Establishing User Requirements for a Recommender System in an Online Union Catalogue: an Investigation of WorldCat.org

Simon Wakeling

Submitted for the degree of PhD in Information Science

Information School, University of Sheffield

Submitted April 2015

ACKNOWLEDGEMENTS

ABSTRACT

This project, undertaken in collaboration with OCLC, aimed to investigate the potential role of recommendations within WorldCat, the publicly accessible union catalogue of libraries participating in the OCLC global cooperative. The goal of the project was a set of conceptual design guidelines for a WorldCat.org recommender system, based on a comprehensive understanding of the systems users and their needs.

Taking a mixed-methods approach, the investigation consisted of four phases. Phase one consisted of twenty-one focus groups with key user groups held in three locations; the UK, the US, and Australia and New Zealand. Phase 2 consisted of a pop-up survey implemented on WorldCat.org, and gathered 2,918 responses. Phase three represented an analysis of two months of WorldCat.org transaction log data, consisting of over 15,000,000 sessions. Phase four was a lab based user study investigating and comparing the use of WorldCat.org with Amazon.

Findings from each strand were integrated, and the key themes to emerge from the research are discussed. Different methods of classifying the WorldCat.org user population are presented, along with a taxonomy of work- and search-tasks. Key perspectives on the utility of a recommender system are considered, along with a reflection on how the information search behaviour exhibited by users interacting with recommendations while undertaking typical catalogue tasks can be interpreted.

Based on the enriched perspective of the system, and the role of recommendation in the catalogue, a series of conceptual design specifications are presented for the development of a WorldCat.org recommender system.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Research context

The emergence of the internet as an unparalleled information source has led to a number of significant and well documented problems that constitute significant areas of research in fields as diverse as economics, sociology, computer and information science, and psychology. One relatively recent development, and one exercising academics in all of these fields, has been the recognition of so called 'information overload' – the difficulty users face navigating and processing the information available online. It has been noted that humans' capacity to find information advances more slowly than the pace at which new information is made available (Cosley et al., 2003), and the current exponential growth of the world wide web, which offers an increasingly vast and heterogeneous repository of information, presents significant challenges to users and service-providers in many types of online environment.

Recommender systems (RS) have emerged as an important means of addressing these challenges, and now constitute a significant area of research in the field of information science (Ricci et al., 2011). Put most simply, personalised recommender systems "analyse user profiles, content items, and the connections between them, and try to predict future user behaviour" (Prekopcsák, 2007: 8). This process results in a presentation of suggested content to the user. Such systems differ from more traditional forms of information retrieval (for example search engines) in the sophistication with which an item's potential utility is calculated, and the extent to which that they explicitly attempt to add value though a personalised approach (Burke 2002). Item-level recommendations, such as Amazon's "Customers who bought this also bought…", stand at one of extreme of this spectrum, but can nonetheless be considered recommendations (Schafer et al., 1999). The earliest development of recommender systems occurred in non-commercial environments, and the roots of the theories underpinning their implementation can be found in areas such as cognitive science, approximation theory, and information retrieval (Adomavicius & Tuzhilin, 2005). However it was the emergence in the late 1990s of e-commerce as a growth area that drove many of the most significant developments in recommender systems. A key benefit of e-commerce to businesses, particularly in retail sectors, came with sheer volume of potential products that could be presented to consumers (Schafer et

1

al., 2001). The commercial application of recommender systems offered a means of presenting specifically those products deemed potentially attractive to a user, leading to improved browser-to-buyer rates, extended cross selling opportunities and increased customer loyalty while also reducing the time and effort spent by consumers on searching (Schafer et al 1999 , Hervas-Drane 2007). A further (and unforeseen) benefit to businesses came with the realisation that recommender systems were a key factor in enabling consumers to explore niche markets, thereby powering an increase of interest in items on the 'Long Tail' of the sales distribution curve.

Development in areas other than e-commerce has also been rapid, and recommender systems have been applied to almost every domain where a perceived information overload problem can be found. Libraries however have been slow to add recommendations to their catalogues. Earlier research by this author found only around 5% of UK libraries included recommendations in their online catalogue, concluding that in an age of shrinking budgets, library professionals are perhaps yet to be persuaded that the effectiveness and utility of recommendation systems are great enough to warrant their implementation becoming a priority (Wakeling et al. 2012).

There are a number of reasons why it is reasonable to imagine that recommendations might be welcomed by users of library catalogues. Aside from numerous studies which report recommendations as a feature requested by library users (e.g. Craven et al. 2010; Connaway, 2007), the huge growth in available online resources has led to a wealth of accessible information with the potential to inform scholarship and facilitate teaching and research, while the provision of that information over the web has dramatically increased the range and scope of services available to users (Webster et al., 2004). Despite these undoubted benefits, however, it has become increasingly clear that traditional models of information retrieval frequently fail to best connect users with potentially relevant material. As Baez et al (2010) note, the sheer volume of available material often demands that searches are narrowed to identify core resources rather than broadened to take in potentially useful but perhaps ill-defined items. This is exacerbated by the increasing preference of users – born out of practice and familiarity - for self-guided web-based searches, bypassing the traditional role of the librarian or information professional (McNee et al., 2004). Furthermore in an increasingly competitive information market-place, institutions are becoming ever more conscious of the need for their services to match the expectations of users in both appearance and functionality.

As well as offering a key differentiating factor for early institutional adopters, a growing body of evidence suggests that recommender systems can help users find and use information more efficiently (Vellino & Zeber 2010, McNee et al., 2004, Webster et al., 2004). This raises several key questions for researchers seeking to apply the principles of recommendation to the academic library domain. Since many such applications are required to serve a diverse community of users, previous implementations have tended to orientate themselves towards a generic user model (Avancini et al., 2005) – a problem given the increasingly diverse user population (Smeaton & Callan, 2005). That this diversity is manifest not only in a demographic sense, but also from a task-orientated perspective, further complicates the development of systems that can truly be described as "useful" – in other words closely attuned to the context of the search (McNee et al., 2004). As was noted as far back as 1986, "on any given system, people will search in different ways, with different levels of success and satisfaction" (Borgman, 1986: 393). This is exacerbated when one considers the diverse make-up of catalogue users, with "a population of information seekers that is heterogeneous in terms of age, language, culture, subject knowledge, and computing expertise" (Borgman, 1996: 494).

1.2 Project Background

This research project represents a collaboration between the University of Sheffield's Information School and OCLC, an international not-for-profit library cooperative. Founded in 1967 by a consortium of Ohio libraries, its purpose was to "create a non-profit membership organization dedicated to the public purposes of furthering access to the world's information and reducing information costs" (OCLC 2012). Most importantly this meant the development of a networked platform on which to build shared resources, thereby increasing access to information while lowering operational costs among member libraries. Today OCLC has more than 72,000 member libraries in 170 countries. OCLC's operations cover a broad range of activities, including the provision of library management systems, library advocacy, research and product development, and cataloguing services. Aside from its ownership of the Dewey Decimal Classification system, it is perhaps best known for WorldCat, the aggregated catalogue of the holdings of its member libraries. Representing the largest bibliographic database in the world, WorldCat contains more than 250 million bibliographic records and more than 1.8 billion holdings. This catalogue is publicly accessible via a web interface found at http://www.worldcat.org.

This project is borne out of work undertaken in 2009 by Professor Paul Clough (Sheffield Information School) and Dr. Lynn Silipigni Connaway (Senior Research Scientist at OCLC and former visiting faculty at the Information School), which was focused on the analysis of transaction log files from WorldCat.org. The richness and potential of the log data, allied with other research conducted by OCLC suggesting a recommender system might benefit the users of online catalogues, led to a proposal being submitted to the AHRC for a PhD Studentship. The accepted project proposed an investigation of the potential utility and optimum design features of a recommender system for WorldCat.org.

1.3 Union Catalogues and WorldCat.org

WorldCat.org is best thought of as a union catalogue. The International Encyclopaedia of Library and Information Science defines a Union Catalogue as "a catalogue that contains not only a listing of bibliographic records from more than one library, but also locations to identify holdings of the contributing libraries" (Feather & Sturges1997, p. 451). Union catalogues are certainly not a recent invention. The concept was widespread as early as the start of the twentieth century (Hartley & Booth 2006), leading to the creation of vast and comprehensive national catalogues (the 764 volume *National Union Catalogue of pre-1956 Imprints* being perhaps the best known). The later emergence of networked technologies naturally heralded the arrival of the online union catalogue, and with the adoption of the Z39.50 standard came the opportunity for distributed virtual catalogues (Hider 2004). Today there exist numerous union catalogues, aggregating collections by geography (for example AMICUS in Canada, OhioLINK for Ohio, US), format (SUNCAT for periodicals, ENCORE for musical scores), and subject (for example EVOCS for Chinese studies, UCABLIS for art books), or myriad combinations of these factors. Thus as new catalogue technologies are unifying collections at the micro (institutional) level, so union catalogues by definition do so at a macro level (be it consortia, national or global) (Hartley & Booth 2006).

WorldCat is the largest union catalogue in the world, holding the combined collections of the many thousands of participating libraries from all parts of the world. It holds not only books, but journals, theses, microfilm and all types of digital media. Newly participating libraries are able to add their catalogue records as batch uploads, while existing contributors can regularly update their holdings to reflect new acquisitions. Thus the catalogue records visible to users of the service represent metadata added at an

institutional rather than central level. Each item in the catalogue is assigned a unique OCLC ID, which is visible on the item's record page in the catalogue, and as part of the URL when accessing the catalogue online. It is also useful to note that OCLC operate a proprietary subject hierarchy known as Conspectus. This scheme operates at three levels (Division > Category > Subject) and allows for the aggregation of call numbers from most classification schemes. Thus items contained within the WorldCat database are assigned as conspectus code based on their classification within other classification schemes. The Conspectus is not in the public domain, and OCLC were unable to provide access to the full scheme for this project. It can be noted though that Conspectus codes are uniformly 8 digits long, with the first two digits relating to one of 24 top-level subject divisions. The Conspectus code of an item is not displayed on WorldCat.org.

The web interface at WorldCat.org is essentially that of a typical next-generation online catalogue. The homepage (**Figure 1-1**) shows a single search box, with the default option to search the entire database. Users can select one of four other tabs to limit their search to books, CDs, DVDs or Articles. A link is also available to the advanced search page, which offers users a range of options for constructing more focussed queries (**Figure 1-2**). The homepage also offers users the opportunity of signing in to their WorldCat account, or registering for one. Users with an account are able to add tags, reviews and ratings, save searches, and create and publish personalised lists.

Users who execute a search are presented with a search results page (**Figure 1-3**). This consists a list of search results with thumbnail book-cover (although this is not always present), and a range of options for filtering and refining the search results. These offer the opportunity of limiting by resource format, as well as using facets to limit results by Author, Year, Language, Content, Audience Level and Topic. Clicking on a result title takes users to the relevant record page in the catalogue. This page (Figure 1-4) includes a range of standard bibliographic details. Of these, the Author and Subject headings are hyperlinked; clicking on these links executes a search of the database by author (au:) or subject (su:) respectively. Also shown is a five star ratings scale based on WorldCat user ratings. A link at the top of the page ("Cite / Export") allows users to generate a citation in one of a number of popular formats, or to export the citation to a reference management agent. Below the bibliographic details, the *Find a Copy in the Library* section of the page presents details of
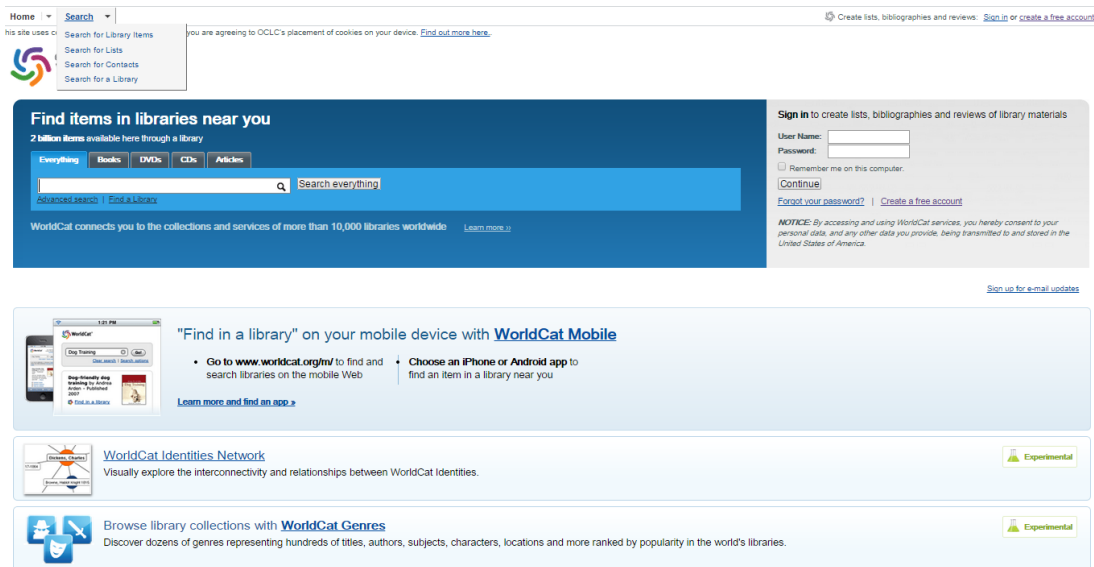
Figure 1-1: WorldCat.org Homepage
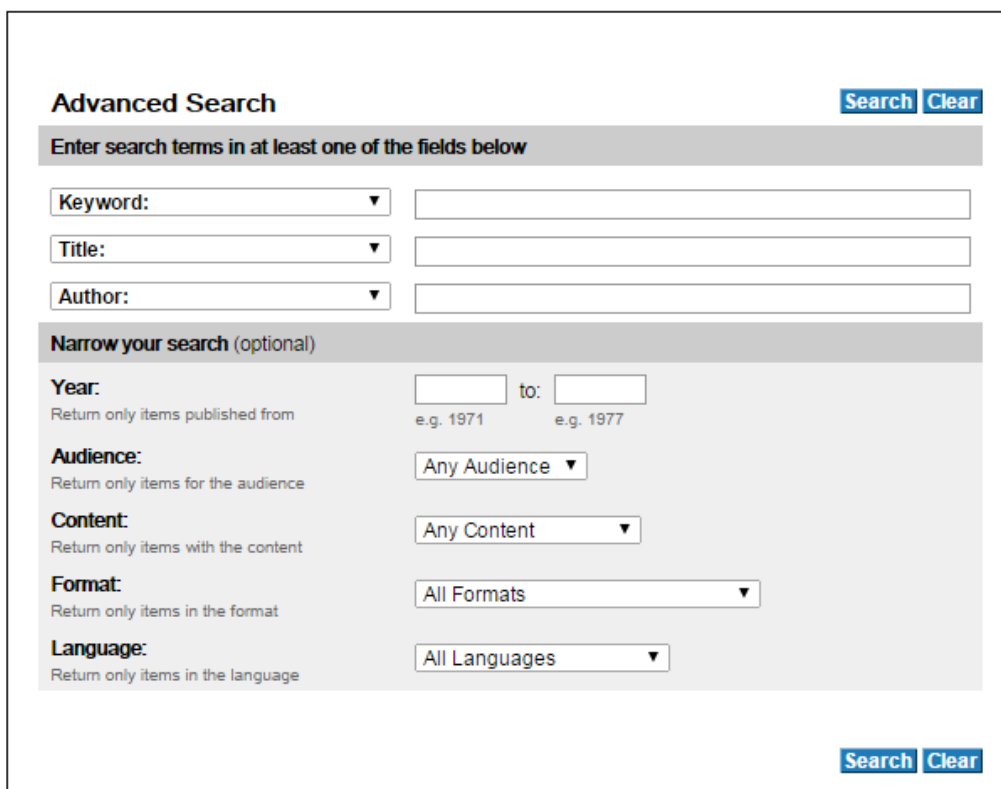
Figure 1-2: WorldCat.org Advanced Search interface
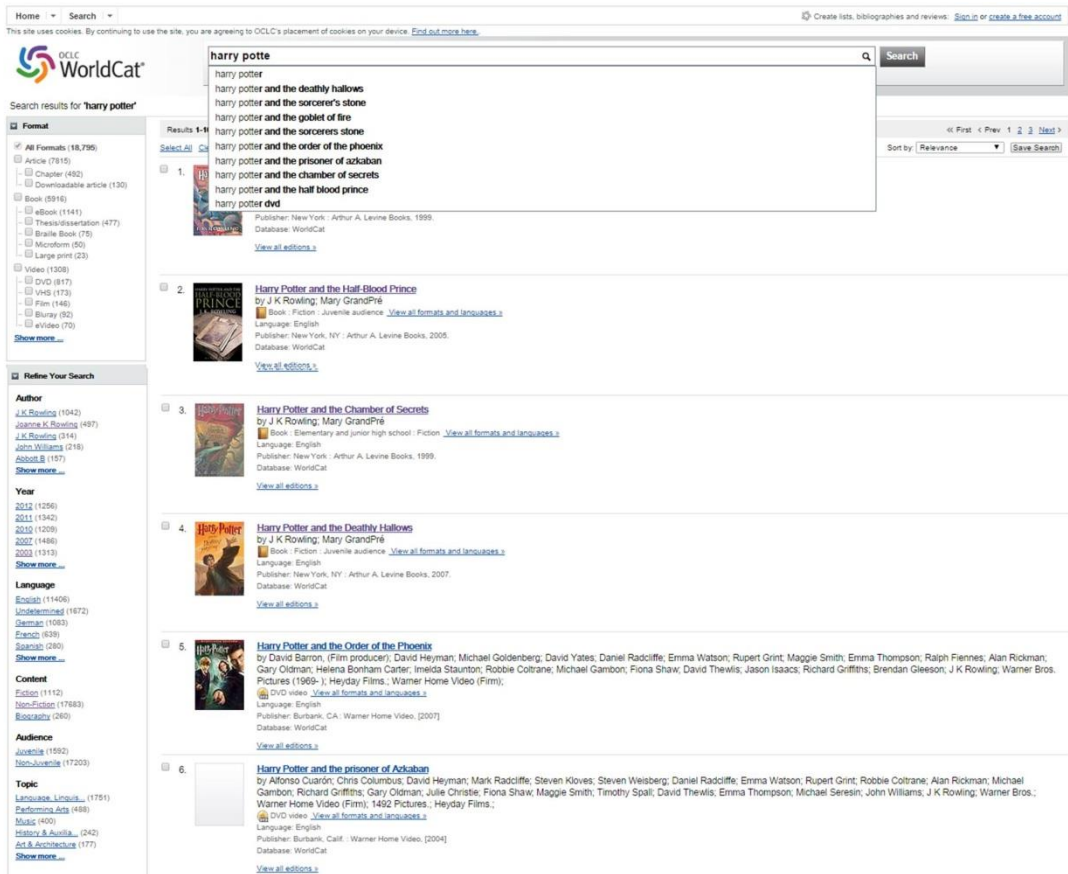
Figure 1-3: WorldCat.org Search Results page

libraries holding the item, ordered by proximity to the user, once they have entered a postcode. The remainder of the page includes links to retailers selling the item, and more detailed bibliographic details. Also included for some items are reviews by WoorldCat.org users, and other reviews imported from GoodReads, and any tags that other users have applied to the item. Finally the page shows all subject headings related to the item, and details any published user lists that include the work. While ratings, tags, reviews and lists are visible to all users, content can only be added by users with a WorldCat.org account.

As well as being indexed by Google and other major search engines since 2004, links to WorldCat.org records appear in numerous places on the web, most notably Wikipedia, and GoodReads.

OCLC WorldCat®

Advanced Search | Find a Library

[Search]

Cite/Export | Print | E-mail | Share | Permalink

Add to list | Add tags | Write a review | Rate this item: ☆☆☆☆☆

Bleak house

Author: Charles Dickens; Hablot Knight Browne
Publisher: New York : Knopf, ©1991.
Series: Everyman's library (Alfred A. Knopf, Inc.), 8.
Edition/Format: Book : Fiction : English View all editions and formats
Database: WorldCat
Summary: Bleak House defies a single description. It is a mystery story, in which Esther Summerson discovers the truth about her birth and her unknown mother's tragic life. It is a murder story, which comes to a climax in a thrilling chase, led by one of the earliest detectives in English fiction, Inspector Bucket.
Rating: ☆☆☆☆☆ (not yet rated) 0 with reviews - Be the first.
Subjects: London (England) -- Fiction.
Illegitimate children -- Fiction.
Guardian and ward -- Fiction.
View all subjects
More like this Similar Items

Nearby libraries

Sheffield Hallam University, Collegiate Learning Centre
Sheffield S10 2BP, United Kingdom
< 1 m / km

Broomhill Library
Sheffield S10 5BR, United Kingdom
< 1 m / km

University of Sheffield, Health Sciences

Librarian? Claim your library

Enter your location: S1 4DP [Find libraries]

Displaying libraries 1-6 out of 1701 for all 97 editions

Show libraries holding just this edition or narrow results by format

« First ‹ Prev 1 2 3 Next › Last »

Library	Held formats	Distance	
1. Doncaster Libraries Doncaster, DN1 3JE United Kingdom	Book	28 km MAP IT	Library info Ask a librarian Add to favorites
2. Kirklees Libraries Huddersfield, West Yorkshire, HD1 2UJ United Kingdom	Book	35 km MAP IT	Library info Ask a librarian Add to favorites
3. Nottinghamshire Libraries Nottingham, NG2 7QP United Kingdom	Book	37 km MAP IT	Library info Ask a librarian Add to favorites
4. Wakefield Libraries and Information Services Wakefield, WF2 0DQ United Kingdom	Book	37 km MAP IT	Library info Ask a librarian Add to favorites
5. Tameside Libraries Central Library Ashton Under Lyne, OL6 7SG United Kingdom	Book	42 km MAP IT	Library info Add to favorites
6. Stockport Libraries Stockport Central Library Stockport, SK1 3RS United Kingdom	Book	45 km MAP IT	Library info Ask a librarian Add to favorites

« First ‹ Prev 1 2 3 Next › Last »

Seller	Price
Better World Books	$8.48

Details

Genre/Form:	War stories Fiction
Material Type:	Fiction, Internet resource
Document Type:	Book, Internet Resource
All Authors / Contributors:	Ernest Hemingway

Find more information about: Ernest Hemingway ▼ [Go]

ISBN:	0684837889 9780684837888
OCLC Number:	36103848
Description:	297 p. ; 25 cm.
Responsibility:	Ernest Hemingway.
More information:	Publisher description Sample text

Abstract:

By turns romantic and harshly realistic, Hemingway's story of a tragic romance set against the brutality and confusion of World War I cemented his fame as a stylist and as a writer of extraordinary literary power. A volunteer ambulance driver and a beautiful English nurse fall in love when he is wounded on the Italian front. The best American novel to emerge from World War I, A Farewell to Arms is the unforgettable story of an American ambulance driver on the Italian front and his passion for a beautiful English nurse. Hemingway's frank portrayal of the love between Lieutenant Henry and Catherine Barkley, caught in the inexorable sweep of war, glows with an intensity unrivaled in modern literature, while his description of the German attack on Caporetto -- of lines of fired men marching in the rain, hungry, weary, and demoralized -- is one of the greatest moments in literary history. A story of love and pain, of loyalty and desertion, A Farewell to Arms, written when he was 30 years old, represents a new romanticism for Hemingway.

Reviews

Tags

Add tags for "A farewell to arms".

All user tags (2)
View most popular tags as: tag list | tag cloud

k (by 1 person)
read (by 1 person)

Similar Items

Related Subjects: (2)
World War, 1914-1918 -- Fiction.
War stories.

User lists with this item (16)
ZZZ Submitted to Judge 0904 2014 Donated (261 items)
by jeffrerjacks updated 2014-09-15

Figure 1-4: WorldCat.org Record page

1.4 Terminology

In referring to the distinction between a book (or other resource) and different editions or versions of that book, the Functional Requirements for Bibliographic Records (FRBR) model will be used. Thus the term *work* will refer to the distinct intellectual or artistic creation, and *manifestation* to a distinct published version. A newly published edition of the novel *Bleak House* would be a *manifestation* of a *work*.

1.5 Research Questions and Objectives

The research questions grew out of the key gaps identified in the review of relevant literature described in Chapter 2. Despite the ubiquity of next-generation features in library catalogues, there is little work which properly investigates the potential utility of recommendations to users of these systems, and indeed the broader recommender systems literature does not clearly address how recommendations are incorporated into users' information search strategies. Furthermore, little seems to be known about the users and uses of union catalogues in general, and WorldCat.org in particular. This project intends to address these gaps by first investigating the WorldCat user base; who they are, and what they are using the system for. Second it attempts to determine the role recommendations can play in supporting information-seeking within the library catalogue, thereby addressing the lack of work that specifically addresses the role of recommendations in library catalogues, the final research questions and objectives were therefore as follows:

Research Questions:

> 1. Who is using WorldCat.org?
>
> 2. For what purposes are users accessing WorldCat.org?
>
> 3. When might a recommender system support users of WorldCat.org?
>
> 4. What effect does the presence of recommendations have on the information seeking process in the library catalogue?
>
> 5. What recommendation characteristics would be most useful to users of library catalogues?

Research Objectives:

1. Identify who is using WorldCat.org, and their reasons for accessing the system

2. Establish user needs and expectations for a WorldCat.org recommender system

3. Develop a set of design specifications for a WorldCat.org recommender system

1.6 Access and Support

OCLC agreed to support the project with the following resources:

- Access to WorldCat.org Transaction Logs
- Access to OCLC Usability Laboratory
- Access to WorldCat holdings table
- 10% of OCLC Senior Research Scientist's time
- 10% of OCLC software engineer's time
- £1,000 per year student award

1.7 Thesis Structure

Following this introduction, Chapter 2 presents a review of the literature relevant to this research. Chapter 3 describes the methodology used to address the research questions, while Chapters 4-7 report the findings of each of the four research phases that constitute this project. Chapter 8 integrates and discusses these finding, and presents a conceptual design of a recommender system for WorldCat.org. Chapter 9 is a summary of the key findings, and offers ideas for future work in this area.

2. LITERATURE REVIEW

2.1 Literature Review Methodology

A literature review can serve a range of purposes. Hart (1998) identifies eleven potential functions of the literature review, of which five constitute aims of this chapter:

- Giving context, both theoretical and historical, for the research questions
- Explaining the significance of the problem
- Understanding how research theory has been applied in practice
- Identifying state of the art developments
- Identifying the principle research methodologies that are suited to the subject

The chapter is divided into four main parts. The first describes research relating to the use and development of union catalogues. This section also summarises research that has focussed on the WorldCat system. The second part details the development of modern library systems, and covers studies relating to the functionality and usability of such systems. The third section reviews literature relating to the field of Information Behaviour and its sub-disciplines, Information-Seeking and Search behaviour. The last section provides a non-technical overview of recommender systems research, particularly identifying the key research areas in the field that relate to this project, and the intersection of recommender systems and library services.

A wide range of databases were consulted during the resource identification phase, including Web of Science, ACM Digital Library, Google Scholar, SCOPUS, and Emerald Insight. In addition key journals in the various fields were consulted, along with the proceedings of relevant conferences. A vast number of search terms were used, beginning with general broad terms (e.g. "union catalogue"; "worldcat"; "information behaviour"; "recommender systems"), which resulted in the discovery of immediately relevant resources from which useful citations could be garnered to widen the search. To ensure completeness increasingly specific search terms were employed (e.g. "recommender system" and "serendipity" and "diversity"; "usability" and "nextgen" and "library").

2.2 Union Catalogues

Broadly speaking, the literature on union catalogues can be divided into the practical and the conceptual. From the conceptual perspective, some authors maintain that the traditional role of the union catalogue is primarily a driver for inter-library loan and resource sharing (Hider, 2004; German, 2007). Others however see potential for union catalogues to play a broader role in the new information landscape. Lass & Quandt (2004) argue that the traditional uses of Union catalogues (shared cataloguing, quality control, inter-library loan) have been expanded to include "the possibility of online search and text delivery" with a single point of access. This intersection with web services is best examined by Gradmann (2004), who notes that while the exposure of union catalogues on the world-wide web is essential, the fundamental differences in approach between library and web systems must be acknowledged. In practice this mean recognising that "library-based information systems are based on the idea of mediated access, whereas the original principle of WWW-based systems is one of direct, instant access" (Gradmann, 2004: 77).

From a practical perspective a number of authors have discussed information architecture issues relating to union catalogues, particularly the relative strengths and weaknesses of distributed and centralized models (Cousins, 1999; Hider, 2004; Cousins, 1999), while there exist a number of case studies detailing the technical and organisation requirements behind establishing new or improved union catalogues (for example Alam & Pandey, 2012; Larsen, 2007; Burnhill & Law, 2005; Mittal, 2011; Boston et al., 2005). A further subset of the union catalogue literature describes more user-orientated studies. Hartley & Booth (2006) describe a study investigating how users use and view union catalogues, comparing COPAC (a union catalogue of more than 70 UK and Irish University and Research libraries) with three UK regional union catalogues. Their methodology utilised observed search sessions (with volunteers completing predetermined tasks), interviews, and focus groups. As the authors note, the search scenarios developed for the research were based on "search types which experience had suggested…are put to union catalogues" (2006: 13), and the study therefore reveals more about user search strategy and perceived system performance than it does about how exactly union catalogues are used in the real world. Results show a wide variety in the search strategies utilised by participants, and the authors identify several key aspects of union catalogue functionality required to serve the varying needs of users, including faceted search, the avoidance of library jargon, information about item availability, and the ability to use a range of search criteria.

Librarians who participated in the focus group section of the study expressed concerns about the currency and data quality of union catalogues, and were sceptical about the number of library users who were aware of union catalogues, a view supported by the study's finding that student participants demonstrated "an almost total lack of awareness of the concept of a union catalogue, even amongst those who were aware of a specific union catalogue." (2006: 14). The authors note the impact of Google and Amazon on the expectations of users of all types, and among their recommendations is the concept of a national union catalogue allowing users to limit their search to customizable geographic areas.

Further work on COPAC is reported by Craven et al. (2010). A total of 12 post-graduate students and academic staff participated in their study, which utilised a combination of focus-groups, interviews and controlled search tasks to examine the usability of the catalogue. The findings reveal a range of perceived usability issues including the use of confusing library terminology, lack of links to external content (particularly reviews and ratings), some navigational difficulties, and the absence of facetted search features. Suggested improvements for the service centred on functionality for managing information that had been found (e.g. user lists, and exporting to reference management software), and additional information about items in the catalogue (extended summaries, reviews and recommendations). The authors also noted that there was considerable discussion among participants as to what the scope of the COPAC service should be, particularly in relation to augmenting the basic searching and locating services with additional functionality. They conclude that most participants preferred COPAC to focus on facilitating the discovery of and access to quality resources, rather than expanding the service with more Web 2.0 functionality. It should be noted that the findings of this study influenced the design of a new user-interface for COPAC (Jeskins & Cousins, 2011).

Goodale and Clough (2012) take a more holistic approach in their user evaluation of the SEARCH25 system, a prototype successor to InforM25, the union catalogue of more than 60 members of Academic Libraries in the South-East of England. Their study includes a survey of users, as well as log file analysis and focus group sessions. The survey reveals the most common tasks for which users frequently utilise the system relate to known-item searches, with 85% of respondents doing this often or very often. Discovery tasks, such as searching by subject, are less popular, although more than half of all users (59%) still regularly conduct these searches. Post-graduate students and academics / researchers were the

group most likely to search for new items in a particular subject area. The survey also indicated that users most valued InforM25 for its item coverage, seeing the system as a potential "one-stop-shop". Negative opinions about the system were found to centre on interface and usability issues, the prevalence of duplicate records, and the absence of holdings and circulation data. Analysis of a sample of the log files reveals the average (mean) number of actions per session to be 3.8 (median = 1, mode = 1), with a majority of sessions (53.8) consisting of just one action, and 85% of sessions consisting of 5 or less. The authors also analyse users' search behaviour, in terms of query formulation, the viewing of search results over multiple pages, and an analysis of the most common search terms and types. The report also highlights some typical use scenarios, gleaned from focus group sessions with users of the system. While not exhaustive, it is suggested that the four scenarios presented represent the majority of use of the system. Two of the scenarios represent a librarian using the system (either undertaking cataloguing and or assisting a patron find an item at a reference desk) while the other two involve a student or researcher (finding a comprehensive and diverse range of material on a topic, and determining which libraries hold certain collections).

Perhaps the most notable aspect of the literature review conducted for this project is how little work has been done to identify who is using union catalogues, and what precisely they are using them for.

2.2.1   WorldCat.org

In a 2006 paper, Lavoie et al. identified a number of areas of research for which WorldCat data could be used, namely collaborative collection management, collection views, library decision making, user behaviour and trend-spotting. A review of the literature suggests that some of these fields have proved more fruitful than others. A significant body of research has utilised the WorldCat holdings data as a means of analysing, benchmarking and assessing collections and collection development, many building on the work of Perrault, whose study of monographic records in WorldCat found that the presence and accuracy of holdings in WorldCat mirrored that of research library collections, therefore concluding that "WorldCat is a rich resource for cataloging records, verification of the existence of titles, and identifying prospective materials for resources sharing" (2002: 2). Subsequent work includes Lavoie et al.'s mapping of the WorldCat "digital landscape", which offers a comprehensive analysis of the WorldCat's digital holdings, a relatively small

but fast growing subsection of the total holdings (Lavoie et al., 2006), Connaway et al.'s analysis of WorldCat holdings to identify "last copies" (2006), and Bernstein's investigation of item coverage and holdings levels (2006). A number of papers also offer practical accounts of using the WorldCat collection development tools (LIST). A further subsection of the literature centres on the concept of a work in the context of the Functional Requirements for Bibliographic Records (FRBR) project, with WorldCat used a both a tool for analysing multiple manifestations and entity families (Bennet et al., 2003; Smiraglia & Leazer, 2004) and as a vehicle for pilot projects to incorporate work level records within the catalogue (Gatenby et al., 2012; Ercegovac, 2006). Other studies have used WorldCat as a point of comparison to Google books. Chen (2012) used random sampling to compare the coverage of both systems, finding that over 80% of WorldCat holdings were retrievable through Google books, (although only 7% of titles had the full-text available for free). Others have compared metadata quality across the two systems (for example Nunberg, 2009; Oder, 2009), finding that despite its imperfections, WorldCat's data quality is generally superior.

User behaviour – another of the potential areas of research identified by Lavoie et al. – has seen very little active research. Nilges reported some usage patterns from the initial launch of Open WorldCat (2006), which focus primarily on the access points to WorldCat and the types of search behaviour exhibited by users. Based on a sample of log files, Nilges shows that users are most likely to access Open WorldCat records via a two to four term keyword search, and that the WorldCat result was on average the sixth result displayed in Yahoo! Search results, although a substantial number of clicks were from results ranked outside the top ten, indicating that "WorldCat does serve a constituency of more determined researchers who tend to dig deeper into results sets" (Nilges 2006; 442-3). Users were also found to click on a "Find a Library" link approximately 4-6% of the time.

The only other significant study on WorldCat.org user behaviour is an OCLC Report – *Online Catalogues: What Users and librarians Really Want* (Calhoun et al., 2009). This research takes a user-centred approach to the question of data quality in WorldCat, and consists of end-user focus-groups, a pop-up browser survey for users accessing WorldCat.org, and a separate survey of librarians. The pop-up survey, which collected 11,151 total responses, showed librarians making up 32% of respondents, with postgraduate (15%) and undergraduate (13%) student making up a further 28%. Teachers and academics constitute 22%, with "Business Professional" and "Other" accounting for the majority of the

remainder. Although the focus of the research was on existing data quality, and potential improvements to the system, the study does distinguish between two typical types of task that users undertake – known-item (i.e. accessing information about a particular pre-identified item) and discovery (i.e. using the system to find and evaluate potentially useful items) – and acknowledges that these tasks make different demands on the system. Overall, users (excluding librarians) prioritised links to full-text content, relevant search results, item availability information, an advanced search option, and "evaluative content" (summaries, tables of content etc.). The expectations and practices of these users were seen to be directly influenced by their experience of searching the web. Librarians, by contrast, identified the merging of duplicate records as their highest priority, followed by the correction of typographical errors and the upgrading of brief records. They did however share other users' preference for more evaluative content. Calhoun notes that librarians' understanding of structured data, and their more nuanced appreciation of cataloguing issues, influenced the forms of improvements they suggested be made to WorldCat. In addition, system functionality such as an advanced search option and facets were found to aid users of all types in their exploration of the catalogue. Overall the study notes that users of all types access WorldCat purposefully, with librarians likely to be carrying out "work responsibilities", and other users seeking resources to address some information need.

While the report provides useful background to this study, particularly in its use of survey and focus groups as a methodology to establish user perspectives on the system, there are some limitations with regard to its relevance to this project. First, although the report discusses other aspects of system functionality, the primary goal is to address what constitutes data quality. Thus the research described in the report is guided by a focus on content rather than functionality. Second it is notable that different methods were used to generate data from different user groups, with librarians canvassed using a separate survey, and not included as part of the focus group research. While the experiences and perspectives described by different user groups offer some post hoc justification for this, it does seem problematic to draw such strongly delineated boundaries between users of the same system. In this sense it is significant that the language of the study excludes librarians from the category "end-users". While perhaps understandable as a means of easily distinguishing the two broad user-groups for the purposes of the study, this categorisation does speak to a curious dichotomy wherein librarians, despite being users of the system,

hold a somehow elevated status. As the report acknowledges, "Librarians' perspectives about data quality remain highly influenced by their profession's classical principles of information organization, while end users' expectations of data quality arise largely from their experiences of how information is organized on popular Web sites." (2009: vi).

Overall we might conclude that whilst WorldCat has proved to be fruitful source of research in a number of areas, there has yet to be work conducted which focusses specifically on the needs and behaviour of its users.

2.3 Evolution and Functionality of the Library Catalogue

2.3.1 Introduction

For all institutions, both academic and public, the catalogue represents the key system for library users attempting to access items within a collection. Norris's *History of Cataloguing* (1939) offers a comprehensive review of the evolution of library organisation and cataloguing, a tradition that might be said to begin with Callimachus's *Pinakes*, an attempt to both list and categorize the holdings of the Library of Alexander. While as late as the 19th Century there remained advocates, on the grounds of simplicity and economy, of uncategorised lists of items held, it was generally understood that "neither the effective disclosure of library materials nor the user's best interests are well served by such a list" (Dempsey, 1999: 3). This recognition that catalogues should support a variety of user needs led explicitly to the development of guidelines for catalogue functionality, most notably as defined by Cutter in his *Rules for a Printed Dictionary Catalog* (1876). Cutter identified three broad objectives of a catalogue – to support the identification of an item based on some prior bibliographic knowledge (the Identifying Objective), to show what content the library has by a given author or on a given subject (the Collocating objective), and to aid the selection of a book by providing details of the edition and subject matter (the Evaluating objective). To support these objectives, catalogues would provide complementary ordering of content, most typically by Author, Title, and Subject, thereby allowing the catalogue user to choose the most appropriate access point for their search (Taylor, 1986). While the earliest catalogues were generally in book format, the twentieth century saw the emergence of the card catalogue, with a card representing each item held in the library, and separate sets of cards ordered according to different metadata attributes. Card

catalogues offered the obvious advantage of being easily expandable, allowing libraries to augment the catalogue with new acquisitions (Chan, 2007).

It is primarily card catalogues that the earliest electronic catalogues, known as Online Public Access Catalogues (OPACs), began to replace. An OPAC can be described as "a database composed of bibliographic records describing the books and other materials owned by a library or library system, accessible via public terminals or workstations" (ODLIS). OPACS were perhaps the first information retrieval systems available for use by non-expert users, and for many of those users interaction with an early OPAC represented a first experience of networked technology and online searching (Fast & Campbell, 2004). The heterogeneity of the user population and the primacy of OPACs as a discovery tools meant that OPACs were once seen as a crucial "test-bed" for addressing user requirements within Information Retrieval systems (Beaulieu & Borgman, 1996). This is surely no longer the case. Indeed it might be said that libraries spurned a glorious opportunity, and that "the promise of online catalogs has never been realized" (Antelman et al., 2006: 128). The complexity of the underlying databases, and the technical and budgetary limitations common to many libraries has meant that the evolution of OPACs has been at times painfully slow (Emanuel, 2009).This fact, when combined with the spectacular emergence of the world wide web as a vast and familiar discovery tool, has meant increasing numbers of users bypass library systems completely (Calhoun, 2006). To better understand the implications of these issues, and to fully appreciate the current state of the art, it is perhaps worth reviewing the history of online catalogue development.

2.3.2   First and Second Generation OPACs

As many authors have noted, the first OPAC systems to emerge in the 1970s and 80s represented little more than automated extensions of card catalogues (Borgman, 1996; Husain & Ansari, 2006; Antelman et al., 2006; Buckland, 1992). In practice this meant that searchers were limited to the same "access points" as traditional catalogues, generally Title and Author fields, and results were often displayed in order of acquisition (with most recently acquired items appearing first) (Husain & Ansari, 2006). Many systems required users to use coded commands (e.g. "a" for author), and searches were left-anchored. The card catalogues on which such systems were modelled had changed little since the 19th Century, and the first OPACs were designed to support users familiar with those catalogues (Buckland, 1992). Since most systems were only able to match exact left-anchored search

strings, they were extremely unforgiving of user mistakes. It also became clear that systems were designed primarily to support known-item searching, and as such were ill-suited to subject searching or browsing (Antelman et al., 2006; Borgman, 1986), meaning that in many situations users were unsure how best to frame their query. For these reasons, it was soon acknowledged that the first generation of OPACs required substantial improvements in order to meet users' needs. In her influential paper "Why OPACs are hard to use" Christine Borgman argued that the online catalogues of the time were difficult to use because their design did not incorporate sufficient understanding of searching behaviour, and that "we do not yet have sufficient knowledge of user behaviour to make major improvements in systems design and training" (1986: 397). She proposed a program of user-centred research that would develop this knowledge, and allow for the design of systems that were better suited to the needs of library patrons.

Writing ten years later, after the emergence of the second generation of OPACs, Borgman was compelled to title her paper "Why OPACs are *still* hard to use" (Borgman, 1996). These second generation systems merged the existing card-catalogue model with Boolean systems developed in other fields of information retrieval, and represented a clear improvement in functionality. Keyword searching was now possible, as well as truncation and wild card support, index term browsing, use of full MARC records, and (limited) subject access to items (Hildreth, 1995; Husain & Ansari, 2006). More access points to the catalogue were offered, and increasingly user-friendly displays and interfaces were developed. This new generation of systems also offered the advantage of being relatively simple to implement, and had modest storage and processing requirements (Antelman et al., 2006). Nonetheless, significant problems remained. Whilst Boolean algebra was a clear advance on the card-catalogue model, it remained difficult to use, and has long been recognised as suitable primarily for experienced searchers (Salton, 1984). Furthermore, significant issues remained with subject searching, which despite the improvements offered by Boolean algebra and index term browsing was still unsatisfactory for users. Once again Borgman was forced to conclude that "online catalogs continue to be difficult to use because their design does not incorporate sufficient understanding of searching behaviour" (Borgman, 1996: 493).

2.3.3   Next Generation Catalogues and Discovery Tools

By the middle of the 2000s a growing body of research was emerging that began to shed light on the behaviour of online catalogue users. Studies of academic library users found that "familiarization, convenience, currency and authority" were valued most by users in their selection of resources (Connaway & Dickey, 2010), but that the "poor usability, high complexity, and lack of integration" of many resources "acts as a barrier to information search and retrieval" (JISC, 2009). The traditional view of user goals as static and self-contained was also challenged, as studies determined that goals were malleable and heavily influenced by the user's interaction with the system, and that furthermore this interaction was informed by a range of contextual factors such as the searcher's experience, mood, attitude and prior knowledge (Hert, 1996). This in turn led to a realisation that catalogue interfaces were ill-suited to supporting a heterogeneous user population with rapidly evolving search goals (Husain & Ansari, 2006; Ballard & Blaine 2011). These interfaces, and indeed the system's functionality as a whole, were unable to support the type of "informal and opportunistic" browsing that users were accustomed to among the physical library shelves (Marchioni, 1995; Poulter, 2003), and that card-catalogues could at least simulate (Antelman et al., 2006). The role of the catalogue was further challenged by a series of studies that showed a declining use of library systems, and an ever-growing reliance on broader web-based tools. Undergraduates in particular were found to engage predominantly in subject searches (Connaway et al., 1997), and to use tools such as Google or Wikipedia to aid resource discovery (Kitalong et al., 2008, Connaway, 2007; Little, 2012), to a point where only 2% of college students were found to start their research on library websites or catalogues (OCLC, 2006). Users' everyday familiarity with the "one-box" searching paradigm of modern web search engines led to a "Googlized" library patron for whom the sophisticated search methods of librarians were anathema (Woods, 2010; Emanuel, 2009; Ballard & Blaine, 2011), and who were found to be increasingly reliant on a small number of information resources within which they could run keyword searches (Head & Eisenberg, 2009). As Calhoun put it in a 2006 Library of Congress Report, "today's library catalogs are long on problems and short on unique benefits for users" (Calhoun, 2006: 9).

This change in user expectations, allied with an increasingly sophisticated understanding of search behaviour has led to the emergence of a new generation of catalogues and associated catalogue search tools (Fagan et al., 2012). Generally termed "Next Generation" (or "nextgen") catalogues, these systems represent a significant advance on previous

OPACs, and can perhaps best be described as "finding aids" (Morgan, 2007). As well as offering improved interfaces and greater scope for customization (Emanuel, 2009), nextgen catalogues introduced a range of features and functionality to the library domain, which Ballard & Blaine (2011) summarise as follows:

- **Enriched content**

  MARC records supplemented with additional data streams e.g. cover art, tables of content, reviews.

- **Faceted navigation**

  The presentation of relevant subcategories, allowing the user to narrow the search results within defined parameters

- **Keyword searching**

  A single search box for initial queries (although "advanced search" functions allow for traditional anchored searches e.g. Author and Title)

- **Relevancy ranking**

  More sophisticated ranking algorithms, and a variety of ranking options given to the user

- **Did you mean. . .?**

  Suggested spelling alternatives

- **Recommendations**

  The presentation of related content to the user

- **Web 2.0 or social network features**

  User generated content (e.g. tags, ratings and reviews) and the incorporation of social media sharing tools such as Facebook and Twitter

To this list we might also add the introduction within some systems of "virtual bookshelves" – essentially replications of the physical library shelf in digital form. In most cases these visualisations are provided at the item level, and present thumbnail cover images of the books that surround the item in question on the actual library shelves (see Figure 2-1). While a number of practitioner blogs have discussed the implementation of bookshelf features (e.g. Tay, 2013; Pattern, June 2008), there appears to be no academic study of their use or utility.
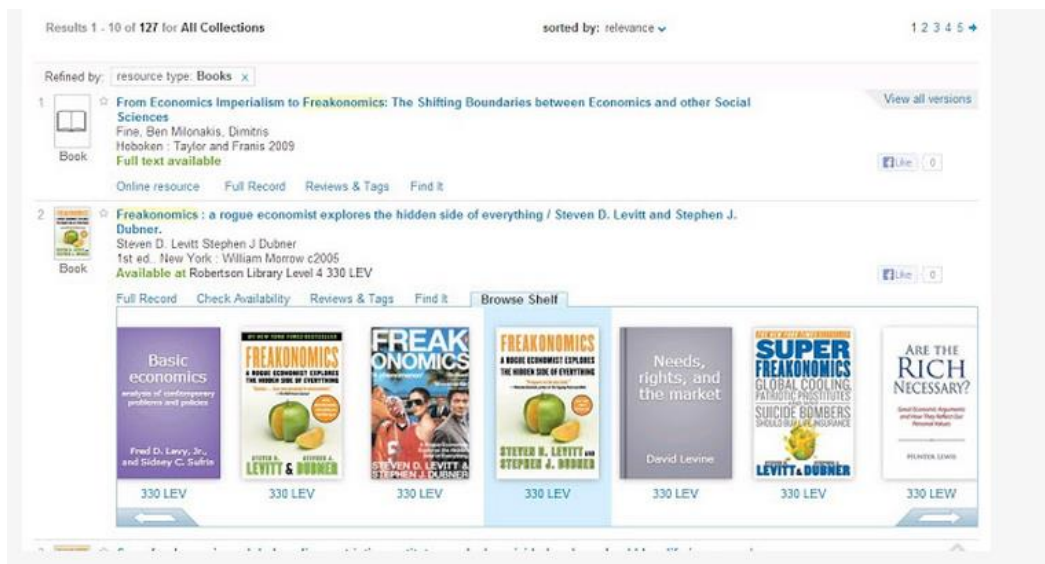
Figure 2-1: Example of a Virtual Bookshelf feature - Curtin University

While these new features undoubtedly represented an improvement on previous OPACs , early usability studies of nextgen catalogues showed that they were still ineffective in connecting users with the full range of library provided resources – specifically the ejournal and ebook collections not held in the catalogue (Fagan et al., 2012; Ballard & Blaine, 2011). Federated and meta-search tools had already emerged as a prospective solution to this issue, systems which allowed users to enter a single search string to query multiple databases. There now emerged a range of products termed "discovery tools" (sometimes "discovery platforms"). While federated search relied on the search algorithm and relevance ranking of the individual databases being queried, discovery tools utilize a centralized search index of all available resources (be they held centrally by the library or available electronically through subscription or non-subscription databases) and their own proprietary search and ranking algorithms to present a unified set of results to the user (Lown et al., 2013). The result is a service that matches the experience users find elsewhere on the web, with improved "speed, relevance, and ability to interact consistently with results" (Fagan et al., 2012: 84). In theory, libraries utilizing a discovery tool negate the need for their users to select an appropriate starting point for their search, since all relevant material is accessible from a single search box. As Breeding puts it, "this new genre of discovery interfaces has revolutionized the library catalog, modernizing it into a form more consistent with other web destinations" (Breeding, 2010: 33). It also of interest to note that for Breeding, any distinction between nextgen catalogues and discovery tools is increasingly blurred, since from a users' perspective much of the functionality is the same.

He therefore prefers the term "discovery interfaces" to describe these new systems which seek to connect library patrons with content, regardless of format or physical or virtual location (Breeding, 2010).

Despite the improvements in functionality and usability, some doubts remain about the utility of these new systems. Are users properly able to distinguish between the different resources types returned by discovery tools, and navigate through the large results sets (Fagan et al., 2012)? Does a single search box discourage users from investigating subject databases that might be more appropriate for their search? It has also been suggested that nextgen catalogues and discovery tools support broader subject searching at the expense of what was once the library catalogue's key purpose – enabling the locating of a specific item (Emanuel, 2009). In considering these issues it is instructive to review the growing body of research examining the practical usability of new systems.

2.3.3.1 Usability Studies of Nextgen Catalogues and Discovery Tools

Borgman's plea, made in 1996, that "online catalogs should be judged by their success in answering questions rather than by their success in matching queries" appears at last to have been heeded (Borgman, 1996: 500). In attempting to assess the utility of nextgen catalogues and discovery tools, recent research has taken an overwhelmingly user-centred approach. A review of the literature shows an increasing preference for studies that engage directly with the user, be that through task-based observation of users (e.g. Youngen, 2010; Swanson & Green, 2011; Denton & Coysh, 2011), focus groups and interviews (e.g. Butters et al., 2009; Bertot et al., 2012; Connaway 2007), log analysis (Ballard & Blaine, 2011), or some combination of the three (Craven et al., 2010). Data acquired through these methods have frequently been supplemented with wider survey data.

The results of these studies suggest that users have reacted positively to the new tools. In his usability study of WorldCat local in Illinois, Youngen (2010) found that users were almost universally positive about the new features on offer, with the list of areas of satisfaction matching almost exactly the features described by Ballard and Blaine (2011) as characterizing nextgen catalogues. Butters et al. (2010) drew similar conclusions in their study of the UK COPAC service, determining that users particularly valued functions that encouraged broader use and exploration of resources. A number of studies also identify

faceted browsing as perhaps the single most appreciated feature of the new catalogue (Emanuel, 2009; Fagan 2010). Numerous studies have also shown that users are overwhelmingly positive about nextgen interfaces, particularly their user-friendly nature, and the ability to sort search results and quickly access relevant bibliographic information (Denton & Coysh, 2011; Craven et al., 2010; Swanson & Green, 2011; Antelman et al., 2006). Ballard & Blaine's study of log data comparing the use of the nextgen Encore system to that of the old catalogue found that users stayed longer, ran more searches, and viewed more pages in Encore than they did in the classic catalogue.

All this is not to say that issues with nextgen catalogues have not been identified. Based on extensive usability testing of two nextgen catalogues- Vu Find and WorldCat Local - Emanuel (2009) concludes that contrary to expectations, nextgen catalogues may be of most use to those users already familiar with the underlying structure of library catalogues. She argues that while the new systems may work well for broad subject searches, they frequently frustrate the user seeking a specific item, who is forced to refine and limit initial result sets to find the required item. While experienced searchers, using advanced search functions or well defined search strings, are able to overcome these issues, the casual user is not. She suggests that nextgen catalogues and discovery tools must seek to improve their algorithms and relevancy rankings to better support these users. Emanuel's research also suggests that in mimicking the interface and usability of web search engines, nextgen catalogues further raise user expectations, meaning that they are less inclined to spend time limiting results and reformulating queries: "If they did not get what they wanted, they quickly assumed the library did not have what they were looking for" (Emanuel, 2009: 120). These arguments are echoed in a number of studies showing that while students are often very positive about the nextgen experience, librarians are less likely to share this enthusiasm (Bertot et al., 2012; Youngen, 2010; Arcolio & Davidson, 2009). While it has been suggested that this may in part be due to a younger demographics' willingness to engage with new technology, it also implies a belief on the part of librarians that the new systems are in some ways inadequate. Other problems which were inherent in first and second generation catalogues have also apparently not been addressed. Swanson & Green, in their review of the usability literature, note that many users demonstrate "a lack of context and familiarity with library-related research" (2011: 223). New systems can still be jargon heavy, with users expected to understand the differences between library-held items, subscription databases, open access material and so on. While this may be an

unavoidable consequence of systems constructed upon a legacy of rigid and proprietary databases, it is still an issue that new catalogues can address (Little, 2012).

2.4 Information Behaviour, Information Seeking and Information Search Behaviour

Tom Wilson, one of the subject's foremost theoreticians, has defined Information Behaviour (IB) as "the totality of human behaviour in relation to sources and channels of information, including both active and passive information seeking, and information use" (2000: 49). This is meant to embrace all situations, with no requirement for agency on the part of the user, or significance on the part of the information. The earliest examples of IB research can be found in studies relating to the use of library services in the first decades of the 1900s, although not until the middle of that century did studies of users begin to address questions relating to information need and use more directly (Case 2012). With the emergence of technology providing hitherto unimagined access to and reliance on information in the 1980s and 90s, approaches to modelling and understanding information behaviour became increasingly diverse, particularly with regard to the level of abstraction with which researchers approached the issue (Saracevic, 2011). By 1999, Wilson was able to propose a nested model of research areas within the general field of IB (Figure 2-2). For Wilson, Information Behaviour is the overarching field concerned with the overriding framework within which Information Seeking Behaviour (ISB) takes place. Information Seeking itself refers to patterns of behaviour within the process of finding and accessing information, while Information Searching behaviour is more focussed yet, and relates to the interactions between user and system that take place within any information seeking endeavour.
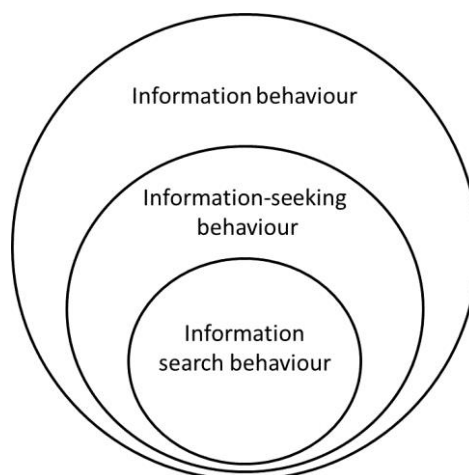
Figure 2-2: Nested Model of Information Behaviour Research Areas (Wilson 1999)

Wilson's nested model has been helpfully adapted by Jansen & Rieh (2010) to demonstrate how the areas of Information Behaviour research can be seen to apply to a similarly nested model of Information Systems (see Figure 2-3). They suggest that the broad field of Information Behaviour can be seen as relating to Information Systems at all levels, with systems supporting the users who access them. At the intermediate level, information seekers make use of the information supplied by the systems they encounter, while at the "micro" level, the features and functionality of electronic IR systems enables users to engage in the searching and browsing activities that constitute information searching behaviour (Jansen & Rieh, 2010: 1518).
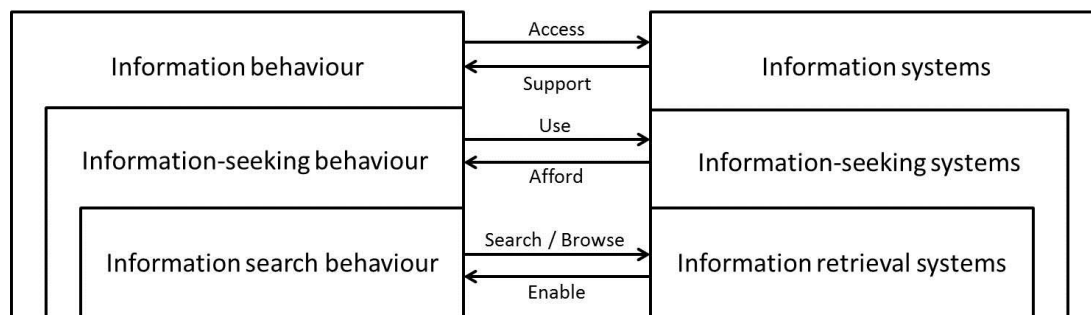
Figure 2-3: Framework of human information behaviour and information systems (Jansen & Rieh, 2010)

The principal aim of this project, which is to investigate how recommender functionality can best serve the users of the WorldCat.org system, sits most naturally in this micro level. The intention in this section therefore is not to critically evaluate the multitude of models and theories relating to IB and ISB. As Case notes, "the diversity of theoretical borrowings makes a single, comprehensive comparison impossible" (2002: 140). Instead, most attention will be paid to research most closely relating to Information Searching Behaviour. However it is important to recognise, as Wilson has shown, that information searching activities occur within a broader theoretical context. The next section therefore outlines some key theoretical models which serve to inform our understanding of users' behaviour when searching.

2.4.1   Theoretical Models

Perhaps the most conceptual approach to explaining information behaviour can be found in Dervin's work on sense-making (1996, 2003). In its broadest form, sense-making

encompasses not just task or problem solving, but "how humans make and unmake, develop, maintain, resist, destroy, and change order, structure, culture, organization, relationships, self" (1996: 1). A highly abstract theory, it argues that information is best considered as a tool for interpreting a reality that can be both structured and unstructured (2003). It identifies four elements to information behaviour – a situation, a gap (in knowledge), an outcome, and a bridge (or means of closing the gap), stresses the perspective of the user rather than the structure or systems within which they sit, emphasising the individual context within a set of abstract variables. Dervin's work is perhaps best used as an intellectual construct against which to assess the mechanics of information behaviour.

Kuhlthau (1991, 1994, 2005) offers an information behaviour process framework founded on empirical observation. She proposes a sequential model that incorporates the personality of both task and user, and covers six distinct stages; task initiation, topic selection, pre-focus exploration, focus formulation, information collection, and search closure. For each of these stages Kuhlthau also identifies the thoughts and feelings that the user typically experiences. The acknowledgement and mapping of these emotional responses to the information process offers an enhanced picture of information use, and echoes the user-centric conceptualisation found in Dervin's work. As Kuhlthau herself notes, "the model … offers an articulation of users' common experiences which, when shared by the user, the intermediary, and the system, may provide a basis for interaction" (1991: 370). It should be noted that Kuhlthau locates her conception of information behaviour within the context of task completion. The model is therefore perhaps most useful as a means of mapping evolving cognitive states onto the linear activities that constitute discrete information seeking contexts.

