physics department explained that he had searched for 2D and 3D nuclear images for the topic ‘myocardial perfusion hibernating’. He initiated the search process with a single phrase topic which was ‘myocardial perfusion hibernating’.

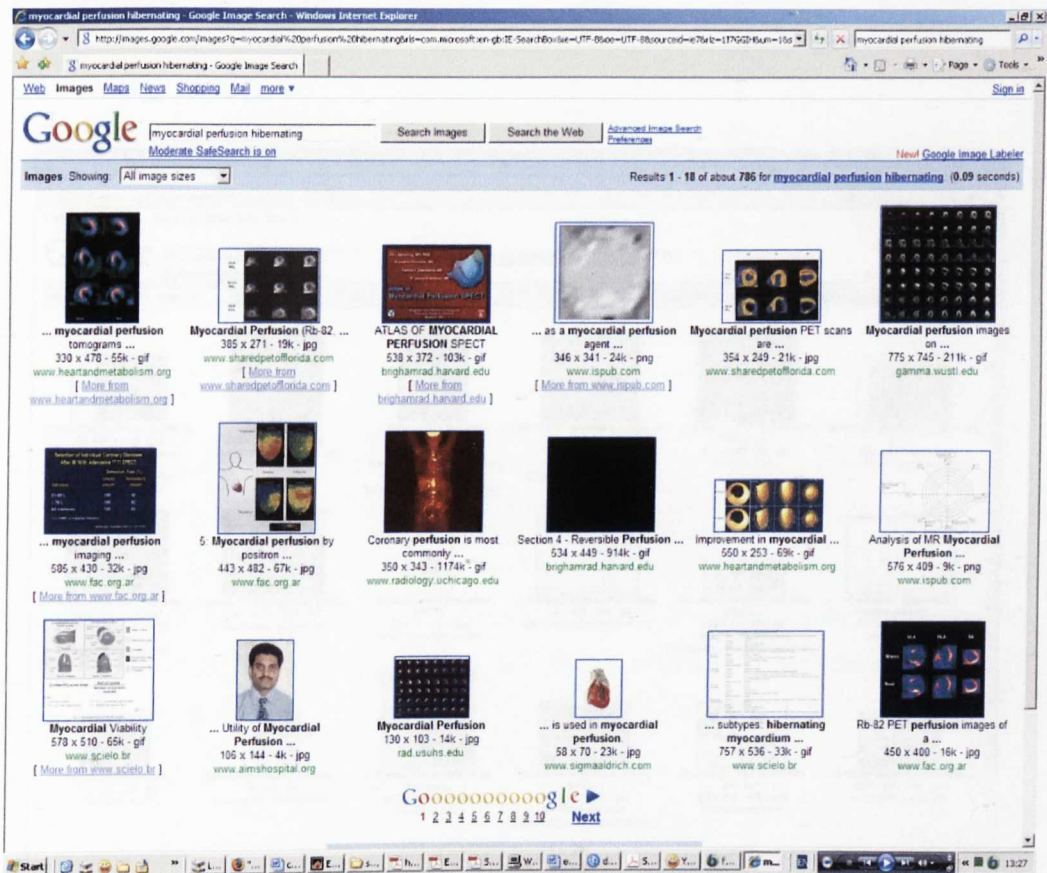


Figure 5-1: Google image search retrieved 786 images for the initial query ‘myocardial perfusion hibernating’ submitted by P16.

He started browsing the images obtained for the initial query ‘myocardial perfusion hibernating’ (Figure 5-1). He stated that few images seemed to be relevant to the topic. Therefore he decided to refine the search results by adding an additional query term. He added ‘PET’ (Positron Emission Tomography) and changed the query to ‘myocardial perfusion hibernating PET’.

After he looked at the results (see Figure 5-2), he was not satisfied with the high number of images and the relevancy of the images obtained. Thus, he added another term which was ‘F18’ (Fluorine-18). He mentioned that F18 is a tracer used to produce nuclear medicine images such as PET scans.

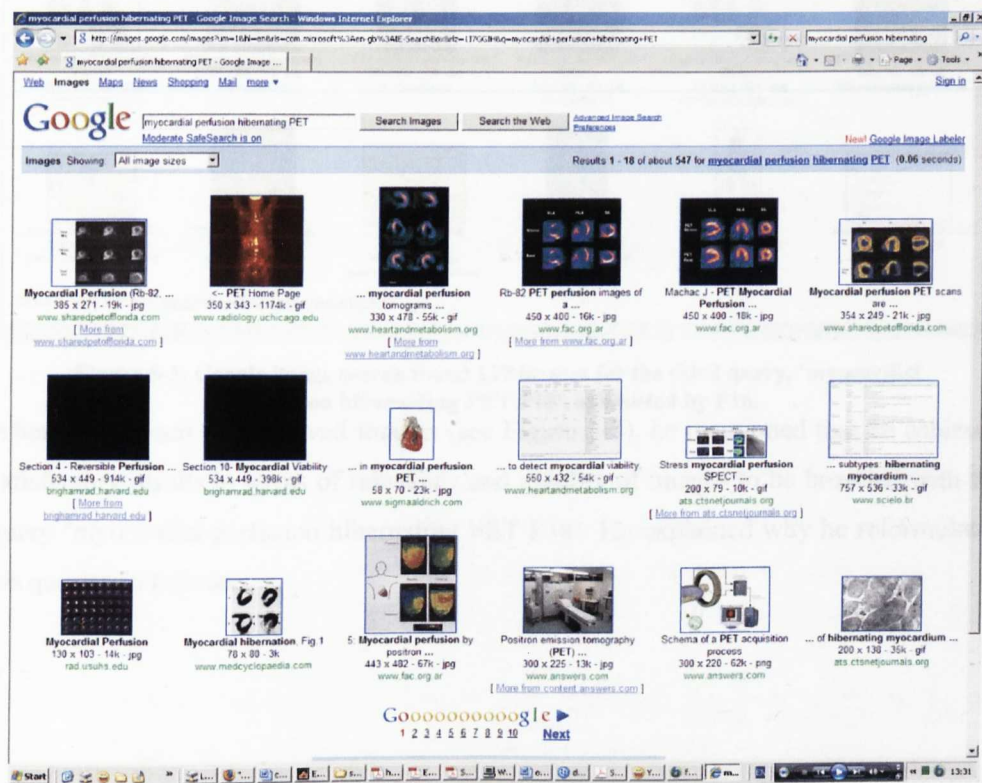


Figure 5-2: Google image search retrieved 547 images for the second query, ‘myocardial perfusion hibernating PET’, submitted by P16.

Relevance criteria for medical images applied by health care professionals

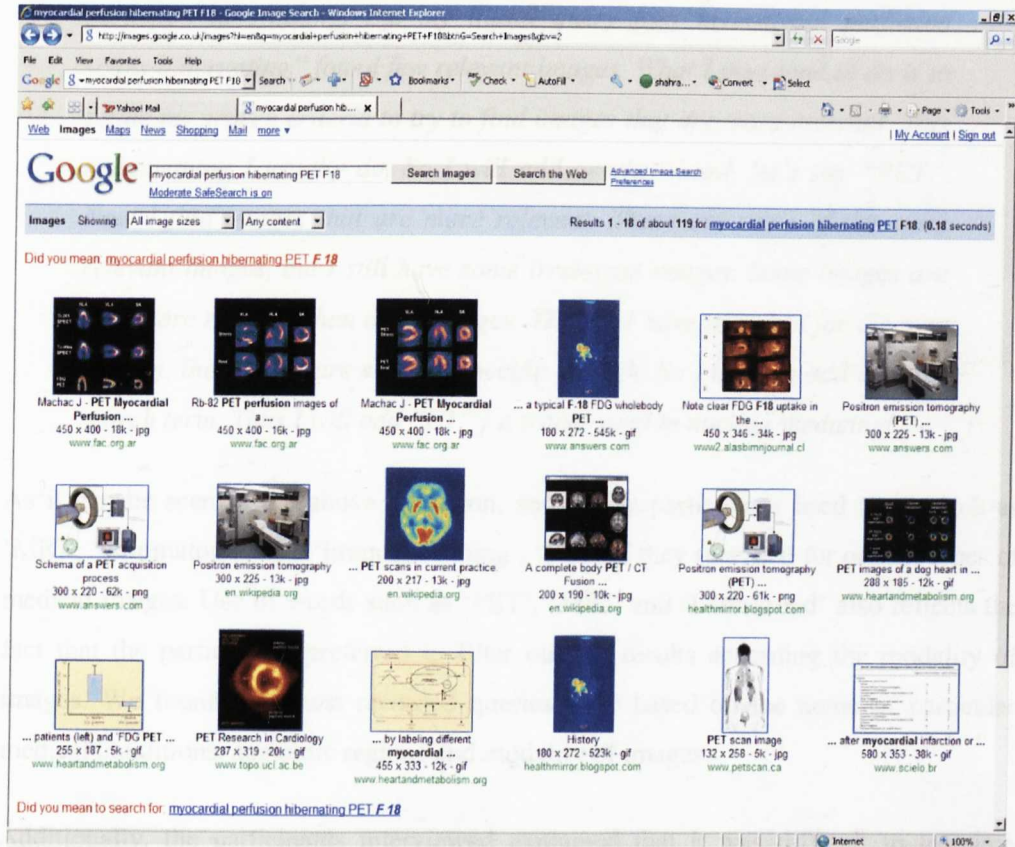


Figure 5-3: Google image search found 119 images for the third query, ‘myocardial perfusion hibernating PET F18’, submitted by P16.

After he browsed the retrieved images (see Figure 5-3), he mentioned that he achieved satisfactory results in terms of relevancy and number of images to be browsed with the query ‘myocardial perfusion hibernating PET F18’. He explained why he reformulated his queries as follows:

P16: Last time, I was looking for some nuclear medicine images; standard 2D plain and 3D nuclear images. These were PET scans for myocardial perfusion hibernating. But my initial query for “myocardial perfusion scans hibernating” found few relevant images. What I then tend to do is to refine my search criteria to try to find images that are more relevant. This is something I usually do. So I will add another word, let's say “PET”. Now I find images that are more relevant. These are some of the more relevant images, but I still have some irrelevant images. Some images are also more relevant than other images. Though I have searched for the type of scan, the images are still not specific enough. So I need to add another search term. Thus I will add “F1” [a tracer used in nuclear medicine].

As it can be seen in the above quotation, sometimes participants used terms such as ‘MRI’, ‘chromatography’, ‘immunostaining’, ‘PET’ if they searched for certain types of medical images. Use of words such as ‘PET’, ‘MRI’ and ‘Ultrasound’ also reflects the fact that the participants preferred to filter out the results according the modality of images. We found that most recorded queries were based on the name of particular medical conditions, anatomic regions and modality of images.

Additionally, the participants interviewed explained that it was difficult to translate visual information into a textual query when they searched for medical images:

P27: Text [text-based image search] is the only available method to search for images. When you are looking for an image, you know what it is or what it looks like. However, text cannot convey your visual imagination properly.

Our findings show that the results of text-based image searches may be affected by the use of abbreviations. For example, one of the participants who was looking for images of the tool-like receptor 3, known as TLR-3 in the medical literature, demonstrated the problems of medical image searching using medical abbreviations and acronyms.

Firstly, he searched for images using ‘TLR-3’ and ‘TLR3’. Then he used ‘tool-like receptor 3’ as his query term. As Figure 5-4,

Figure 5-5 and Figure 5-6 indicate, searches for each query yielded different sets of results.

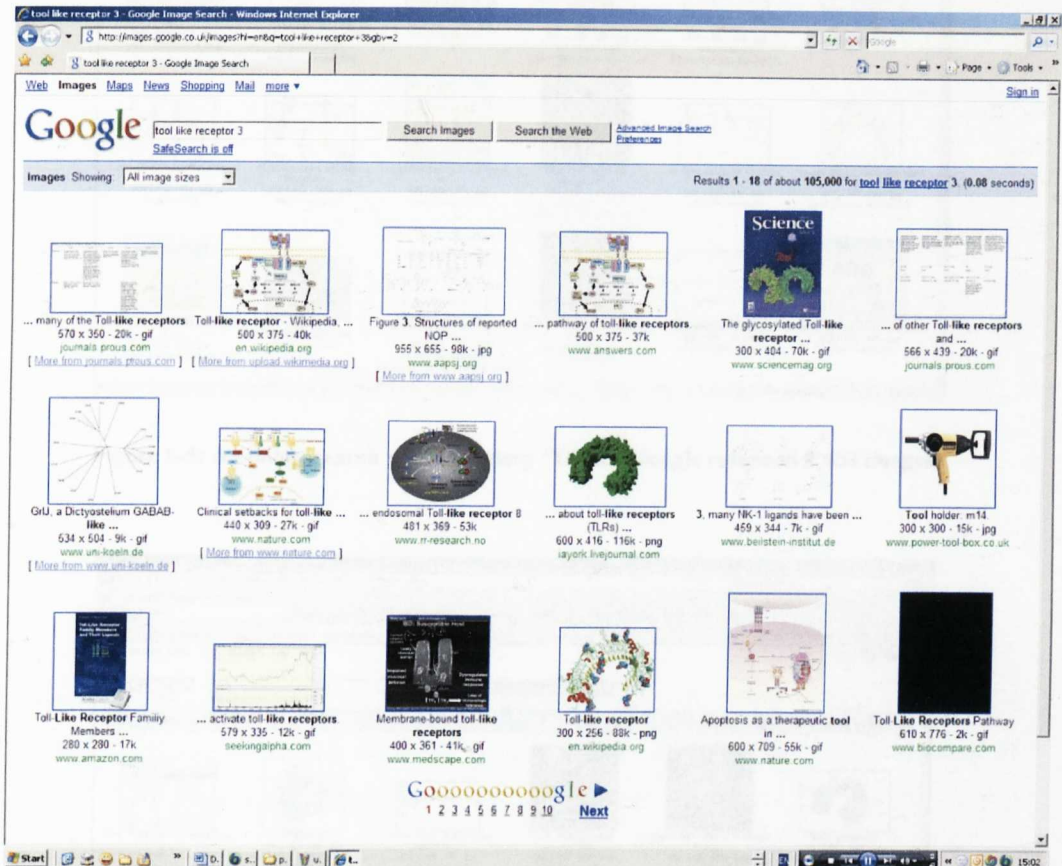


Figure 5-4: Google image search results with the query ‘Tool-like receptor 3’¹.

The same problems arose when there were different terms used for referencing the same clinical problem or concept, or a term had different spellings. For example, searches for images using ‘Hemophilia’ and ‘Haemophilia’ produced different sets of results; Google image search returned 7,890 hits for Haemophilia and 32,100 for Hemophilia.

¹ TLR3

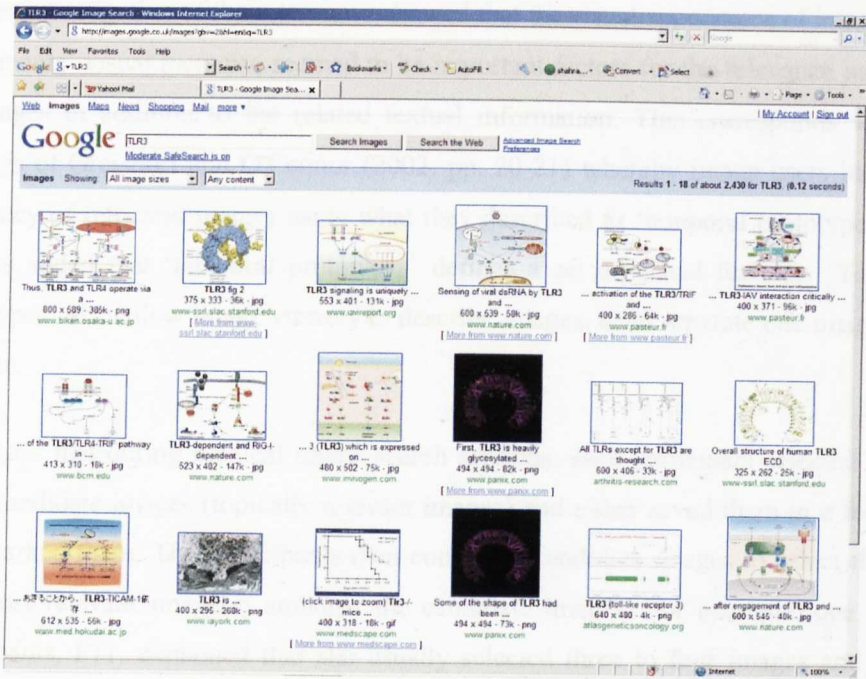


Figure 5-5: An image search with the query 'TLR3', Google returned 2,430 images.

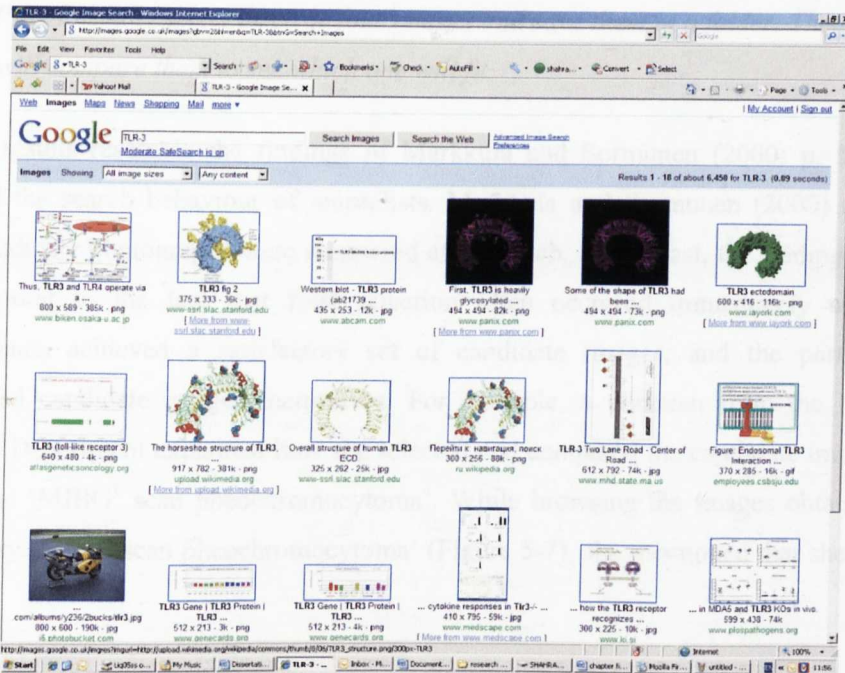


Figure 5-6: An image search with the query 'TLR-3', Google returned 6,450 images.

Relevance criteria for medical images applied by health care professionals

Further investigation of the interviews showed that the visual appearance of images, and participants' visual memory, seemed to be important factors for the relevance judgment of images in addition to the related textual information. This corresponds with the findings of Greisdorf and O'Connor (2002: pp. 20-21) whereby image users judge the relevancy of retrieved images using what they described as 'temporal prototypes'. The authors added that 'temporal prototypes' defines a set of visual features. 'Temporal prototypes' also allow image viewers to describe images, or to separate one image from another.

We found that during medical image search sessions, each participant selected one or more candidate images (topically relevant images) and either saved them in a folder or bookmarked them. The participants then compared candidate images to select the best and most relevant image(s) amongst the candidate images. For example, one of the participants, P11, explained that she usually selected three to four images and saved them in a folder. She then compared the saved images, and selected one that fitted her information needs. She said:

P11: I usually select three or four images and save them in a folder. Then I will compare them to see which one will fit.

These results resemble the findings of Markkula and Sormunen (2000: p. 22) who studied the search behaviour of journalists. Markkula and Sormunen (2000) reported that candidate photographs were compared after search. In contrast, the findings of our study point to the fact that final selection often occurred immediately after the participants achieved a satisfactory set of candidate images, and the participants evaluated candidate images themselves. For example, a clinician from the Medical Physics Department explained how she selected and compared the candidate images for the topic 'MIBG¹ scan pheochromocytoma'. While browsing the images obtained for the query 'MIBG scan pheochromocytoma' (Figure 5-7), she mentioned that she would

¹ Meta-Iodobenzylguanidine

select candidate images if the images visually resembled the topic. She described her visual memory of the topic ‘MIBG scan pheochromocytoma’ as follows:

P18: It is in my mind and I need a picture like that. When I am searching an image I know what I am looking for.

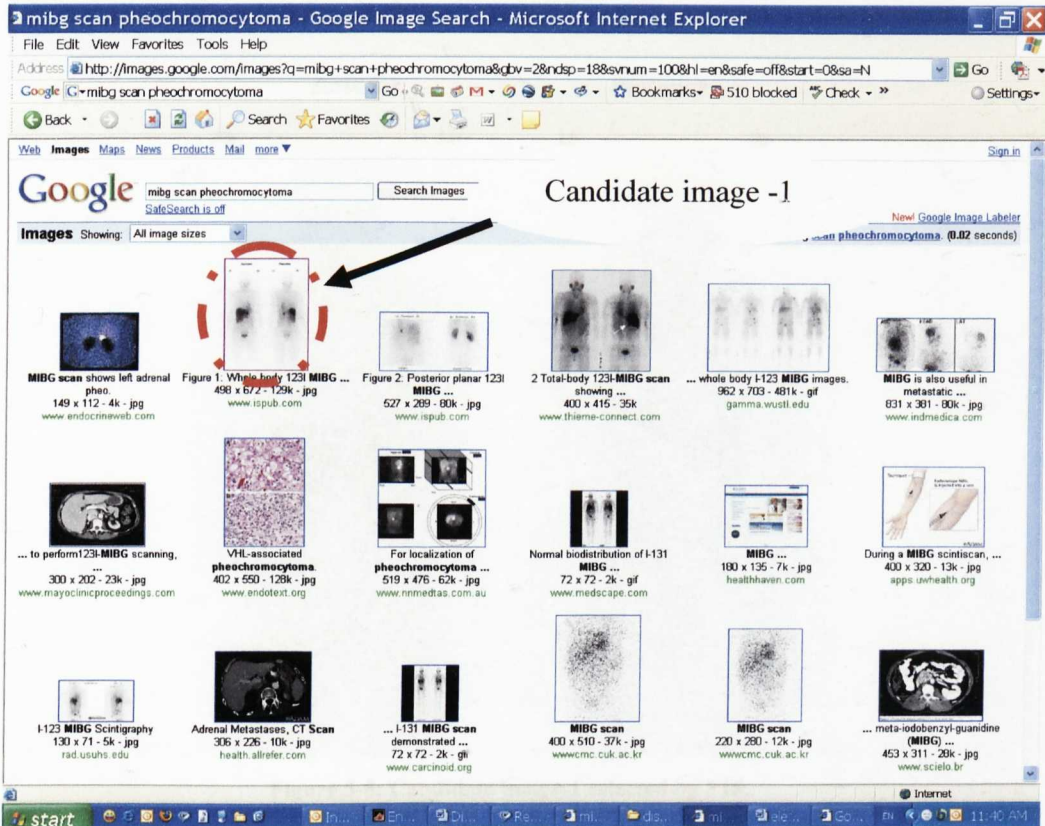


Figure 5-7: Images obtained for the query ‘MIBG scan pheochromocytoma’.

While interacting with and reformulating the query, she selected three candidate images and stated the reasons for choosing each candidate image. As it can be seen in Figure 5-8, this interviewee mentioned that she selected candidate image-1 due to its topical relevancy and its modality. By saying “*that’s the kind of image I would usually be looking for*” she implied that the candidate image-1 visually resembled the query

'MIBG scan pheochromocytoma'. This indicates that she had a visual representation for the images she needed. She said:

P18: I clicked on this image because it is a nuclear medicine image and I am usually looking for this type of image. Then I would check other attributes of the image.

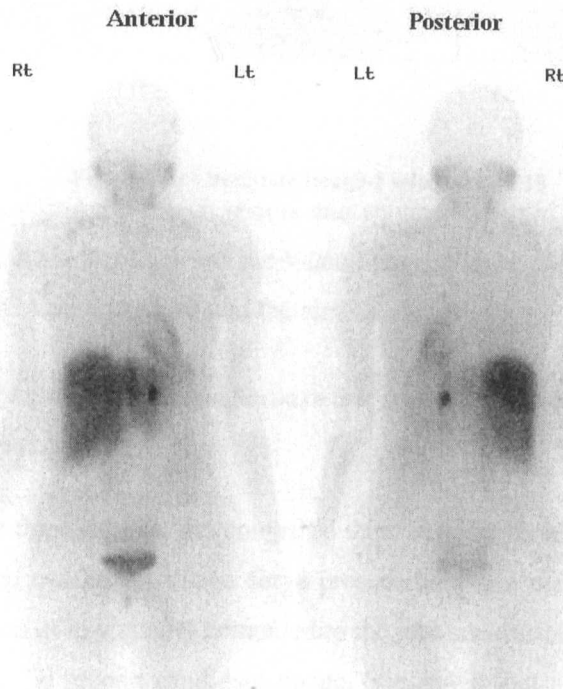


Figure 5-8: Candidate image-1 selected by P18.

After she browsed some images retrieved for the query 'MIBG scan pheochromocytoma', she reformulated the query as 'MIBG organ zuckermandl' and continued the search session. She initially selected candidate image-2 (Figure 5-9), but then said:

P18: Nuclear medicine images are not always of high resolution and this kind of image is quite hazy.

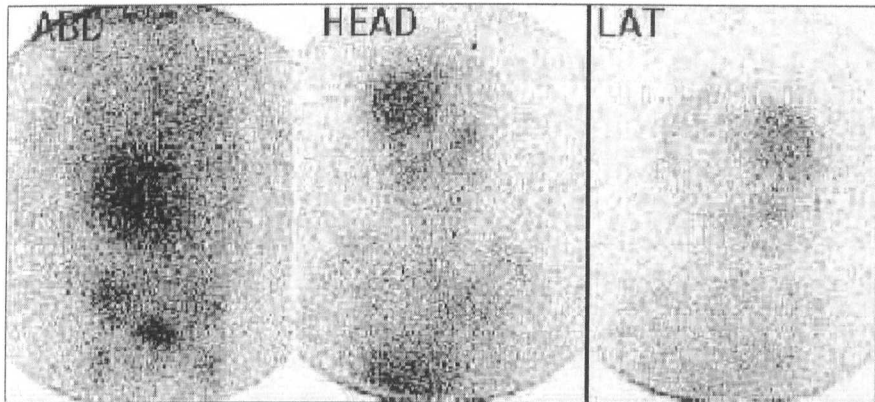


Figure 5-9: Candidate image-2 selected by P18

She continued browsing the search results and spotted the third candidate image. She stated that she would select the third candidate image (Figure 5-10) because the image was labelled MBIG scan and illustrated the tumour clearly:

P18: I would select this image, because it is a clear image and it shows the abnormality very well.

After selecting the three images, she compared them in order to select one of them. She mentioned that she needed an image for a presentation to a class. Since the second candidate image was of low quality compared to the first and third candidate images, she decided to discard the second candidate image. She stated that she could not use the second candidate image because of its poor quality:

P18: Amongst the three images I have chosen, this one is deeply disappointing. It is just awful! It is so pixely, so I would not use this one.

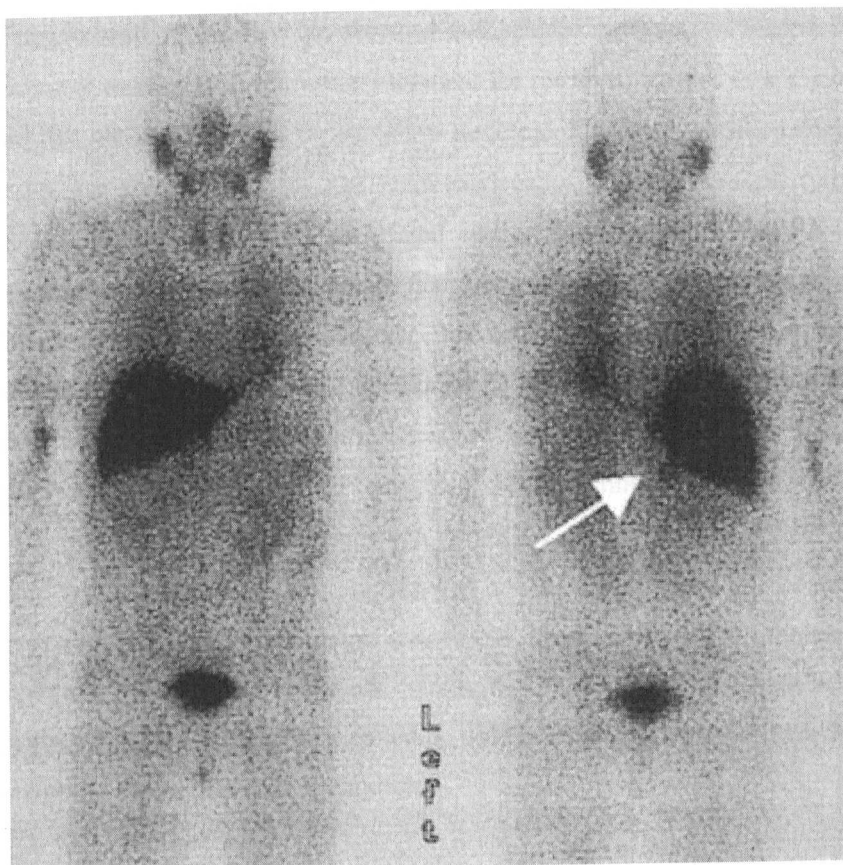


Figure 5-10: Candidate image-3 selected by P18

She compared the first and third candidate images and decided to use the third one as it was labelled MBIG scan and she felt that it was better in terms of size, appropriateness and quality.