A comparison of two models proposed by Wilson reveals the increasing focus on user personality as a core component of information behaviour (1981, 1996). The 1981 model presents information seeking behaviour as means of resolving an information need through the interaction with both formal and informal information sources. Changes to the mode and frequency of these interactions are influenced by their success and failure. But as Wilson himself later noted, the model lacks a means of incorporating or evaluating the contextual effects on the use – i.e. "whether the various assumed barriers have similar or different effects upon the motivation of individuals to seek information" (Wilson 1999: 253). Influenced by models such as those by Kuhlthau (1991, 1994) and Ellis (1989), Wilson

attempts to remedy this in his later model expanding the types of information seeking behaviour and classifying the 'intervening variables' – which are now acknowledged as potentially positive agents (Wilson 1999). Absolutes such as "success" and "failure" no longer feature in the model, suggesting a more nuanced approach to the issue of task completion, while the context which stimulates need is also separated from the decision to commence the information seeking-process, which again allows a greater role for influencing external and internal variables (see Figure 2-4). The information seeking process is expanded to include four modes of information acquisition. As defined by Wilson, *Passive attention* describes situations in which information is acquired without being directly sought (for example listening to the radio). Wilson also includes three modes of search; *passive*, which relates to occasions when information that the user is not specifically seeking, but is still in some way relevant, is obtained during a search; *active*, which describes the more prosaic process of intentionally seeking some certain information; and *ongoing*, which refers to the practice of monitoring or updating an information seeker's state of knowledge that has already been informed through other information acquisition modes.
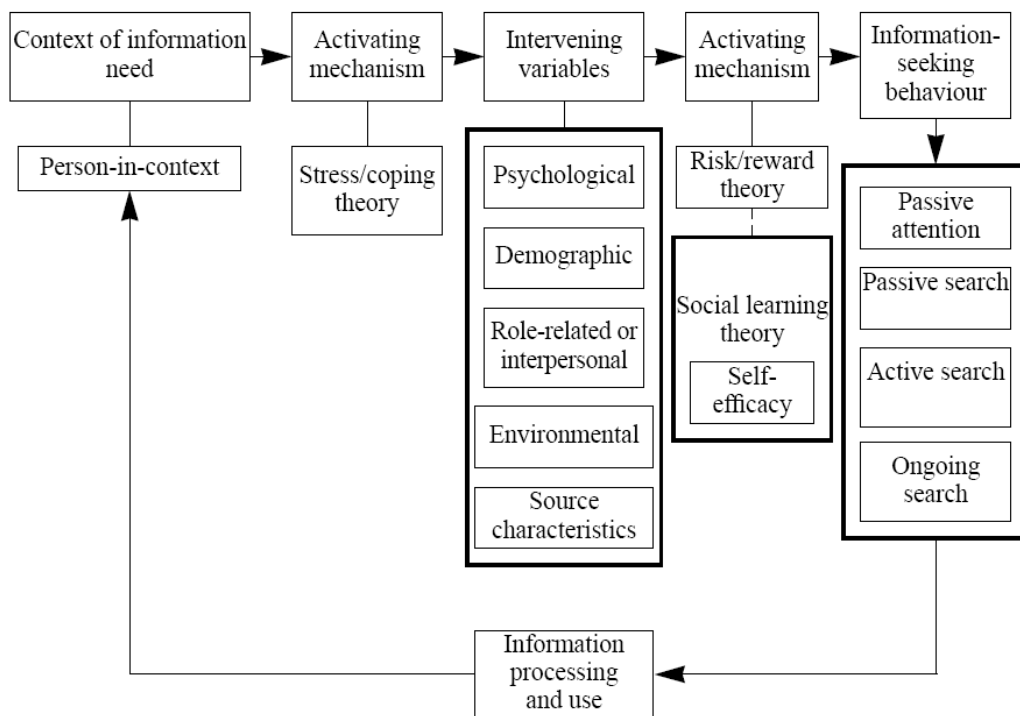
Figure 2-4: Wilson's model of Information Behaviour (1999)

While Wilson's later model acknowledges the role of context in understanding information behaviour, it stops short of any more detailed exploration of how contextual issues

28

mediate the process of information seeking. This point is addressed in great detail by Ingwersen, who offers a cognitive framework that maps the causal links between the cognitive space of both user and information system (Ingwersen, 1996). The framework is founded on the notion that both the creation and reception of information represent "acts of information processing" (p. 5), and that both therefore are inexorably tied to the cognitive states of the agents involved. An individual's information need can be seen as a direct result of their cognitive state, the problem space they inhabit, and the broader task and domain within which the individual is located. Furthermore, the characteristics of the information retrieval system, and its interface, are in a sense manifestations of the cognitive states of their creators. Thus the linear process described by Wilson and others might be seen as a series of cognitive transformations, with each moment of transformation representing the interaction of a plurality of cognitive states (see Figure 2-5). A great strength of Ingwersen's model is its incorporation of the IR system, and more importantly its recognition that the characteristics of the system play a key role in influencing the cognitive transformations of the user and the information they seek. It is perhaps most usefully used as an adjunct to Wilson's model, representing a further layer of understanding outside the three areas of research shown in Wilson's nested model.

Figure 2-5: Cognitive Framework of Information Retrieval (Ingwersen 1994)
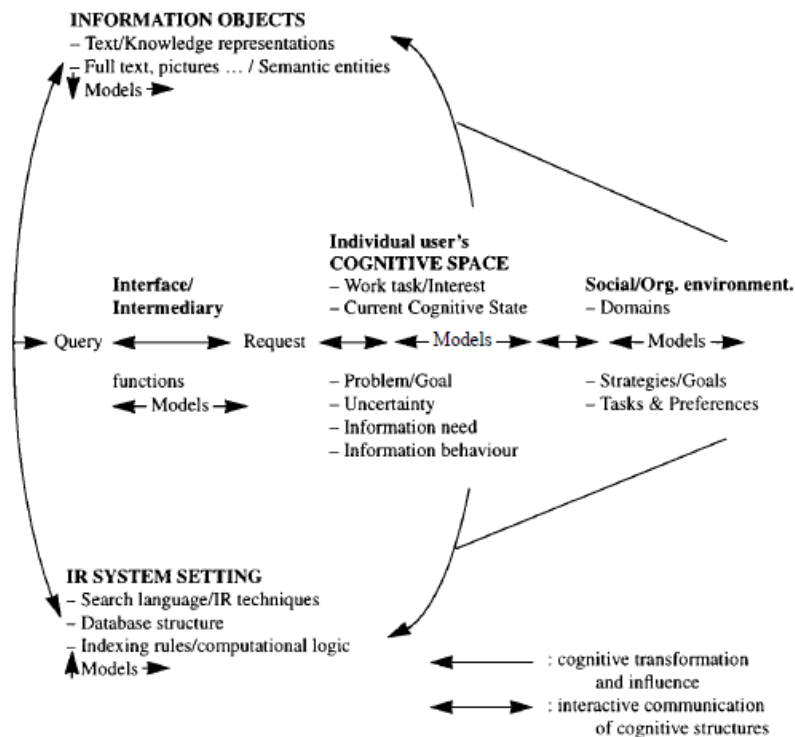
Wilson's model is not without its issues. The essentially linear process it depicts has been challenged by some researchers who perceive human information behaviour as more chaotic and interrelated than represented in the model (e.g. Foster 2004, Dresang 2005). It is also, as Wilson himself notes, a model of "macro-behaviour", and it might be argued that certain features of the model (for example *Information Processing and Use*) are so complex as to defy easy aggregation as a single construct. Nonetheless, in a theoretical landscape sorely lacking in consensus, Wilson's model offers a generally accepted overview of the key factors influencing general human information behaviour, and as such will serve as a basic conceptual underpinning to this thesis.

In seeking to better understand perspectives on Information Search Behaviour, it is perhaps easiest to start with activities that might be said to fall within Wilson's Active Search element. Within IR research, a standard model of behaviour has emerged that attempts to represent the user interaction process with a typical IR system (Broder, 2002; Shneiderman et al., 1997; Sutcliffe & Ennis, 1998). This process has been described as the *query-response paradigm* (White & Roth, 2009), and while different authors include varying levels of contextual detail, the essential act of searching is consistent - a cycle of 1) query formulation, 2) execution, 3) reviewing of results, and (if necessary) 4) query reformulation (Shneiderman et al., 1997; Broder 2002). Suttcliffe & Ennis (1998) expand this basic pattern by including elements relating to problem identification and the articulation of information need. They also recognise a variety of strategies employed by system users at each of the stages. Of particular interest here is their acknowledgement of varying strategies for query formulation, ranging from very detailed and specific strings intended to obtain the appropriate result immediately, to broader more general search terms which can be refined over each iteration. They note that the selection of appropriate search terms, and the effectiveness of reformulations, is to an extent dependent on a user's domain knowledge. Since there is a requirement "to find lexical terms which express the searcher's goal", if the user lacks the conceptual knowledge to produce a range of search terms they are more likely "to acquire search terms from the environment" (Suttcliffe & Ennis, 1998: 328).

These classical models are perhaps best viewed as the most micro-level depictions of the search process. Indeed one might suggest that the classical model says less about how searchers use IR systems, and more about how such systems were *designed* to be used. This issue is exacerbated when we consider the limited functionality of the classical IR

system when compared to more feature-laden information retrieval support systems commonly found on the web (Yao et al, 2012). There is also a sense in which the classical model is best viewed as charting the process focused searching; that is the circumstance of a searcher having a clear idea of both their information need and, more importantly, the most effective way of utilising the system to satisfy that need (White & Roth, 2009). The last thirty years of Information Search Behaviour research has increasingly been driven by an understanding that information seekers frequently operate in circumstances where the information need is ill-defined or evolving, and that an iterative cycle of query, response and reformulation fails to fully represent searchers' attempts to address these needs. Some attention has been paid, for example, to situations when searchers are unable to articulate their information need in the form of effective queries. O'Day & Jeffries, in their analysis of this phenomenon, suggest that users in these circumstances employ a strategy they term *orienteering*. This consists of conducting broad searches that users believe will take them to "a part of the information space containing potentially relevant documents", within which the searcher can utilise their "recall and recognition skills to locate relevant information" (White & Roth, 2009: 17-18).

Work in this area has been most heavily influenced however by two frequently cited models, both of which view information searching as analogous to ecological processes. Marcia Bates' berry-picking model (Bates, 1989) is founded on a rejection of the classical IR model on the basis that "the query is treated as single unitary, one-time conception of the problem" (Bates, 1989: 409). She argues instead that as the searcher encounters information during the search process, their conception of the query (and the specific information need it represents) evolves. Thus the information required to satisfy the need does not constitute a single discrete document or set of documents; instead the need is satisfied by "a series of selections of individual references and bits of information at each stage of the ever-modifying search" (1989: 410). It is this "bit-at-a-time" retrieval that Bates likens to the act of gathering berries from different bushes, and a number of studies testing the model using naturalistic and observational methods have presented results supporting it (e.g. Ellis, 1989; O'Day and Jeffries, 1993; Borgman, 1996). In considering the implications of the model, Bates also notes the variety of strategies used to locate information in electronic catalogue environments (citation chaining, author searches etc.), and concludes that the classical model of IR only really applies to the querying of indexed document collections. Considering the optimal design of systems to support searching she concludes

that "if we want to meet users' needs, we should enable them to search in familiar ways that are effective for them" (1989: 414).

While the berry-picking model generally establishes the evolving nature of the information need, and the dispersed nature of relevant information, others have attempted to examine and explain in more detail the factors that influence searcher behaviour when engaged in berry-picking. Pirolli & Card (1995, 1999, 2007) do so with a similar appropriation of an ecological idea – this time optimal foraging theory. This biological theory states that an organism's foraging strategy is governed by a desire to maximise the ratio of energy gained to time taken. Pirolli & Card co-opt this concept in their Information Foraging Theory, arguing broadly that a searcher's behaviour while navigating an information system adapts to the perceived value of the information, and the perceived costs in terms of time and effort of locating and evaluating that information. More precisely, the theory attempts to rationalise the strategic information search process through the introduction and modification of three concepts from the originating biological theory: *patches*, *diet*, and *scent*. The notion of an *information patch* is similar to Bates's core analogy – that relevant information is distributed unevenly within an environment. Patches can be thought of at varying levels of granularity, with different patches representing different information access methods (asking a librarian or using a catalogue), different systems (using a library database or Google Scholar), or different sets of search results. For Pirolli & Card, the information seeker must constantly evaluate the richness of the patch they currently inhabit, and the time and effort required to locate other fruitful patches (1999). Having located a useful patch, searchers can utilise techniques to enrich the environment, be that through the formulation of more precise and effective queries, or by utilising filtering features to raise the proportion of relevant results. More broadly, the information system might be enriched to offer reduce the time and effort required to navigate between patches.

Within an information patch, searchers consume an *information diet*, i.e. the selection of documents or resources that are perceived as relevant. Once again, different sources "will differ in their access costs or prevalence, and they will differ in profitability", and searchers must determine the most effective strategy to maximise their overall information consumption (Pirolli & Card, 1999: 11). *Information scent* represents the perceived value of and cost of both different information sources within a patch, and different information patches themselves, as based on the "imperfect proximal cues" encountered by the

searcher (1999: 10). In a system environment, examples of cues are citations, document summaries or abstracts, or other text snippets. The activity of the information forager can therefore be summarised as the efficient navigation within and between systems, guided by the incidental information encountered along the way, with the purpose of locating resources that most effectively satisfy the information need. In relating Information Foraging Theory to systems development, Pirolli & Card note that as the expectation of searchers increases, particularly in terms of the perceived range and effectiveness of available systems, they are less inclined to stay in any one information access environment.

Both berry-picking and information foraging are theories that seek to illuminate the processes of non-focused searching. While both theories are applicable to a variety of information environments, relating them to a hyperlinked system naturally brings us to another key concept in information seeking, namely browsing. In this context, browsing has been most broadly defined as movement in a connected space (Kwasnik, 1992; Palay and Fox 1981; White & Roth, 2009). It is important to note however the term has been used widely but inconsistently in the LIS literature: sometimes as a means of describing a particular form of document or list scanning, and at others as a more comprehensive means of describing information interaction within a system (Rice, 2001). For the purposes of this study we adopt a definition more closely aligned with the latter interpretation, and see browsing as a series of encounters with information snippets, which in turn can lead to further examination of the resource, or a continuation of the surveying process (Bates, 2004).The literature relating to browsing seems to agree on three general types of browsing: *directed*, where the browser is seeking a particular item; *semidirected*, where the browser has a defined purpose, but a less definite result in mind, and; *undirected*, where the browser has no specific goal at all (at least as relates to a result) (Choo et al., 2000). In each case, Bates argues that the act of browsing consists of four elements:

1. glimpsing a field of vision;
2. selecting or sampling a physical or representational object from the field;
3. examining the object; and
4. physically or conceptually acquiring the examined object, or abandoning it.

(Bates, 2007)

These elements are derived from a cognitive and behavioural analysis of the browsing process, a process applicable to any environment or context. Within an electronic environment, depending on the type of browsing being undertaken, users will be more or

less likely to employ certain systematic search and evaluation tactics – most typically the formulation of general conceptual queries to ensure broad results sets, the following of trails between potentially related items, and the consumption of information snippets, (Marchionini, 1995; White & Roth, 2007 & 2009; Choo et al., 2000).

There are circumstances when browsing and searching are used in concert, and these types of interactions have been termed exploratory search. Building on the work of Marchionini (2006), White et al. identify two senses in which the term *exploratory search* can be applied to information seeking: 1) "to describe an information-seeking problem context that is open-ended, persistent, and multi-faceted"; 2) "to describe information-seeking processes that are opportunistic, iterative, and multi-tactical" (White et al., 2006: 6). While the first sense can be seen to have some relevance to our overall understanding of search-tasks, it is the second part of the definition that it most relevant to this discussion. Exploratory search in this sense occurs when an information seeker is unable to achieve their goal through the traditional IR process model. The result is that the user employs "a combination of searching and browsing behaviour to navigate through (and to) information" (White & Roth, 2009: 10). Browsing activity is employed first for the purposes of locating some
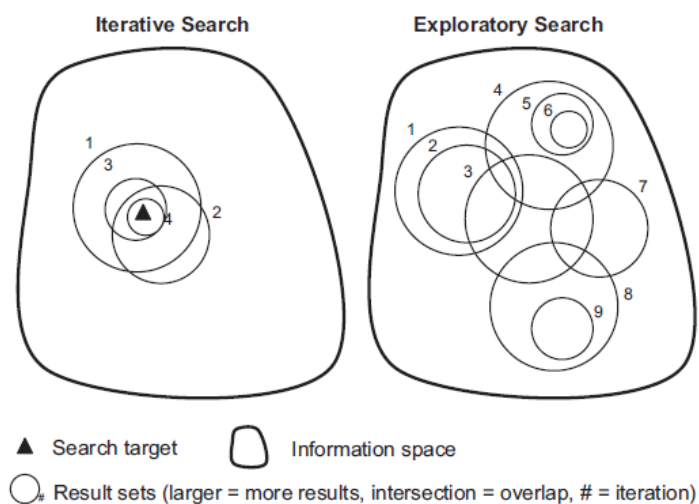
Figure 2-6: Iterative search versus exploratory search strategies (White & Roth, 2009)

resources or information that might illuminate the information need, and offer context for further investigation of the information content. Once searchers are better informed about the field, this further investigation is more and more likely to take the form of focused searching, of the type described in the standard models of IR. This exploratory approach is

more likely to result in the exposure to a greater number and range of potential results than the cyclical query-response paradigm (see Figure 2-6).

An acknowledgement of the importance of exploratory search as mode of information seeking has led in turn to the discussion and development of system functionality specifically designed to support it. Such tools can take a number of forms. Most generally, maximising the opportunity for link navigation has been shown to aid users' exploration, and remove the need for constant query reformulation (Marchionini 2006). Similarly it has been argued that interface tools designed to encourage user interaction with the system (for example sliders and dynamic screen updates) serve to encourage exploratory behaviour (Shneiderman & Plaisant, 2005). Other tools include support for query formulation and refinement (for example suggesting common search terms), the incorporation of facets and metadata-based filtering, the use of visualisations (as a means of allowing users to better comprehend the collection as a whole), supporting multi-session search episodes through storage and sharing functionalities, and leveraging the search context (user, situation and task information) to optimise the material presented (White & Roth, 2006). Support for collaborative search, including synchronous searching episodes, has also been suggested (Pickens et al., 2008).

2.4.1.1 Information Search Behaviour in the OPAC

Although it has been noted that the information search behaviour of students in the library catalogue is not uniform, or easily generalizable beyond the institutions within which research takes place (Rowlands et al., 2008), there are nonetheless common themes across the literature. It has been noted that despite their often vast experience of searching on the web, users of library catalogues do not use sophisticated search strategies (Connaway & Dickey, 2008; Rowlands et al., 2008). This may in part be due to a relative unfamiliarity with catalogue interfaces, and it has been observed that users spend a large proportion of sessions navigating the system (Rowlands et al., 2008). While next-generation catalogues are often feature rich, users have regularly been shown to make little use of advanced search functionality (Lau & Goh, 2006; Babu & Tamizhchelvan, 2003; Jansen & Pooch, 2000), perhaps because many systems do not offer sufficient support for the functionality (Connaway & Dickey, 2010). Similarly, users rarely use Boolean operators and other search limiting techniques (Favart, & Passerault, 2004; Lau & Goh, 2006; Ballard & Blaine, 2011), and the number of query terms used is typically small (Villén-Rueda & Senso, 2007; Jansen

& Pooch 2000). Although students have been shown to value a clear search input and results display, even at the expense of other features (Kani-Zabihi et al., 2008), users do make regular use of faceted browsing features (Kules et al., 2009), and the use of faceted browsing has been observed to improve the range of resources discovered (Olson, 1997). Post-graduate and undergraduate students have also been observed to exhibit different search behaviour, with postgraduate students exhibiting behaviour closer to that of academics in the breadth of their search, use of advanced search operators, and evaluation of relevance (Spink, 1993; Connaway & Dickey 2008).

Research has also extensively examined the utility of subject searching in OPACs. It is important to note here that in the context of the OPAC, subject searching relates to querying a controlled list of subject classifications (often the Library of Congress Subject Headings) (Villén-Rueda & Senso, 2007).Although recognized as a potentially powerful tool, subject searching is often utilized poorly, with users frequently attempting web-search style keyword searches which yield poor results (Connaway et al., 1997; Antell & Huang 2008; Grey 2012). A consequence of this is a perception by users that the system simply doesn't contain relevant items, meaning that rather than attempt to reformulate their query they abandon the search and seek resources elsewhere (Griffiths & Brophy, 2005; Kumar, 2011).Unsurprisingly, users want OPACS that are easy to use, and that produce reliable and relevant results (Kani-Zabihi et al., 2008). In reality students often see OPACs as "closed" and "rigid" environments that often return too many results or too few (Fast & Campbell, 2004; Villén-Rueda & Senso, 2007).

2.4.2   Satisficing

The term "satisficing", which originates in the economic theorising of Herbert Simon (1955), has been appropriated for the information science domain by Prabha et al., for whom the term signifies "an information competency whereby individuals assess how much information is good enough to satisfy their information need" (2007: 75). This concept is of some relevance to this study, since understanding the mechanisms by which users determine the end point to information usage cycles is likely to strongly influence our appreciation of the various information behaviour models, and perspectives on system effectiveness. Prabha et al. see their work relating closely to the sense-making framework, in that Dervin's work acknowledges an incomplete reality within which any information seeking action can only ever be partially fulfilled. This implies that no user can ever

36

completely satisfy an information need, but must instead determine the point at which the information gathered is sufficient. Naturally this point is considered variable depending on the context and requirements of the task at hand (Krikelas, 1983; Taylor, 1968), while others have emphasised the influence of internal factors (motivation, existing knowledge, search skill) in the completion equation (Foster, 2004). Ellis (1997) has observed that users may still seek information even at the very end of a task. This may imply that any decision to stop searching is an uneasy one, or instead be a symptom of working practices that do not conform to linear information process models – a suggestion supported by Wai-Yi (1998), who observed auditors in Singapore approaching tasks in a much more fluid way.

Testing these theories in an academic setting, Prabha et al have found that the criteria for determining when information satisfices vary according to role, with students driven primarily by objective assignment requirement (for example a minimum required number of sources), while academics are likely to use both qualitative and quantitative methods depending on the circumstances – they were far more likely to see time as a limiting factor for work related to teaching than for research, for which the search would only cease when a quality standard had been met. This latter approach supports work by Zach (2005), which demonstrated that the essential element in deeming a search complete was a certainty that the relevant task could be successfully completed.

2.4.3   Categorising Search-Tasks

A common feature of the models of information seeking behaviour described in the previous section is the recognition that the information seeking process is essentially "the advance from uncertainty to certainty" (Wilson 1999: 265). This intention is represented explicitly in some models (e.g. Wilson's *information need*, Ingwersen's work-task/interest), while for others it serves as the predicating condition (Kulthau, Ellis, Dervin), but all the models acknowledge that the modes of behaviour exhibited by the user will naturally be influenced by the desired outcome of the information seeking process. For those responsible for developing systems to support information seeking, that outcome is related to "the perceived need for information that leads to someone using an information retrieval system" (Schniederman et al., 1997: Appendix 1). It follows therefore that for researchers seeking to improve IR system performance and user experience there is clear value in better understanding and classifying those needs (Rose & Levinson, 2004; Gisbergen et al., 2007).

In discussing attempts at classifying users' information seeking interactions, we should note that there is some variation in terminology across (and sometimes within) the literature. The terms *goal*, *intent*, *task*, and *need* are all commonly found, and their use is far from universally consistent. Before continuing the review of literature relating to classifying user needs, it is helpful to briefly define the key terms, and the relationships between them. A key issue in untangling this terminology relates to the complications of determining the level of granularity within the overall information seeking process to which each term can appropriately be applied (Bystrom & Hansen, 2005). For the purposes of this chapter we follow Toms's interpretation of the literature of task-based searching (Toms, 2011). For Toms, the predicating condition of any information seeking process is some work function, with work here understood in its broadest sense i.e. relating not only to economic but any other "extrinsic benefit" (Toms, 2011: 44). Within this work context, an individual is likely to undertake tasks, which are here thought of as defined by Hackman; "a set of assigned a) goals to be achieved, b) instructions to be performed, or c) a mix of the two" (Hackman, 1969, in Toms, 2011: 45). Thus a *goal* can be said to represent the desired outcome or objective of a task.

Understanding tasks within a work context leads naturally to the conception of the term "work-task", a term used by a number of authors to represent an overarching unit within which information seeking activities are undertaken (Vakkari, 2001 & 2003; Bystrom & Hansen, 2002). As Toms notes, the work function can consists of any number of work-tasks, and each of the tasks may themselves consist of sub-tasks. One such sub-task, and the one particularly relevant to the fields of information seeking and information retrieval, is the search-task, which represents the motivating external factors influencing user interaction with an information retrieval or support system. The word external is important here, since it helps distinguish task from the information need – the former being an extrinsic and objective set of requirements to be satisfied, the latter the personal and subjective gap in knowledge to be bridged (Pirolli, 2007). In the context of the search-task we also follow Jansen et al.'s characterisation of user *intent* as "the affective, cognitive, or situational goal as expressed in an interaction" (Jansen et al. 2008: 1255). In other words, *intent* is analogous to *goal*, but incorporates the way the goal is expressed in the user's interaction with the system.

Since the aim of this project is to better understand user requirements for a recommender system in an online catalogue, it is the term *search-task* that seems most relevant in

discussing categorisations of purposes for using library systems. However, in the discussion of relevant studies found below, the original terminology used by the authors is retained. While we have seen that each term has its own precise meaning, the categorisation schema described below are generally applicable to a variety of levels of granularity – be it session or task – and so can be said to relate to understanding types of search-tasks.

Perhaps the largest body of research in this area relates to classifying the search tasks that inform web queries. A comprehensive review of attempts at such classification schema can be found in Jansen et al. (2008).The most widely cited scheme is that of Broder (2002), whose taxonomy of web search categorises the "need behind the query" as *Informational*, *Navigational* or *Transactional*. Informational queries are defined as those attempting to locate content on a particular topic to address a specific information need, while navigational intent represents interactions with a goal of locating a particular website. Transactional interactions are those which are undertaken with the intent of reaching a site where some further activity is intended. Subsequent work in the field has tended to expand and refine this trichotomy. Rose & Levinson (2004) developed a more detailed framework based on an analysis of logs from the AltaVista search engine. They broaden the Transactional element to incorporate any session undertaken with the intention of locating a particular resource, renaming the element of the framework as Resource, and sub-divide the Informational classification to incorporate different information seeking needs (see Table 2-1). Subsequent work by Jansen et al. (2008) essentially validates this hierarchy, proposing only minor changes to the Resource element. Several attempts have been made to develop similar taxonomies relating to user search tasks in online library catalogues. The most basic form of classification distinguishes simply between searches for "known-items" (i.e. an item that the users knows to exist), and those for resources relating to a topic or subject (Matthews et al. 1983, Buckland 1979). Lewandowski (2010) maps these search tasks on to Broder's taxonomy, likening a known-item search to Broder's *Navigational* classification, and a topic search to an *Informational* intent. For Lewandowski, the OPAC equivalent of the *Transactional* search is the *search for sources*, during which a user attempts to locate a source from which to continue their information seeking (e.g. another database).

Empirical studies of catalogue use have developed alternative schemes. Hert (1996) based her analysis of user search tasks in the OPAC on her observations of students interacting

Table 2-1: Search Goal Hierarchy (Rose & Levinson. 2004)

Classification	Intent
1. Navigational	To access a particular known website
2. Informational	To learn something about a topic
2.1 Directed	To learn something specific about a topic
2.1.1 Closed	To obtain a single answer to a specific question
2.1.2 Open	To obtain an answer to an open-ended question
2.2 Undirected	To learn anything or everything about a topic
2.3 Advice	To obtain advice or instructions about a subject
2.4 Locate	To determine where some product or service can be found
2.5 List	To collate a list of potentially useful websites
3. Resource	To obtain a resource (not information)
3.1 Download	To download a useful resource
3.2 Entertainment	To be entertained by consuming a resource
3.3 Interact	To use an online service to interact with a resource
3.4 Obtain	To obtain a resource that does not require a computer to use

with the OPAC at Syracuse University. Library patrons seen approaching an OPAC terminal were invited to undertake their search in a laboratory setting, with post-session interviews to gather qualitative data about their intent. The various goals articulated by participants are reduced to four overarching types: a search for a specific known-item; a search for an unknown-item (i.e. a single resource on a particular topic); a search for information about an item (e.g. the start date of a journal); or a general search for information with no specific number or type of resource in mind. The notion of an *unknown-item* search is also found in Slone (2000), who attempted to categorize the search tasks of searchers using public library OPACs. Based on data collected from surveys, interviews and observations of students, she identifies three key types of tasks: *known-item*, *unknown-item*, and *area.* For Slone, the *unknown-item* category encompasses what other authors have termed subject or topic searches, but also incorporates search tasks that would only uncomfortably fit into the topic search category (e.g. searching for a single textbook). The *area* search relates to users who use the catalogue to determine the area of the physical library items on a particular topic are held, and then continue their searching there.

It is interesting to note at this point the similarities between these classifications of catalogue search-tasks, which emerge from user observation and interview, and the professional literature relating to the design of catalogues and cataloguing rules, in particular the user objectives that they should support. As noted in section 2.3.1, Cutter was among the first to formalise these objectives, categorising them as Identifying, Collocating, and Evaluating. Subsequent work in this area has been generally cumulative rather than revisionary, with additional authors and professional bodies adding further specifications (e.g. ICCP, 1961; Lubetzky, 1986; Svenonius, 2000). The culmination of this work can be found in the comprehensive "Objectives and Functions of the Catalogue", found in the IFLA *Statement of Cataloging Principles* (2009). They posit that a catalogue should enable a user:

1. To find bibliographic resources in a collection as the result of a search using attributes or relationships of the resources:
      1a. to find a single resource
      1b. to find sets of resources representing

- all resources belonging to the same work
- all resources embodying the same expression
- all resources exemplifying the same manifestation
- all resources associated with a given person, family, or corporate body
- all resources on a given subject
- all resources defined by other criteria (language, place of publication, publication date, content type, carrier type, etc.), usually as a secondary limiting of a search result;

2. To identify a bibliographic resource or agent
3. To select a bibliographic resource that is appropriate to the user's needs
4. To acquire or obtain access to an item described
5. To navigate within a catalogue and beyond
(IFLA, 2009: 3-4)

While this list is clearly broader in scope than a simple classification of search-tasks, it is possible to map the core concepts detailed in other models (known-item, unknown-item / topic search) onto many of the IFLA objectives. Indeed it is helpful to view the IFLA objectives as a more detailed manifestation of the other schemas– while noting that these objectives are not the result of empirical study. One point worthy of particular attention is the last item in the list – enabling users to "navigate within a catalogue and beyond". This item originated in the work of Svenonius (2000), and describes the potential need for users to recognise relationships between items, and be aided in their exploration of the

catalogue "through the logical arrangement of bibliographic and authority data and presentation of clear ways to move about" (IFLA 2009: 4).

The location of a known-item within the catalogue is recognised as a core task within the classification schema described above, and a number of studies of catalogue use identify accessing a known-item as the most common search task in library catalogues (Larsson, 1991; Yee & Layne, 1998). Given this prevalence of the term in the LIS literature it is worth perhaps considering the concept in greater detail. As Lee et al. note, "most researchers articulate their own conceptual and operational definitions of a known-item search, making little effort to explicitly connect these to the general concept and rarely providing citations to sources or authorities" (Lee et al. 2007: 3). Their discussion focuses on analyses of various definitions of the term, including:

> "A search for some item for which either the author or title is known" (Lancaster, 1991)

> "A situation in which a user is trying to find an item previously read, and consequently in which the user's memory of the item is of primary importance. (Allen, 1989; p.247)

> "A known-item search occurs when the user has a limited but correct description of an existing document. The user is sure of the fact that the document exists, that its title and author are explicitly stated somewhere in the document, and these assumptions are true to the actual state of the docuverse." (Dahlström & Gunnarsson, 2000)

> "Some people who approach a library catalog have a particular item in mind, and they want to determine whether the library holds that item and where in the library it is located. Such a person would conduct a known-item search. A known-item search may include the author, the title, the subject, or a combination of these and other pieces of information to identify the item in the catalog." (Wildemuth & O'Neill, 1995; p.265)

> "…a specific work which he knows to exist - possibly one *with which he has had previous contact*." (Swanson et al., 1968; p.1)

> "…a situation in which a user is trying to find an item previously read, and consequently in which the *user's memory of the item is of primary importance"* (Allen, 1989; p.247)

> "Research has shown that one of the most common searches done by our users is a *known-item* search (actually, a search for a particular work, or a *known-work* search)." (Yee & Layne, 1998; p.74)

> (All quoted in Lee et al., 2007)

Lee et al. highlight the uncertainties and assumptions inherent in these definitions. They note, for example, that inherent in most definitions is a "requirement that the user is searching for a "known" object or an object "known to exist", and argue that the nature of "knowing" is complex in this context. They give the example of a student seeking to find an introductory textbook on differential equations; the student might reasonably be said to know that such an item exists, even if they have not previously encountered it, or have particular title and author in mind. More strikingly, some definitions (e.g. Swanson et al.) require that "the user has a distinctly close relationship to the object sought, one closer than simply knowing it exists" (Lee et al., 2007: 5). Lee et al. question whether this should reasonably be considered necessary. The relationship between known-item and subject searches is also perhaps more complex than often assumed. A number of authors have noted a potential contradiction between overall user intent, and their search activity. For example, a known-item search might serve as a first step in a subject search, offering a user a starting point for their wider search. (Lewis, 1987; Brinkley & Burke, 1995; Hancock-Beaulieu, 1990). Alternatively the user might access a known-item in order to gather some information relating to a topic – information that is not uniquely related to the item itself.

Whilst acknowledging the nuances inherent in arguments advanced by Lee et al., this study will adapt Slone's definition of a known-item search, and define it as an interaction with the system wherein the searcher is seeking to locate in the catalogue the record of a specific item, about which some data is known. This is contrasted with an unknown-item search, which we define as an interaction with the system where the searcher is seeking to locate in the catalogue one or more items that offer some potential utility, without knowing the specific items in advance.

2.5 Recommender Systems

2.5.1 Introduction

The preceding sections have reviewed the literature relating to union catalogues, library catalogues in general, and information seeking behaviour. From this review it is apparent that in order to support a diverse range of users, search-tasks, and modes of behaviour, library catalogues must seek to offer a range of features, one of which is potentially recommendations. This section turns to the field of recommender systems, and attempts

to provide a general overview of current research into recommendation generation, presentation and use.

While RS uses and applications are undoubtedly diverse, the recommender problem can essentially be reduced to the question of how best to forecast a rating value for an item that a user has not seen (Adomavicius & Tuzhilin, 2005). A number of different technical approaches to this problem have emerged, and there follows an overview of the major types of recommender systems, and their hybrids. We will then examine current developments in some of the key areas of recommender systems research.

2.5.1.1 Collaborative Filtering

Different recommender system models can best be distinguished by the differing forms of data used as the basis of the quality measure. A collaborative filtering (CF) system "models the social process of asking a friend for a recommendation, making suggestions for some target user u from the items that are liked by users similar to u (user-based) or from items that have received similar ratings to the items that u likes (item-based)" (O'Donovan et al., 2008: 1085). This approach emerged as a means to aid users with the exploration of informal text repositories (for example discussion lists), for which traditional information retrieval methods produced sub-optimal results (Schafer et al., 2007). The TAPESTRY system, developed at Xerox in the early 1990s, allowed users to annotate items in a message database, with the annotations visible as a guide to future users. As the power and value of utilising user generated ratings emerged, more sophisticated algorithms for manipulating the ratings data were developed, most of which can be classified into one of two groups. Memory-based algorithms use the entire set of previously acquired ratings and user data to calculate their recommendations, while model-based methods use this dataset to develop a prediction model that can be used for future recommendation (Prekopcsák, 2007; Su & Khoshgoftaar, 2009). While it is beyond the scope of this review to undertake a technical analysis of specific algorithm methods, it should be noted that a wide range of approaches have been utilised, including similarity/neighbourhood based approaches , personality diagnosis, Bayesian networks, clustering models, and matrix factorization techniques (Takacs et al., 2009).

CF recommender systems have a number of advantages for users. They offer a rich means of discovering new items, obtaining advice about a selected item, and connecting with other like-minded users. Research also indicates that CF systems tend to offer greater

recommendation diversity than other RS models, presenting users with unfamiliar but potentially valuable niche material (Burke, 2007). There are however some significant problems associated with CF systems (see Table 2-2).

Table 2-2 - Problems associated with Collaborative Filtering Recommender Systems

Problem	Description
New User	CF systems have little data about new users, making the identification of similar users problematic.
New Item	Some CF systems have no way of incorporating new items, or those with few if any ratings, into their recommendations.
Sparsity	Most systems have a relatively small number of ratings compared to the total number of items. Matching users based on a small similarity weighting can lead to inaccurate predictions.
Scalability	Computational requirements grow quickly as the user base expands.
Privacy	Since CF systems rely on information about their users, they face challenges securing and safeguarding that data.
Grey Sheep	Users whose tastes are unusual or particularly varied are difficult to match with similar users.
Security	CF systems can be vulnerable to 'Shilling' attacks – the creation of fake profiles in order to artificially affect item ratings.
Explainability	It can be difficult to adequately explain to users why they have been recommended an item.
Synonymy	Variability in descriptive terms associated with an item can adversely affect some CF systems.
Stability vs. Plasticity	Once a user's profile has been established it can be difficult for the system to adapt to changing tastes or needs. Systems can incorporate a temporal element to discriminate against older items, but this has the potential to damage recommendation stability

(Based on Adomavicius & Tuzhilin, 2005; Lekakos & Giaglis, 2006; Su & Khoshgoftaar, 2009; Massa & Avesani, 2007; Li &Murata, 2010; Resnick & Varian; 1997, and Burke, 2007)

It should be noted that not all of these issues are unique to CF systems – for example the New User and Gray Sheep problems are common to other models. Nor should it be assumed that these problems are insurmountable, as the examination of current RS research found below will show. The number and range of the issues though do explain why, as a problem rich area, interest in RS research is high.

2.5.1.2 Content Based

Content-based recommender systems can be linked closely with traditional information retrieval research, and approach recommendation as a user-specific classification problem (Burke 2007). However whereas information retrieval systems such as search engines require the active participation of the user to formulate and refine queries, content-based recommender systems facilitate the passive presentation of items. CB models create a profile of a user drawn from their previous interactions with the system, and match that profile against static data held about items. This is most commonly done using a relational database approach, although some systems working with textual data will utilise keyword information retrieval techniques. The effectiveness of CB systems relies on the richness of data held about content, and the amount of information it can obtain about the user (O'Donovan & Smyth, 2005). This content metadata can be problematic, since it must come either in machine readable form or be manually entered, and often requires significant maintenance. Since knowledge about the user is limited solely to data acquired directly from the user, pure content based systems are most likely to recommend items similar to those preferred by the user in the past, and therefore offer less opportunity for diversity and serendipitous discovery than collaborative methods.

2.5.1.3 Others

Knowledge-Based (KB)

As with Content Based systems, knowledge-based recommender systems utilise information about users and content to generate recommendations. What characterises knowledge-based systems is best represented as a third layer of data representing domain knowledge, which allows the system to infer functional links between the user's needs and items that might fulfil them (Resnick& Varian, 1997). In this sense KB systems are particularly adaptive to context, since they attempt to tailor recommendations to a discrete scenario (Li &Murata, 2010). In order to ascertain user need, KB systems will often require

explicit action on behalf of the user (for example selecting preferences). For this reason such recommenders are often referred to as "conversational systems" (Burke, 2000). Since preferences are elicited at the point of need, it is not necessary for the system to hold large amounts of historical user data to provide accurate results, and research has shown that the knowledge layer of the system does not have to be prohibitively large to produce accurate results (Burke, 1999). As well as mitigating common RS issues such as the new user problem, these techniques have proved particularly useful in conjunction with other methods.

Utility-Based

Utility-based systems make recommendations based on a computation of utility. While in some case this might equate to the most practically useful item for the user, the utility to other parties can also be incorporated (e.g. the profitability of items might influence recommendations on an e-commerce site) (Shani et al., 2005). As with knowledge-based systems, utility-based recommenders do not attempt to build user models over time, but instead build suggestions based on a particular need (Chen et al., 2008). Clearly the central issue with utility-based recommenders is how the utility function is derived (Resnick& Varian, 1997).

Demographic

Demographic recommender systems attempt to categorise users according to their personal attributes, and then aim to make recommendations according to extrapolated demographic classes (Resnick& Varian, 1997). While it is generally accepted that demographic data alone fail to offer sophisticated enough personalisation for high quality recommendations, it has been suggested that it has particular value in overcoming the new user problem in hybrid systems since users can be linked by factors other than previously rated items (Lekakos & Giaglis, 2006; Burke, 2002).

2.5.1.4 Hybrid Recommender System

As the name implies, hybrid recommender systems are those which utilise at least two of the models outlined above in order to generate recommendations. Adomavicius & Tuzhilin (2005) identify four ways in which hybrid recommenders can be constructed:

1) Implement CB and CF methods separately, and combine the results

2) Use a predominantly collaborative approach with some content based input

3) Use a predominantly content-based approach with some collaborative input

4) Construct a general unifying model

Table 2-3 - Hybrid Types (Resnick & Varian 1997)

Type	Description
Weighted	An items predicted rating is calculated from the results of all available methods
Switching	The system switches between recommendation methods depending on the item /user under consideration
Mixed	Recommendations from multiple methods are presented simultaneously
Feature Combination	Use collaborative information as an additional data-set to be utilised by a CBH method.
Cascade	Recommendation methods are used in sequence
Feature Augmentation	One technique is used to derive data to augment a second technique
Meta-level	A model generated by one technique is input into another technique

Resnick & Varian (1997) go further, identifying seven potential hybrid types (see Table 2-3). The anticipated advantages of hybrid methods are twofold. While they do not alleviate all of the issues concerned with individual methods, the combination of methods can help address data sparsity and ramp up problems. They also fundamentally have the potential to offer improved recommendations, if skilfully implemented, by correctly balancing different approaches. Given the seven types of hybrid defined above, and the five basic recommendation techniques available, a huge range of possible hybrid systems can be envisioned. Burke (2007) characterises 53 hybrid systems, and evaluates 41 of these in a laboratory setting, identifying significant disparities in recommendation type between hybrid models and concluding that system designers must carefully tailor hybrid type to system function. His work also revealed the efficiency of knowledge-based systems as a

contributory or secondary component to hybrid systems, and identified cascade hybrids as an underused but potentially valuable model. As has been noted, attempts to develop the recommender systems field will certainly utilise hybrid systems (Yager, 2003).

2.5.1.5 Item-level recommendations

The recommender systems literature overwhelmingly focuses on the provision of personalised recommendations (Ricci et al., 2011), and the preceding discussion of recommender types reflect that. It should be noted however that recommendations need not necessarily be personalized. Whilst few would go as far as Demiriz (2004), who suggests that keyword search interfaces on e-commerce sites can be considered simple non-personalised recommenders, systems such as Amazon's "customers who bought this also bought" are commonly cited in early recommender systems papers as examples of recommender systems, albeit ones in which knowledge of the user is limited to their viewing a particular product (Schafer et al., 1999). Such systems are however effective, with research showing they significantly boost sales when implemented on e-commerce sites (Pathak et al., 2010; Chen et al., 2004).

While there is little else in the literature relating specifically to item-level recommendations, this is perhaps an issue of semantics. Recommendation personalisation is perhaps best viewed as a spectrum, with systems offering zero personalisation at one end (an example here might be some list of editor's picks, or top selling items, which appear unchanged to all users accessing the system), and totally bespoke and completely adaptive systems at the other (one could argue whether such a system could ever be more than hypothetical). It might then be possible to place any RS somewhere on this spectrum, depending on the extent to which the level of detail inherent in the user profile influences the recommended items. Indeed we might take this further, and locate different users' interactions with the same system at different points on the spectrum, depending on the extent to which their profile is more or less detailed, and therefore the recommendations are more or less accurately personalised. While recommendations such as Amazon's "Customers also bought" are not founded on a user profile of the type that exercises the majority of systems at the centre of RS research, it can be argued that the system is in some sense presenting recommendations based on the behaviour of the user – i.e. their decision to access the item-level page for a particular product. Since such systems are usually dynamic, we are perhaps justified in placing them somewhere to the right of the

completely non-personalised system. Thus for the purposes of this thesis, the term "item-level recommendation", rather than "non-personalised recommendation", will be used to describe systems that employ RS techniques to present recommendations to users at an item level following some explicit action on behalf of the user (e.g. clicking a link to the record page).

2.5.1.6 Feedback / Data Acquisition

For CF and CB recommender systems to predict appropriately, it is essential that they are able to draw on sufficient data to accurately model the user – indeed for most systems it can be said that the greater the amount of data, the better the recommendations are likely to be (Schafer et al., 2007). Methods for acquiring data can be broadly categorised as 'implicit' or 'explicit' (or sometimes 'extensional' or 'intentional') (Yager 2003). Implicit methods gather data from standard engagement with the system – for example purchasing an item or withdrawing a book – and are not noticed by the user. Explicit methods are those requiring the active participation of the user – for example rating a movie (Prekopcsák 2007). Implicit methods have the advantage of making no demands on the user, but the relationship of the data gathered with optimal ratings may be imprecise – for example a user may purchase an item as a gift for someone with hugely different taste. One area of research has examined the role of temporal information – the time spent viewing an item – as a means of inferring ratings. Parsons et al., (2004) note that a correlation between viewing time and preference has been broadly established in other fields, and such a method has proven effective as a means of augmenting e-commerce recommender systems (Lee et al., 2008). Other web usage data (for example click-through rates and query log analysis) has also been successfully applied to such systems (Cho et al., 2002).

Explicit methods are potentially much more valuable to recommender systems, since they usually represent an unambiguous statement of preference (Chen & Pu, 2007). The most common type of explicit data collection comes in the form of user ratings. These can be either unary ("good" or "don't know"), binary ("good" or "bad"), or integer based (a Likert numeric scale) (Schafer et al., 2007). Since users are generally assumed to avoid or seek to reduce cognitive effort, collecting such data was assumed to be problematic, since the user must be persuaded to actively participate in the process (Gretzel & Fesenmaier, 2005). This has led some to observe that the drive for increased recommendation accuracy must be

balanced with the levels of user effort required (Rashidet al., 2008). Recent research, however, has indicated that users are frequently willing to engage with ratings processes – motivated by the opportunity to improve their profile (and therefore recommendations they receive), express themselves, and help or influence others (Herlocker et al., 2004). If these factors can be considered intrinsic motivation, some systems also attempt extrinsic motivation methods – the offer of some tangible reward (e.g. store credit) to the user for carrying out ratings (Farzan & Brusilovsky, 2011). Research has also examined the ratings process itself, establishing that users prefer more detailed or "fine-grained" scales (Cosley et al., 2003), while Gretzel & Fesenmaier (2005) suggest that the structure, content and layout of a ratings process can substantially influence a user's perception of subsequent recommendations. Sinha & Swearingen (2002) posit that systems should seek to elicit ratings at particularly opportune moments, suggesting a 'conversational and collaborative' model by which additional ratings requests can be triggered at moments when the user is assumed to have particular motivation to do so (for example when the user is surprised by a particularly high or low rating). Chen & Pu (2007) expand this conversational element to create an ongoing dialogue with the user to refine and develop ratings and ratings ranking. The development of unobtrusive yet comprehensive interfaces for ratings collection has been identified as a crucial area for future research in recommender systems (Perugini et al., 2004).

2.5.2   Trust

Work in other fields has categorised different forms of trust (Abdul-Rahman & Hailes, 1997). Two of these categories can be found to influence current recommender systems research. *System / impersonal trust* relates to the trust users have in a system. Despite advances in recommender system technology, it is noticeable that users are selective about the extent to which automated recommendations influence behaviour. To use O'Donovan & Smyth's (2005) comparison, the risk associated with following a movie recommendation is considerably less than committing to an expensive holiday suggested by a vacation recommender. The extent to which users trust a rating is also influenced by their understanding of why and how particular recommendations have been given. Methods of overcoming this problem are discussed in section 2.5.3. It is also true that users are aware that the ratings that drive many recommender systems are not immune from manipulation – so called Shilling attacks by self-motivated rogue users (Schafer et al., 2007).

This leads to the second category of trust relating to recommender systems. *Context-specific interpersonal trust* describes an occasion when a user has to put their trust in another person, but only for that specific instance. Recent research suggests that collaborative filtering systems could augment user connections by incorporating a trust measure. A number of researchers have shown that predictions are improved by excluding from recommendation calculations the ratings of users who have been deemed somehow untrustworthy or unreliable, and suggest a variety of methods for developing a computational model of trust (O'Donovan & Smyth, 2005; Massa & Avesani, 2007; Victor et al., 2009). Massa & Bhattacharjee (2004) offer a more straightforward solution, suggesting that a simple 'friend' function allows a system to infer trust ratings on a large proportion of the user population, while Ziegler (2004) describes how the integration of existing trust networks born out of social networking environments have the potential to rapidly improve the quality of recommendations to new users.

2.5.3   Explanation

As noted by McSherry (2005), "the importance of intelligent systems having the ability to explain their reasoning is well recognised in domains such as medical decision making and intelligent tutoring" (179). The early recommender system has been characterised as a 'black box' offering the user no information about how suggestions were computed (Schafer et al., 2007). The result is a system more akin to a search engine, and far removed from the type of word-of-mouth recommendations that users are accustomed to (Bonhard & Sasse, 2006). Recent research has shown that users interact far more effectively with systems that clearly explain the relationship between ratings and recommendations. This was most effectively demonstrated by Sinha & Swearingen's findings that users wanted an explanation even for items they were recommended and already liked (2002). Herlocker et al., (2000) have also successfully demonstrated four key benefits to incorporating an explanation function into a recommender system: justification (helping users understand why a recommendation has been made), user involvement (making the user feel more involved in the process), education (so the user better understands the scope of the system) and acceptance (greater confidence in recommendations, thereby mitigating trust issues discussed above). They also identify that explanations offer a means of handling recommendation errors – be they process or data based.

While explanations are beneficial in principle, it has also been shown that the form of the explanation is important (Tintarev & Masthoff, 2007). Some explanation methods trialled by McSherry (2005) were found to have a negative effect - particularly those of a technical nature. By contrast the second most effective form of explanation consisted solely of a statement of past performance – e.g. 'the system has predicted correctly for you 80% of the time'. Other research has experimented with a confidence display – essentially a graphic representation of the system's faith in the recommendation (McNee et al., 2003). McSherry has also suggested a new model combining explanation and rating elicitation through a conversational case-based reasoning system in which an ongoing dialogue between human and machine enables the system and user to refine both ratings and recommendations.

2.5.4 Context

Context has been defined in the recommender systems field as "the discourse which informs the users' current behaviour in the system - their current requirements, their motivation, their previous experience, their preferences and the knowledge available to them" (Hayeset al., 2002, 4). Numerous studies of information behaviour have identified context as a critical factor in the understanding of a user's interaction with an information system, and it has also been observed that context is instrumental in consumer decision making (Adomavicious & Tuzhilin, 2008). It has been clearly noted that an inherent weakness in most existing recommender systems is their inability to factor context into the recommendation process, and that addressing this failing is a key challenge for recommender systems research (Adomavicius & Tuzhilin, 2005). Herlocker & Konstan (2001) have proposed a task-focused approach. Operating in tandem with a traditional recommender system, the task-focused system would attempt to match recommendations to a particular context. The creation of a task-profile (either implicitly or explicitly determined) represented by a few task-relevant items allows the system to query a pre-populated item association database, and call up potentially relevant items. These are then ranked according to the users profile based on the standard interest ratings database. Testing this system on the MovieLens dataset yielded positive results, although the system fared poorly when presented with complex contexts.

Adomavicious & Tuzhilin (2008) also identify a potential approach to using contextual information, which could be determined by a user's selection from pre-defined domain

specific options in two ways; querying and searching, or preference elicitation. Focusing on the latter approach, they identified three further categories - pre-filtering (where contextual information is used to filter the dataset, after which a standard recommendation can be processed), post-filtering (where the recommendations are generated as normal, and the results filtered according to the contextual data), and contextual modelling (where contextual factors are incorporated into the recommender technique itself). They also acknowledge that while several small-scale context aware systems have been developed, none have proved robust enough to influence the wider RS field.

2.5.5   Serendipity and Recommendation Diversity

Serendipity refers to "the action of, or aptitude for, encountering relevant information by accident" (Case, 2012: 390). It is important to note that for something to be serendipitous both facets of the definition must hold true; the discovery should be accidental, and the information relevant. On the grandest scale serendipity has been seen to play a significant role in the progression of human knowledge and understanding, with numerous scientific and technical advancements resulting from apparently serendipitous discoveries.[1] More prosaically, serendipity can be viewed as "an important component of the complex phenomenon that is information seeking" and ". . .a method for achieving breadth and identifying information or sources from unknown or partially unknown directions" (Foster, 2006: 157; Foster & Ford, 2003: 337). We might also link the notion of serendipity to Wilson's Passive Search construct, in the sense that it signifies an instance of unplanned information gathering (Wilson, 1996).

It has been argued that serendipitous discovery is under threat from the increasing prevalence of digital systems as information intermediators, primarily because ever-more effective search engine algorithms and personalised filters are so adept at delivering content which matches our perceived need (Gup, 1997; Pariser; 2011). As McCay-Peet notes, "information-rich environments such as digital libraries primarily support targeted search interactions" (2011: 398). To introduce the type of novel and diverse material that might encourage serendipity to an environment where the user might expect and value only results relevant to their query, is to risk leaving the user dissatisfied (Ge et al., 2010).

---

[1] Oft cited examples include Fleming's discovery of penicillin, Nobel's invention of Gelignite, Pasteur's discovery of molecular chirality, Columbus reaching America, not to mention the invention of (among many other things) Velcro, Viagra, and vulcanised rubber.

Thus the challenge for researchers is to develop systems that support chance discovery without impacting everyday usability. In considering how information systems can facilitate this, it is first essential to acknowledge that any particular serendipitous encounter relies not only on the system presenting a novel item, but also on the user recognising that item as being potentially relevant or interesting. Thus serendipity requires a range of precipitating conditions relating to the user's emotional, intellectual and motivational characteristics (Heinström, 2006; Roberts, 1996; Erdelez, 1999). In addition, two paradoxes lie at the heart of understanding serendipity in a digital information seeking context. First, although the act of serendipitous discovery is considered (by definition) beneficial, at the same time it is "elusive and unpredictable", and therefore impossible to utilise as a "conscious information seeking strategy" (Foster & Ford, 2003: 321) – in other words it is both useful and unusable. Second, any system that is successfully engineered to induce serendipity might then no longer be considered serendipitous, since "the system is likely to reduce the amount of chance and insight involved in the serendipitous discovery simply by helping to encourage it" (Makri et al., 2011: 728).

Table 2-4: Core elements of support for serendipity in a digital environment (McCay-Peet & Toms, 2011)

Factor	System
1. Enabled connections	The system allows the user to make new and useful connections that change how the user conceptualises the task at hand
2. Encountered the unexpected	The system presents content that is novel or unexpected
3. Presented variety	The system provides a variety of content, and offers the user the means to explore it
4. Triggered divergence	The system calls attention to interesting or surprising content through visual cues
5. Induced curiosity	The system both ignites and stokes the curiosity of the user

McCay-Peet & Toms (2011) offer perhaps the most rigorous analysis of how digital library environments can encourage serendipitous discovery (see Table 2-4). Their large-scale study invited 124 participants to use the wikiSearch system in a simulated free browsing scenario, after which a detailed questionnaire examining system attributes was administered. The results suggest five core factors that are crucial to maximising the

chance of serendipitous encounters: Enabled connections, Encountered the unexpected, Presented Variety, Triggered Divergence, and Induced Curiosity. It is significant that three of the five factors (1-3) are directly related to content diversity.

An acceptance that serendipity is a desirable consequence of online information seeking, and that diversity of content is key driver of serendipitous discovery, raises significant issues for recommender systems research. Not least, it exposes an inherent conflict within Recommender systems design and evaluation; that whilst a common proposed benefit of recommender systems is as a means of exposing users to new and interesting content (Herlocker et al., 2004), improvements in algorithms are generally directed towards recommending objects based on user or item similarity (Zhoua et al., 2010). Recommender systems essentially assume that the most interesting and useful items are those with the highest calculated rating or utility value (McNee et al., 2006). The result can be a sort of positive feedback loop, with users being recommended ever-more specialised content at the expense of the novel and unexpected (Fleder & Hosanager, 2007; Loed & Panagos, 2011). This so-called "similarity problem" is especially acute in content-based system, since the system is limited to recommending items that match the preferences implied by that users profile (Iaquinta et al., 2008). While collaborative filtering techniques mitigate the issue somewhat, it is still acknowledged that accurate recommendations are most likely to be based on the preferences of other users most like us, whereas the most valuable recommendations might occasionally be found in the preferences of "weak ties" - users with whom we share little in common (Zhoua et al., 2010).

It is accepted therefore that recommender systems should present a more diverse range of items than would result from a ranking of best fit items (McSherry, 2002; Smyth & McClave, 2001; Herlocker et al., 2004; Swearingen & Singh, 2001; Hu & Pu, 2011).While this is likely to facilitate serendipity, studies have shown a range of benefits related to diverse recommendations sets. Variety appears to reduce choice difficulty or overload (Willemsen et al., 2011; Hu & Pu 2011; Bollen, 2010), and deliberately diversified recommendation sets can even be perceived as more accurate (Konstan et al., 2006). This latter and perhaps counter-intuitive finding is explained by the tendency of users to judge the accuracy of recommendation sets rather than individual recommendations; diverse recommender sets perhaps more accurately reflect the diverse tastes of users (Knijnenburg et al., 2012). Another advantage of recommendation diversity is the potential for the resulting list to meet a broader range of possible user preferences and contexts (Price, 2005), or help users

easily narrow down their requirements (Hu & Pu, 2011: 43). There is also the possibility, as noted by Ge et al. (2000), that a small number of recommendations which are obviously different to the majority of the recommendation set will naturally arouse curiosity on the part of the user, a possibility that maps directly onto the *Induced Curiosity* factor identified by McCay-Peet & Toms.

2.5.6 Interface

How recommendations are presented to users, both in terms of the information presented and the visual form of that presentation, is a key consideration for recommender systems designers (Herlocker et al., 2004). Recommendations are typically presented in a rank ordered list, often with some form of additional information explaining why the item has been recommended (Pu & Chen, 2007), and research suggests that interface considerations can have a greater effect on user experience than algorithm performance (Knijnenburg et al., 2012). Given the importance of the user interface, it is perhaps somewhat surprising that with a few notable exceptions, interface design has remained relatively traditional (Chen 2011). Within this existing paradigm of recommendation layout, several studies have provided some basic guidelines for presentation. In the field of e-commerce, recommended products are best shown alongside the main product being viewed (Ozok et al., 2010), with users preferring to see some general information about the item being recommended (Swearingen & Sinha, 2001).