As mentioned in chapter 2, topicality has always been the most important and most frequently cited relevance criterion for the users of information retrieval systems, mostly for textual documents (Barry (1994); Borlund (2003a); Ingwersen and Järvelin (2005); Maglaughlin and Sonnenwald (2002); Mizzaro (1997); Schamber (1994)). Similarly, our study found that topical relevancy was the most frequent and most important criterion for medical images sought by health care professionals. In order to investigate how health care professionals judge the topicality of medical images, we asked participants to explain how they assessed the images. The following example explains how one of the participants, P12, searched for images using the query 'diabetes and retinopathy' in

Google image search, and how he assessed and selected images he needed. With his permission, we captured the relevance judgment for retrieved images (see Figure 5-11). He stated that all images might be related to the topic. However, he mentioned that he became interested in nine images that seem to visually illustrate the topic (we marked them as 1-9 in Figure 5-11). As we stated earlier, participants assessed the topical relevancy of images as the first step, thus this person became interested in images that he thought were topically relevant to the topic. The initial assessment of images was based on textual description and the visual appearance of images. Then he started to evaluate each image using other criteria. He stated image 7 was more diagrammatic while he was looking for actual photographs. He classified the remaining seven images into two groups: the first group included images 4, 5 and 6; the second group included images 1, 2, 3, 8 and 9. Images in the first group showed how a person with diabetic retinopathy appears in real life. The participant stated that he is looking for photographs of haemorrhage in the retina of a diabetic person. He explained that he was looking for photographs taken through a device called an ophthalmoscope; therefore, he decided to check images in the second group. He said:

P12: The one that is interesting for me at the moment is this one [image number 1 in Figure 5-11] This is because I was looking for some photographs taken through a device called an ophthalmoscope. Those photos will show you what the retina of a diabetes patient, who is developing retinopathy diseases, will look like. Those three images at the top [1, 2 and 3 in Figure 5-11] and those [8 and 9] in the bottom line are promising. I know roughly what image I need. I have seen pictures like this one [he showed an image on the screen] in textbooks, and I have looked at the eyes of retinopathy patients with an ophthalmoscope in the past. I know what sort of image I need and I want to find something like that.

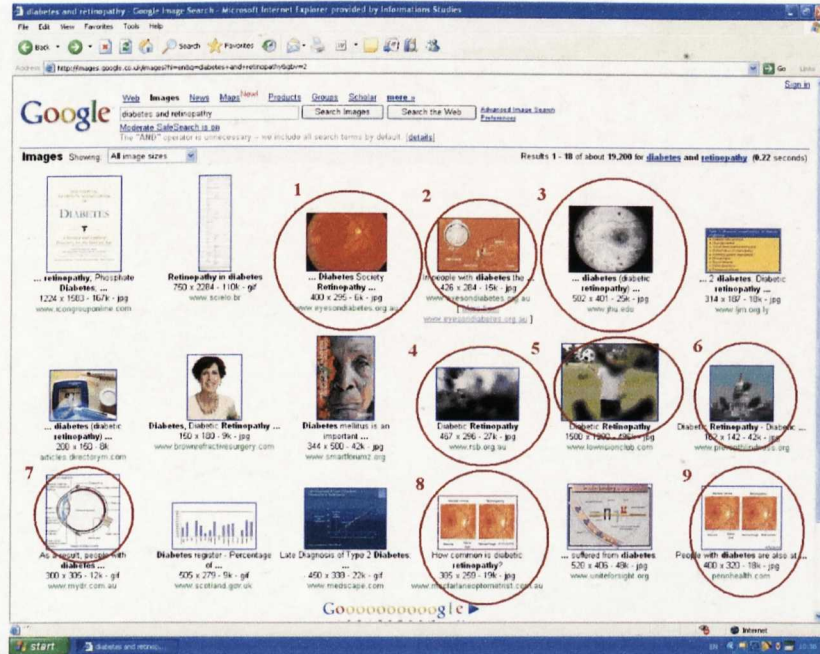


Figure 5-11: Results obtained for diabetes and retinopathy in Google image search.

We noted that this participant paid regard to the visual appearance of images by saying “I’ve seen pictures like this one in textbooks” to select nine images topically or potentially relevant to his image information needs. It was evident that he knew what type of image he needed. In other words, he had a visual memory of images for retinopathy in diabetic patients. Then he applied another criterion, ‘modality’ (the type of image), and selected five images which were taken through an ophthalmoscope¹ and seemed to fit his information needs. Using ‘colour’ as a criterion he discarded image 3 since it was black and white. By clicking on images 1, 2, 8 and 9 he was focusing on images that seemed to be most relevant to the topic. He mentioned that he would select image 1 (Figure 5-12) as it shows the haemorrhage and blood vessels in the retina better than images 2, 8 and 9, and that due to its quality, it shows the details of the lesion.

¹ An ophthalmoscope is a medical instrument with a special mirror that allows doctors to examine the interior of the eye.

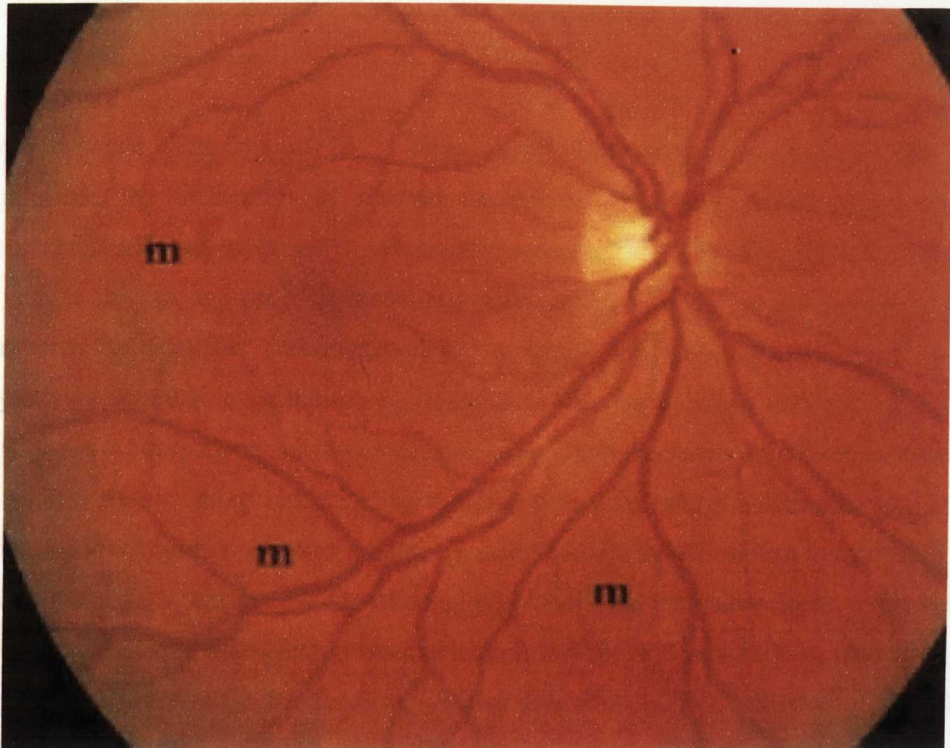


Figure 5-12: Image selected by P12 for the topic 'diabetes and retinopathy'.

The following sections now discuss the relevance criteria identified by our experiment. By identifying the criteria applied by health care professionals, we may be able to incorporate those criteria into the medical image indexing and retrieval process. The research literature identifies two distinct approaches to image indexing: text-based (concept-based) and content-based approaches (Müller et al., 2004a; Lehmann et al., 2000; Rui et al., 1997; Brandt, 1999).

Text-based: This approach can be traced back to the 1970s (Bach et al., 1996; Rui et al., 1997). A traditional approach for text-based systems of that time was to manually construct representations of the images by assigning descriptive terms to each image and then use database management systems (DBMS) for retrieval. Manual annotation of images is time consuming, costly, and tiresome (Glatard et al., 2004; Deselaers, 2003). Sometimes creating annotations by hand can become difficult because image databases

can be very large. For example, it is practically impossible to annotate manually all the images stored in the web.

Content-based: The emergence of large scale image databases in the 1990s has emphasized the difficulties of manual annotation. Therefore content-based image retrieval (CBIR) was proposed to overcome the outstanding need for efficient indexing of images. In this approach, features of images, such as colour, shape or texture are automatically identified and extracted by computer software. The extracted visual features are stored in the database as Figure 5-13 illustrates. In the retrieval stage, and when the user presents one or more example images that represent the information need, the system should return similar images. CBIR thus addresses the situation where the image retrieval system searches images by using visual features of the images such as colour, shape, and texture, instead of using text labels and attributes (more information about the CBIR and visual features is available from Ahmad Fauzi (2004); Glatard et al. (2004); Deselaers (2003); Müller et al. (2004a); Rui et al. (1997); Müller et al. (2004b); Brandt (1999).

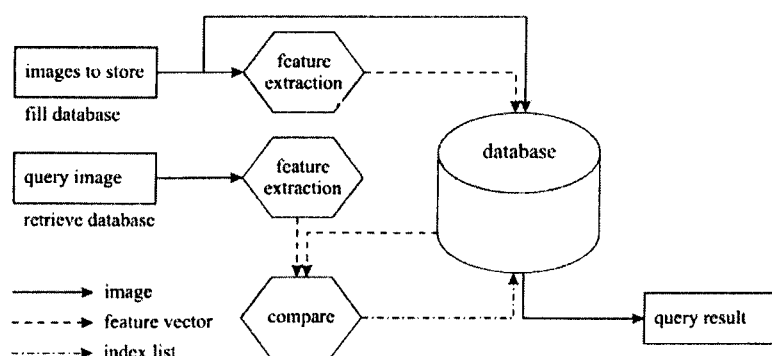


Figure 5-13: General architecture of CBIR systems
(from Lehmann et al., 2000: p.133).

CBIR allows users to find images that are visually similar to the presented query image; however, this approach presents a number of challenges, primary among which is the

Relevance criteria for medical images applied by health care professionals

problem of 'similarity'. Brandt (1999); Deselaers (2003); Müller et al. (2004a) report that 'similarity' may mean different things for different people in different situations and that measures of similarity must be defined. For example, a radiologist may have different criteria of similarity for x-ray radiographs from those of a journalist.

5.2. Relevance criteria

The objective of this study is to elicit the relevance criteria health care professionals apply when searching for medical images. In this section, we list criteria regardless of the role of the participant, and the activities carried out in their associated medical departments. All participants were aware of the legal issues involved and the need for protection of patient privacy. Officers from The University of Sheffield and Sheffield Teaching Hospitals NHS Foundation Trust monitored the study to ensure that neither researchers nor participants breached the rules. We noticed that participants could access anonymous images through the health information systems they used, and in addition used images in different ways for a variety of reasons (e.g. viewing images from medical websites or electronic journals for clinical purposes). Some participants may use patient images only if they have prior written consent from the patient.

In total, fifteen relevance criteria were elicited from the participants (see Appendix 9), as shown in Table 5-2, together with the number of participants who specified each criterion. Participants employed diverse criteria in their image evaluation processes. Among them, topicality, image quality and dimensional size of the image were the three most frequently used relevance criteria. These fifteen criteria emerged during the first stage of data analysis using the Straussian version of grounded theory: open coding. During the axial coding stage we grouped the emerged concepts (relevance criteria in this study) to form the categories. We merged the fifteen criteria into three groups: visual (or non-textual), textual, and other criteria. Visual criteria are associated with the visual/photographic attributes of an image, such as orientation, image quality, magnification and size (dimensional). Textual criteria relate to text attached to an image. We also established the 'Other' category to group four criteria that were not easily attributable to the previous two categories. Some criteria could be classified under more

than one category. For example, modality was considered as a visual criterion; however, participants sometimes identified the modality of medical images based on textual information attached to images. Therefore, based on the overlap, we established an intersecting group of criteria: textual-visual, which is detailed in

Figure 5-14.

Table 5-2: Relevance criteria employed by participants
The number indicates the number of participants who applied each criterion (No=15).

Criteria	Frequency
Topicality	29
Image Quality	27
Size (dimensions)	21
Age and Gender	20
Modality	18
Orientation	15
Credibility	13
Targeted Audiences	12
Technical Information	11
Magnification	7
Colour	7
Copyright	6
Availability	4
Recency	3
Originality	2

After forming the categories, the main category was identified to generate a theory which describes the relevance judgment of medical images and the image seeking behaviour of health care professionals. We selected the main category at the selective coding stage, and the theory was generated after we reached saturation in data collection and data analysis. The main category we selected was visual (non-textual) criteria, as we noted that health care professionals paid more attention to visual attributes when

selecting medical images. We discuss further the reasons for choosing the main category in this chapter.

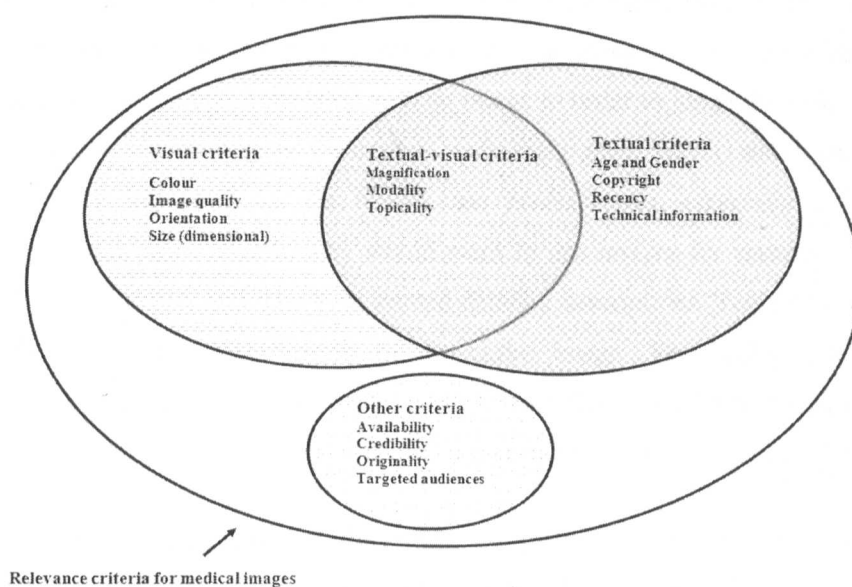


Figure 5-14: Groups and subgroups of relevance criteria we identified.

We describe the criteria in sections 5.2.1, 5.2.2 and 5.2.3. As we stated, we arranged the criteria in three main groups, and ordered them by the number of participants who used them.

5.2.1 Visual criteria

As stated earlier, visual (non-textual) criteria arose when the participants considered the visual/photographic attributes of an image in making relevance judgments. They are topical relevancy (or topicality), image quality, size (dimensional), modality, orientation, magnification and colour.

Topicality¹

This study found that the most frequent and most important criterion used by all of the participants was topical relevancy. Although this finding is consistent with previous research outlined in chapter 2, we noted that participants used both textual description and the visual appearance of the image when judging topical relevancy. All participants of the study stated that they required textual information such as medical history and diagnoses in the text associated with an image in order to judge the topical relevancy of the image. The findings of the current study highlighted the fact that textual information was very important for the health care professionals, especially a comprehensive description, since usually they would want to use images for research, educational or clinical purposes, and thus they wanted detailed descriptions. Topicality reflected to the 'aboutness' or subject matter of the entire image. The health care professionals consistently read the text surrounding the images that seemed relevant as they wanted to know what was actually illustrated in the image, and what its original context had been.

One of the participants of the study, P17, highlighted the importance of textual information for the relevance judgment of medical images as follows:

P17: It is important to have detailed information about the images. Sometimes you find very similar images with very different text. For instance, when you need T-2 weighted MRI scans you might find T-2 weighted scans with different spin echo types. When you have different echo types, MRI scans change in a graded fashion. You need to know the pathology, the echo time, and other sequence details as explained in the text.

Although text seemed to be an important source of information for judging the topical relevancy of images, we noted that the participants could not access relevant textual

¹ Although we established topicality as a criterion under the visual criteria category, we noted that this criterion can also be classified as a textual criterion as the participants used textual information attached to the image to judge its topical relevancy.

information when images were obtained from web-based resources or when they used departmental image collections.

We also note that health care professionals judged the topical relevancy of images based on the visual appearance of medical images. Our findings demonstrated that participants had a visual representation of the objects they sought in their mind, and that participants used this visual representation when assessing retrieved images. The participants used expressions such as ‘*visual knowledge*’, ‘*something in my mind while I am looking for an image*’, ‘*a mental image of what I want*’, ‘*it is already in my mind and I need a picture like that*’, ‘*I probably already know in my head what I want*’, ‘*what I want is something that I have seen before or I know what it is*’, ‘*visual memory*’, ‘*I have got draft sorts of image in my head*’. The following quotation shows one participant’s opinions about visual memory and its importance for the assessing the topicality and making relevance judgments of retrieved images.

P12: I know what I am looking for. I have seen pictures like this one [he showed an image on the screen] in textbooks, and I have looked at the eyes of retinopathy patients with the ophthalmoscope in the past. Therefore, I know roughly what sort of image I need and I want to find something like that.

Our findings show that to make relevance judgments, the user must visually inspect the image in order to know if the retrieved image is topically relevant, and whether the image contains the requested visual information. To find out more about the visual appearance of images, the participants were asked: “What do you mean by the visual appearance of images?” Participants replied used phrases such as “*it visually illustrates what I want*”, “*what I want is something I have seen before*”, “*a visual representation of things you are trying to show*”, “*image must present what I want*”, “*the image must illustrate what I want properly*”, “*it visually presents what I need*”, “*I want an image like those I have seen before*”. We noted that, based on their knowledge and experience, participants had a ‘mental image’ of what they were looking for (i.e. a part of human anatomy, certain type of image or a medical instrument), and they were trying to find

something similar. The following quotation from a participant describes how he evaluated relevancy of images to his image information needs:

P8: The image must be relevant and it must illustrate the point that you are trying to make. I have knowledge of the subject. I will compare and select images using my knowledge. If you ask me about the knowledge, I would describe it as 'visual knowledge', because I have an image in mind of the topic. Let me explain it for you. Blind people are unable to describe or understand the image, because they do not have visual memory of, or visual information on, the objects or people around them. We compare images at the visual level based on our visual knowledge or memories of the topics.

Further investigation of the data showed that the visual appearance of images, and the participants' visual memory, seemed to be the most important sources of information in judging the topical relevancy of images. Two participants explained this:

P3: I would select the image if it illustrates what I want. I know from my experiences and my education how the image would look for this particular medical condition. That is in my mind, so I will look at the images to select images that are similar to what is in my mind.

P16: It visually illustrates what I want, and that is the most important thing. There is something in my mind, and I want something similar to it. I know what I am looking for and I know what I want to see.

The interviews in this study also showed that after topical relevancy of an image was established, participants applied other criteria. For example, twenty-five participants said that if the image was topically relevant and illustrated what they wanted, they would then check the quality of the image to see whether they could use it or not.

P7: First, does the image illustrate what I am after? Then the quality of image is important. It must also give me the best view, and it should be from a valid source.

Similarly thirteen participants stated that they would also consider the credibility of images if the images were topically relevant.

Topical relevancy was applied by all participants in the current study. We noted that topical relevancy referred to whether an image was relevant or not based on its visual content and appearance as participants such as P12 stated. However, as we stated earlier in this section, this criterion was in part dependent upon the textual description of the images, the ability to recognise images visually similar to those which were previously relevant, and also to recognise the meaning of visual content, e.g. the visual appearance of human anatomy, cells and organs. Tagare et al. (1997) suggest that similarity from a medical perspective is predominantly context dependent. Lehmann et al. (2000) support Tagare et al. (1997) and report that the interpretation of similarity in medical images is inherently knowledge-based and dependent on both image and query context. Lehmann et al. (2000) also add that the medical knowledge arises from anatomic region and physiological information, which is quite often obtained by the clinician simultaneously with the diagnostic process. The findings of this study on the judgment of topicality of medical images indicate that visual features are required to support medico-diagnostic queries. Moreover, the context of queries is unknown when images are indexed and entered into the database. Tagare et al. (1997) suggest that the database scheme must be generic and flexible. The authors also note that medical image interpretation is a complex and poorly understood process. Diagnostic deductions derived from medical images such as X-ray radiographs rest on an incomplete, continuously evolving model of normality. Therefore, indexing of medical images is required to support medico-diagnostic requests on a higher level of image interpretation.

Seven out of the fifteen criteria we identified were in the visual category. It also appears from the results of this investigation that when users select images, they first look to see if the image is topically relevant and whether it visually illustrates the problem, before

any other attribute of the image is considered. This was made evident by finding that medical image users were more focused on the visual attributes when they saw images, and less on other attributes. They usually select three to five candidate images which seem to be topically relevant to their information needs. The other criteria were important when they compared candidate images. This supported the findings of Greisdorf (2000); and Markkula and Sormunen (2000) that the relevance judgment of retrieved images was composed of multiple sessions and one image will be chosen over another if it is better on at least one attribute. Our investigation demonstrated that health care professionals appeared to judge the relevance of retrieved images as a stepped process leading from the topical relevancy of retrieved images (results set) to comparison of other attributes of the candidate images.

Image quality

The photographic quality of an image, including image resolution, contrast, and brightness, was the second criterion applied by the participants. We asked participants to define the quality of images. One participant defined quality as:

P8: The quality of image is also important. If you ask me to define the quality, I would tell you that the quality of image means that the image should be clear and legible. For example, if you want to see microscopic images you will gain more information if they are high quality images. Additionally, you could see visual details in a high quality image.

One of the participants emphasized the importance of image quality and said:

P22: Images, especially ultrasound images, should be high quality images. High quality images represent the point that I want to make properly. It should show the visual details because details are important too.

As Table 5-1 indicates, the quality of an image criterion (e.g. image resolution, quality of printed image, contrast, and brightness) was applied by twenty-seven participants.

The participants seemed to check the quality of images after topical relevance was established.

P1: The image should be topically related to my aim. If it was similar to what I wanted I would check the quality and visual details of the image.

Sometimes the participants who looked for images in web-based resources using Google image search complained about the low quality of images. Therefore they preferred to find images in resources other than web-based resources. For instance, P8 mentioned that he was unable to use most images from the web found by Google:

P8: I wish the Google image search engine would return me all high quality images. Sometimes you can find images but you cannot use them, because most of them are low quality images.

Similarly, another participant criticised the quality of images obtained from web-based resources and said that he never expected to find high quality images using Google image search:

P15: If I am using Google image search and if I am looking for images on the internet I am not expecting to find images greater than 1280 x 1020 pixels. Because anything greater than that, even if it is JPEG, starts to get quite big.

Some participants used images to illustrate their presentations, therefore they emphasized that they were interested in copyright-free and high quality images to show the desired visual details in their presentations. One of them said:

P10: Since I need the image to use in a presentation, image quality and the actual resolution of the image are important.

Sometimes participants printed images before using them to verify the quality of the images. For example, one of the participants, P11 who needed some images to illustrate

her scientific paper, mentioned that she wanted to see how the images would appear when printed. Thus she would check the quality of printed images:

P11: Definitely I will not choose this one because I am currently writing a review article. I would print the image and see how it appears on paper.

According to Chaffey (1996) quality of a digital image is directly related to a number of factors including resolution, compression method and colour depth. Researchers such as Ivkovic and Sankar (2004) developed algorithms for automated image quality assessment. However, quality assessment of the images is required to compute those factors, and limited success has been achieved in automated quality assessment of images as Wang et al. (2002) state. However, as Wang et al. (2002) suggest it might be possible to use a “*universal image quality index, which is easy to calculate and applicable to various image processing applications*”. Such an index can also be used to assess the quality of digital medical images and integrate it in the medical image retrieval process.

Size (dimensions)

The selection criteria for medical images were closely connected to the participants' individual tasks. Often participants looked for medical images to illustrate their publications and presentations; therefore, they wanted to make sure images were an appropriate size. For example, one of the participants, P9, enlarged images using a projector. She stated that she must see images in their full size to decide whether she can use them or not.

P9: I have to take a closer look at images to select them. I have to see the full size images. Sometimes they are quite small. When you enlarge them, you cannot use them.

Sometimes the participants stated that they might change the size of images using graphics editing software, such as Adobe Photoshop. However, they preferred to find images already in the desired dimensions or size:

P11: I usually select three or four images and copy them in a folder. Then I compare them to see which one will fit. When I compare them, I will consider the size and quality of image, if I want to use it for a presentation or in printed material. Some images are too big, so I will reduce the size of them. When you reduce the size, the quality will be degraded.

P20: If you find a small image and change its size, you cannot use it. But if it is a large image you can manipulate and reframe it. However, it is time consuming and I hate doing that.

Some of the participants mentioned that they might find a small size image which seemed to be the most relevant. In those cases, they tended to reproduce larger versions of those images.

The participants considered the dimensional size of image as a criterion. This is a criterion independent from the context and totally dependent on the image as an object. According to Stern and Richardson (2003) the dimensional size of digital images is usually expressed as the number of pixels per inch (PPI), or as the number of dots per inch (DPI). Dimensional size of image is a criterion that could be implemented in the design of image retrieval systems including medical image retrieval systems. For example, the 'Measurements' element in the Visual Resources Association core categories (see Table 5-3 and Table 5-4) can be used to define dimensions in the metadata set recorded for an image. Additionally, the dimensional size of images embedded in web pages is stored in the HTML¹ code. For example, `` means that an image with the dimensional size of 624*563 has been used in the web page (Raggett, 1997). This information can be extracted by image retrieval systems (e.g. image search engines) automatically from the HTML code and would allow users to find images in the desired dimensional size. For example, Google image search allows end users to classify images according to their size (extra large, large, medium and small) as Wang et al. (2006) state. The participant of the current study often

¹ hypertext mark-up language

used this feature in Google image search to filter the images in the results set according to their dimensional size.

Modality¹

There are different types of medical images such as radiographs, MRI scans, microscopic images, diagrams or general photographs. Many of the participants wanted to search for specific types of medical image, and regarded the type of medical image as a criterion. We used 'modality' to describe this criterion.

P10: I will try to limit the search to MRI scans, I mean I need MRI scans.

P12: The one that is interesting for me is this one [image number 1 in Figure 5-11], because I am looking for some photographs taken through a device called an ophthalmoscope.

Eighteen participants used this criterion when searching for medical images, and eight of them considered modality as the first criterion used.

P13: A relevant image for me is generally an MRI scan covering the entire region I am interested in.

P17: I am looking for MRI scans, but it depends what I am looking for. In this particular case, I am looking for a T-2 weighted image of the brain.

P24: I would consider the type of image first. Then I will see which one is a high quality image.

Sometimes the participants were interested in including the modality of images in the image retrieval process, and wanted to retrieve certain types of images as they specified. However they had some difficulties when they wanted to narrow down the image search to a particular type of image. For instance, when participants used Google image search

¹ Modality was among those criteria that could be classified in both visual and textual criteria. However, we discuss this criterion in the visual criteria section.

to search for images, they used the modality of images as a search term. They added terms such as 'PET', 'MRI' and 'CT' to the query (Table 5-1).

P10: I do not know what will happen if I put MRI aneurysm. That gives me quite a good selection, because I have got mainly MRI scans.

P16: Few images seem to be relevant. What I tend to do is to refine my search criteria trying to find more relevant images. I will add another word, let's say PET. Now there are more relevant images. There are some relevant, but still some irrelevant images.