One area of recommendation interface design that has attracted particular attention is the elicitation of feedback and ratings. A common concern here is the type and granularity of the rating scale. Although it has been shown that users rate relatively consistently using different scales (Cosley et al., 2003), the process of relating personal preferences to quantitative scale measures can be problematic for users (Nguyen et al., 2013). Pommeranz et al. stress the importance of a user-centric design process for developing rating elicitation interfaces, claiming that better understanding "the mental models of users' preferences" is essential in creating systems that "support the process of human preference construction" (Pommeranz et al., 2012: 390). Their study shows users value designs that provide visual feedback, and support the exploration of the system during the rating process. Similarly Webb & Kerne (2008) suggest a "fluid in-context interface for interest expression", arguing rating sliders should be carefully integrated within the wider system (as opposed to being stand-alone pop-ups), thereby reducing effort and improving feedback. Other work

suggests that the incorporation of personalized tags can help users better recall the characteristics of items being rated, and that presenting users with other items they have rated similarly increases the consistency and quality of ratings (Nguyen et al., 2013). McNee et al (2003) found that engineering a more detailed, and therefore more cognitively demanding, ratings interface for a movie recommender resulted in greater user retention, even though there was no resulting improvement in recommendation accuracy.

While a list remains the most common form of recommendation interface, other layouts are often utilised. Chen & Tsoi (2011) compare three typical designs: list, grid and pie (see Figure 2-7). An analysis of click-through data during a controlled experiment with
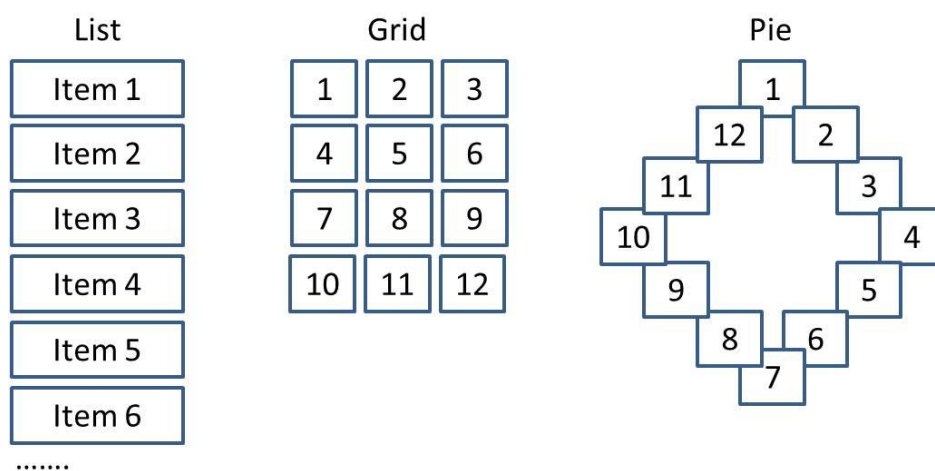
Figure 2-7: Common Recommendation set displays (Chen & Tsoi, 2011)

standardised recommendation sets showed that the grid and pie displays encouraged users to click on a wider range of recommendations, with clicks most evenly distributed in the pie interface. By incorporating qualitative data gathered in post-task interviews, they suggest that the pie layout was most popular with users, with that interface improving perceived recommendation quality and decision confidence. A more elaborate technique is suggested in a series of papers by Chen & Pu (2008, 2010 & 2011). They propose and develop an *Organization Interface* "to compute and categorize recommended products, and use the category title (e.g. "these products have cheaper price and longer battery life, but slower processor speed and heavier weight") as the explanation of multiple products." (2008: 76). As well as finding that this interface improved user perception of recommendation quality compared to a traditional list view, they also showed that it improved decision making efficiency (2008). Additional studies using eye-tracking equipment demonstrated that users adopted different techniques when interacting with list and organisation layouts, with the

latter seeming to encourage users to view a greater number of more diverse recommendations (2010). They also suggest that the different recommendation categories are best displayed in a quadrant (i.e. with the first and second categories presented on the same horizontal level, with the third and fourth categories underneath) rather than vertical arrangement (2011).

Figure 2-8: Example of the TopicLens Recommendation Interface (Devendorf et al. 2012)

Interfaces that allow users to interact with recommendation sets are increasingly seen as a means of boosting user engagement with recommendations, thereby potentially increasing their effectiveness (Verbert et al., 2013). While In its simplest form that might mean including filters allowing users to customize recommendation sets (Swearingen & Sinha, 2001), others have seen the potential of creating interactive visualisations. Devendorf et al. (2012) describe a visualisation model created for the TopicLens, an exploration and recommendation tool allowing users to interact with large datasets (Figure 2-8). Using a wheel metaphor to display recommendations, the interface allows users to explore recommended content, with the interface dynamically adjusting to the desired level of granularity. A comparable, if less visually impressive interface is proposed by Verbert et (2013), who demonstrate how the underlying data powering academic talk recommendation on the Conference Navigator site can be presented in an interactive way in the *TalkExplorer* system (Figure 2-9). Users are presented with the underlying networks of users, bookmarks and ratings, with the facility to explore connections. In evaluating the

prototype system they found the system greatly increased users understanding of and trust in the recommendations, and that users valued the ability to actively explore recommendations.
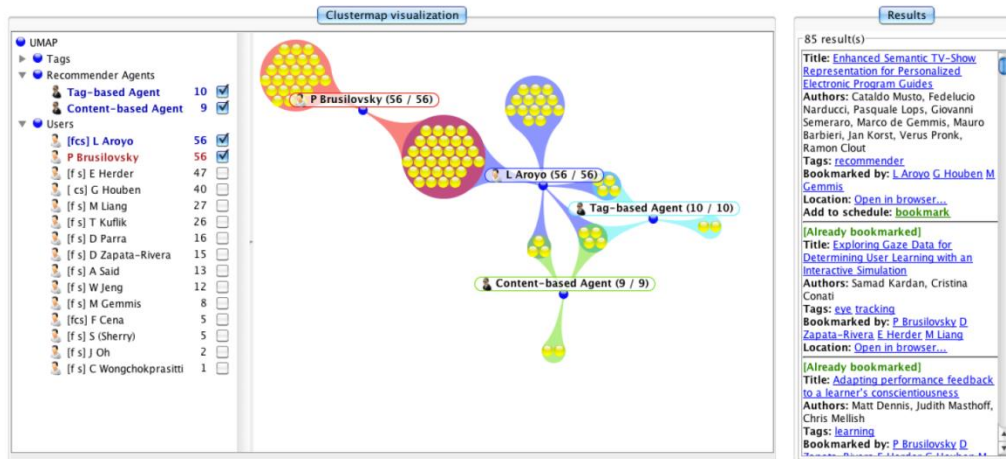
Figure 2-9: TalkExplorer Interface (Verbert et al. 2013)

2.5.7   Evaluation and User Experience

The issue of how best to evaluate the output of recommender systems has been identified as a key problem. While it is agreed that evaluation is critical to the success of an information system, it has been noted that the absence of a universally accepted method has limited the analysis of many recommendation techniques (Perugini et al., 2004). Most evaluations of recommender systems are conducted off-line and are based on determining system accuracy (Fleder & Hosanagar, 2007; Herlocker et al., 2004). Primarily this means predictive accuracy – the extent to which a system can accurately predict a user's rating. Typically this is assessed using a 'leave-n-out' approach, whereby a user generated rating is withheld, and the system asked to predict its value (McNee et al., 2006; Schafer et al., 2007). Accuracy is then calculated using mean absolute error (MAE), which identifies the variation between predicted and actual rating. Other techniques include classification accuracy metrics (the rate at which systems make correct recommendations), precision and recall measures, rank accuracy metrics (comparing a system's ranking of recommendations to that based on actual ratings), error rate measures (which determine the rate at which recommendations are made in error), prediction-rating correlation, half-life utility metric (based on the difference between the users rating and a default rating for the item), and receiver operator characteristic (ROC) curves (measuring how efficiently a system can distinguish between relevant and irrelevant data) (Herlocker et al., 2004, del Olmo &

60

Gaudioso 2008, Schein et al., 2005). Other measures of algorithm accuracy used in live system environments include click logs (Konstan and Riedl 2012), consumption volume (Pedersen 2000), and decision time (Haubl et al. 2004).

A number of problems have been identified with the predominance of accuracy metrics. The range of accuracy metrics themselves makes it difficult to compare systems that have been assessed using different measures. Even more crucially, accuracy metrics fail in a number of crucial ways to fully capture and assess the actual user experience (Ziegler et al., 2005). A number of scenarios have been highlighted to illustrate this: accuracy metrics, for example, have no way of penalising similarity even though a user will not necessarily appreciate a system that recommends only within a narrow range of content (McNee et al., 2006a & 2006b). Likewise studies have shown that users are less forgiving of erroneous recommendations that appear highly ranked – yet most accuracy metrics have no way of assessing this (Schafer et al., 2007). As has been noted, the increasing prevalence of recommender systems in diverse information seeking environments has exacerbated these issues (Konstan et al., 2006).It seems clear therefore that offline accuracy metrics can only partially evaluate a recommender system (Tintarev & Masthoff, 2007). Other frequently used and non-labour-intensive methods – for example click through rates and other web metrics – have also been shown to be flawed (Zheng et al., 2010; Pu & Chen, 2006).

 In light of these problems a number of researchers have attempted to develop alternative methods that better test a systems ability to meet user requirement. These are based more closely on the broader aims of recommender systems such as transparency, trust, scalability, persuasiveness and satisfaction – factors that do not necessarily correlate with accuracy (Tintarev & Masthoff, 2007; Drineas et al., 2002). Mirza et al.'s (2003) *Jumping Connections* model evaluates algorithms based on the number of people they connect, a potentially key metric also noted by Perugini et al., (2004), while McNee et al., (2006a) propose both a serendipity metric – that might attempt to judge the true value of variety within a recommendation set – and a means of assessing recommendation lists as a whole (thereby ensuring the evaluation identifies similarity issues). Evaluating recommendations in terms of list diversity and coverage is also proposed by Ge et al., (20010), while Oku & Hattori (2011) propose a means of calculating recommendation utility. Another approach has been suggested by del Olmos & Gaudioso (2008), who present a general framework in which each recommender system is divided into two sub-sections: a filter subsystem (which selects the items for recommendation) and a guide subsystem (which presents the

data to the user). This technique would allow designers to identify appropriate evaluation mechanisms for each part.

Others insist that to be truly effective, evaluations must directly involve users, a method familiar to many areas of Information Science research (Herlocker et al., 2004). This in turn has led researchers to think more holistically about user interactions with recommender systems, and develop models of user experience that might serve as a conceptual basis for evaluation (Knijnenburg et al., 2012). User experience (UX) has been defined as

> "a momentary, primarily evaluative feeling (good-bad) while interacting with a product or service. Good UX is the consequence of fulfilling the human needs for autonomy, competence, stimulation (self-oriented) through interacting with the product or service (i.e. hedonic quality)." (Hassenzahl, 2008: 12)

A number of attempts have been made to model recommender system user experience. Zins & Bauemfiend (2005) surveyed users of three recommender services (two related to the travel industry, and one for electronics) in order to better understand the factors influencing system satisfaction. They showed that perceptions of system trustworthiness, as well as browsing patterns and levels of system engagement, were heavily influenced by the individual characteristics of users. These factors were also shown to influence levels of satisfaction with the system, irrespective of objective algorithm performance. While this work represents an important and early shift towards user centred evaluations of recommender systems, it has been noted that the in largely ignoring aspects of system performance the model fails to address the complexities of human system interaction (Knijnenburg et al., 2012).

Another approach is that of Xiao & Benbasat (2007). Their conceptual model is based on a comprehensive review of the literature relating to e-commerce recommender systems, and attempts to identify the factors that influence to recommender system user experience, and the relationships between those factors. These factors might be grouped into three broad categories – System (e.g. algorithm type, interface, use case), User (e.g. familiarity with RS, product expertise, expectations), and Commercial Context (e.g. product type, provider credibility). The model describes how these factors influence the consumer decision making process, which itself influences user perceptions of the recommender system in terms of trust, usefulness, ease of use and general satisfaction. While the model does address the interaction of human and system features, it does not specifically attempt to integrate subjective aspects of the user experience. It should also be noted that the

model was developed as a means of summarising and conceptualising existing research on recommender systems use, rather than as a framework for practical evaluation.

McNee et al., (2006) propose a new field of research – Human Recommender Interaction (HRI), which they describe as "a framework and a methodology for understanding users, their tasks, and recommender algorithms using a common language" (1105). They suggest that a better understanding and means of categorizing the interactions a user has with a recommender system will eventually lead to systems that better match user needs. To facilitate this they present three pillars of HRI – Recommendation Dialogue, Recommender Personality and End User's Information Seeking Task, with each pillar consisting of a set of "aspects" (see Figure 2-10). McNee et al. claim that by selecting the most appropriate aspects from each pillar, a user's needs and expectations can be comprehensively categorised. It is important to note that while the model might be used for evaluative purposes, it was developed primarily as a tool for user-centric design. Thus the Three pillars are complemented by an Analytic Process Model, which seeks to represent the recommender system within a wider information seeking context, and as one side of an ongoing dialogue with the user. The authors suggest that by carefully analysing user requirements certain key HRI aspects might be identified. They envisage that these aspects might be mapped to key system metrics, which in turn might map to recommender algorithms. Despite the clear potential of such a process model, however, little work

Recommendation Dialogue	Recommender Personality	End User's Information Seeking Tasks
Correctness Transparency Saliency Serendipity Quantity Usefulness Spread Usability	Personalization Boldness Adaptability Trust / First Impressions Risk Taking / Aversion Affirmation Pigeonholing Freshness	Concreteness of Task Task Compromising Recommender Appropriateness Expectations of Recommender Usefulness Recommender Importance in Meeting Need

Figure 2-10: The Three Pillars of HRI (McNee et al. 2006)

algorithms. Despite the apparent potential of such a process model, however, little work appears have been done formalising the frameworks required for such mappings, and this author could find no examples in the literature of the process being utilised for an actual recommender system implementation. It should also be noted that McNee et al. provide

no formal methodology for the identification of the various "aspects" presented in the model.

Two further models have evaluation as a primary purpose for their creation. Perhaps the most comprehensive attempt at a universal recommender evaluation method comes from Pu et al. (2011). Their *ResQue* framework serves two purposes – first it identifies and groups the key constructs relating to the user experience of recommender systems, and second it suggests the relationships between these constructs (Figure 2-11). Using a survey of recommender system users, Pu et al. were able to both validate and refine the framework to incorporate the causal relationships between individual evaluation constructs. They found perceived quality of recommendation accuracy, novelty and diversity directly correlated with the perceived usefulness of the system, which in turn correlated with Trust and confidence in the system. These findings were used to develop an evaluation survey tool consisting of thirty-two questions exploring the constructs identified in the model.

Perceived System Qualities	User Beliefs	User Attitudes	Behavioural Intentions
Explanation	Transparency	Trust and confidence	Purchase intentions
Interaction adequacy	Control	Overall satisfaction	Use intentions
Recommendation accuracy	Usefulness		
Recommendation novelty	Ease of use		
Recommendation diversity			
Information sufficiency			
Interface adequacy			

Figure 2-11: ResQue Constructs (Pu et al., 2011)

Knijnenburg et al. (2012) similarly present and validate a framework to represent system and user interaction, which they suggest can be used as a guiding model for diverse RS evaluations. In comparison to previous models, theirs more clearly defines and incorporates contextual elements of the information seeking process. User Experience (EXP) and Interaction (INT) are seen to be influenced by not only Objective System Aspects

(OSA), but by a subjective perception of those system aspects (SSA), in conjunction with the Situational (SC) and Personal Characteristics (PC) of the interaction. In using field trials and controlled experiments to validate the framework, Knijnenburg et al. demonstrate that "subjective system aspects and experience variables are invaluable in explaining why and how the user experience of recommender systems comes about." (2012: 442). A clear strength of their model is its explicit acknowledgement of the existence of both objective and subjective system qualities, and its implicit placing of human recommender interaction within a wider information seeking context.
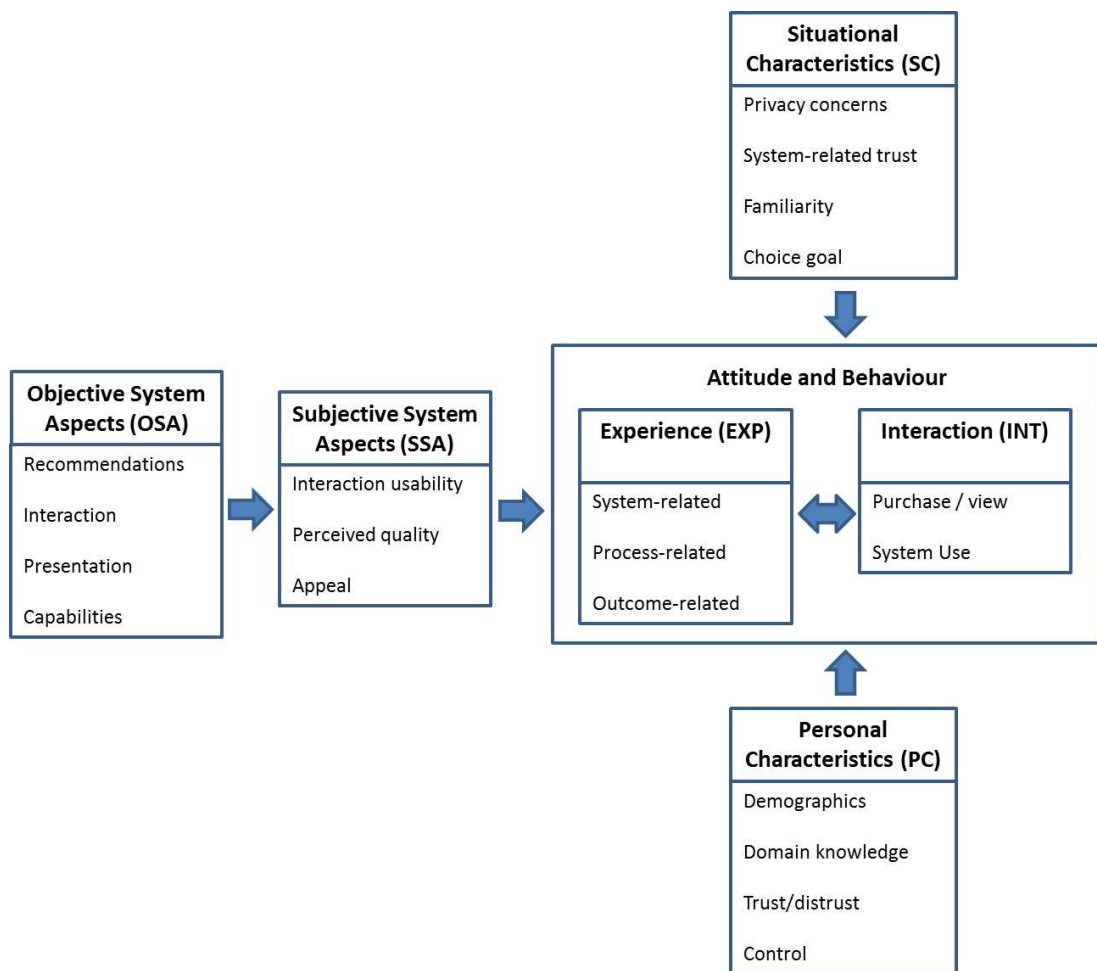
Figure 2-12: Recommender System User Experience Framework (Knijnenburg et al. 2012)

In presenting their models, Pu et al. (2011) and Knijnenburg et al. (2012) provide comprehensive reviews of other user-centred studies that have attempted to examine specific aspects of and factors effecting user experience. An examination of those studies, along with research not included in those papers, reveal some important details about the factors effecting user experience not yet covered in this chapter. Domain knowledge /

expertise has been shown to heavily influence both interaction behaviour and system satisfaction (Knijnenburg et al., 2012). In general experts are less likely to use recommender systems, and find them harder to use (Kamis & Davem, 2004). Hu & Pu (2010) show that the level of domain expertise impacts the perceived accuracy of recommendations, with experts rating accuracy lower than non-experts. They also suggest that users with a moderate level of domain expertise are likely to view the system as more effective than both expert and novice users. This can be seen as particularly significant given other research into satisfaction levels, which suggests that the perceived recommendation quality directly affects the level of user satisfaction with the product that is eventually purchased or consumed (Bharati & Chaudhury, 2004).   Castagnos et al. (2010) conducted a controlled user study to observe user interactions with a recommender system during a simulated perfume buying task, with eye-tracking and log data collected. They found that users made greatest use of recommendations at the start of the tasks, when their domain knowledge was weakest. In the early stages of tasks users were seen to be twice as likely to find products via recommendations as from searches. As the task progressed, and users became better acquainted with the products available, reliance on recommendations lessened.

While these models undoubtedly offer valuable insights into the way recommendations effect user experience, there nonetheless remains little work that addresses in more practical terms the ways in which recommendation effect user searching behaviour. It is hoped that this study can begin to address this gap in the context of library services.

2.5.8   Recommender Systems and Library Services

2.5.8.1 Introduction

In discussing the current theoretical and practical landscape of recommenders in a library context we can identify three key areas, and there follows a discussion of each. First we look at systems designed to recommend journal articles and papers, followed by recommender systems applied to digital libraries. Finally we examine some literature relating to recommendations in the library catalogue, and examine some practical implementations. Whilst clearly these broad areas are not necessarily mutually exclusive (for example some digital libraries hold journal articles, and catalogues themselves might

be considered a form of digital library), they nonetheless allow us to examine a variety of recommender techniques in differing contexts.

2.5.8.2 RS and Journal Articles

A significant portion of research relating to recommender systems in a library context is focused on the retrieval of journal articles. While to some degree this merely reflects the increasingly central role the journal article plays in scholarly discourse, the nature of the journal article itself – relatively short, highly focused, domain specific – lends itself particularly to the recommender system. While conventional information retrieval methods such as web search engines allow users to search with high precision and increasing levels of sophistication, they fail to adequately address one of the issues mentioned above – namely the exposure of potentially valuable papers from different domains (Vellino, 2010). This limitation is principally due to the heavy reliance on keyword, since terminology often varies across disciplines. These limitations are exacerbated in so called *emergent* systems – databases or libraries of machine harvested content (for example CiteSeer). Such systems are frequently found to contain items of highly variable quality (since they have no professional or expert human agent managing the acquisitions process), and therefore further expose the limitations of a keyword search strategy (Torres et al., 2004).

Two distinct approaches to generating journal recommendations have been proposed. The first of these relies on bibliographic citations as a means of evaluating an item's potential utility. In the language of recommender systems, 'the idea is to consider an article as a "user" and view articles that it cites to be the articles' "preferences"' (Vellino, 2010). Systems such as TechLens (Torres et al, 2004) and Synthese (Vellino & Zeber 2007) supply the user with recommendations though traditional collaborative filtering algorithms utilizing citation data in lieu of user generated ratings. Evaluation of the TechLens system in particular has revealed a broadly positive response to the recommendations, with a key finding that researchers are so keen to discover potentially interesting articles that they will judge the recommender system to be successful even if only one article out of five was relevant and novel (Konstan, 2006). Despite this positive assessment, problems with citation based recommendations remain to be conquered. The citation data itself can be difficult to obtain, relying on sophisticated automatic extraction or manual curation. More significantly perhaps, newly published articles lack sufficient citations to drive

recommendations, and hence are discriminated against in the recommendation process (Pohl et al., 2007).[2]

The second major approach to journal recommendation is based on usage or access data. Perhaps the most impressive example of such a system is Ex Libris's *bX* service, which mines SFX logs to identify articles downloaded by individual users during the same session, and uses this data to infer relationships between items (Vellino, 2010). Building on the work of Bollen & Sompel (2006), who developed a set of standards for the cross-institutional harvesting of such data, the bX system is able to channel intelligence from geographically disparate communities and diverse subject areas into potentially valuable recommendations. It has been suggested that recommender systems powered by access data generate more diverse and interesting recommendations – particularly for newly published papers (Pohl et al., 2007). Interestingly it has also been shown that the two methods – citation based and usage based – demonstrate a high level of complementarity; not only does testing show that they rarely provide the same recommendations, but they can rarely both provide recommendations related to a given particular article.

In a sense both methods are an attempt to overcome the problem of data-sparsity – a particularly pertinent issue given the small number of scholars relative to journal articles within a subject area. These users can be further divided into levels of "maturity", with experts (Professors, Researchers etc.) demonstrating considerably lower satisfaction rates than students when evaluating recommendations (Torres et al., 2004). One means of assuaging these users is through the development of more sophisticated user profiles, for example by applying concept-based profiling based on click histories (Kodakateri et al., 2009).

2.5.8.3 RS and Digital Libraries

As digital libraries become an increasingly common feature of the academic landscape, some have called for them to better reflect the expectations of their users who see them not simply as content repository but as virtual research environments – "community-based services which require personalized service offerings" (Vellino & Zeber, 2010). This requires digital libraries to become more adaptive to their users, proactively offering and

---

[2] While discussing recommender systems and citations, it is worth noting work done by He et al. (2010) on citation recommendation – essentially a system that can analyze the text of a paper during its composition, and suggest appropriate resource to cite. This might be regarded as the *non plus ultra* of context-specific recommendation.

tailoring information to their users (Avancini et al., 2005). Several researchers in the field have provided more specific visions of how recommender systems can play an integral role in improving the user experience. Renda & Straccia (2005) describe a digital library that offers not only sophisticated document retrieval but also a virtual space for collaborative working. The base algorithms of collaborative filtering might be used to establish networks of users, from which explicit human recommendations might evolve. Automated recommendations can also be seen as the end product of a functional chain incorporating information seeking, personalized information management, and collaboration (Avancini et al., 2005). Torres et al (2004) also envision a multifaceted environment, with recommendations tailored to particular user tasks, experience levels and even language.

The scope of these ambitions is due to the localized nature of the DL – at least in the sense that the system is working with a defined corpus. A consequence of this is that unlike a system such as bX, which operates by manipulating large amounts of basic access data, a DL can both obtain and manipulate more sophisticated usage data. A number of practical studies have examined how this data might be used to drive more useful recommendations, whether through multi-phase algorithm application (Chen & Chen 2006), continuous learning models (Kruschwitz et al., 2010), or hybrid content/collaborative methods (Huang et al., 2002).
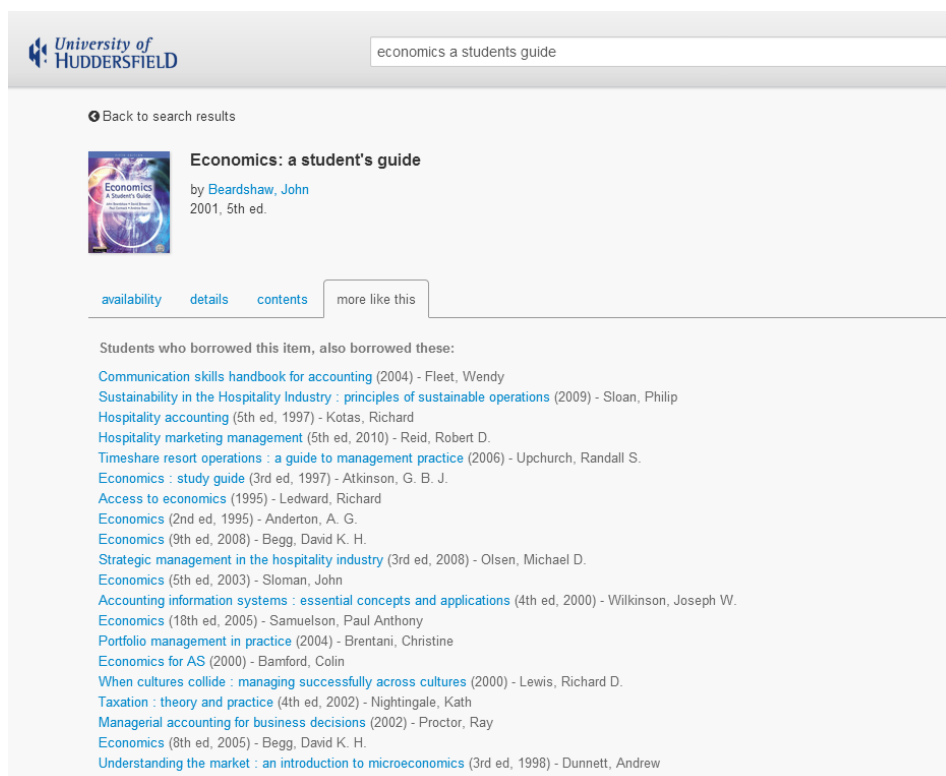
2.5.8.4 RS and the Online Catalogue

While the development of recommender systems for libraries has been slow, there exist a number of practical implementations which it is instructive to consider. Given its apparent relevance to this project, we can begin by discussing the Kindred Works project, an OCLC research project focused on the creation of a content-based recommendation system. The system recommends items based on similarity to a given item, subject headings, classification numbers, and genre terms. A prototype of the system is available at a publicly accessible URL (http://experimental.worldcat.org/kindredworks/), and the service is also available as an API for integration with existing library catalogues and services. This project is relatively recent (2013), and as yet has not been formally evaluated. Its primary aim appears to be supporting the browsing and discovery of items for leisure reading.

Academic institutions have also experimented with recommender systems implementations. The University of Huddersfield has developed an in-house system (see Figure 2-13) that mines circulation logs to present users viewing an item on the catalogue

69

with suggested alternatives (Pattern, 2008). Their approach matches the item being viewed in the OPAC with a list of all the library users who have withdrawn that item. The full circulation records of those users are then collated, and the most commonly withdrawn items are presented back to users under a "more like this" tab on the record page. Even with this relatively simple method, circulation has increased significantly following the system's introduction (Pattern, 2009).

Figure 2-13: Borrowing Recommendations at the University of Huddersfield

The use of circulation records as a means of generating item-level recommendations has also been explored by Tsuji et al. (2012). They compared two methods of generating such recommendations from circulation records of an institutional library (by collaborative filtering and through an association mining rule) with recommendations generated by Amazon. Using student participants, the researchers evaluated the recommendations generated by each method, finding that in general users preferred Amazon's recommendations to either of the recommendation sets generated by the circulation-based systems.

An alternative approach is taken with the BibTip project (Monnich & Spiering, 2005 & 2006). They suggest that the use of circulation data alone is insufficient for a

comprehensive recommender system, since reference items cannot be properly integrated, and the availability of items to be loaned is not considered. BibTip therefore uses three software agents to build recommendations based on implicitly inferred data. An *Observation Agent* identifies all the titles selected within each discrete user session, passing the data to an *Aggregation Agent* which collates co-occurrences and builds links between items. Finally a R*ecommendation Agent* mines these links to provide recommendations to users in the form of hypertext links in the browser (see Figure 2-14). At each stage statistical evaluations of the data are employed to refine connections between users and

Figure 2-14: Example of BibTip Recommendations (text translates to "Others have found interesting")

titles viewed. Users of the prototype system at the University of Karlsruhe rated the quality of the recommendations as 4.21 on a 1-5 Likert scale, and the system is now employed by a number of German Universities.

Another model for providing OPAC recommendations can be found with *Library Thing For Libraries* (LTFL). *Library Thing* is an online service allowing members to catalogue their book

collections, and supplement this catalogue with ratings, reviews and tags. With more than a million members, this represents a significant amount of explicit feedback for collaborative filtering recommendations. LTFL is sold to libraries as an OPAC overlay, using a small piece of JavaScript to query *Library Thing*'s database for the ISBN of the item being viewed in the OPAC. Relevant tags, reviews and recommendations are then exported to the OPAC interface and can be viewed by the user (see Figure 2-15). Since the overlap of the Library Thing aggregated catalogue with University OPACS is around 50%, this offers a readymade means of adding significant value to a large part of the corpus (Westcott et al., 2009). While no research yet has focused specifically on the utility of the LTFL recommendations, Mendes et al. compared the resources discovered by users through the LTFL tag cloud with those found using traditional Library of Congress Subject Headings. They determined that "for every new book a user discovers using LCSH headings they will discover 4 books using LTFL" (Mendes et al., 2009: 10).

Figure 2-15: Library Thing for Libraries Recommendations
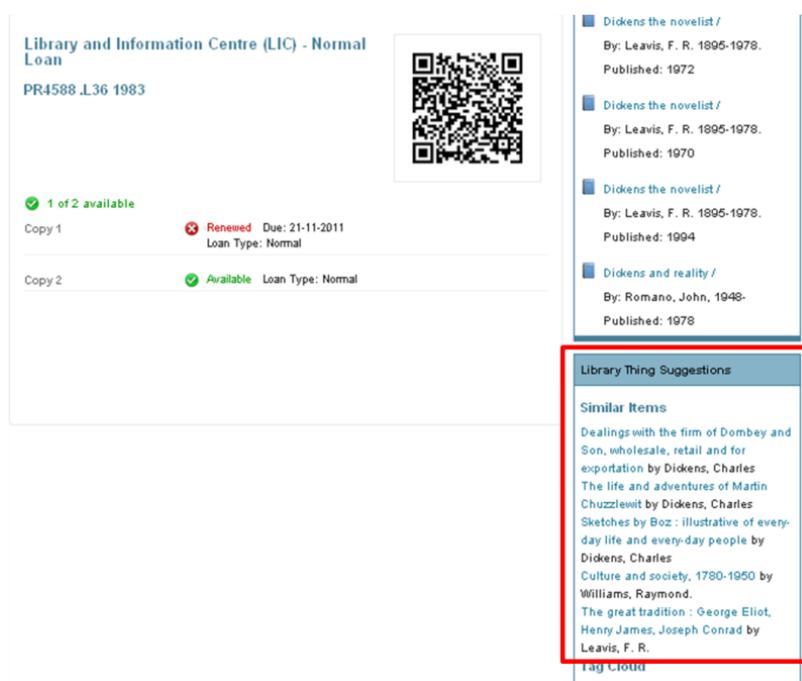
Other examples include the University of California's Melvyl Recommender project (Whitney, 2006), which investigated the utility of recommendations based on UCLA circulation data, using a method similar to that employed by the University of Huddersfield. In conducting a detailed evaluation of their prototype system they found that recommendations successfully supported academic tasks and helped with query

reformulation problems. They also determined that subject expertise heavily influenced the way that recommendations were evaluated.

Perhaps the work to date most closely related to this thesis is the SALT (Surfacing the Long Tail) project. This JISC funded initiative had as its goal the development of a recommender system for COPAC (an online union catalogue of 90 of the UK and Irelands major academic and research libraries). It should be noted here that the author could find no peer reviewed papers relating to the projects, and unfortunately was unable to contact the researchers to discuss in person. The following review of SALT is therefore based on the project blogs (http://salt11.wordpress.com/ and http://copac.ac.uk/innovations/activity-data/). While these are comprehensive, they are clearly not intended as academic publications, and therefore details of methodologies – particularly relating to evaluation data collection and participant demographics – are frequently missing.

The primary objective of the initial SALT project was to develop and evaluate a recommender system originally based on the circulation records of the John Rylance University Library (JRUL) at the University of Manchester, and subsequently expanded to incorporate similar data from the Universities of Cambridge, Huddersfield, Sussex and Lincoln. With the involvement of Dave Pattern, the librarian responsible for the University of Huddersfield recommender system described above, a similar algorithm was used to collate items borrowed by users who had withdrawn the particular item in question (Pattern, March 2011). The resulting table of the ranked items are additionally weighed according to relative subject area (based on the stock collection), and the rankings are further tweaked by the incorporation of a value designed to measure the strength of the link between the requested item and the potential recommendations. This is calculated by dividing the number of common borrowers by the number of users who have borrowed the recommender book. The higher this value, the greater implied connection between the requested item and the recommendation. The resulting top ranked items are presented to users as "SALT recommendations" on the record page in the catalogue (see Figure 2-16). A key variable in the algorithm, and one that drove much of the evaluative research on the

Figure 2-16 : Example of SALT recommendations in the COPAC catalogue

project, is that of the recommendation "threshold" – the "minimum number of unique users who must have borrowed both the item in the request and any recommendation for the recommendation to be included in results" (COPAC 2012 - http://copac.ac.uk/innovations/activity-data/?page_id=227). Lower threshold values result in a greater number of generally less accurate recommendations, and maximise the chances of a given item generating recommendations, while raising the threshold makes it less likely for recommendations to be found for a particular item, but increases the accuracy of those that are presented.

Prototype systems were developed for incorporation into the JRUL and COPAC catalogues (see Figure 2-16). Evaluations of the systems were conducted at various stages. Prior to completion of the prototype system, a focus group was conducted at the University of Manchester to garner perspectives on the role recommendations might play in their academic work Charnock, June 2014). All four students spoke positively about the potential for recommendations, although some concerns were raised – particularly the potential for "irrelevant" material to be "pushed to them", and the need for the recommendations to be somehow "trustworthy". The prototype systems were also evaluated in a lab setting on a number of occasions, with a protocol involving a focus group of student users, followed by task based evaluation of the systems (Rigby, July 2011, September 2011, August 2012).

74

Here participants were first tasked with finding a book that was well known to them, and then asked to execute a subject search in a topic they knew "a little about, but have little expertise in". In each case they were asked to rate recommendations on a five point Likert scale in terms of their relevance, novelty, and the prototype, interviews were conducted with four students at the likelihood of participant borrowing the item. This evaluation was run three times, with the principle controlled variable across session being the threshold value. The first evaluation (Rigby, July 2011) was completed by 18 humanities postgraduates using the JRUL prototype, configured with a low threshold value of 3. Assessing the results of the study is complicated somewhat by a lack of clarity in reporting. While responses to the question "Do any of the recommendations look useful" are reported as "Yes" = 77.5%, "No" = 22.5%, it is not clear whether this is calculated at a task level or participant level. It is also unclear whether these figures are extrapolated from the assessment of individual recommendations encountered during each task, or whether the question was asked in a post-session questionnaire. Overall through the results showed a strong support for the idea in principle, with a caveat that a significant number of recommendations were deemed irrelevant, and sometimes confusing.

The second user study (Rigby, September 2011) has 6 student participants, and tested both the JRUL and COPAC prototypes, with threshold values of 15 for JRUL, and 8 for COPAC. Results here show that 92.3% of recommendation sets in JRUL, and 100% in COPAC included at least one item considered "useful". Over half (56.2%) of recommendation sets in JRUL included an item that the participant subsequently intended to borrow. The final evaluation by 11 undergraduates (Rigby, August 2012) was of the prototype incorporating the circulation data from the additional four Universities. The blog post relating to this evaluation states that "91.4% of the recommendations looked useful", although once again it is not clear whether this figure refers to individual recommendations, or recommendation sets. In interpreting these results, as well as recognising the uncertainty surrounding the clarity of reporting, we might also question the potential ambiguity of the term "useful". As the authors state, some participants categorized an item as "useful" even if they didn't need to borrow it (either because they had already done so, or because it was too "niche"), but felt the item might be useful to others.

While each of the three studies reports almost uniformly positive comments from the participants, they also uncovered some competing user preferences for the system. Most significant were differences relating to the age of recommended resources, with some

researchers, particularly Law students, demanding relatively up to date material, whilst others (e.g. historians) preferred older material (Rigby, August 2012). Further research also included interviews with academics who had used the system, finding that they viewed the recommendations as valuable tools for reading list development, as well as serendipitous discovery (Manista, 2012).

2.6 Conclusion

This chapter has attempted to review the literature in three general fields which can be seen to directly influence this project. From a review of the literature relating to Union catalogues in general, and WorldCat.org in particular, we can conclude that while several assumptions are made about the demographics and search-tasks of users of such systems, there has been little research designed to systematically investigate the role union catalogues can or should play in academic resource discovery and acquisition. It is hoped that this thesis will serve to address this gap. We also noted the ongoing evolution of library catalogues, and the next-generation features and functionality offered by the newest systems. As well as revealing the methodological approaches taken to evaluate such systems, it also seems clear that students and academics are still drawn extensively to other web-based systems as a means of discovering useful resources. It is hoped that this study can further investigate the usability of next-gen catalogue features, specifically with regard to how they influence users engaged in the information search process. Information search behaviour was shown in this chapter to be located within the broader fields of Information Behaviour and Information Seeking Behaviour, and in reviewing key models in these disciplines the intention was that the micro-level of interactions between IR system and user be better understood. The final section of this chapter dealt with recommender systems, and offered an overview not just of the key types of recommender system, but the areas which might influence the development of design specifications for a union catalogue recommender. We note the relative paucity of literature evaluating recommender systems in the library domain, and indeed the general lack of research into how recommendations influence information search behaviour, and it is hoped that this thesis can make a significant contribution to better understanding how recommendations can best be designed and integrated into catalogue systems.

3. METHODOLOGY

3.1 Introduction

An introduction to mixed-methods research is given, along with a discussion of the philosophical context for the research (Pragmatism). There follows an overview of the research methodology and detailed discussion of the four phases of the research: WorldCat.org user Focus Groups, a WorldCat.org pop-up survey, analysis of WorldCat.org transaction logs, and a User Study investigating comparing the use of Amazon and WorldCat.org for typical library tasks.

It may be helpful for the reader to be reminded at this point of the Research Questions and Objectives outlined in the Introduction chapter:

Research Questions:

1. Who is using WorldCat.org?

2. For what purposes are users accessing WorldCat.org?

3. When might a recommender system support users of WorldCat.org?

4. What effect does the presence of recommendations have on the information seeking process in the library catalogue?

5. What recommendation characteristics would be most useful to users of library catalogues?

Research Objectives:

1. To identify who is using WorldCat.org, and their reasons for accessing the system

2. To establish user needs and expectations for a WorldCat.org recommender system

3. To develop a set of design specifications for a WorldCat.org recommender system

3.2 Mixed Methods Research

In order to properly address the chosen research questions, a mixed-methods approach was adopted. As Creswell (2010) notes, there has been some debate in recent years as to how best to define mixed-methods research. This reflects the relatively recent emergence of mixed-methods as a formal methodological approach. The origins of mixed-methods can be traced back to the work of various sociologists and anthropologists in the early twentieth century, although it was not until the 1990s that the approach began to be formalised (Denscombe, 2008; Johnson et al., 2007). Thus it is only since then that discussions of a definition for mixed-methods research have crystallised, with the key battle-ground being the extent to which mixed-methods as a concept encompasses elements of the research process beyond the use of complementary form of data collection. Here a clear distinction can be drawn between "methods" (different forms of data collection and analysis) and "methodology" (the entirety of the research experience, including the researcher's philosophical outlook) (Creswell, 2010). Indeed a study by Johnson et al. (2007), in which a number of researchers were asked to define "mixed methods research," uncovered significant variations in the extent to which definitions incorporated methods and methodologies, in addition to the stage at which the mixing occurred, and the purpose for mixing methods. Their resulting composite definition is as follows:

> "Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches … for the purposes of breadth and depth of understanding and corroboration" (Johnson et al., 2007: 123).

Advocates of mixed-methods research identify a number of benefits to such a blending of methods. As summarised by Johnson & Onwuegbuzie (2004: 21), in addition to the benefits associated with individual quantitative and qualitative techniques, mixed methods research offers the potential for complementary data sources to expand the range of research questions, improve generalizability, provide stronger evidence for conclusions, and add insight and understanding. The same authors also identify potential weaknesses of the mixed methods approach. Aside from what might be termed practical considerations (e.g. the time and expense associated with running multiple data collection strands, and the challenges associated with a single researcher developing the expertise to conduct multiple types of research), they also identify the philosophical challenges of synthesising methods drawn from opposing research paradigms (of which more later), and note that mixed-

methods studies are not necessarily immune from the respective weaknesses of qualitative and quantitative research. This last point dovetails neatly with Johnson & Turner's *fundamental principle of mixed research* (2003), which asserts that data collection methods should be combined to in ways that maximise complementary strengths and avoid overlapping weaknesses.

In designing mixed-method studies, two principle considerations have been identified; the relative emphasis placed on the complementary quantitative and qualitative methods, and whether the phases are to take place sequentially or concurrently. Based on these considerations, Johnson & Onwuegbuzie (2004) present a matrix representing possible study designs (Figure 3-1). It should be noted that the matrix is illustrative rather than exhaustive, and that study designs incorporating more than two phases are possible (Creswell and Plano Clark, 2011).

Time Order Decision

		Concurrent	Sequential
Emphasis Decision	Equal	QUAL + QUAN	QUAL → QUAN QUAN → QUAL
	Dominant	QUAL + quan QUAN + qual	QUAL → quan qual → QUAN QUAN → qual quan → QUAL

Capitals indicate high priority, lower case indicates lower priority, "+" indicates concurrent, "→" indicates sequential.

Figure 3-1: Johnson & Onwuegbuzie's Mixed Methods Design Matrix (2004: 22)

The combination of methods, and their relative weight and chronology, should be determined based on the study's research questions and the rationale for combining methods. Greene et al. (1989) identify five major reasons for conducting mixed methods research (see Table 3-1):

Table 3-1: Purposes of Mixed Method design (Greene et al., 1989: 259)

Purpose	Intent
Triangulation	Convergence, corroboration and correspondence of results
Complementarity	Elaboration, enhancement, illustration and clarification of results
Development	Use results from one method to develop or inform another method
Initiation	Seek paradoxes and contradiction to inform new perspectives
Expansion	Extend the breadth and range of inquiry

3.3 Research Paradigm

The term paradigm was first introduced to academia by Thomas Kuhn, who in broad terms used it to characterise the differing worldviews of scientists (Kuhn, 1962). Kuhn believed that paradigm shifts occurred when prevailing philosophical and methodological techniques were no longer able to address key research question, and new approaches emerged to tackle them. These ideas were later adapted and expanded by Guba & Lincoln (1989), who categorised the notion of a paradigm as including not just perspectives on methodology and the nature of reality, but also ethics and epistemology. Furthermore, they posited that multiple paradigms, with dramatically differing assumptions about the nature of scientific enquiry, might coexist (Mertens, 2012).

Since the widespread adoption of this interpretation, two key paradigms are usually cited as fundamental, yet competing, underpinnings of systematic enquiry. On the one hand we have Positivism, which expounds the belief that there exists a singular reality that can be revealed by objective study, and represents the philosophical context for quantitative research methodologies. Positivists tend towards hypothetic-deductive inquiry, view their results as universally generalizable, and aim to study relationships between well-defined variables while viewing ethical and moral issues as beyond the scope of the scientific method (Neuman, 2011; Morgan, 2007). In contrast Constructivism (or Interpretivism) is closely related to traditional qualitative methods, and posits that there is in fact no such objective reality, but rather a social reality constructed from subjective experience. Constructivist researchers aim to uncover meaning though inductive reasoning, view their results as context specific, and will typically acknowledge and account for the values and feelings of both researcher and study participants (Schwandt, 2000).

A key issue then for mixed-methods research is the extent to which these two paradigms can coexist, if at all, within a single study. Howe's incommensurability thesis (1988) argues that the fundamental differences in ontology and epistemology between the two paradigms are in effect unbridgeable. Strict interpreters of this position (or to use the terminology of Rossman & Wilson (1985), *purists* advocate mono-method researchers whose body of work is distinctly qualitative or quantitative. Others – *situationalists* – contend that specific research questions and contexts require researchers to adopt the more suitable of the two paradigms, although individual studies should remain mono-method. Finally an alternative view is offered by the pragmatists, who reject the notion that qualitative methods are necessarily Constructivist, and quantitative Positivist, and therefore view the integration of methods within a single study as methodologically sound (Onwuegbuzie & Leech; Miller & Fredericks, 1991; Creswell, 1995).

The use of the term pragmatist to characterise this last group is significant, in that Pragmatism itself has emerged as a guiding Paradigm for mixed methods research. Based on the work of John Dewey, William James and Charles Sanders Peirce, the central tenet of their thinking is best expressed in the Pragmatic Maxim:

> "Consider what effects, that might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object" (Peirce, [1904]: 402)

Pragmatism focuses on the examination of the practical effects of an action or idea, and posits that value and meaning is best judged by an assessment of the impact of that action or idea in the real world (Johnson & Onwuegbuzie, 2004). This allows pragmatism to inhabit a spectrum of perspective that encompasses the ground between the ontological and epistemological dualisms posited by Positivism and Constructivism. While these paradigms contend that reality is necessarily objective or subjective, the pragmatist belief in an "existential reality" (Dewey, 1925: 40) demands an *intersubjective* approach – one in which a researcher "has to work back and forth between various frames of reference" (Morgan, 2007: 71). Likewise, the deductive vs. inductive dichotomy is superseded in the pragmatic paradigm by the notion of *abductive* reasoning, which might best be summarised as a form of inference that seeks to identify the most likely explanation for a phenomenon (Ivankova, Creswell, & Stick, 2006). Finally the pragmatic approach advocates a middle-ground between the context-bound constructivism and the universally generalizable positivism, instead acknowledging the propensity of different types of inferences to be

more or less applicable in different circumstances   - what Morgan terms *transferability* (2007: 72). Thus researchers working within the pragmatic paradigm are freed from the ''forced choice dichotomy between postpositivism and constructivism'' (Creswell & Plano Clark, 2007: 27), and instead offered a philosophical context for their research that justifies their utilisation of different methods and analytical techniques.

The final point worth stressing in the discussion of pragmatism is its emphasis on the relationship between inquiry and practice. The research project described in this thesis can be said to incorporate elements of library science, information retrieval, and interactive information retrieval, and a common theme of research within these areas is a predominance of "applied" research; i.e. research with a focus on addressing some real-life problem (Connaway & Powell, 2010; Kelly, 2009). Since a research objective of this project is to produce design specifications for a WorldCat.org recommender system, it seems appropriate to work within a research paradigm that "advocates an informative relationship between inquiry and practice where inquiry simply informs practice" (Greene & Hall, 2010: 132).

3.4 Overview of Methodology

Based on an analysis of the research questions and available quantitative and qualitative methods, a pragmatic multi-phase mixed-method methodology was devised, with phases running mostly sequentially. A mixed-methods approach was deemed suitable for this project for a number of reasons. Properly addressing the research questions was found to require elements of exploration and analysis. The exploratory aspects (for example determining the purposes for which WorldCat.org is used, and the characteristics of recommendations that would be most useful to users) are most naturally aligned with qualitative methods, while analytic aspects of the research (for example understanding the effects of recommendations on the information seeking process) are best addressed through quantitative methods. Thus both qualitative and quantitative methods emerged as the optimal means of addressing these questions. Furthermore, the research questions can also be viewed holistically as a means of better understanding the relevance of recommendations to WorldCat.org, and as a means of guiding and optimising the design of a recommender system. Thus a combination of research methods was deemed likely to broaden the perspective offered on these more general goals, while the collection of different types of data might potentially mitigate the weaknesses inherent in any one data

set. The resulting heterogeneous datasets also offered the potential for varying forms of integration (see section 3.7).

Figure 3-2 provides a diagrammatic model of the research phases, which are described in detail in sections 3.5.1-4, while Figure 3-3 shows a Gantt chart of the chronology of the project. It should be noted here that aside from the philosophical pragmatism introduced in section 3.3, elements of the project were necessarily guided by what Biesta calls "everyday pragmatism" (2010: 96) Since three of the intended data collection methods relied heavily on the cooperation of OCLC (providing access to the log data, implementing the survey, funding student participants in the focus groups) elements of the research process were guided to an extent by the everyday realities of collaborating with a large and diverse organisation. In particular, flexibility was required in the timing of the various phases, especially the analysis of WorldCat.org transaction logs. The chronology of the project can therefore be seen as a compromise between intended best practice and practical considerations.

A number of alternative methods of data collection were considered, and in some cases pursued, two of which merit mention here. The first was a potential extension to work conducted by this author (Wakeling et al., 2011), discussed in the introduction to this thesis, and would have centred on canvassing the opinion of Library professionals to better understand the reason for the relatively small number of recommender system implementations in operational catalogues. While this would have potentially yielded valuable data relating to librarian perspectives on catalogue functionality in general, and recommender systems in particular, it was decided that a) the phase 1 focus groups would offer an opportunity to canvass practicing librarians on the subject, and b) that the user-focused nature of the research questions suggested that time and resources were better spent elsewhere.

Secondly, and perhaps most significantly, considerable attention was paid to the possibility of conducting user evaluations of existing library recommender systems. A difficulty with this approach, which is discussed further in section 3.5.4, relates to the importance of context and corpus size. Thus while, for example, the University of Huddersfield's OPAC is publicly accessible, and so available for evaluation by researchers outside that institution, such an evaluation would effectively take place in isolation – i.e. without a point of comparison against which to measure the utility of the recommendations (Kelly, 2009: 29).

Overall Design: (QUAL + quan) → QUAN → QUAN → (QUAN + qual)

Phase 1: QUAL + quan
WorldCat.org user Focus Groups
RQs addressed: 1, 2, 3, 5

Phase 1 results inform survey design.

Data Analysis
Coding; qualitative content analysis.

Interpretation
Discussion of use of WorldCat.org, perceived strengths and weaknesses and suggested improvements.

Phase 2: QUAN
WorldCat.org pop-up Survey
RQs addressed: 1, 2, 3, 5

Phase 1 and 2 results inform queries run on log data.

Data Analysis
Quantitative analysis of survey results.

Interpretation
Discussion of use of WorldCat.org system, use of Library 2.0 functionality, and recommendation preferences.

Phase 3: QUAN
WorldCat.org Log Analysis
RQs addressed: 1, 2

Phase 1 and 2 results inform user study design.

Data Analysis
Quantitative analysis of log data. Manual coding of sample sessions.

Interpretation
Discussion of WorldCat.org usage. Use of WorldCat.org features (facets, advanced search etc.) and session types.

Phase 4: QUAN + qual
WC Vs. Amazon User Study
RQs addressed: 3, 4, 5

Data Analysis
Quantitative analysis of study data; coding and analysis of post-task interview data.

Interpretation
Discussion of the role of recommendations and the effect on information search behaviour.

Integration of Phases 1 to 4
Integration of findings from each phase to corroborate, enhance and develop understanding. Identification of key themes emerging from the data. Development of holistic perspective on research questions. Review of practical impact on recommender system design specifications.

Figure 3-2: Diagram of Mixed-Methods Study Design

Literature review

Develop Methodology

Ethics Review

OCLC Visit

Phase 1 - Focus group design

Phase 1 - Australia & New Zealand focus groups

Phase 1 - UK Focus Groups

Phase 1 - US Focus Groups

Phase 1 - Focus group data preparation and analysis

Phase 2 - Survey design and implementation

Phase 2 - Survey live

Phase 2 - Survey data analysis

Phase 3 - Preparation of log data

Phase 3 - Analysis of log data

Phase 4 - User study experimental design

Phase 4 - User study data collection

Phase 4 - User study data analysis

Write up thesis

Year 1 (2010-11) | Year 2 (2011-12) | Year 3 (2012-13) | Year 4 (2013-14)

O N D J F M A M J J A S

Figure 3-3: Gantt Chart of Research Project

It was decided therefore that any such evaluation would necessarily need to be conducted *in situ*, with participants drawn from regular users of the system, and that any data collected should ideally be augmented by available usage statistics (e.g. institutional OPAC log files). Contact was made with a number of institutions whose OPACs were found to offer recommendations to discuss the potential for such a study, but these approaches were unsuccessful. In light of this the study design that forms Phase 4 of this project was created.

3.5 Research Phases

3.5.1 Phase 1: Focus Groups

Focus group interview research offers "a way of collecting qualitative data, which – essentially – involves engaging a small number of people in an informal group discussion (or discussion), "focused" around a particular topic or set of issues" (Wilkinson, 2004: 177). It is a group interview intended to explore the beliefs, experiences and feelings of the subjects, thereby offering insights into behaviour and perception and allowing the investigators to develop ideas and hypotheses (Walden, 2006). Although focus group interviews were originally developed as a means of gathering market research, their use spread to the Social Sciences in the late 1980s (Peek and Fothergill, 2009; Madriz, 2000). As the name implies, sessions begin with a broad discussion and gradually focus on the subject of the research. A number of advantages of focus group interviews as a research method have been identified, including economies of scale that allow for a greater number of participants, the likelihood of participant interaction leading to more data for collection, and the opportunity to record spontaneous opinions (Onwuegbuzie et al., 2009; Evans & Kotchetkova, 2009). Focus group interviews also constitute an established methodology within Library and Information Science (Seggern & Young, 2003; Connaway & Powell, 2010: 173-174; Connaway, 1996), and a number of previous studies have used the methodology to investigate the use of online catalogues (e.g. Berger & Hines, 1994; Connaway et al., 1997).

3.5.1.1 Design

The intention of this phase of research was to gather qualitative data from users of WorldCat.org relating to their use of and perspectives towards the system. Since there is no literature detailing the specific groups constituting WorldCat.org's user population, the

selection of groups to be targeted in the research was driven by discussion with OCLC. In particular, reference was made to an internal document which outlined a number of user personas. These identified the demographics and characteristics of key types of system user, including: librarians of different types (cataloguing, public access, systems; public and academic); antiquarian booksellers; school, undergraduate and postgraduate students; rare booksellers; and academics (arts and social sciences). Since it was considered impractical to investigate all of these groups, the decision was made to focus on user groups that might offer a range of perspectives on the system, whilst also limiting the practical difficulties of data collection. The user groups selected were librarians (public access and cataloguing, university and non-university), students (postgraduate and undergraduate), antiquarian booksellers, and academics (historians). Further, it was decided to undertake focus groups in three distinct geographic locations: the US, the UK, and Australia and New Zealand. Whilst not comprehensive, it was hoped that this geographic diversity might reveal differing perspectives on WorldCat.org.

3.5.1.2 Focus Group Questions

The questions asked during the focus group interview sessions were carefully designed to ensure that participants had the opportunity to address a broad range of issues and experiences with WorldCat.org. A total of five questions were developed (see Table 3-2).

Table 3-2: Focus Group Questions

Question	Purpose
1. Tell us about your experiences with WorldCat.org	A broad introductory question intended to reveal the extent to which users have engaged with WorldCat.org, and the information-seeking contexts within which they use the system.
2. Describe a time when you used WorldCat.org that you considered a success.	Explores the features and functions of WorldCat.org that participants view positively. Requiring participants to discuss a particular instance provides richer data about the range of uses of the system.

3. Describe a time when using WorldCat.org was unsuccessful – i.e., you did not get what you wanted.	Explores the features and functions (or lack thereof) of WorldCat.org that participants view negatively.
4. Describe a time when you found something that wasn't what you were looking for originally, but was still of interest or useful to your work?	Intended to encourage discussion about the role of serendipity in information seeking, and the extent to which WorldCat.org facilitates resource discovery.
5. If you had a magic wand, what would your ideal WorldCat.org provide?	Encourages participants to discuss potential improvements to WorldCat.org. The use of the phrase "magic wand" ensures that participants are not restricted by what they believe to be practical or realistic.

In addition to the initial five questions, detailed follow up queries or probes were developed. These can be found in APPENDIX 1, and were intended to help the moderator stimulate the discussion if participants were struggling to engage with the broader questions.

3.5.1.3 Participant Recruitment

Given the lack of precise quantitative data regarding the usage rates for the different user groups under investigation, and the logistical difficulties of recruiting participants across three continents, it was determined that probabilistic sampling methods were impractical for this study. Two non-probabilistic methods were therefore applied: convenience sampling (with participants selected based on expediency) and snowball sampling (a cumulative approach that draws on the social and professional networks of participants to expand the sample size (Connaway & Powell, 2010: 117-119).

Different approaches were required for the recruitment of the various user groups under consideration. Since initial contact was with libraries, the recruitment of librarians was relatively straightforward. Key contacts at participating institutions provided email addresses for suitable staff members, or distributed the invitations themselves. In both cases a standard email was used that informed potential participants of the focus group

interview details, explained the aims of the research, and specified what would be required of participants.

The recruitment of the other user groups proved more problematic. Library contacts were used to identify the subject liaison librarian for History departments, which resulted in a limited number of introductions to historians. Invitation emails were sent to interested parties, but overall recruitment numbers were low. Antiquarian Booksellers required a different approach, since in most cases library contacts were unable to help with this group. Booksellers were identified through their membership of professional bodies (the Australian & New Zealand Association of Antiquarian Booksellers, the Antiquarian Booksellers Association (UK), and the Antiquarian Booksellers' Association of America), whose websites include contact details for members. Emails were sent to all booksellers included in the online membership directories, explaining the research and inviting them to attend a session. While this approach was unsuccessful in Australia and the US, we were able to recruit enough UK-based booksellers to conduct a focus group interview session.

A number of methods were used for the recruitment of students. In some cases libraries were willing to send a mass email on behalf of the investigators to student mailing lists, which proved highly successful. Other students were recruited through fliers distributed around libraries, via related academic departments, and from the ranks of student library assistants. In addition to refreshments, students were offered a cash incentive for attending, amounting to £15 in the UK and $25 in the US, which aided recruitment. It should be noted that due to the limited time available for conducting Focus groups in Australia and New Zealand, it was not possible to schedule student sessions in these regions. A summary of participants by user group can be found in Table 3-3, while a complete breakdown of each session by location and number of participants can be found in APPENDIX 2: Breakdown of Focus Groups by Location and Participants.

Table 3-3: Focus Group Participants by User Group and Location

	Aus / NZ	UK	US	Total
Librarians	23	18	20	61
Students	0	15	25	40
Booksellers	0	10	0	10
Historians	4	3	0	7
Total	27	46	45	118

	University	Non-University	Total
Public Access	37	10	47
Cataloguing / Other	8	6	14
Total	45	16	61

	Undergraduate	Postgraduate	Total
Arts	12	2	14
Social Science	3	13	16
Science	10	0	10
Total	25	15	40

3.5.1.4 Data Collection

The research was conducted in three stages, each relating to a geographical location: Australia and New Zealand (21 March – 8 April 2011), the UK (9 – 17 May 2011) and the US (25 – 27 October 2011), with 21 sessions conducted in total. Due to the costs of this author travelling to Australia and New Zealand being prohibitive, sessions in these locations were conducted solely by Dr. Connaway, and were initially intended as a pilot study to test the effectiveness of the questions and the focus group format. Since the data collected proved useful and complete, and no changes were subsequently made to the research protocol or questions, the data from these sessions is included in the final study. In each location the same general process was followed for the recruitment of participants and scheduling of the sessions. As far as possible, existing contacts with key staff members within academic and National libraries were used as a starting point for recruiting participants and securing appropriate venues for the focus group interview sessions. Initial contact with a number of institutions was made by email, and the broad goals of the research were explained. Based on their willingness to participate, and the availability of key staff and facilities, a preliminary schedule was drawn up for the research in each location.

The focus group interviews were conducted according to the principles laid out in Connaway & Powell (2010). Sessions were held in informal environments (usually a library meeting or conference room), and efforts were made to ensure that each session included between six and twelve participants, in line with guidance on optimum group size found in

the literature (e.g. Millward, 1995; Connaway & Powell, 2010; Berg & Lune, 2011) . In all cases potential attendees were sent a copy of the questions to be asked during the session in advance, and received a reminder email two days before the scheduled session. Despite this, on a number of occasions poor attendance rates meant that the final number of participants fell below the optimal minimum, including several instances where only one participant was present. In these circumstances a standard interview was conducted using the same questions.

Sessions lasted between one and two hours, and refreshments were served, with every effort made to ensure the participants felt comfortable and relaxed throughout. To ensure that all available data were captured, the sessions were recorded on a digital audio recorder. The majority of Focus group sessions were conducted with the assistance of Dr. Lynn Sillipigni Connaway, OCLC Senior Research Scientist. During each session one investigator acted as moderator, initiating the discussion and probing for more detailed answers where appropriate, and encouraging everyone in the group to participate. Another investigator acted as note taker, providing a detailed record of the session in case of equipment failure, and registering the individuals or groups for whom certain ideas were particularly important. The investigators alternated between roles.

3.5.1.5 Analysis

In order to analyse and interpret the results of the focus group interviews, Qualitative Content Analysis (QCA) was employed. Content Analysis describes the coding of human communications, with coding being "the process of transforming raw data into standardized form (Babbie, 2001: 309). It has traditionally been viewed as a quantitative analytical tool, since the frequency and distribution of codes within the corpus is a primary means of inferring meaning from the data (Krippendorf, 1989).There are however a number of difficulties associated with applying this quantitative method to focus group data. Sampling issues (i.e. whether the focus group participants truly representative of a broader population) can affect the validity of findings, whilst the method has also been criticised for a blindness towards the context of textual components and distinct individual cases (Kohlbacher, 2006). Qualitative Content Analysis, as developed by Mayring (2000), aims to address these weaknesses by applying qualitative-interpretive tools to the analytic process (Mayring, 2000). It has been defined as "a research method for the subjective interpretation of the content of text data through the systematic classification process of

coding and identifying themes or patterns" (Hseih & Shannon, 2005: 1278). QCA addresses the potential weaknesses of classical content analysis by acknowledging that the data should be considered within its wider context. Thus it is better able to recognise the potential significance of individual cases that might be marginalised during a classical content analysis process that utilises a purely quantitative approach to the data. Similarly, by locating the analysis primarily within the qualitative paradigm, researchers are less inclined to generalise the results of QCA, meaning that potential sampling issues are mitigated.

While other common methods of qualitative data analysis were considered, particularly Grounded Theory and Thematic Analysis, a number of factors influenced the selection of QCA. A full Grounded Theory methodology is best applied to a large scale research project (Pidgeon & Henwood, 1997), and aims to produce a theoretical framework for understanding the results (Birks & Mills, 2011). In this case it was determined that since the Focus group research constituted only one of four phases of research, and the goal was not to produce a stand-alone theory based solely on this phase, it was not appropriate to employ the full set of grounded theory practices. Thematic analysis is a broad term used to describe a number of similar methods of qualitative data analysis which identify themes, and the relationship between themes, in the data (Braun & Clarke, 2006). While it has been noted that many modes of thematic analysis closely resemble QCA (Vaismoradi et al., 2013), it is generally held that QCA offers the potential for some quantitative interpretation of the results, particularly counts of code occurrences (Morgan, 1993, Gbrich, 2007). While this has attendant risks, particularly relating to the validity of participant sampling, it was felt that the potential for this form of analysis was beneficial, and QCA was therefore adopted for the analysis.

 Zhang and Wildemuth (2009) describe a seven-stage process for conducting a Qualitative content analysis, with the stages outlined as follows:

1. **Prepare the Data**

    Transcribe the data or validate and augment notes

2. **Define the Unit of Analysis**

    Unit of analysis is generally at the thematic rather than linguistic level

3. **Develop Categories and Coding Scheme**

    Can be developed deductively or inductively

4. **Test Coding**

Validate the code book by coding a sub-set of the data.

5. **Code all Text**

   Apply coding rule to entire data set.

6. **Assess Coding Consistency / Intercoder reliability**

   Verify that coding practices have not developed over the coding process, and that there is consistency between coders if applicable.

7. **Draw Conclusions from the Coded Data**

   Make sense of the themes or categories identified

This process was closely followed in the analysis of the focus group data. The notes from each session were augmented and clarified after a review of the audio recording (step 1). Full transcription of the sessions was not deemed necessary, both for practical reasons (the large quantity of recorded audio data would have made transcription an impractically time-consuming or expensive process), and since the unit of analysis (step 2) was at the thematic rather than linguistic level (Zhang & Wildemuth, 2009). Both investigators closely examined these notes, highlighting all ideas and terms that related to participants' engagement with WorldCat.org. These terms were then rationalized, merged as appropriate, and arranged into a hierarchical structure within five main categories: Work-Tasks, Search-Tasks, Strengths, Challenges / Difficulties, and Suggestions for Improvement (stage 3). To test the code book, two researchers coded the same five randomly selected transcripts, and compared results. After discussion, the code book was amended to reflect the final agreement on coding terms and organization (step 4). The final code book can be found in APPENDIX 3: Focus Group Code Book.

Following finalization of the code book, the transcripts from all the focus group interview sessions were coded (step 5) by this author. To ensure the highest possible level of accuracy, coding was done manually on printed copies of the transcripts, with no more than four sessions coded in any one sitting. Once all coding was complete, five sessions (20%) were randomly selected, and coded by Dr Lynn Connaway. The coding of these five sessions were then compared for inter-coder reliability (step 6). Overall inter-coder reliability was calculated using Cohen's kappa coefficient to be .85, above the .80 required to indicate reliable coding (Yardley, 2008). This confirmed that the coding scheme was suitable for effective analysis.

The notes documents for all sessions were coded using the NVIVO qualitative research tool. A profile was also created for each participant, which captured their specific user group

(e.g. US-based Arts and Humanities Undergraduate student, or UK-based public access university librarian). Each coded item was then also assigned to the relevant participant, allowing the researcher to run complex queries on the data, and to identify potential trends attributable to specific user groups (step 7).

3.5.2   Phase 2: Survey

Surveys have long been utilised as a means of gathering data, and since the emergence of the World Wide Web in the mid-1990s researchers have increasingly preferred web-based questionnaires rather than more traditional mail or face-to-face surveys. Web surveys now constitute an established data-collection method in many domains. This is reflected in the large number of studies within the library and information science community that have utilised survey methodology to investigate a diverse range of subjects including the use of library websites (Schonfeld & Housewright, 2010; Fry & Rich, 2011), digital library evaluation (Hill et al., 1997; Choudhury et al., 2002), the information–seeking behaviour of college students (Head & Eisenberg, 2009; De Rosa, 2006), ebook usage (JISC & UCL, 2009), and the usability of online catalogues (Denton & Coysh, 2011; Bertot et al., 2012; Calhoun, Cantrell, Callagher, & Hawk, 2009).

Connaway and Powell (2010) identify a number of advantages offered by surveys; participant anonymity (encouraging honest answers), the elimination of interviewer bias, freedom for the participant to answer at their leisure (encouraging thought-out and accurate responses), and the relative ease of collecting large amounts of quantitative data at little or no cost. However, they also note some attendant disadvantages, including the effect of responder and non-responder bias, high non-response and non-completion rates in general, and the potential for question ambiguity and miscomprehension by participants. These difficulties though can be mitigated through careful survey design, for example by avoiding complex language and technical jargon.

3.5.2.1 Design

To achieve a comprehensive understanding of users' interactions with WorldCat.org, the survey questions were developed to cover three broad areas:

1. Demographic information about the user:
   - Gender, Age, Location, Occupation

2. Details of the user's information-seeking context and task, and the type of resource required:
   - o Purpose and reason for using WorldCat.org
   - o Format of required resource
   - o Importance of full-text
3. The User's experience with and perspective on certain features of the system:
   - o Creating a profile
   - o Creating and / or using lists, tags, reviews and ratings
   - o Preferred type of reviews and recommendations

It should be noted that certain questions, specifically those relating to the use of tags and lists, were inserted at the request of another OCLC researcher. OCLC were keen to maximise the value of the survey by collecting data for use across a number of projects. Furthermore, in developing the questions it was agreed with OCLC that in order to ensure as a high a completion rate as possible, and minimise any potential inconvenience to WorldCat.org users, the number of questions should not exceed fifteen. Whilst this naturally limited the opportunity to explore the research questions relating to this project, it is recognised that the implementation of the survey proved to be a significant undertaking requiring the approval at a number of levels within the organisation, and significant technical assistance. It was therefore entirely understandable that OCLC should seek to maximise the value of the survey.

Careful consideration was given to establishing the most important areas for investigation, and where possible questions were structured so as to provide the richest possible data. The questions and answer options were heavily influenced by the results of the focus group research, particularly with regards to the types of tasks users were likely to be completing. A variety of question and answer types were utilised. In most cases participants were able to select one or more answer from a predefined set of options, with the capacity to manually enter "other" answers. Other questions utilised Likert scaled responses, or simple YES / NO answers. Care was taken at this stage to avoid technical jargon wherever possible, and where such jargon was essential (e.g. the use of the word "tag") an explanation was provided. It also was noted that a high proportion of traffic to the site enters directly at a record page from a set of search engine results; therefore, it was determined that wherever possible, questions would refer to "this site" rather than "WorldCat.org", so as not to confuse users who were unaware of the service where they had arrived.

A number of platforms for delivering the survey were considered, with *SurveyMonkey*[3], an online survey and questionnaire tool, considered the most appropriate. Their "Gold Plan" service supports a wide range of question types and results formats, and allowed sufficient flexibility in layout. Short introductory and concluding passages were composed to bookend the survey.

3.5.2.2 Testing & Revision

Prior to the survey going live, a link to the *SurveyMonkey* questionnaire was sent to fifteen participants to pre-test the survey. The makeup of the testing group was as follows:

- 8 OCLC colleagues (with domain and systems knowledge)

- 3 University of Sheffield colleagues (with domain but not systems knowledge)

- 4 Others (without any specialist domain or systems knowledge)

Testers were asked to complete the survey, and respond to the following specific feedback requests:

- The time it takes to complete the survey

- Are the questions clear and easy to understand?

- Are the answer options comprehensive enough (i.e. are there any answers you think we should add)?

- Are there any technical terms or other words that you did not understand, or that you feel require further explanation?

- Are there any technical issues with the survey (e.g. pages not loading properly)?

No technical issues with the survey were reported. As a result of the feedback from the test participants, the expected completion time for the survey was set at five minutes. In addition, several questions were refined to resolve issues of clarity and potential overlaps between answer options. The finalised survey questions, along with the introductory and concluding passages, can be found in APPENDIX 4: Pop-Up Survey Questions.

3.5.2.3 Implementation

In order to capture potential differences in behaviour and opinion between users accessing the site through the WorldCat.org homepage, and those landing directly at detail record

---

[3] http://www.surveymonkey.com

pages, two identical questionnaires were created in SurveyMonkey, with invitation pop-ups appearing from the respective pages (i.e. invitations from the homepage going to one survey, and from the record pages to another). An important consideration in the implementation of the survey was that regular users of WorldCat.org should not be subjected to repeated invitations to complete the survey. Initially it was hoped that the use of cookies would ensure that an invitation would only appear once for each IP address accessing the site during the period the survey was live. However, it became apparent that this would be difficult to properly implement. Instead it was decided that the invitation to complete the survey only would appear on every 100th record page accessed, and every 100th time the homepage was loaded. This meant that the probability of a single user receiving multiple invitations was extremely low.

The survey went live at 00:00 hours EST on Thursday 5 April 2012. Screenshots of the survey's invitation and question screens can be found in APPENDIX 5: Screenshots of Pop-Up Invitation and Survey. After a week, a review of completed responses revealed extremely low response numbers from the WorldCat.org landing page. It was therefore decided that the invitation would be set to appear every time the homepage was loaded (rather than every 100th time) for the remainder of the survey period. Invitations at the record pages remained at 1/100. The survey ran with these invitation ratios until 00:00 hours EST on Thursday 19^{h} April 2012. Whilst ideally the survey would have run for longer, the needs of this project had to be balanced with the potential inconvenience to users of the service. It was therefore agreed with OCLC that the survey would run for a maximum of two weeks.

3.5.2.4 Data Preparation and Analysis

A total of 980 responses were collected from the WorldCat.org page survey and 2,669 from the record pages survey. Of these 3,649 responses, 731 were incomplete, leaving 2,918 completed surveys (894 from the .org page, 2,024 from record pages). Since it proved impossible to obtain WorldCat.org transaction logs for the exact period the survey was live (see Section 3.5.3.1), a calculation of response rates is problematic. Based on the traffic to WorldCat.org shown in the logs for October 2012 a rough estimate can be made which puts the response rate at 1.6%. While this is low for traditional survey instruments, it is not uncommon for online pop-up surveys to record response rates well below 5% (Ockuly, 2003; Chirabnov, 2011).

A number of questions included an "Other" answer option, with a free text field for respondents to manually enter their answer. This was most significant for Q6 – "What is your Occupation?" – where 714 respondents (25% of the total) had written in an answer. In order to capture these data, the responses were manually coded. This resulted in the identification of 14 additional occupation types, each of which had been specified by a minimum of 10 respondents, and which taken together accounted for 461 respondents. The remaining 270 responses were coded as either "Other professional" (where the user had specified a recognisable profession – e.g. "Probation Officer", "Chef") or "Other" (where the user's response could not be coded in any other way – e.g. "Unemployed", "Religious Humanitarian").

The "Other" responses for several other questions also were manually analysed. These were answers relating to questions about the users reasons for visiting the site, and their perspectives on recommendations and reviews. In many cases respondents had written in answers that clearly matched one of the pre-defined answer options, in which case the response was re-assigned accordingly.

All responses were collated and imported to SPSS (version 18), where a number of analytical tools and tests were applied. Cross-tabulation was used to obtain descriptive statistics, with Pearson's Chi Square test used to confirm the significance of categorical data. Since a number of the questions required Likert scale responses, consideration was given to the appropriate means of interpreting and analysing such data. In particular, the author notes significant debate in the literature regarding whether they should be treated as ordinal data (where the numerical data is viewed as having no value apart from as a ranking indicator) or interval / scale data (where the numerical value represents a standardised degree of difference between levels) (Pallant, 2010). While it has been argued that an interval interpretation can be valid in certain circumstances (e.g. De Winter and Dodou, 2010), a majority of researchers tend towards viewing Likert scale data as ordinal (Pallant, 2010). Thus distributions and medians were used to analyse these data.

3.5.3   Phase 3: Transaction Log Analysis

Jansen (2006) defines a transaction log as "a file (i.e., log) of the communications (i.e. transactions) between a system and the users of that system" (2006: 408). These logs are commonly stored by organisations using server applications, and represent a record of interactions between users and online IR systems. Transaction log analysis (TLA) describes

the methodical and comprehensive investigation of queries and other actions executed by a user, and the resulting system response (Phippen et al, 2004; Blecic et al., 1998). Thus TLA "can be conceptualized both as a form of system monitoring and as a way of observing, usually unobtrusively, human behaviour" (Peters, 1993: 42). Mat-Hassan & Levene (2005) chart the development of TLA through three distinct fields of research; web information retrieval, data mining, and library science. Within each of these areas, TLA techniques provide a means of developing in-depth accounts of user behaviour when interacting with a given IR system (Jones et al. 2000).

TLA typically focuses on one or more of three levels; term level, query level or session level (Jansen, 2006). At the term level, analysis is concerned with the frequency, diversity, or co-occurrence levels of particular text strings in user queries. Query level analysis broadens this approach to take entire queries as the base unit of analysis, and might seek to investigate patterns of query reformulation, query structure and complexity, or repeated queries. The session level widens the scope still further to encompass the entirety of interactions within a period of user/system interaction (Hancock-Beaulieu, 2000), and therefore offers an opportunity for analysis that investigates issues of user intent and information seeking behaviour. Given that the intention of this phase of the research is to better understand who is using WorldCat.org, and the search-tasks they are engaged in, it is natural that the focus of the analysis of the WorldCat.org logs should be at the session level.

In defining what exactly constitutes a session from the perspective of TLA, we encounter contrasting views that resemble disagreements found elsewhere in the literature relating to definitions of tasks and goals. Jansen et al. define a session as "a series of interactions by the user toward addressing a single information need" (2007: 862). This definition is however problematic, since the usefulness of defining a session in relation to a "single information need" is called into question by studies showing the frequency with which users are found to address multiple work- and search-tasks in a single continuous period of interaction (Spink et al., 2006). A safer definition is perhaps a broader one, whereby a session constitutes the sequence of searches and other actions undertaken by a user within a single episode of engagement (Beitzel et al., 2004; Park et al., 2005).

Jansen's practical guide to conducting TLA describes three key stages; collection, preparation and analysis (Jansen 2006). There follows a description of the activities

undertaken for each of these stages. Since this researcher lacks advanced programming skills, it should be noted that the technical process of collecting, preparing, and querying the log data was undertaken by others, namely Jeremy Browning (OCLC), Professor Paul Clough (University of Sheffield) and Dr David Tomas (University of Alicante).

3.5.3.1 Data Collection

OCLC ultimately supplied files containing log data for two months of WorldCat traffic; October 2012 and April 2013. The initial intention had been that the log data supplied should cover two months: the first covering the period for which the WorldCat.org pop-up survey had been live (April 2012), and the second six months later (to offer the opportunity for exploring changing usage patterns). However although these data were collected by OCLC, transferred to disc, and posted to the author, the package was lost in transit. Unfortunately OCLC did not retain a copy of all the data, which is not stored as standard, and therefore the April 2012 dataset was lost. It was therefore decided to replace this with the transaction log for April 2013 instead. The data that was received consisted of two tab delimited files, one for each month. The dataset included the following fields shown in Table 3-4. In addition to the log files, OCLC provided a supplementary file providing additional information about each unique OCLC ID found in the logs. This file included the fields detailed in Table 3-5.

Table 3-4: Fields in log data supplied by OCLC

Field	Description
Anonymised IP Address	A random code assigned to each unique IP address present in the log
Country of origin	The country of origin of the IP address, as determined by an IP lookup service
Date	The date of the server interaction
Time	The time of the server interaction (hh:mm:ss)
URL	The URL executed by the server
OCLCID	The OCLC ID of the item being viewed (if applicable)
Referrer URL	The page from which the URL was executed
Browser	Technical information about the browser type and version

Table 3-5: Fields in supplementary item file provided by OCLC

Field	Description
OCLCID	The OCLC ID of the item
Title	The title of the item
Author(s)	The Author(s) of the item
Publisher	The publisher of the item
Media Type	The item's media type (book, DVD etc.)
Summary	Summary of the item (not present for most items)
Total Holdings	The number of OCLC member libraries holding the item
Academic Holdings	The number of libraries holding the item which are classified as "Academic" by OCLC
Public Holdings	The number of libraries holding the item which are classified as "Public" by OCLC
Occurrences in logs	The number of occurrences of the item in the logs.

3.5.3.2 Data Preparation

3.5.3.2.1 Removing non-human traffic

As part of the process of preparing the log files prior to sending, OCLC filtered the dataset to remove a large amount of robot traffic (primarily web search engine crawlers). This was done using a current list of robot traffic identifiers maintained by the University of Sheffield IR group. Since such a list is never completely comprehensive, a further filtering process was undertaken to remove sessions consisting of more than 100 queries. This follows best practice guidelines offered by Jansen (2006). The removal of robot traffic reduced the number of lines in the combined logs by more than half, from over 100,000,000 to 56,243,702.

3.5.3.2.2 Session cut-off time

Since a key aim of this research phase was to analyse the logs at a session level, an important consideration in preparing the log data was determining a means of reconstructing user sessions. Simply sorting the data by masked IP address is clearly insufficient, since a user or users from a single IP address may conduct multiple search

sessions over the period of the logs. It is therefore necessary to use some method of reconstructing sessions from the logs originating from each unique IP address. A number of methods of achieving this have been developed. While some researchers have advocated methods based on query reformulation (Jansen et al., 2007), navigation patterns (Kapusta 2012), and combinations of various metrics (Jones & Klinkner, 2008), such methods are often complex and time-consuming. The simplest and most widely used method is the adoption of a session cut-off time, which segments sessions according to a set period of inactivity. Thus a new session is applied to logs originating from a single IP address if server transactions attributable to that IP address are separated by a set time period. Most commonly this period is set at 30 minutes (Jones & Klinkner 2008), a figure closely linked to early research suggesting the average web search session length is around 25 minutes (Catledge and Pitkow, 1995). Researchers have suggested both lower and higher periods; He and Goker (2000) show that 10 minutes is the minimum inactive duration to be used, while others suggest durations ranging from 15 minutes (Jansen and Spink, 2003; He et al., 2002) to 125 minutes (Montgomery and Faloutsos 2001), the latter authors also suggesting that in practice the inactive period selected has little effect on session segmentation.

It must be noted however that the majority of studies in this area focus on web search engine logs. Since one might reasonably assume that users interact differently with an online union catalogue than with a search engine, further investigation into the effect of applying different inactive durations to the WorldCat logs was conducted. A sample of 10,000 lines of the logs ordered by IP address was generated, which represented 721 unique IP addresses. These logs were imported into excel, and analysis run to identify all inactive periods of between 10 minutes and 1 hour for the same IP address (i.e. instances where lines of the log originating from the same IP address were separated by between 10 and 60 minutes). A total of 487 such instances were observed. Each instance was then manually examined in the context of the full logs to determine whether the activity either side of the inactive period might reasonably be considered part of the same session, and coded accordingly ("Same session" or "Different Session"). This judgement was primarily based on the subject area of the queries executed and items viewed either side of the inactive period. Since this judgement was inherently subjective, to limit the likelihood of incorrect judgements a third code was also used ("Unknown"). This was applied in circumstances where there was no reasonable way of judging whether the inactive period constituted a new session or not.

The resulting data consisted of 487 inactive periods of between 10 minutes and 1 hour, and the code assigned to each period. 99 of these were coded "Unknown", and were not considered for further analysis. It was subsequently possible to simulate the effectiveness of a variety of potential session timeout durations based on the codes assigned to the 388 remaining inactive periods. Where i=the inactive period in the log, t= the proposed timeout duration, s="Same session" and d="Different session", we observe four potential outcomes:

1. $i>t, s$ = Incorrect session split
2. $i>t, d$ = Correct session split
3. $i<t, s$ = Correct session collation
4. $i<t, d$ = Incorrect session collation

Outcomes were calculated for each of the coded inactive periods in the logs sample (n=388) for cut-off times at 30 second intervals between 10 minutes and one hour. Figure 3-4 shows the distribution of these outcomes, with outcomes 2 & 3 aggregated.
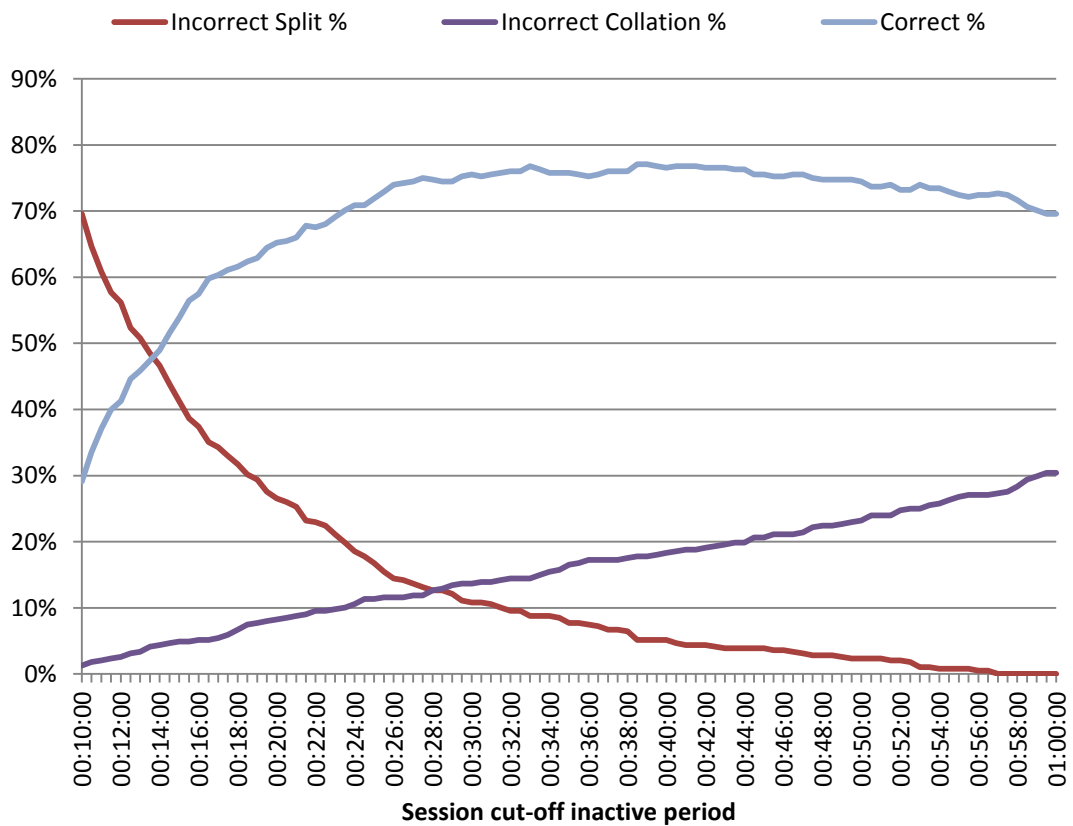
Figure 3-4: Effect of different session cut-off periods on session segmentation accuracy (based on manual coding of a sample of 388 inactive periods in the WorldCat.org logs)

A session cut-off time of 39 minutes was found to provide the highest proportion of correctly sessions (77.1%), although there was little variation in the proportion of correctly sessions between 26 and 57 minutes, with each cut-off time producing correct outcomes for over 75% of inactive periods. It is also interesting to note the distribution of the two error types. The results indicate that using a 10 minute cut-off time results in a high proportion of sessions (70%) being incorrectly split. Naturally as the session cut-off period is extended, an increasing number of sessions are incorrectly collated. A session cut-off time of 28 minutes was found to produce an equal number of the two error type (incorrect split = 13%, incorrect collation = 13%). Thus we can conclude that while session cut-off times of between 26 and 57 minutes have little effect on the overall accuracy of session segmentation, there is variation in the distribution of error types. A session cut-off time of around 28 minutes is shown to reduce the likelihood of one error type predominating.

Noting that this exercise was conducted on a relatively small sample of the logs, the results are perhaps best viewed as supporting the prevailing consensus that 30 minutes is the most appropriate session-cut off time. This cut-off period was duly used to assign unique session IDs to the full worldcat.org log data sets, and the final logs were found to constitute 15,799,727 sessions.

3.5.3.2.3   Identifying Actions and Referrers

The logs as provided by OCLC included a field representing the URL executed by the server. A key step in preparing the logs for further analysis was identifying the type of system interaction represented by each URL. The goal here was to produce a set of regular expressions that could be used to map each URL found in the logs to a list of defined system interactions. This was achieved by a three step process.

First, WorldCat.org was systematically explored to identify all possible system interactions, and the URLs those actions generated. These data were used to create a table of actions and example URLs. Second, a regular expression was created for each line of this table. These regular expressions were then used to interrogate the complete log data, assigning an action to each line in the log which matched a pattern represented by one of the regular expressions. Coverage was found to be only 35% (i.e. only 35% of lines in the log matched one of the regular expressions). Examples of URL types found not to match a regular expression were generated, and these were manually checked to determine the action they represented. A very large proportion of these were found to relate to servlet actions

involved in the generation of dynamic page content (e.g. links to retailers selling the item being viewed). Once actions had been assigned to URLS not previously matched to a pattern, the process was repeated until coverage was 100%. The resulting list of 83 actions can be found in APPENDIX 6: Actions Identified in WorldCat.org Logs.

Table 3-6: Classification of WorldCat.org Referrers

Referrer Type	Description
Search Engine	The referrer URL represents a web search engine. The final list comprised the following search engines: Google, Bing Yahoo, Yandex, Baidu, Sogou, Daum, Babylon, Delta-search, Ask.com, So.360.cn, Mysearchresults, Mywebsearch, and Searchmobileonline.
Library	The referrer URL represents a Library. This was captured using a regular expression to identify instances of a series of library related keywords within the referrer URL.
WorldCat.org other page	The referrer URL represents another WorldCat.org page. These might be part of the WorldCat identities service, or other pages with a worldcat.org url that do not constitute the catalogue itself. It is also likely that a number of sessions assigned this classification will relate to lines from the log relating to a single IP address that have been split into two or more sessions. The second of these sessions would appear to have a WorldCat.org referrer url.
WorldCat.org home	The session starts directly at the WorldCat.org homepage (i.e. the first page loaded in the session is WorldCat.org, with no other referrer URL provided).
Citation Service	The referrer URL represents a citation service (easybib.com, bibme.org, citefast.com, redlightgreen.com/org, or mendeley.com)
Goodreads.com	The referrer URL represents a GoodReads page
Wikipedia.org	The referrer URL represents a Wikipedia page
OCLC Services	The referrer URL represents an OCLC page
Other	The referrer URL is present in the logs, but does not map to any of the above categories
Not specified	The referrer URL is absent, or improperly formed in the logs. This most likely represents a web service or crawler that has blocked their referrer details.

Additional work was conducted to exploit the "Referrer URL" field included in the logs. Specifically a further series of regular expressions were developed to categorise the referring URL for the first line of each session, since this could be used to determine how the user accessed WorldCat.org. The development of the final list of referrers, and the regular expressions used to represent them, was through an iterative process whereby basic expressions were created to capture the most likely source of referrals, namely popular search engines and library domains. In order to expand this list, the logs were processed to identify referrer URLs that did not match either of these referrers. A list of the URLs not captured by the initial two regular expressions was generated, and sorted by occurrence. This was then manually analysed to classify the referrer type of any URL with more than 5,000 instances in the log. This resulted in the creation of 10 referrer categories (see Table 3-6), for which regular expressions could be created and applied to the logs. Each session could therefore be seen to originate from one of these referrer types.

3.5.3.3 Analysis

Analysis of the worldcat.org logs proceeded from general queries of the aggregated data, to more specific queries relating to the use of various system features and functionality, to manual coding of sample sessions. Initial analysis focused on general descriptive statistics of the aggregated data, specifically the calculation of average session duration, items viewed per session, and queries executed per session (mean, median and mode). This was supplemented by the production of four tables, each of which captured data about the use of each action identified in the logs:

Table 1: For each unique origin country identified in the logs, the number of instances of each action.

Table 2: For each unique Country identified in the logs, the number of sessions including **at least one instance** of each action.

Table 3: For each session referrer classification, the number of instances of each action.

Table 4: For each session referrer, the number of sessions including **at least one** instance of each action.

Analysis of these tables was conducted in Excel, and allowed for the identification of different usage patterns based on geography and referrer type. Initial analysis of these

tables suggested high volume of traffic to WorldCat.org, particularly sessions originating from search engine referrals, was engaged in only very limited interaction with the system. Averages (session duration, number of queries, items viewed) were therefore re-calculated for the subset of sessions that included at least one of the 13 actions representing the executing a search within WorldCat.org. It should be noted no analysis of user queries was conducted, despite such analyses (for example the number of query terms, instances of query reformulation, and most common search terms) being included in many studies of transaction logs. In this case it was determined that given the time and resources available for this phase of the project, the focus should instead be on understanding behaviour at a session and search-task level.

The final stage of analysis involved the manual coding of three sets of sample sessions. The intention here was to compare the types of session and behaviour exhibited by users interacting with the system having directly accessed the WorldCat.org homepage, with users whose sessions originated from Search Engine and Library referrals. Sessions originating at the WorldCat.org homepage were selected since it was clear from the overall log statistics that these users were more likely to undertake extended interactions with the system. Samples from search engine referrals were selected for analysis since they constitute almost 50% of the sessions in the logs, while samples from library referrals, as well as constituting a significant proportion of the logs (14%), were considered likely to represent interactions from academic users involved in search-tasks similar to those undertaken in institutional catalogues. In order to capture sessions that involve some level of system interaction, sample sessions that included at least one search action were extracted from the log. Four hundred random sessions from each of the relevant referrer types (WorldCat.org home page, Search Engine, and Library) were extracted. A sample size of four hundred for each referrer type was deemed sufficient based on precedents set in the literature relating to session classification (e.g. Jansen et al. 2008; Broder 2002). The main aim of the coding was to judge whether a session constituted a known-item or unknown-item search task, or some combination of the two. The criteria used to determine the type of search task was based on existing literature relating to known-item query formulation and detection. A number of authors have observed the frequency and effectiveness of known-item queries which combine author name and title (Slone, 2000; Kilgour et al., 2001). Kan & Poo (2005) highlight six characteristics of known-item queries that can aid identification. They posit that as well as being longer than topic search queries,

known-item queries are more likely to contain determiners ("the", "a" etc.), proper nouns, mixed cases, advanced search operators, and object identifying keywords such as "textbook" or "article". Since the analysis was conducted at a session rather than query level, it was also possible to identify occasions when the query terms precisely matched the title of an item subsequently viewed. Since the intention was to explore differences in session duration, number of queries, and item views between different task-types, care was taken not to use these measures as a means of judging tasks type.

After an initial inspection of the sample logs, seven codes were identified as suitable for classifying the sample sessions (see Table 3-7).

Table 3-7: Codes used for classification of sample sessions

Symbol	Description
K	Session consists of one or more known-item searches
U	Session consists of one or more unknown-item search
K→U	Session consists of a known-item search followed by one or more unknown-item searches
AU	Session consists of a query consisting solely of an author's name, and cannot be reasonably classified as known- or unknown-item)
ACC	The session solely consists of a user managing their WorldCat.org account
LIB	The session consists solely of the user accessing information about a library
NA	The session cannot reasonably be classified with any code

Further data about the sample sessions was automatically generated, including the number of queries executed, number of facets used to refine search results, and the number of clicks on related subject and author links. Attention was also paid to the number of items viewed. Two additional fields were automatically generated, representing the number of items viewed for each session (based on OCLC ID), and the total number of item views (i.e. including repeat views of the same item). As part of the manual coding process, a final field was included to capture the number of titles viewed. This last field was employed to identify occasions when users viewed multiple editions of the same work. It was also apparent from an initial inspection of the samples that many sessions consisted of multiple sub-tasks, the most commonly occurring example being a session consisting of several

distinct known-item searches. A further field was therefore added in which to record the number of sub-tasks observed within the session.

The coding process itself involved essentially "replaying" each session by following the URLs contained in the log, where necessary loading the page in a web browser to better understand the user's interactions. On completion of the coding, a random 20% of the raw sample sessions were extracted and re-coded by a colleague using the same scheme. Inter-coder reliability was calculated using Cohen's kappa coefficient, and found to be .89, above the .80 required to indicate reliable coding (Yardley, 2008).

3.5.4   Phase 4: User Study

Phase 4 of the research project had two principle aims: to evaluate users' interactions with WorldCat.org, and to test for the effects of recommendations on the information search process. The study design therefore drew heavily from the field of Interactive Information Retrieval (IIR). IIR methodology is perhaps best viewed in light of the once prevailing method of IR evaluation – the so-called Cranfield paradigm (Voorhees 2001). Such studies seek to evaluate IR systems in an isolated laboratory setting, thereby "freed as far as possible from the contamination of operational variables" (Cleverdon, 1991: 9). Thus users are removed from the evaluation process (with the potential exception of test collection development and relevance judgements), and instead system performance metrics such as precision and recall are collected in isolation. Kelly (2009: 9-14) locates IIR studies in the spectrum between these system focused studies, and the contextualised user-centred studies found in the field of Information Behaviour. Thus IIR research "comprises studies of people's information search behaviours, their use of interfaces and search features, and their interactions with systems" (Kelly & Sugamoto, 2013: 745). The results of these studies can help inform both theoretical models of information-seeking and search behaviour, and the design and development of systems better able to facilitate information access (Ruthven, 2008).

It is perhaps a reflection of this broad definition of the field that some authors bemoan the lack of "standardized methods and measures" (Kelly & Sugamoto, 2013: 746) for use in IIR studies. As Toms et al. note:

> "Complicating the research is the fact that the IIR process and its outcome are affected by many factors, including the knowledge and experience of the participant, the types of search tasks, the information retrieval engine, the type of

interface, and aspects of the interface in use. Thus, studies have intricate designs and require rich, varied data sets to assess the effect of the stimulus on the IIR process, and to add to our understanding of that process" (Toms et al., 2004: 656)

A review of the literature offering practical direction on methods of IIR evaluation confirms both the range of available approaches, and the multitude of methodological factors to be considered. As a starting point for IIR study design, Kelly (2009) identifies three potential goals for research. Exploratory studies are used to reveal more about a little understood phenomenon, while Descriptive studies aim to describe a phenomenon for benchmarking purposes. Explanatory studies seek to explore the relationship between two or more variables (Kelly 2009: 25). The precise methodology employed by an IIR study will therefore depend on the scope of the research questions, the types of systems under investigation, and the practical realities of the researcher (for example the availability of a usability laboratory, access to participants etc.).

3.5.4.1 Study Design

The primary goal of Phase four was to address research questions 3, 4 and 5, namely:

3. When might a recommender system support users of WorldCat.org?

4. What effect does the presence of recommendations have on the information seeking process in the library catalogue?

5. What recommendation characteristics would be most useful to users of library catalogues??

Results of the Phase 1 focus groups (see Chapter 4) revealed that students stated a preference for "Amazon-like" features in the library catalogue, echoing claims made elsewhere in the literature, while Amazon is also frequently cited as demonstrating state of the art use of item-level recommendations. Thus in addition to providing a benchmark for an evaluation of the current WorldCat.org system (Kelly, 2009: 27), a user study involving Amazon was identified as having the potential to both test whether the stated preference of users for certain features correlated with performance, and to examine the impact of item-level recommendations on the information search process. A further advantage of using Amazon as a point of comparison with WorldCat.org relates to corpus size. While a number of library catalogues that incorporate recommendations might have been chosen for the study, a major difficulty in interpreting the results of tasks performed on the two systems would have been assessing the impact of corpus size on the results. Since Amazon

and WorldCat both offer almost universal content coverage in terms of published books, it was felt that this potential variable was controlled, and that differences in specific data points (e.g. number of books found) could be related to factors other than corpus size.[4]

The study also sought to investigate differences in behaviour and performance between two related user-groups, namely undergraduate and postgraduate students. This study therefore also aimed to examine whether these groups exhibited differences in their interactions with and perspectives on recommendations.

3.5.4.2 Tasks

In order to address the research questions, a task-based methodology was adopted. As noted by Borlund (1997), task design can offer significant challenges to researchers. The development of tasks for this study drew heavily on findings from Phases 1 and 2, in particular the intersection of tasks typically carried out on WorldCat.org and those likely to benefit most from a system offering recommendations. Two tasks were therefore designed. The first was intended to simulate a key use of WorldCat.org as identified by research Phases 1 and 2 – namely a broad subject search, where the participant has some level of domain expertise. The second task was intended to explore the same type of subject search, but this time in a subject where the participant lacks domain expertise. This second task was developed in such a way as to incorporate a second distinct type of task – namely a known-item search. The two tasks were therefore set as shown in Table 3-98

Table 3-8 - User Study Tasks

TASK 1	Find a range of books that would be useful for your studies in a module you are currently taking. (If you are a PhD student, please search instead for books relating to a particular aspect of your research).
TASK 2	Imagine you have been recommended a book by a friend. The book is called [TITLE] by [AUTHOR]. a) Find out when the book was published b) Now imagine that you want to take the book out of the library, but all the copies are out on loan. Use the system to find a range of other books on the same subject that you could get out instead.

---

[4] It should be noted that a major difference in the print collections of the two systems is the presence of Theses, Journals and Conference Proceedings in the WorldCat.org catalogue. However since the tasks developed for this study were explicitly focused on books, this was not considered to be a significant difference.

Participants were allowed up to 10 minutes for the completion of each task. A within participant design was adopted, with subjects completing each of the two tasks on each of the two systems. In order to ensure that participants were not searching for the same books in both systems, which would undoubtedly generate a learning effect, the tasks were amended slightly to ensure that the tasks performed on each system would require different items to be found (see Table 3-9). A final consideration for Task 2 was the selection of the books to be used. A number of factors demanded consideration here. The books should relate to a topic where participants might be expected to have little or no knowledge, whilst not being so obscure as to unduly hinder the discovery of similar items. It was also necessary that there be a sufficient number of books on the subject available for discovery in each of the systems. It was determined that local history would provide a suitable topic, and some exploratory searching of both systems revealed sufficient numbers

Table 3-9: User Study Tasks with System Variations

TASK 1	SYSTEM 1	Find a range of books that would be useful for your studies in a module you are currently taking. (If you are a PhD student, please search instead for books relating to a particular aspect of your research).
	SYSTEM 2	Find a range of books that would be useful for your studies in a **different** module you are currently taking. (If you are a PhD student, please search instead for books relating to **another** aspect of your research).
TASK 2	SYSTEM 1	Imagine you have been recommended a book by a friend. The book is called [TITLE 1] by [AUTHOR 1]. a) Find out when the book was published b) Now imagine that you want to take the book out of the library, but all the copies are out on loan. Use the system to find a range of other books on the same subject that you could get out instead.
	SYSTEM 2	Imagine you have been recommended another book by a friend. The book is called [TITLE 2] by [AUTHOR 2]. a) Find out when the book was published b) Now imagine that you want to take the book out of the library, but all the copies are out on loan. Use the system to find a range of other books on the same subject that you could get out instead.

of resources relating to the histories of Birmingham and Leeds. The two books selected for the tasks were ***Birmingham: A History of the City and Its People*** by **Malcolm Dick,** and ***A History of Leeds* by W.R. Mitchell.**

3.5.4.3 Task Order

To equally distribute the impact of order effects, the sequence of tasks was determined by a Randomized Latin Square design (Kelly 2009). The variables were defined as shown in Table 3-10.

Table 3-10: Variables for Latin Square Design

Task	Task 1	Find books relating to your studies
	Task 2	Find books relating to a given title
System	System 1	WordCat.org
	System 2	Amazon.co.uk
Book	Book 1	History of Birmingham
	Book 2	History of Leeds

First, a matrix was created with each row relating to a different subject, and each column representing the chronological order of tasks. The first integer in each cell represents the Task, and the second integer (in brackets) the System. For each new row, values for the above row are rotated one cell to the left (Table 3-11):

Table 3-11: Basic Latin Square Design

Subject	Order of Tasks			
	1	2	3	4
S1	1 (1)	1 (2)	2 (1)	2 (2)
S2	1 (2)	2 (1)	2 (2)	1 (1)
S3	2 (1)	2 (2)	1 (1)	1 (2)
S4	2 (2)	1 (1)	1 (2)	2 (1)
…	…	…	…	…

As can be seen, this design does not satisfactorily control for the sequence of tasks (for example Task [1 (2)] follows [1 (1)] in 75% of rows). To address this problem, the order of columns was randomized. A random sequence generator (http://www.random.org/sequences/) was used to generate a random sequence of the numbers 1 to 4, and the order of columns in the Latin Square design was amended accordingly. The sequence given was 4, 2, 1, 3, thus as shown in Table 3-12:

Table 3-12: Randomized Latin Square Design

Subject	Order of Tasks			
	1	2	3	4
S1	2 (2)	1 (2)	1 (1)	2 (1)
S2	1 (1)	2 (1)	1 (2)	2 (2)
S3	1 (2)	2 (2)	2 (1)	1 (1)
S4	2 (1)	1 (1)	2 (2)	1 (2)
…	…	…	…	…

Finally, it was necessary to ensure that the Book assigned to the two iterations of Task 2 varied evenly across the systems. Thus the order was switched on alternate rows. In this Table, the second digit in brackets for Task 2 represents the book assigned to that task (Table 3-13):

Table 3-13: Randomized Latin Square with Book assignations

Subject	Task Order			
	1	2	3	4
S1	2 (2, 2)	1 (2)	1 (1)	2 (1, 1)
S2	2 (1, 2)	1 (1)	2 (2, 1)	1 (2)
S3	1 (2)	2 (2, 2)	2 (1, 1)	1 (1)
S4	1 (1)	2 (1, 2)	1 (2)	2 (2, 1)
…	…	…	…	…

The full randomized Latin Square design for the study can be found in APPENDIX 7: Full Latin Square Design for User Study.

3.5.4.4 Sampling and Participant Numbers

Since the aim of this study was to investigate the performance and preferences of two distinct sub-sets of the wider population (undergraduate and postgraduate students), as opposed to generating generalizable data relating to the entire WorldCat.org user population, non-probabilistic sampling was deemed appropriate. A combination of convenience and quota-sampling was therefore used. As defined by Kelly, convenience sampling is "relying on available elements to which one has access," while quota-sampling involves "dividing the population into more refined groups" (Kelly 2009: 66). Participants were drawn from the general student population of the University of Sheffield (convenience sampling), and volunteers were screened according to their level of study to

ensure an even split of Undergraduate and Postgraduate students (quota sampling). The recruitment process is described in detail in section 3.5.4.6.

The question of the appropriate number of participants for IIR studies so far lacks a definitive answer, despite an acknowledgement that "reliability, validity and efficiency are directly related to the number of participants … contained within the experiment" (Toms et al. 2004: 658). While single system usability studies require only very small numbers of participants (no more than five according to Nielsen (2000)), the complexity of IIR studies dictates more intricate experimental design, and therefore greater numbers of participants. Kelly & Sugamoto's review of published IIR studies found that "most studies had less than 30 subjects, with the plurality having between 11 and 20 subjects" (2013: 757). In determining participant numbers for this study theoretical considerations relating to the validity of resulting statistical analyses were balanced against practical limitations – specifically the time and expense associated with conducting the study, and the availability of the laboratory setting. The number of participants was therefore set at 36, with postgraduates (n=18) and undergraduates (n=18) each representing half of the total sample.

3.5.4.5 Data Collection Methods

The study design utilised three principle forms of data collection; Questionnaire, Logging, and Interview. The questionnaire has been defined as "any structured research instrument which is used to collect social research data … It consists of a series of questions set out in a schedule, which may be on a form, on an interview schedule on paper, or on a Web page" (Bulmer, 2004: xiv). Questionnaires constitute "a vital part of interactive IR studies since [they are] one of the primary vehicles for eliciting data from subjects" (Kelly et al., 2008: 123).   Four separate questionnaires were used, as summarised in Table 3-14. For the most part Likert-type scales were employed, wherein participants were asked to state their level of agreement with given statements, and 5 or 7 point scales were used to ensure a mid-point was available to participants (Kelly, 2009).

Table 3-14: User Study Questionnaires

Type	When	Description
Demographic	Start of session	Age, Gender, Education
Pre-Task	Prior to first task	Use of Library systems and Amazon
Post-Task	After each task	Perceived success of task Usefulness of system features
Exit	After completion of final task	Overall system preference Usefulness of Recommendations

The demographic questionnaire was designed to capture standard demographic information about the subject. The pre-task questionnaire was designed to collect data about participants' previous experience with the two systems. Since it was anticipated that most participants would not have any prior experience with WorldCat.org, they were also asked about their prior interactions with the University of Sheffield's OPAC, StarPlus. The questionnaires are included as APPENDIX 8: Demographic Questionnaire and APPENDIX 9: Pre-Task Questionnaire.

The post-task questionnaire was administered after the participant completed each of the four tasks, and consisted of two sections. The first examined the participant's perceptions of the system used for the preceding task, and their satisfaction with the quantity and quality of books found. The second section asked participants to rate how useful all available system features were in helping them complete the task. In order to aid completion of this section, illustrated screenshots of the systems were provided indicating each of the features and functionality referred to in the questionnaire. The questionnaire is included as APPENDIX 10, and a copy of the screenshots used as APPENDIX 11.

On completion of the final task, participants were asked the following question: "If you had to choose one of the two systems to find books on a particular topic, which would it be?"

In addition to the data collected through the questionnaires, the study design allowed for the collection of large amounts of interaction data through client-side logging. Client-side logging refers to the collection of system interaction data that "happens on a user's local machine via a client-side application", and is generally considered more robust and comprehensive than server-side logging (Kelly, 2009: 89-90). This was facilitated by the use of Morae Usability Testing software (http://www.techsmith.com/morae.html) to record

every task completed by the participant. In addition to logging all participant clicks, and queries entered, the software also allows for the recording of a video of the participant's computer screen during the session. Morae then allows the researcher to code the sessions with their own "markers," a process which can happen either in real-time (i.e. as the participant completes the task) or post-experiment. A total of 17 markers were defined for this study (Table 3-15). These included system interactions common to both systems (for example running a Search or Advanced Search), markers unique to Amazon (for example the use various types of item level recommendation), one marker unique to WorldCat.org (viewing a user-created list), and finally markers relating to the completion of the tasks. Here a distinction was made between items found from searches, and those found through the use of item-level recommendations.

The final data collection method employed was a post-session interview. Connaway & Powell (2010) observe that researchers must choose the type of interview to conduct "along the continuum between structured and unstructured" (216). Researchers should utilise structured interviews when the boundaries of their knowledge are clear, and they can therefore design questions to explore a well-defined area. Unstructured interviews, in contrast, are best used when the researcher in unclear about what is known and what is unknown, and must allow the subject the scope to inform the researcher (Lincoln & Guba, 1985). For this study semi-structured interviews were used, whereby a series of questions were designed to direct the interview, but the structure allowed for the investigation of interesting tangents that emerged during the course of the discussion. This is line with the view of semi-structured interviews as facilitating "'discovery' rather than 'checking'" (Denscombe, 1998: 113).

The interview questions were intended to better understand participants' views of the two systems, and to learn more about their strategies for completing the two tasks. They also sought to explore the factors that encourage the serendipitous discovery of resources, and to better understand user perceptions of the potential utility of recommendations. It was hoped that recent exposure to a system offering item-level recommendations, while undertaking tasks that were potentially aided by them, would encourage participants to offer more reflective perspectives on these questions.

Table 3-15: Morae Markers for User Study Coding

Type	Morae Marker Shortcut	Description
Markers common to both systems	S	Run a search
	A	Run an advanced search
	J	View next page of search results
	C	View an item from a search results lists
	E	Click on a "Related Subject" link
	G	Click on a Facet
	L	Use Look Inside / Preview feature
Markers unique to Amazon.co.uk	D	View an item from a 'Frequently Bought Together' Recommendation
	B	View an item from a 'People Also Bought' Recommendation
	V	View an item from a 'People Also Viewed' Recommendation
	K	View an item from another type of Recommendation
	M	View more Recommendations
Marker unique to WorldCat.org	H	View a 'User List'
Task-related markers	P	Get publication date
	R	Find Item from Recommendation
	I	Find Item from a Search
	W	Author Search
	X	Interesting

The complete interview questions were as shown in Table 3-16:

Table 3-16: User Study Interview Questions

1	In tasks X and X you were asked to find books that would be useful in your studies for one particular module. Describe how you went about finding these books. [Prompts: Did you try different search terms? What features of the systems were useful? What features were not useful? Did you find anything frustrating? How did you evaluate books you found?]
2	In tasks X and X you were asked to find books that could be used instead of a particular book. Describe how you went about finding these books. [Prompts: Did you do anything different to the other task? Did you try different search terms? What features of the systems were useful? What features were not useful? Did you find anything frustrating?]
3	Is it easier to find books on Amazon or WorldCat.org? Why?
4	Can you think of a time when you've been searching a library catalogue, and have found an interesting or useful item that WASN'T what you were looking for originally?
5	You will probably have noticed that the last part of the post-task questionnaire was focused on recommendations. Do you think recommendations would be useful in the Library Catalogue? [Prompts: Do you think they would help you discover resources? Would you trust these recommendations? If not, why not?]
6	Would you want personalised recommendations in the library catalogue?
7	What do you think makes a good book recommendation?

3.5.4.6 Implementation

Participants were recruited via the University of Sheffield Student Volunteers mailing list. An email was sent inviting volunteers to participate in the study (see APPENDIX 12). However only 3 responses were received. It was therefore decided that an incentive should be offered to participants in an attempt to boost volunteer numbers. The incentives were funded by the University of Sheffield Information School Information Retrieval research group, and were set at £10 per participant, and a second email was sent to the Student Volunteer list including the offer of the incentive (APPENDIX 13). This yielded more than 100 responses. Participants were assigned appointments principally on a first come first served basis, although the researcher ensured that the distribution of undergraduate and postgraduate students was even. A total of 40 students were offered appointments, with a further 10 placed on a standby list in case of no-shows and cancellations.
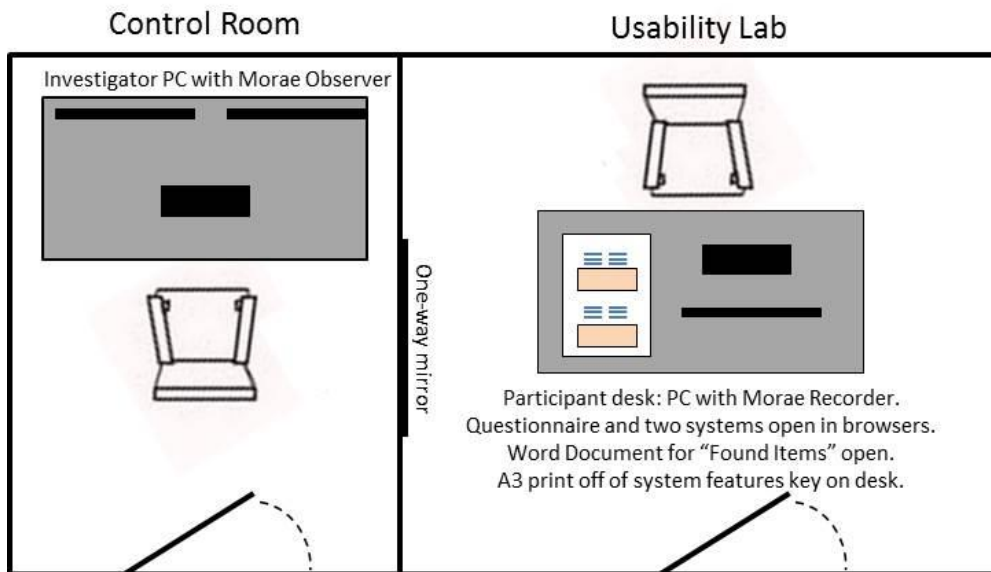
Figure 3-5: Layout of iLab and Experimental set-up

The experiments were conducted in the University of Sheffield's iSchool iLab (see Figure 3-5) between 7th and 30th May 2013. A comprehensive protocol for the study was created and used for all sessions. The pre-task, post-task and exit-questionnaires were administered using an online questionnaire tool called PyQuest, which was developed by Dr. Mark Hall at the University of Sheffield. Morae usability research software was used for task data collection. The Morae set-up used required three separate components:

- **Morae Recorder**: Installed on the participant PC, the software records all system interactions.
- **Morae Observer**: Installed on the control room PC, the software allowed the investigator to remotely start and stop Morae Recorder as the participant began and finished the tasks. The investigator was also able to view the participant's desktop to observe the tasks being completed.
- **Morae Manager**: Used for study set-up, data preparation and analysis, this component allows an investigator to code sessions recorded by the Recorder component. Use of this component is discussed in detail in section 3.5.4.3.

Prior to each session the browsing data on the desktop PC in the Usability Lab was cleared, and the system set up as follows:

- Firefox Browser window with PyQuest questionnaire at first page
- Firefox Browser window pre-loaded with http://amazon.co.uk
- Firefox Browser window pre-loaded with http://worldcat.org

- Word Document opened with a pre-created template for copying titles of books found during tasks (see APPENDIX 14: User Study Template Word Document)
- Morae Recorder opened and the window minimized.