Image modality is a fundamental characteristic of medical images that can be used to aid in the medical image retrieval process. However, the descriptions or captions associated with medical images often do not capture information about the modality of images. Therefore, in a number of studies researchers tried to automatically extract the modality of a given image using its visual information, and to index a given medical image based on its modality. As a result, the system allows for image retrieval by modality in both text-based and content-based image retrieval approaches. This allows the end user of medical image retrieval systems to restrict the query results set to one containing just images with the same modality as the query image, or as requested by the user, and allows identification of more relevant images. For example, users can retrieve images by modality keyword (i.e. 'CT images of lung') and by content (i.e. 'find all similar images to the query image').

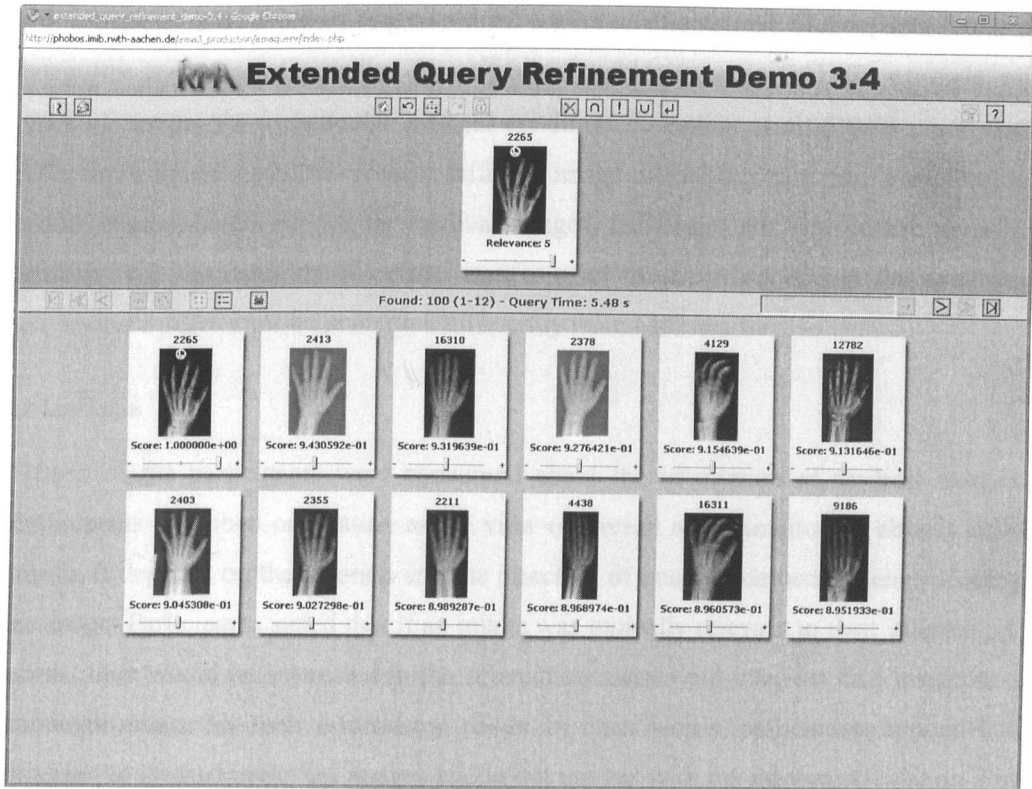


Figure 5-15: IRMA search results
 Top: Image query is an X-Ray of left hand. Below: Images retrieved by IRMA
 adapted from http://phobos.imib.rwth-aachen.de/irma3_production/irmaquery/index.php

According to Müller et al. (2004a) the efficiency of medical image retrieval based on the modality of images has already been demonstrated. One of the few projects with at least a partial implementation is the image retrieval in medical applications (IRMA¹) project that allows for a relatively robust classification of medical images according to the anatomic region, image modality and body orientation. Such a framework enables users to retrieve a certain type (modality) of image (see Figure 5-15). In the sample search, the image query is an X-ray of the left hand. The IRMA system retrieved radiographs taken of the left hands of patients. Lehmann et al. (2000) report that IRMA codes allow each image to be linked to several categories. Moreover, different types of medical images

¹ Detailed information about the IRMA project is available at <http://www.irma-project.org>

will share the same category (e.g. anatomic region or orientation). Hence, data sets and categories are related internally. This situation enables users to find, for example, all types of images for a particular medical condition. Lehmann et al. (2000) report that different imaging modalities require different image processing methods. Therefore, to enable content-based queries for medical images, the image retrieval system must be familiar with the modality of current images prior to query processing. For example, ultrasound images must be processed differently from MRI scans.

Orientation

Fifteen of the participants were concerned about the orientation of medical images. Participants described orientation as the view or overall representation of objects in an image. It depends on the location and the direction of imaging devices when producing an image. Participants stated that if an image was topically relevant to their information needs, they would be interested in the orientation and would consider that image as a candidate image for their information needs. In other words, participants applied this criterion to discard irrelevant images and select images with the desired orientation. For example, P13 was looking for MRI scans of the 'hot cross bun sign' in the brain acquired in an axial plain. She used the orientation of the image to distinguish between relevant and irrelevant images.

P13: This image is not useful for me as I am looking for an axial plain.

Moreover, participants reported that there are some standard and predefined orientations, such as Sagittal and Coronal sections, for taking medical images. Therefore, they know which orientation is the best for their needs:

P17: In MRI scans there are only Trans, Sagittal or Coronal views and it depends on what you are looking for and what you want. I mean we never look at MRI scans of the lung from a coronal or sagittal view because it will be too small. [This participant showed an example and explained that sagittal MRI scans of the lung do not present an appropriate view for the user].

Orientation is one of the most widely used attributes for medical images in the retrieval process. For example, Pietka and Huang (1992) present a computerized procedure to determine the orientation of 976 computed radiography images of the chest automatically. They report that their system could determine the orientation for 95.4% of radiographs correctly. In another study, Lehmann et al. (2003) propose an advanced computerized method based on a Nearest-Neighbour (NN) classifier to identify the orientation of chest radiographs and classify radiographs based on their orientation. They report that correctness of classification of radiographs was 99.6%.

According to Müller et al. (2004a) one of the successful examples of content-based medical image retrieval is the IRMA¹ project. The IRMA system not only categorizes images automatically, but also annotates each image automatically (see Figure 5-15). This allows the end user to retrieve images using text queries or image queries. In IRMA, a multidimensional code is created automatically to annotate each image in medical image databases with axes for modality, body orientation, body region examined, and the biological system under examination. Müller and Geissbuhler (2005) report that IRMA codes are in the form of TTTT-DDD-AAA-BBB. The number of letter in each part of the code shows the level of detail (e.g. modality has four levels of detail). T, D, A and B stand for modality, orientation, anatomic region and biological system axis, respectively. For example, the sample code 1123-211-520-3a0 presents an image as radiography, projection radiography, analogue, high energy (T) – sagittal, left lateral decubitus, inspiration (D) – chest, lung (A) – respiratory system, and lung (B).

Regarding the importance of orientation as a relevance criterion for relevance judgment of medical images, the classification of images based on orientation is a basic requirement for the indexing and retrieval of medical images. This is a criterion that cannot be addressed by text-based (concept-based) image search tools. Additionally, this is a visual attribute of medical images and such a criterion cannot be incorporated in

¹ Image retrieval in medical applications (IRMA) is a content-based medical image retrieval system. See <http://irma-project.org> for detailed information.

medical image retrieval process unless a content-based image retrieval system such as IRMA is applied.

Colour

The colour of images, including colouring techniques and composition, was also a criterion identified by the participants. They thought that colour could help differentiate different parts of an image. For example, one of the participants, P3, stated that he could show certain type of proteins using green staining (a colouring technique in medicine). He added that he could differentiate between two types of proteins using the colouring technique utilized in the images.

P8: Sometimes colour is important and it can affect the selection of an image. Sometimes the colour of an image can help you to see the differences.

The interviews in this study also showed that participants tended to select colour images. They believed that colour images could be understood easier than the single colour or black and white images. The participants such as P23 and 27 seemed to prefer colour images.

P23: It is a representative image for what I am looking for. It is a clear image and it is a colour image, so I can easily understand and distinguish between the different parts.

P27: I prefer to have a colour image rather than black and white images, because you can understand colour images more easily.

Although there are some colour imaging techniques in the medical domain (i.e. pathology images), not many medical images are colour. Using techniques such as colour histograms, images can be classified into colour and black and white images as Tomasi and Manduchi (1998) state. Currently, general image search engines such as Google image search, use colour information of images such as colour histograms and allow categorization of images according to their colour (Wang et al. (2006)). Using

such approaches might allow end users of medical image retrieval systems to filter medical images according to their colour. However, as Muller et al. (2005b) report, medical images such as radiographs or ultrasound images are mainly greyscale images and categorization of medical images based on their colour is impractical.

Colour is also a powerful visual feature, which simplifies object identification. Thus, this is one of the most frequently used features for content-based image retrieval. To extract the colour features from the content of an image, a proper colour space and an effective colour descriptor have to be determined. For example, a colour histogram, which represents the distribution of the number of pixels for each quantized colour, is an effective representation of the colour content of an image (Müller et al. (2004a)). According to Flickner et al. (1995) IBM's QBIC¹ (Query by Image Content), was the first content-based image retrieval system that allows the user to generate queries based on example images, sketches, colour and texture patterns. After the colours of interest have been chosen and added to the right side bar, the area of each colour band can be adjusted in the 'colour specification area' to indicate their respective weights.

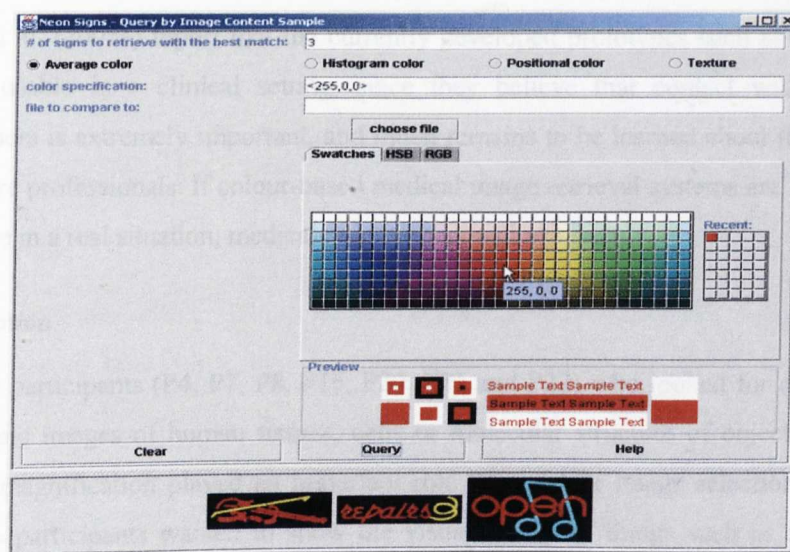


Figure 5-16: QBIC query based on colour

From <http://www.ibm.com/developerworks/data/library/techarticle/0202cox/0202cox.html>

¹. QBIC: Query Image by Content – see: <http://www.qbic.almaden.ibm.com/>

Using colour as a visual feature for medical image retrieval is restricted due to the nature of medical images. As Zhou et al. (2008) state, medical images are mainly greyscale images and no colour features are thus available. In the field of medicine, pathology images have been proposed for content-based access as the colour and texture properties can be detected. For example, Comaniciu et al. (1998) used colour as an attribute for the retrieval of pathology images. Although they emphasized the importance of colour in content-based pathology image retrieval, they state that pathologic images will need to be normalized in some way as different staining (colouring technique used to produce pathology images) methods can produce different colours. MedGIFT is another content-based medical image retrieval based on GNU Image Finding Tool (GIFT¹), an open source system developed in the University of Geneva. Muller et al. (2005a) report that MedGIFT retrieves images using four groups of attributes related to the colour features of images: 1- global colour and grey level histogram; 2- local colour blocks at different scales and various fixed regions; 3- global Gabor filter response histogram using several scales and directions; 4- local Gabor blocks in fixed areas of the image in several scales and directions.

Muller et al. (2005a) report that the currently developed prototypes such as MedGIFT are not usable in a clinical setting, since they believe that contact with medical practitioners is extremely important, and much remains to be learned about the need of health care professionals. If colour-based medical image retrieval systems are useful and applicable in a real situation, medical practitioners will use them.

Magnification

For those participants (P4, P7, P8, P15, P20, P21 and P22) who looked for detailed or microscopic images of human tissues, cells or molecular structure of objects such as proteins, magnification played an important role during their image selection process. When the participants wanted to show the visual details of things such as cells they would pay attention to the magnification of the images they retrieved.

¹: <http://www.gnu.org/software/gift/>

P8: Size and magnification are very important especially for images at a microscopic level. In fact, it is a major requirement. You cannot show a microscopic image without giving information on the magnification.

Participants such as P22 were able to zoom on the whole, or any specific part, of high quality images using graphics editing software, such as Adobe Photoshop; thus they seemed able to find images in the proper magnification or size. Sometimes they mentioned that changing the magnification of an image was not appropriate as it might affect the quality of the image and its visual appearance.

P22: You can change the magnification of an image if it is a high quality image. I mean using image-editing software you can zoom on a specific part of the picture, but most of the time I prefer to find a picture which is focused on the anatomic region I am interested in and is already suitably magnified.

5.2.2 Textual criteria

Participants considered a range of textual criteria when judging the relevancy of medical images. Textual relevance criteria referred to those related to textual information such as technical information. We ordered textual criteria based on how many participants applied each criterion when assessing images.

Age and gender

The age and gender of the case (i.e. the patient) was a common criterion and participants mainly obtained such information from the textual information attached to images in case notes. For instance, one of the participants stated: "*diseases of children are different from those of an adult.*"(P5). Age is an important criterion for some medical conditions and sometimes is included implicitly in the query text itself, such as 'epiphyseal closure', as participant 3 explained:

P3: The age range is also important. If I want an image of epiphyseal closure I will put age 18 [in the query] because it [epiphyseal closure] is common at age 18.

The same person commented further on this issue, and reported that age is important for some medical conditions that change with age:

P3: If I am looking for images of the foramen of the skull when it is open or closed, age is important. The foramen of the skull is open in newborn babies but changes with age.

Another interviewee also expressed similar concerns:

P29: All criteria such as age, gender and anatomic region are important in medicine. There are several diseases common in children and there are diseases common in adults.

The gender of the patient was also important for some interviewees when selecting a relevant image.

P21: This [ultrasound images of Endometrium] is something that relates to females. In some cases, all images are female related images, because I am looking for ultrasound images of ovaries.

In reply to the question: 'Which one is the most important criterion: Age or Gender?' one participant said:

P6: Some injuries affect adolescents and some children ... if you have a specific injury it does not matter if you say male or female, but you have different injuries in a 12 years old compared to someone older. Thus, gender is not so important in that sense ... also Anterior Cruciate Ligament of the knee was common in men [Football and Rugby players] in the 1960s, but now it is also common in women, because they started playing football and rugby.

According to Duque et al. (2003), images in DICOM (Digital Imaging and Communications in Medicine) allow health staff to specify age, gender, background and medical history in patient-related metadata. Such metadata can be used to filter the image search result according the age and gender of patients. The findings of this study show that not all images stored on the internet contain information about the age and gender of patients in associated text. Additionally Lehmann et al. (2000) report that not all health staff add patient-related metadata to the medical images stored in medical image archiving systems such as DICOM. The visual content of medical images can be considered as an alternative for metadata and textual information on the patients' age and gender. Currently it is possible to detect the age of patients automatically using images of certain anatomic areas. Automated age detection systems in medicine are mainly based on the bone age detection approach. In this approach, researchers develop computerized systems to analyse X-Ray images of certain anatomic regions (e.g. hands) to detect the age of patients (see for example Sinchai et al., 2008; Tanner and Gibbons, 1994; Pietka et al., 2001; Zhang et al., 2007). Stein et al. (1999) report that at the human femoral diaphysis, the femur differs little across age and sex. They proposed a technique for analysis of intracortical porosity in human femoral bone to determine the age and gender of patients automatically. Neeb et al. (2006) also developed an algorithm to study cerebral water content changes in the MRI scans of twenty-two patients and use age- and gender-related H₂O patterns for automated age and gender detection. The authors report their system could detect the gender of 68.2% of patients correctly; the success rate for age detection was 87.5%.

As participants of the study mentioned, age and gender were critical factors for making clinical decisions. Medical databases such as PubMed allow health care professionals to limit their search by gender or age. Such search facilities are required to limit the image search to a certain age and gender. However, the diversity of medical images is the main problem for automatic detection of age and gender. Another problem is the anatomic region illustrated in the image. For example, Sharif et al. (1994) report that epiphyses and hand are the most appropriate anatomic regions for automatic age detection in radiographic images.

Technical information

This is a criterion used by eleven participants. This criterion typically relates to information included in the text associated with the images, such as annotations. Participants applied this criterion in situations where they needed detailed information on materials, methods, the laboratory situation in which an image was taken, phases and stages of a test or a mechanism, and stages prior or subsequent to the stage presented in the image. One of the participants made a clear statement showing how the availability of technical information was important to them.

P8: You have to know how, and using which technique, that image has been produced. For example, you should know whether images were produced using fluorescence or an optical microscope.

Sometimes participants stated that technical information was important and required supplementary information for a better comparison of their work (and research findings) with prior cases. Therefore, in order to access technical information relevant to the images for better comparison of their findings, participants preferred to find images in the published literature; in particular in academic papers.

P4: I use electronic and online journals, because in addition to the image you have the text of the article, which explains and clarifies all aspects of the work presented in the image. The text provides information about materials, methods, the laboratory situation in which that image was taken, technical information, phases and stages of a mechanism or test prior to the presented stage in the image. All information is important and we need supplementary information for a better comparison of images.

Copyright

There were many legal issues concerning the use of copyrighted medical images. For example, patients are the owner of their images in the medical context (Tranberg et al., 2003; Moskop et al., 2005). Copyright was an issue raised by some participants, especially if they wanted to use images in their publications and presentations.

P20: All lectures have to be published electronically and accessible online, and none of the graphics would go with them. I cannot breach copyright rules.

In particular, participants were concerned about copyright issues related to using images from web-based resources. As the following quotation indicates, sometimes participants could not access any copyright information for images obtained using Google image search:

P28: I would not use licensed images. Though it is not always clear in the images obtained from the Internet and you have to be careful.

Sometimes participants preferred to find images in papers and books rather than searching for images in web-based resources. For example, one of the participants in the quotation below expressed his personal preference for using images from books and papers for educational purposes. He said:

P20: For images obtained from published literature, it is easy to get permission and use images, because there is a publisher or an agency that you can ask and get the permission. The only time that I have to go and get copyright permission myself is for images I use for exams. Once we used colour photographs of oncology conditions and they had come out in books. We realized that some photographs in our exams would be probably copyright protected. Thus, I wrote some letters to a number of publishers saying that for the purposes of our exam, our exam only, can we use this image for one day only.

Medical images contain valuable information that can be used by physicians and other health care professionals for clinical and other purposes. Digitized medical imaging provides many advantages, such as ease of use and ease of transmission via the internet as stated by Oh et al. (2005) and Borgman et al. (2004). Our findings indicate that copyright is an overarching disincentive to using images obtained from web-based resources. As the participants of this study mentioned, explicit permission may be required to use an image for academic or research purpose. However, it might be possible to restrict the image search to copyright free images. A solution to the copyright issue in web-based resources is offered by Creative Commons¹. According to Pass and Zabih (1996), unlike traditional copyright law, which severely limits the potential uses of an image, the Creative Commons licence allows the owners of websites to license the content of their website including images with varying degrees of rights such as Attribution, Non-commercial, No Derivative Works or Share Alike (see <http://www.creativecommons.org/> for detailed information about Creative Commons licences). According to Loy and Eklundh (2005) Creative Commons is used by many online image collections and websites including Flickr². This allows users to restrict their image search to copyright free images, and so they can download and distribute images freely.

¹ <http://www.gnu.org/software/gift/>

² <http://www.flickr.com>

Another solution is to state the image rights through metadata. However, expressing the rights in the metadata requires certain tools as Gadd et al. (2004) state. Digital Rights Expression Language (DREL) has been proposed to obtain machine-readable rights specifications from metadata. Gadd et al. (2004) report that currently two DREs are used to express rights specifications: eXtensible Rights Mark-up Language (XrML¹) and Open Digital Rights Language (ODRL²). XrML provides a universal method to assign usage rights and conditions associated with all kinds of documents including images automatically. XrML, in fact, is a type of DREL based on XML³ (eXtensible Markup Language). ODRL is mainly used by the open source software and educational communities since it is free. Like XrML, ODRL is based on XML; however it is not as well developed as XrML. ODRL is a standard vocabulary for assigning terms and conditions over both digital and physical content as Gadd et al. (2004) state.

Recency

Sometimes participants stated that they would select images based on recency, wanting the latest images for a particular topic:

P5: I want to find the latest images, since everything changes in medicine rapidly.

P5: If you want to do research on a topic, you prefer to find the latest and the best images from the most reliable resources and well-known authors. Then you can compare your image, which represents your findings, with those images.

Recency is another criterion which exactly matches the findings of the studies discussed in chapter 2. For example, recency in Wang and White,1999; Maglaughlin and Sonnenwald, 2002 and currency in Schamber,1991 refers to the extent to which users judged documents to be current or recent. Due to the importance of the date in

¹ www.xml.org

² www.odrl.net

³ More information about XML is available at http://www.w3schools.com/XML/xml_whatIs.asp

information retrieval, in a number of studies researchers proposed techniques to extract the date automatically and thus allow end users to retrieve documents published in a certain time period. Using the date as a search feature allows end users of IR systems to sort the results sets chronologically (Glover et al., 1999). Lewandowski (2004) report that search engines such as Google and Yahoo obtain the date from different resources:

- date automatically added by the server which hosts the document;
- date that document was indexed by a search engine for the first time;
- date obtained from the metadata of the document added by the document creator; and
- date given in the contents of the web-based documents.

Lewandowski (2004) states that the date obtained from the metadata and content of web-based documents including images is more accurate than the date provided by the server or document's date of being indexed to determine the recency of documents. The date can be extracted from metadata attached to images; however, extraction of the date from the content of images requires CBIR techniques. For example, Xi and Xinggang (2003) developed a CBIR technique to extract the shooting date from the date imprinted by cameras on digital photographs. Garcia-Mateos et al.) also developed an OCR (Optical Character Recognition) method for automatic recognition of time and date in CCTV¹ videos.

5.2.3 Other criteria

Participants also considered other criteria that we could not group under visual and textual criteria. Therefore, we established a third group: other criteria.

¹ Closed Circuit Television

Credibility

This criterion was considered important because participants often wanted to make sure that images were from reliable and valid sources. Sometimes participants used image search engines to locate medical images from online repositories; however, general-purpose search engines (e.g. Google) do not distinguish between medically credible and less-credible websites. The findings of this study show that participants always debated the credibility and trustworthiness of images retrieved from web-based resources.

P7: When I look for images using the Internet and Google image search, the source of image is important for me. For example, I never use images from Wikipedia because everybody can put information or images in Wikipedia, and I cannot trust it.

Therefore, sometimes participants (such as P27) preferred using medical databases, such as PubMed¹, to locate relevant articles and then look for images in the articles. He mentioned that he preferred to find images in high-ranked medical journals.

P27: The credibility of the source of images I need, is also important. Since I prefer to use images from papers, I want to find them in top ranked journals of our field and widely cited publications.

We noted that the credibility of the image source was important for the participants; in particular, for participants such as P4 who looked for images to compare their research findings with the findings of other researchers.

P4: The source of the image is important, because we use images to compare the results and findings with previous work. Without images, we cannot evaluate our work, and show the differences between our results and the previous work.

¹ <http://www.ncbi.nlm.nih.gov/pubmed/> (site accessed: 18/06/2008).

Participants could be sure of the credibility of images when they used images from their personal collections or when they searched for images in books and papers. In other words, credibility was used as a criterion to decide about the potential image resources to use for image retrieval. This findings of the current study imply that developers of image retrieval systems should focus on the criteria or aspects of images which seem to be important for the users. The health care professionals studied in this study relied heavily on the criterion 'credibility' when they searched for images in web-based image resources. It is assumed that health care professionals would not value images that they deem 'incredible'. Thus, an image retrieval system intended to support health care professionals should take this into account. Search engines such as Google use algorithms such as PageRank¹ to compute the credibility of each page as Benincasa et al. (2006) report. This algorithm is based on the idea that the importance of an academic paper can be assessed by the number of citations the paper has from other academic papers. Based on this approach, a web page can be judged by the number of links the web page has from other web pages. Google calculates the importance of each page based on the number of links to that page and uses the number as a credibility measure for the content of that particular web page. Using this approach it might be possible to arrange the results set including images according to their PageRank number. However as the results of this study show, users might use images obtained from web-based resources in their presentations. Calculation of PageRank in such conditions would be rather difficult, unless the images were used in other web-based resources, or image users publish their presentations on the web.

The existing literature also shows that users consider a number of factors to assess the credibility of web-based resources. For example Kim et al. (1999) reviewed twenty-nine published articles and rating tools to identify the criteria proposed or employed specifically to evaluate the credibility of health related web sites. The authors identified 165 explicit criteria for assessing the credibility of health related web sites and grouped

¹ PageRank is exclusively registered for Google (BENINCASA, C., CALDEN, A., HANLON, E., KINDZERSKE, M., LAW, K., LAM, E., RHOADES, J., ROY, I., SATZ, M. & VALENTINE, E. (2006). Page Rank Algorithm. Available at <http://www.math.umass.edu/~law/Research/PageRank/Google.pdf>. Accessed date 12 March 2009).

the criteria under twelve categories including: content, design and aesthetics, disclosure of authors, sponsors, developers, currency of information, authority of source, ease of use and accessibility, and availability. Kim et al. (1999) also suggest that many “*authors agree on key criteria for evaluating health related websites, and that efforts to develop a set of key criteria may be helpful*”.

Kim et al. (1999) report that there are several factors which users consider to judge the credibility of health information obtained from web-based resources. We identified credibility as a relevance criterion for the relevance judgment of medical images; however, further research is required to know which factors affect health care professionals’ decisions concerning the credibility of image resources. Toms and Taves (2004) recruited eighty participants to investigate the user perceptions of web site reputation. They asked each participant to search on the internet for twelve topics and evaluate the credibility of the first twenty websites for each topic. Based on the findings of their study, Toms and Taves (2004) identified eleven factors affecting the assessment of credibility of websites: trustworthiness, authoritativeness, aboutness, willingness to return, will recommend, previous visit, interest level in topic, highly rated in relation to other sites assessed on the same topic, age of participants, education, and web browsing experience of the participant.

Rieh and Belkin (1998) interviewed six faculty members and eight doctoral students to find out how users judge the quality and authority of information obtained from the internet. From the analysis of the interview data, Rieh and Belkin (1998) identified seven factors used for the judgment of information quality: source, content, format, presentation, currency, accuracy, and speed of loading. They report that source was the most important factor mentioned by twelve out of fourteen participants. Rieh and Belkin (1998) add that the participants evaluated the source of information at institutional and individual levels. The institutional level involved the institutional characteristics of the source such as domain, type of institute and name of institute. The individual level involved the identification of the name of the author, and the author’s affiliation.

The factors identified by Rieh and Belkin (1998); Kim et al. (1999); Toms and Taves (2004) can be used by website developers as guidelines in developing credible resources. One possible solution is using metadata as Eysenbach et al. (2000) suggest. Eysenbach et al. (2000) propose MedCERTAIN¹ (MedPICS Certification and Rating of Trustworthy Health Information on the Net), a collaborative system for assessing the quality of health information on the internet. Bearing in mind the fact that the quality of health information on the internet cannot and should not be controlled by a central body or authority, Eysenbach et al. (2000) recommend that health information can be evaluated and labeled by metadata in a decentralized manner. They proposed a set of descriptive medical core metadata (known as medPICS (Platform for Internet Content Selection)) including authorship, qualification of authors, sources of funding and content keywords. Eysenbach et al. (2000) explain that the idea behind medPICS was to allow individuals and organizations to add metadata including credibility ratings to the information. The assigned metadata can be used by health information consumers to filter retrieved information according to the credibility rating levels. Eysenbach et al. (2000) propose four levels of credibility that can be added as metadata to the health information. They describe these levels as follows:

- Level 1 labels mean that the site is in 'good standing'.
- Level 2 labels mean that the website is monitored by a third party.
- Level 3 labels mean that the content of a website has been evaluated by medical experts or organizations.
- Level 4 labels mean that the content of a document (a web page) has been peer-reviewed by an independent third party.