On arrival, subjects were seated at a desktop PC in the Usability Lab, asked to read the information sheet, and if willing to continue asked to sign a consent form. The investigator then briefed the participant on what was required of them before moving to the iLab control room. In-built microphones and speakers allowed for two way communication between the rooms. Participants were left a paper copy of the Task instructions (see APPENDIX 15), and the Feature Screen Shot documents (APPENDIX 11) as a point of reference for the post-task questionnaires. Participants then completed the questionnaires and tasks as shown in Figure 3-6. While Morae offered the researcher the ability to code participant session in real-time, in practice this proved too difficult given the number and complexity of the markers. Markers were therefore added as part of the data preparation phase (see section 3.5.4.4)
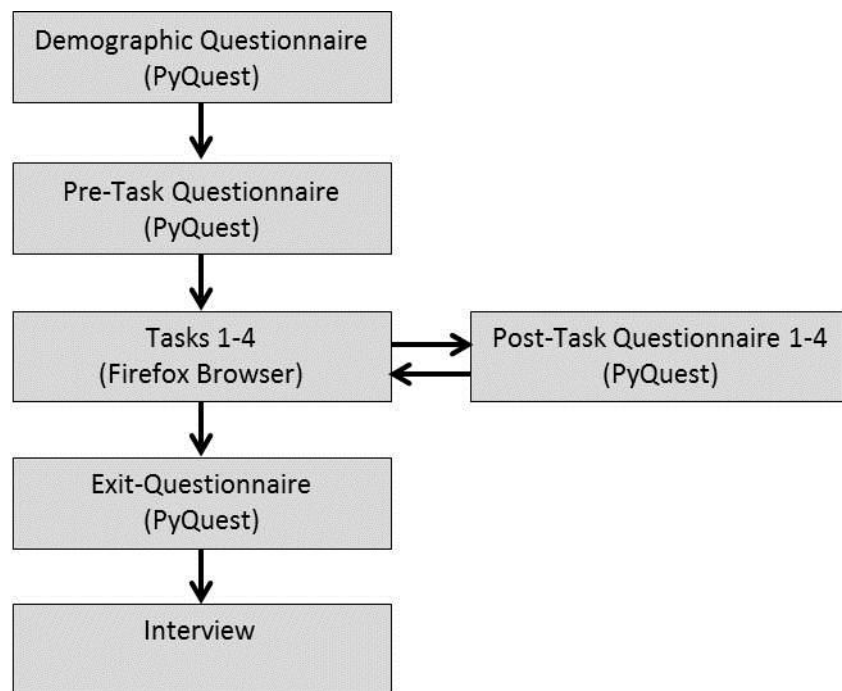
Figure 3-6: Flowchart of User Study

On completion of the tasks and questionnaires the interview was conducted with participants. Interviews were recorded using a digital Dictaphone, and extensive notes were taken by the investigator. After the interview participants were paid £10, and asked to sign a form confirming receipt of the money.

3.5.4.7 Data Preparation and Analysis

3.5.4.7.1 Morae and Questionnaire Data

The first stage of data preparation was to code the Morae recordings of participant sessions using the markers defined in section 3.5.4.1.4. This was done using the Morae Manager component and involved the researcher watching the video screen-capture of the sessions and assigning markers to match subjects' actions. Once complete, this data was downloaded from Morae in the form of a comma separated value (csv) file, and imported into Excel. In excel, some work was required to account for the variation in task order caused by the Latin-square experimental design.

Table 3-17: Session and Task variables for analysis of User Study

Symbol	Description	Formula
SR	Number of searches run	n/a
AS	Number of advanced searches run	n/a
NP	Number of times next page of search results viewed	n/a
VS	Number of times an item was viewed from a search results list	n/a
RS	Number of clicks on a "Related Subject" link	n/a
CF	Number of clicks on a facet	n/a
LI	Number of times the Use Look Inside / Preview feature was used	n/a
VF	Number of times an item was viewed from a 'Frequently Bought Together' Recommendation	n/a
VB	Number of times an item was viewed from a 'People Also Bought' Recommendation	n/a
VV	Number of times an item was viewed from a 'People Also Viewed' Recommendation	n/a
VO	Number of times an item was viewed from another type of Recommendation	n/a
VM	Number of times additional recommendations / next page of recommendations was viewed	n/a
VL	Number of times a User List was viewed	n/a
VA	Number of times an Author Search was run	n/a
GA	Time taken for user to obtain publication date	n/a
TD	Task duration	n/a
TA	Total Number of Actions	SR+AS+NP+VS+RS+CF+LI+VF+VB+VV+VO+VM+VL+VA
IX	Number of Items found from a Recommendation	FR+PR+NR
IY	Number of Items found from a search	FS+PS+NS
IF	Total Items found	IX+IY
TI	Time to find each item	TD/IF
RV	Total number of recommendations viewed	VF+VB+VV+VO
AI	Number of Actions per item found	TA/IF
AM	Number of Actions per minute	TA*(1 HOUR/TD)/60
TS	Total searches run	SR+AS
SA	Proportion of searches that are advanced searches	AS/TS
IR	Number of Items found from searches per search run	IY/TS
RF	Number of Items found from recommendation per recommendation viewed	IX/RV

The resulting spreadsheet showed a row for each participant, with the task and session level data shown in the rows. This was expanded by the calculation of a number of derivative variables, meaning the final data set included 48 variables. Table 3-17 shows all variables, and how the additional variables were derived.

All questionnaire data was downloaded from PyQuest, also in the form of a csv file, and uploaded to the same excel table. This table was then imported into SPSS for analysis. The Likert-type data was treated as described in section 3.5.2.4, i.e. as ordinal rather than scale data. Thus the dataset included a combination of parametric and non-parametric data, and careful consideration was therefore given to the appropriate statistical tests to use on different sub-sets of the data. Initial analyses also examined the distribution of scores for continuous variables. This was done through a calculation of Skewness and Kurtosis values, and the examination of histograms. These results of these tests informed the subsequent selection of parametric or non-parametric inferential statistical tools.

Chi-Square tests are applied to categorical data, and are designed to test for statistically significant differences in the distribution of mutually exclusive events between different groups. The result indicates the extent to which the observed data differs from an expected distribution. The test was used for comparing results between different subject groups (e.g. postgraduate and undergraduate students). Two further tests were employed to calculate levels of correlation between variables. Pearson's *r* is a parametric test designed to measure the correlation between continuous variables, and was therefore used in specific circumstances in the analysis where this was required (e.g. correlation between books read in the past year and books found during the tasks) (Kelly, 2009). Spearman's *rho* also measures the correlation between variables, but can be applied to both continuous and ordinal data (Pallant, 2010). Thus this test was used particularly for evaluating correlation between performance metrics and questionnaire responses.

To compare subject performance between the two systems, the Wilcoxon signed rank test was used. Considered a non-parametric alternative to the repeated-measures t-test, Wilcoxon signed-rank is applicable to non-parametric data and was therefore used to test for statistically significant differences in a range of variables relating to subject performance and system perception (Lazar et al., 2010). For comparisons between UG and PG performance, the Mann-Whitney U test was employed.

3.5.4.7.2   Interview data

Recordings of the post-session interview were transcribed, and analysed using the seven step Qualitative Content Analysis process described in section 3.5.1.2. Recordings of the interviews were transcribed and uploaded to NVivo, and a coding scheme inductively developed at a thematic level through a careful review of the transcripts. The coding scheme was tested on a sub-set of the interview data, and revisions made to the thematic categories and coding rules. The resulting data was then analysed to better understand user perspectives about the systems used for the study, factors influencing the serendipitous discovery of resources, and the role and preferred characteristics of recommendations within library catalogues.

3.6 Ethics

Each of the four Phases of the research process were subject to the University of Sheffield's Information School Ethics Review Process. This process involved the researcher submitting a comprehensive research proposal, along with copies of proposed supporting documentation (Information Sheets and Consent Forms). All submissions to the Ethics Review committee were approved, with only some minor alterations required for the Phase 4 study. These alterations related to the Information Sheet, and advised the following:

1. Refining the language to remove replace specialist terminology (e.g. "known item") with layperson's language
2. Providing greater clarity about the data to be collected (especially with regards to the use of Morae software)
3. Providing a breakdown of how long component parts of the study session would take.

The requested changes were made to the Information Sheet, and the revised submission was subsequently approved.

3.7 Phase Integration and discussion of combined data

While the research literature of many disciplines abounds with discussions of mixed-methods research design and implementation, it has been noted that far less attention has been paid to the theory and practice of integrating the results of multi-phase projects (Greene, 2007). Some authors have noted the prevalence of published works which claim to present integrated results of mixed-methods research projects, but which either fail to adequately assimilate findings from the attendant methodological strands, or do not properly discuss the techniques employed to achieve integration (Bryman, 2008; Woolley, 2009).

Perhaps the most commonly cited theoretical underpinning to mixed-methods integration is triangulation. The use of the term as a methodological concept dates back to the 1960s, when Webb et al., building on earlier work by Campbell & Fiske, noted that "the most persuasive evidence comes through a triangulation of measurement processes" (1996: 3). Initially this argument was most usually applied to quantitative forms of research, and therefore closer in spirit to the original meaning of the term as a surveying methodology involving the taking of multiple measurement readings. Later however the concept was popularised as a mixed-methods approach by Denzin (1970, 1978), who outlined four modes of triangulation: data triangulation (capturing data from diverse subjects at diverse points in time and space), investigator triangulation (the use of more than one researcher to collect data), theory triangulation (utilising multiple theoretical constructs to interpret the data) and methodological triangulation (using different methods to collect data). Of these by far the most influential in social sciences research is methodological triangulation, where it is frequently cited as a justification for and conceptual underpinning of mixed-methods research. A difficulty arises though in the extension of triangulation to a point where it is cited as a model for integrating qualitative and quantitative data. As Denzin himself has noted (2012), this interpretation is somewhat beyond the defined scope of his earlier work. The primary purpose of methodological triangulation is to use multiple data sources as a means of validating findings (Greene 2007), rather than a method of integrating complementary findings.

The component phases of this study do offer some limited potential for the utilisation of triangulation as a validation tool, most notably in combining and interpreting results of the survey and the transaction log analysis. The bulk of the assimilation of research strands, however, is guided by Bazeley & Kemp's metaphors for integrative analysis (2011). Their work combines ideas from throughout the methodological literature into a set of approaches to data integration, which they express as metaphors. These are presented in Table 3-18. The result is a loose framework of methods which the authors encourage researchers to interpret imaginatively. They further suggest eight principles for integration, which emphasise the flexibility required to maximise the outcomes of research, and the importance of integration being an ongoing and iterative process.

Many of the techniques described by Bazely & Kemps can be applied to the integration of results from the four phases of this research. The phase one focus groups provide a rich source of qualitative data against which to evaluate the quantitative findings from other phases. Participant quotes are used to illustrate points made in discussion of other strands of the research, and the understanding of user perspectives of the system can aid the process of making sense of the transaction log analysis results. The focus group data can also be transformed in quantitative data,

Table 3-18: Overview of Bazeley & Kemps metaphors for integrated analysis (2011)

Complementary Approaches	Description
Completion: Bricolage, Mosaics, and Jigsaws	Constructing a "patchy" aggregate based on the available data, or more carefully amalgamating all findings into a unified whole.
Enhancement: Sprinkling and Mixing/Stirring	Augmenting meaning by incorporating small data points, or mingling diverse but complementary findings together.
Detailing a More Significant Whole: Triangulation and Archipelago	Revealing unknowns through the combination of known points, or reveal a broader picture through snapshots of evidence.
Generative Approaches	**Description**
Exploration Through Transformation Involving Blending, Morphing, or Fusion of Data Elements	Developing new variables, or otherwise transforming or combining data
Conversation and DNA as Iterative Exchange	Re-assessing initial interpretations in light of subsequent findings, and identifying and linking "sense strands"

through the interpretation of code occurrences among different user groups. The phase two survey data, while relatively limited, provides quantitative data from a subset of users that can both be better understood in light of phase three findings, and in turn help draw out significant findings from phase one. The results of the transaction log analysis (phase three) help in the identification of key aspects of the phase one data, whilst themselves offering the opportunity for the development of new variables with which to examine data from phase four. The phase four findings themselves can best be viewed in light of information gleaned from the preceding phases, whilst offering an opportunity for validating both focus group and log data.

In general, all of the research questions are best answered through the amalgamation of findings from each phases, whether that amalgamation be formerly structured or a more creative patching together of disparate data. Naturally the research questions relating to the use of WorldCat.org are answered most fully by synthesising the results of the first three phases of research, revealing a broader picture than that offered by the findings of any single stage. Research questions relating to

the role and optimal design of a recommender system are answered best through an iterative process of re-assessing evidence from each phase, identifying key factors and charting their relationship to each other and to theory.

3.8 Conclusion

This chapter described the methodology used to address the project research questions. An introduction to mixed-methods research was given, along with a justification and explanation for the use of pragmatism as a philosophical context for the research. The overall multi-phase mixed-methods research process was described, and the chapter presented details of the design, implementation and data analysis of four constituent phases: WorldCat.org user focus groups, a WorldCat.org pop-up survey, WorldCat.org transaction log analysis, and an Interactive information retrieval user study comparing the Amazon and WorldCat.org. A brief summary of the ethics review process was described, Finally, the methods used to integrate the four research phases were described.

4. PHASE ONE: WORLDCAT.ORG USER FOCUS GROUPS

4.1 Introduction

This chapter presents the results of the focus groups involving four key Worldcat.org user-groups; librarians, students, booksellers and academics. The intention was for this phase of research to address research questions 1, 2, 3 and 5, namely:

> 1. Who is using WorldCat.org?
>
> 2. For what purposes are users accessing WorldCat.org?
>
> 3. When might a recommender system support users of WorldCat.org?
>
> 5. What recommendation characteristics would be most useful to users of library catalogues?

The methodology utilised is described in section 3.5.4. The results of the focus groups are presented and discussed in five sections, each relating to a major element of the code book that emerged from the quantitative content analysis process. These broadly correlate with the five general questions asked during the focus group sessions, although the coding scheme was flexible enough to incorporate answers to all questions within each category if relevant. While discussion of the results is principally conducted in the overall discussion chapter (Chapter 8), a brief discussion section highlighting the most significant issues to emerge from this phase of the research concludes this chapter.

4.2 Uses of WorldCat.org

The first question asked of participants in the focus groups was "Tell us about your experiences with WorldCat.org", with a number of follow up questions designed to explore participants' reasons for using the system. Responses to this question were found to generate responses at two levels – those describing work-tasks (i.e. the general activity prompting the interaction with the system) and those describing the search-task (the specific purpose of the interaction). The coding scheme that emerged during the qualitative content analysis reflects this distinction, and the majority of participant responses to the opening question were therefore coded with one of the codes shown on the next page. The following sections describe these responses, beginning with a brief summary of the work-tasks mentioned by participants as driving use of the system.

Section of Code book relating to uses of WorldCat.org[5]

1. Uses
   a. Work-Tasks
      i. Academic
         1. Essay / Assignment
         2. Research
      ii. Leisure
         1. Hobbies / personal research
         2. Reading for pleasure
      iii. Professional
         1. Acquisitions / Collection Development
         2. Cataloguing
         3. Inter-library loan (ILL)
         4. Instruction / Training
         5. Reading-list development
         6. Valuation

   b. Search-Tasks
      i. Institutional Information
         1. Location
         2. Policies
         3. Specializations
      ii. Known-item
         1. Bibliographic details
         2. Editions
         3. Format
         4. Location
         5. Holdings
      iii. Unknown-item
         1. Related
            a. Author
            b. Manifestation
            c. Similar item

         2. Topic
            a. Completeness
            b. Monitoring
            c. Multiple items
            d. Single item

---

[5] Please note that the order of elements within headings and sub-headings is organised alphabetically

4.2.1 Work-Tasks

The focus group participants described three broad contexts for using WorldCat.org; Professional, Academic, and Leisure. As might be expected, librarians and booksellers were the most likely to use WorldCat.org for professional purposes. Several of the librarians who participated in the focus-groups were cataloguers, and they spoke of using WorldCat.org as a means of establishing the bibliographic details of items they were required to catalogue for their institution. Booksellers described using the system for similar reasons; in their case adding book descriptions and metadata to their stock lists. It was interesting to note that in both cases WorldCat was often used to assist with the cataloguing of rare or esoteric material. For example:

> *"We look to derive records from OCLC to put into our cataloguing workflows. This is particularly useful for Asia-Pacific material."* (BL Librarian)

> *"I purchase from companies that publish the results of engineering projects, and these are not widely available publications. I can go to WorldCat to get bibliographic details, which makes cataloguing a lot easier."* (Auckland University Librarian)

> *"I used [WorldCat.org] recently to find out the illustrator details for a 1710 text I'd acquired."* (London Bookseller)

Other librarians, particularly those working on reference desks or in other patron facing roles, spoke of how they used WorldCat.org to assist students and faculty with Inter-Library Loan (ILL) requests, while others with responsibility for collection development and acquisitions explained how they used WorldCat.org as a source of data to direct their strategic buying or collection optimisation decisions. Again bookseller used the system for similar purposes, using the system to determine whether or not to acquire items. Booksellers also mentioned using WorldCat.org to assist in the valuation of rare items (*"to get a sense of relative rarity"* - London Bookseller). One academic also mentioned using the system during the process of developing and updating student reading lists. Finally, librarians involved in information literacy or other library training programmes mentioned their use of the system during training and instruction sessions for demonstration purposes. This last work-task can be distinguished from the previous three in that it incorporates no subsidiary search-task.

Several work-tasks were described by students and academics. All of the academics and several post-graduate students spoke generally of using the system to aid their research. The responses of undergraduate students to the question of why they accessed the system showed that it was almost without exception for the purposes of aiding a defined academic assignment such as an essay or presentation. While it was clear that most viewed WorldCat.org as primarily an academic or

professional resource, a small number of participants from all groups also mentioned using the system for leisure purposes, either as means of finding books to read for pleasure, or in support of their own hobbies.

4.2.2   Search-Tasks

While the work-tasks summarised above offer an interesting perspective on the broader contexts of users' interactions with WorldCat.org, the focus groups proved most useful in understanding the specific search-tasks users undertook on the system. Analysis of the transcripts led to the coding of search-tasks to be classified into three broad categories: those relating to Institutional Information, searches for Known-Items, and searches for Unknown-Items.

4.2.2.1 Institutional Information

A number of participants told of occasions when they had used WorldCat.org to ascertain information about libraries. Several librarians spoke of using WC to find out the address of a library, usually for the purpose of sending some correspondence. Students also spoke of using the system to find a library's address, although this was usually in order to facilitate a visit. Librarians also described using the system to determine other libraries' ILL policies. Several participants spoke of undertaking more sophisticated search-tasks on the system which were related to understanding individual library specialisations. Librarians tended to use such searches as way of staying up to date with collection development policies at rival institutions, and to gather information that might influence future collection development decisions. The only academic to mention this type of task explained that they were keen to understand which libraries would be most beneficial to visit:

> *"I sometimes use WorldCat to work out which libraries seem to specialize in certain types of material. That can be very useful as I can try and arrange a visit."* (Wellington Historian)

> *"[WorldCat.org]'s very useful as a tool for seeing which libraries are good in which subjects".* (LSE Librarian)

4.2.2.2 Known-Item Searches

As discussed in section 2.4.3, for the purposes of this study a known-item search is considered an interaction with the system wherein the searcher is seeking to locate in the catalogue the record of a specific expression of a work, about which some bibliographic data is known. As might be expected, the focus group participants described a wide range of search-tasks that required such interaction. Among the most commonly mentioned of these, particularly by librarians and booksellers, was the

task of determining the bibliographic details of an item. A number of variations of this type of task were described. Participants told of using the system to check bibliographic details as part of a standard validation process (*"We use WorldCat to verify if the bibliographic details are correct"* - Auckland Librarian), or confirming details about which the searcher had some doubt:

> *"I sometimes acquire rare pamphlets and other ephemera and I don't always trust the standard reference details that come with them."* (London Bookseller)

> *I'd catalogued a copy of Wilde's de Prefundis, it was a London edition I think. I'd used the card that came with the copy from the bookseller, but there were variant title pages from the same year. I couldn't understand why another publisher was coming up when I looked online. Eventually I found both editions on WorldCat and could know for sure which edition I had.* (London Bookseller)

A number of librarians also spoke of using the system to confirm a reference based on incomplete or incorrect information:

> *"I knew of a book that I wanted, but couldn't remember the title or the author. I could remember the editor of the book series though, so with a bit of creative searching I could find it in WorldCat and remind myself of the title."* (Northeastern Librarian)

> *"People sometimes come up to me on the reference desk and say they can't find a book they are looking for. Quite often I suspect they have the title or something wrong, so I can use WorldCat to find the correct details."* (Simmons Librarian)

> *"When I use WorldCat it's when people are convinced they've seen a book, but maybe only remember a bit of the title."* (BL Librarian)

Interestingly, although a number of librarians described occasions when they had used WorldCat to verify a reference given to them by a patron, no students mentioned using the system for this purpose.

Another very frequently mentioned known-item search-task was related to determining locations where a particular item is held. Students, librarians and academics all described situations in which they utilised WorldCat.org's "Find a Copy in the Library" function to ascertain which library or libraries held the item (*"It's a tool for locating things"* - Nottingham Historian, *"WorldCat is often the best option for locating a book outside the library"*, Northeastern Librarian). Some participants described using this service as a means of determining which libraries they could submit ILL requests to:

> *"Our colleague here is using it here for ILL for holdings."* (Nottingham Librarian)

> *"I was looking for a journal on autism, but couldn't find it in the stacks so just found it on WorldCat and ordered it through inter-library loan."* (Northeastern Student)

> *"When I search for a book on WorldCat my hope is that I can end up getting it delivered to Northeastern."* (Northeastern Student)

Participants also described using the service to check whether particular books were held in libraries that they could potentially visit in person, thereby saving the time required to arrange an ILL:

> *"I was trying to get a couple of books that were French literature but with English translation. I had work due on Monday and could not get it on the Friday. I went to Amazon and Blackwells, but they didn't stock it. Then I went to WorldCat and found it in a nearby university and a friend got it for me."* (Nottingham student)

> *"I had a grad nursing student looking for a book. I checked BLC [Boston Library Consortium] and nobody had it in Boston. It was available in Wisconsin though, the only place. She happened to be going there for thanksgiving! It was a great result, if very lucky."* (Northeastern Librarian)

> *"Our first suggestion is always ILL, but if people need things today we will go to WorldCat and say, oh X college has it, you can try there, or the public library or Northeastern."* (Simmons Librarian)

A Nottingham Historian also explained that there was value in knowing which libraries held an item even if there was no immediate opportunity to obtain the item: *"If it's something obscure, knowing it is in the US is still helpful. It means I know there's a copy out there."* (Nottingham Historian). In general though participants most valued the potential for locating a copy close by:

> *"I use WorldCat to see what titles I can get locally, places I can get to easily. Location is crucial."* (Sydney Librarian)

> *"I like the libraries close to you function, it's very helpful."* (Northeastern Student) Student)

> *"It's good to know if there is a library close by that holds a book … Using WorldCat is much easier than checking several different libraries."* (Nottingham student)

Another important use of WorldCat.org described by librarians and booksellers related to using the system to determine the number of libraries holding a particular item. For librarians, this was often spoken of as aiding decisions relating to acquisitions. Some librarians spoke generally about comparing their own collections to those of other libraries: *"Collection overlap is a key focus area"* (Melbourne Librarian). There was a strong sense here that knowing whether other local libraries held an item would influence the likelihood of acquisition:

"If several libraries nearby don't hold [a work] then maybe we should have it, to make available." (Auckland Librarian)

"If loads of places in Boston have it then I might not need to buy it." (Northeastern Librarian)

"I use WorldCat to see how well an item is represented in local libraries." (BL Librarian)

In other cases, knowing that other libraries held an item was considered an argument for acquisition:

"If lots of places have it then maybe we should too." (Simmons Librarian)

"If I see it has been acquired by a US library then I'm more likely to decide to get it." (BL Librarian)

Librarians also spoke of establishing the number of libraries holding an item in order to assist academic authors seeking to gauge the impact and sales of their publications (*"We check for academics who has bought their publications"* – Auckland Librarian, *"Sometimes I use WorldCat to measure impact – how many libraries hold a book written by one of our faculty members"* – Sydney Librarian). Other search tasks relating to library holdings were more concerned with determining the rarity of a particular item, either for valuation purposes, or to establish conservation requirements:

"The last time I used WC it was to check something in America and work out how many copies in the world there were of this thing. Turns out it was very rare, so that helped me assess its conservation needs." (BL Librarian)

"I like using WorldCat to prove booksellers wrong. They tell me that something is extremely rare, and I can show them that in fact lots of libraries hold it." (BL Librarian)

"I use WorldCat as a means of establishing rarity, and therefore price." (London Bookseller)

"WorldCat can be the only way to confirm if a particular item is valuable." (Northeastern Librarian)

The final categories of search-tasks described by participants related to different manifestations, expressions and items of a work. A number of participants described search-tasks in which they sought to identify all the editions of a particular resource:

"I had to list all the editions of certain texts. Very hard to know if everything has been covered. I typed the book into WorldCat and it brought up two or three more editions than I had." (Nottingham Historian)

"I was asked to find some resources relating to an Indonesian general, and this specific set of speeches he made, they're published in about fifteen different ways. I had three to start, by

the time I'd finished using WorldCat I had all fifteen. Without WorldCat it would have been extremely difficult." (Auckland Librarian)

"I recently did search for Catcher in the Rye, *wanted to know all the versions and editions. We were able to do a pretty complete search on WorldCat."* (Simmons Librarian)

Other participants were seeking a single specific edition of a work: *"I was looking for a specific edition of Moby Dick that I'd read about and knew had interesting illustrations. I was able to find it on WorldCat."* (Simmons Student). A bookseller also spoke of using the system to locate and evaluate individual copies of rare items: *"When I've found an item on WorldCat and found that my copy has slightly more plates or illustrations or something that's great because it means my copy is better!"* (London Bookseller).

Academics and students were particularly interested in locating electronic versions of a particular book, something made clear not only by their own comments *("I'm checking WorldCat to check if there's a digital version"* – Nottingham Historian; *"Quite often I go to WorldCat to see if there's an ebook that I can try and get access to"* – Northeastern Student), but also from the comments of librarians who had assisted them:

> *"Students are very interested in the format. They almost always want instant access, and feel electronic versions can provide that. If a student comes up to me at the desk and asks about an item that we don't have in electronic form, WorldCat is somewhere I can go to see what e-versions are out there."* (Waikato Librarian)

In summary, participants described a range of known-item search tasks that can broadly be categorised as seeking either the bibliographic details, locations, editions, format, or number of libraries holding an item. We will now proceed to a discussion of the various search tasks relating to unknown –items.

4.2.2.3 Unknown-Item Searches

As discussed in section 2.4.3, for the purposes of this study an unknown-item search-task is considered an interaction with the system where the searcher is seeking to locate in the catalogue one or more manifestations that offer some potential utility, without knowing the specific items in advance. Finding unknown-items emerged as an important use of the system. As one student put it: *"I think that's my primary use of WorldCat, to find things I did not know existed"* (Nottingham student). Analysis of the data generated from the focus groups revealed a range of unknown-item search tasks undertaken by participants on WorldCat.org. The first of these tasks relates to the identification of titles by a known author. This was spoken of by librarians, historians and students as

an effective means of discovering useful resources, and a task that they frequently used WorldCat.org for:

> *"WorldCat can be quite useful as a way of checking what other titles are written by same author. I do that quite a lot."* (Auckland Librarian)

> *"I use [WorldCat.org] to find new items for research by looking for authors that I know and seeing what else they have published."* (LSE Historian)

> *"I like the way WorldCat lets you easily find all the books by an author. That's really helpful and a good way to find things that could be useful."* (Nottingham Student)

> *"On a business course I was doing I used WorldCat to draw up a list of titles by authors that I knew were relevant to the course."* (Northeastern Student)

A point to note about these quotations is that in all cases the participant states that the specific goal of the system interaction is to identify titles by a given author. This can be contrasted with other instances of unknown-item search described by participants where this approach was used as a tactic in a broader topic search. For example:

> *"I was trying to find a load of stuff on a particular subject. I did a broad search and found that a load of the stuff that came up was by the same author, so I started to look for all the things that author had written."* (Nottingham Student)

The key distinction to be made here is whether the author is known to the searcher prior to the search session starting. We can therefore distinguish between identifying unknown-items by a particular author as a defined search task, and the same activity forming a sub task of a topic search.

Topic searches represented the most frequently mentioned form of unknown-item search. The typical approach to these searches was summed up by one student: *"I put in keywords and find useful things"* (Nottingham Student). Students and librarians frequently described situations where they used WorldCat to identify multiple items on a topic:

> *"I mostly use [WorldCat.org] to try to find initial sources of material for an assignment. I had to find sources about rescue helicopters and there were quite a few books about them on WorldCat.org."* (Simmons Student)

> *"I was doing an essay on speech therapy and needed to find a whole range of books and articles about all sorts of related things. I did a subject search for "Speech" and the results were really broad. I loved it because I could find a whole range of books that were useful."* (Northeastern Student)

> *"I use it for broad research on a subject."* (Nottingham Student)

Librarians also spoke of directing students seeking additional material on a topic to WorldCat: *"we often suggest WorldCat to students after they've used our own catalogue, particularly for topic searches"* (Northeastern Librarian). It was also apparent that for some participants, WorldCat was perceived as particularly useful for more obscure subject areas:

> *"I'd purposely use WC if I'd exhausted other major resources."* (Nottingham Historian)

> *"Good for obscure subject areas that don't bring up many results on other databases."* (Melbourne Historian)

> *"I had to find a framework that argues against a regulation but could not find anything about it anywhere. I tried google and loads of other places. I typed in 3 keywords to WorldCat and found a lot of items about it."* (Nottingham Student)

Sometimes participants described search-tasks that did not require the identification of multiple resources, but just one unknown-item. In these cases the searcher was most often looking for a single item on a topic that met some strict criteria relating to audience level or specific subject:

> *"A Professor wanted to read a story to his son's 2nd grade class. He wanted a book on kayaking suitable for 7 year olds. To maintain street cred I checked WorldCat and was able to find something appropriate."* (Simmons Librarian)

> *"For an assignment I wasn't sure whether I needed to include something about non-fiction reader-response theory. I searched on WorldCat to find a book or something that I could use to find out more about it."* (Northeastern Student)

> *"A student needed to find a book about maths to use in a kids' classroom. We looked together on WorldCat and found something that seemed appropriate."* (Nottingham Librarian)

Other examples of seeking a single unknown-item on a topic came as participants described another search-task they used WorldCat.org for: finding a similar item. Students described in general terms how they sometimes found it useful to try and find items that were similar to resources that had previously proved useful, and more specifically spoke of occasions when they had been required to find alternatives to a known item:

> *"Quite often I'll use [WorldCat] to find things that look similar to stuff that's been helpful in the past."* (Nottingham Student)

> *"If I need a particular book right now, but there are no copies left in the library, that can be an issue. So sometimes I'll see if there's anything I can find on WorldCat that I can get out instead from another library."* (Northeastern Student)

> *"It's frustrating when the things you want are out on loan. WorldCat can be really helpful for finding an alternative book."* (Simmons Student)

Other descriptions of topic searches related to finding everything available on a given topic. Librarians spoke of how PhD students and academics viewed WorldCat as an ideal system for ensuring the completeness of their searches. For PhD students this was often to make sure they had identified all the literature in their area, while for academics it was frequently related to ensuring nobody had covered the precise subject of their research:

> *"We use WorldCat with researchers who want to know if the subject of their research has been done before."* (Melbourne Librarian)

> *"I know that PhD students use WorldCat to trawl through and see what has been written so they can find everything about a subject."* (Northeastern Librarian)

> *WorldCat is a great resource for checking to see what others have done. You're checking to see if there is a gap there."* (Wellington Historian)

> *"WorldCat would be the clean-up. I might run a search to see if there is anything I've missed."* (Simmons Student)

Research students and academics also both told of another type of topic search for which they regularly use WorldCat.org, namely monitoring new publications on a topic. While again this is in essence a similar task to a standard topic search, a crucial aspect here is date of publication, with searchers familiar with the existing body of work and only interested in new additions to the corpus:

> *"I know there's not a lot of material about 18th century surgeon's logbooks, which are one of my research areas. So I do regular searches on WorldCat to see what comes out, and what's new."* (Wellington Historian)

> *"I use WorldCat as one way of keeping up to date with what's been published. I need to know if there's anything new in my area, so I'll do some searches every now and then to make sure that there's nothing recent that I've missed."* (Nottingham Student)

We have therefore identified from the focus group data a range of search-tasks where the information seeker is attempting to locate an unknown-item. Taken in conjunction with the set known item tasks identified in section 4.2.2.2, and the institutional information tasks related in section 4.2.2.1, there emerges a taxonomy of search tasks for which users employ WorldCat.org. This taxonomy will be further discussed in the final section of this chapter.

4.3 Perceived Strengths of WorldCat.org

The second question asked the focus group participants to describe a time when they used WorldCat.org that they considered a success. Follow up questions and prompts were designed to

explore the aspects of the system that contributed to that success. The responses to these questions and prompts provide for the basis of the next section, which describes the perceived strengths of WorldCat.org. In analysing these responses, a basic distinction could be made between participant comments relating to the design and functionality of the system, and the content of the WorldCat.org database. This is illustrated by the relevant section of the code book, which is presented below. This section will begin by presenting the strengths as identified by participants that relate to system content, before similarly describing positive perceptions of the system's functionality and design.

Section of the code book relating to perceived strengths of WorldCat.org

2. Perceived Strengths

    a. Content
        i. Articles
        ii. Different Editions
        iii. Different Formats
        iv. Foreign Language
        v. Full Text
        vi. Global Scope
        vii. Metadata
        viii. OAIster
        ix. Obscure Authors
        x. Obscure Items
        xi. T.O.C

    b. Design and Function
        i. "Find a copy in the library"
        ii. Citation Export
        iii. Ease of use
        iv. Easy to teach
        v. Filters and sorting
        vi. Interface
        vii. Reviews
        viii. Timeline
        ix. User Account

4.3.1 Content

One of the biggest strengths of WorldCat.org to emerge from the sessions was its content, particularly the global scope and comprehensive nature of the catalogue. The ability to access the

collections of thousands of libraries from a single system was understandably seen by librarians as WorldCat's principal selling-point:

> "[WorldCat] is giving us access to a world that we could not do on our own. You couldn't go library to library even if you knew 15 libraries with extensive collections, the time it would take." (Auckland Librarian)

> "The global aspect is its power." (Nottingham Librarian)

> "WorldCat really is a one-stop-shop." (Sydney Librarian)

The listing of multiple editions and versions of works was viewed as a major strength by some librarians, particularly as a means of establishing the identity of items they held, and for managing potential issues with duplicates:

> "I find the fact that WorldCat gives information on all editions, things like co-published stuff, I find that really useful because it means we don't end up buying duplicate editions." (BL Librarian)

> "An example is whether to catalogue certain items as books or serials. Increasingly publishers, philosophy publishers in particular, they're publishing their serials as books as well. All around the world people make these choices. If we've catalogued something as a serial and someone asks us for this book, then we don't know what it is til we look it up on WC." (Auckland Librarian)

While students spoke more generally about the successes they had had finding material through WorldCat (e.g. "WorldCat is very useful for finding material about French authors that are obscure" – Nottingham Student), and these successes were often a result of the
scale and reach of the WC database, students did not explicitly cite the size of the catalogue as a strength. This can perhaps be explained by the fact that students are far less likely than librarians to understand the aggregated nature of WorldCat. Students did however note the range of material available in the catalogue, particularly in terms of the different types of media for which holdings are listed, a view shared by other user groups:

> "I like that [WorldCat] gave me a lot of options like videos and music. We like to see things like that happen as students." (Northeastern Student)

> "When I was searching for my dissertation I got CDs and DVDs which were quite helpful. It's good being able to search for all that stuff in one place." (Nottingham Student)

> "It's great that records for microform items are available. I've introduced my students to it, our postgrads, because it could be very useful to them." (Wellington Historian)

Users of all types also spoke positively of the inclusion of Tables of Content for some items. While one historian mentioned the inclusion of serials and journal articles as a strength of the system, most participants felt that WorldCat was not the place they would go to access this type of material. A number of users were however pleased with the availability of links to full text e-versions of resources:

> *"I found a lot of e-books that I could access which would be good for my dissertation next year."* (Nottingham Student)

> *"Some things I found had links to full-text versions which I could read online. That's the perfect situation really; it would be great if there were more links like those."* (Northeastern Student)

> *"I really value the links to scanned copies of rare items held by the Ashmolean."* (London Bookseller)

As noted by one librarian, these links are frequently a result of the integration of OAIster[6] records, which several librarians felt was a positive development for WorldCat.

Librarians and booksellers also recognized the importance of the metadata available through the system: *"the depth of cataloging is beautiful"* (Auckland Librarian). While some participants spoke favourably of the standard of cataloguing and range and accuracy of metadata available (*"the system has the precise metadata missing from a lot of other catalogues"* – BL Librarian), it should be noted that this was not a universal perception (see section 4.4.1). As one librarian put it, *"WorldCat has opened my eyes to the crap cataloging some people do, compared to the beautiful cataloging of others"* (LSE Librarian).

4.3.2 Design and Function

Many participants spoke positively about the "Find a copy in the library" function, with a consensus that this is potentially a very valuable service. While some operational issues were identified (see section 4.4.2) the service was considered a unique strength of the system:

> *"Being able to see which libraries hold an item, out of the ones close to you, is amazing. It's the main reason I use [WorldCat]."* (Simmons Student)

> *"It really can be an invaluable tool for helping students and faculty get what they need. It's something that a lot of us here make use of all the time."* (Northeastern Librarian)

---

[6] OAIster is a union catalogue of open access digital material. Its records are harvested from compliant collections using the Open Archives Initiative Protocol for Metadata Harvesting. Since 2009 OAIster has operated in partnership with OCLC, which explains its integration with WorldCat.

The filtering and sorting functionality offered by WorldCat was also perceived as useful, and while a number of students and librarians noted that it was standard functionality for library catalogues, the consensus was that WorldCat did it well. Other features were also praised, with some users speaking highly of the timeline function, while the citation export service was deemed a particular strength by a number of participants:

> *"The citation export is good, means we don't have to type a bibliography the way we used to have to do in the old days."* (Wellington Historian)

> *"Being able to export citations, that to me is another selling point for WorldCat."* (Waikato Librarian).

> *"I use Endnote so being able to export citations straight to there is a big help."* (Simmons Student).

While very few of the participants mentioned creating a WorldCat.org account, those that had done so were positive about its benefits, particularly the students who had tried it, although most saw an account as primarily beneficial for non-academic use of the service. Participants who had taken time to fully explore the service before the focus group interview session were often surprised at the range of features on offer. The opportunity to select favourite libraries and save searches was particularly valued by some participants *("I have three favourite libraries and can search all three at once. Don't have to mess around"*, - Northeastern Student; *"Creating a profile has been one of the most useful things for me. It saves your searches so I can go back and repeat searches and find useful books again"* - Nottingham Student). Other features linked to creating an account, such as tags and reviews, were also mentioned. While it should be noted that such features were not seen as useful by other user groups (particularly librarians), for students they seemed to make for a more engaging user experience, although it was notable that very few participants mentioned using content such as reviews and tags that had been created by other users; instead they spoke positively about the opportunity to add their own tags and reviews:

> *"For the first time I'm actually excited to write a review for one of these things. I've never been inspired before, now I write reviews on WorldCat. I don't know what it is but it makes me want to get more involved, interact more."* (Northeastern Student)

> *"I really like these features. My local library in Minnesota allows you to create tags and that stuff, so it's good to be able to do it on WorldCat too."* (Simmons Student)

Many users, particularly students, also spoke positively about the general usability of the system, and the interface in particular. A clear discrepancy emerged between the perceptions of students

and librarians on this point, with librarians seeming to believe that students need an intermediary for WorldCat while students want to and believe they are capable of accessing it themselves:

> *"I feel like WorldCat requires an intermediary to figure out what happens on the screen. I think students will be confused and not know what to click on."* (Northeastern Librarian)

> *"I wish we had more time to train students on [WorldCat]. It's a complicated database and I'm not sure many students really know how to use it." (LSE Librarian*)

> *"It's very efficient, user friendly visually and also in terms of options."* (Northeastern Student)

> *"The interface in WorldCat was easier to use than the links that I was sent to by my lecturer."* (Nottingham Students)

In general then participants were relatively enthusiastic about the functionality and design of WorldCat, although only the "Find a copy in a library" function was considered something that would not typically be found in a modern library catalogue.

4.4 Perceived Weaknesses of WorldCat.org

In general the focus group sessions spent longer discussing the system's weaknesses than strengths, although this is perhaps unsurprising; as one participant put it, *"the only time I think about it is when it doesn't work"* (UK Bookseller). As with the discussion of system strengths, a distinction can be made between comments relating to system design and functionality, and those relating to the content of the system. In addition another category emerged from analysis of the focus group data, namely criticism of the marketing of worldcat.org as a service. The section of the codebook detailing system weaknesses can be found on the following page. This section will present the results of discussions relating to each category of perceived weaknesses in turn.

4.4.1   Content

The inclusion of multiple listings of what appeared to be the same item was mentioned by many participants as a weakness of the system. In some cases this was attributable to participants not understanding the differences between works and manifestations, with these users stating a preference for clustering similar items in the displays. Other users recognized the value of showing different manifestations, but complained that there were too many duplicate records (i.e. multiple identical records for the same edition), with librarians in particular raising this as an issue:

> *"The main issue though is the duplication of records – difficult to distinguish between items that we consider identical."* (BL Librarian)

"Duplicate records are the biggest problem, the system seems swamped at times with multiple records for identical items." (Wellington librarian)

"It's irritating when you're trying to work out how many copies there are of something. You find the same library is using multiple entries. So duplicates, they make the process more difficult." (UK Bookseller)

Librarians also felt strongly that there with issues with the accuracy of WorldCat's holdings data. This appeared to be based on experiences of locating an item in the catalogue, and

---

Section of the code book relating to perceived weaknesses of WorldCat.org

3. Perceived Weaknesses

   a. Content
      i. Duplicate records
      ii. Holdings
      iii. Metadata
      iv. Multiple editions
      v. Primary sources
      vi. US centric

   b. Function
      i. Citation Function
      ii. Dead links
      iii. Log-in
      iv. Interface
      v. Location
      vi. Navigation
      vii. Reviews
      viii. Search
         1. Ranking
         2. Results
      ix. Tags

   c. Marketing
      i. Existence of service
      ii. Range of services
      iii. Membership terms

---

noticing obvious omissions in holding libraries (i.e. the absence of libraries they knew to hold the item). While some librarians acknowledged that this could sometimes be a result

of delays by individual libraries in updating WorldCat with current holdings information, in general the perception seemed to be that such omissions somewhat damaged the credibility of the service:

> *"The holdings in WorldCat don't match Libraries Australia records. When I notice things like that it makes me less likely to trust WorldCat for other things."* (Melbourne Librarian)

> *"Sometimes I'm suspicious when [WorldCat] says the nearest version of a journal is in the Middle East – I'm surprised there is nothing closer because really I know there is."* (Simmons Librarian)

> *"Some items I knew they were in the University of Nottingham library but they weren't in WorldCat, which is frustrating."* (Nottingham Librarian)

In addition to inaccurate holdings, some librarians and booksellers also raised the issue of inaccurate metadata. Again it was acknowledged by several participants that this is not a fault attributable directly to WorldCat or OCLC. As one librarian put it: *"The metadata problems are down to input error by libraries, rather than as issue with database management per se"* (Auckland Librarian). Nonetheless participants were keen to point out the types of errors they commonly encountered, and these typically related to typographical mistakes, inconsistent cataloguing rules, and the absence of key metadata elements:

> *"Sometimes there are no subject headings, or other fields are missing, or some people follow other cataloging rules."* (Simmons Librarian)

> *"The detail in the records is sometimes lacking for non-English language items."* (LSE Librarian)

> *"Something I've noticed with cataloguing is that the errors are because someone has just copied another catalogue, including their errors."* (UK Bookseller)

> *"There are a hell of a lot of typos, I'm sometimes embarrassed that librarians make so many mistakes."* (Sydney Librarian)

Although some participants mentioned the lack of primary sources as a weakness of WorldCat (*"I find it frustrating that more primary sources aren't catalogued. I quite often need to consult old Australian newspapers and you really can't get at them through WorldCat.org"* - Melbourne Historian), the majority of the remaining complaints about the system related to its perceived US centricity. While this view was perhaps influenced by the issues with the functionality of the "Find a copy in a library" feature outside the US, there was a sense too that the perceived high proportion of contributions from North American libraries could make it a less effective system:

> *"The coverage is very US centric. I find that if someone is looking for something specific to New Zealand then it probably isn't going to be in there."* (Waikato Librarian)

> *"Sometimes I'll end up at WorldCat but it is very North American focused – sometimes that's useful, but at times it can feel limiting."* (Nottingham Historian)

It is important to note in these discussions, as some librarian participants acknowledged, that the accuracy, range and scope of both holdings and metadata is almost entirely dependent on the libraries contributing to WorldCat. Each record accessible through WorldCat.org has been catalogued and uploaded by a library. While OCLC do attempt to monitor and correct issues such a duplicate records, the rate at which the catalogue grows makes this an almost Sisyphean task. Although one might argue that the global reach and coverage of the catalogue is a function of the success OCLC have in attracting non-US libraries as contributors, it must be noted that OCLC (and by extension WorldCat) began as US institutions. IT is therefore perhaps understandable that a relatively high proportion of records originate from North American libraries. The fact that over 50% of WorldCat holdings are now non-English language suggests that considerable progress in this area has been made.

4.4.2   Design and Functionality

A common complaint mentioned by participants related to issues with the search functionality of WorldCat.org. These complaints fell broadly into three categories: issues with the ranking of search results, the inclusion of non-relevant items in result sets, and the sheer number of results returned. As participants put it:

> *"Searching is problematic. In a normal catalogue, entering the title of a book in a search will pop it up at the top of the results list. In WorldCat the book title will not appear, it'll be buried in a list of millions of things"* (Northeastern Librarian).

> *"If you type in a search, some really random stuff comes up. It's really weird."* (Nottingham Student)

> *"Searching can get confusing, it's hard to know where it goes wrong sometimes."* (Simmons Student)

> *"Often stuff is in there but you wouldn't know it from the results that come back."* (LSE Librarian)

> *"Sometimes your search can be swamped with articles when you're looking for a book, but that is not what you use it for."* (Northeastern Librarian)

> *"Sometimes it's hard because there is so much stuff, it can be overwhelming."* (Sydney Librarian)

It was noticeable that in general librarians tended to demonstrate a more sophisticated understanding of the challenges of delivering an effective search system, and were more likely to

utilise technical terminology in their analysis of the problem (for example *"Relevancy ranking is the big problem for these web-scale systems"* – Northeastern Librarian). Students, in contrast, were more likely to simply be confused by the results, and generally couldn't understand why the relevance of results failed to match those they experienced on web search engines.

While the overall performance of the system was generally perceived to be good, some UK-based students noted minor difficulties logging into their account, and issues with the citation export function. Participants also mentioned encountering dead-links, these being most often links from WorldCat.org to member library catalogues. These participants also spoke of difficulties returning to search results after following broken links:

> *"I followed an item link to a library and ended up with a screen basically saying 'this is where it would be if it existed.' Not sure how that's supposed to help me."* (Simmons Student)

> *"The number of links that don't work, or go to wrong place, is huge problem."* (Simmons Librarian)

> *"I clicked on a link and it transferred me to the Minuteman catalogue, but I got an error message. Then when I tried to go back I somehow lost all my search results and had to do the search again."* (Northeastern Student)

For users outside the US, the "Find a copy in a library" function often appeared frequently to perform poorly, with the issue apparently related to the system's inability to properly recognize non-US zip codes. Several participants who have used the system in both the US and abroad noted that the system is better suited to American users:

> *"If I use it here it rarely works, and when I used it in Lebanon it was even worse. But the times I have been in the States it seems to work really well."* (Nottingham Historian)

> *"The UK postcodes are utterly and completely wrong."* (London Bookseller)

Other comments focused on the design and functionality of the WorldCat.org interface. Some participants felt the interface could be cleaner and simpler, although it should be noted that other participants praised the interface for these very attributes. A significant issue that emerged related to the design of the search results page, which some participants felt made it difficult to identify and evaluate different resources:

> *"The first screen of a result is the title and author, then it's broken up by broad blue lines, and then there is another bit of info, then more lines, then finally the bibliographic data. I guess it's a UNIX program, it looks very out of date and is slightly difficult to use."* (UK Bookseller)

"I think it's very research-looking. I don't think it looks like a fun web site" (Nottingham Student)

"I have a hard time telling difference between records. Have to click on each one, find out who has copies, it's all very slow. You cannot tell the difference between records at a glance." (Northeastern Librarian)

"People can't tell the type of item from the initial search result. So they don't know what to expect." (Sydney Librarian)

Perspectives on some of the next-generation catalogue features incorporated into WorldCat such as tags and reviews were also negative in many cases. These perspectives seemed in some cases influenced by a vision of the purpose of the catalogue, which reviews and tags might serve only to confuse, and by a mistrust of the subjective nature of these forms of user-generated content. Some librarians were vociferous in their assessments:

"I'm starting to feel about social media like I used to feel about genealogists, like they just get in your way. That's why I probably won't engage with tags or reviews until I get over this. All those puffballs who venture forth with their opinions, and you don't want to know." (Wellington Historian)

"The reviews are totally subjective and I'd rather just turn them off completely." (Waikato librarian)

"It gives me a negative feeling actually when I look at WorldCat and see tags and reviews and what have you. If I wanted that I'd go to Amazon. I just want the hard info." (BL Librarian)

"I don't like the tags. It's literal thinking and I suppose a background as a librarian does have a very different implication … I think of cataloging and description as a precise, targeted description ... People tagging things see labels used by librarians as being approximate, multifunctional, amorphous. When someone isn't using tags in a precise and descriptive way, it can be more work than it's worth." (Auckland Librarian)

Students in contrast seemed much more likely to see the potential benefits of tags and reviews, and instead found the lack of actual content the major weakness of WorldCat:

"Tags can sometimes be helpful for finding stuff again, and I will occasionally look at reviews if any are available. It's sometimes useful to know what other people think, it can save me wasting my time. I don't ever really see any in WorldCat though." (Nottingham Student)

"Sometimes tags can be useful, like if you're searching for stuff that doesn't have very good subject headings like 'race'. It would be good if WorldCat had more tags like this." (Simmons Student)

> *"It's good in principal, but I think everything I've ever looked at in WorldCat has said zero reviews, and I don't really remember using any tags."* (Northeastern Student).

While they spoke of the value of such content, very few students admitted to creating it themselves. Principally this was down to the time and effort required to add reviews and tags (*I'm generally in a hurry and don't have time to tag"* – Simmons Student), something exacerbated by the requirement to be logged in to a WorldCat account in order to do so.

One final weakness of the system mentioned by participants was a perceived lack of support for browsing. Students and academics spoke of frustrations associated with topic searches, in which they were felt there was little option but to try multiple different search terms in the hope of bringing up fresh results. The process of finding and evaluating the relevance of items was perceived as laborious by some, and seemed to compare negatively to other systems they experienced on the web:

> *"I do try browsing WorldCat but it's just too big, it takes so long to find things and I just end up trying to think of different ways of searching for the kind of stuff I want."* (Simmons Student)

> *"I wouldn't advise students to do topic searches on WorldCat. It's just too big and it doesn't really support exploratory searching."* (Northeastern Librarian)

> *"When I'm looking for stuff online I hardly have to do any searches, I can just follow links. You can't really do that in library catalogues; you just have to do searches and then go through loads of pages of results."* (Nottingham Student)

It is significant that these comments were often linked to suggested improvements to the system (see 0).

4.4.3   Marketing

A final challenge to emerge from the focus group sessions relates to the marketing of WorldCat.org. Perhaps understandably, this surfaced most strongly in the sessions outside the US. Two principal issues were mentioned – a lack of awareness of the existence of the service, and a lack of understanding of the range of services available. Librarians also spoke of not understanding the scope of the system, or how best to use it. Students in particular were frustrated that more had not been done to introduce them to the system, with some seeing to criticize library staff for not bringing it to their attention:

I'm a bit annoyed that I did not know about WorldCat sooner. I did not want to use Google and went to see a librarian and she never told me about WorldCat, I ended up just stumbling across it from a Google search or something." (Nottingham Student)

"It's a shame that no one told us about [WorldCat], you'd think librarians would be encouraging us to use it." (Northeastern Student)

The difference in perceptions of the service between librarians and students (as described in section 4.3.2) offers one explanation for this, with librarians apparently viewing the system as unsuitable for disintermediated student use.

4.5 Serendipity

The fourth question asked during the Focus Group sessions asked participants to describe a time when they had found an item in WorldCat that wasn't what they were originally looking for. Although the intention of the question was to explore instances of serendipity in the catalogue, in practice participants tended to interpret the question differently, and generally recounted times they had found useful items that were closely related to an original search (for example: *"A faculty member wanted something, we didn't have it so looked on WorldCat and ended up finding newer items that were actually better"* – Northeastern Librarian). Some participants also described sessions where they had followed citation chains or linked subject headings, and discovered useful items that way. After further prompting, participants still generally found it difficult to think of occasions when serendipitous discovery had occurred, and even those who had experienced it struggled to remember specific examples:

"Serendipity sometimes happens. I can't remember a specific example but it has

happened, and that's partly what keeps me coming back!" (Nottingham Historian).

It is questionable whether any of the examples cited by participants above could truly be called serendipitous. Recall that serendipity requires the act of discovery to happen by chance, and the information discovered to be relevant. In the case of finding titles closely related to an original search target, one might argue that little chance is at play in formulating search terms that lead the searcher directly to the new item. Similarly the act of citation chaining, while theoretically having the potential to lead to serendipitous discovery, also represents a formal (and taught) mode of resource discovery. The discovery of relevant information after utilising such a tactic perhaps owes more to sound searching strategy than chance. There was little sense in the examples given by participants of being surprised by their discoveries.

Interestingly several participants did refer to a more emotionally charged form of fortuitous discovery, but these examples took place not within WorldCat (or indeed any catalogue), but among physical collections of books. In these cases participants described chance encounters with potentially useful items as they moved amongst the shelves of libraries and bookshops:

> *"In the library or a bookshop, as I'm walking around I'll quite often see interesting stuff that piques my interest, stuff that's completely unrelated to the reason I went in there but that suddenly grabs my attention."* (Nottingham Student)

> *"If I have time I will wander round the physical library. I wouldn't go to a database for surprises, but you get that kind of thing happening when you look at the actual shelves."* (Northeastern Student)

4.6 Suggested Improvements to WorldCat.org

The final question addressed to the focus groups asked participants to imagine their ideal WorldCat system, and the features and improvements it would have. As might be expected, many of the suggested improvements can be linked directly to perceived weaknesses of the system, and in developing the code book for responses to this question it became apparent that the same broad classifications could be used, namely Content, Design and Functionality, and Marketing. The full coding scheme question can be found on the following page. It should also be noted that a number of suggestions for improvement involved features that are actually already present in WorldCat. These suggestions have been included in the presentation of the results that follows, with the existence of the feature noted.

4.6.1   Content

From a content perspective, improving the quality and scope of the metadata was suggested by a number of librarians (*"I suppose if we really had a magic wand we'd want 100% totally accurate records"* – LSE Librarian; *"Fuller bibliographic details would be ideal, if it had absolutely comprehensive metadata for every item"* – Nottingham Librarian). It was also noted by users across all user groups that the holdings data could be improved, both in terms of its accuracy and universality; as one user put it, WorldCat could be *"the OPAC of OPACs, so everything comes up from every library"* (London Bookseller). Several librarians felt that merging or removing duplicate records would offer an immediate improvement to the system. Another suggestion made by several participants was that the currency of holdings information could be improved, while others

Section of the code book relating to suggested improvements of WorldCat.org

5. Suggestions for Improvements
   a. Function
      i. Customization
         1. Favourite libraries
         2. Interface
         3. Notification of new items
      ii. Hyperlinking
      iii. Recommendations
      iv. Search
         1. Highlight search terms
         2. Ranking
         3. Sorting
         4. Spelling Variations
      v. Work level display
      vi. WorldCat Local functionality
   b. Content
      i. Full text
      ii. Granularity
      iii. Holdings
         1. Accuracy
         2. Item availability
         3. Universal
         4. Up-to-date
      iv. Merge duplicate records
      v. Links to related information
         1. Amazon
         2. Antiquarian Booksellers
         3. Author pages
      vi. Metadata
         1. Accuracy
         2. Completeness
      vii. Supplementary information
         1. Authoritative reviews
         2. Book covers
         3. Popularity metrics
         4. Ratings
         5. Summary
   c. Marketing
      i. Clarify membership terms, privileges and contributions
      ii. Encourage word-of-mouth advocacy
      iii. OCLC representative visits
      iv. Training podcasts

expressed a preference for WorldCat.org to make them aware of item availability *("Holdings and availability in the institution, that could be made available through WorldCat"* – BL Librarian). Several participants also discussed the possibility of greater content granularity, with comprehensive records and metadata for book chapters (or even sections of chapters), and journal articles:

> *"In terms of long-term wishes we want every article title in every journal, and every chapter in every book having its own record, or at least its own metadata. I think content should be discoverable at a much finer level."* (BL Librarian)

Increased access to full-text was another frequently mentioned potential improvement, although many seemed conscious of the difficulties in realizing this improvement. While some participants particularly embraced the idea of having a magic wand in this regard (*"It should give us full text access to everything ever"* – Simmons Student), others were more realistic:

> *"It's the biggest turn off to students that they cannot get full text. Having a large amount of full text … e-books today are important. Having e-books in there, even if there are limits on the number of people using it, would be massively valuable."* (Waikato Librarian)

> *"The missing pieces of the pie are e-resources. Obviously there are huge issues with implementing it but the more access you can give users to full text online the better."* (BL Librarian)

A range of suggestions related to the incorporation of supplementary information. Students in particular described how useful book-covers can be beneficial, both in getting a sense of the item's characteristics and matching the experience of sites like Amazon. While WorldCat does include book-covers, these are not universally present. Other participants suggested summaries (*"I would like additional information. Not reviews, more helpful would be more info about content, a decent summary for example"* – BL Librarian), and authoritative reviews (*"I find a review really important in scholarly situation, as long as it's a proper serious review"* – Nottingham Student). Other ideas for supplementary information such as ratings and reviews represent features already present in WorldCat, although there were some ideas for refining this content: *"It would be good to have star ratings and to have an average and to separate by students and others"* (Nottingham Student). Another suggestion related to an alternative means of assisting resource evaluation; the addition of data showing the popularity of items: "You could show how many people had accessed a record or borrowed the item recently, that would help us know how popular something was" (Northeastern Student). Participants also suggested supplementing record pages with links to related information such as author web pages, and links to vendor such as Amazon (these links are already present), or if more appropriate antiquarian booksellers.

4.6.2 Design and Functionality

A number of suggestions related to the customization of WorldCat.org. These included the ability to turn certain features on and off, set automatic notifications of new items, and allow for the personalization of the interface. The selection of favourite libraries was also mentioned (something already available to users with a WorldCat.org account), and the idea of filtering the database geographically was also taken up by participants who suggested the incorporation of WorldCat Local functionality (often without using that terminology), specifically the automatic filtering of search results by institution and location:

> *"When people think of discovery systems they think of them as the same system or an integrated front end. It's hard to talk about a discovery system decoupled from your local catalogue. If WorldCat could allow users to limit their search to local institutions then that could be hugely beneficial."* (Northeastern Librarian)

> *"Say I'm searching on the LSE campus and access WorldCat. It would be great if it said 'I'm going to default search holdings to LSE first.' So it would only give you LSE results first, then other locations nearby. So confined first, then lets the user make a wider search"* (LSE Librarian)

> *"You could use like mileage, show the closest stuff first."* (Nottingham Student)

Other improvements to the general search function were suggested, including better handling of spelling errors and variations, but primarily to ensure more relevant and transparent ranking of results:

> *"WC should highlight the keywords that I search so that we know what fields its searching since we don't know. It gives people a bit of confidence in your database."* (Waikato Librarian)

> *"If we could say to someone wanting to start a search 'type keywords into WorldCat and the very best resources will appear at the top' that would be great."* (Simmons Librarian)

> *"Searching just needs to work better. If you enter the title of a known item then that item must come at the top of the results."* (Northeastern Librarian).