Such frameworks can be used to add metadata indicating the levels of credibility of web-based image resources. Search engines can access and use such metadata to filter

¹ <http://www.medccertain.org/>

image search results according the levels of credibility desired by health care professionals.

Target audiences

Twelve participants were concerned about their target audience when selecting images. For example, if they wanted to show the images to their students, they would want to make sure that the images used were suitable for that particular group of individuals:

P20: It depends on your target audience. For example, if you want to teach first year medical students you want something typical. If you are teaching junior doctors you will need something settled. It depends on your audience.

Participants emphasized that an image can be more influential than text for educational purposes, and believed that some images convey the message more effectively to the target audience. It seemed that participants considered other criteria such as quality of image before they decided whether an image is suitable for the target audiences or not. One of the participants explained that after she ensured the topical relevancy, orientation, image quality and dimensional size of the image, she would decided whether the selected image was appropriate for the target audience:

P22: This is a classic picture of what I am looking for. I mean PCO [polycystic ovarian] ultrasound images. The picture is occupying the whole area and it is clearly illustrative. If I use it in the presentation the students will never forget it.

Availability

It was important for health care professionals to access images they found on the internet. Sometimes participants who used search engines, electronic journals and medical databases to find images, could not access full-sized versions of the images retrieved from a search, or the full text of papers. Google image search presents thumbnails of images stored in its cache, but participants wanted to see full size images

in order to assess and select them. Sometimes web pages containing images they requested had been removed or had been replaced by different images. Therefore health care professionals could not access those images. Although participants preferred to access images that were free of charge, sometimes they said they would pay to access images they needed:

P15: Unfortunately, the cost of journals is now very high for individual subscribers and the NHS no longer has an organizational subscription. There are many online journals. The ones I am more interested in are already accessible.

Originality

This is a criterion mostly used by two participants, P14 and P15, who worked in a Medical Physics Department. Since they wanted to analyze the content of actual medical images using techniques they had developed, they wanted original versions of the medical images, i.e. without any form of manipulation:

P14: If I am doing an analysis, I have to select PET¹ scans. I have to use all of the data, and I cannot pick the best one. I have to say I have ten patients and I analysed ten patients' original images.

The findings of the current study, and the research reviewed in chapter 2, support the assumption that topicality does not automatically result in relevant images and that users evaluate images with qualities that go beyond topicality. However, this does not mean that image retrieval based upon topicality is an inappropriate mechanism. As we discussed earlier, topicality plays an important role in the relevance judgment of retrieved documents including images. However, we believe that information retrieval systems can be designed to incorporate criteria other than the topicality. The results of

¹ Positron emission tomography

this study, and the studies we discussed, indicate that users can determine whether those criteria are applicable for a retrieved document, or they can decide whether or not the document will provide the information they need. By identifying the criteria applied by the users, we may be able to incorporate the relevance criteria into the image indexing and retrieval process and consequently take image retrieval beyond the topical approach. Hollink et al. (2004) report that practical improvements in document indexing and retrieval have been achieved through the introduction of the Dublin Core Metadata standard¹ and Visual Resources Association² (VRA) Core Categories. Metadata, i.e. data about data, have been widely discussed within the information retrieval community for many years (Geueke and Stausberg, 2003). A collection of metadata is comparable to catalogues in libraries. Geueke and Stausberg (2003) report that metadata include objective and subjective data. For example, one could describe the format of a document and another one describes its theme. The Dublin Core Metadata standard is used to add metadata to a wide variety of web-based resources such as video, sound, image, text, and composite media like web pages; however, Visual Resources Association (VRA) Core Categories is only a standard for adding metadata to images.

¹ <http://dublincore.org>

² <http://www.VRAweb.org>

**Table 5-3: Dublin Core metadata element set and definitions
From The Dublin Core Metadata Initiative (2008).**

Elements	Definition
Title	A name given to the resource.
Creator	An entity primarily responsible for making the resource.
Subject	The topic of the resource.
Description	An account of the resource.
Publisher	An entity responsible for making the resource available.
Contributor	An entity responsible for making contributions to the resource.
Date	A point or period of time associated with an event in the lifecycle of the resource.
Type	The nature or genre of the resource.
Format	The file format, physical medium, or dimensions of the resource.
Identifier	An unambiguous reference to the resource within a given context.
Source	A related resource from which the described resource is derived.
Language	A language of the resource.
Relation	A related resource.
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.
Rights	Information about rights held in and over the resource.

The fifteen metadata elements¹ of the ‘Dublin Core’ described in Table 5-3 are part of a larger set of metadata vocabularies and technical specifications maintained by the Dublin Core Metadata Initiative² (The Dublin Core Metadata Initiative, 2008).

According to the Visual Resources Association (2009), the VRA Core Categories consist of a single elements set that can be applied as many times as necessary to create records to images. The elements set in VRA Core Categories contains seventeen elements and describes various aspects of the context of images (see Table 5-4). For instance, all information about the production or manufacturing processes, techniques, and methods incorporated in the fabrication of the image must be recorded under the element ‘technique’.

¹ Note that each element in the Dublin and VRA metadata is optional and can be repeated several times for a document.

² DCMI: see <http://dublincore.org/documents/dcmi-terms/>

Table 5-4: Visual Resources Association elements set and definitions

Elements	Definition
Record Type	Identifies the record as being either a work record, for the physical or created object, or an image record, for the visual surrogates of such objects.
Type	Identifies the specific type of work or image being described in the record.
Title	The title or identifying phrase given to a work or an image. For complex works or series the title may refer to a discrete unit within the larger entity (a print from a series, a panel from a fresco cycle, a building within a temple complex) or may identify only the larger entity itself. A record for a part of a larger unit should include both the title for the part and the title for the larger entity. For an image record this category describes the specific view of the depicted work.
Measurements	The size, shape, scale, dimensions, format, or storage configuration of the work or image. Dimensions may include such measurements as volume, weight, area or running time. The unit used in the measurement must be specified.
Material	The substance of which a work or an image is composed.
Technique	The production or manufacturing processes, techniques, and methods incorporated in the fabrication or alteration of the work or image.
Creator	The names, appellations, or other identifiers assigned to an individual, group, corporate body, or other entity that has contributed to the design, creation, production, manufacture, or alteration of the work or image.
Date	Date or range of dates associated with the creation, design, production, presentation, performance, construction, or alteration, etc. of the work or image. Dates may be expressed as free text or numerical.
Location	The geographic location and/or name of the repository, building, or site-specific work or other entity whose boundaries include the work or image.
ID Number	The unique identifiers assigned to a work or an image.
Style/Period	A defined style, historical period, group, school, dynasty, movement, etc. whose characteristics are represented in the work or image.
Culture	The name of the culture, people (ethnonym), or adjectival form of a country name from which a work or image originates or with which the work or image has been associated.
Subject	Terms or phrases that describe, identify, or interpret the work or image and what it depicts or expresses. These may include proper names (e.g. people or events), geographic designations (places), generic terms describing the material world, or topics (e.g. iconography, concepts, themes, or issues).
Relation	Terms or phrases describing the identity of the related work and the relationship between the work being catalogued and the related work. Note: If the relationship is essential (i.e. when the described work includes the referenced works, either physically or logically within a larger or smaller context), use the title. Larger entity element.
Description	A free-text note about the work or image, including comments, description, or interpretation, that gives additional information not recorded in other categories.
Source	A reference to the source of the information recorded about the work or the image. For a work record, this may be a citation to the authority for the information provided. For an image, it can be used to provide information about the supplying agency, vendor or individual; or, in the case of copy photography, a bibliographic citation or other description of the image source. In both cases, names, locations, and source identification numbers can be included.
Rights	Information about rights management; may include copyright and other intellectual property statements required for use.

Dublin core metadata and VRA core categories are used to add metadata to documents including images. We investigated whether the Dublin Core Metadata Element Set and Visual Resources Association (VRA) Core Categories could address the criteria we identified. The investigation was based on the definitions of each element and the definition of relevance criteria as presented in Table 5-3, Table 5-4 and Appendix 9. We

note that some criteria can be addressed by more than one element, or a combination of elements is required to address a single criterion. For example, the criterion 'recency' can be addressed simply by the element 'date', whereas credibility of the image source is dependent on the creator, publisher, source and publisher elements. Table 5-5 shows the possible incorporation of the relevance criteria in the image retrieval process. We found that some of the relevance criteria can be addressed by metadata. This enables users to incorporate some characteristics of images in the retrieval process. We also note that some elements can address more than one criterion. For instance, the format element can address technical information, size, magnification, originality, image quality, colour and modality of medical images (see Table 5-5).

Table 5-5: Addressing medical image relevance criteria by Dublin and Visual Resources Association (VRA) Core Categories.

Dublin Core Metadata Element Set, Version 1.1	Equivalent element in VRA Core Categories	Medical image relevance criteria addressed
Title	Title	Topicality
Creator	Creator	Credibility
Subject	Subject, Style/Period	Topicality, technical information, modality, age and gender
Description	Description	Topicality, modality, magnification, age and gender, technical information
Contributor	Creator, Location	Credibility
Date	Date	Recency
Type	Record Type, Type	Modality, colour
Format	Measurements, Material, Technique	Technical information, size, magnification, originality, image quality, modality, colour
Identifier	ID Number	
Source	Source	Credibility, technical information
Relation	Relation	
Coverage	Date, Location, Style/Period, Culture	Topicality
Rights	Rights	Copyright
Publisher		Credibility
Language		

The relevance criteria we identified highlighted the importance of descriptive metadata that health care professionals felt would be particularly useful for image search engines or medical image collections: image quality, size, technical information, colour, copyright status, and modality of images. The participants tried to overcome the

problem of missing metadata, though most of the time their attempts were not successful (e.g. including modality of images in the search terms).

Moreover, Dublin core metadata and the Visual Resources Association elements sets are introduced for general purposes. In medical care, metadata define demographics, diagnoses and care of patients for the purposes of documentation, communication, transaction and monitoring. Therefore, such metadata can be created and used by standards such as Digital Imaging and Communication in Medicine (DICOM¹). Although DICOM is a standard for archiving and transferring medical images, it allows medical staff to add metadata (known as DICOM tags) to medical images. According to Duque et al. (2003) each record in DICOM is made up of a header containing metadata followed by single or several image slice(s). DICOM metadata can be classified into two groups:

- 1- Patient-related metadata such as the patient name, gender, age, the radiologist's name, the name of hospital or health centre. All clinical metadata that are not originally part of the image acquisition such as notes by medical experts can be added as patient-related metadata.
- 2- Image-related metadata such as the image acquisition device, constructor and parameters, the acquisition date, the number of images stored, the size of images and the orientation.

Lehmann et al. (2000) report that systems such as DICOM and PACS² allow adding metadata and textual information about the examination carried out; however, most of the time medical staff do not enter sufficient data into the systems. Thus, text-based retrieval will neither result in complete nor in sufficient information. For example, a physician wants to compare all images of the patient's abdominal region regardless of modality or orientation of images. Lehmann et al. (2000) described this type of query as a 'primitive query' in medical diagnostics, and state that primitive query can be processed successfully by CBIR systems such as IRMA, which categorizes images according to anatomic region, image modality, body orientation and biological system (e.g. respiratory system). Lehmann et al. (2000) also suggest that CBIR can be used for

¹ See <http://medical.nema.org/dicom/> for more information about DICOM.

² Picture Archiving and Communication Systems

searching for representative images of known diseases and medical conditions, described as a 'semantic query'. Lehmann et al. (2000) mention that content-based retrieval by examples of the search patterns is one of the main activities during diagnostic procedures, especially under unknown circumstances. For instance, a radiologist wants to retrieve all X-ray images showing Monteggia fracture. Lehmann et al. (2000) add that CBIR techniques can be used to categorize images in the results set based on criteria such as modality since reviewing a large number of retrieved images is time-consuming for health care professionals. Lehmann et al. (2000) described this type of application as browsing.

The results of the current study are also favourable for applying content-based image indexing and retrieval methods to medical images. However, the difficulty of developing content-based indexing and retrieval tools is due to several factors. Firstly, in most cases, medical images are intensity-only images containing less information than non-medical images. Additionally, different types of images (e.g. MRI and ultrasound images) may be acquired of the same anatomic area in particular medical conditions. Each type of medical image needs an additional registration procedure (Glatard et al., 2004). Secondly, medical images are usually low resolution and high noise images. Thus, it is difficult to automatically analyze and extract their visual features. Medical images obtained with different devices, even using the same modality, may have significantly different properties (Lehmann et al., 2000). Thirdly, medical images could be indexed on medical criteria that are extremely variable depending on the kind of image acquisition considered (e.g. imaged anatomic area and clinical context). However, medical images interpretation is often difficult even for trained medical doctors. A holiday picture might bear enough information without text; however, for medical images the text is essential. For deeper analysis of medical images, detailed textual information is required because medical images themselves are dependent on the context (Muller et al., 2005). Low-level visual attributes of images such as colour or size of image were expressed as the main criteria by the participants. As Figure 5-17 indicates nearly half of the criteria were visual (e.g. modality), though sometimes the main focus of the medical image search was on the context information (e.g. technical information)

requiring high-level human reasoning. Indexing on a high level of abstraction is currently possible only by using textual descriptions, as Markkula and Sormunen (2000) reported.

The health care professionals in our study used Google image search and medical databases supporting traditional textual query operations. Thus, it is difficult to explain how they would change their medical image searching behaviour if they could execute queries based on the visual similarity of photographs. It is also difficult to prejudge common uses for purely visual queries without textual search keywords. Perhaps the first challenge would be how to formulate a visual query for medical image retrieval. Moreover, the number of medical images on the internet and the heterogeneity of medical images might make content-based medical image retrieval problematic. Nevertheless, a content-based approach can be used in the classification of the set of images retrieved by search keywords or index terms. Within the retrieved set of images, visually similar images could be grouped together and the output of image retrieval could be organised by these groups. Thus, the end user can see different image categories contained in the retrieved set of images. The selection criteria applied by the health care professionals also suggest some ideas for organising the image set. For instance, if the user was looking for T-2 weighted MRI scans of the brain, they could be grouped according to their dimensional size or modality.

5.3. Quantitative analysis of relevance criteria

Figure 5-17 gives a summary of all the criteria mentioned by participants during the study. Fifteen relevance criteria were summarized from the interviews and we calculated the percentage for each criterion we identified. The percentage give for each criterion in Figure 5-17 means the number of participants who mentioned each criterion. We believed that this figure show the weight of each relevance criterion mentioned.

Among the criteria used by all twenty-nine participants, the largest proportion went to visual (non-textual) criteria which related to visual attributes of medical images. The next was textual criteria, which related to the participants' domain knowledge, descriptions of images and textual attributes of an image including age and gender,

copyright, recency and technical information. Finally participants relied on other criteria. Thus, it could be concluded that participants relied mostly on the visual attributes of medical images to make relevance judgments. Topical relevancy (100%) and image quality (93.1%) were the two most frequently mentioned visual criteria. However, age and gender (68.97%), grouped in textual criteria, was also an important criterion.

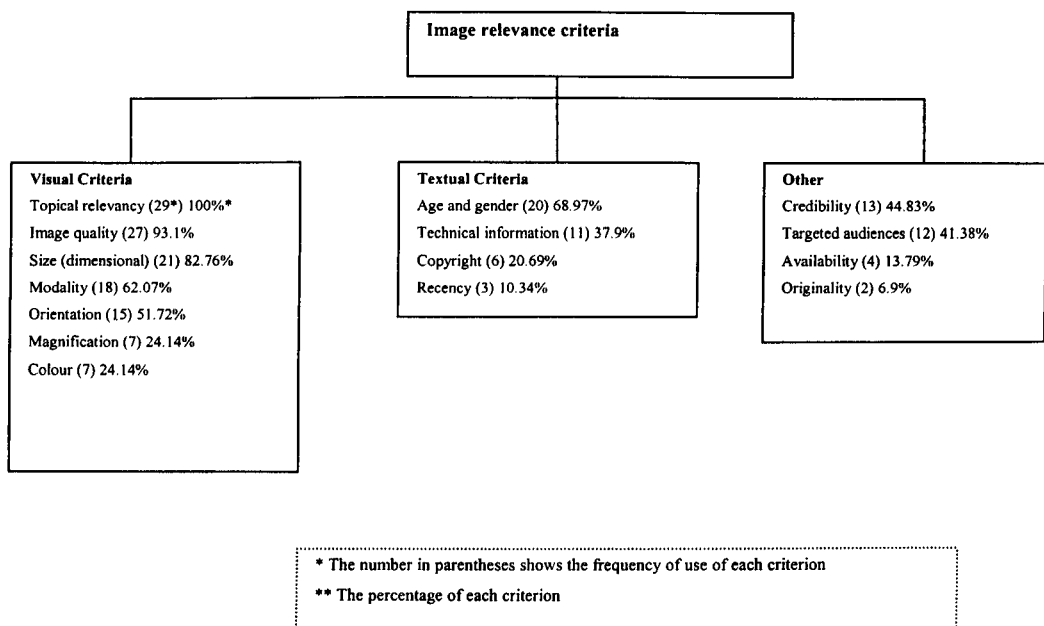


Figure 5-17: A summary of the criteria mentioned by all of the participants.

It can be seen from Figure 5-18 that topical relevancy, image quality, size (dimensional), age and gender, and modality are the top five most frequently applied criteria by all the participants. These five criteria were mentioned by more than two-thirds of the participants.

What can be concluded from the study of relevance criteria for medical images applied by health care professionals is that our findings show the diversity of the selection criteria which health care professionals apply when judging the relevancy of images. The common criterion health care professionals used was a suite of criteria grouped under the term, 'visual criteria'. Our interviews show that the participants generally use

topicality as the first criterion to judge the relevancy of images, before applying other criteria such as image quality and dimensional size of the image for their final selection. In the research literature on relevance, topicality is mainly defined as the relationship between the topic (query terms) and content (theme) of documents. Our findings show that the participants judged the topicality of medical images using textual descriptions and the visual appearance of images. We believe that topicality is an independent criterion and can be used to judge the relevancy of all types of documents in general and medical images in particular.

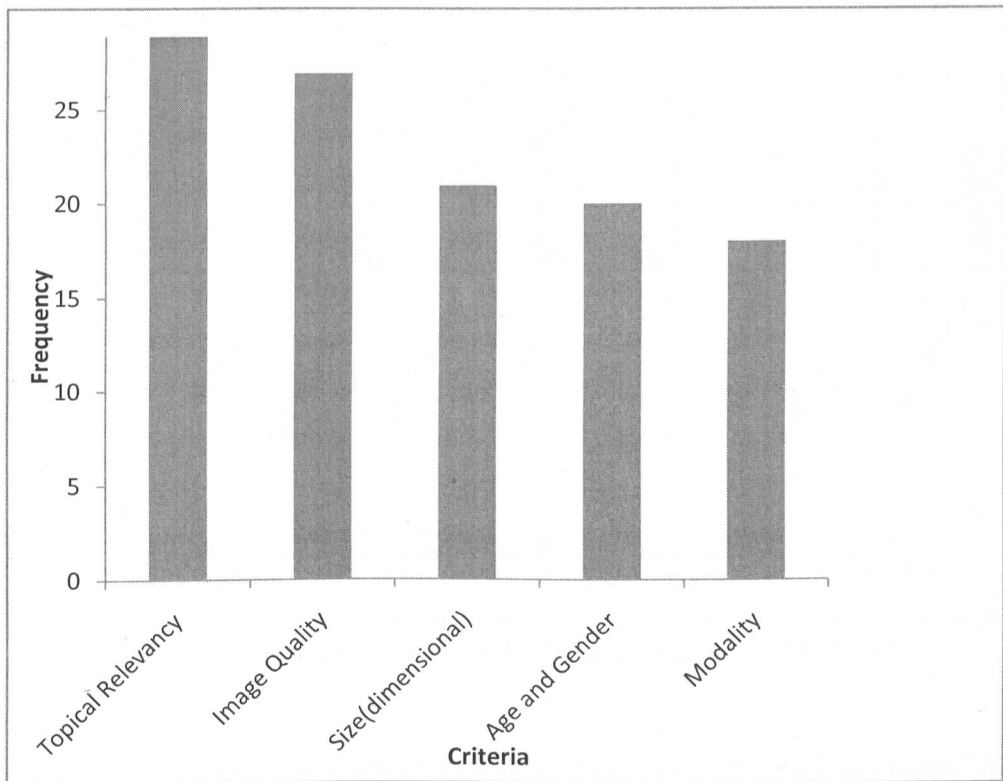


Figure 5-18: Top 5 most frequently mentioned criteria by participants.

5.4. An experiment with ImageCLEFMed topics

ImageCLEFMed is an international evaluation campaign for medical image retrieval. It is the medical track of ImageCLEF which was established as a part of the Cross Language Evaluation Forum (CLEF¹) in 2003. CLEF itself is an offspring of the Text Retrieval Conference (TREC²). The topics of the track were created after conducting surveys and examining the search logs of a number of medical search systems. The type of relevance judgments applied in ImageCLEF is generally referred to as topical relevance (Müller et al., 2006).

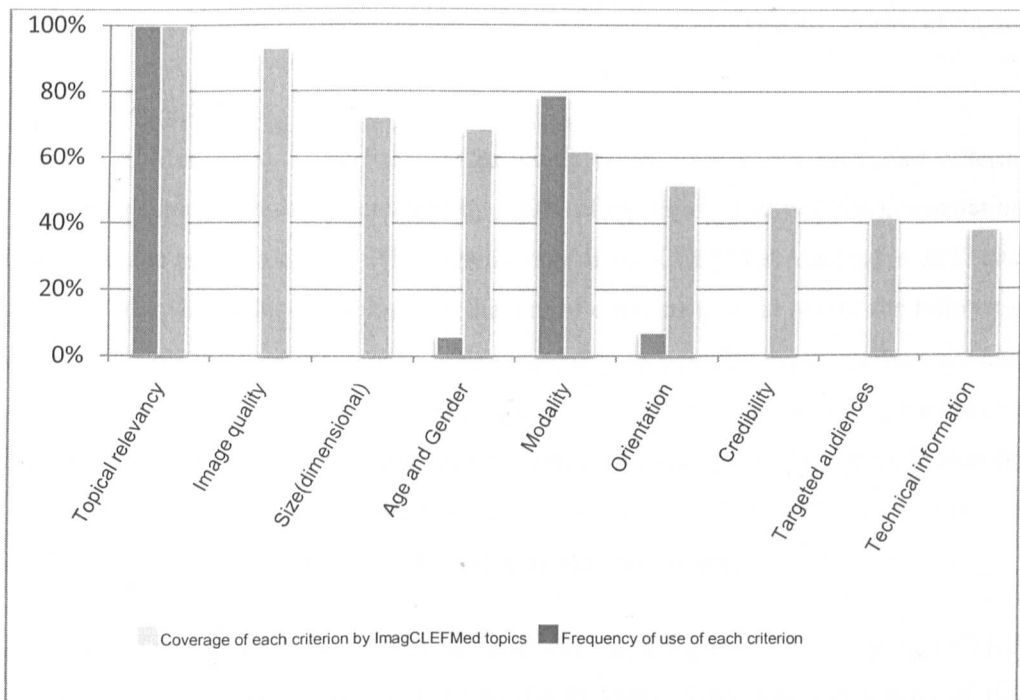


Figure 5-19: Coverage of relevance criteria by the topics of ImageCLEFMed.

Our findings showed users apply an apparently wider range of criteria to evaluate the relevancy of medical images to their situational medical image needs. We decided to

¹ <http://www.clef-campaign.org> (site accessed: 16/08/2008).

² <http://trec.nist.gov/>

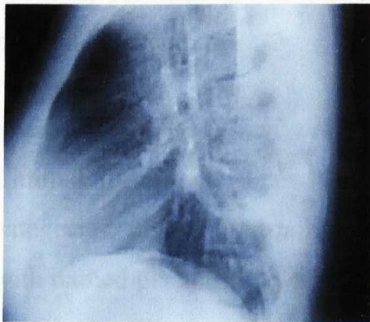
contrast 85 topics used in ImageCLEFMed 2005, 2006 and 2007 with the relevance criteria we identified.

The assessment of the coverage of criteria was based on the text of each ImageCLEFMed topic. For example, if in the text the modality of an image was mentioned, we recorded that the criterion 'modality' was covered. There were also topics where particular criteria were implied. For example, a query about cancer of the ovary implied a particular gender and for this topic, gender was recorded. We focused on the nine most common of the fifteen criteria that health care professionals applied (specified by >25% of the participants). The results of our analysis are shown in Figure 5-19.

ImageCLEFMed topics were found to explicitly cover four of our identified criteria. They were topical relevancy (covered by 100% of topics), modality (79%), orientation (7%), and age and gender (6%). The organizers of ImageCLEFMed reported in 2007 and 2006 that they had selected topics with the aim of covering at least two of the following criteria (which they referred to as axes): modality, anatomic region, pathology, and visual observation (Müller et al., 2008; Müller et al., 2007). Accordingly, the criteria such as topicality and modality were given significant coverage by the topics. However, the following criteria were by design not covered by ImageCLEFMed topics: image quality, size (dimensional), credibility and technical information.

According to Müller et al. (2007) the majority of images in the ImageCLEFMed collection contain annotations. As we can see in Figure 5-20, age and gender of the patient has been mentioned in the annotation of the image. In addition, in the abstract, we found some technical information. Moreover, the type of image has been stated, which is a conventional radiograph. Therefore, it might be possible to create annotation tasks that address at least some of these criteria. We also noticed in our study that users regarded the credibility of images as important due to the diverse range of sources being searched. However, ImageCLEFMed ensures that its image collection is drawn from teaching files which are considered as credible sources. Therefore, this is not a criterion that could currently be tested in ImageCLEFMed. However, it is clear that searchers

often retrieve images using collections where this is an important issue: for example, the searchers in our study often searched on Google Image, where credibility is critically important. In such situations, building a medical image search engine that ensures only highly credible images are retrieved could be an important research challenge. If ImageCLEFMed was looking to expand its campaign to new fields of research, a study of the credibility of an image source might add a challenging new line of exploration. This could be achieved by building a collection of medical images sampled from a wide range of web sites where the task for system developers is to locate both credible and relevant images.



Title: PNEUMONIA RLL — DOES THE CT HELP GIVE US SOMETHING OTHER CAUSE THAN ROUTINE BACTERIAL PNEUMONIA

Abstract: 15-year-old white female who presented to the ER October 30, 2004 with chest pain and shortness of breath. She was feeling well until Thursday, October 28, 2004, when she developed a sore throat at a football game. On October 29, 2004, she had decreased appetite and began developing some shortness of breath as well as low back pain and cough. The cough was noted to be productive with green yellowish sputum. It was thought at that time that she had some viral illness. She has a known past history of asthma. She also noted increasing fatigue.

There was some questionable history of calf pain in both of her calves while running last week; however, she denies any recent travel or trauma

She does have an old meniscal tear on her right knee. Also of note, she was started on the oral contraceptive pill approximately five days ago which was started for dysmenorrhoea and menorrhagia symptoms. Due to her continued chest pain and shortness of breath, she was seen in the ER. At that time, chest x-ray was normal and chest CT showed a right pulmonary artery density thought to be a clot in a right lower lobe, questionable infarct. On admit, white count of 6.4, hemoglobin 12.4, hematocrit 36, platelets 192. PT of 15.8, PTT is 30. Sodium 141, potassium 2.3, chloride 101, CO2 26, BUN 10, creatinine 0.7, glucose 117, calcium 9.3, unconjugated 0.3, conjugated 0.0. Alkaline phosphatase 71, albumin 4.4, protein 7.3, AST 22, ALT 18. UA clean catch was normal. During hospitalization, she has been on Rocephin and Azithromycin but still has been spiking temperatures. Her CRP was 8.9 on 10/31 and was 13.9 on 11/2.

Findings: Even though there are less air bronchograms than usual, the appearance is more suggestive of bronchogenic than hematogenous origin.

Discussion: No specific signs are seen pointing to the organism. Unfortunately, many atypical organisms that confound the clinicians have nonspecific appearances. The patient has no adenopathy, effusion, cavitation that delay the response. Unfortunately, gray images don't make up for gram stains.

Pathology: Infection

Anatomy: Chest

Modality: CT, Conventional Radiograph

Figure 5-20: A sample image from ImageCLEFMed image collection
(The original size of image was 10.16 x 13.55 cm, and we extracted annotation from the XML file for this image. Image and annotations adapted from the ImageCLEFMed 2007 test collection (used with permission)).

5.5. Image information needs

In a number of studies, researchers investigated the users' image queries. Yang (2005) cites Panofsky (1972) and reports that Panofsky distinguished between three levels of comprehension for images: pre-iconographical, iconographical and iconological. Pre-iconographical understanding is related to the general subject matter of the image (e.g., a dog). It requires only an everyday familiarity with objects and events (Panofsky, 1972). The second level, iconographical understanding, relates to the specific subject matter of the image (e.g., John F. Kennedy), and requires additional domain knowledge or linguistic cues. Hence, iconographic meaning is culturally determined. Iconological meaning relates to emotional, abstract meanings and symbolic aboutness and denotes the intrinsic, personal meaning of an image (e.g., he is my idol). Thus, iconology is both personal and cultural.

Enser et al. (1993) analyzed users' image requests at the Hulton Deutsch Collection Limited in Europe. They found that users' requests can be grouped in two main categories: unique, and non-unique. Unique requests were defined as those concerned with named persons, one-off events, objects or locations. An example of a unique query is 'George III'. An example of a non-unique query is '5-6 year-old boy in silhouette'. They report that both classes of query require refinement in terms of time, location, action, event or technical specification. Enser et al. (1993) add that almost 70% of the image requests were for a unique person, object or event, and 34% of the requests were refined (mostly by time).

Fidel (1997)'s study also provides valuable classifications of users image queries. She categorizes images retrieved for 100 actual requests, submitted in an agency, which had a large collection of stock photos, into two groups: the 'data pole' and the 'objects pole' (see Table 5-7). She explains that images are considered as sources of information at the 'data pole'. For example, a physician may need to use a slide of a normal foot to help decide if a patient's foot is flat. On the other hand, images are used as objects at the 'objects pole'. For example, a slide librarian may be asked to find slides that represent a specific idea or object. There are also some in-between cases such as medical instructors

and art historians who want to retrieve images both as information sources and as objects. Fidel (1997) believes that a medical instructor may look for a good slide of a normal foot for a class she teaches. She wants the slide to have the information required about a normal foot, but at the same time, she is looking for the best slide as an object: the one taken from a useful angle or with an image big enough to be projected in a classroom.

Table 5-6: Summary of Data Pole and Objects Pole (Fidel, 1997).

Data pole	Objects pole
Images provide information	Images are objects
Relevance criteria can be determined ahead of time	Users will recognize relevance criteria when they see them
Relevance criteria are specifications of which the user is aware	Relevance criteria are latent, and are invoked when viewing images
It is possible for users to explain why an image is relevant	It might be difficult for users to explain why an image is relevant
Images can be retrieved with textual and other verbal clues	It might be difficult to find verbal clues for retrieval, clues are often visual
Colour, shape, and texture can convey information and, therefore, are important for retrieval	No evidence exists that colour, shape, and texture are important for retrieval
Images must include similar information to satisfy the same need	Two very different images may satisfy the same need
Ofness often equals aboutness	Ofness is likely to be different from aboutness
Biographical attributes are not likely to play a role	Biographical attributes are important for relevance assessment
To satisfy requests may require sets of more than one image	Requests are usually satisfied with one image
May not require browsing through the whole answer set	Requires browsing the whole answer set
Browsing is time consuming	Browsing can be done rapidly

Markkula and Sormunen (2000) analyse journalists' routine illustration tasks and also their queries in an image archive in Aamulehti, a daily newspaper in Finland. They classified journalists' queries in four categories: concrete objects (i.e. named persons or places), themes or abstractions interpretable from the images, background information

on the photograph (such as documentary information), and known images. They point out that requests for concrete objects dominated the use of the photograph archives (81% in their requests, and 56% in their illustration tasks). Choi and Rasmussen (2002) analyze requests submitted by faculty and students of American history searching for images in the American Memory image collection. Their investigation demonstrates that most users looked for general/nameable images. They also report that date, title, and subject descriptors are key factors representing images.

The findings of the current study, however, indicate that the image information needs of health care professionals could be categorised into two broad categories, general and specific medical images, based on the type of images sought:

General medical images: this is when participants looked for general medical images on a particular topic, for example looking for images of anatomic organs (such as images of female reproduction organs). Thus we believe that this type of image request is equal to pre-iconographic as Panofsky (1972) mentioned. This type of image query also could be classified as non-unique as suggested by Enser et al. (1993). Participants mentioned that they mostly looked for this type of image for educational purposes, and they used general medical images to illustrate their presentations. They described this type of image by expressions such as '*generic images*', '*common medical images*', '*classic picture*', '*general images*', '*images of common issues*', '*educational images*', '*a typical picture*', '*common medical conditions*' and '*something popular and well-known*'. The findings of our study show that participants could find general medical images more easily than finding specific medical images.

P18: If I want to show common medical conditions I could find images easily.

P21: It depends on what you are looking for. If it is something popular and well known it is easy to find what you need.

Specific medical images: this is when participants looked for images to illustrate their research findings and publications. They also used specific medical images for clinical purposes. Participants used expressions such as 'specialized medical images', 'something rare or not common', 'odd things', 'rare cases', 'detailed images', 'specific images', 'clinical images' and 'images for rare conditions' when they wanted to distinguish between general and specific medical images. Since the health care professionals were concerned with the specific subject matter of the image, this type of image resembles the iconographical or unique type of image query suggested by Panofsky, (1972) and Enser et al., (1993).

Participants usually looked for specific medical images to compare their research findings to, or to make clinical decisions. Thus the participants wanted to obtain the best and most recent images available to them. For example, a participant who looked for images of TLR-3¹ in fallopian tubes mentioned that he needed these images to compare his research findings with the findings of other researchers. He mentioned that due to the novelty of the topic he could not find images in web-based resources, but he found them in medical journals. He said:

P8: I was looking for images [of TLR-3 in fallopian tube] in published papers because it is something new. I know there are few published papers about this topic. I could not find these images in books or the Internet. We are one of the first research groups who worked on this topic. I use images to compare our findings with the findings of other research groups.

Participants seemed to have difficulties when they looked for specific medical images. The difficulties can arise for various reasons including the time that participants might spend to find images, the availability of the relevant published papers in the field,

¹ Toll-like receptors

novelty, and rareness of the topic. As one of the interviewees explained, while he could find general medical images for common topics such as tooth filling easily, for specific medical images on topics such as amalgam restoration he would spend a lot of time searching for images in medical journals.

P1: It is easy to find general images [common issues] like tooth, tooth filling. However, if I need specialized images I would go to find them in papers. To find special images for amalgam restoration I have to spend a lot of time because it takes time to review the articles to find the images that I need.

Other interviewees also made similar remarks. Three participants mentioned that they could not find the images they needed due to novelty and the lack of relevant literature discussing the topic:

P3: It is not always easy to find what I want. For example, if it is something rare or not common it will only be available in a few Web sites or articles. However, if you are looking for images of common medical conditions they will be easy to find.

P4: Yes, it was difficult to find images, because few people worked in this area and they published few articles. The number of images was limited, so it was difficult to find relevant images from those articles.

Our findings support the idea that participants might make decisions about the image resources they use with regard to the type of image they required. Further investigation of the interviews revealed the fact that the health care professionals tended to use images published in papers and personal collections when they looked for specific medical images.

P1: If I want to look for common medical images, I would search on the Internet, but for specialised medical images I would use my personal collection.

Other findings indicated that participants cared about visual criteria more than other criteria when they looked for general medical images. For example, participants such as P3 and P8 who looked for general medical images for educational purposes, tend to have a better ability for interpreting the images. Therefore, they considered visual criteria for the relevance judgment of general medical images:

P3: Sometimes I want to show an image to my students and explain a medical condition to them. In this situation, the image itself is sufficient, because I can explain what the image illustrates.

However, when they looked for specific medical images they seemed to consider criteria such as textual and other criteria in addition to visual criteria. Sometimes the participants mentioned that they considered criteria such as credibility of the image source to compare and decide about the appropriateness of images for their information needs. We noted that participants wanted to ensure that they had found the most relevant images when they looked for specific medical images to illustrate their scientific publications or when they needed similar specific medical images to compare their research findings with other researchers' findings. For example, a researcher (P8) from the Department of Reproductive and Developmental Medicine stated that he would consider criteria such as the credibility of image sources when he used images for comparing the findings of his study, as presented in the images he had produced, with those of other research groups:

P8: The criteria that I use to evaluate images will depend on my purpose. For example, if I do research on a topic, I prefer to find the latest and the best images from the papers published in the most reliable journals. Then I will compare my images, which represents my research findings, with the findings of other researchers. However, if I want to find an image for educational purposes, I just want to show the image and I do not need to find a detailed image.

We conclude this section by saying that the type of images requested, the queries used, and the relevance criteria applied, indicate that health care professionals use medical images as sources of information and as objects. This finding corresponds with the results of Fidel (1997)'s study that image users evaluate images as a source of information or as an object. For example, the participants considered medical images as a source of information when they applied topicality as the minimum requirement for the relevance judgment of images. Medical images were also regarded as objects when the participants applied criteria such as 'image quality' or dimensional size of images.

5.6. Motivation for image searching

Although we classified the type of images health care professionals needed into two types (general and specific images), we also examined how images were used (Table 5-7). The results of our study show that health care professionals required images for the following purposes: to meet clinical aims, education, research, publication and documentation. Table 5-7 presents the motivation of health care professionals for searching for images. Twenty participants (68.96%) mentioned education and research as the two most frequent motivators. This indicates the importance of medical images for educational and research purposes in health care and biomedicine.

Table 5-7: What motivates health care professionals to look for medical images.

Motivation	Frequency (percentage)
	Total=29
Education (self-instruction, patient education, public education, teaching students and colleagues)	20 (68.96%)
Research (compare research findings with prior works, obtain technical information about a test or an operation)	20 (68.96%)
Patient care (diagnostics, treatment and follow up)	13 (44.83%)
Publication (books, articles, posters, reports and presentations)	12 (41.38%)
Documentation of operations and medical examinations and tests (for legal reasons, quality assurance, recording the results)	8 (27.59%)
*Note that the respondents mentioned more than one motivator and the cumulative percentage is more than 100 percent.	

Some participants stated that images help students to learn more effectively compared to text. One of the participants, an orthopaedic surgeon, expressed the importance of images in medicine and said that:

P3: Images are important in medical education. Using images is the best way of showing a medical condition to students. If I want to teach students who have not seen this fracture [Monteggia fracture]. I cannot explain it unless I show them an image of Monteggia fracture.

Most of the time participants expressed the importance of images for educational purposes.

P20: Images are increasingly important. An image is worth a thousand words, isn't it? You cannot teach without them [images].

P19: I use images because a picture is worth a thousand words. It is much easier to communicate with images than with words. If you show pictures to students, they absorb the information much more quickly.

Further investigation of the interviews indicated that health care professionals mostly looked for general medical images when they used images for educational purposes. The findings also showed that health care professionals applied fewer criteria when judging the relevancy of images for educational purposes. Additionally, we noticed that they tended to use images stored in web-based resources when they searched for images for educational purposes. However, further research is needed to substantiate this claim.

As Table 5-7 illustrates, health care professionals used images for a variety of research purposes (twenty out of twenty-nine). The findings of the study showed that health care professionals used images to present their research findings and compare their research findings with the findings of other researchers (e.g. P4).

P4: We also use images to present our findings. In molecular science and biotechnology you use images to record your data and for documentation. Without images, we cannot evaluate our work and show the differences between our results and findings of other researchers.

We noted that health care professionals looked for specific medical images when they needed images for research purposes. We also found that there was a preference for using images from personal collections and from journal articles when the health care professionals required images for research purposes. However, additional research is required to confirm these results. The education and research motivators were followed by clinical aims (44.83%) and publication (41.38%). Studies exploring the information seeking behaviour of health care professionals have found that patient care is the most frequent reason for seeking information (Covell et al. (1985); Gorman (1995); Shelstad and Clevenger (1996); Ely et al. (1999)). This study found that the image seeking behaviour of health care professionals differs from the previous studies of information seeking behaviours of health care professionals, as patient care had been the main reason for searching for information.

We compared the roles of participants and their image seeking motivation to investigate the overlap between image information needs and roles of participants. As can be seen in Figure 5-21, there were significant overlaps between the role of participants and their motivation for image search; in particular for education as their motivation and the educational role of participants. We did not investigate the main role of participants who participated in this study. We also did not ask participants to rank their motivations for image seeking. Therefore we were unable to make any relationship between the role of participants and their motivation. Further investigation is required, however, to determine how health care professionals prioritized their motivation. We found that the image information needs of health care professionals influence the relevance judgment and medical image selection process conducted by health care professionals; however, further research is needed to investigate whether there is a relationship between the main roles of health care professionals and their pictorial information needs.

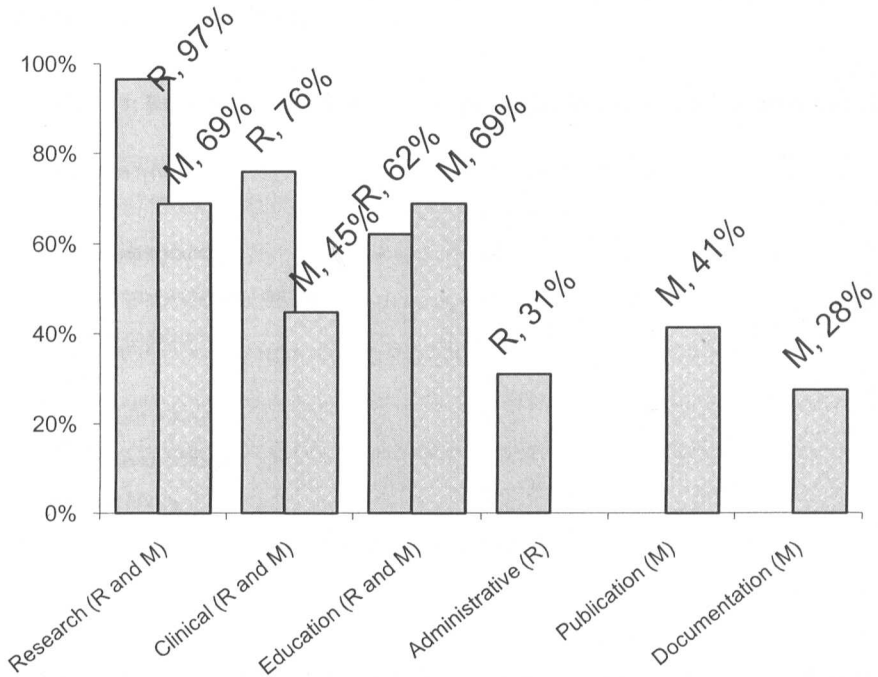


Figure 5-21: A comparison between the roles and image seeking motivators of the participants.

5.7. Medical image resources

Participants used a variety of resources to find the medical images they wanted. One of the major findings of our study is that online medical image collections were not a common resource for images used by health care professionals. As stated earlier, in this study we did not focus on any specific image collection. Based on the findings of the present study, it appears that despite the recent progress in medical image collections, health care professionals preferred to look for medical images in web-based resources using Google image search, or use images published in medical journals. Table 5-8 gives the list of resources used for medical images. The total is not based on a hundred

percent because more than one resource might have been used. As we can see in Table 5-8, the web-based resources have been used the most, followed by using images from papers and personal collections.

Table 5-8: Resources used by health care professionals for medical image retrieval.

Image resources*	Number of Participants (total=29)
Web-based resources	26
Papers from journals and databases	24
Personal collection	18
Books	12
Friends and colleagues	8
Departmental collection	7
CDs and DVDs	2

* Participants used more than one resource to find images they required.

In addition to web-based resources, health care professionals also used papers published in medical journals to find the medical images they required. They typically looked for relevant papers in medical databases such as PubMed, and then looked inside the articles to find images that they needed. Personal collections were always the third resource participants chose to find images. The majority of participants mentioned that they would use images from their personal collections to present their research findings. Books seemed to be another important resource for health care professionals to obtain medical images. Although images obtained from the books were not new, they could be used for educational purposes.

5.7.1 Web-based resources

In response to the question: 'Why do you look for images in web-based resources?', two participants stated:

P3: Looking at the images on the screen is better than looking at images in books. Because it is closer to reality.. When they take the x-ray, scan it or put its photo in the book and print it, the contrast is not like real x-ray sheets, but on the screen the image is natural. I personally prefer to find images in websites and see them on the screen because you see the real picture.

P5: The first source for the images I need is the Internet. The online image searching is interactive and you have the chance to interact with the system, try more keywords, and modify the query easily.

We noticed that participants obtained images from web-based resources in two different ways: direct and indirect. In the direct method, images were obtained from web-based medical image collections or professional websites. Our findings show that only 31.03% (nine out of twenty-nine) of participants were aware of web-based medical image collections. Medical image collections seemed to be unpopular among participants due to the lack of diversity of images in web-based image collections, and the low number of images available. One of the participants stated that for her the number of images in web-based medical image collections is important. She maintained that due to the limited number of images relevant to her field in such collections she preferred to look for images in other resources.

P9: There are a few collections, one I think is the website of a drug company, but I usually begin with Google. I know that there are some collections but I do not know how extensive they are. As you can see on their website [she went to one of the websites she uses] there is a picture gallery here for bone research. So here are some images and here are some teaching slides. There is not one I use because as you can see it is fairly limited. I am just trying to show you. They are limited, so I have not used it for a long time..

The indirect method was by far the most popular method for medical image retrieval from web-based resources and was applied by 89.65% of participants. In this method, participants used image search engines such as Google. Image search engines always seemed to be the nearest place for participants to search for images especially for general or common medical images. This was because of the easy accessibility of image search engines and the fact that participants could visually browse the results and select images quickly. For instance, one of the participants emphasized the use of web-based resources:

P25: [Google image search] is just quick and fast. It is easy to use Google image search. You just type a keyword like 'blood supply for brain' and you get the images you needed.

There are many search engines and many ways to gain access to the images stored in online resources. At the moment, however, what is evident from our findings is the high reliance of participants on Google image search. It was also the most frequently used tool for identifying the images that participants needed. This high reliance on Google image search, and the role it plays in medical image seeking, merits further investigation. Past studies have also raised questions on the use of Google by health care professionals. The results of the survey of doctors by Tang and Ng (2006) showed the importance of Google. Tang and Ng suggested that in difficult diagnostic cases, doctors used Google to find diagnoses and/or treatment. The authors also reported that search engines such as Google are becoming the latest tools in clinical medicine and doctors in training need to become proficient in their use. When participants were asked whether they used Google image search for finding images, one of them answered as follows:

P9: My first line, and probably usually my only line, is to go to Google images. It is probably sitting there and it is easy to use. I think Google is quite good because you can run down images quickly.

From the interviews, we found several reasons for the popularity and use of Google image search. The main reasons were:

Google image search is very straightforward to use. Participants mentioned that this image search engine has a simple interface which allows them put in keywords and retrieve images. It also controls and corrects spellings of the search terms to rectify your search.

It is easy to check the relevancy of images on the screen. Thus participants preferred to find images in web-based resources rather finding and browsing papers or books to find images.

P28: I go to Google image search and type in the keywords. It is easy, isn't it! You can also check a page full of images quickly to see if any of them is useful or not.

Google image search is useful for finding medical images for educational purposes. Several participants stressed that Google is a helpful tool if one wants to see an image for self-education or wants images for presentations:

P10: I would use Google image search for educational purpose and if I want to show something in general.

P25: I use Google image search for all sorts of illustrations I need to show to medical students. For instance, yesterday I was teaching medical students and I found some images for blood circulation in the brain using Google image search.

P26: I use Google image search to find images I use in my lectures that I give to my students and colleagues.

Health care professionals demonstrated a high reliance on the Google web search engine in their daily life. Participants believed that Google has made everything searchable even online medical image collections; therefore they believed that using Google image search they could find the images they needed.

P2: I prefer to use image search engines instead of using [web-based] image collections. You may obtain images from those collections through Google image search.

Participants, however, were also strongly critical in their use of Google image search in that they were aware of the issues concerning the credibility and quality of medical images. They mentioned that Google image search presents too many hits, most of which are irrelevant images for a search, and health care professionals need to be able to filter through the results to find images that suit their information needs. They also emphasized that using Google image search they were unable to find images of rare medical conditions, specialized images, or up-to-date images. There were five main criticisms of Google image search when used for medical image retrieval which emerged from our interviews:

Lack of specialty. Participants believed that Google image search is a general image search engine. Thus it is not a suitable tool for finding specialized medical images.

P1: It is easy to find general images [common issues] like tooth, tooth filling. But if I need specialized images I will go to find them in papers. To find special images for amalgam restoration I will have to spend a lot of time because it takes time to review the articles to find the images that I need.

Lack of detailed textual information for medical images. Although participants stressed the importance of textual information for medical images, using Google image search they were not sure whether they could access detailed information related to the images.

P26: Discussion and background information are important and I am not sure whether I can find images with detailed information when I use Google image search. If I need detailed discussion on images I would look for them in papers.

Poor quality of images stored in web-based sources. Although the merits of Google make it an ideal search tool, participants experienced a significant problem. Most often, medical images such as MRI¹ and CT scans are high quality images and participants had difficulty finding high quality images using Google.

P15: If I am using Google image search and if I am looking for images on the internet I am not expecting to find images greater than 1280 x 1020 pixels. Because anything greater than that, even if it is JPEG, starts to get quite big.

Low precision rate in image retrieval and the large number of irrelevant hits. Google image search tends to produce a large number of irrelevant hits because image retrieval in this search engine is text-based. Although participants could use advanced search features to improve results, the image retrieval process is based only on word or phrase matches, without any filtering process. Participants were frustrated at not being able to find the right images quickly and easily. Sifting through a large number of irrelevant images to find the images was not only frustrating, but was time-consuming for busy health care professionals.

Credibility of images obtained from web-based sources. Participants always debated the credibility and trustworthiness of images retrieved from web-

¹ Magnetic Resonance Imaging

based sources. Sometimes they declared that they would not use images obtained from websites such as Wikipedia¹.

The fact that Google image search was the first means by which images were found might be because of the high availability of medical images on the web. We see that web-based sources contain a great quantity of images that may be useful for health care professionals. However, locating the required image remains a challenging task, especially when a general search engine cannot distinguish between an image of a flower and a chest X-ray. This makes it particularly vital to develop medical image retrieval tools to assist health care professionals in specifying image information needs and retrieving images from web-based resources.

5.7.2 Articles

Twenty-four participants obtained medical images from published papers in medical journals. The participants searched for relevant papers through scientific databases such as Medline, Elsevier, Ovid and Springer.

P27: I prefer to use PubMed to find images I need. I mean for specific images. I may use Google image search, but Google searches for images on a general theme. In PubMed, first I search for relevant papers. After I have found papers, I read the abstracts to ensure that the papers are relevant. Then I look for the full text of papers and check whether they contain images I need.

A number of reasons were mentioned by health care professionals for obtaining images from academic papers. The main reason was that participants believed that using web-based resources they would be unable to find the specialized medical images they needed.

¹ www.wikipedia.org

P2: Health and biomedical databases like PubMed are the other source of images, especially specialized medical images or images of rare medical conditions. I use keywords to find articles. Then I look for images in relevant articles to find images that I need. I always find images I need in articles.

P3: The reason for searching images in the articles is that sometimes I cannot find images in image collections, or images I find on the Internet, are not relevant [he searched for images of Monteggia fracture in a web-based image collection, and there was not any image for this topic].

The other reason was the participants' preference for locating background information:

P4: I use electronic journals, because, in addition to the image, you have the text of the article which explains and clarifies all aspects of the work presented in the image.

For some, especially those involved in research projects, academic papers were an important source for current images and the latest information in their field. Participants used images from papers for comparing the findings of their studies to those of other research groups.

P22: I use images from the published literature to compare the results of my study with other studies. The other reason for using papers is that you cannot find this type of staining [pathology images] in the Internet, I mean Google. Most of the time I know the papers, so I can take the pictures from them.

Although papers were the second resource for participants to obtain medical images, our interviews revealed that participants did not use Google scholar to look for journal papers. The interviewees were specifically asked if they used Google scholar for finding papers and the majority answered no. When they knew that they were looking for

papers, they used medical databases, and particularly PubMed¹ to find relevant literature. It should be born in mind that the participants chose databases such as PubMed intentionally as a search tool for identifying papers, but finding images was a by-product of paper searching. The following quotation explains this. When one of the interviewees was asked why he used PubMed to find images he replied:

P29: Using the search facilities of PubMed you can limit the search in different ways. As you can see, I can limit the search to a specific language, journal or author. In Google image search, you cannot do it. I found 5,353 articles in PubMed. Now I can limit the language of articles to English or human anatomy. I also select review article in the type of publication section. Now there are nine articles. I will find the full-text of those nine papers and I will see the images used by the authors.

The fact that PubMed was used by participants to find papers might be because of the high credibility of medical databases such as PubMed for health care professionals, and its popularity for finding published literature in the field of medicine.

5.7.3 Personal collections

Eighteen out of twenty-nine participants mentioned that they would first use images from their own image collections if they had one. Participants used their own personal collections for several reasons:

Firstly, for presenting the results of their research projects.

Secondly, as a source for original medical images that could not be found in a web-based resource.

Thirdly, their own images were more relevant than images obtained from other resources.

¹ <http://www.ncbi.nlm.nih.gov/pubmed/>

Fourthly, due to legal issues such as copyright, participants were not sure whether they could use images obtained from other sources such as web-based sources.

Fifthly, when a high level of credibility was required, participants preferred to use images from their own collection.

Sixthly, participants' personal preferences for using images from their own image collection.

P14: My own images are more relevant. Occasionally from journals if it is relevant background work.

P16: When I want to use an image for publications, I cannot use other researchers' images. It has to be our own images. We require them to be from our own studies.

5.7.4 Books

Sometimes participants used books to find medical images, mainly for educational purposes.

P7: I use images from books for self-education, because of the accuracy and validity of the information and images. Images from books are not helpful for my research project. I prefer to use my personal collection for my research or presentations.

However, as one participant stated it took time to find images in the books.

P2: Books can be a source of images but it takes time to find images in books. I mean to find an image in a book, you have to go to the library and then find images in the book. Most of the time, I am not sure whether I can find the image in the book. Sometimes the authors discussed the topic, but they did not put the image in the book.

The majority of participants mentioned that images from books are out-of-date images; therefore, they looked for images in papers.

P23: I do not use images from books, because they go out-of-date rapidly.

We do not rely so much on books.

5.7.5 Other resources

There were also some other resources that health care professionals used. For example, eight participants (P2, P5, P6, P9, P14, P20, P28, and 29) asked their colleagues and friends for images. Seven participants (P1, P2, P9, P10, P14, P20, and P28) used departmental collections. Two (P2 and P5) mentioned that they might use images from CDs and DVDs accompanying medical textbooks.

P5: Sometimes a CD accompanies the book, so if I could find the image in the CD I would use it.

P28: Last time I needed a diagram and I used my colleague's images. It was a diagram about the radiation interaction with cells.

This study found that the image seeking behaviours of health care professionals does not resemble the information seeking behaviours of physicians, surgeons, specialist doctors and nurses as shown in previous studies. Studies by Gorman (1995); Shelstad and Clevenger (1996); Ely et al. (1999); Covell et al. (1985) found that health care professionals depended on textbooks and medical journals for information; our findings indicate that books were not a common medical image resource used by health care professionals.

Summary

The findings of this study showed that the participants mainly considered the visual attributes of medical images when judging the relevancy of retrieved images. This raises an important issue: how can image search tools match users' pictorial information needs and their relevance criteria for the judgment of images.

By examining the relationship between the structure of indexing tools, end-users' queries and their criteria for the judgment of images, the designers will be able to improve image retrieval effectiveness through maintenance and development of the image indexing tools. End users of image retrieval systems such as health care professionals would be able to construct queries with features (relevance criteria) identified in this study. Such design may help end users clarify their visual information needs, and submit their queries more accurately and efficiently.