Improvements were also suggested for the search results display. Some participants felt that the system should more clearly distinguish between different types of media, and that this could be achieved by the use of clear icons representing different resource types. Participants also spoke often about the issue of seeing multiple editions of the same work in their search results. Students and librarians shared similar perspectives on how this could be improved, although librarians were able to articulate their ideas using FRBR terminology:

"I had to go through a lot of different publications for the same book. I get that it's useful to have several editions since we may need a specific one, but we should be able to search by title and not have to see all the different versions." (Nottingham Student)

"There's got to be some way of simplifying the search results so you don't see 15 versions of the same book. I don't really care about all these editions; I just want to find out who has a copy." (Northeastern Student)

"When you are looking from an ILL perspective, it would be good to find one record for the work, then see another list showing which libraries hold the item." (LSE Librarian)

"The ideal solution would be to use the work for search purposes, then let the user drill down to the manifestation, expression, and then the item." (BL Librarian)

"I think moving to the work level would make it much easier, it would really streamline things." (Simmons Librarian)

Other suggested improvements included the integration of full ILL capabilities, and the increase of working links between WorldCat and individual library catalogues. The remaining significant suggestion was the potential role of recommendations. It should be noted here that while this project focusses on the potential role of recommendations in the catalogue, this was not made clear to participants in advance. Recommendations were not mentioned by the moderator of focus groups, although some additional questions were asked of participants once the subject had been independently raised. All the student focus groups and several librarian focus groups included such discussions.

In some cases discussion of recommendations was prompted by recognition of the limitations of the existing "Find Similar Items" feature, and Subject Heading functionality:

"The link is called 'similar items'. You expect to see book covers, recommendations basically, not subject headings." (Simmons Librarian).

"When you're within a subject area, it would be good to have other titles to look at. You can click on the subject headings themselves but I'd like to see actual related titles displayed that you could click on." (Nottingham Librarians).

Some students proposed recommendations of a similar type to those they had encountered elsewhere online, particularly Amazon:

"Basically like those Amazon recommendations on different products I guess, the ones that give you different options from a product page." (Northeastern student)

"I use recommendations on Amazon a lot, the ones that say "people also bought this", I use them for leisure and academic reading. I think a similar thing on WorldCat could be really helpful." (Northeastern Student)

Participants generally described three forms of recommendations that they perceived as potentially useful. Several spoke of the usefulness of what might be termed expert recommendations, and wondered if there was some way of incorporating these into the system: *"Basically the people whose opinion I trust most are professors, so the ideal thing would be to get suggestions about what they think I should be reading. But I've no idea how that would be possible"* (Northeastern student). Others spoke of systems similar to the familiar Amazon recommendations that could suggest content based on what other people had clicked on or borrowed:

"When you are looking at a known item, you could see recommendations based on what other people had taken out or looked at. A bit like "people who have also bought" – that's very helpful." (Northeastern Librarian)

"What other people clicked on would be useful, that might be a way of discovering new resources. Also subject related recommendations. Different users might want different types." (Simmons Librarian)

As shown in the last example, some participants also saw the potential value of subject related recommendations. The consensus here seemed to be that these recommendations should be very closely related in subject to the item being viewed:

"It would be useful if you searched one specific book and it comes up and then if you received information about similar books on the same subject, that would be good." (Nottingham Student)

There were also discussions in several groups about not just the type of recommendations, but more generally why they would be useful. Among students in particular there was a sense that this kind of feature is standard on the web, and they expected more ways of exploring systems than simply searching:

"Recommendations would give me more options, they'd mean I wouldn't have to just keep doing searches." (Simmons Student)

"It's really difficult to just browse a system like WorldCat, it's really difficult to just click on things. I reckon recommendations are just a way of exploring." (Nottingham Student)

It should also be noted that certain user-groups were much less enthusiastic about the prospect of recommendations being added. In librarian focus group sessions, it was noticeable on occasion that

while one participant might raise their potential value to students, others would state a preference for their not being added. This view was generally driven by a belief that the system was primarily a tool for known-item search-tasks, and that additional features and functionality would simply add clutter which might hinder those core tasks.

4.6.3   Marketing

A limited number of suggestions were made for how to improve WorldCat.org's marketing. Specific ideas ranged from OCLC representative visits to libraries, encouraging word of mouth advocacy, training and news podcasts, and the implementation of better links to WorldCat.org from other library catalogues. While these ideas tended to come from librarians, several students also felt that they should be better informed about the service. However they felt that librarians in their institution were best placed to publicise the service to patrons.

4.7 Summary of Trends

While the analysis of focus data was primarily qualitative, an advantage of Qualitative Content Analysis is the emergence of quantitative data relating to the frequency and distribution of codes among participants. This section will therefore serve as a summary of the major themes to emerge from the Focus Group research, augmented by some quantitative data relating to code assignment. It should be noted here that any discussion of relative code frequencies is potentially problematic for two reasons. First, the nature of focus group data collection means it is not always possible to accurately determine the number of participants to which any one code assigned. Second, the recruitment process used means the participants cannot be considered completely representative of the wider user base. Thus it should be stressed that code frequencies are intended only as a means of giving a general sense of the relative significance of the emergent themes among the specific users involved in the focus groups.

The work-tasks that emerged from the focus groups were found to fit broadly into three categories – Academic, Leisure, and Professional. Naturally student and researcher participants were most likely to describe academic tasks, and librarians professional tasks. Leisure tasks were described least often, with only 17 of the 118 participants (14.4%) mentioning using the system for this purpose. Of these 17 participants, 6 were librarians, and 11 students. It is perhaps more instructive to review the distribution of codes relating to specific search-tasks. Here we note that three clear categories emerged; Institutional Information, Known-item tasks, and Unknown-item tasks. Table 4-1 shows the percentage of participants within each user group who were assigned at least one code from

each search-task category. We note that overall, the system appears to be used more often for known-item than unknown-item tasks, and only very rarely to access institutional information.

Table 4-1: Percentage of participants from each user group assigned at least one code from each search-task category

	Bookseller (*n*=10)	Historian (*n*=7)	Librarian (*n*=61)	Student (*n*=40)
Institutional Information	10.0%	0%	3.3%	2.5%
Known-item	80.0%	100.0%	78.7%	65.0%
Unknown-item	10.0%	85.7%	49.2%	40.0%

Summarising the perceived strengths of the system, we observe that these were categorised broadly as relating to either content, or design and function. 53.4% (*n*=63) of participants identified at least one strength relating to content, while 32.2% (*n*=38) were assigned a code relating to design and function. Based on the frequency of code assignments, the most commonly mentioned strengths were the system's interface (11.0% of participants, n=13), the presence of obscure items (16.1%, n=19), the global scope of the catalogue (19.5%, n=23) and the breadth and quality of the metadata (11.9%, n=14). 20% of students (n=8) also spoke of the system's ease of use as a major strength. Overall, codes relating to strengths of the system were assigned 274 times. This compares to 461 instances of codes relating to perceived weaknesses of the system, although it should be noted that 64 of these represented comments about the marketing of the service, rather than the system itself. Of the codes relating to system weaknesses, 57.6% (n=68) of participants were assigned at least one code relating to a functional weakness, and 50.8% (n=60) relating to some issue with the system's content. It is notable that in general students were much less likely than other groups to be assigned codes relating to content weaknesses. 20.0% of students were assigned such codes, compared with over 70% of other participants. Overall, the most frequently assigned codes were regarding the system's search function (28.8% of participants, n=34), problems with the interface (16.1%, n=19), issues with the "find a library" feature, broken links (12.7, n=15), inaccurate metadata (13.6%, n=16), and duplicate records (28.0%, n=33). Almost a quarter of all participants (24.6%, n=29) were assigned a code relating to issues with the marketing of WorldCat.org. It is interesting to observe some differences between user groups, in particular that librarians were much more likely than other groups to perceive duplicate records as an issue (42.6% of librarians, 12.2% of other users), and to perceive the search function as being problematic (39.3% of librarians, 17.5% of other users).

Looking finally at suggestions for improvement, we note that again codes were categorised according to whether they related to system design and functionality, or content. The frequency with which codes were assigned to these categories was very similar, with suggestions relating to functionality being made by 71.2% of participants (n=84), and those relating to content by 68.6% (n=81). The most commonly made suggestions are found in Table 4-2, along with the percentage of participants assigned each code. It is notable that just over a quarter of participants mentioned the addition of recommendations, a figure boosted by the large number of students (55.0%, n=22) assigned the code. It is also significant that three of the most popular codes (recommendations, hyperlinking, and links to related information) relate to tools to help users navigate and explore content in the system and beyond.

Table 4-2: Most commonly assigned codes relating to suggested improvements

Improvement	Participants assigned code (n=118)
Full text	33.1%
Customization	29.7%
Recommendations	26.3%
Marketing	25.4%
Search	24.6%
Add supplementary information	21.2%
Holdings	19.5%
Merge duplicate Records	13.6%
WorldCat Local functionality	11.9%
Links to related information	11.0%
Better hyperlinking	11.0%

4.8 Discussion

While the main discussion of the focus group findings, and their impact on the research questions, will be addressed in Chapter 8, it is perhaps useful to briefly consider here some of the key points to arise from this phase of the research project.

4.8.1   Search-Tasks

The classification of search tasks that emerged during the process of subjecting the focus group to qualitative content analysis bears some similarities to existing models, particularly those of Slone

(2000) and Hert (1996). In particular the distinction between known-item and unknown-item searches was found to hold true in the analysis of these results. It was noticeable however that no participants spoke of what Hert called "a general search for information." Participants seemed to very clearly understand that WorldCat.org was a means of resource discovery rather than a source of general information. Nor did it seem necessary in classifying search-tasks to categorise Hert's "search for information about an item" separately to "a search for a known-item." The scheme developed here considers the former a subsidiary of the latter, and instead distinguishes the different and specific search-tasks that relate to know-items. In general terms then this scheme is closest to Slone's, in that the two key distinctions are between known- and unknown-item searches. The difference lies in Slone's inclusion of area-search as a third class of search-task. While there is no reason to doubt that this is a search-task users undertake on institutional catalogue systems, the nature of WorldCat means such searches are more or less impossible, since no detail on physical location is given in a WorldCat record. We have instead identified search-tasks relating to institutional information as a third class of search task.

That known-item searches are clearly vitally important to certain groups (one thinks particularly of cataloguers), and form the basis of many search-tasks for which users employ WorldCat.org, comes as no surprise. It was however perhaps surprising to identify such a range of unknown-item tasks, and that participants spoke so frequently of using WorldCat to undertake them. This is certainly at odds with assumptions made in some of the literature relating to union catalogues, where unknown-item search tasks are rarely considered a typical use scenario.

4.8.2   Student and Librarian Disconnection

While only a minor point in relation to the broader project, it was nonetheless striking to note the discrepancy between some librarians' perceptions of students' ability to use WorldCat.org, and students' own confidence in their competence to do just that. In assuming that students would be unable to use WorldCat.org without some form of intermediation, be it training or hands on assistance, some librarians were perhaps revealing a disconnection with the capabilities and confidence of today's students in interacting with catalogue systems. This is potentially an issue of some significance to libraries, and the development of library systems. There seemed little doubt from the student responses that they were generally comfortable using WorldCat, and were able to complete many search-tasks on the system with little difficulty. Indeed a feature of student responses to this study was the self-assurance in their own searching ability, and a sense that difficulties encountered during a search task were a result of system failings rather than their own. Given the range of information sources available to students online, and the simplicity of their use, it

is perhaps dangerous for librarians to assume that students should be taught to adapt to systems rather than adapting systems to them.

4.8.3   Unknown-Item Searching

It was noticeable that many of the student comments relating to weaknesses of WorldCat.org related to issues with its supporting browsing and other means of unknown-item searching. Many of the comments suggested that for these tasks, system interactions are out of necessity located firmly within the query-response paradigm. The lack of functionality supporting exploratory search would seem to be an issue in this regard, and the result is that users are tied to a cycle of iterative queries that fail to connect them with diverse but potentially useful items. The relative paucity of examples offered by participants of serendipitous discovery within WorldCat.org can be noted here too. This suggests the well documented challenges for system designers in creating electronic environments conducive to discovery apply to those responsible for developing WorldCat.org. Thinking back to Bates' description of the browsing process, with its emphasis of the role of glimpsing and sampling information snippets, it is interesting to note that several participants contrasted the likelihood of discovering new and interesting items in a physical space, with the chances of doing so in a digital environment. We might suggest that systems such as WorldCat.org fail to offer users the opportunities to visually encounter information in the way that rows of bookshelves do.

That many of the suggestions for improvements to WorldCat related to aiding unknown-item search is therefore not surprising, although it is notable that many such suggestions represented functionality already in place within WorldCat (for example filtering and sorting search results). This perhaps suggests that despite the confidence of users, catalogues such as WorldCat are still failing to support users in maximising the tools already at their disposal for resource discovery.

4.8.4   Recommendations

It is of course particularly relevant to this project to consider participants' views on recommendations. In general it was notable that all student focus groups, and many librarian sessions, included unprompted discussion of the incorporation of recommendations as a potential improvement to WorldCat.org. The concept of recommendations seemed to be well understood, and it was assumed by most participants that recommendations would be most valuable at the item-level. It was also notable that at no stage did participants consider personalised recommendations as something that would be useful. This is perhaps related to the relatively low number of participants who admitted to having a WorldCat.org account, in that the system did not appear to be perceived by many participants as one in which a personalised experience was required.

Amazon recommendations were referred to on a number of occasions, and appeared to represent a familiar benchmark for many participants. It was notable here that several different articulations of how Amazon recommendations are presented emerged, and it was clear at times that some participants, despite valuing the service, were often unclear how exactly recommendations were generated. One suggestion here might be that users are relatively uninterested in how the recommendations are generated, so long as they are useful. This would seem to contradict much of the Recommender Systems literature relating to explanation and trust, but it should be noted that this literature relates to the domain of personalised recommendations. It might therefore be the case that item level recommendations are viewed more as links to explore than deliberate product recommendations.

5. PHASE TWO: SURVEY OF WORLDCAT.ORG USERS

5.1 Introduction

This chapter presents the results of the survey of WorldCat.org users conducted in April 2012, following a methodology discussed in section 3.5.2 . The intention was to address research questions 1, 2, 3 and 5, namely:

1. Who is using WorldCat.org?

2. For what purposes are users accessing WorldCat.org?

3. When might a recommender system support users of WorldCat.org?

5. What recommendation characteristics would be most useful to users of library catalogues?

Pop-up invitations to complete the survey were loaded to appear on the WorldCat.org homepage, and on item records pages. A total of 2,918 complete responses were received, with 894 from the .org page, and 2,024 from record pages. Results are reported in three sections: Demographics, which covers the location, age, gender and occupation of respondents; User goals, which details responses relating the reasons respondents gave for using the system; and Features, which presents the results of questions relating to a variety of system features. While the key finding as related to the research questions are discussed in Chapter 8, this chapter concludes with a brief discussion highlighting some of the key findings and limitations of the survey.

Percentages are reported to one decimal place in the text, although for layout purposes some figures only include the whole number. Pearson's Chi Square test was used to determine the significance of differences between categorical variables.

5.2 Demographics: Location, Gender, Age, and Occupation

Respondents from 128 different countries completed the survey, with the United States accounting for almost half (49.9%) of the 2918 responses. The top ten countries represented can be found in Table 5-1.

Table 5-1: Country of origin of survey respondents

	Country	Number of Responses	% of Total
1	UNITED STATES	1457	49.9%
2	CANADA	141	4.8%
3	CHINA	136	4.7%
4	GERMANY	109	3.7%
5	UNITED KINGDOM	80	2.7%
6	AUSTRALIA	76	2.6%
7	BRAZIL	59	2.0%
8	INDIA	55	1.9%
9	MEXICO	51	1.7%
10	ITALY	50	1.7%

Invitations to complete the survey were generated at two points in the system; at the WorldCat.org homepage, and at the detail (i.e. item level) pages for individual records. Canada (48.2%, *n*=68) and the United States (43.9%, *n*=639), were found to have the highest proportion of users responding to the survey at the home page (WorldCat.org) as opposed to a record page (see Figure 5-1).
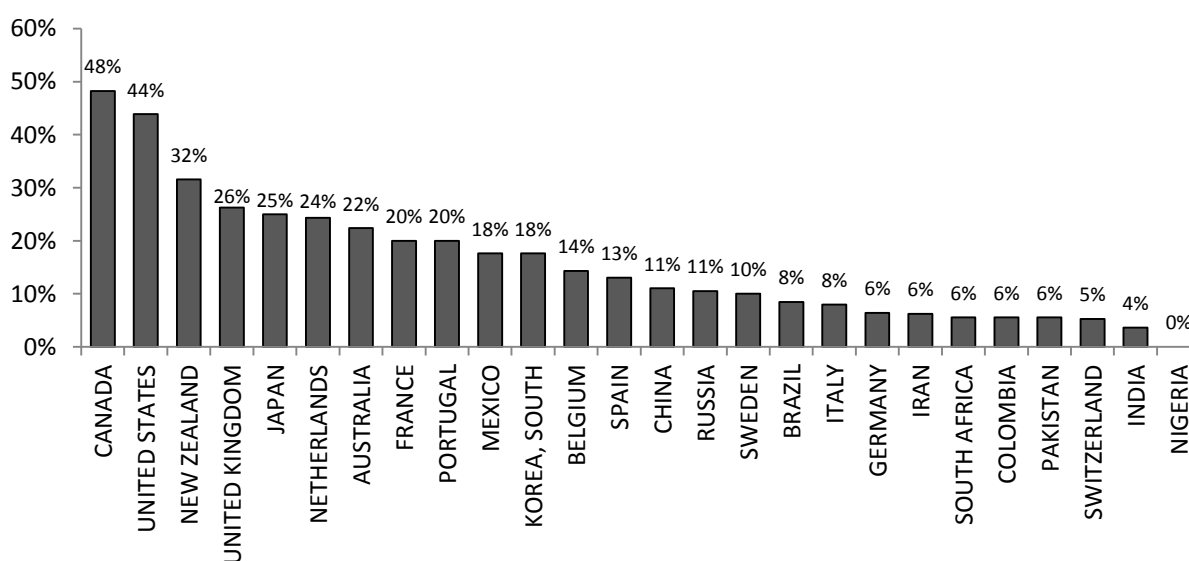
Figure 5-1: % of respondents from each country completing the survey at the home page (minimum of 15 total responses)

A slightly higher number of females than males complete survey (Female = 55.2%, *n*=1,611; Male = 44.8%, *n*=1,307). The age of participants (see Figure 5-2) was found to be relatively high, with 63.5% of respondents (*n*=1,852) giving their age as 36 or above, and the 50+ age group being the best represented (39.0%, n= 1,137).
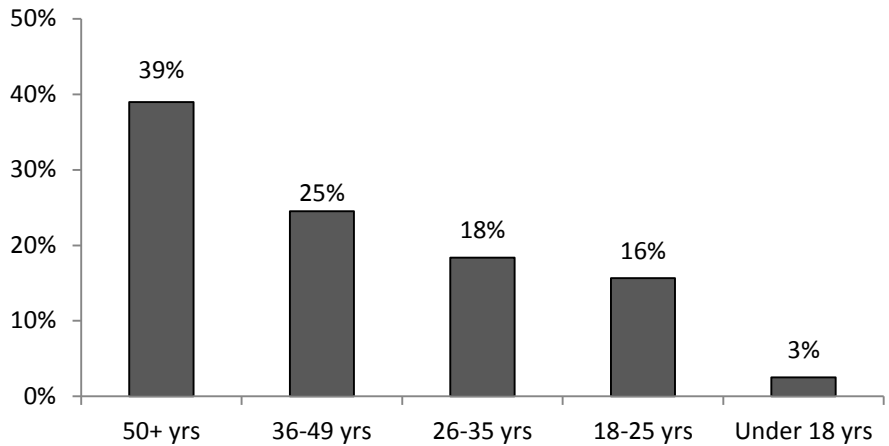
Figure 5-2: Age of Respondents

Respondents were also asked to provide their occupation, with four options provided (undergraduate student, postgraduate student, librarian, and faculty / researcher) as well as an option to manually enter an alternative occupation. As described in the methodology section (3.5.2.4), those alternative responses were reviewed and aggregated. It is interesting to note that a significant number of respondents (n=67) chose not to select the "Librarian" option, but manually entered a job title clearly related to library work (e.g. "Library Assistant", "Cataloguer"). These were coded as "Other Library Staff". Since a comparison of results for "Librarians" and "Other Library
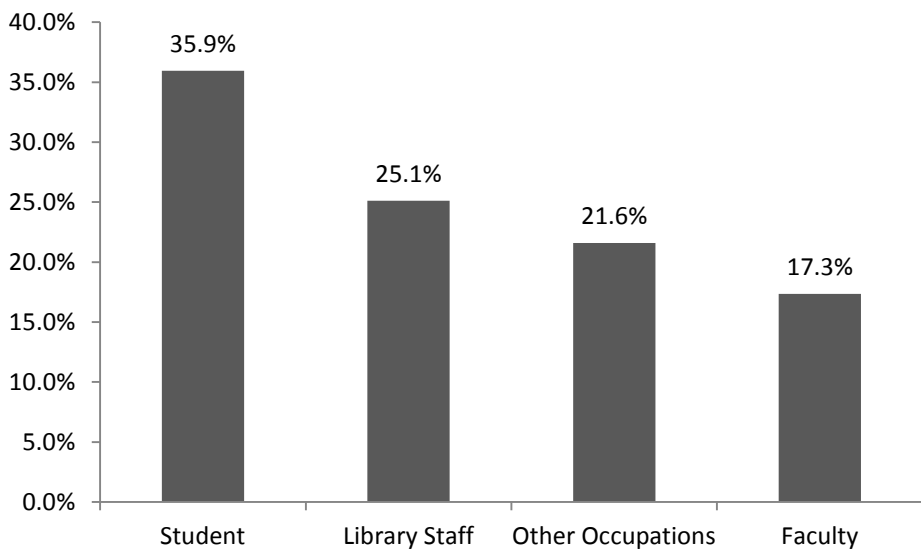
Figure 5-3: Respondent Occupations (aggregated categories)

Staff" revealed no significant differences, the data for these two groups have been collated and presented as "Library Staff" in the subsequent reporting of results. Likewise, although student respondents were able to select either "Undergraduate Student" or "Postgraduate / Graduate Student" as their occupation, analysis revealed that in most cases the responses of these two groups
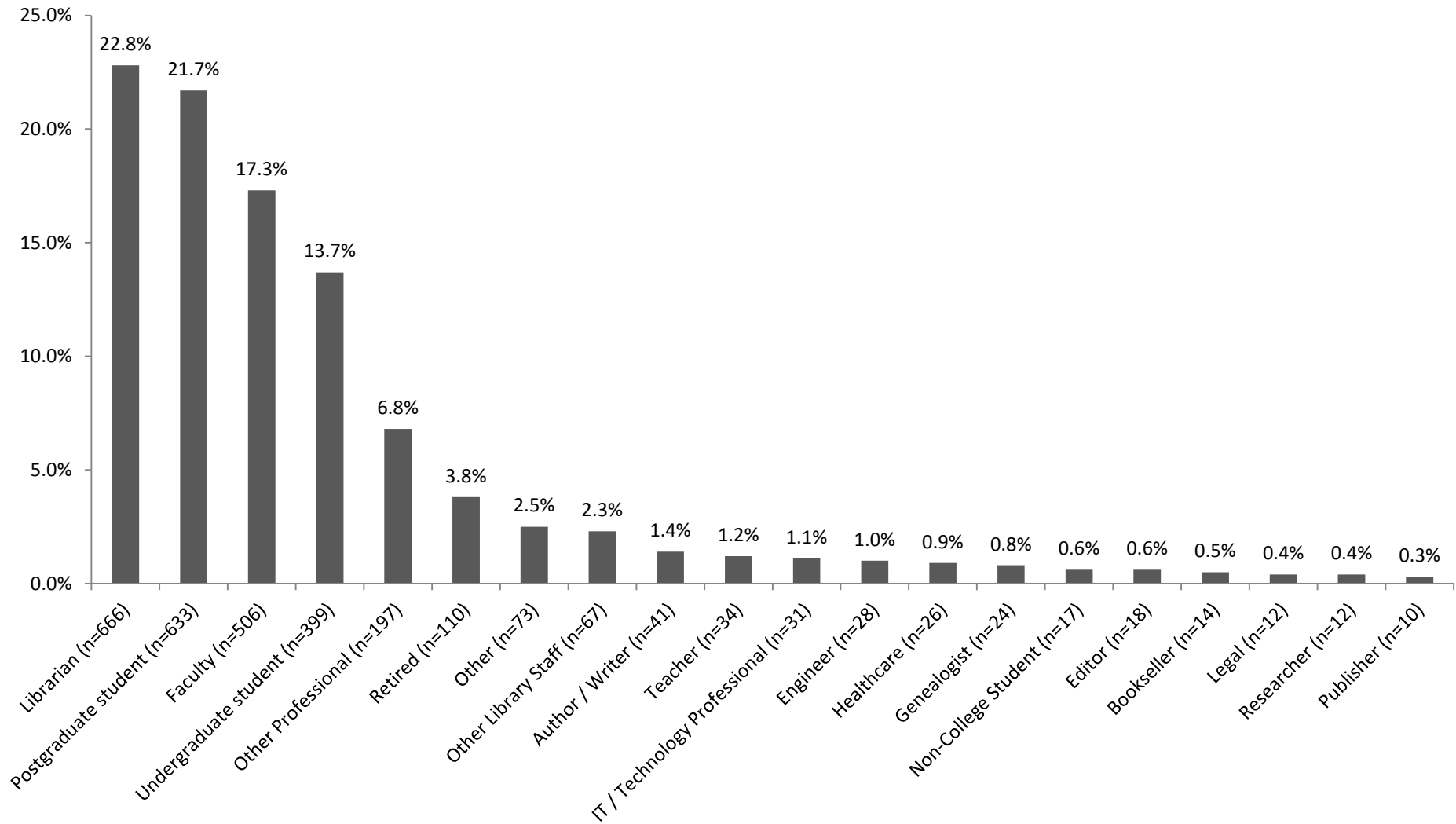
Figure 5-4: Respondent Occupations (% of all responses)

were not significantly different. They have been aggregated into a "Student" category for the presentation of most results, with the exception of those where notable differences in the two groups were observed. The general breakdown of respondent occupations is therefore presented in Figure 5-3, and shows that students represent the largest single respondent group (35.9%, *n*=1,049). Library staff account for a quarter of all respondents (25.1%, *n*=733) and academic staff under a fifth (17.3%, *n*=506), with respondents identifying as other occupations making up the remainder (21.6%, n=630). A detailed breakdown of all occupations, including coding categories, can be found in Figure 5-4.Figure 5-5 shows a breakdown of the occupations of respondents from the ten best represented countries in the survey. It shows the US and Canada as the only two countries to have a higher proportion of library staff respondents than students.
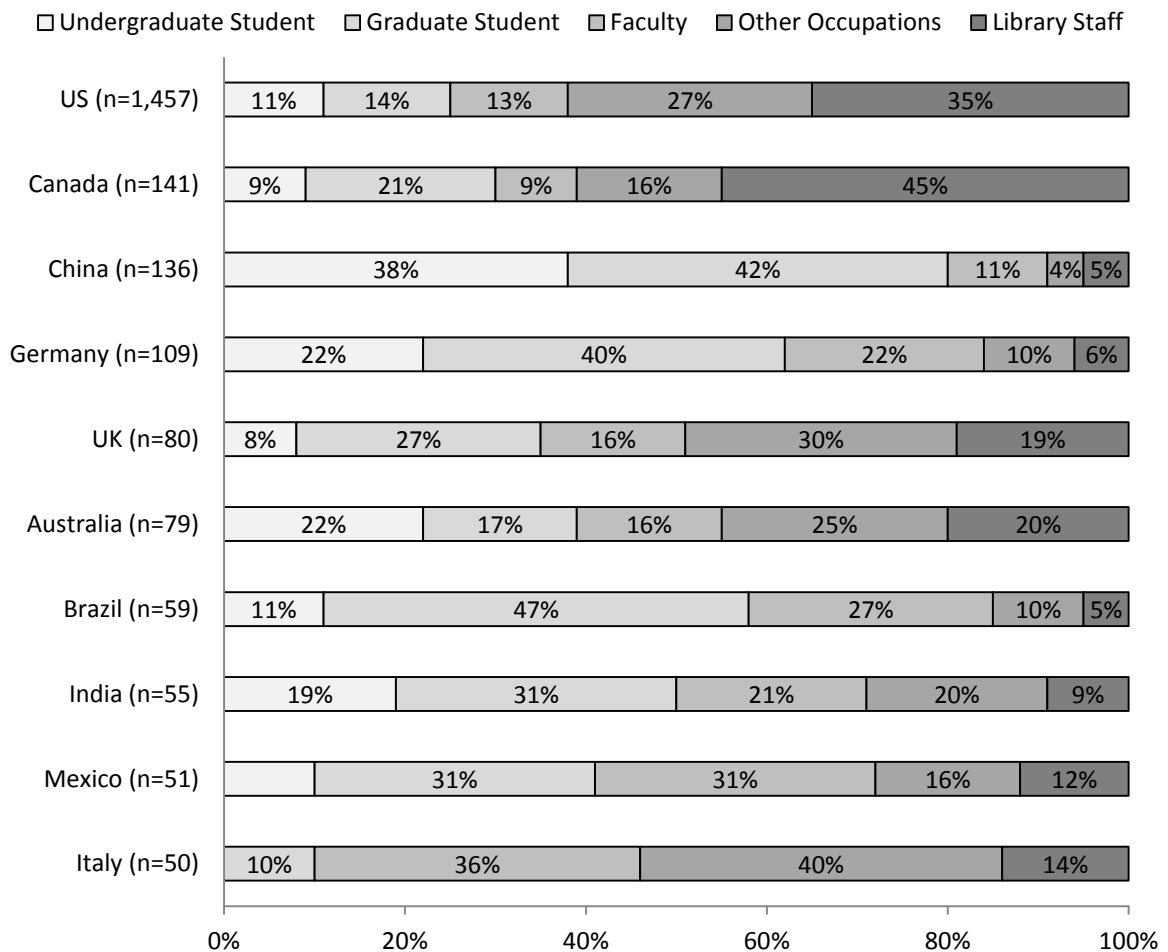
Figure 5-5: Respondents by occupation group for top 10 countries

In comparing the number of respondents taking the survey at the two entry points (WorldCat.org home page and record pages), it must be noted that since an invitation to complete the survey appeared for every user accessing WorldCat.org, as opposed to every 100th user accessing a record page, the results of this survey do not provide an accurate representation of the total number of users entering the site at these points. However, the results do allow us to compare the relative proportion of users within each occupation group taking the survey at each point (see Figure 5-6), and differences between groups were found to be statistically significant (X^2 (3, N = 2,918) = 476.12, p <.001). Librarians were significantly more likely than any other user group to complete the survey at the homepage, with 62.6% (*n*=459) of librarian respondents taking the survey there compared to only 18.3% (n=192) of students and 19% (n=96) of faculty.
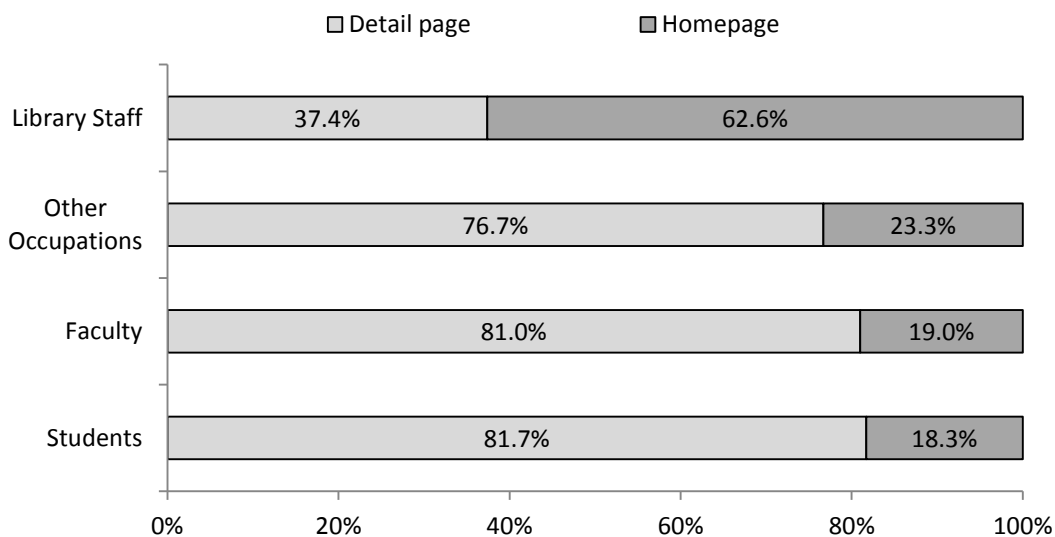
Figure 5-6: Proportion of each occupation group taking the survey at each invitation point

Splitting the total respondents from each access point by occupation (see Figure 5-7) shows that library staff made up over half (51.3%, *n*=459) of all respondents to the survey at the WorldCat.org home page, while students and faculty represent 626% (*n*=1,267) of all surveys taken at a record pages.
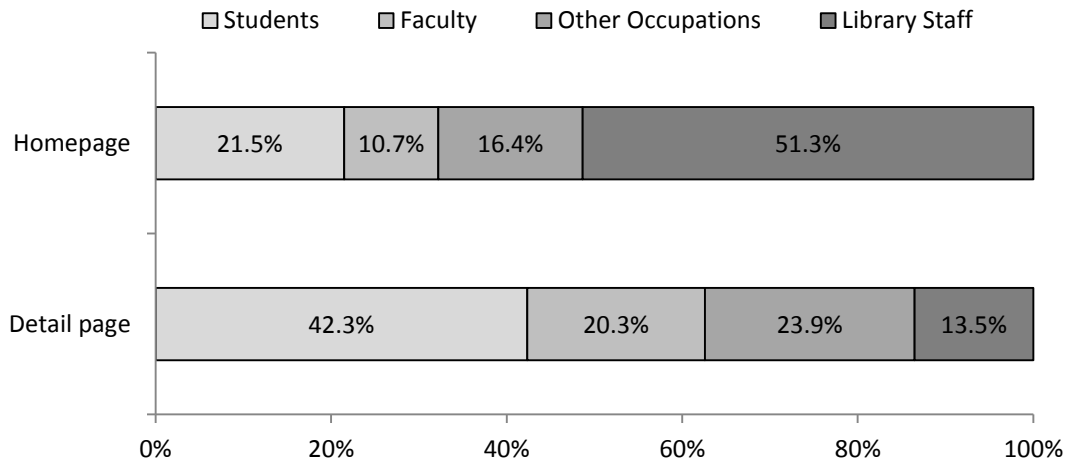
Figure 5-7: Total responses for each invitation point by occupation group (2)

5.3 User Goals: Search Tasks and Information Needs

Respondents were asked to classify their purpose for visiting the site as one of three options; "Educational", "Professional" or "Recreational." It is noted that for a significant number of users (particularly academics and library staff) it may have been difficult to distinguish between professional and educational purposes, and the results are therefore presented with these categories combined (see Figure 5-8). Only 13% (*n*=378) of respondents had a recreational purpose for visiting the system, and with the figures for WorldCat.org's key users groups – students (7.5%, *n*=79), faculty (6.5%, *n*=33) and library staff (6.0%, *n*=44) – even lower. In contrast, 59.1% (*n*=65) of retired respondents stated they were using the system for recreational reasons.

Perhaps unsurprisingly, books constituted the most sought items on WorldCat.org (see Figure 5-9), with 83.1% (*n*=2,422) of respondents stating they were seeking this format, as opposed to 29.8% (*n*=875) for journal articles, 10% (*n*=293) for video, and 9% for audio material (*n*=260). Other formats identified in the free text field included theses, sheet music, and websites, although each of these represented less than 1% of the total answers given. Faculty and students appeared slightly more likely to be seeking journal articles, but otherwise the results were relatively consistent across user groups.
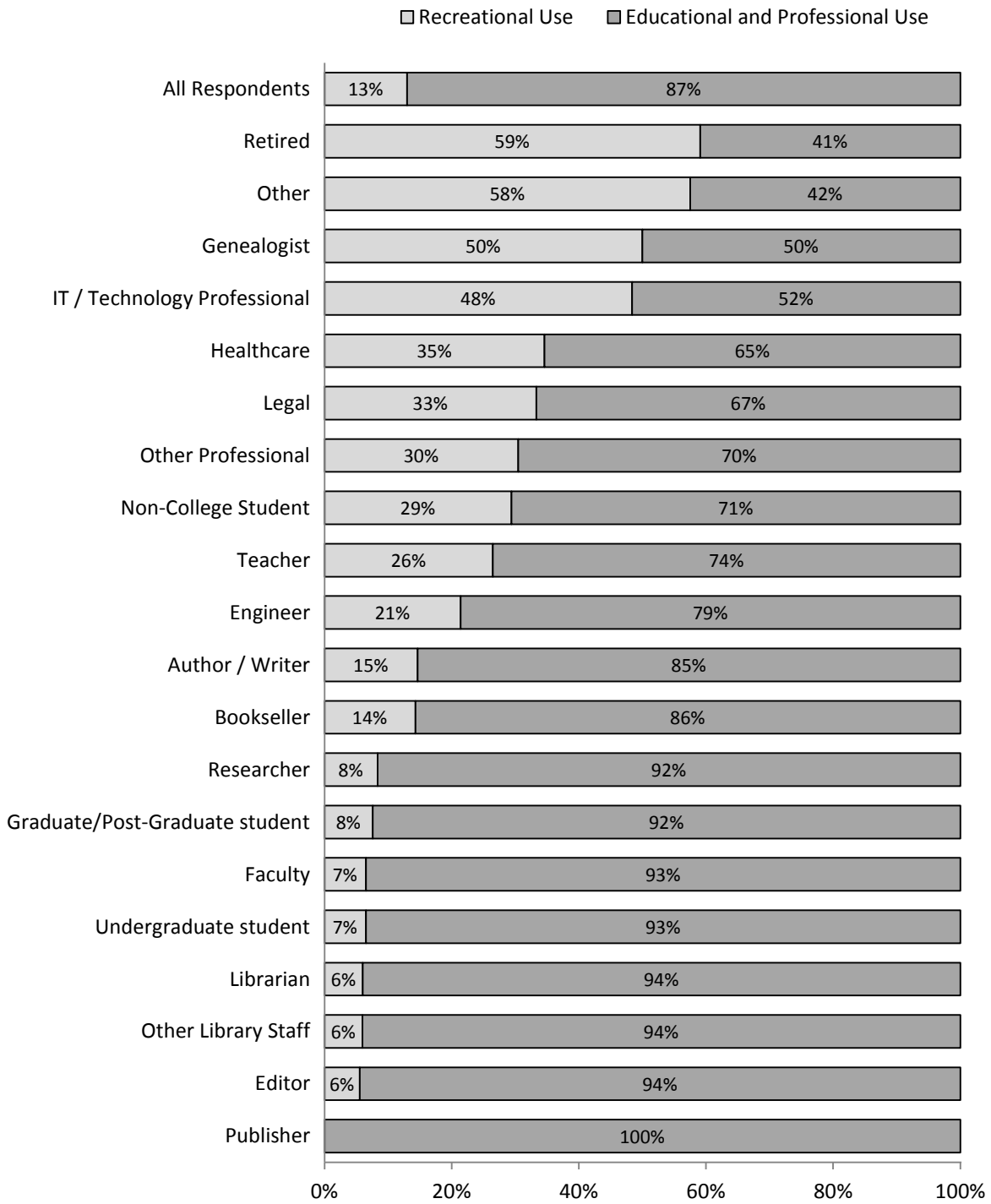
Figure 5-8: What is your purpose for using WorldCat.org? Respondent Groups ordered by Recreational Use
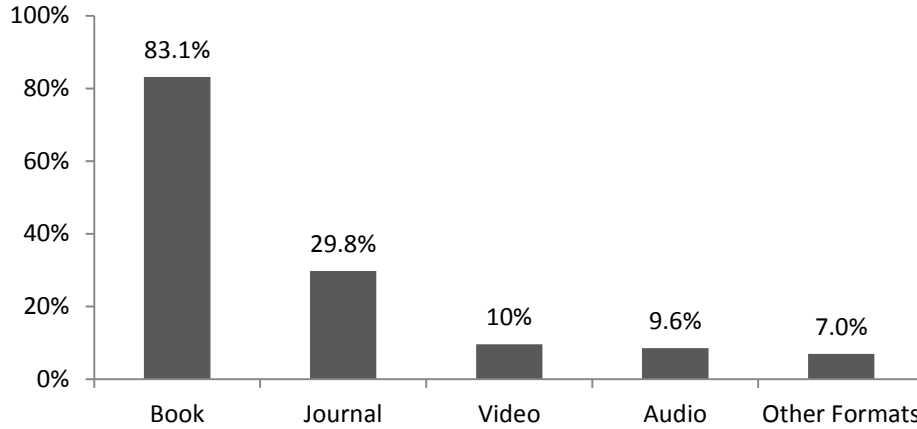
Figure 5-9: "What material are you looking for? (Select all that apply)"

Respondents also were asked to rate the importance of being able to access full text versions of resources online, with four possible answer options: "Not Important", "Somewhat Important", "Important", and "Very Important". Overall, 52.9% (n=1,543) of respondents selected "Very Important", and 76.0% (n=2,218) either "Important" or "Very Important". Differences between occupation groups were found to be statistically
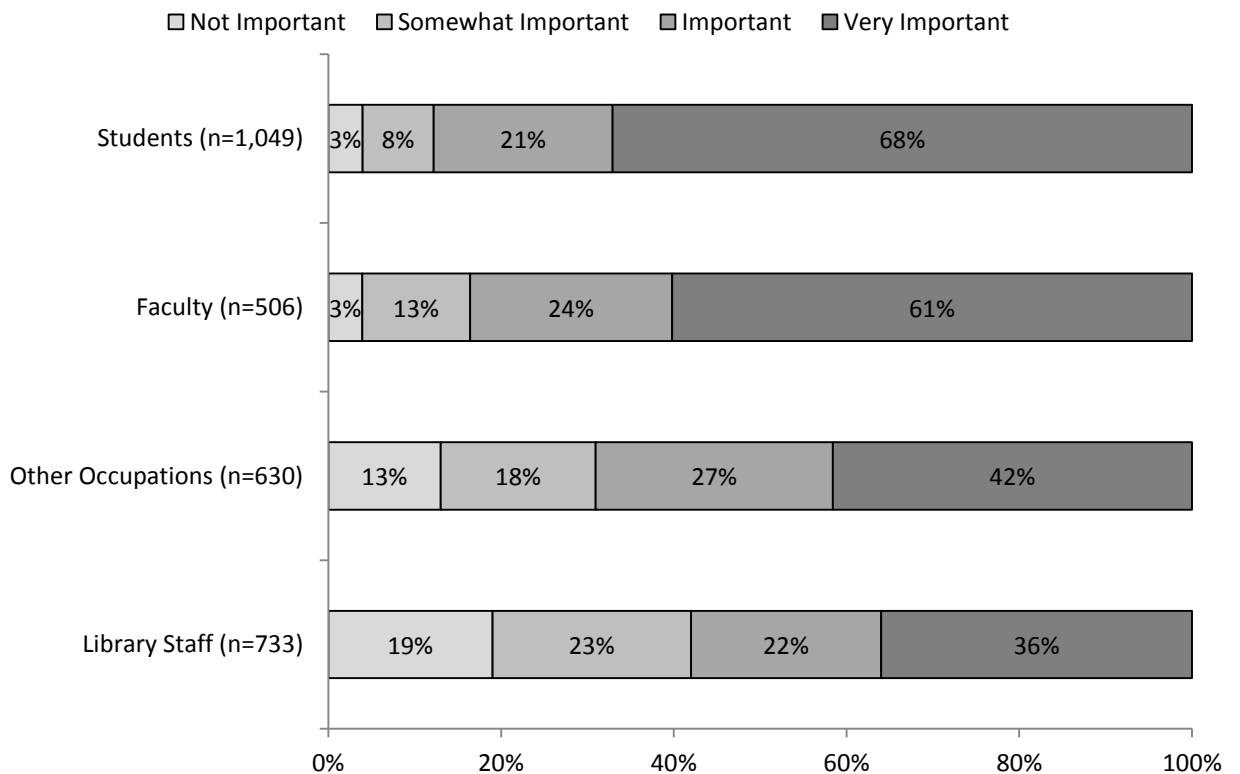
Figure 5-10: Importance of full text by occupation

significant (X^2 (9, N = 2,918) = 330.49, p <.001). Breaking results down by occupation, students and faculty valued full-text access most, with 88.5% (n=928) and 84.8% (n=429) respectively selecting either "Important" or "Very Important", compared to only 58.1% (n=426) of library staff (see Figure 5-10). The possibility of age being a factor was considered, with statistically significant differences found between the perceived importance of full text to different age groups (X^2 (12, N = 2,918) = 65.68, p <.001). In general younger respondents appeared to value the availability of full text more highly, although it should be noted that for all age groups a large majority (90.9%, n=2,652) felt online access was at least "somewhat important".
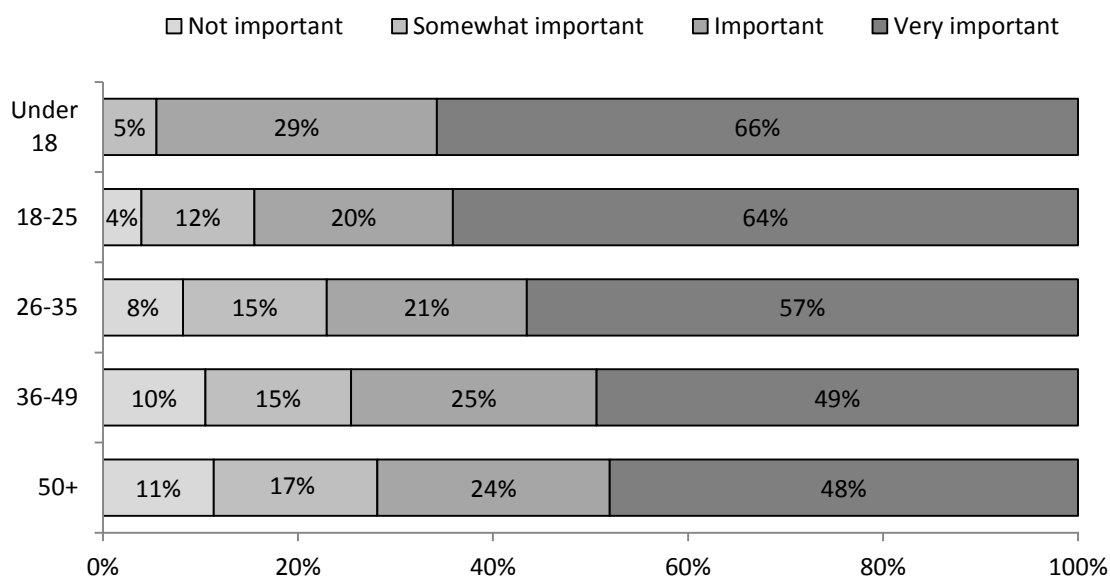
Figure 5-11: Importance of full text by age

Participants were also asked to select sites and services they would normally use to search for full text online (see Table 5-2). Overall, users were most likely to seek online versions of text through an academic library website or catalogue, with 55.2% (n=1,610) of all respondents acknowledging this behaviour. This number even was larger within student and faculty groups; 67.7% (n=710) and 66.2% (n=335) respectively. While at first glance these numbers appear high, it should be noted that they suggest around a third of students and faculty do not typically attempt to find full-text resources through their institution's catalogue or databases. While it is highly likely that academic libraries facilitate access to full text material found through other systems (through IP address or some other recognition of user affiliation with a subscribing library), these results suggest that for a

substantial number of users library services are not seen as a helpful means of discovering (rather than accessing) electronic content.

Table 5-2: "Where do you normally go to search for full text online?"

	Students (*n*=1,049)	Faculty (*n*=506)	Library Staff (*n*=733)	Other Occupations (*n*=630)
Academic Library	67.7%	66.2%	53.5%	27.5%
Google Books	47.4%	52.6%	38.7%	44.6%
Search Engine	37.8%	41.1%	36.6%	52.2%
Public Library	33.2%	32.6%	47.1%	38.7%
Google Scholar	35.6%	32.0%	29.2%	14.0%
Online Retailer	16.0%	17.2%	20.5%	25.2%
Wikipedia	19.9%	14.4%	8.5%	26.0%
Other Sources	5.8%	8.5%	13.1%	14.8%

37.8% of all respondents (*n*=1,102) use public library online services to discover full text. Interestingly, use of public libraries for this purpose is highest among librarians themselves (47.1%, *n*=345). Google services were found to be widely used, with Google Books (45.5%, *n*=1,328) proving more popular than Google Scholar (28.7%, *n*=837). Respondents identifying as "other occupations" were the most likely to turn to Wikipedia (26.0%, *n*=164), and it is noticeable that a higher proportion of students (19.9%, *n*=209) were likely to use this resource than faculty (14.4%, *n*=73) and library staff (8.5%, *n*=62).

Question 10 asked 'What is your reason for using / visiting this site?', with respondents invited to select all that applied from the following options:

1. *To find information about a particular topic (e.g. to find information about cats or World War II)*
2. *To search for one or more unknown books or resources about a particular topic (e.g. to find a book or books about cats or World War II that you weren't previously aware of)*
3. *To find a location (either in a library or online) of a specific known item (e.g. to see whether a local library has a particular item that you are already aware of)*
4. *To find more information about a specific known item (e.g. to check the publication date or author of an item that you are already aware of).*

The intention had been to differentiate between the act of looking for information itself (1), and the act of looking for an unknown-item from which to gather that information (2). As Figure 5-12 shows, only 16% (*n*=454) of respondents answered with option 2, compared to 38% (*n*=1,123) for option 1, despite it intuitively seeming likely that most users of a library catalogue with a specific information need would be seeking a resource to address that need. This contrast was even starker within certain user groups, with 60% (*n*=240) of undergraduate students seeking information, and only 19% (*n*=75) searching for unknown-items. It seems likely that some users perhaps failed to properly distinguish between these two answer options, and it is acknowledged here that the question might have been more

.
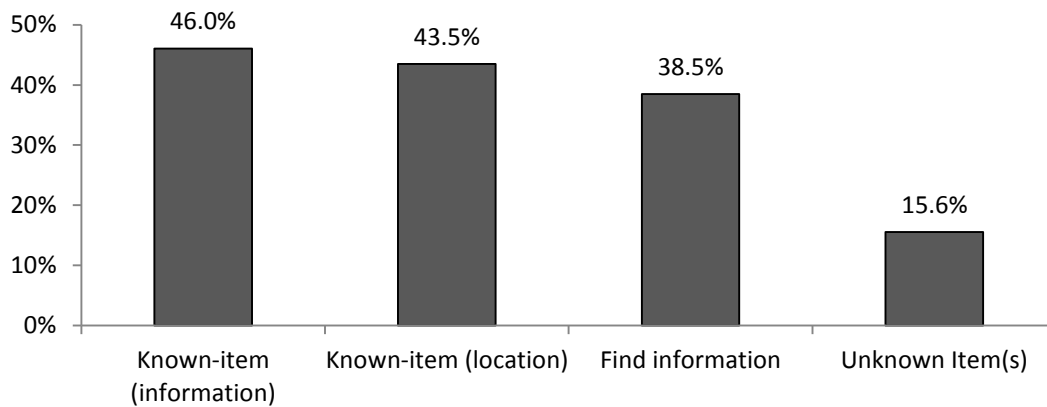
Figure 5-12: "What is your reason for using / visiting this site? (Select all that apply)"

clearly defined. In particular it is possible that respondents were confused by the distinction between the more abstract notion of addressing an information need, and the more focussed act of attempting to identify information sources to address that need. The phrasing of the questions might also have been improved, especially in that unfamiliar terminology such as "unknown" could have been avoided[7].

Analysis revealed significant differences in search tasks between user groups (all differences were found to be significant to *p*<.01), and this was one of the few questions to reveal a significant difference between undergraduate and postgraduate students (see Table 5-3. Noting that respondents were able to select multiple tasks, 28.1% (*n*=112) of

---

[7] One respondent added a particularly pithy critique of this term to the closing comments section of the survey: "As long as a book's "unknown" I certainly won't search for it, neither here nor anywhere else - except for where it exists."

undergraduates said they were seeking the location of an item, and 34.3% (*n*=137) that they were attempting to find out more information out about an item, compared with 42.3% (*n*=268) and 41.2% (*n*=261) respectively of postgraduates (see Table 5-3). Library staff appeared the least likely to be engaged in the search for an unknown-item. Among Other occupations, finding the location of an item was the most popular task (48%).

Table 5-3: Reason for Visiting WorldCat.org by Occupation

	Find information	Unknown-item(s)	Known-item (location)	Known-item (information)
Undergraduate Students	60.2%	18.8%	28.1%	34.3%
Postgraduate Students	48.2%	19.9%	42.3%	41.2%
Faculty / Researcher	43.3%	17.6%	43.7%	43.9%
Library Staff	16.4%	8.6%	48.4%	62.3%
Other Occupations	36.3%	15.4%	48.1%	41.4%

The data were further analysed to identify respondents engaged in at least one known task (i.e. those selecting one of options 3 and 4), and also those stating they were only undertaking know-item tasks (i.e. respondents who selected one or both of options 3 and 4, but neither option 1 nor 2). As Table 5-4 shows, significant differences were found between user groups (for "at least one known item task": X^2 (4, N = 2,918) = 185.37, p <.001; for "only known item task(s)": X^2 (4, N = 2,918) = 279.80, p <.001).

Table 5-4: Proportion of participants engaged in known-item tasks

	At least one known-item task	Only known-item tasks(s)
Students	60.4%	37.1%
Faculty	72.1%	49.2%
All Library Staff	89.5%	77.1%
Other	74.6%	52.1%

Library staff were found to be much more likely to be undertaking some form of known-item search, with 89.5% (n=656) respondents from this group engaged in this activity, compared with 60.4% (n=634) of students. While this shows an interesting difference, it must be noted that these percentages include respondents who may also have been conducting informational or unknown-item searches, but intended to obtain the location of

the item(s) they discovered. The proportion of respondents engaged solely in known item tasks is therefore more revealing, in that over three quarters (77.1, n=565) of library staff responding to the survey were determining either the location or some bibliographic information about a known-item. In contrast less than half of students said they were only conducting a known item search (37.1, n=389).These results were statistically significant (X^2 (3, N = 2,918) = 279.80, p <.001) with a large effect size (Cramer's V = .310).

5.4 Features of WorldCat.org

15.5% (n=451) of all respondents said they had created a WorldCat.org account. Library staff (24.3%, n=178) were significantly more likely to have done so as students (11.7%, n=123) and faculty (both 12.5%, n=63) (X^2 (6, N = 2,918) = 64.487, p <.001).
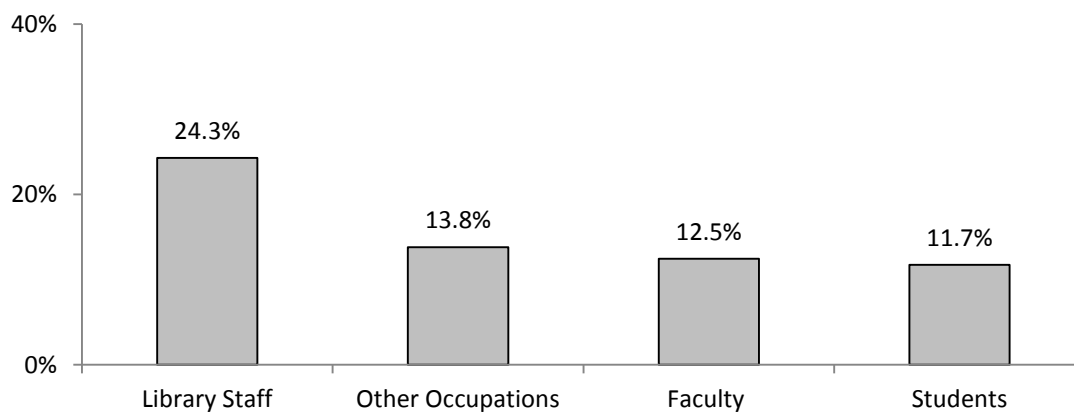
Figure 5-13: Proportion of respondents who have created a WorldCat.org account

Questions 12 and 13 investigated the role of several next-generation catalogue features; Lists, Reviews, Tags, and Ratings. Participants were first asked whether they had ever used each one of these features of WorldCat.org to help find or evaluate an item, with a further option to specify that they had never used any of them. No single feature was found to

Table 5-5: Proportion of respondents by occupation group who have user next-generation features for resource evaluation and discovery

	Lists	Ratings	Reviews	Tags	None
Students	16.9%	9.8%	13.7%	12.7%	60.8%
Faculty	13.0%	6.1%	13.4%	10.7%	66.8%
Library Staff	16.4%	7.0%	16.6%	17.3%	64.9%
Other Occupations	10.3%	6.0%	10.3%	10.5%	75.4%

have been used by more than 20% of respondents, with 66.0% (n=1,927) stating that they had never used any. While use of the features was found to be low, some variation in the usage levels of features between occupations was observed (all found to be significant to *p*<.01), with library staff more likely than other groups to use tags and reviews (see Table 5-5).

Further analysis was conducted to determine if respondents with a WorldCat.org account were more likely to make use of these features than those without. Differences in use between account holders (n=2,467) and non-account holders (n=451) were found to be statistically significant for each feature (p<.001), with account holder much more likely to utilise the features for resource discovery and evaluation (see Table 5-6). Most notably, whereas 79.1% (*n*=1,767) of non-account holders have not used any of the features, this figure was only 35.5% (*n*=160) for account holders, meaning roughly two thirds of the latter group had previously engaged with tags, review, lists or ratings.

Table 5-6: Proportion of respondents with and without a WorldCat.org account who have used next-generation features

	WorldCat Account?	
	Yes	No
Lists	42.6%	10.6%
Ratings	12.0%	7.6%
Reviews	21.5%	13.5%
Tags	25.9%	11.8%
None	35.5%	79.1%

An important point to note here is the relative sparsity of tags and reviews in WorldCat.org. The survey data unfortunately do not allow us to determine for certain whether the relatively low use of these features by general users is a consequence of simply not encountering tags and reviews in their interactions, or more fundamental evidence of an unwillingness or inability to utilise them. We have seen however that users with an account are much more likely to use next-generation features, which implies that non-use of tags etc. is not solely a sparsity issue. Therefore we might suggest that WorldCat.org account holders employ different information-seeking strategies within the system, in that they make greater use of all available resource discovery and evaluation tools. The disparity between the use of features by account holders and non-account holders may support the idea of a set of repeat WorldCat.org users who are prepared to utilise user-generated content.

Since adding lists, tags, reviews and ratings requires the user to have a WorldCat.org account, analysis of data relating to the creation of this content was limited to respondents who had stated they had such an account (*n*=451). This served to filter out data from a number of respondents (*n*=63) who stated they didn't have an account, but had created one or more types of user generated content.[8] Almost half (48.3%, *n*=218) of respondents with an account stated they had never created lists, tags, ratings or reviews on the site. 39.9% (n=180) of users with an account have created a list, making it by far the most
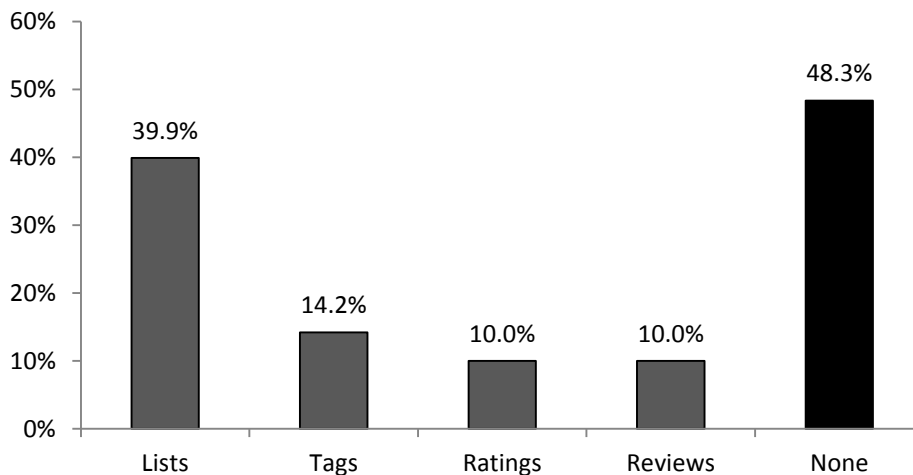
Figure 5-14: Proportion of respondents with a WorldCat.org account who have added user-generated content

popular feature (see Figure 5-14). It is likely that this can be explained by the personal utility of list creation: while user-generated lists often are of use to other users, they also constitute a valuable means of managing and ordering resources. A small minority of respondents stated they had added tags (14.2%, *n*=64), ratings (10.0%, *n*=45) and reviews (10%, *n*=45).