Moreover, the results of our study revealed that different groups of users used different relevance criteria related to their speciality and their information needs. Our investigation on the potential range of relevance criteria covered by the information needs (topics) specified within the ImageCLEFMed benchmark revealed that it might be possible to include some of our identified criteria in this kind of evaluation setting, thereby testing more realistic criteria which affect searching in a real-life setting. In particular, there may be potential for ImageCLEFMed to address some of these criteria in the future. Firstly, creating topics that cover a wider set of criteria (the existing topic, 'Fetal MRI', could be adapted to '24 weeks male fetal MRI T2 weighted coronal scan'). Secondly, one could adapt the existing ImageCLEFMed image annotation task to become a task that tests automated annotation of criteria such as age, gender and modality from the semi-structured or unstructured text of ImageCLEFMed test collection annotations. In its three years of running, the modality of image, the body region shown in the image, and the orientation of the body, are key parts of the task (Muller et al., 2007). (These classes correspond closely with the criteria that our study identified, confirming the importance of them in the annotation task. It might be possible to extend future annotation tasks to include medical annotation from text).

Earlier studies by Covell et al. (1985); Ely et al. (1999); Gorman (1995); Shelstad and Clevenger (1996) indicated that health care professionals looked for information to respond to the clinical questions raised by patient care; however, our study found that health care professionals looked for medical images for a variety of reasons. Although we classified their medical image information needs in two groups of general and

specific medical images, the health care professionals mentioned education, research, patient care, publication and documentation as the main reasons for searching for medical images. The health care professionals emphasized the importance of medical images as an educational tool in health and biomedical education. The participants stated that using images for educational purposes facilitated the learning process and helped the audiences to understand the concepts effectively. Additionally, images were used widely for research purposes. Health care professionals in this study mainly used images they had produced to present their findings. They also used images produced and published by other researchers to compare their research findings to. Clinical purpose was the second reason (mentioned by thirteen out of twenty-nine persons) why the health care professionals looked for medical images.

Our investigation also showed that health care professionals also looked for medical images in diverse resources. The studies exploring the information seeking behaviours of health care professionals found that medical textbooks and journals were the most frequently used resources for the information sought by health care professionals. However, our findings revealed that web-based resources (i.e. Google image search), journals and personal collections were the most frequently used medical image resources by health care professionals.

The attributes of medical images can be classified into three groups: textual, visual, and other criteria. Textual attributes such as the age and gender of patients provide semantic information about the image content. Visual attributes represent the visual content of images. The examples are modality, size (dimensional), image quality and the orientation. How to integrate these three groups to create better medical image searching and browsing interfaces, is crucial to the success of a medical image retrieval system. To accomplish this goal, it is helpful to learn how people apply, and how they perceive and conceive, these criteria from both theoretical and empirical points of views as Enser et al. (2006) suggested.

CHAPTER 6-DISCUSSION AND CONCLUSIONS

Introduction

In the course of meeting its aims and objectives, this grounded theory study has examined the relevance criteria for medical images and the image seeking behaviour of health care professionals. The overall purpose has been to gain an understanding of where health care professionals seek images, and how they judge the relevance of images with respect to their image information needs. The scope of the research was limited to health care professionals from the Sheffield Teaching Hospitals NHS Foundation Trust. This chapter summarizes the major findings of the study, and discusses those findings with respect to the study's main and other research questions.

6.1. Main research question: What criteria do health care professionals use to make relevance judgments when searching medical images?

The aim of grounded theory research (including this study) is not intended to verify the findings of a study, but to locate the study within the existing literature. Our research fills a gap within the existing literature by exploring relevance criteria for medical images as applied by health care professionals, and explains how health care professionals search for and select medical images they need. Using the Straussian version of grounded theory, fifteen relevance criteria (classified in three groups as illustrated in Table 5-1) were elicited from health care professionals who participated in this study. These are: age and gender; availability; colour; copyright; credibility; image quality; magnification; modality; orientation; originality; recency; size (dimensional); targeted audiences; technical information; and topical relevancy. In particular, topical relevancy was a criterion applied by all of the participants, while criteria such as originality were mentioned by only two participants. It was found that each participant in the current study used on average 7.51 criteria to judge the relevancy of an image. The analysis of the interviews found the following factors affected the relevance judgment of medical images and image seeking behaviour of health care professionals:

The image resources used by health care professionals. The findings of this study revealed that the application of relevance criteria was significantly

affected by the choice of medical image resources used by health care professionals. For example, health care professionals seemed to apply more criteria for images obtained from web-based resources compared to images obtained from journal articles or their personal collections. Health care professionals also used certain criteria when they obtained images from a particular resource. For instance, credibility and ‘image quality’ were considered as important criteria when the images were obtained from web-based resources.

The type of image information that health care professionals needed. It seemed that the health care professionals applied more criteria when they needed specific medical images. In contrast, fewer criteria were used when they looked for general medical images. Other findings of the study showed that the type of medical image information needs also affected the choice of image resource and, consequently, the relevance criteria used by health care professionals. Participants seemed to look for general medical images in web-based resources and books, whereas they looked for specific medical images in journal articles or they used images from their personal collections.

The size of results set. The larger the set of results retrieved, the more criteria were applied. We noticed that when the participants looked for images in web-based resources using Google image search, the size of the results set was an important factor affecting the relevance judgment process. Further studies are required, however, to elucidate the relationship between the size of results set and the number of relevance criteria applied.

The purpose for using the images. We found that health care professionals applied less relevance criteria when they needed images for educational purposes. In contrast they applied more criteria when they needed images for research purposes. As stated earlier, health care professionals looked for general medical images when they wanted images for educational purposes. Specific medical images were required when health care professionals needed

images for research purposes. These findings support the idea that the type of relevance judgment conducted by the health care professionals was situational and depended on their image information needs. We also believe that when health care professionals needed images for research or publication purposes, they applied many relevance criteria to ensure that the selected images fitted their situational pictorial information needs. Nevertheless, additional studies are required to determine whether there is a difference between the relevance judgment of images and motivation (listed in Table 5-7) of health care professionals for searching images.

Work experience of image users. The findings indicate that more experienced health care professionals applied fewer criteria or used a different set of relevance criteria for the judgment of images. For example, experienced health care professionals did not need technical information or detailed textual information.

Stage of image selection process. The findings of this study revealed that health care professionals seemed to consider the topical relevancy of images together with a few other criteria to select candidate images at the beginning of their image search. After they selected candidate images, they evaluated the selected images more precisely. It indicated that image relevance judgment consisted of multiple sessions and that in each session image users applied different sets of criteria.

Findings showed that health care professionals considered topical relevancy as the minimum requirement for beginning the judgment process. These findings highlighted the importance of topical relevancy for the judgment of images; however, other criteria such as quality of images were also applied. The critical criterion was applied when the health care professionals wanted to select the best and most relevant images from among the candidate images.

We conclude this section by reporting that the health care professionals applied relevance criteria in different ways. Although they applied criteria such as topical relevancy or dimensional size of images based on the visual appearance of images or the textual information attached to images, they also applied criteria such as ‘recency’ that were not evident, or were even absent, in the images. Additionally the participants appeared to apply certain criteria to compare images which were topically relevant to their information needs, e.g. the quality of an image is regarded as important if derived from a printed publication. Similarly, participants applied certain criteria regarding the sources used, e.g. credibility was considered as an important attribute when looking for images in web-based resources.

6.2. Other research questions

6.2.1 Research question 1: Are the relevance criteria we identified different from those criteria suggested in the literature?

We believe that it is not easy to compare the medical image relevance criteria identified in the current study to those found in previous literature: Park (1992); Cool et al. (1993); Barry (1994); Wang (1994); Schamber and Bateman (1996); Tang and Solomon (1998); Bateman (1998a); Spink et al. (1998); Maglaughlin and Sonnenwald (2002) investigated the relevance criteria used for the judgment of textual documents (mostly bibliographic information). Moreover, textual documents contain generally-agreed textual features such as title, keywords and abstract. Using these textual features, users can make a decision on the relevancy of textual documents to their information needs; these textual features also facilitate the retrieval process. However, tangible and shared features such as these do not exist for all images. Thus comparison could not be meaningful unless we compare the criteria we found to those found for images; unfortunately, few studies investigated the relevance criteria for images.

The second reason is the setting of the current study. We investigated the relevance criteria applied for real image information needs of our participants in their real-life context; however, studies such as Barry (1994); Park (1993); Maglaughlin and Sonnenwald (2002); Hung et al. (2005) investigated the relevance criteria applied by

users for predefined tasks, or studied the relevance criteria in artificial test situations. The criteria we identified were not obtained in the same manner as those found in the literature, and so cannot be compared.

The third reason is the different names used for criteria found in the literature. We noticed that researchers used different names for the relevance criteria and categories of criteria they identified (see Maglaughlin and Sonnenwald (2002: p.330)). For example, recency and recentness are two different names for the same criterion.

The fourth reason is that health care professionals looked for medical images in different resources including Google image search, journal articles, personal collections, books and departmental collections. However in studies by Choi (2000); Barry and Schamber (1998); Park (1994); Markkula and Sormunen (2000) the users made use of only certain databases to retrieve the documents they required.

In order to see whether the criteria we identified were different from those criteria suggested in the literature, we contrasted¹ our criteria against those found in the literature (see Table 6-1). In order to access a comprehensive list of criteria, we used the lists of criteria collected by Schamber (1994)²; Mizzaro (1997)³; Maglaughlin and Sonnenwald (2002)⁴. We also used the relevance criteria for images identified by Hung et al. (2005); Markkula and Sormunen (2000); Choi and Rasmussen (2002)⁵; and video relevance criteria documented by Yang (2005).

1 Unfortunately not all authors defined the relevance criteria. Therefore the contrast between the criteria we identified and those suggested in the literature was based on the meaning overlap of the criteria.

2 Schamber, (1994) reviewed the literature on relevance published between 1983 and 1994.

3 Mizzaro, (1997) reviewed the relevance studies published between 1959 and 1996.

4 Maglaughlin and Sonnenwald, (2002) compared the relevance criteria identified in their studies with the criteria suggested in ten major relevance studies conducted between 1991 and 1999.

5 As stated in chapter 2, Choi and Rasmussen, (2002) did not study the relevance criteria applied by image users. They asked images users to use a list of ten widely used criteria in the literature (mostly those criteria were applied for relevance judgment of textual information) and judge the relevancy of images to their information needs.

Table 6-1: The appearance of criteria in the literature

Relevance criteria for medical images; ordered according to frequency	The appearance of criteria in the literature
Originality (2)	
Recency (3)	✓
Availability (4)	✓
Copyright (6)	✓
Magnification (7)	
Colour (7)	✓
Technical information (11)	
Targeted Audiences ¹ (12)	✓
Credibility (13)	✓
Orientation (15)	
Modality (18)	
Age and Gender (20)	
Size (dimensional) (21)	✓
Image Quality (27)	✓
Topical Relevancy (29)	✓
*The number in parenthesis shows the frequency of each criterion	Total number of criteria appearing in the literature: 9

As Table 6-1 shows nine out of fifteen criteria we identified are suggested in past literature. Table 6-1 also indicates that six out of fifteen criteria were applied by more than half of the participants (total=29 participants).

Overall we identified six criteria which were not suggested in the literature, of which four criteria were applied by more than half of the participants. The most frequent and most important criterion we documented was topical relevancy. In the research literature on relevance, topicality is mainly defined as the relationship between the topic (query terms) and content of a documents. Our findings show that participants judged the topical relevancy of images based on the description and visual appearance of images related to their image information needs. In response to the first research question of the study, and based on the findings of our experiment, we believe that the set of relevance

¹ Sometimes researchers used terms such as 'difficulty level' to name this criterion.

criteria that health care professionals applied for the judgment of medical images, differed from the sets suggested in the literature.

6.2.2 Research question 2: What are the core relevance criteria used for judging the relevance of medical images?

We noted that health care professionals relied more on visual criteria than on other groups of criteria. This category includes seven out of the fifteen criteria we identified. (Three out of seven criteria in the visual criteria group were common between this group and the textual group of criteria).

Quantitative analysis of the relevance criteria (see Figure 5-17) supported our decision concerning the selection of visual criteria as the main category. Among the criteria used by all the twenty-nine participants, the largest portions of those were visual criteria (7 out of 15 criteria) which related to the visual attributes of medical images. The most frequent criteria were ‘topical relevancy’ and ‘quality of images’, which belonged to the main category. ‘Topical relevancy’ was mentioned by all twenty-nine participants of the study, and was a common criterion between visual and textual groups of criteria. Twenty-seven participants considered the quality of images when they selected medical images, whereas size was another frequently used criterion applied by twenty-one participants. ‘Age and gender’ was another frequently used criterion and belonged to the textual group of criteria. Modality was the fifth most frequent criterion mentioned by eighteen interviewees.

The findings of the study showed that ‘topical relevancy’ and ‘modality’ have been clearly identified as the most important factors for the evaluation and selection of medical images. The ‘topical relevancy’ of medical images was not only the most frequent criterion applied by all participants of this study, but was also selected as the most important criterion by fifteen participants. Figure 5-18 shows that health care professionals mainly considered the visual attributes of images for the relevance judgment of medical images. This is another affirmation for the selection of visual criteria as the main category in our grounded theory study.

While prior research provides considerable evidence for topicality as the most important and most frequently used relevance criterion for the judgment of retrieved documents (including images), it provides little insight how image users select the images they need. As we discussed in chapter 5, an image may be perceived as topically relevant, but may not be selected for a particular situation the user is dealing with at that time. (This study filled this void in the literature by highlighting the importance of the visual appearance of images, and the visual memory of users for the relevance judgment of images). Analysis of the data collected for this investigation confirm that topical relevancy is a major attribute of images retrieved for the purpose of resolving the pictorial information needs of health care professionals. As a single attribute of such images, however, topical relevancy seems to be more important to health care professionals for deselecting items than for choosing them. That is, unless some other attributes of the retrieved images are considered, topical relevancy by itself will not determine if an image is relevant; it is only an indication of potential relevance. Thus the participants of this study used this criterion to select candidate images that appeared to be topically relevant to their image information needs, and visually resembled the participants' visual memories of an object or a condition. By considering topical relevancy, health care professionals not only select or deselect candidate images, but also evaluate the pertinence or comprehension of retrieved images. Their visual memory helped them to make a cognitive connection between the retrieved images and their image information needs. Our research demonstrates that without this cognitive connection, no further evaluation can take place.

The results of our investigation suggest that until the issue of topical relevancy has been determined by the image viewers, it is difficult to predict which criterion among the fifteen criteria will be applied by participants. Indeed, other criteria may be said to be posterior criteria for selection and comparison of candidate images. It is important, though, to expand the discussion so that the application and importance of posterior criteria was dependent upon the image resources used by health care professionals and the illustration task of health care professionals. For example, credibility was regarded

as a critically important posterior criterion when the health care professionals looked for images in web-based resources using Google image search.

6.2.3 Research question 3: Where do health care professionals look for medical images?

This study found that the image seeking behaviours of health care professionals does not resemble the information seeking behaviour of physicians, surgeons, specialists and nurses as shown in previous studies. Studies by Gorman (1995); Shelstad and Clevenger (1996); Ely et al. (1999); Covell et al. (1985) found that health care professionals depended on textbooks and medical journals for information, whereas our findings indicate that health care professionals look for images in web-based resources (by using Google image search), journals and personal collections. Overall, journals were considered as an important source of images for medical image users; in particular, when they looked for specific medical images (although the participants mentioned that they prefer to use images from their personal image collections). Books seemed to be less important resources when the health care professionals looked for images.

Although we detailed the motivations and barriers of accessing the medical image resources used by health care professionals in chapter 5, analysis of the interviews showed that the choice of medical image resources used by health care professionals was affected by the purpose for which the medical images were collected. As stated earlier in chapter 5, we classified the medical image information needs of health care professionals into two groups: general and specific medical images. The health care professionals also mentioned that they needed medical images for the following purposes: education, research, patient care, publication and documentation.

The health care professionals emphasized the importance of medical images as an educational tool in health and biomedical education. Participants stated that using images for educational purposes facilitated the learning process and helped the audiences understand the medical conditions effectively. Additionally, images were used widely for research purposes. Health care professionals mainly used the images they produced to present their findings. They also used images produced and published by

other researchers to compare their research findings to. The clinical purpose was the second reason (mentioned by thirteen out of twenty-nine persons) why the health care professionals looked for medical images. While earlier studies by Covell et al. (1985); Ely et al. (1999); Gorman (1995); Shelstad and Clevenger (1996) indicated that health care professionals looked for information to respond to the clinical questions raised by patient care.

One of the major findings of our study is that only nine of the participants of the current study used online medical image collections. Our investigations indicate that while there is willingness and interest in using medical image collections, participants believed that they could not find the images they needed. In addition, health care professionals believed that the number of images in the medical image collections on the internet was limited. Moreover, health care professionals emphasized that any web-based medical image collection should be easy to use. Although the potential usefulness of medical image collections is apparent, barriers or difficulties in using these collections put health care professionals off. The same thing happened when they could not find images in medical image collections suggested by the researcher, resulting in them wholly neglecting other aspects of medical image collections. There are various reasons why medical image collections might be useful for health care professionals to find medical images, but this will only happen if we can provide medical image collections that offer clear benefits for health care professionals. These image collections must also support the way health care professionals currently work, rather than asking them to adopt new and alien ways of image retrieval.

6.2.4 Research question 4: What difficulties do health care professionals face when searching medical images?

Previous studies identified barriers to information by health care professionals (Ely et al., 1999; Shelstad and Clevenger, 1996; Gorman, 1995). These were: lack of time, lack of access to electronic journals and databases, and lack of information seeking skills. This study, however, found different barriers to seeking medical images. These barriers include: low precision in the image retrieval; lack of domain specific image searching

tools; the large number of images in the results set; lack of credibility; low quality of retrieved images; lack of image search facilities; lack of time; difficulties of image search with medical abbreviations and acronyms; and legal issues (such as copyright of images). These barriers affected the choice of medical image resources used by health care professionals. For example, health care professionals tended to use journal articles or their personal collections when they could not find specific medical images. We also mentioned earlier that health care professionals applied medical image relevance criteria according to the image resources they used. Thus we believe that the barriers to seeking medical images can affect the relevance judgment and the criteria applied by health care professionals. Although further studies are required to determine whether there is a relationship between the barriers to seeking images and the relevance judgment of images by users, our findings also showed improvements in the barriers identified by previous studies. These improvements include: better access to electronic journals and databases, and improved information literacy skills. Lack of time still remained an important barrier to seeking medical images.

A large number of image retrieval systems, such as image search engines, have been developed. End users, including health care professionals, use keyword search tools to find the images they need. Using text to search images, including medical images, could be successful if sufficient text such as annotations is available for searching, or if end users use the same keywords associated with images in their queries as Keister (1994) reported. She analyzed the queries received by the Prints and Photographs Collection of the National Library of Medicine, and reported that sometimes it had been extremely difficult to use verbal queries (text) to retrieve non-textual documents (images). She explained the difficulty as follows:

The difficulty is not only that the communication elements in an image are visual, another level of difference exists. Each of these documents acts as its own unique information vehicle, loaded with data, both obvious and subtle, all presented simultaneously in one neat visual rectangle. (Keister, 1994)

Keister (1994) highlighted a gap between what the end users of an image retrieval system wanted and what the system retrieved. In fact this is a gap between how users present their medical image information needs and how they evaluate the retrieved images for their information needs. The findings of this study supported Keister's ideas concerning the difficulties that image users face when searching for images. The participants of this study stated that currently search by text is the best way to search for medical images. This is also evident from the outcome of the interviews, which revealed that most participants were unable to describe their visual information needs using text. This is a common problem in image retrieval and has been referred to as the 'semantic gap' in the literature (e.g. see Smeulders et al. (2000)). The aim of this study was to contribute to knowledge about one side of the gap, the relevance criteria applied. The next step would be a link between these criteria and the technical possibilities. An experiment by Sedghi et al. (2008) showed that the gap between users' visual information needs and their textual queries might be reduced through the inclusion of relevance criteria a user may specify, in the image retrieval process. Further investigation is required to see how relevance criteria such as 'modality' could be included in the medical image search process. Using content-based image retrieval techniques which retrieve images based on visual features such as colour and shape, is another solution to bridging the semantic gap in medical image retrieval.

6.3. Conclusions

This study is the first of its kind to investigate the relevance criteria applied by health care professionals when searching for medical images. Our research highlighted the importance of relevance studies in a particular domain for better understanding of relevance judgment and image seeking behaviour. To the best of our knowledge, there has not been a comprehensive study of relevance criteria applied by health care professionals. Thus the study helps to contribute to our understanding of relevance criteria applied to medical images. The current study also considers the concept of relevance and the information seeking behaviour of health care professionals.

This thesis is also one of the few studies that have focused on different aspects of image seeking behaviour and the relevance judgment process, including the similarities and dissimilarities between the image seeking and image selection behaviour of image users based on the image resources used, and the information needs mentioned by health care professionals.

Twenty-nine health care professionals from different health and biomedical departments (e.g. Medical Physics, Radiology, Haematology, Genetics and Immunology) participated in this study. In general, fifteen criteria were applied by health care professionals when they looked for medical images and, based on our findings, the criteria we identified were grouped in three main categories: visual criteria, textual criteria and other criteria. Figure 5-17 in chapter 5 show the criteria, the groups of criteria, and how frequently they were mentioned by all interviewees. Obviously, most of the criteria we identified were in the visual category such as topical relevancy, image quality and dimensional size of image, and as Figure 5-18 shows these criteria were the three most frequently mentioned criteria. Overall topical relevancy was the most frequently applied criterion (used by all participants), which is consistent with the findings from image relevance studies by Markkula and Sormunen (2000); Choi and Rasmussen (2002); Hung et al. (2005); video relevance criteria by Yang (2005); and relevance criteria for textual documents identified by Barry (1994); Park (1993); Tang and Solomon (1998); Maglaughlin and Sonnenwald (2002); Chamber and Bateman (1996).

Health care professionals main concern was the usefulness of the images found for their situational pictorial information needs. Topicality, by definition, refers to the relevancy of the query used by users to the theme of the retrieved documents. This implies the utilization of tangible and textual features of the document. Since such features do not exist for medical images, health care professionals judged the topical relevancy of these based on the visual appearance of images and related textual information. Medical images can be interpreted and used in different ways by different users. Though we may not know exactly which attributes of medical images, or which specific combination of

attributes, determine the relevance of an image at a given point in time, we know that health care professionals judged the relevancy of images according to the visual appearance of images and using what they described as 'visual memory'. This is consistent with the findings of Greisdorf and O'Connor (2002: pp.20-21)'s study that image users judge the relevancy of retrieved images using what they described as 'temporal prototypes'. Although we explained how health care professionals compared and selected images they needed, further investigation is required to determine how image users use their visual memory to search for and select images they need.

The information needs and the medical image resources available seemed to be important factors influencing relevance judgment of images in the current study. For example, the health care professionals looked for medical images using Google image search for educational purposes, and mainly paid attention to the visual attributes of images when they evaluated the images retrieved. Health care professionals also did not consider criteria such as credibility when they used images from journal articles. There were also other factors such as the size of results set (when participants used Google image search), the work experience of health care professionals, and the quality of search results which affected the relevance judgment of medical images, and made the medical image relevance judgment a multidimensional and situational process. The participants searched for both general and specific types of medical images, and used medical images for education, research, patient care, publication and documentary purposes. Health care professionals obtained medical images from web-based resources (mostly using Google image search), journal articles, personal collections, books, image collections of colleagues and departmental collections. Although all participants began image relevance judgment by applying the criterion 'topical relevancy', they applied different criteria such as 'quality of images' or 'credibility' when they wanted to make their final decision concerning the relevancy of images to their information needs. The findings highlighted the fact that the medical image selection process would only begin after the topical relevancy of images was ensured by health care professionals.

The results of the current study also contribute to the user-oriented and relevance studies in the information retrieval field. The findings not only deepen our knowledge on the concept of relevance and enrich the relevance criteria identified in previous studies, but also prove that relevance is a multidimensional phenomenon involving complex sets of factors such as personal and situational factors. We tried to identify some dimensions of this phenomenon. We believe that we achieved what we set out to do, which was to document the relevance criteria and relevance judgement process for medical images applied by health care professionals.

6.4. Further research

We have suggested further areas of research in chapters 5 and 6, and have identified additional topics for future research. This study implies the importance of an inductive approach for relevance studies. It was important to study the relevance judgments and relevance criteria as perceived by participants as had been mentioned by some other researchers in the past (Barry, 1993; Schamber et al., 1990; Park, 1993; Saracevic, 2007). All the criteria we identified emerged directly from the data we collected, using interviewing and think-aloud techniques. Therefore, adopting a grounded theory approach for the study of image relevance judgment in other disciplines could be appropriate for future work.

This has been the first study of its kind to investigate the relevance criteria for medical images applied by health care professionals with a grounded theory approach. However, due to the limited number of participants interviewed, we expect that more similar image relevance studies will help to examine the credibility and trustfulness of the criteria we identified. In particular, we suggest that researchers conduct more medical image relevance studies, and investigate the relevance criteria applied by specific groups of health care professionals. We also suggest that image retrieval system developers and designers could use the findings of this experiment to improve image retrieval systems, and thus respond better to the image information needs of users.

One potential area for further research is the way that health care professionals look for images in the image resources they use, their search strategies, the way that they interact

with image search tools such as Google image search, and the differences and driving factors of their image seeking behaviour. The current research showed that health care professionals looked for images in different resources, and that there were variations between health care professionals not only in the relevance criteria they applied, but also in terms of the way they searched for images in the image resources they used. Further research is needed to clarify these variations. Future study should focus on the image resources used by health care professionals, their interaction with image searching tools, and the factors which drive their image information needs..

Another potential area for further research is to evaluate some of our identified criteria in relation to the future design of image retrieval systems and medical image retrieval. For example, dimensional size of images was a criterion mentioned by 72.41% of participants. Inclusion of criteria such as dimensional size of images and modality will improve the design of medical image retrieval systems and thus better satisfy health care professionals' pictorial information needs.

Although we identified topical relevancy as the most important and most frequent relevance criterion for relevance judgment of medical images, we still need further research about the nature of topical relevancy as a relevance criterion for the relevance judgment of images.

6.5. Final words

Though the limitations of this study were discussed in section 3.12, we should bear in mind here that one must be cautious in generalisation of the results of the current study. Like many other qualitative and user studies in the field of library and information sciences, the current study was limited to a small group of participants from different health and biomedical departments and, as the results themselves showed, the domain related factors (such as modality of medical images or orientation) affect relevance judgment of medical images. We also would like to mention that grounded theory has been used by researchers in library and information science, but few of them mentioned whether they reached saturation. Researchers who did mention they had reached

saturation, did not state clearly when saturation was reached. We have reported that in our study, we reached saturation after we had interviewed fourteen participants.

Summary

This chapter has drawn conclusions from our study, and discussed them in the light of past research. It has also presented suggestions for further investigation. The chapter has explained the main characteristics of relevance judgment, the relevance criteria applied by health care professionals, and their image seeking behaviour. It has shown that differences exist within the relevance criteria and medical image resources used by health care professionals. The chapter has also illustrated that health care professionals applied criteria beyond topicality (e.g. modality of image) for the relevance judgment of images. Finally, this chapter has explained that relevance judgment of medical images can be influenced by factors such as the image resources available, and the work experience of participants.

BIBLIOGRAPHY

AHMAD FAUZI, M. F. (2004). Content-based image retrieval of museum images. *School of Electronics and Computer Science*. Southampton, University of Southampton.

ANDERSON, T. D. (2006). Studying human judgments of relevance: interactions in context. *Proceedings of the 1st international conference on Information interaction in context*. Copenhagen, Denmark, ACM.

BARRY, C. L. (1993). A Preliminary Examination of Clues to Relevance Criteria within Document Representations. *Proceedings of the ASIS Annual Meeting*, 30, 81-86.

BARRY, C. L. (1994). User-defined relevance criteria: An exploratory study. *Journal of the American Society for Information Science*, 45, 149-159.

BARRY, C. L. & SCHAMBER, L. (1998). Users' criteria for relevance evaluation: a cross-situational comparison. *Information Processing and Management: an International Journal*, 34, 219-236.

BATEMAN, J. (1998a). Changes in Relevance Criteria: A Longitudinal Study. *ASIS Annual Meeting*. Medford, Information Today, Inc.

BATEMAN, J. A. (1998b). Modeling changes in end-user relevance criteria: an information seeking study. University of North Texas.

BENINCASA, C., CALDEN, A., HANLON, E., KINDZERSKE, M., LAW, K., LAM, E., RHOADES, J., ROY, I., SATZ, M. & VALENTINE, E. (2006). Page Rank Algorithm. Available at <http://www.math.umass.edu/~law/Research/PageRank/Google.pdf>. Accessed date 12 March 2009.

BERG, B. L. (2001). *Qualitative research methods for the social sciences*, Boston, Allyn and Bacon.

BORGMAN, C. L., LEAZER, G. H., GILLILAND-SWETLAND, A., MILLWOOD, K., CHAMPENY, L., FINLEY, J. & SMART, L. J. (2004). How geography professors select materials for classroom lectures: implications for the design of digital libraries. ACM New York, NY, USA.

BORLUND, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 56, 71-90.

BORLUND, P. (2003a). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, 54, 913-925.

BORLUND, P. (2003b). The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research*, 8, 1-38.

BORLUND, P. & INGWERSEN, P. (1998). Measures of relative relevance and ranked half-life: performance indicators for interactive IR. *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. Melbourne, ACM Press.

BRANDT, S. (1999). Use of shape features in content-based image retrieval. *Laboratory of Computer and Information Science*. Helsinki, University of Technology.

BRYMAN, A. (2001). *Social Research Methods*. Oxford, Oxford University Press.

CHAFFEY, D. J. (1996). Design and implementation factors in image retrieval across a wide area network--A commercial example. *International Journal of Information Management*, 16, 381-390.

CHOI, Y. (2000). The characteristics of users' queries and users' relevance criteria in a image retrieval in American history. University of Pittsburgh.

- CHOI, Y. & RASMUSSEN, E. M. (2002). User's relevance criteria in image retrieval in American history. *Information Processing and Management: an International Journal*, 38, 695-726.
- CLEVERDON, C. (1967). The Cranfield tests on index language devices. *Aslib Proceedings*, 19, 173 - 194.
- CLEVERDON, C. W. (1960). ASLIB Cranfield research project: report on the first stage of an investigation into the comparative efficiency of indexing systems.
- CLEVERDON, C. W. & MILLS, J. (1963). The testing of index language devices. *Aslib Proceedings*, 5, 106 - 130.
- COMANICIU, D., MEER, P., FORAN, D. & MEDL, A. (1998). Bimodal system for interactive indexing and retrieval of pathology images. *Applications of Computer Vision, 1998. WACV '98. Proceedings., Fourth IEEE Workshop on*.
- COOL, C., BELKIN, N. J. & KANTOR, P. B. (1993). Characteristics of texts affecting relevance judgments. *Proceedings of the 14th National Online Meeting*. Medford, Learned Information, Inc.
- CORBIN, J. M. & STRAUSS, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13, 3-21.
- COVELL, D. G., UMAN, G. C. & MANNING, P. R. (1985). Information needs in office practice: are they being met? *Ann Intern Med*, 103, 596-9.
- CRYSTAL, A. & GREENBERG, J. (2006). Relevance criteria identified by health information users during Web searches. *Journal of the American Society for Information Science and Technology*, 57, 1368 - 1382.
- CUADRA, C. A. & KATTER, R. V. (1967). Experimental studies of relevance judgments: Final report. Santa Monica, System Development Corporation.

CUNNINGHAM, S. J., BAINBRIDGE, D. & MASOODIAN, M. (2004). How People Describe Their Image Information Needs: A Grounded Theory Analysis of Visual Arts Queries. *Proceedings of the 4th Joint ACM/IEEE Conference on Digital Libraries*. Tuscon.

DESELAERS, T. (2003). Features for image retrieval. *Lehrstuhl für Informatik VI*. Aachen, RWTH Aachen University.

DUQUE, H., MONTAGNAT, J., PIERSON, J. M., BRUNIE, L. & MAGNIN, I. E. (2003). DM2: A Distributed Medical Data Manager for Grids. *Proceedings of BioGrid*, 3, 606–11.

EAKINS, J. P. & GRAHAM, M. E. (1999). Content-based Image Retrieval: A report to the JISC Technology Applications Programme. Institute for Image Data Research, University of Northumbria at Newcastle.

ELY, J. W., OSHEROFF, J. A., EBELL, M. H., BERGUS, G. R., LEVY, B. T., CHAMBLISS, M. L. & EVANS, E. R. (1999). Analysis of questions asked by family doctors regarding patient care. *British Medical Journal*, 319, 358–361.

ENSER, P. G. B., MCGREGOR, C. G. & BRITISH LIBRARY, B. (1993). *Analysis of Visual Information Retrieval Queries*, British Library Board.

ENSER, P. G. B., SANDOM, C. J. & LEWIS, P. H. (2006). Surveying the Reality of Semantic Image Retrieval. *Lecture Notes in Computer Science*, 3736, 177.

EYSENBACH, G., YIHUNE, G., LAMPE, K., CROSS, P. & BRICKLEY, D. (2000). Quality management, certification and rating of health information on the Net with MedCERTAIN: using a medPICS/RDF/XML metadata structure for implementing eHealth ethics and creating trust globally. *Journal of Medical Internet Research*, 2.

FERNÁNDEZ, W. D. (2004). The Grounded Theory Method and Case Study Data in IS Research: Issues and Design. *Information Systems Foundations Workshop: Constructing and Criticising, Canberra, Australia.*

FIDEL, R. (1997). The image retrieval task: implications for the design and evaluation of image databases. *New Review of Hypermedia and Multimedia*, 3, 181-199.

FLICKNER, M., SAWHNEY, H., NIBLACK, W., ASHLEY, J., HUANG, Q., DOM, B., GORKANI, M., HAFNER, J., LEE, D., PETKOVIC, D., STEELE, D. & YANKER, P. (1995). Query by Image and Video Content: The QBIC System. *Computer*, 28, 23-32.

GADD, E., OPPENHEIM, C. & PROBETS, S. (2004). RoMEO studies 6: rights metadata for open archiving. *Program: electronic library and information systems*, 38, 5-14.

GARCIA-MATEOS, G., GARCIA-MERONO, A., VICENTE-CHICOTE, C., RUIZ, A. & LOPEZ-DE-TERUEL, P. E. Time and Date OCR in CCTV Video.

GEUEKE, M. & STAUSBERG, J. (2003). A meta-data-based learning resource server for medicine. *Computer Methods and Programs in Biomedicine*, 72, 197-208.

GILBERT, G. N. (2001). *Researching Social Life*, Sage.

GLASER, B. G. (1992). *Basics of Grounded Theory Analysis*, Mill Valley, CA, Sociology Press.

GLASER, B. G. (2002). Constructivist grounded theory. *Forum: Qualitative Social Research*, 3, 2-10.

GLASER, B. G. (2003). *The Grounded Theory Perspective II: Description's Remodeling of Grounded Theory Methodology*, Mill Valley, Sociology Press.

GLASER, B. G. & HOLTON, J. (2004). Remodeling grounded theory. *Qualitative Social Research*, 5, [?].

GLASER, B. G. & STRAUSS, A. L. (1967). *Discovery of Grounded Theory: strategies for qualitative research*, Chicago, Aldine.

GLATARD, T., MONTAGNAT, J. & MAGNIN, I. E. (2004). Texture based medical image indexing and retrieval: application to cardiac imaging. *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*. New York, NY, USA, ACM Press.

GLAZIER, J. D. & POWELL, R. R. (1992). *Qualitative research in information management*, Englewood, Libraries Unlimited.

GLOVER, E. J., LAWRENCE, S., BIRMINGHAM, W. P. & GILES, C. L. (1999). Architecture of a metasearch engine that supports user information needs. *Ann Arbor*, 1001, 48109.

GORMAN, P. N. (1995). Information needs of physicians. *Journal of the American Society for Information Science*, 46, 729-736.

GOULDING, C. (1998). Grounded theory: the missing methodology on the interpretivist agenda. *Qualitative Market Research: An International Journal*, 1, 50-7.

GREEN, R. (1995). Topical relevance relationships. I: why topic matching fails. *Journal of American Society of Information Science*, 46, 646-653.

GREISDORF, H. (2000). Relevance thresholds: a conjunctive/disjunctive model of end-user cognition as an evaluative process. Denton, TX., University of North Texas.

GREISDORF, H. & O'CONNOR, B. (2002). Modelling what users see when they look at images: a cognitive viewpoint. *Journal of Documentation*, 58, 6-29.

GREISDORF, H. & O'CONNOR, B. (2002). Modelling what users see when they look at images: a cognitive viewpoint. *JOURNAL OF DOCUMENTATION*, 58, 6-29.

GULL, C. D. (1956). Seven years of work on the organization of materials in the special library. *American Documentation*, 7, 320-329.

HEATH, H. & COWLEY, S. (2004). Developing a grounded theory approach: a comparison of Glaser and Strauss. *International Journal of Nursing Studies*, 41, 141-150.

HERSH, W. (1994). Relevance and retrieval evaluation: Perspectives from medicine. *Journal of the American Society for Information Science*, 45, 201-206.

HERSH, W., JENSEN, J., MÜLLER, H., RUCH, P. & GORMAN, P. (2005). A qualitative task analysis of biomedical image use and retrieval. *MUSCLE/ImageCLEF Workshop 05*. Vienna, Austria.

HERSH, W. R. (2003). *Information Retrieval: A Health and Biomedical Perspective*, Springer.

HIRSH, S. G. (1999). Children's relevance criteria and information seeking on electronic resources. *Journal of the American Society for Information Science*, 50, 1265-1283.

HOGGE, B. Creative Commons. Available at http://www.own-it.org/assets/library/documents/creative_commons_factsheet.pdf. Accessed date 24 February 2009.

HOLLINK, L., SCHREIBER, A. T., WIELINGA, B. J. & WORRING, M. (2004). Classification of user image descriptions. *International Journal of Human-Computer Studies*, 61, 601-626.

HUNG, T. Y., ZOELLER, C. & LYON, S. (2005). Relevance judgments for image retrieval in the field of journalism: A pilot study. *Lecture Notes in Computer Science* 3815, 72-80.

INGWERSEN, P. (1992). *Information retrieval interaction*. London, Taylor Graham.

INGWERSEN, P. & JÄRVELIN, K. (2005). *The Turn: integration of information seeking and retrieval in context*, Dordrecht, Springer.

IVKOVIC, G. & SANKAR, R. (2004). An algorithm for image quality assessment.

KAGOLOVSKY, Y. & MOEHR, J. R. (2004). A New Look at Information Retrieval Evaluation: Proposal for Solutions. *Journal of Medical Systems*, 28, 103-116.

KEISTER, L. H. (1994). User types and queries: impact on image access systems. In FIDEL, R., HAHN, B., RASMUSSEN, E. M. & SMITH, P. J. (Eds.) *Challenges in indexing electronic text and images*. Medford, Learned Information.

KIM, J. (2006). *Relevance Judgments and Query Reformulation by Users Interacting with a Speech Retrieval System*. PhD. University of Maryland.

KIM, P., ENG, T. R., DEERING, M. J. & MAXFIELD, A. (1999). Published criteria for evaluating health related web sites: review. *British Medical Journal*.

LANCASTER, F. W. (1979). *Information retrieval systems: characteristics, testing, and evaluation*, New York, Wiley.

LEHMANN, H. (2001). *A Grounded Theory of International Information Systems*. Auckland, University of Auckland.

LEHMANN, T., WEIN, B., DAHMEN, J., BREDNO, J., VOGELSANG, F. & KOHNEN, M. (2000). Content-based image retrieval in medical applications-A novel multi-step approach. *SPIE 2000*.

LEHMANN, T. M., GÜLD, O., KEYSERS, D., SCHUBERT, H., KOHNEN, M. & WEIN, B. B. (2003). Determining the View of Chest Radiographs. *Journal of Digital Imaging*, 16, 280-291.

LEWANDOWSKI, D. (2004). Date-restricted queries in web search engines. *Online Information Review*, 28, 420-427.

LEWIS, C. & RIEMAN, J. (1994). *Task-centered user interface design: a practical introduction*, Colorado, University of Colorado

LEWIS, R. B. (2004). NVivo 2.0 and ATLAS. ti 5.0: A Comparative Review of Two Popular Qualitative Data-Analysis Programs. *Field Methods*, 16, 439-469.

LOY, G. & EKLUNDH, J. O. (2005). A review of Benchmarking content based image retrieval.

MAGLAUGHLIN, K. L. & SONNENWALD, D. H. (2002). User perspectives on relevance criteria: A comparison among relevant, partially relevant, and not-relevant judgments. *Journal of the American Society for Information Science and Technology*, 53, 327-342.

MARKKULA, M. & SORMUNEN, E. (2000). End-User Searching Challenges Indexing Practices in the Digital Newspaper Photo Archive *Journal Information Retrieval* 1, 259-285.

MIZZARO, S. (1997). Relevance: The whole history. *Journal of the American Society for Information Science*, 48, 810-832.

MIZZARO, S. (Ed.) (1998) How many relevances in information retrieval?

MORSE, J. M. & RICHARDS, L. (2002). *Read Me First for a User's Guide to Qualitative Methods*, Sage Publications Inc.

MOSKOP, J. C., MARCO, C. A., LARKIN, G. L., GEIDERMAN, J. M. & DERSE, A. R. (2005). From Hippocrates to HIPAA: Privacy and confidentiality in Emergency Medicine--Part II: Challenges in the emergency department. *Annals of Emergency Medicine*, 45, 60-67.

MÜLLER, H., CLOUGH, P., HERSH, B. & GEISSBUHLER, A. (2006). Variation of Relevance Assessments for Medical Image Retrieval. *In Proceedings of 4th*

International Workshop on Adaptive Multimedia Retrieval. Geneva, Switzerland, University of Geneva.

MÜLLER, H., DESELAERS, T., DESERNO, T., CLOUGH, P., KIM, E. & HERSH, W. (2007). Overview of the ImageCLEFmed 2006 Medical Retrieval and Medical Annotation Tasks. *Lecture Notes in Computer Science*, 4730, 595-608.

MÜLLER, H. & GEISSBUHLER, A. (2005). Lung CT segmentation for image retrieval. *Medical Imaging and Telemedicine (MIT 2005)*. WuyiShan, China.

MÜLLER, H., KALPATHY-CRAMER, J., HERSH, W. & GEISSBUHLER, A. (2008). Using medline queries to generate image retrieval tasks for benchmarking. *Studies in health technology and informatics*, 136, 523-528.

MULLER, H., LOVIS, C. & GEISSBUHLER, A. (2005a). The medGIFT project on medical image retrieval.

MÜLLER, H., MICHOUX, N., BANDON, D. & GEISSBUHLER, A. (2004a). A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *International journal of medical informatics*, 73, 1-23.

MULLER, H., ROSSET, A., GARCIA, A., VALLEE, J.-P. & GEISSBUHLER, A. (2005b). Informatics in Radiology (infoRAD): Benefits of Content-based Visual Data Access in Radiology. *Radiographics*, 25, 849-858.

MÜLLER, H., ROSSET, A., VALLEE, J. P., TERRIER, F. & GEISSBUHLER, A. (2004b). A reference data set for the evaluation of medical image retrieval systems. *Comput Med Imaging Graph*, 28, 295-305.

MYERS, M. D. (1997). Qualitative Research in Information Systems. *MIS Quarterly*, 21, 241-242.

- NEEB, H., ZILLES, K. & SHAH, N. J. (2006). Fully-automated detection of cerebral water content changes: Study of age- and gender-related H₂O patterns with quantitative MRI. *NeuroImage*, 29, 910-922.
- OH, G., LEE, Y. B. & YEOM, S. (2005). Security Mechanism for Medical Image Information on PACS Using Invisible Watermark. Springer.
- PALING, S. & MISZKIEWICZ, M. (2005). Digital resources for dentistry database: A needs assessment. *Proceedings of the American Society for Information Science and Technology*, 42, NA.
- PANOFSKY, E. (1972). Studies in Iconology: Humanistic Themes in the Art of the Renaissance, Westview Press.
- PARK, T. K. (1992). The nature of relevance in information retrieval: An empirical study. Unpublished doctoral dissertation, *School of Library and Information Science*. Bloomington, Indiana University.
- PARK, T. K. (1993). The Nature of Relevance in Information Retrieval: An Empirical Study. *Library Quarterly*, 63, 318-51.
- PARK, T. K. (1994). Toward a theory of user-based relevance: A call for a new paradigm of inquiry. *Journal of the American Society for Information Science*, 45, 135-141.
- PASS, G. & ZABIH, R. (1996). Histogram refinement for content-based image retrieval.
- PICKARD, A. J. (2007). *Research Methods in Information* London Facet Publishing.
- PIETKA, E., GERTYCH, A., POSPIECH, S., CAO, F., HUANG, H. K. & GILSANZ, V. (2001). Computer-assisted bone age assessment: image preprocessing and

epiphyseal/metaphyseal ROI extraction. *IEEE transactions on medical imaging*, 20, 715-729.

PIETKA, E. & HUANG, H. K. (1992). Orientation correction for chest images. *J Digit Imaging*, 5, 185-9.

POWELL, R. R. (1999). Recent trends in research: A methodological essay. *Library & Information Science Research*, 21, 91-119.

PUNCH, K. F. (2005). *Introduction to Social Research: Quantitative and Qualitative Approaches*, London, Sage Publications Inc.

RAGGETT, D. (1997). HTML 3.2 reference specification. Available at <http://www.w3.org/TR/REC-html32.html>. Accessed date 10 March 2009.

REES, A. M. & SCHULTZ, D. G. (1967). A field experimental approach to the study of relevance assessments in relation to document searching. *NSF Report*.

REVERE, D., TURNER, A. M., MADHAVAN, A., RAMBO, N., BUGNI, P. F., KIMBALL, A. & FULLER, S. S. (2007). Understanding the information needs of public health practitioners: A literature review to inform design of an interactive digital knowledge management system. *Journal of Biomedical Informatics*, 40, 410-421.

RIEH, S. Y. & BELKIN, N. J. (1998). Understanding judgment of information quality and cognitive authority in the WWW. *Journal of the American Society for Information Science*, 35, 279-289.

RUI, Y., HUANG, T. S. & CHANG, S. F. (1997). Image retrieval: Past, present and future. IN LIAO, M. (Ed. *Proceedings of the International Symposium on Multimedia Information Processing*. Taipei, Taiwan.

SALTON, G. (1992). The state of retrieval system evaluation. *Information Processing and Management*, 28, 441-449.

SARACEVIC, T. (1996). Relevance reconsidered. In INGWERSEN, P. & PORS, N. O. (Eds.) *the 2nd Conference on Conceptions of Library and Information Science*. Copenhagen, Royal School of Librarianship.

SARACEVIC, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. *J. Am. Soc. Inf. Sci. Technol.*, 58, 2126-2144.

SARACEVIC, T. M. (1975). Relevance: a review of and a framework for the thinking on the notion in information science. *Journal of the American Society of Information Science*, 26, 321-343.

SCHAMBER, L. (1991). Users' Criteria for Evaluation in a Multimedia Environment. *Proceedings of the ASIS Annual Meeting*. p126-33.

SCHAMBER, L. (1994). Relevance and information behavior. *Annual Review of Information Science and Technology*, 29, 3-48.

SCHAMBER, L. & BATEMAN, J. (1996). User criteria in relevance evaluation: Toward development of a measurement scale. *Proceedings of the American Society for Information Science, Baltimore, MD*, 218-225.

SCHAMBER, L., EISENBERG, M. & NILAN, M. S. (1990). A re-examination of relevance: toward a dynamic, situational definition. *Information Processing and Management*, 26, 755-776.

SEDGHI, S., SANDERSON, M. & CLOUGH, P. (2008). A study on the relevance criteria for medical images. *Pattern Recognition Letters*.

SHARIF, B. S., ZAROUG, S. A., CHESTER, E. G., OWEN, J. P. & LEE, E. J. (1994). Bone edge detection in hand radiographic images.

- SHEFFIELD TEACHING HOSPITALS NHS FOUNDATION TRUST (2007). ANNUAL REPORT AND ACCOUNTS 2006 - 2007. Sheffield, Sheffield Teaching Hospitals NHS Foundation Trust.
- SHELSTAD, K. R. & CLEVINGER, F. W. (1996). Information retrieval patterns and needs among practicing general surgeons: a statewide experience. *Bull Med Libr Assoc*, 84, 490-7.
- SINCHAI, T., ARKADIUSZ, G., AIFENG, Z., BRENT, J. L. & HAN, K. H. (2008). Automated bone age assessment of older children using the radius. In KATHERINE, P. A. & KHAN, M. S. (Eds.). SPIE.
- SLATER, M. (1990). *Research methods in library and information studies*, London, Library Association.
- SMEULDERS, A. W. M., WORRING, M., SANTINI, S., GUPTA, A. & JAIN, R. (2000). Content-Based Image Retrieval at the End of the Early Years. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 1349-1380.
- SONNENWALD, D. H., WILDEMUTH, B. M. & HARMON, G. L. (2001). A research method to investigate information seeking using the concept of information horizons: an example from a study of lower socio-economic students' information seeking behaviour. *The New Review of Information Behaviour Research*, 2, 65-86.
- SPINK, A., GREISDORF, H. & BATEMAN, J. (1998). From highly relevant to not relevant: examining different regions of relevance. *Information Processing and Management*, 34, 599-621.
- STEIN, M. S., FEIK, S. A., THOMAS, C. D. L., CLEMENT, J. G. & WARK, J. D. (1999). An Automated Analysis of Intracortical Porosity in Human Femoral Bone Across Age. *Journal of Bone and Mineral Research*, 14, 624-632.

STERN, E. J. & RICHARDSON, M. L. (2003). Preparation of digital images for presentation and publication. *Am Roentgen Ray Soc.*

STRAUSS, A. L. & CORBIN, J. M. (1990). *Basics of qualitative research: grounded theory procedures and techniques*, Newbury Park, Sage Publications.

STRAUSS, A. L. & CORBIN, J. M. (1998). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, Sage Publications Inc.

SWANSON, D. R. (1986). Subjective versus Objective Relevance in Bibliographic Retrieval Systems. *Library Quarterly*, 56, 389-98.

TAGARE, H. D., JAFFE, C. C. & DUNCAN, J. (1997). Medical Image Databases: A Content-based Retrieval Approach. *J Am Med Inform Assoc.*, 4, 184-198.

TANG, H. & NG, J. H. K. (2006). Googling for a diagnosis--use of Google as a diagnostic aid: Internet based study. *BMJ*, 333, 1143-1145.

TANG, R. (1999). Use of relevance criteria across stages of document evaluation: A micro level and macro level analysis (UMI# 9954723). *Ann Arbor, MI: UMI.*

TANG, R. & SOLOMON, P. (1998). Toward an Understanding of the Dynamics of Relevance Judgment: An Analysis of One Person's Search Behavior. *Information Processing and Management*, 34, 237-256.

TANNER, J. M. & GIBBONS, R. D. (1994). Automatic bone age measurement using computerized image analysis. *The Journal of pediatric endocrinology*, 7, 141.

THE DUBLIN CORE METADATA INITIATIVE (2008). Dublin Core Metadata Element Set Version 1.1. Available at <http://dublincore.org/documents/dces/>. Accessed date 14 February 2009.

TOMASI, C. & MANDUCHI, R. (1998). Bilateral filtering for gray and color images.

- TOMBROS, A., RUTHVEN, I. & JOSE, J. M. (2005). How users assess Web pages for information seeking. *Journal of the American Society for Information Science and Technology*, 56, 327-344.
- TOMS, E. G. & TAVES, A. R. (2004). Measuring user perceptions of Web site reputation. *Information Processing & Management*, 40, 291-317.
- TRANBERG, H. A., ROUS, B. A. & RASHBASS, J. (2003). Legal and ethical issues in the use of anonymous images in pathology teaching and research. *Histopathology*, 42, 104-109.
- TSAI, C. F. (2007). A review of image retrieval methods for digital cultural heritage resources. *Online Information Review*, 31, 185-198.
- VISUAL RESOURCES ASSOCIATION (2009). VRA Core 3.0. Available at <http://www.vraweb.org/resources/datastandards/vracore3/index.html>. Accessed date 10 February 2009.
- WANG, B., CHEN, Y., LI, Z. & LI, M. (2006). Compact Representation for Large-Scale Clustering and Similarity Search. *Lecture Notes in Computer Science*, 4261, 835.
- WANG, P. (1994). A cognitive model of document selection of real users of information retrieval systems. research directed by College of Library and Information Services. University of Maryland at College Park.
- WANG, P. & WHITE, M. D. (1999). A cognitive model of document use during a research project. Study II. Decisions at the reading and citing stages. *Journal of the American Society for Information Science*, 50, 98-114.
- WANG, Z., BOVIK, A. C. & LU, L. (2002). Why is image quality assessment so difficult? , IEEE; 1999.

WILSON, T. D. (1999). Models in information behaviour research. *Journal of Documentation*, 55, 249-270.

WILSON, T. D. (2006). On user studies and information needs. *Journal of Documentation*, 62, 658-670.

XI, Z. & XINGGANG, L. (2003). Automatic date imprint extraction from natural images. Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on.

XU, Y. C. & CHEN, Z. (2005). Relevance Judgment—What Do Information Users Consider beyond Topicality? *Journal of the American Society for Information Science and Technology*, 57, 961 - 973.

YANG, M. (2005). An exploration of users' video relevance criteria. PhD, THE UNIVERSITY OF NORTH CAROLINA.

ZHANG, A., GERTYCH, A. & LIU, B. J. (2007). Automatic bone age assessment for young children from newborn to 7-year-old using carpal bones. *Computerized Medical Imaging and Graphics*, 31, 299-310.

ZHOU, X., DEPEURSINGE, A. & MÜLLER, H. (2008). Hierarchical classification using a frequency-based weighting and simple visual features. *Pattern Recognition Letters*, 29, 2011-2017.

Appendix 1: Ethics approval from the NHS


National Research Ethics Service
South Sheffield Research Ethics Committee

Upstairs, 1st Floor, Victoria Corridor
Northern General Hospital
Herries Road
Sheffield
S5 7AU

Telephone 0114 271 4297

22 August 2007

Mr Shahram Sedghi Ikhaniar
Research Student
The University of Sheffield
Department of Information Studies
Regent Court, 211 Portobello St,
Sheffield, S1 4DP

Dear Mr Sedghi Ikhaniar

Full title of study: **Relevance Criteria for Medical Images Used by Health
Care Professionals**
REC reference number: **07/H1309/76**

Thank you for your letter of 03 August 2007, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Chair

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Ethical review of research sites

The Committee has designated this study as exempt from site-specific assessment (SSA). There is no requirement for other Local Research Ethics Committees to be informed or for site-specific assessment to be carried out at each site.

Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

Approved documents

The final list of documents reviewed and approved by the Committee is as follows

Document	Version	Date
Application		19 July 2007
Investigator CV	Shahram Sedghi	
Protocol		
Covering Letter		
Letter from Sponsor	University of Sheffield	26 April 2007

This Research Ethics Committee is an advisory committee to Yorkshire and The Humber Strategic Health Authority
The National Research Ethics Service (NRES) represents the NRES Directorate within
the National Patient Safety Agency and Research Ethics Committees in England

Letter from Sponsor	University of Sheffield	02 July 2007
Peer Review	Nigel Ford	14 June 2007
Compensation Arrangements	University of Sheffield	20 April 2007
Questionnaire: Interview Questions		19 July 2007
Letter of invitation to participant	2.0	02 August 2007
Participant Information Sheet	2.0	02 August 2007
Participant Consent Form		19 July 2007
Response to Request for Further Information		03 August 2007
A20 Application Form		02 August 2007
Supervisor's CV	Paul Clough	
Supervisor's CV	Mark Sanderson	

R&D approval

All researchers and research collaborators who will be participating in the research at NHS sites should apply for R&D approval from the relevant care organisation, if they have not yet done so. R&D approval is required, whether or not the study is exempt from SSA. You should advise researchers and local collaborators accordingly.

Guidance on applying for R&D approval is available from
<http://www.rdforum.nhs.uk/rdform.htm>.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

Feedback on the application process

Now that you have completed the application process you are invited to give your view of the service you received from the National Research Ethics Service. If you wish to make your views known please use the feedback form available on the NRES website at:

<https://www.nresform.org.uk/AppForm/Modules/Feedback/EthicalReview.aspx>

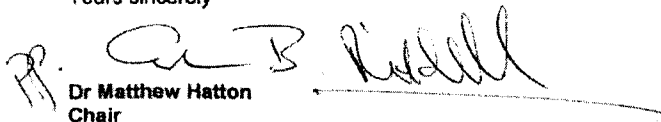
We value your views and comments and will use them to inform the operational process and further improve our service.