Respondents were generally very positive about the idea of recommendations being added to the site. Question 14 asked "What type of recommendations (i.e. suggestions of related items) would be useful on this site?", with participants invited to select all that applied form the following list:

1. *Recommendations based on expert opinion*
2. *Recommendations based on items with similar content or authors*
3. *Recommendations based on library circulation data*

---

[8] It is suspected that these respondents may have mistaken the question to refer to the creation of such content in any library catalogue, or indeed any system.

4. *Recommendations based on the most viewed record pages*

5. *Recommendations based on the ratings of other users*

6. *None – I have no use for recommendations*

33% (*n*= 963) respondents selected option 6, indicating that they had no used for recommendations on the site. If responses representing any combination of options 1 to 5 are collated, we find that 67.5% (*n*=936) of all respondents perceive at least one form of recommendation being useful. Breaking that dataset down by occupation group, we observe small but statistically significant differences (X^2 (3, N = 2,918) = 35.02, p <.001). As seen in Figure 5-15, Students were most likely to perceive some potential benefit of recommendations (73.1%, *n*=767), followed by Faculty (68%, *n*=343) and Library staff (65%, *n*=478.
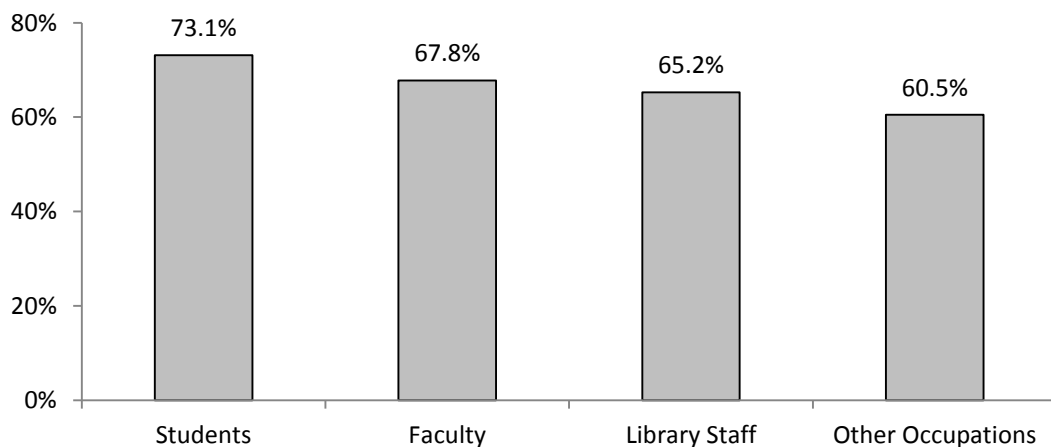
Figure 5-15: Proportion of respondents from different occupation groups who view at least one form of recommendation as potentially useful

Similarly, small but significant differences were observed in the proportion of respondents from different age groups that perceived at least one recommendation type as being potentially useful (X^2 (4, N = 2,918) = 35.42, *p* <.001). Respondents aged over 50 were less likely than the 18-49 age groups to indicate a preference for recommendations (see Figure 5-16).

179
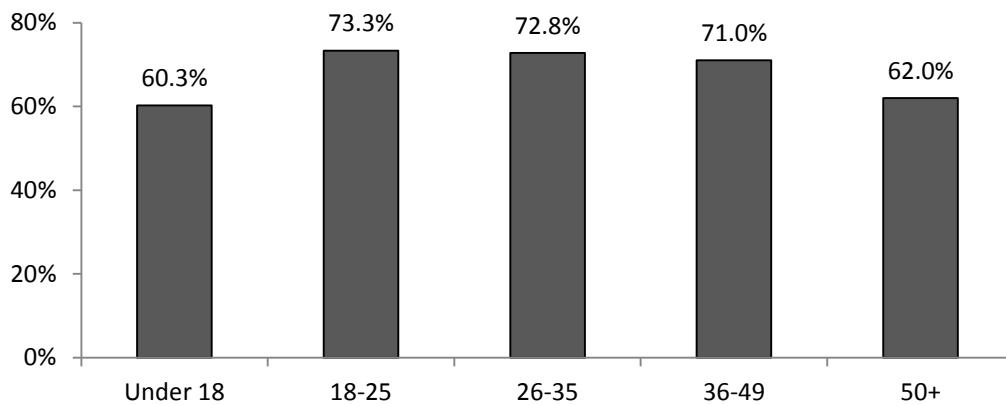
Figure 5-16: Proportion of respondents from different age groups who view at least one form of recommendation as potentially useful

Regarding the types of recommendations that respondents felt would be most useful, recommendations based on expert opinion and those based on similar content were clearly identified as being potentially the most useful type of recommendations, with 41.5% (n=1210) and 38.6 (n=1,126) of respondents respectively selecting those answers (see Figure 5-17. Recommendation types that draw on collaborative filtering methods were significantly less popular, with the idea of utilizing user ratings (16.9%, n=492), circulation data (20.2%, n=589), and most viewed items (10.2%, n=298) all finding relatively little support. Perhaps surprisingly, there were no significant differences between occupation groups regarding the perceived usefulness of the different recommendation types.
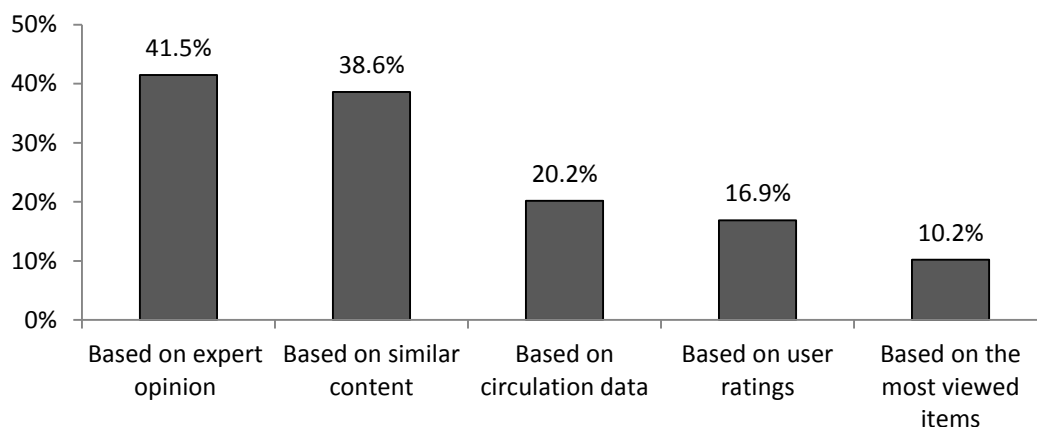
Figure 5-17: "What type of recommendations would be useful on this site? (Select all that apply)"

Respondents also were enthusiastic about the idea of reviews. While reviews are already a feature of the service, as noted above they are far from widely created or used. The survey asked respondents to judge whether two types of review would be useful on the site: "Reviews by experts (e.g. academics or critics)" and "Reviews by other users." In total, 74.6% (*n*=2077) of respondents stated a preference for at least one of these review types. As with the discussion of recommendation types, the idea of expert content proved most attractive to users from all occupation groups (see Figure 5-18). User generated reviews were perceived as most valuable by non-academic users, perhaps because a significant number of these users do not place the same importance on the authority of reviewers.
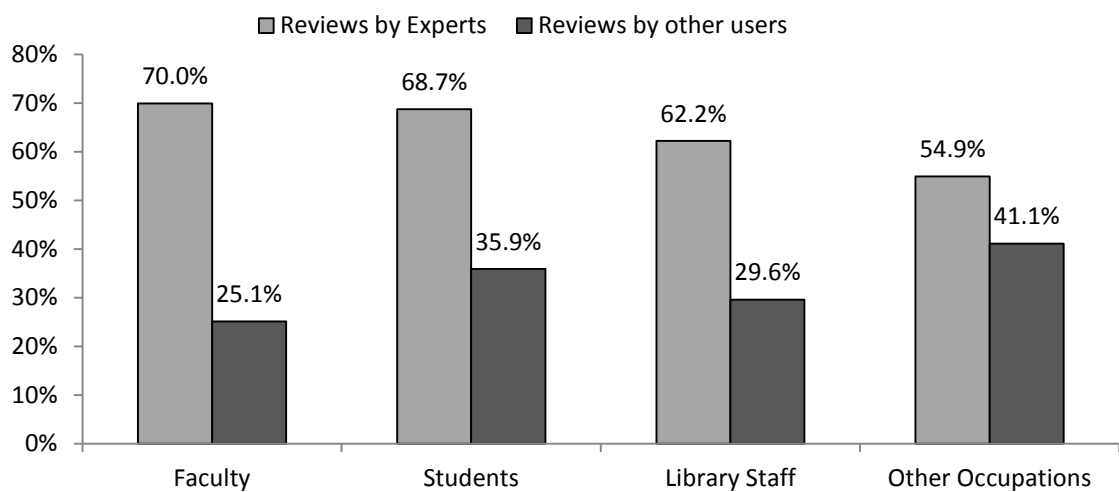
Figure 5-18: "What types of reviews would be useful on this site?"

One final trend to emerge from the data related to the behaviour and perceptions of respondents from three loosely related occupations – publishers, editors and booksellers. As Table 5-7shows, respondents from these groups were significantly less likely to have used any of the WorldCat.org features (Lists, Tags, Reviews, Ratings, and Recommendations) to find or evaluate an item, and were much more likely to see no use for recommendations and reviews on the site (all significant to *p*<.01). Although the sample size is relatively small (a total of 42 respondents from the three occupation categories), these results do suggest a strong resistance to any changes within the catalogue from user groups that in one sense might be expected to welcome new means of discovering, sharing and evaluating books.

Table 5-7:- Perspectives on and interaction with WorldCat features by Publishers, Editors and Booksellers

	Publishers (10)	Editors (18)	Booksellers (14)	All Respondents (2918)
Have not used next-gen features (e.g. Tags) to find or evaluate an item	70%	72%	79%	**66%**
Do not believe Recommendations would be useful	80%	61%	57%	**32%**
Do not believe Reviews would be useful	80%	33%	43%	**25%**

5.5 Discussion

5.5.1 Sample Validity and Non-Response Bias

Before discussing the survey findings in more general terms, it is important to consider the validity of the results, particularly in terms of how representative they are of the views and behaviours of the wider WorldCat.org user base. The problem of non-response bias (i.e. that the answers of respondents differ from the potential answers of those who did not respond) is likely to have affected the results, and there is no way of determining precisely either the scale of the problem or the direction in which results may have been skewed. Nonetheless it seems logical that four types of users are potentially less likely to have completed the survey:

- **First time or occasional users**
  Regular users of WorldCat.org potentially have a greater interest in helping improve the service, and are therefore incentivised to participate. It is possible therefore that the survey results over sampled users such as librarians and academics, and under-represented casual non-expert users.

- **Users engaged in time critical tasks**
  It might be expected that users facing time-constraints would be less likely to complete the survey. One implication might be that recreational users, and to a lesser extent academic users, are over represented, while professional users are under sampled.

- **Users engaged in clearly defined, close-ended tasks**

  Similarly, it seems likely that users conducting specific known-item tasks – for example, verifying bibliographic details, or checking an author's name – would be less likely to interrupt their session to complete the survey than users engaged in a more open-ended discovery tasks.

- **Non-English speakers**

  It is reasonable to assume that users for whom English in not a first language may be under-represented in the survey results.

The analysis of WorldCat.org transaction logs, described in Chapter 6, goes some way to addressing these concerns, and the implications of those results will be discussed in Chapter 8.

5.5.2 Comparison with 2008 Survey

Although there is not a great deal of earlier data against which to benchmark our results, one study does offer an interesting point for comparison. The 2009 OCLC Report *Online Catalogs: What Users and Librarians Want* (Calhoun et al., 2009) presents data from of a pop-up survey of users, which ran between May 12 and July 9 2008, and garnered 11,151 responses. While the focus of the 2008 survey was to elicit suggested improvements to data, the results also include demographic details. Figure 5-19 compares the occupations of respondents to both surveys. If one assumes that issues of non-response bias are likely to

Figure 5-19: Comparison of respondent occupations: 2008 Survey v 2012 Survey
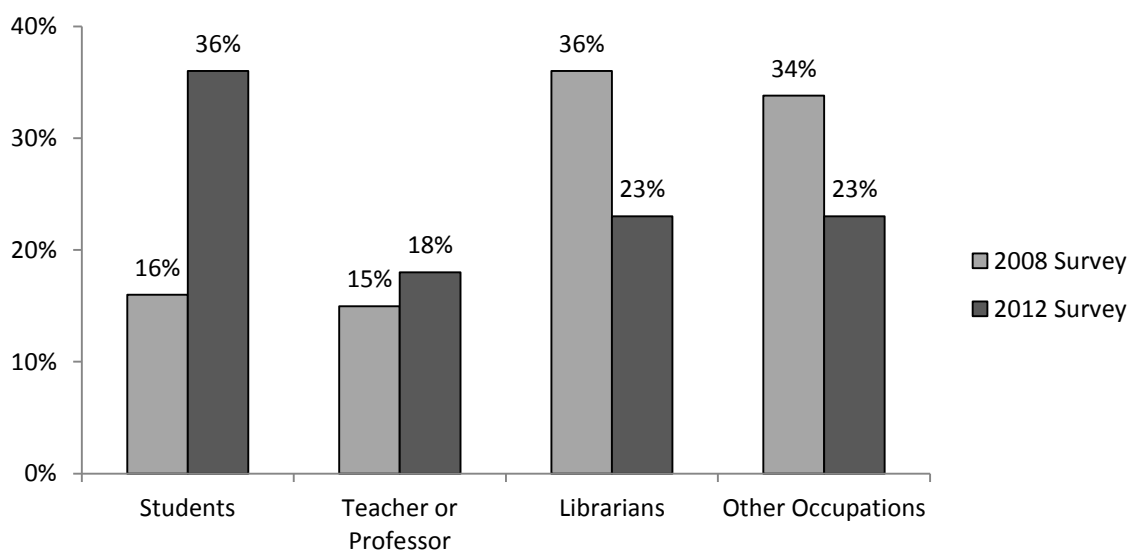
be more or less equivalent in both cases then these results reveal two interesting trends; an increase in the proportion of student respondents (from 16% to 36% of the total) and a decline in the number of librarian s (from 34% to 22%).

It also is possible to compare the ages of respondents of the two surveys. The 2009 report did not report age data for librarian respondents, so the data shown exclude library staff. There also are slight differences in the age-bands used in each survey, meaning the comparisons are not exact. Nonetheless, even accounting for these minor differences, it seems clear that that the age of users has remained more or less constant (see Figure 5-20). This is perhaps surprising given the rise in student numbers reported above.
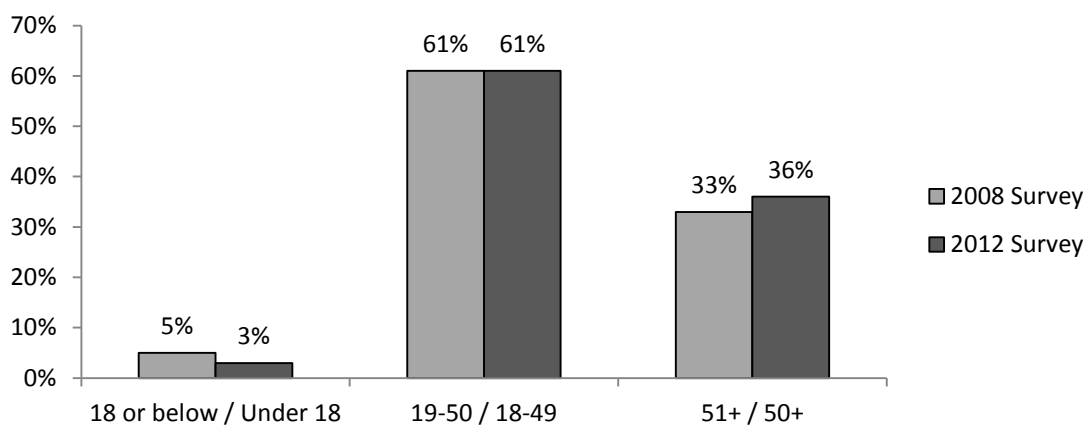
Figure 5-20: Comparison of respondent ages: 2008 Survey vs. 2012 Survey (both excluding library staff)

5.5.3 User Groups and Search-Tasks

While noting the difficulties in determining how representative respondents to this survey were of WorldCat users in general, it seems reasonable to draw some broad conclusions about the make-up of the user population. Librarians, academics and students seem to represent significant bodies of users, even if the proportions of each group are questionable. WorldCat.org has also been shown to attract users from a variety of other professions. We can also conclude that while the majority of users are US based, the system has a huge geographic reach, with respondents coming from well over 100 countries. Analysis of the location of respondents to the different survey invitation points (homepage and record page) suggest that users from North America are more likely than others to access the site from the homepage, indicative perhaps of their greater awareness

of the service. It seems probable that significant numbers of users from outside the US and Canada arrive at WorldCat.org via search engine and library referrals. Survey respondents may similarly be split by occupation, it appearing that librarians are more aware of the WorldCat.org service, and therefore more often deliberately arrive at the homepage.

It also seems reasonable to conclude that WorldCat.org primarily supports academic and professional work tasks, although clearly a sizeable number of users do utilise the system for recreational purposes. While it is regrettable that the survey was unable to gather clearer data on respondent's search tasks, in identifying the proportion of users from different occupational groups engaged solely in known-item tasks we can observe that although such tasks form the most likely use-case for a very many users, there are considerable numbers of students and academics for whom WorldCat is being used as a resource discovery tool.

5.5.4   Next-Generation Features

Interpreting the results of questions relating to the use and creation of tags, reviews, ratings and lists, is made problematic by the known scarcity of such content on the system. It is therefore unclear whether the relatively low usage rates are a result of users choosing not to utilise them, or simply not having an opportunity to do so. The higher rates of use among account holder might suggest that scarcity alone is not to blame, but regardless of the reason, the fact remains that the survey results suggest that the features are by and large unused. Comparing the proportion of respondents who perceive a benefit of review, with the number who actually use reviews, reveals an interesting contradiction; that while users are broadly enthusiastic about the idea of incorporating the feature into the system, those same users do not actually tend to use or create this content in its current form. This contradiction may in part be explained by the nature of the survey methodology, and indeed it is a common problem in the elicitation of software requirements that users will tend towards supporting the introduction of a feature. Equally however it may be that the issue lies in the effectiveness of the implementation; that users see the value such features could bring, but systems such as WorldCat.org fail to deliver them in meaningful ways.

5.5.5 Recommendations

The survey results relating to recommendations show broad support for their introduction in some form. While the notion of expert recommendations is most attractive to users, other more practical forms of recommendations were supported by small but significant numbers of respondents. Overall though the results perhaps say less about the type of recommendations that would offer the most utility in practice, and more about users' perceptions of recommendations. It is likely that very few respondents to the survey understand the complexities of recommender systems design, or appreciate the richness of the various available data sources. Recommender systems research indicates that users value recommendations more highly if they are thought to originate from a human source (McNee et al. 2003; Konstan et al., 2006), and it is therefore understandable that the idea of "expert" recommendations, however impractical to implement, should be particularly attractive. It is also perhaps unsurprising that without an appreciation of the potential power of collaborative filtering algorithms, users should be suspicious of recommendations based on the preferences of others. The lack of engagement with tags and reviews suggest that visibility and coverage might be key considerations in the design of a recommender system for WorldCat.org.

6. PHASE THREE: WORLDCAT.ORG TRANSACTION LOG ANALYSIS

6.1 Introduction

Phase three of the research project consisted of an analysis of two months of transaction logs for WorldCat.org (October 2012 and April 2013), which were supplied by OCLC. The intention was to address research questions 1 and two, namely:

1. Who is using WorldCat.org?

2. For what purposes are users accessing WorldCat.org?

The final logs were filtered to remove non-human traffic, and segmented into sessions using a 30 minute cut-off period. The resulting data set consisted of 56,243,702 lines, each representing a server action, segmented into 15,799,727 sessions (October 2012 = 7,996,172, April 2013 = 7,803,555). No significant differences were observed between the data sets for the two months, and therefore the results presented here for the most part represent analysis of the aggregated data.

As described in section 3.5.3.3, as well as the generation of general statistics, analysis consisted of the generation of tables representing counts of actions, and counts of sessions including at least one instance of an action. The table also allowed for the analysis of the data broken down by referrer type and originating country of the IP address. Other work was conducted to analyse the supplementary file provided by OCLC, which contained bibliographic data and counts of the unique record pages for which record views were found in the logs. Finally, manual sampling was employed on a random filtered sample of sessions including a search action from three key session originating points; Search Engines, Libraries and the WorldCat.org homepage. The chapter concludes with a brief summary of the key findings to emerge from this phase of the research, with the bulk of discussion undertaken in Chapter 8.

Temporal values are reported in this chapter using the formats mm:ss and hh:mm:ss.

6.2 General Analysis

6.2.1 Descriptive Statistics

Descriptive statistics were generated for the aggregated log files, with means, median and mode values calculated for session duration, number of actions per session, the number of queries per session, and the number of records viewed per session. These are presented in Table 6-1. It is important to note some caveats when considering these data. First, large maximum values were found for some measures, suggesting that not all non-human traffic has been removed from the logs. The presence of this noise in the data potentially biases the averages, particularly the mean. Also, since the session duration represents the time between the first and last action logged for a session, and therefore does not account for time the user spends viewing with the last page visited, values for session duration cannot be said to perfectly capture the behaviour of users. Even considering these caveats, reviewing these data reveals that a significant proportion of sessions are clearly very brief, with over a million sessions lasting just 3 seconds.

Table 6-1: Descriptive statistics for aggregated log files

Session Duration	
Arithmetic mean	00:05:14
Standard deviation	00:17:19
Median	00:00:24
Mode	00:00:03
Mode occurrences	1,038,423
Actions per session	
Arithmetic mean	3.66
Standard deviation	5.08
Median	2
Mode	1
Mode occurrences	6,276,531
Records viewed per session	
Arithmetic mean	1.99
Standard deviation	99
Median	2
Mode	1
Mode occurrences	6,154,046

Well over a third (39.7%) consist of a single line in the log, which represents the loading of the landing WorldCat home or record page. The mode values for actions per session and

reinforce the argument that a large number of users are visiting but not engaging further with the site.

Preliminary analysis also included a determination of the number of sessions originating from different referrer types (see Table 6-2). We can first note that around 20% of all sessions were assigned the referrer type "Other" (13.6%) or "Not specified" (6.8%). Whilst it had been hoped that this figure would be lower, the fact that c.80% of sessions were assigned a referrer means the discussion of referrer types, and difference in behaviour exhibited by users from those referrers, is based on analysis of a large proportion of the dataset.

Sessions originating from a search engine are by far the most common type found in the logs, representing almost half of all traffic to WorldCat.org (47.1%). Referrals from libraries account for a further 14.4% of sessions, while traffic from other WorldCat pages (6.0%), and sessions originating at the WorldCat homepage (5.3%) in total account for around one in ten sessions in the logs. While the overall proportion of sessions from citation services, GoodReads and Wikipedia are low, they still represent a significant number of visitors to WorldCat.org.

Table 6-2: Sessions originating from each referrer type

Referrer	Sessions	% of total sessions
Search Engine	7,439,433	47.1%
Library	2,277,215	14.4%
Other	2,149,130	13.6%
Not specified	1,078,661	6.8%
WC Other	946,696	6.0%
WC Home	829,546	5.3%
Citation Service	578,133	3.7%
GoodReads.com	250,293	1.6%
Wikipedia	155,427	1.0%
OCLC Services	95,193	0.6%

6.2.2   Geographical Origin of Sessions

The log data supplied by OCLC included a field representing the originating country for each IP address. It was therefore possible to determine the geographical distribution of sessions found in the logs. The top 20 originating countries, along with the proportion of total

Table 6-3: Top 20 Originating Countries (based on IP address look-up)

Rank	Country	Sessions	% of total traffic
1	United States	7,076,767	44.8%
2	China	830,510	5.3%
3	Canada	825,261	5.2%
4	United Kingdom	584,724	3.7%
5	Germany	504,407	3.2%
6	France	365,478	2.3%
7	India	281,966	1.8%
8	Italy	267,477	1.7%
9	Indonesia	264,749	1.7%
10	Spain	242,923	1.5%
11	Netherlands	239,468	1.5%
12	Mexico	212,101	1.3%
13	Australia	205,023	1.3%
14	Brazil	201,074	1.3%
15	Poland	185,828	1.2%
16	Japan	145,215	0.9%
17	Malaysia	142,139	0.9%
18	Korea, Republic of	109,208	0.7%
19	Russian Federation	107,758	0.7%
20	Singapore	102,968	0.7%

session that traffic from each country was found to represent, are presented in Table 6-3. The United States is responsible for almost half of all traffic to WorldCat.org (44.8%), with more sessions originating there than in the rest of the top 20 countries combined. China and Canada each represent around 5% of total traffic, while the remainder of the top 20 consists of developed or newly industrialized countries. A total of 240 countries registered at least one session in the logs, with the British Indian Ocean Territory and Saint Helena sharing last place in the ranked list with one session each.

It was further possible to compare the distribution of referrer types originating from each country. Table 6-4 shows these distributions for the top ten countries. It is unsurprising to note that the US and Canada have the lowest proportion of their sessions originating from a search engine (29.8% and 30.2% respectively), and the highest beginning directly at the WorldCat.org homepage (7.5% and 5.2%), reflecting increased awareness of the service in North America. Indeed, traffic from the US accounts for 87% of all sessions originating at the WorldCat.org homepage. For all other countries, the majority of sessions are referred

Table 6-4: Distribution of referrals for top 10 countries (% of sessions originating from each country that come from each referrer)

	Search Engine	Library	Other	Not specified	WC Other	WC Home	Citation Service	GoodRead	Wikipedia	OCLC service
US	29.8%	27.5%	13.4%	8.0%	4.1%	7.5%	6.9%	1.4%	0.8%	0.6%
China	50.3%	13.9%	13.2%	1.7%	18.8%	1.3%	0.1%	0.1%	0.3%	0.3%
Canada	30.2%	4.5%	40.1%	8.4%	3.2%	5.7%	5.2%	1.4%	0.8%	0.4%
UK	58.7%	5.5%	10.1%	9.0%	5.9%	5.5%	1.1%	1.7%	2.0%	0.5%
Germany	59.6%	0.7%	14.5%	12.7%	6.2%	3.9%	0.3%	0.4%	1.1%	0.6%
France	67.0%	0.9%	12.2%	8.3%	5.4%	3.2%	0.2%	0.3%	1.7%	0.8%
India	71.1%	1.7%	8.0%	3.4%	3.8%	1.8%	0.3%	6.7%	2.1%	1.2%
Italy	77.5%	1.1%	8.1%	5.1%	3.6%	2.5%	0.3%	0.3%	0.8%	0.7%
Indonesia	69.9%	0.3%	10.8%	2.3%	10.8%	0.7%	0.1%	4.3%	0.4%	0.4%
Spain	71.2%	2%	10.6%	6.4%	4.8%	2.6%	0.3%	0.8%	1.1%	0.7%

from a search engine, with over 70% of traffic from some countries (India, Italy and Spain) originating from that source. It should be noted that despite the proportion of US traffic originating from a search engine being relatively low, separately computing the distribution of search engine referred sessions between countries shows that over a quarter (28.4%) of all such traffic originates in the US.

We can also note some apparent anomalies in the data. 40% of Canadian sessions were assigned the "Other classification", far higher than for any other country. This implies that some site or service specific to, or particularly popular in Canada, was not captured in our referrer classification scheme. Given the relatively low proportion of Canadian traffic found to originate from library referrers, it is probable that some Canadian library services were not picked up by the relevant regular expression. We also note the high proportion of Chinese sessions that originate from WorldCat.org urls excluding the homepage (the "WC Other" classification). As mentioned in Chapter 2, this classification is likely to be assigned to sessions originating from WorldCat pages outside the catalogue, such as the WorldCat Identities service, or represent session splits, where application of the session cut-off time has segmented sequential lines of logs. The explanation for the high number of Chinese sessions assigned this code is not clear.

One final point relating to Chinese traffic emerged from the analysis of sessions by geographical location and referrer. In comparing the total number of sessions found in each

month's data set (October 2012 = 7,996,172, April 2013 = 7,803,555), it is noticeable that more sessions were recorded in October 2012 than April 2013, which would appear to be contrary to the upward trend of WorldCat.org's usage reported in recent years. The geographical analysis revealed that the fall in usage is attributable entirely to a dramatic drop in search engine referral traffic from China (October 2012 = 393,159 sessions, April 2013 = 24,737 sessions). While it has proved impossible to verify, it seems plausible to suggest that some change in indexing or ranking method by one or more Chinese search engines might have caused this drop in traffic.

6.3 Session Activity

6.3.1 Queries

A count of action types representing the user submitting a query to WorldCat.org reveals 9,545,861 such actions in the logs. The distribution of types of search are presented in Table 6-5. Of the search tabs available on the WorldCat.org homepage, the default option (search everything) is the most used, with these searches accounting for almost 10% of all queries submitted. The most common search action types identified "Query submitted from the search results page" (26.3% of all searches), and "Query submitted from elsewhere on site (including record page)" (41.0%). This latter action type is necessarily unspecific; since a search bar is available from every page of WorldCat.org, discerning the origin point of a search is only possible when the URL of the

Table 6-5: Distribution of types of search action

Action Type	% of all search actions
Query submitted from elsewhere on site (including record page)	41.00%
Query submitted from the search results page	26.30%
Query submitted from WorldCat.org homepage Search Everything	10.90%
Query submitted from "Advanced search" page	9.20%
Query submitted from "No results" page	7.30%
Query submitted from WorldCat.org homepage Search Books	3.10%
Query submitted from WorldCat.org homepage Search for Library Items	0.80%
Query submitted from WorldCat.org homepage Search Articles tab	0.40%
Query submitted from WorldCat.org homepage Search DVDs	0.40%
Query submitted after clicking on spelling suggestion	0.30%
Query submitted from WorldCat.org homepage Search CDs	0.20%

action stored on the log includes a consistent reference identifying it. Such references are present for many pages (e.g. the search results page, the "No results found" page), but not for several others (e.g. an individual record page, sign-in page, view a tag page). Thus the "Query submitted from elsewhere on site (including record page)" action represents searches conducted from many locations within WorldCat.org.

Around one in ten queries (9.2%) were observed to be advanced searches. Additional analysis of the proportion of sessions including at least one instance of these actions reveals differences in the likelihood of users from different referral points undertaking advanced searches. While 2.7% of all sessions include an advanced search, 9.5% of sessions originating at the WorldCat.org homepage, and 5.8% of Library referred session do so, compared with only 0.7% of search engine originating sessions. It was also possible to determine the type of advanced search that were undertaken. The three default options available on the WorldCat Advanced Search page are Title, Keyword, and Author, and these three search types were found to be clearly the most used (see Table 6-6). It should be noted here that since advanced searches allow more than one operator to be used within the same query, meaning percentages shown in Table 6-6 can sum to greater than 100%.

Table 6-6: Use of advanced search types

Search operator	% of Advanced Searches containing operator
Title	47%
Keyword	39%
Author	37%
ISBN	7%
Subject	4%
Accession Number	1%
Journal Source	1%
ISSN	1%

In attempt to refine the descriptive statistics presented above, various averages were recalculated on the subset of sessions that were found to include at least one of the actions representing a user submitting a query. The intention was that this would better capture the patterns of behaviour for users spending time engaging with the site. These statistics were also split by referrer type, to reveal if users from different referrers were likely to exhibit different usage patterns. Of the 15,799,727 sessions in the logs, 3,576,271 (22.6%) include at least one query. However the proportion of sessions including a query was found

to vary widely between referrer types (Table 6-7). By far the most likely sessions to include a query were those originating at the WorldCat.org homepage, with 80.3% of such sessions including a search. This compares with 42.8% of library referrals, and just 8% of Search Engine traffic. It is perhaps reasonable to assume that users arriving at a WorldCat.org page from a search engine results page (or indeed from GoodReads or citation service) will land at a record page. These figures suggest that only a small minority of such users choose to engage further with the system.

Table 6-7: Proportion of sessions from each referrer that include one or more query

Referrer Type	% of sessions including at least one query
WC Home	80.3%
Library	42.8%
Not specified	31.7%
Wikipedia	30.7%
Other	29.4%
WC Other	28.0%
OCLC Services	24.6%
All Referrers	22.6%
Search Engine	8.0%
Goodreads.com	4.9%
Citation Service	3.8%

The mean number of queries for sessions including at least one search was found to be 1.5 (SD=0.7), and the median and mode 1. In total 2,244,562 of the 3,576,271 sessions (62.8%) included just one query. These figures were relatively constant for all referrer types. There were however noticeable differences in the session durations for some referrer types. While the overall median value was 00:02:21, sessions originating at the WC homepage had a median duration of 00:02:24, shorter than the median values for Search Engine (00:02:46) and Library (00:03:10) sessions.

6.3.2 Catalogue features

Analysis of the occurrences of various actions relating to the use of catalogue features was also conducted. Looking first at the use of features to refine and filter search results (see Table 6-8), we note that refining results by format was used most, with the action found almost a million times in the logs (984,627), and used in 413,740 sessions (2.6% of all sessions). Other facets were used much less frequently, the most popular being refinement by year, author and language. In total, filtering by format represented 76.1% of all refinement actions.

Table 6-8: Use of features to refine and filter search results

	% of sessions including action	Occurrences in log	% of all refinement actions
Filter search results by format	2.62%	984627	76%
Refine search results by year	0.60%	129848	10%
Refine search results by author	0.43%	85984	7%
Refine search results by language	0.31%	59384	4%
Refine search results by topic	0.10%	19714	2%
Refine search results by content	0.05%	10093	1%
Refine search results by audience	0.02%	3928	0%

Gauging the use of features such as tags, ratings and reviews proved difficult, since in most cases a user consumes the content without requiring an explicit system interaction. The few actions relating to these features that were captured in the logs related to activities such as clicking on a tag on a record page (66,358 sessions; 0.42% of all sessions), clicking on "View all tags for this item" (15,799; 0.1%), clicking on "Write a review" (4,740; 0.03%), clicking on "Add to List" (94,798; 0.6%), and viewing another user's list(s) (61,619; 0.69). In all cases the proportion of sessions undertaking these actions was very low.

Perhaps the most widely used system feature aside from the search bars was found to be the "View all editions and formats" function. This feature is available as a link from both the search results page and the record page itself, and presents the user with a list of all versions and editions of the title held in the catalogue. 7.2% of sessions used the feature from a record page, and 5.7% from the search results page, while combined these actions represented over 5% of all actions (3,348,471) recorded in the logs. Relatively frequent use was also made of the system's citation functionality, which presents a formal academic citation in a number of standard styles, with 279,927 sessions (1.8% of all sessions) including a click on the "Cite/Export" link on a record page. Interestingly these sessions

were responsible for 644,157 instances of the action, meaning that users are likely to check several citations during a session. The final feature investigated was the Author and Subject hyperlink from a record page. These links automatically generate an advanced search using operators for the author or subject in question. These features represent WorldCat.org's only real exploratory functionality beyond searching. Each feature was used in less than 1% of sessions, with the Author link used a little more often than the subject link (Author = 116,675 sessions, Subject = 95,068 sessions).

Finally, the frequency of actions relating to user accounts was investigated. Across the two months of the logs, 16,016 sessions included a click on the "Create and Account" link, although since the logs were filtered by OCLC to exclude POST data there is no way of knowing what proportion of these users completed the registration. 105,055 sessions included a click on the "Sign in" link, although similarly it is impossible to know the proportion of those that actually completed the login process, while 25,737 sessions included a click on the "My WorldCat link". Since this link is only available to users who have signed in, it establishes a minimum number of sessions by account holding users.

6.4 Analysis of Coded Sample Sessions

As described in Chapter 3, three sets of random sample sessions were generated for manual coding and analysis. These consisted of 400 sessions originating from three sources: Search Engine, Library, and WorldCat.org homepage. Rather than generate samples from the entirety of the logs, the samples were filtered to ensure that each session included at least one instance of a user executing a query. This meant that the manual analysis was focused on sessions in which the user engaged with the system. Coding primarily consisted of judging the likely task-type represented by the interactions shown in the logs, with a coding scheme consisting of seven categories; Known-item (K), Unknown-item (U), Known-item then Unknown-item (K→U), Author search (AU), Account activity (ACC), Library search (LIB), and unclassified (NA). A number of other measures were captured for each session, including session duration, number of titles and individual items viewed, use of facets and refinements, and clicks on "Related Subjects" and "Author" links. Finally the number of subtasks observed within each session was recorded.

Table 6-9 presents the distribution of Task codes for each of the three referrer types. In total 169 sessions (14.1%) proved impossible to confidently code The majority of sessions for each referrer type were coded as known-item, with 63.6% (*n*=763) of the combined

Table 6-9: Sample session Task-type coding by Referrer type

Task-type	WC Home	Library	Search Engine	Combined
Known-item	69.2%	62.4%	59.0%	63.6%
Unknown-item	11.9%	20.8%	24.0%	18.8%
Known-item then Unknown-Item	2.2%	0.0%	0.0%	0.7%
Author	2.9%	2.0%	3.0%	2.6%
WC Account	0.2%	0.0%	0.0%	0.1%
Library Info	0.2%	0.0%	0.0%	0.1%
Not classified	13.3%	14.9%	14.0%	14.1%

sample set assigned this code. Results were relatively consistent for each referrer type, with no statistically significant differences. Unknown-item tasks represented the next largest proportion of sessions, with almost a fifth (18.8%, n=226) of all sample sessions allocated this code. Differences were observed in the number of unknown-item sessions for each referrer, with almost a quarter of search engine sessions (24.0%, n=96) ascribed the code compared to 11.9% of WorldCat.org homepage sessions (n=47) and 20.8% of Library sessions (n=83). These results were found to be statistically significant (X^2 (2, N = 246) = 3.28, p <.001). All other codes were very rarely assigned, with author searches representing less than 3% of all sessions (n=31), and the other codes combined accounting for less than 1% (n=11).

The annotation process also including noting the number of subtasks involved in a session. Table 6-10. The highest number of subtasks seen in a session was 9, which occurred four times. Sessions judged to include known-item search tasks were much more likely to involve multiple subtasks than session coded as unknown-item.

Table 6-10: Number of subtasks observed by task-type

Number of Subtasks	Known-item		Unknown-item	
1	569	74.5%	221	97.9%
2	108	14.2%	2	1.1%
3	34	4.4%	2	1.1%
4	17	2.2%	0	0.0%
5	15	2.0%	0	0.0%
6	7	1.0%	0	0.0%
8	6	0.7%	0	0.0%
9	4	0.5%	0	0.0%

Mean session durations were found to be almost identical for each referrer type (WC Homepage = 08:45, Library = 08:46, Search Engine = 08:37), although the median values revealed these means are heavily biased by outlier values. The median session duration for sessions originating at the WorldCat.org homepage was found to be 02:10, compared with 03:35 for Library referred sessions, and 03:28 for search engine referred sessions. Since the distribution of known- and unknown-item sessions was different for these referrers, further analysis was undertaken to explore differences in duration and other measures between these different task-types. Table 6-11 presents the results of this analysis, all of which were tested for statistical significance using the Mann Whitney U Test.

Table 6-11: Comparison of key session and subtask metrics for Known- and Unknown-item search tasks

Session duration					
Task Type	Mean	*SD*	Median	*r*	*p*
Known-item	00:08:08	00:13:43	00:02:16	.21	.000
Unknown-item	00:14:08	00:26:12	00:05:05		
Duration per subtask (session duration / number of subtasks)					
Task Type	Mean	*SD*	Median	*r*	*p*
Known-item	00:05:08	00:08:39	00:01:49	.29	.000
Unknown-item	00:13:54	00:26:11	00:05:05		
Number of queries per session					
Task Type	Mean	*SD*	Median	*r*	*p*
Known-item	2.19	1.93	2	.16	.000
Unknown-item	2.99	3.09	2		
Number of queries per subtask (number of queries / number of subtasks)					
Task Type	Mean	*SD*	Median	*r*	*p*
Known-item	1.48	.93	1	.32	.000
Unknown-item	2.94	3.08	2		

These results indicate significant differences between known-item and unknown-item sessions. Sessions in which the user is conducting a topic search last longer than those in which the user is seeking a known item or items. Indeed when the number of subtasks is factored in, it becomes apparent that users spend considerably more time seeking

unknown-items than known ones. We also note that the number of queries per task is significantly higher for unknown-item tasks than known-item, with the difference again increased when viewed at a subtask level.

We can also observe significant differences in the number of titles viewed per session between the two search-task types (see Table 6-12). Known-item tasks appear to be completed relatively efficiently, with most queries resulting in a single record page view. Unknown-item searchers, in contrast, view more titles in general, and more items per query.

Table 6-12: Number of titles viewed and items viewed per query

Number of works viewed					
Task Type	Mean	*SD*	Median	*r*	*p*
Known-item	1.56	1.44	1	0.36	0
Unknown-item	3.78	3.11	3		
Records viewed per query (number of records viewed / number of subtasks)					
Task Type	Mean	*SD*	Median	*r*	*p*
Known-item	1.19	1.11	1.00	0.16	0
Unknown-item	2.08	1.98	1.42		

It is also revealing to compare some key measures by both search-task type and referrer type. This reveals some interesting differences between search engine and other referrer types, with users referred by search engines appearing to take significantly longer to complete their tasks, both for known-item and unknown-item search tasks (Table 6-13).

Table 6-13: Subtask duration and items viewed by search-task type and referrer type

		Known-item	Unknown-item
WC Home	Median Subtask duration	00:01:31	00:05:09
	Median Records viewed	1	3
Library	Median Subtask duration	00:01:48	000:04:35
	Media records viewed	1	4
Search	Median Subtask duration	00:03:23	00:08:27
Engine	Media records viewed	1	4

The remaining measures used to assess the sample sessions were the instances of clicks on Author and Subject links from record pages, and the use of filtering and refinement features. Only three clicks of a Subject link were observed during any of the sessions coded as known-item, compared with 91 during unknown-item searches. These 91 clicks were distributed across 34 sessions, meaning 14.9% of unknown-item sessions included the action. In contrast there was no significant difference between the use of any of the refinement tools or facets. The majority of sessions (91.0%) did not use any such feature, and the figures for the proportion of each session type that saw more than one refinement are almost identical.

Table 6-14: Distribution of refinements and filters of search results

Number of Refinements	Known-item	Unknown-item	Combined
0	91.4%	90.4%	90.6%
1	4.2%	7.4%	5.2%
2	2.0%	2.1%	2.1%
3	0.7%	0.0%	0.7%
4	1.0%	0.0%	1.0%
5	0.5%	0.0%	0.3%
6	0.0%	0.0%	0.0%
7	0.0%	0.0%	0.0%
8	0.2%	0.0%	0.2%

6.5 Discussion

Interpreting the results of transaction log analysis is inherently problematic. Such is the scale and richness of the data that it is tempting to draw wide-ranging conclusions. Yet that same scale and richness can make meaningful inferences difficult, not only because of the noise inherent in the data, but also because broad conclusions often involve assumptions about user intent and behaviour that lie beyond the lines of the log. The analysis conducted on the WorldCat.org logs suggests a number of important points of consideration for the wider project, both in terms of the quantitative analysis of action occurrences within sessions and by referrer and geographical location, and in terms of the manual coding of large sample sets. While the coding of sample sessions as part of this process was unavoidably subjective, it is hoped that the rigour with which the process was conducted means that the results can be used to inform understanding of the log data as a whole.

6.5.1   System Engagement

The results show that WorldCat.org is a truly international site, with traffic originating from 240 countries around the world. While this is testament both to the scope of the service, and the efforts made to globalise OCLCs library membership base, it is likely primarily a function of being indexed by major search engines, which drive almost half of all traffic to the site. It seems clear from the logs that a large proportion of users arriving at WorldCat.org from search engine results pages limit their engagement with WorldCat.org to the page (most likely a record page) at which they arrive. Sessions originating from citation services, GoodReads and Wikipedia appear to be similar in some ways to search engine traffic, in that sessions are most often short, and involve little interaction with the system. It seems likely, given the lack of public awareness of WorldCat beyond North America, that many users arriving at the site are ill-informed as to its purpose or use, and it is possible that this ignorance is a contributory factor to so many sessions ending without further engagement with the site.

While this characterisation of much search engine traffic is somewhat speculative, there is strong support in the data for the notion that users beginning their session at the WorldCat.org homepage exhibit different behaviour to other users. Clear differences can be observed in the logs between the characteristic of sessions starting at the homepage and sessions from other referrers, not least in the likelihood that the user goes on to interact with the system. Arriving deliberately at the WorldCat homepage implies an awareness of the service, and it is therefore unsurprising that 87% of these sessions originate from the US. Examining the results of the sample coding, we observe that WC session were generally more efficient, taking less time and fewer queries for each record page viewed. This offers some support to the notion of users familiar with the site being more expert searchers, and more effective users of its features – which is no great surprise. In terms of understanding the current and potential utility of the site, however, there are some interesting questions raised by this interpretation. Primarily, we might wonder about the difficulties of balancing system functionality and usability such that the site is better able to support transient traffic whilst continuing to serve its regular and practiced users.

6.5.2   Search-Tasks

Many of the conclusions to be drawn from analysis of the coded of sample sessions are perhaps unsurprising. We note that sessions classified as known-item are shorter, and are

frequently constituted of more than one discrete subtask, each requiring a fresh query, although query reformulation is rarely required. Unknown-item searches, in contrast, are longer, generally involve more queries, and result in a greater number of record page views. It seems certain that the majority of search-tasks undertaken on WorldCat.org are of the first kind, and, the coding of the samples sessions offers some mechanism for estimating the number of sessions involving unknown-item search tasks in the wider logs. Extrapolating the proportions of known-item and unknown-item search tasks found within the samples session, we can calculate that on a monthly basis around 71,000 search engine referred sessions, 100,000 library referred sessions, and 40,000 sessions starting at the WorldCat.org homepage include unknown-item searches. The resulting total of 211,000 sessions represents the estimated number of users each month from those referrer types engaged in unknown-item searches within WorldCat.org, and when one considers that this does not include traffic from the seven other referrer types (who constitute 33% of all traffic) it seems reasonable to suggest the figure may be between 250,000 and 300,000 sessions.

Thus while supporting unknown-item search may not be WorldCat.org primary goal, there appears to be a significant number of users who do use the system for this purpose, and thus motivation for OCLC to explore potential means of improving the discovery process.

7. PHASE FOUR: USER STUDY

7.1 Introduction

The Phase four user study was intended to address research questions 3, 4 and 5. These are:

3. When might a recommender system support users of WorldCat.org in their interactions with the system?
4. What effect does the presence of recommendations have on information seeking behaviour in the library catalogue?
5. What recommendation characteristics would be most useful to users of library catalogues?

The study adopted an IIR Task based methodology, as described in section 3.5.4.1, with 36 participants (18 undergraduates and 18 postgraduates) drawn from the University of Sheffield's student population . Each participant was required to complete two tasks on each system, with the task and system sequences determined by a randomized Latin square arrangement. Task 1 asked participants to find a range of books that would be useful for a module they were currently studying, while Task 2 first asked participants to find the publication date of a specific book (*A History of Leeds* or *Birmingham: a History of the City and its People*), and then to find a set of books that the student could take out in instead in the event of it being on loan. Data was captured using the Morae screen recording tool, and a series of markers were added after completion of the session to indicate moments participants had undertaken interactions with the systems. Participants were required to complete pre-and post-task questionnaires, and a post-study interview was conducted.

The following chapter present the results of the study, before drawing some general conclusion about the findings. In presenting results the following abbreviations are used: T1 for Task 1, T2 for Task 2, WC for WorldCat.org, AM for Amazon, UG for undergraduate and PG for postgraduate. Figures in bold in tables represent statistically significant results to $p<.05$. An abridged version of this chapter was presented as a full paper at *The Information Interaction in Context conference (IIiX)* 2014 (Wakeling et al, 2014).

7.2 Task Completion Times

Although a ten-minute time limit was applied to each task, participants were allowed to move to the next task earlier if they felt they had completed the task to their satisfaction. Results showed no significant differences in the time taken between the two systems (i.e., T1 WC vs. T1 AM, and T2 WC vs. T2 AM). There were, however, differences in the time taken to complete task by the UG and PG participants (see Table 7-1). In both cases PG students were found to take longer, with the difference for both iterations of Task 1 being statistically significant ($p<0.05$).

Table 7-1: Median time taken to complete tasks (statistically significant differences between UG and PG groups in bold)

	UG	PG	*r*	*p*
T1 WC	**08:49**	**10:00**	**0.41**	**0.014**
T1 AM	**08:47**	**09:51**	**0.49**	**0.003**
T2 WC	07:40	09:24	0.16	0.335
T2 AM	08:05	09:12	0.22	0.184

7.3 Actions

Table 7-2 presents the mean frequencies with which the pre-defined action markers were applied to the participant sessions in Morae, with the exception of the "Find Item" and "Get publication date" markers (these not recording functional system interactions). Clicking on a search result was the most frequent action logged for three of the four task iterations (T1 WC $n=307$, T1 AM $n=200$, T2 WC $n=272$), with clicking on a recommendation the most frequently logged action for T2 AM. The number of queries executed was similar across both systems, although fewer queries were logged for T2 than T1. Participants were, however, much more likely during both tasks to view additional pages of search results in WC than AM. It is also noticeable that the mean number of clicks on an Amazon recommendations was significantly higher for Task 2 than Task 1 (T1 AM $M=2.23$, T2 AM $M=5.56$, t(35)=5.90, $p<0.001$).

Table 7-2: Mean frequency of marker allocation (statistically significant differences between WC and AM iterations in bold)

	Task 1		Task 2	
Marker	WC	AM	WC	AM
Q (Search)	3.39 (3.84)	3.64 (3.36)	2.83 (1.7)	2.36 (1.46)
A (Advanced Search)	0.47 (0.94)	0.22 (0.72)	**0.69 (1.24)**	**0.22 (0.48)**
N (Next page of search results)	**2.94 (3.53)**	**1.25 (2.32)**	**2.97 (3.03)**	**0.92 (1.87)**
C (Click Search Results)	**8.53 (5.17)**	**5.56 (3.75)**	7.56 (4.17)	4.11 (3.65)
E (Click on Related Subject)	**1.00 (1.49)**	**0.14 (0.42)**	**1.06 (1.04)**	**0.03 (0.17)**
G (Click on a facet)	1.53 (1.7)	1.17 (1.54)	**2.33 (2.37)**	**0.5 (0.81)**
L (Look Inside / Preview)	**0.03 (0.17)**	**0.94 (1.37)**	0.00 (0.00)	0.31 (0.62)
W (Author Search)	0.25 (0.65)	0.36 (1.02)	**0.56 (0.81)**	**0.31 (0.58)**
H (View a user list)	0.11 (0.67)	n/a	0.00 (0.00)	n/a
R (Click a Recommendation)	n/a	2.33 (2.89)	n/a	5.56 (3.23)
M (See more Recommendations)	n/a	2.03 (3.6)	n/a	1.83 (2.4)

Interpreting the feature markers as participant actions, the median number of actions by UG and PG participants for each task and system were calculated (see Table 7-3). No significant differences were observed in the number of actions undertaken for the two iterations of task 1. The median number of actions was identical for both UG and PG groups for Task 2, with participants using significantly fewer actions on Amazon than WorldCat.org

Table 7-3: Median number of actions for each task iteration (statistically significant differences between WC and AM iterations in bold)

		WC	AM	*r*	*p*
UG	T1	20	18.5	0.01	0.948
	T2	**18**	**14**	**0.38**	**0.023**
PG	T1	17	19.5	0.09	0.6
	T2	**18**	**14**	**0.36**	**0.029**
All	T1	19	19.5	0.04	0.717
	T2	**18**	**14**	**0.37**	**0.002**

(WC *Mdn*=18, AM *Mdn*=14, *r*=0.37, *p*<0.005).

Also of interest was the temporal distribution of actions within sessions, particularly the points at which participants tended to use recommendations. To determine this, the time stamps for each action were converted into a percentage representing the proportion of the session that had passed before the action took place (i.e., time of action divided by total session time for the task). The distribution of actions was then determined by quartile. The results of this analysis relating to markers Q (execute a query) and R (click a recommendation) are shown in Table 7-4. For both tasks almost half of all queries executed occurred in the first quarter of the session, with a relatively even distribution thereafter. The distribution of recommendation clicks for T1 tends toward later in the session, whilst for T2 this distribution is reversed.

Table 7-4: Temporal distribution of Queries and Recommendation Clicks by quartile

		Q1	Q2	Q3	Q4
Q (Execute a query)	T1	45%	19%	17%	19%
	T2	47%	15%	25%	13%
R (click a recommendation)	T1	18%	25%	26%	31%
	T2	35%	28%	24%	14%

7.4 Number of Books Found

When completing the study participants were asked to save the title of any books they considered relevant to the task to a Word document. While it is acknowledged that the number of books found is not in itself a measure of task success, it is nonetheless revealing to compare the numbers of books copied to the Word document by participants across the tasks and systems (Table 7-5). Looking first at Task 1, across all participants the number of books found were almost identical in both systems (WC T1 *Mdn*=7.5; AM T1 *Mdn*=7.5). Further analysis, however, reveals different results from UG and PG participants, with the former finding more items through Amazon (WC *Mdn*= 7.0, AM *Mdn*=9.0, *p*=0.08), and the latter more through WorldCat.org (WC *Mdn*=8.5, AM *Mdn*=6.5, *p*<0.05).

Table 7-5: Mean and median number of books found (statistically significant differences between WC and AM iterations in bold)

		WC	WC	AM	AM		
		M (*SD*)	*Mdn*	Mean (*SD*)	*Mdn*	*r*	*p*
UG	T1	7.94 (5.32)	7	9.28 (4.56)	9	0.29	.083
	T2	**4.72 (2.24)**	**4.5**	**6.50 (2.70)**	**6.5**	**0.47**	**.004**
PG	**T1**	**9.22 (5.78)**	**8.5**	**7.7 (5.87)**	**6.5**	**0.34**	**.04**
	T2	6.50 (3.47)	5.5	7.61 (4.10)	7	0.27	.101
Total	T1	8.58 (5.51)	7.5	8.53 (5.23)	7.5	0	1
	T2	**5.61 (3.02)**	**5**	**7.00 (3.48)**	**7**	**.36**	**.002**

Although the UG result does not reach statistical significance, further analysis was conducted to establish the differences in the numbers of books found for each task at a participant level (i.e., the number of books found in Amazon minus the number of books found in WorldCat, with a positive value indicating a greater number of books found in Amazon). Results showed that undergraduates were indeed likely to find more books on Amazon, and postgraduates on WorldCat.org (UG T1 *Mdn*+1.0, PG T1 *Mdn*=0.5, *r*=0.43, *p*<0.01). This did not hold true for T2, with both PG and UG students finding more books on Amazon than WorldCat.org. This is reflected in the overall figures for T2, which show significantly more books found on Amazon (WC *Mdn*=5.0, AM *Mdn*=7.0, *r*=0.36, *p*<0.05).

Combining the books found and action data, the number of actions per book found were calculated (number of actions/number of items found) (see Table 7-6). Median values were calculated for this derived variable. Results were lower for the Amazon iteration of both tasks, with the exception of PG performance for T1 (T1 UG WC *Mdn*=2.5, AM *Mdn*=2.2, *r*=0.23, *p*=0.177, PG WC *Mdn*=2.4, AM *Mdn*=2.9, *r*=0.44, *p*<0.01, T2 all participants WC *Mdn*=3.3, AM *Mdn*=2.2, *r*=0.56, *p*<0.001).

Table 7-6: Actions per book found (statistically significant differences between WC and AM iterations in bold)

		WC		AM			
		M (SD)	Mdn	M (SD)	Mdn	r	p
UG	T1	3.03 (2.15)	2.5	2.29 (1.85)	2.1	0.23	0.177
	T2	**4.65 (3.32)**	**4.0**	**2.81 (1.64)**	**2.3**	**0.4**	**0.016**
PG	**T1**	**2.37 (0.81)**	**2.4**	**3.72 (2.54)**	**2.9**	**0.44**	**0.009**
	T2	**3.73 (2.54)**	**2.6**	**2.40 (0.99)**	**2.1**	**0.41**	**0.013**
Total	T1	2.70 (1.64)	2.4	2.98 (2.23)	2.3	0.15	0.363
	T2	**4.31 (3.17)**	**3.3**	**2.61 (1.35)**	**2.2**	**0.56**	**0.001**

Finally, the time taken per item found was calculated (task duration / number of books found). From the results presented in Table 7-7 we once again observe that discovery was more efficient for the Amazon iteration of T2, with the median time taken per item found almost 30 seconds lower (T2 WC median = 01:34, AM median = 01:08, *r*=0.4, *p*=.001). While the data show PG participants taking longer to find each item on Amazon for T1, and UG less time, the results did not reach statistical significance.

Table 7-7: Time taken per book found (statistically significant differences between WC and AM iterations in bold)

		WC		AM			
		M (SD)	Mdn	M (SD)	Mdn	r	p
UG	T1	01:17 (00:49)	00:56	01:00 (00:39)	00:54	0.29	0.085
	T2	**02:04 (01:08)**	**01:43**	**01:26 (00:57)**	**01:08**	**0.46**	**0.006**
PG	T1	01:19 (00:42)	01:09	02:00 (02:10)	01:23	0.24	0.147
	T2	01:52 (01:21)	01:31	01:21 (00:40)	01:08	0.32	0.055
Total	T1	01:18 (00:45)	01:02	01:30 (01:39)	01:02	0.014	0.905
	T2	**01:58 (01:14)**	**01:34**	**01:24 (00:48)**	**01:08**	**0.4**	**0.001**

7.5 Impact of Recommendations

Further analysis focused on the apparent utility and effectiveness of Amazon's item-level recommendations. Given the disparities between participant performance in Tasks 1 and 2, we first compare the interactions with recommendations across the two tasks (Table 10). Participants view significantly more recommendations and find more items from recommendations during Task 2 than Task 1. Furthermore, the "conversion" rate (i.e., the likelihood of participants considering a recommended item relevant) is significantly higher for Task 2, and a far greater proportion of total saved items are discovered via a recommendation. All differences are found to be significant at $p<0.001$, and no significant differences were found between the UG and PG sub-groups. Results indicate that recommendations are more heavily used and considered more relevant for T2 compared with T1.