07/H1309/76

Please quote this number on all correspondence

With the Committee's best wishes for the success of this project

Yours sincerely


 Dr Matthew Hatton
 Chair

Email: Gordon.Riddell@sth.nhs.uk

Enclosures: *Standard approval conditions*

Copy to: *Mrs Gretta Pearman, University of Sheffield*

RESEARCH IN HUMAN SUBJECTS OTHER THAN CLINICAL TRIALS OF INVESTIGATIONAL MEDICINAL PRODUCTS

Standard conditions of approval by Research Ethics Committees

1. Further communications with the Research Ethics Committee
 - 1.1 Further communications during the research with the Research Ethics Committee that gave the favourable ethical opinion (hereafter referred to in this document as "the Committee") are the personal responsibility of the Chief Investigator.
2. Commencement of the research
 - 2.1 It is assumed that the research will commence within 12 months of the date of the favourable ethical opinion.
 - 2.2 In the case of research requiring site-specific assessment (SSA) the research may not commence at any site until the Committee has notified the Chief Investigator that the favourable ethical opinion is extended to the site.
 - 2.3 The research may not commence at any NHS site until the local Principal Investigator (PI) or research collaborator has obtained research governance approval from the relevant NHS care organisation.
 - 2.4 Should the research not commence within 12 months, the Chief Investigator should give a written explanation for the delay. It is open to the Committee to allow a further period of 12 months within which the research must commence.
 - 2.5 Should the research not commence within 24 months, the favourable opinion will be suspended and the application would need to be re-submitted for ethical review.
3. Duration of ethical approval
 - 3.1 The favourable opinion for the research generally applies for the duration of the research. If it is proposed to extend the duration of the study as specified in the application form, the Committee should be notified.
4. Progress reports
 - 4.1 Research Ethics Committees are required to keep a favourable opinion under review in the light of progress reports and any developments in the study. The Chief Investigator should submit a progress report to the Committee 12 months after the date on which the favourable opinion was given. Annual progress reports should be submitted thereafter.
 - 4.2 Progress reports should be in the format prescribed by NRES and published on the website (see <http://www.nres.npsa.nhs.uk/applicants/review/afterprogress.nfm#submission>)

- 4.3 The Chief Investigator may be requested to attend a meeting of the Committee or Sub-Committee to discuss the progress of the research.
5. Amendments
- 5.1 If it is proposed to make a substantial amendment to the research, the Chief Investigator should submit a notice of amendment to the Committee.
- 5.2 A substantial amendment is any amendment to the terms of the application for ethical review, or to the protocol or other supporting documentation approved by the Committee, that is likely to affect to a significant degree:
- (a) the safety or physical or mental integrity of the trial participants
 - (b) the scientific value of the trial
 - (c) the conduct or management of the trial.
- 5.3 Notices of amendment should be in the format prescribed by NRES and published on the website, and should be personally signed by the Chief Investigator.
- 5.4 A substantial amendment should not be implemented until a favourable ethical opinion has been given by the Committee, unless the changes to the research are urgent safety measures (see section 7). The Committee is required to give an opinion within 35 days of the date of receiving a valid notice of amendment.
- 5.5 Amendments that are not substantial amendments ("minor amendments") may be made at any time and do not need to be notified to the Committee.
6. Changes to sites (*studies requiring site-specific assessment only*)
- 6.1 Where it is proposed to include a new site in the research, there is no requirement to submit a notice of amendment form to the Committee. The Site-Specific Information (SSI) form of the application form together with the local Principal Investigator's CV should be submitted to the relevant REC for site-specific assessment (SSA).
- 6.2 Similarly, where it is proposed to make important changes in the management of a site (in particular, the appointment of a new PI), a notice of amendment form is not required. A revised SSI form for the site (together with the CV for the new PI if applicable) should be submitted to the relevant REC for SSA.
- 6.3 The relevant REC will notify the Committee whether there is any objection to the new site or Principal Investigator. The Committee will notify the Chief Investigator of its opinion within 35 days of receipt of the valid application for SSA.
- 6.4 For studies designated by the Committee as exempt from SSA, there is no requirement to notify the Committee of the inclusion of new sites.

2 of 4

SOPs version 3.0 dated June 2005; v3.1 dated 1 April 07 minor editing to NRES name change
SL-AC2 Approval conditions (research other than CTIMP)

7. Urgent safety measures

- 7.1 The sponsor or the Chief Investigator, or the local Principal Investigator at a trial site, may take appropriate urgent safety measures in order to protect research participants against any immediate hazard to their health or safety.
- 7.2 The Committee must be notified within three days that such measures have been taken, the reasons why and the plan for further action.

8. Serious Adverse Events

- 8.1 A Serious Adverse Event (SAE) is an untoward occurrence that:

- (a) results in death
- (b) is life-threatening
- (c) requires hospitalisation or prolongation of existing hospitalisation
- (d) results in persistent or significant disability or incapacity
- (e) consists of a congenital anomaly or birth defect
- (f) is otherwise considered medically significant by the investigator.

- 8.2 A SAE occurring to a research participant should be reported to the Committee where in the opinion of the Chief Investigator the event was related to administration of any of the research procedures, and was an unexpected occurrence.

- 8.3 Reports of SAEs should be provided to the Committee within 15 days of the Chief Investigator becoming aware of the event, in the format prescribed by NRES and published on the website.

- 8.4 The Chief Investigator may be requested to attend a meeting of the Committee or Sub-Committee to discuss any concerns about the health or safety of research subjects.

- 8.5 Reports should not be sent to other RECs in the case of multi-site studies.

9. Conclusion or early termination of the research

- 9.1 The Chief Investigator should notify the Committee in writing that the research has ended within 90 days of its conclusion. The conclusion of the research is defined as the final date or event specified in the protocol, not the completion of data analysis or publication of the results.

- 9.2 If the research is terminated early, the Chief Investigator should notify the Committee within 15 days of the date of termination. An explanation of the reasons for early termination should be given.

- 9.3 Reports of conclusion or early termination should be submitted in the form prescribed by NRES and published on the website.

10. Final report

- 10.1 A summary of the final report on the research should be provided to the Committee within 12 months of the conclusion of the study. This should include information on whether the study achieved its objectives, the main findings, and arrangements for publication or dissemination of the research including any feedback to participants.

11. Review of ethical opinion

- 11.1 The Committee may review its opinion at any time in the light of any relevant information it receives.
- 11.2 The Chief Investigator may at any time request that the Committee reviews its opinion, or seek advice from the Committee on any ethical issue relating to the research.

12. Breach of approval conditions

- 12.1 Failure to comply with these conditions may lead to suspension or termination of the favourable ethical opinion by the Committee.

Appendix 2: The invitation letter for participation in the study Services



The University of Sheffield
 Department of Information Studies
 Regent Court, 211 Portobello Street, Sheffield, S1 4DP, UK
 Tel: (0)114-22 22530 Fax: (0)114-278 0300
 Relevance Criteria for Medical Images Used by Health Care Professionals
 Researchers: Dr. Mark Sanderson, Dr. Paul Clough, and Shahram Sedghi

Invitation Letter

Dear Madam or Sir

This letter is an invitation to consider participating in an investigation I am conducting as part of my doctoral dissertation in the Department of Information Studies at the University of Sheffield under the supervision of Dr Mark Sanderson and Dr Paul Clough.

The aim of this study is to find out how health care professionals search for and select medical images that they need. Thus, I am looking for health care professionals who:

- 1- use any type of medical images.
- 2- are skilled and knowledgeable internet and computer users.
- 3- hold a degree in health and bio-medical sciences.
- 4- have access to the internet.

The following health professionals are excluded from the study:

- 1- health care professionals who do not have an academic degree.

Your participation is of course voluntary and it would involve an interview of approximately thirty minutes to take place in a mutually agreed location. Your work will not be affected in any way by your participation or lack of participation in this study.

This study has received approval through NHS South Sheffield Research Ethics Committee. Confidentiality of the survey participants will be maintained; all data obtained in the study will be rendered anonymous at the earliest possible time. All data collected during this study will be in a locked and secure place. 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.

Before you decide whether or not you would like to take part, please read the information sheet that I enclosed with this letter. If you have any questions regarding this study, or would like additional information to assist you in reaching a decision about participation, please contact my supervisors or me at 0790 844 1425 or by email at s.sedghi@shef.ac.uk.

I hope that the results of my study will be of benefit to all health care professionals and medical image retrieval system designers. When the project is completed, I will send you a summary of the findings. I very much look forward to speaking with you and thank you in advance for your assistance in this project.

Yours Sincerely
 Shahram Sedghi
 Research Student

Reference No: 07/11308/76 (South Sheffield Research Ethics Committee)
 Version 2.0, Date: 28/8/2007
 Name of Applicant: Shahram Sedghi Ilkhanlar

Appendix 3: The invitation email for participation in the study

Dear Madam or Sir,

As part of my PhD, I am conducting research looking at how health care and biomedical professionals search and select images they need. I would like to interview health care and biomedical professionals for about forty minutes in a mutually agreed location.

Delicious chocolate rewards are provided after completing the interview!

If you are interested, please contact me, Shahram Sedghi, at 0790 844 1425 or by email at s.sedghi@shef.ac.uk. You can also contact the study supervisors Dr Mark Sanderson on m.sanderson@shef.ac.uk and Dr Paul Clough on p.d.clough@shef.ac.uk. We will be more than happy to answer your questions.

All data collected will be anonymised and kept confidential as the study was approved by both NHS Research Ethics Committee (NHS REC) and the University of Sheffield Research Ethics Committee ('U- REC').

The results of my study will be of benefit to all medical image retrieval system designers and users in general and health & biomedical community in particular.

Warmest Wishes,

Shahram Sedghi, MSc, PhD(C)

Department of Information Studies,

Information Retrieval Research Group

The University of Sheffield, UK

Fax: +44 (0)114 2780300

Appendix 4: The information sheet of the project



The University of Sheffield
 Department of Information Studies
 Regent Court, 211 Portobello Street, Sheffield, S1 4DP, UK
 Tel: +44 (0)114-22 22630 Fax: +44 (0)114-278 0300
 Relevance Criteria for Medical Images Used by Health Care Professionals
 Researchers:
 Dr. Mark Sanderson, Dr. Paul Clough, and Shahram Sedghi

Information Sheet

1. Project Title

Relevance Criteria for Medical Images Used by Health Care Professionals

2. Invitation paragraph

Thank you for your cooperation. We are Sheffield University researchers based at the department of Information Studies and are asking your help to find out what criteria of relevance Health care professionals use when searching for medical images and how they select medical images that they need. Shahram is doing this research as part of his PhD degree. Before you decide it is important for you to understand, why the research is being done and what it will involve. Talk to others about the study if you wish. Please read this information sheet carefully and if there is anything that is not clear or you need more information, please do not hesitate to ask us. Even if we cannot speak to you immediately, one of our team will get back to you. Please, take time to decide whether you wish to take part or not. If you decided not to take part in the study, you do not have to return the reply slip. Thank you for reading this.

3. What is the purpose of the project?

The aim of this research is to elicit and document the relevance and selection criteria that users apply when seeking for medical images and provide recommendations and strategies that could assist medical image retrieval research groups to improve medical image retrieval systems. This study is a part of a PhD research project.

4. Why have I been chosen?

You have been chosen because we are looking for health care professionals who:

- 1- use any type of medical images.
- 2- are skilled and knowledgeable internet and computer users.
- 3- hold a degree in health and bio-medical sciences.

Reference No: 07/h1309/76 (South Sheffield Research Ethics committee)
 Version 2.0
 Date: 2/08/2007
 Name of Applicant: Shahram Sedghi Ilkhanlar

1



The University of Sheffield
Department of Information Studies
 Regent Court, 211 Portobello Street, Sheffield, S1 4DP, UK
 Tel: +44 (0)114-22 22630 Fax: +44 (0)114-278 0300
Relevance Criteria for Medical Images Used by Health Care Professionals
 Researchers:
 Dr. Mark Sanderson, Dr. Paul Clough, and Shahram Sedghi

The following Health care professionals are excluded from the study:

1-health care professionals who do not have an academic degree.

5. Do I have to take part?

No. Your participation in this study is of course voluntary. You may decline to participate. If you decide to participate, you will be given this information sheet to keep, and you will be asked to sign a written consent form. You can still withdraw at any time and your decision about not participating will not affect your job. You do not have to give a reason if you withdraw from the study prior to its completion. If you withdraw, your data will be returned to you or destroyed.

6. What will happen to me if I take part?

You will be asked to participate in an interview and to perform a real medical image search based on your information needs. Before starting, participants will be asked to complete a consent form. The data collection (both interview and medical image search sessions) will last about 30-40 minutes (although the interview can be longer if you wish). Data collection can take place in your department, in a convenient room and at a mutually convenient time for you.

During the interview, participants will be asked to describe one of their recent medical image needs. Investigator also will ask participants to describe their specific medical image needs, the sources that they obtained images, image searching queries, medical image selection process, and their final medical images uses.

Each participant then will be asked to perform a real medical image search based on information needs that is developed independently by the participant and s/he discussed it in the interview session. The participants will first be asked to go to a website they usually use to search medical images. If participants do not know online medical image collections available, the researcher will suggest some online medical image collection to the participants, based on the types of medical images they are looking for. Researcher will ask

Reference No: 07/h1309/76 (South Sheffield Research Ethics committee)
 Version 2.0
 Date: 2/08/2007
 Name of Applicant: Shahram Sedghi Ilkhanlar

2



The University of Sheffield
Department of Information Studies
Regent Court, 211 Portobello Street, Sheffield, S1 4DP, UK
Tel: +44 (0)114-22 22630 Fax: +44 (0)114-278 0300
Relevance Criteria for Medical Images Used by Health Care Professionals
Researchers:
Dr. Mark Sanderson, Dr. Paul Clough, and Shahram Sedghi

some questions about the criteria that participants apply when they choose a certain medical image as a relevant image to their information needs.

While the process of note taking and writing the memos is paramount, the researcher will record interview on audiotape with your permission. The data collection session does not include any sensitive issues, and you are free not to respond any question that you do not like without having to give a reason. All recorded data including audiotapes will be hold in a locked drawer in the department of Information studies at the University of Sheffield. After Shahram completes his PhD, all data will be destroyed.

7. What do I have to do?

If you are interested in participating, please complete the enclosed reply slip, indicate your interest and provide your details, and contact information so that the researcher from the University of Sheffield can contact with you. You will be asked to participate in an interview soon after completing and returning the reply slip. You will be asked to describe one of their last experiences about medical image search during the interview. Researcher also will ask you to perform a medical image search.

8. What are the possible disadvantages and risks of taking part?

There are no known disadvantages or risks associated with this experiment, and we do not anticipate any disadvantages or specific risks will be associated with this study.

9. What are the possible benefits of taking part?

Researchers in information retrieval and computer science departments can benefit from this research by gaining a better understanding of how Health care professionals judge the relevance of medical images and which criteria of relevance are used by them. The results of this study can help researchers in information retrieval and computer science departments to design and improve medical image retrieval systems, therefore people in medical departments also benefit from this experience.

10. What happens if the research study stops earlier than expected?

Reference No: 07/h1309/76 (South Sheffield Research Ethics committee)
Version 2.0
Date: 2/08/2007
Name of Applicant: Shahram Sedghi Ilkhanlar

3



The University of Sheffield
Department of Information Studies
 Regent Court, 211 Portobello Street, Sheffield, S1 4DP, UK
 Tel: +44 (0)114-22 22630 Fax: +44 (0)114-278 0300
Relevance Criteria for Medical Images Used by Health Care Professionals
 Researchers:
 Dr. Mark Sanderson, Dr. Paul Clough, and Shahram Sedghi

If this is the case, the reason(s) should be explained to the participant.

11. What will happen if something goes wrong?

Since there are no known disadvantages or risks associated with this study, it is unlikely that you will be harmed. If you wish to raise a complaint about the way you have been dealt or any possible harm you might suffer in the first instance you can contact Dr Mark Sanderson or Dr Paul Clough in the department of Information Studies at the University of Sheffield via email (m.sanderson@shef.ac.uk , p.d.clough@shef.ac.uk), phone (0114 2222 648, 0114 2222 664) or send a letter to:

Dr. Mark Sanderson/ Dr. Paul Clough,
 Department of Information Studies,
 The University of Sheffield,
 Regent Court, 211 Portobello St.
 S1 4DP, UK, Sheffield

If you are not satisfied with a response, please contact Dr David Fletcher, the Registrar at the University of Sheffield:

Dr David Fletcher
 Firth Court,
 Western Bank,
 Sheffield,
 S10 2TN.
 Tel: 0114 22 21100
 Email: D.E.Fletcher@shef.ac.uk

12. Will my taking part in this project be kept confidential?

Apart from when we first write to you and if we contact you to confirm the time to visit, we will not be using your name, address or telephone number. The data obtained in the study will be kept confidential. All information provided by you or recorded by the researcher will be kept under code number and will not be linked to your name. Data will be available only to

Reference No: 07/h1309/78 (South Sheffield Research Ethics committee)
 Version 2.0
 Date: 2/08/2007
 Name of Applicant: Shahram Sedghi Ilkhanlar

4



The University of Sheffield
Department of Information Studies
 Regent Court, 211 Portobello Street, Sheffield, S1 4DP, UK
 Tel: +44 (0)114-22 22630 Fax: +44 (0)114-278 0300
Relevance Criteria for Medical Images Used by Health Care Professionals
 Researchers:
 Dr. Mark Sanderson, Dr. Paul Clough, and Shahram Sedghi

the persons conducting the study unless you specifically give permission in writing to allow us to do otherwise. All information will be in a locked drawer at the Department of Information Studies in the University of Sheffield. All information will be destroyed immediately after Shahram completed his PhD.

13. What will happen to the results of the research project?

You will receive a summary of the results. The results will also form a part of a PhD thesis, which will be available from the University of Sheffield library or the department of Information Studies. Additionally, researchers may present the results of the study in research conferences or publish in peer reviewed journals.

14. Who is organising and funding the research?

Shahram is conducting the study as part of his PhD research. This study is sponsored by the Ministry of Health and Medical Education of Iran.

15. Who has reviewed the project?

This study has been reviewed by research committee of the department of Information Studies and South Sheffield NHS Research Ethics Committee.

16. Contact for further information

If you would like further information about the study, please contact Dr Mark Sanderson, Dr Paul Clough, or Shahram Sedghi by using following contact information:

Title Dr First Name/Initials: Mark Last Name: Sanderson
 Post Reader Department: Information Studies
 Address
 Department of Information Studies
 The University of Sheffield,
 Regent Court, 211 Portobello St,
 S1 4DP, UK
 Email: m.sanderson@shef.ac.uk Telephone: +44 114 2222 648

Title Dr First Name/Initials: Paul Last Name: Clough
 Post Lecturer Department: Information Studies
 Address
 Department of Information Studies
 The University of Sheffield,
 Regent Court, 211 Portobello St.

Reference No: 07/h1309/76 (South Sheffield Research Ethics committee)
 Version 2.0
 Date: 2/08/2007
 Name of Applicant: Shahram Sedghi Ilkhanlar

5



The University of Sheffield
Department of Information Studies
Regent Court, 211 Portobello Street, Sheffield, S1 4DP, UK
Tel: +44 (0)114-22 22630 Fax: +44 (0)114-278 0300
Relevance Criteria for Medical Images Used by Health Care Professionals
Researchers:
Dr. Mark Sanderson, Dr. Paul Clough, and Shahram Sedghi

S1 4DP, UK

Email: p.d.clough@shef.ac.uk Telephone: +44 114 2222 664

Title: Mr First Name/Initials: Shahram Last Name: Sedghi Ilkhanlar

Post: PhD Student Department: Information Studies

Address:

Department of Information Studies

The University of Sheffield,

Regent Court, 211 Portobello St,

S1 4DP, UK

Email: s.sedghi@shef.ac.uk Telephone: +44 (0)790 844 1425, +44 114 2222 666

Researcher and his research supervisors are thankful for considering taking part and taking time to read this sheet.

Reference No: 07/h1309/76 (South Sheffield Research Ethics committee)
Version 2.0

Date: 2/08/2007

Name of Applicant: Shahram Sedghi Ilkhanlar

6

Appendix 5: Reply slip



The University of Sheffield
Department of Information Studies
 Regent Court, 211 Portobello Street, Sheffield, S1 4DP, UK
 Tel: +44 (0)114-222 22630 Fax: +44 (0)114-278 0300
Relevance Criteria for Medical Images Used by Health Care Professionals
 Researchers
 Dr. Mark Sanderson, Dr. Paul Clough, and Shahram Sedghi

Reply Slip

Dear Madam or Sir

I am a PhD student in the Department of Information Studies at the University of Sheffield. I am conducting a research project on "Relevance Criteria for Medical Images Used by Health Care Professionals" and I would like to invite you to participate in this study.

I have enclosed an Information Sheet about the project with this pack; please take time to read it. If you agree to participate in this research, please sign below. Your help is very much appreciated.

Signature: Date:

Name:

Email address:

Address:

.....

.....

.....

.....

Post Code:

Telephone: Fax:

* There is also a postage paid reply envelope provided in which to return the reply slip. If you are potentially interested in participating, please send back the reply slip or fax it to 0114 278 0300 or contact me at 0790 844 1425 or send an e-mail to s.sedghi@shef.ac.uk. You may also contact my research supervisors directly, Dr Mark Sanderson at (0)114 2222 648 or by email m.sanderson@shef.ac.uk or Dr Paul Clough at (0)114 2222 664 or by email p.d.clough@shef.ac.uk. As soon as you contact us, I will contact you to arrange for the interview.

Yours Sincerely
 Shahram Sedghi
 Research Student

Reference No: 07/h1309/76 (South Sheffield Research Ethics Committee)
 Date: 19/07/2007
 Name of Applicant: Shahram Sedghi Ikhanter

Appendix 6: Consent form



The University of Sheffield
 Department of Information Studies
 Regent Court, 211 Portobello Street, Sheffield, S1 4DP, UK
 Tel: (0)114 22 22630 Fax: (0)114 278 0300
 Relevance Criteria for Medical Images Used by Health Care Professionals
 Researchers: Dr. Mark Sanderson, Dr. Paul Clough, and Shahram Sedghi

Consent Form

Title of Project: Relevance Criteria for Medical Images Used by Health Care Professionals

Name of Researcher: Shahram Sedghi Ilkhanlar

Participant Identification Number for this project: _____

Please initial box

1. I confirm that I have read and understand the information sheet for the above project and have had the opportunity to ask questions
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason.
3. I understand that the interview will be audio taped and my responses will be anonymised before analysis. I give permission for members of the research team to have access to my anonymised responses.
4. I agree to take part in the above research project.

_____	_____	_____
Name of Participant	Date	Signature
_____	_____	_____
Lead Researcher	Date	Signature

Reference No: 07/h1309/76 (South Sheffield Research Ethics Committee)
 Date: 19/07/2007
 Name of Applicant: Shahram Sedghi Ilkhanlar

Appendix 7: Interview protocol



The University of Sheffield
 Department of Information Studies
 Regent Court, 211 Portobello Street, Sheffield, S1 4DP, UK
 Tel: +44 (0)114-22 22630 Fax: +44 (0)114-278 0300
 Relevance Criteria for Medical Images Used by Health Care Professionals
 Researchers:
 Dr. Mark Sanderson, Dr. Paul Clough, and Shahram Sedghi

Interview questions

1. Demographic information (profession, status, frequencies of image search)

- Profession
- Work experience
- Frequencies of medical image search

2. Information needs analysis:

- Please, think about one recent job-related situation in which you needed to find some medical images.
- Could you please describe in detail why you needed medical images? What kind of medical image did you want?

3. Information source analysis:

- Where did you search for the image information? Why?(online , offline or physical)
- Why do you use these resources to find medical images that you need?
- Do you use any online medical image collection to find medical images that you need?
 - If yes, please name the image collections that you use.
 - If not, why you do not use medical image collection?
- Do you think that there should be image retrieval systems specially designed for medical images? If so, what are your requirements for an ideal medical image collection? What kind of information and preview facilities /tools (such as Browse through collection, thumbnails, hierarchical Classification) do you expect to have? Would you like to have? Why?

4. Image search questions:

- How did you begin your search? (e.g., the tools that you used, the actions you took)

Reference No: 07/h1309/76 (South Sheffield Research Ethics committee) 1
 Date: 19/07/2007
 Name of Applicant: Shahram Sedghi Iikhanlar



The University of Sheffield
 Department of Information Studies
 Regent Court, 211 Portobello Street, Sheffield, S1 4DP, UK
 Tel: +44 (0)114-22 22630 Fax: +44 (0)114-278 0300
Relevance Criteria for Medical Images Used by Health Care Professionals
 Researchers:
 Dr. Mark Sanderson, Dr. Paul Clough, and Shahram Sedghi

- Can you think of other search methods? (e.g., search by example image, sample sketch, colour, navigational) Will they work or not? Why?

5. Browsing surrogates/relevance judgment questions:

- When you are looking for medical images, what kind of information (Diagnosis, technical information, size of images, visual attributes, thumbnails, description and discussion, clinical summary, demographic information such as age, sex) did you look for first or seemed most important to you? Why?
- If both textual (all annotations including Diagnosis, treatment, description and discussion, clinical summary, demographic information, technical attributes of images, size of images, date, place, etc) and pictorial (visual) information was available when viewing an individual image, which information was more important to you; Textual or Visual? Why?
- Was it difficult for you find what you wanted? Why?
- Do you think that the system that you used provided enough information for you to search medical images? why?
- If not, what other kinds of information did you want to have? Why?
- What relevance criteria, you applied in some specific cases? Can you please describe them in detail?
- In those cases, what kind of image attributes and surrounded information (e.g., demographics, annotations) seemed most important to you? Why?

6. After seeing and selecting images

- After you saw and chose the images, were you satisfied with the results? why?
- If not, what disappointed you?
- What information did you expect to have in the surrogates / metadata?
- In general, how do you describe a relevant image to your information needs?

7. Image use

Reference No: 07/h1309/76 (South Sheffield Research Ethics committee)
 Date: 19/07/2007
 Name of Applicant: Shahram Sedghi Ilkhanlar

2



The University of Sheffield
Department of Information Studies
Regent Court, 211 Portobello Street, Sheffield, S1 4DP. UK
Tel: +44 (0)114-22 22630 Fax: +44 (0)114-278 0300
Relevance Criteria for Medical Images Used by Health Care Professionals
Researchers:
Dr. Mark Sanderson, Dr. Paul Clough, and Shahram Sedghi

- How did you use the image?(edit image, change the size, add captions, etc)

8. In general, when you search medical images, what criteria /metadata/ surrogate do you think are most important? (Please name the top five criteria)

9. Is there anything else that you would like to say?

Thank you very much!

Reference No: 07/h1309/76 (South Sheffield Research Ethics committee)
Date: 19/07/2007
Name of Applicant: Shahram Sedghi Ilkhanlar

3

Appendix 8: Publications

SEDGHI, S., SANDERSON, M. & CLOUGH, P. (2008) A study on the relevance criteria for medical images. *Pattern Recognition Letters*, 29, 2046-2057.

Appendix 9: A summary of image relevance criteria	
Age and Gender	Whether the age and gender of patient was explained or not.
Availability	Whether the image was available or not.
Colour	Whether the image was a colour image, monochrome or blank and white image.
Copyright	The copyright of an image. E.g., whether the image could be presented in public or the content of the image could be used for publication purposes.
Credibility	Whether the image was obtained from a reliable source or not. The participants wanted to choose images from resources or people whom they trusted.
Image quality	The photographic quality of image content; whether the image content was clear to see.
Magnification	What was the magnification scale of an image, and if it was a magnified image the magnification was shown in the image or text.
Modality	The type of an image such as MRI, X-Ray, PET, CT scan, Pathology image and Electron microscope images.
Orientation	How the objects were presented in the image and from what angles or direction images were taken.
Originality	Sometimes the participants wanted to find medical images in original format without any manipulation. Original medical images are large files and stored in electronic medical image archiving systems such as DICOM.
Recency	The date of production of an image. Participants wanted to find more recent images to use.
Size	The dimensional size of an image the participant was interested in.
Targeted audiences	Whether the image was targeted at certain audiences the participant wanted to focus on
Technical Information	Whether the methods and materials used to produce an image was explained or not.
Topicality	The aboutness of an image. Whether the image visually illustrates what the participants were interested in. The participants used visual appearance of images and textual information to assess the topicality of images.