Table 10: Amazon recommendations viewed - T1 vs. T2

M (*SD*)	T1	T2	*r*	*p*
Recommendations Viewed	2.33 (2.89)	5.56 (3.23)	0.52	0.000
Number of Items found from a Recommendation	1.5 (1.96)	4.31 (2.857)	0.50	0.000
Items found from a recommendation / Recommendations viewed	0.10 (0.08)	0.79 (0.25)	0.47	0.000
Items Found from Recommendation / Total Items Found	0.20 (0.24)	0.62 (0.31)	0.56	0.000

A similar comparison for UG and PG participants during Task 1 was performed, but no statistically significant differences were observed. However, a notable observation here was a very unbalanced distribution of recommendation clicks for PG participants. Whilst the mean number of recommendations clicked by PG students was 2.6 (*SD*=3.634), higher than for undergraduates (*M*=2.1, *SD*=2.56), exactly half of the PG participants (*n*=9) did not click on a single recommendation during Task 1, whilst four participants clicked on 7 recommendations or more. This suggests significant individual differences among PG students in the use of recommendations.

7.6 Known-Item Task

The first part of T2 asked participants to find the publication date of one of the two local history books. Completing the task on WorldCat, 75% participants (*n*=27) completed the task with just two actions; "Execute a Query" (Q) and "Click Search results" (Q→C). Of the remainder: three used the advance search function rather than the standard search before clicking a result (A→C); two were able to ascertain the publication date from the snippet shown on the search results page (Q); two participants ran two queries before clicking a search result (Q→Q→C); and two ran advanced searches after an initial search (Q→A→C). The mean time taken to find the publication date was 48 seconds, with the quickest participant taking just 10 seconds, and the longest almost three minutes (02:53). No significant differences were observed in either the speed or behaviour of undergraduate and postgraduate participants. On Amazon, 67% (*n*=24) of participants found the item following a query and a click (Q→C), and one from an advanced search and click (A→C), while another five participants were able to find the publication date from the search results page (i.e. Q only). Four participants needed to reformulate their query to find the relevant result (Q→Q→C), with one participant trying a further advanced search before locating the item (Q→A→Q→C). One other participant initially selected an incorrect item from the search results list (Q→C→C). The mean time taken to find the publication date on Amazon was 45 seconds, with the fastest participant taking 00:19 and the slowest 02:18. Once again there were no significant differences between UG and PG populations.

These results suggest that both systems, and the participants, were able to deal comfortably with the known-item search tasks they were set. The vast majority of participants were able to locate the required information in less than a minute (WC 81%, n=29; AM 83%, n=31) using only a single query (WC 89%, *n*=31; AM 86%, *n*=32).

7.7 Action Sequences

7.7.1   N-Grams

To better understand search behaviours, common sequences of actions (n-grams) were calculated for each task iteration, whereby *n* consecutive actions are treated as a unit (see Table 7-8). It should be noted here that n-grams containing repeated actions represent two possible types of system interaction. For example, the 4-gram RRRR might represent a participant following a trail of item-level recommendations (i.e., clicking on a recommended item, then clicking an item recommended on the subsequent page and so

Table 7-8: Top 5 n-grams of participant action sequences including % of total actions for each task and mode rank position of each n-gram sequence (in parentheses)

	T1 WC		T1 AM		T2 WC		T2 AM	
	2-gram sequences							
1	CC	22.5 (8)	CC	12.0 (8)	CC	17.7 (7)	RR	22.6 (4)
2	QC	9.2 (1)	QC	10.5 (1)	QC	8.8 (1)	QC	10.8 (1)
3	CN	8.5 (11)	QQ	7.2 (1)	CN	7.7 (13)	CC	9.4 (15)
4	NC	7.4 (7)	MM	7.2 (7)	NC	7.0 (16)	CR	6.6 (2)
5	QQ	6.6 (1)	RR	5.2 (3)	FF	5.6 (4)	RM	6.1 (3)
	3-gram sequences							
1	CCC	11.1 (2)	CCC	5.5 (7)	CCC	8.5 (7)	RRR	15.5 (3)
2	CNC	5.1 (6)	MMM	4.4 (5)	CNC	4.5 (15)	QCR	5.5 (1)
3	QCC	5.1 (1)	QQQ	3.7 (1)	NCN	3.1 (16)	CRR	5.3 (2)
4	QQQ	4.8 (1)	QCC	3.2 (1)	QCC	3.0 (1)	CCC	4.5 (9)
5	CCN	4.3 (8)	QCR	3.0 (1)	CCN	3.0 (7)	RMR	3.9 (6)
	4-gram sequences							
1	CCCC	6.1 (2)	CCCC	2.8 (6)	CCCC	4.8 (7)	RRRR	10.0 (3)
2	QQQQ	3.8 (1)	MMMM	2.5 (17)	CCNC	2.6 (22)	CRRR	4.9 (2)
3	CCNC	2.9 (8)	QQQQ	2.3 (-)	CNCC	2.2 (7)	QCRR	4.7 (1)
4	CQCC	2.6 (7)	MMMR	1.5 (-)	CNCN	1.8 (15)	RRRQ	2.5 (3)
5	QCCC	2.4 (1)	CQCC	1.5 (6)	CECC	1.7 (8)	CCCC	2.5 (14)

on). However, it may also represent a participant utilizing multi-tab browsing (i.e., opening links in new tabs to a number of recommendations from a single item). Despite this uncertainty the n-grams reveal some important differences in user behaviour across the two systems and tasks. Common n-grams relating to tasks undertaken on WC are more likely to include executing a query and clicking on a search result than those relating to Amazon. Also, whilst each set of WC n-grams includes at least one action relating to viewing the next page of search results, this action does not occur in any of the most common Amazon n-grams. These results might suggest that the presence of recommendations has lessened participants' reliance on the usual query-click sequence. Comparing the Amazon n-gram sets for T1 and T2 also reveals an increasing proportion of the total n-grams including clicks on recommendations, and it is interesting to note that 10% of all 4-grams for T2 AM are repeated clicks on recommendations (RRRR).

The final set of n-grams (Table 7-9) represent the most common 2-grams that occur immediately prior to the participant finding an item they deem relevant to the task

(presented with the proportion of total items found that each 2-gram represents). The most common actions leading to an item being found for T1 is the standard query-click on both WorldCat.org and Amazon. Both sets of n-grams suggest a reliance on the search function. For the WorldCat.org iteration of T2 we observe the most common sequence begins with the participant viewing the next page of search results. A much higher incidence of sequences involving recommendations on Amazon can be observed for Task 2 than Task 1. Thus the behaviour of favouring recommendations over search implied by the other n-gram may be more noticeable for T2 than T1.

Table 7-9: Top five 2-grams leading to "Find Item" markers (F) and % of total Find Item markers (n-gram count / total "Find Item" markers)

	T1 WC		T1 AM		T2 WC		T2 AM	
1	QC	12.3	QC	4.3	NC	13.7	CR	2.8
2	NC	8.1	MR	1.4	CC	6.0	MR	2.6
3	CC	8.1	CC	1.3	QC	4.3	QC	2.4
4	FC	4.5	FC	1..3	FC	4.3	RR	2.2
5	EC	2.9	CL	1.0	EC	4.2	CC	1.3

7.7.2   Transition Probability Matrices

Transition probability matrices were also calculated for each task iteration. These matrices show the probability that each action is followed by each other action, with the rows representing each action, the columns the subsequent action, and the elements of the matrix the probability of the transition. The probability of transition between all Morae markers were calculated, with the exception of the "Find Item" and "Get publication date" markers, since these do not represent functional system interactions. These markers were therefore removed from the action sequences used to calculate the transition

Table 7-10: Transition Probability Matrices for all actions, tasks and systems

Task 1 WorldCat

	A	C	E	G	H	L	N	Q	W
A	0.18	0.71	0.00	0.06	0.00	0.00	0.00	0.06	0.00
C	0.01	0.49	0.11	0.04	0.01	0.00	0.18	0.12	0.03
E	0.03	0.66	0.00	0.06	0.00	0.00	0.14	0.11	0.00
G	0.04	0.39	0.00	0.37	0.00	0.00	0.17	0.04	0.00
H	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
L	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
N	0.00	0.48	0.01	0.08	0.00	0.00	0.38	0.03	0.01
Q	0.03	0.48	0.02	0.11	0.00	0.00	0.03	0.34	0.00
W	0.00	0.44	0.11	0.00	0.00	0.00	0.00	0.44	0.00

Task 1 Amazon

	A	C	E	G	L	M	N	Q	R	W
A	0.25	0.50	0.00	0.13	0.00	0.13	0.00	0.00	0.00	0.00
C	0.01	0.39	0.01	0.03	0.12	0.07	0.07	0.14	0.15	0.02
E	0.00	0.60	0.00	0.00	0.00	0.00	0.00	0.40	0.00	0.00
G	0.00	0.46	0.00	0.24	0.00	0.00	0.24	0.05	0.00	0.00
L	0.00	0.38	0.03	0.03	0.07	0.14	0.07	0.07	0.07	0.14
M	0.01	0.03	0.01	0.01	0.00	0.60	0.01	0.04	0.28	0.00
N	0.03	0.38	0.00	0.18	0.00	0.00	0.35	0.08	0.00	0.00
R	0.01	0.06	0.01	0.05	0.13	0.15	0.00	0.16	0.39	0.04
S	0.01	0.50	0.01	0.10	0.00	0.00	0.04	0.34	0.00	0.01
W	0.00	0.46	0.00	0.00	0.00	0.00	0.00	0.23	0.23	0.08

Task 2 WorldCat

	A	C	E	G	H	L	N	Q	W
A	0.13	0.54	0.04	0.13	0.00	0.00	0.04	0.13	0.00
C	0.04	0.43	0.12	0.07	0.00	0.00	0.19	0.10	0.06
E	0.03	0.50	0.08	0.18	0.00	0.00	0.08	0.11	0.03
G	0.01	0.35	0.00	0.41	0.00	0.00	0.15	0.07	0.00
H	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
L	0.03	0.54	0.02	0.14	0.00	0.00	0.09	0.18	0.00
N	0.01	0.44	0.01	0.08	0.00	0.00	0.33	0.10	0.02
Q	0.11	0.32	0.11	0.05	0.00	0.00	0.16	0.21	0.05
W	0.04	0.44	0.06	0.14	0.00	0.00	0.17	0.11	0.03

Task 2 Amazon

	A	C	E	G	L	M	N	Q	R	W
A	0.17	0.67	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00
C	0.02	0.37	0.00	0.02	0.03	0.04	0.16	0.07	0.26	0.03
E	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
G	0.00	0.40	0.00	0.20	0.00	0.07	0.07	0.13	0.00	0.13
L	0.00	0.27	0.00	0.00	0.00	0.09	0.00	0.00	0.64	0.00
M	0.02	0.03	0.00	0.00	0.00	0.40	0.00	0.08	0.46	0.02
N	0.00	0.56	0.00	0.11	0.00	0.00	0.19	0.15	0.00	0.00
R	0.01	0.02	0.01	0.01	0.04	0.17	0.00	0.10	0.64	0.01
S	0.00	0.71	0.00	0.07	0.00	0.00	0.05	0.11	0.02	0.04
W	0.00	0.44	0.00	0.11	0.00	0.00	0.11	0.11	0.22	0.00

probabilities.[9] The probability of finding an item subsequent to each action was calculated separately. The complete transition matrices for both iterations of both tasks can be found in Table 7-10. It was noted that in several cases, the relative scarcity of certain markers within the data set make the aggregated calculation of transition probability of questionable value (for example there was only one incidence of user clicking an Amazon "related subject" link. Since that user subsequently ran a query, the transition probability of E→Q was calculated as 1 for that task). Therefore analysis was principally directed towards matrices reduced in both of two ways; 1) showing only transitions from action types representing 10% or more of the total number of aggregated actions recorded for a given task iteration, and 2) showing only transitions with a probability of greater than 0.05. While these figures are to a certain degree arbitrary, calculations show that the reduced matrices still represent over 80% of all 2-grams for each of the four task iterations.

Visualisations of the reduced transition matrices relating to all participants were created, and these are shown in Figure 7-1 and Figure 7-2 . The size of the nodes is proportional to the number of occurrences of that action in the total data set, and the edge labels show the probability of transition between two nodes. The diagrams provide a useful visualization of the additional opportunities for discovery offered by Amazon recommendations. Whilst using WorldCat.org, participants are restricted to using the search functionality (nodes Q, C and N). In Amazon, however, recommendation functionality offers users a route out of the query-click-reformulate cycle. The probability of transitions from the recommendation functions (nodes R and M) to the search functions are also relatively low (T1 R→Q=0.16, T2 R→Q=0.10). The diagrams also reveal the likelihood of clicks on search results and recommendations leading directly to the discovery of a relevant item (F). For T1 these probabilities are relatively constant (WC C→F=0.61, AM C→F=0.56 and R→F=0.57), with the search results of both systems and Amazon's recommendations equally likely to present items the participant considers relevant. For T2, however, Amazon's recommendations are much more likely to be considered relevant than the search results of both systems (WC C→F=0.44, AM C→F=0.39 and R→F=0.77).

---

[9] The full sequence of markers for each task included many sequences such as this: … Q→C→F→Q … (i.e. Query, Click, Find an Item, Query). Removal of the Find Item markers allowed the two system transitions (Q→C and C→Q) to be captured.

Figure 7-1: Task 1 Transition Probability Diagram (WorldCat.org on left; Amazon on right)
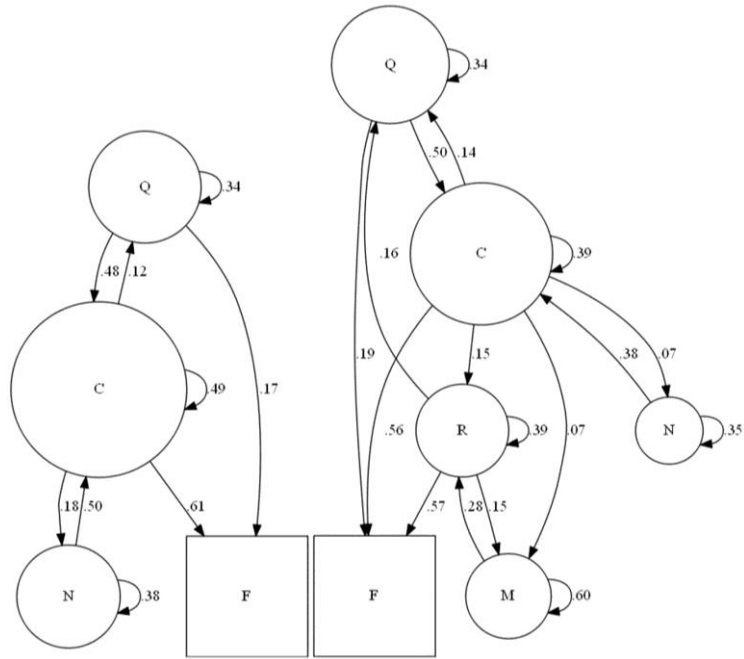
C = Click a Search Result, F = Find an Item, Q = Execute Query,
N = View Next Page of Search Results, M = Click "See more Recommendations",
R = Click a Recommendation

Figure 7-2: Task 2 Transition Probability Diagram (WorldCat.org on left; Amazon on right)
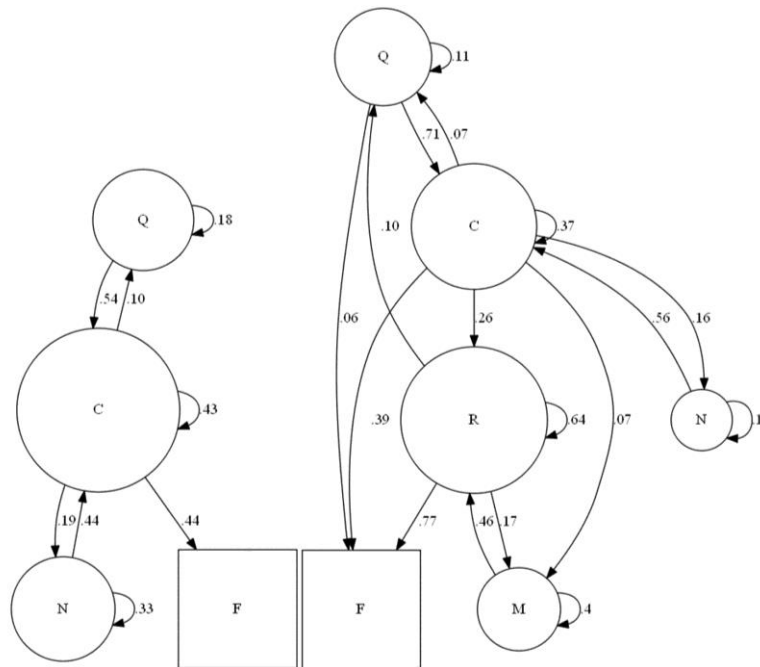
C = Click a Search Result, F = Find an Item, Q = Execute Query,
N = View Next Page of Search Results, M = Click "See more Recommendations",
R = Click a Recommendation

Separate transition matrices were also created for UG and PG users, in order to identify any differences between these participant groups in their use of the systems. Consideration was given to determining the best means of calculating the significance of apparent differences in the value of matrix elements between UG and PG participants. While t-tests have been used for these purposes elsewhere in the literature (e.g. Marchioni 1989), the relative sparsity of the data within the transition matrices makes this test problematic. Ultimately a simple subtraction was conducted to determine the difference between each element of the comparable matrices (UG and PG).

Table 7-11: Differences in the probability of transition for UG and PG participants. A negative value indicates the probability is greater for PG students.

Task 1 WorldCat						
	Q	C	N	R	M	F
Q (Search)	**-0.12**	0.07	-0.01	n/a	n/a	**-0.18**
C (Click Search Results)	0.04	0.04	-0.05	n/a	n/a	**-0.14**
N (Next page of search results)	0.00	-0.01	0.04	n/a	n/a	0.10

Task 1 Amazon						
	Q	C	N	R	M	F
Q (Search)	**-0.19**	**0.18**	-0.04	0.00	0.00	-0.07
C (Click Search Results)	0.04	-0.03	-0.04	0.01	0.07	0.09
N (Next page of search results)	0.08	**-0.21**	**0.25**	0.00	0.00	0.02
R (Click recommendation)	0.01	0.04	0.00	0.00	**-0.12**	0.08
M (See more Recommendations)	0.02	0.00	-0.03	-0.05	**0.10**	-0.05

Task 2 WorldCat						
	Q	C	N	R	M	F
Q (Search)	-0.03	0.02	0.04	n/a	n/a	0.05
C (Click Search Results)	0.05	-0.03	-0.03	n/a	n/a	-0.08
N (Next page of search results)	0.04	-0.07	0.02	n/a	n/a	**-0.14**

Task 2 Amazon						
	Q	C	N	R	M	F
Q (Search)	-0.01	0.01	-0.02	0.00	0.00	-0.06
C (Click Search Results)	-0.02	-0.03	-0.02	0.03	0.01	-0.08
N (Next page of search results)	-0.05	**-0.16**	**0.11**	0.00	0.00	**-0.12**
R (Click recommendation)	-0.02	-0.01	0.00	-0.01	0.03	-0.06
M (See more Recommendations)	0.03	0.02	0.00	**-0.10**	0.06	0.05

Table 7-11 presents the results of this calculation on the reduced matrices. Since the PG probability was subtracted from the UG value, a negative value indicates that the transition was more common for PG participants, and a positive value that it was more common for undergraduates. While values greater than 0.10 and -.10 have been highlighted in the table, this is for illustrative purposes, and does not denote statistical significance. In general

the matrices for both groups are very similar. It is interesting to note that a number of the elements with greater differences in probability relate to the "Next page of search results" action on Amazon, with UG students more likely to continue viewing additional pages of results, while PG participants more often click on a result.. The results also suggest that PG participants are more likely to follow a query with another query (i.e. immediate reformulation), while there is a higher probability that UG students will click a result.

7.7.3 Temporal Sequence Visualisations

While *n*-grams and transition probability matrices offer helpful means of interrogating the data collected during the study, it is important to note that it does so in aggregated form. In attempting to draw meaningful conclusions about the role recommendations play in individual users' searchers, it is also instructive to consider the micro-level of individual interactions with the system. Comparing and conceptualising individual participant sessions can be problematic if relying on simple lists of actions or other forms of data output. To address this, a form of visualisation was developed that allows for the easier interpretation of an individual's interactions with the system over the duration of the session. These Temporal Sequence Visualisations (TSVs) use a horizontal bar to represent a participant's session, with the length of the bar representative of the total length of the session. Each action taken by the participant during the session is represented as a small coloured block on the bar, with each type of action given a distinctive colour. As with the transition diagrams, for the sake of clarity the visualisations produced here include only those actions accounting for greater than 10% of total actions.

The four TSVs relating to each task and iteration can be found in Figure 7-3 to Figure 7-6. Perhaps the most striking aspect of these visualisations is the sheer range of strategies employed by participants in completing the tasks. While the diagrams relating to WC tasks appear at first glance to be reasonably uniform, a closer look reveals great variety in the methods used to discover books through the system. In particular we note the number and position of queries within the action sequences, and the number of search results viewed. While the WorldCat TSVs in particular show participants located squarely within the query-response paradigm, it is also clear that there are distinct variations in user behaviour within that paradigm. For the tasks completed on Amazon, we notice a tendency for bunches of red recommender-related actions to be interspersed among groups of blue blocks representing searching.

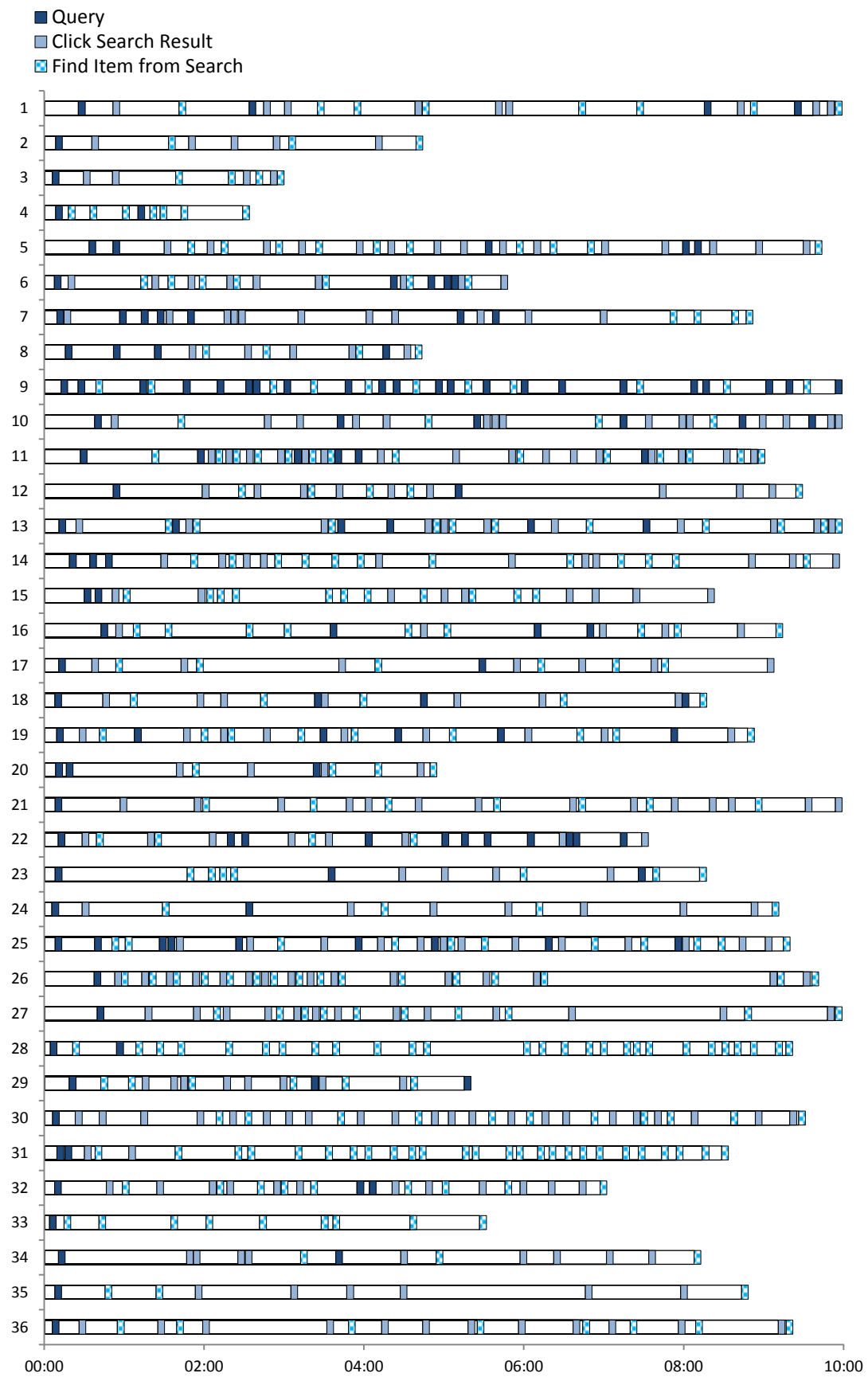**Figure 7-3: Temporal Sequence Visualisation for Task 1 WorldCat**

Figure 7-4: Temporal Sequence Visualisation for Task 1 Amazon
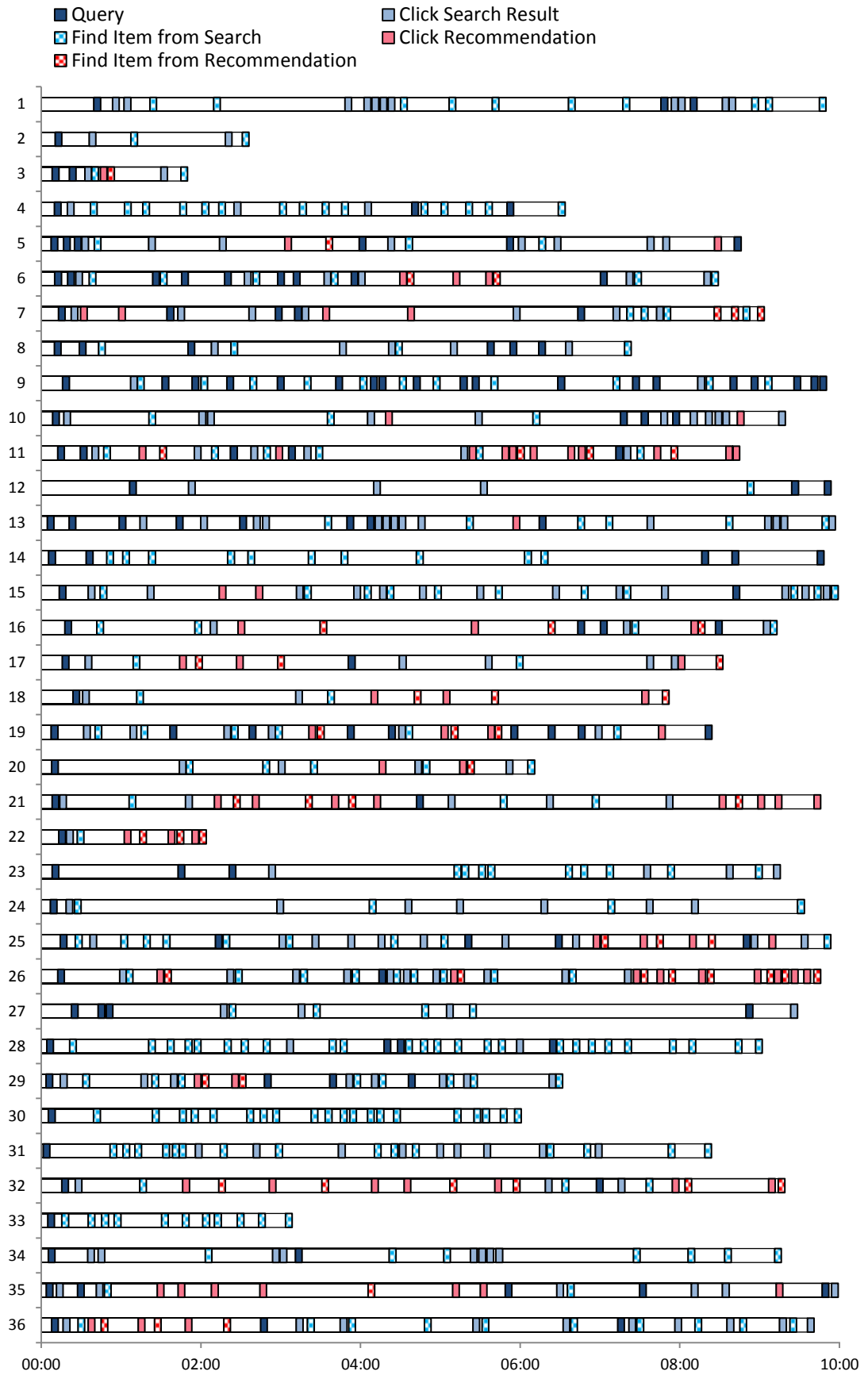
Figure 7-5: Temporal Sequence Visualisation for Task 2 WorldCat
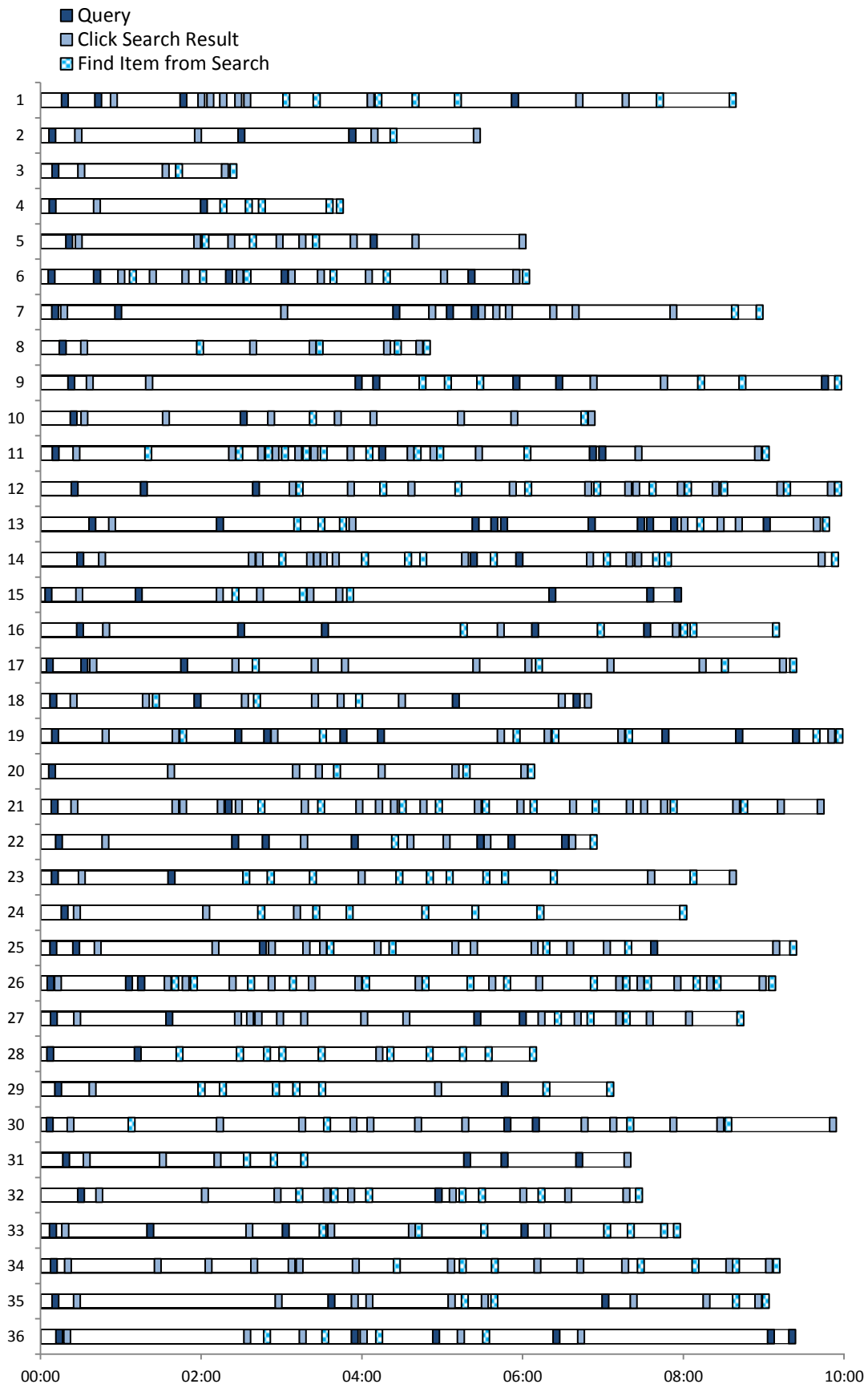
Figure 7-6: Temporal Sequence Visualisation for Task 2 Amazon
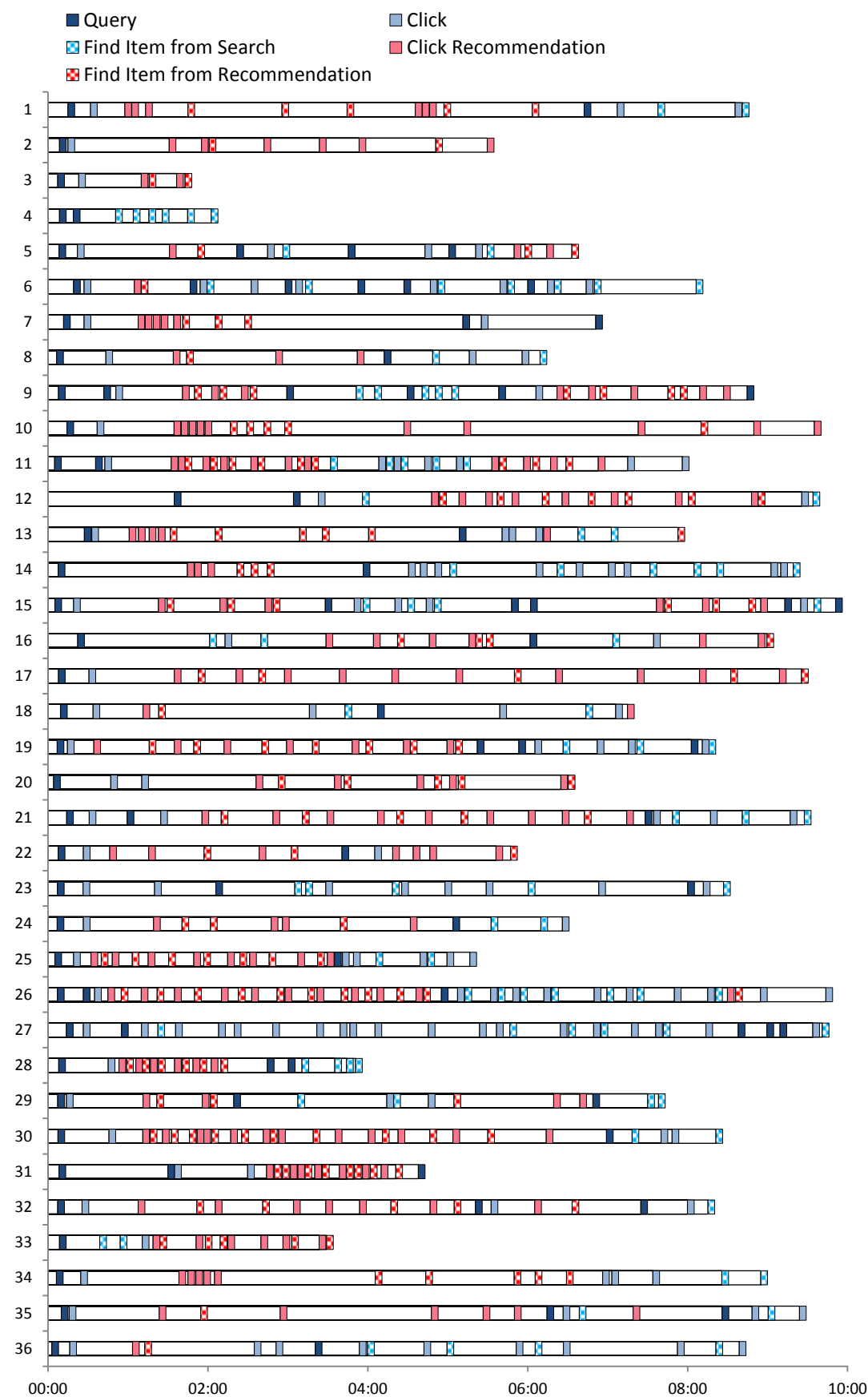
It is also revealing to examine some individual sessions in more detail. One thing that becomes clear in doing so is that not all users exhibited different behaviour across the two systems. Participant 13, for example, utilised a similar strategy for both systems when completing the two iterations of Task 1 (see Figure 7-7). The participant executed a similar number of queries on both system (WC Q=6, AM Q=7), and viewed a comparable number of items (WC C=11, AM C=14). Aside from a lone interaction with a recommendation, there was no utilisation of any additional functionality offered by Amazon.
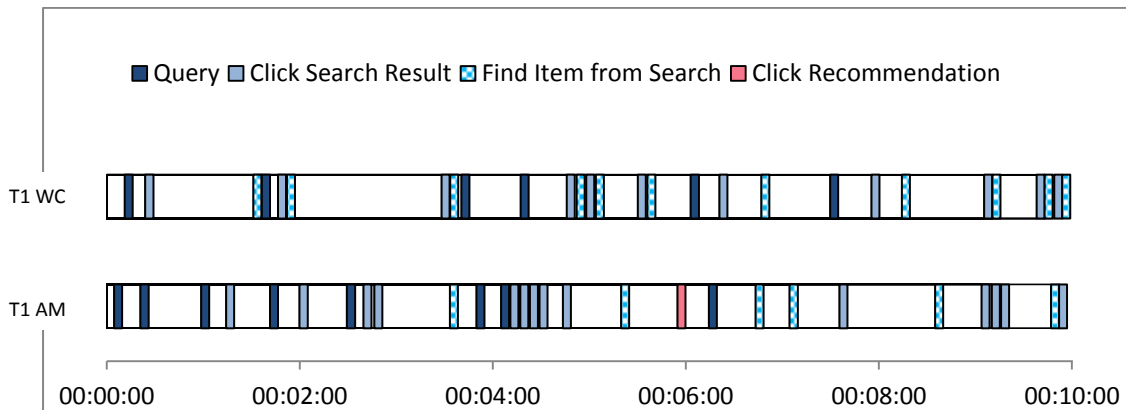
Figure 7-7: TSVs for Participant 13, T1 WC and T1 AM

In contrast, participant 11 exhibited quite different behaviour on the two systems, principally through interacting extensively with recommendations. It is notable too that the bulk of recommendation views are in the second half of the session. Again we observe the
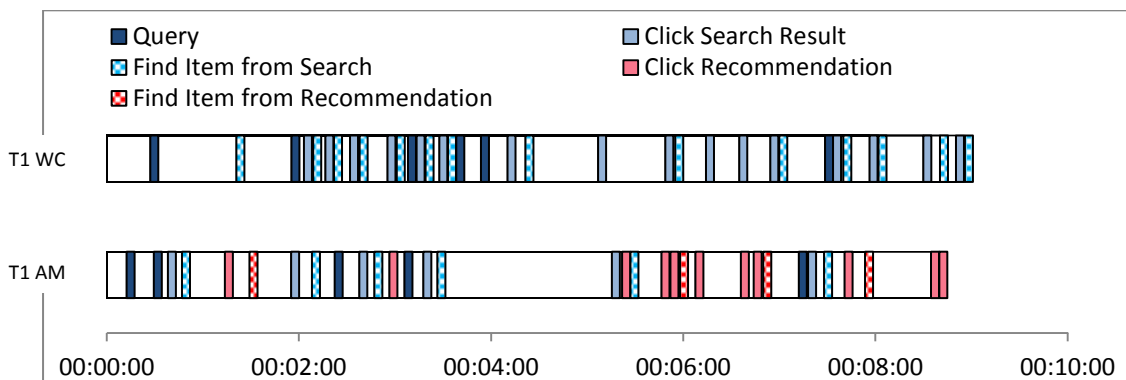
Figure 7-8: TSVs for Participant 11, T1 WC and T1 AM

recommender interactions bunched together, with the participant following chains of recommendations. It is worth considering at this point the different levels of cognitive effort required to engage with recommendations compared with the query-click process.

222

Recommendations offer the user a manageable number of easy options, accessible with a single click, that in turn take that user to another item, and another set of recommendations. The ease with which this browsing behaviour can be accomplished stands in contrast to the effort that searching in the traditional sense requires.

Some participants also demonstrate dramatic differences in behaviour between the two tasks (see Figure 7-9). Participant 9 did not follow a single recommendation during T1 on Amazon, preferring instead to refine their query numerous times (T1 AM Q=19). We also observe that the participant was happy to determine an item relevant without clicking on a link to the record page. This may be a reflection of the participant's confidence in their domain knowledge, and by extension ability to judge relevance from a relatively small amount of information; it might also reflect the effectiveness of Amazon's search results page in providing enough information to judge relevance, since we see a similar pattern for Task 2, where it is safe to assume the level of domain knowledge is much lower. Once again, for T2 on Amazon we observe the chaining of recommendations, as the participant explores relevant titles without recourse to executing queries.
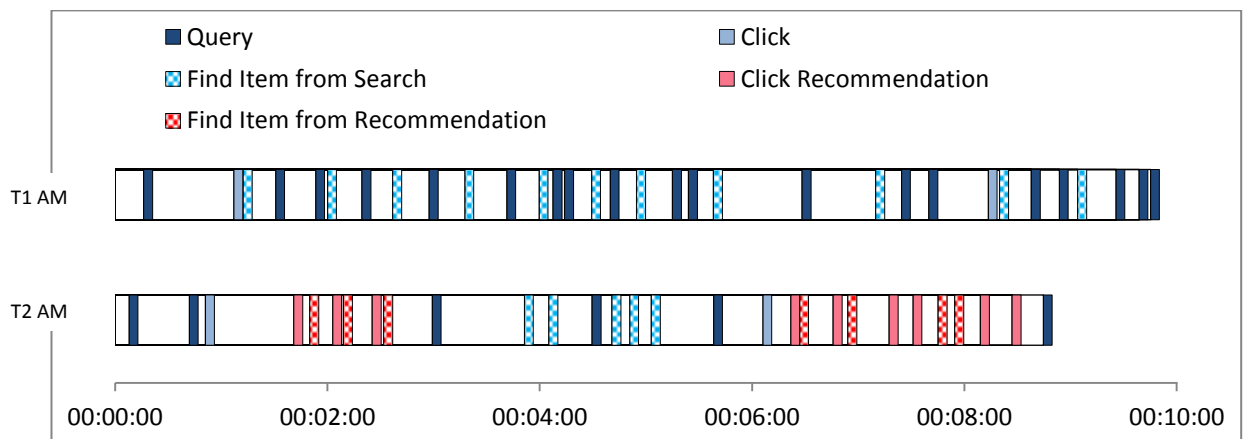
Figure 7-9: TSVs for Participant 9, T1 AM and T2 AM

7.8 Participant Perceptions

On completion of each task, participants were asked to rate the usefulness of various system features. Results are displayed in Table 7-11, and were similar for both tasks. On WorldCat.org the search box and search results display were considered the most useful features. These features also scored highly on Amazon, but recommendations were considered almost as useful for Task 1, and more useful than the search results display for

Task 2. Participants were also asked about their satisfaction with their performance, and the performance of the system, after each task. Participants were asked the extent to which they agreed or disagreed with four statements (see Table 7-12).

Table 7-12: Perceptions of feature usefulness - median scores (1 = not at all useful, 5 = very useful)

Feature	T1 WC	T1 AM	T2 WC	T2 AM
Advanced search	4.0	n/a	3.0	n/a
Format filter	3.0	1.5	3.0	1.0
Related subjects	3.0	2.5	2.5	3.0
Overview	4.0	4.0	3.0	4.0
Reviews	1.0	3.0	1.0	3.0
Search box	5.0	5.0	5.0	5.0
Search results display	5.0	5.0	4.0	5.0
Recommendations	n/a	4.0	n/a	5.0

Satisfaction with the number and quality of books found was relatively constant across the systems, with participants only slightly less satisfied with the quality of books found for Task 2 on WorldCat.org than on Amazon (see Table 7-13). The third statement ("I feel there are books in the system I couldn't find") was intended to measure any potential frustration with the systems, but results were constant for both iterations of both tasks (*Mdn*=4 "neither agree nor disagree"). We observe a significant difference between systems for the final statement ("The system made completing this task easy"). Here participants agreed more strongly with the statement for the Amazon iterations of both tasks (T1 WC *Mdn*=5, AM *Mdn*=6, *p*<0.05, T2 WC *Mdn*=4.5, AM *Mdn*=6, *p*<0.005).

Table 7-13: Perceptions of system and task performance - median scores (1= strongly disagree, 7 = strongly agree)

	T	WC	AM	*p*
I am satisfied with the number of books I found	T1	6	6	0.241
	T2	5	6	0.108
I am satisfied with the quality of books I found	T1	6	6	0.079
	T2	5	6	0.102
I feel there were books in the system I couldn't find	T1	4	4	0.204
	T2	4	4	0.283
The system made completing this task easy	T1	**5**	**6**	**0.032**
	T2	**4.5**	**6**	**.001**

No correlation was found between perceptions of the system's ease of use for a task, and the number of items found during that task. Further investigation did, however, reveal relationships between perceived ease of use of WorldCat.org and the number of actions per item found (calculated as number of actions / items found). For both T1 WC and T2 WC a moderate negative correlation was observed (Table 7-14).

Table 7-14: Correlation (Spearman's *rho*) between perceived ease of use and actions per item found

	Correlation Coefficient	*p*
T1 WC	**-.364***	**.029**
T1 AM	-.298	.077
T2 WC	**-.348***	**.038**
T2 AM	-.302	.074

After completing all tasks, participants were asked a final question regarding their system preferences (Q: "If you had to choose one of the two systems to find books on a particular topic, which would it be?"). The results showed a strong preference for Amazon over WorldCat.org (WC=25% (*n*=9), AM=75% (*n*=27)), with no significant difference between undergraduates and postgraduates.

7.9 Post-Study Interviews

Post-study interviews were conducted with all participants on completion of the four tasks and post-session questionnaire. The first two questions asked participants to describe their search strategies for completing the two tasks on the two systems, with a series of prompts and follow up questions designed to elicit perspectives on the features of the two systems that supported searching. The third questions asked participants for their preferred system, in terms of ease of use, while the fourth question repeated the question relating to serendipity that was developed for the Phase 1 focus groups. The final three questions focussed on recommendations, specifically whether participants would welcome Amazon style recommendations in the catalogue, whether they supported the idea of personalised recommendations, and what they felt makes a good book recommendation. Transcriptions of the interviews were subjected to Qualitative Content Analysis, and relevant sections of the code book are included in the following sections, which broadly follow the structure of

the interview. In reporting the results, incorrect pronunciations of WorldCat (e.g. WorldCat) have been corrected in quotations.

7.9.1   Search Strategy

A review of the transcripts of participant responses to the first two questions reveals that many had initial difficulties articulating the steps and decisions they took in the process of discovering relevant items, and in a number of cases the interviewer was required to ask a number of follow up questions to better understand the user's strategy. The result was a relatively rich set of descriptions of the tactics and strategies used to find items through both systems. The coding scheme to emerge from these descriptions is presented below.

Section of the code book relating to participant search strategy

1. Search Strategy
   a. Advanced Search
   b. Author
   c. Facets
   d. Keywords
      i. Query formulation issues
   e. Recommendations
      i. Chaining
      ii. Optimum starting point
   f. Related
      i. Known-item then related
      ii. Unknown-item then related
   g. Subject

Unsurprisingly, most participants described executing a keyword search as their first step for Task 1. A revealing slip of the tongue from one participant perhaps explained the prevalence of keyword searching:

> *"Firstly I literally just googled the topic, sorry not googled, searched using the search box for the topic."* (P29,)

Or as another student put it:

> *"I'm so used to google I just put the words in without thinking about it."* (P19, UG)

The process of finding useful books was driven by this activity for the vast majority of participants, particularly in WorldCat.org. As might be expected, participants generally described this keyword search being followed by an evaluation of the search results, with potentially useful items being viewed in more detail and assessed for relevance using descriptions, summaries, and previews. A small number or students described beginning with advanced searches, and the advanced search was also mentioned as a means of finding items by authors known to have produced relevant material, and a way of narrowing the search results:

> *"I ended up using advanced search for both systems. I found that the first few results weren't very useful when I did a keyword search, I had to scroll through quite a few. With the advanced search I could be really specific about what I wanted."* (P15)

> *I used the advance search to look for authors that I knew were reliable or important, ones that I'd heard mentioned by lecturers, I knew they were more reliable or important."* (P3)

Many participants described this issue of search results being too broad, or not seeming relevant. While several students spoke of using facets to limit the range of results, most participants described the need at some stage to formulate new queries:

> *"I started off with the topic itself and searched for that. That brought up lots of stuff that's not relevant so I had to go back to the search and add more detail to what I was searching for."* (P28)

Participants described their tactics for reformulating or augmenting query terms, which was also a requirement when participants felt they had exhausted the relevant results from an initial query. Several approaches were spoken of, including the addition of related words to the original query, and the use of synonyms. Some students also spoke of deliberately developing their queries according to the terminology and language found in useful items they had discovered:

> *"I used the titles and description of things that looked good to come up with new searches."* (P12)

> *"I looked at the blurbs of the books I'd found to try and get ideas for other searches to do."* (P22)

Issues of query reformulation appeared to be particularly common during to Task 2 *("I wasn't sure how to change my search keywords to get better results"* – P8; *"Because I don't*

really know much about history I just didn't know what keywords to put in" – P16), although in general participants described using similar strategies for both tasks.

Some participants described a slightly different approach to topic searching, using a known-item as a starting point:

> *"I looked for something I was aware of or familiar with and then used that as a point to go from, either looking for ones that were related, or by the same author."* (P1)

> *"I found it helpful to find a standard text book as a place to start."* (P33)

Most participants using this strategy felt that it was more effective on Amazon than on WorldCat principally because Amazon provided recommendations, which allowed for the easy exploration of related items. Indeed some students spoke of utilising different strategies on the two systems:

> *"I searched for a key book that I knew about, then on Amazon I looked at what other people bought or viewed. In Worldcat it was a bit more difficult; I found it a bit harder, just sort of searched for general words and looked through the main page of results."* (P22)

> *"I started with the core textbook, then in amazon used the 'customers also bought' to find related items. In WorldCat I used more keywords that I knew would get hits."* (P7)

> *"Generally the search is a starting point, in Amazon anyway you find something you know is good and from there you can follow the 'customers have also bought' links. You can't do that in WorldCat."* (P35)

A similar strategy was employed by other students, although in this case rather than specifically look for a known-item to anchor the search, participants identified a particularly relevant but previously unknown-item to use as a starting point:

> *"I did a broad search using keywords, then tried a few different search terms and saw what looked useful. Then when I found something that looked good I could try and find related books."* (P6)

> *"I tried performing searches on the search bar. If I couldn't find lots of things about the topic I would try and find one that seemed related, and find similar books to that. I liked the way amazon made that easy through what other people did."* (P21)

While several students described using WorldCat's "similar subjects" feature to identify related books (*"with WorldCat I quite quickly got academic books that seemed appropriate,*

228

then used the subject headings to work outwards" – P5), most participants felt that Amazon's recommendations greatly assisted the process of exploring the system:

> *"On amazon it was much easier cos you could see related books that people had bought or looked at or whatever, and it spiralled from there."* (P25)

There was strong sense from the interviews that recommendations provided an uncomplicated means of exploring the system. A key difference in behaviour across the two systems was the way Amazon was felt to provide options unavailable on WorldCat. Users described reaching "dead-ends" on WorldCat; occasions when they had exhausted easy options and were left pondering how to continue. Both systems offered easy access to the "low hanging fruit" of items returned by simple searches, but Amazon was felt to offer much better support for exploration beyond these easy pickings.

> *"You have to just scroll through search results, that's all you can do on WorldCat really. You can find a few things but then you reach these dead ends when you're on a page and you have nowhere to go."* – (P35)

> *"The recommendations just made things a lot easier. There was always something to click on."* (P5)

For Task 2, some participants described a strategy centred on the use of recommendations. The approach here seemed to be to deliberately identify a book of the greatest general relevance, since that would likely have the most appropriate recommendations, and then use those recommendations to find other titles:

> *"I thought the recommendations would be the easiest way of finding things on the same subject, so I found the most basic history book because I thought that would probably have the most useful recommendations."* (P24)

> *"I realised that I didn't know much about the subject, so finding things through what other people had bought was probably the best way to go. I tried to get to books that looked the most useful and look at the 'other people bought' for those."* (P6)

Not all participants however felt that using recommendations was an effective strategy, particularly for Task 1. Some postgraduates either doubted the usefulness of the

recommended items, or felt that their subject knowledge was such that they could navigate the system via search:

> *"I tried to use recommendations but found it quite difficult to know whether they were going to be useful or random. Basically I thought I could get better results by just searching. I know the subject pretty well so I was pretty confident that I could do the right searches to find things."* (P8)

> *"Personally given how much I know about my subject I didn't really think about using recommendations. I know the right searches to do and didn't really think what a bunch of people had bought on Amazon would be much use."* (P34)

For others less confident in their domain knowledge or search skills, the collaborative nature of the recommendations served as a means of reducing effort, or tapping into the more effective searching of other users:

> *"The customer viewed or bought is good because if people are looking for a similar thing and they've looked at other things it's easier than you trawling and doing all the searching if they've just done it."* (P8)

> *"On Amazon I would definitely use the 'who looked at this book' or authors. That can be really useful because if you miss a keyword it won't be brought up in the actual search but people will have bought things relevant to it, they will have done the right search."* (P26)

7.9.2   Preferred System

The third question in the interviews asked participants to state the system they found easiest to use, and the reasons for their preference. While the intention had not been to elicit detailed views on the perceived strengths and weaknesses of each system, these clearly emerged from the coding process. The full range of responses are best summarised by the relevant section of the code book which are included on the next page. In their direct answer to the question of which system they found easier to use, 27 participants (75%) stated unequivocally that it was Amazon, while 5 (14%) answered WorldCat.org. Two participants felt that the systems were equally easy to use, while another two though that it varies according to the search-task being undertaken. Of those offering this last opinion, the perception was that Amazon provided a more accessible experience for non-expert users:

"I found more books on amazon, but it's harder to filter out the stuff that's too populist. If you know a lot about the subject WC is easier. For everything else Amazon is better." (P5)

Sections of the code book relating to perceived strengths and weaknesses of the systems

2. Amazon or WorldCat
    a. Amazon
    b. WorldCat
    c. Equal
    d. Varies according to task
3. Positives
    a. Amazon
        i. Cover images
        ii. Descriptions
        iii. Hyperlinking
        iv. Look inside
        v. More relevant results
        vi. Recommendations
        vii. Related subjects
    b. WorldCat
        i. Academic focus
        ii. Author Search
        iii. Related Subjects
        iv. TOC
        v. User lists
        vi. Wide range of books
4. Negatives
    a. Amazon
        i. Inaccurate recommendations
        ii. Too commercial
    b. WorldCat
        i. Inaccurate search results
        ii. Lack of hyperlinks
        iii. Lack of item details
        iv. Lack of recommendations
        v. No cover images
        vi. Hard to explore
        vii. Too academic
        viii. Too many search results

"It's easier to find a wider selection of books on WorldCat, but Amazon gives a smaller list of more relevant things. So I guess it would depend on what you were using it for." (27)

The various strengths and weaknesses identified by participants as justification for their answers bears striking resemblance to many of the answers given by participants in the Phase One focus groups. Particular strengths of WorldCat were perceived to be its broad scope and academic focus, while features such as tables of content, user lists and related subjects link were also cited as helpful features. With regard to Amazon, the prevalence of cover images, detailed item descriptions, look inside feature and search ranking were praised. Some students once again found issues with WorldCat's search functionality, feeling that the search results were too broad or otherwise inaccurate. The principal issues with Amazon were perceived to be a lack of academic content, and the occasional presence of inaccurate recommendations.

7.9.3 Serendipity

The fourth interview question asked students to describe a time they'd found an item that was useful, but not what they'd originally been searching for. As with the Focus Groups, participants generally struggled to recall specific examples, but spoke more generally about the role of serendipitous discovery, and the environments in which they'd experienced it. The limited section of the code book developed from these answers is presented below.

Section of code book relating to serendipity

5. Serendipity
    a. Happens in Amazon
    b. Happens in another system
    c. Happens in physical library
    d. Happens with keyword searches
    e. Important

Several participants described how keyword searches of the Sheffield Library catalogue could sometimes yield surprising results, and lead to resources that offered unexpected

benefits. Other participants described experiences using Amazon when they felt they had discovered unexpected but useful items:

> *"A number of times I've just been cruising round Amazon and found some really weird but interesting stuff. I've got about six books on my Kindle that I'd never have bought deliberately, I just got them through bumping into them on Amazon."* (P28)

> *"It happens in Amazon a lot, you'll be looking for something but then think 'I'll see what other people got' and sometimes they are really interesting, like things you'd never have thought of getting otherwise."* (P4)

As well as Amazon, other systems were mentioned by individual users, including Web of Knowledge, Science Direct, and the Frankfurt University online catalogue. The last of these was described by a participant as featuring a "browse the shelf" function which they felt greatly aided unexpected discovery:

> *"You can browse the shelf based on the book numbers, like the dewey decimal numbers, with little pictures of the books. I'd use that quite a lot and it's the best system I've used for being able to find unusual things."* (P5)

As in the Focus Groups, perhaps the largest number of participants described serendipitous discovery happening most often in the physical library:

> *"It's much more likely to happen in the actual library by glancing at the shelf than in the library system."* (P26)

> *"Sometimes it happens in the library, you spot books on the next shelf or on the shelves behind you or something, those neighbouring books can be useful but not what you went in to get."* (P24)

In general participants felt that the ability to discover interesting but unexpected items was important. The exceptions tended to be undergraduates from scientific disciplines, and these students often struggled to see why such discovery would be necessary:

> *"I'm an engineer so it's hard to imagine a situation where I'd find something that completely surprised me or whatever. We pretty much know what we need to read."*

7.9.4 Recommendations in the Library Catalogue

The penultimate interview question asked participants whether they thought Amazon-style item-level recommendations would be useful in the library catalogue, and the code book for this section of the interview is presented below. In general participants felt strongly that such recommendations would be a useful addition to library catalogues. For the most part recommendations were perceived as a potentially useful way of finding resources,

---

Section of the code book relating to perceptions of recommendations in the library catalogue

6. Recommendations in the library catalogue
   a. Would be useful
      i. Alternatives to books on loan
      ii. Completeness
      iii. Discovering resources
      iv. Getting a balanced view
      v. New acquisitions
      vi. Novelty
      vii. Text books
      viii. Would save time
   b. Not useful
      i. Prefer to manage own search
      ii. Lead astray
      iii. Trust
   c. Type
      i. Other people viewed
      ii. Other people borrowed
   d. Presentation
      i. Needs covers
      ii. Explanation
      iii. Not too many

---

particularly novel items that might not otherwise be located through a search:

> *"They'd definitely be useful. Definitely. I think the most useful thing is that they could show you related books that you didn't know existed."* (P32)

> *"It's a good idea. I'm not really sure why they're not there already. I think the best case is that they'd help you find things you wouldn't find without the recommendation, I'm pretty sure that would happen."* (P15)

"Anything that helps us find things is helpful. It can be really hard to find varied things when you search, you just get the same kind of stuff. I'd use recommendations for sure." (P22)

"Yes I think they'd be good. It's a way of discovering things, and if they don't look useful or they're a bit weird like happens on Amazon sometimes, you can always ignore them." (P4)

Other participants felt that recommendations might be particularly useful in circumstances when a specific item required by the user was on loan. Here students felt that recommendations would be a way of identifying similar items:

"They'd be very useful, especially if the book you want is out. If they could offer other things then you'd know what else to get out that could help for the essay." (P10)

"Sometimes the book we want is out, so it would help us find alternatives. That would be its main use I think." (P31)

As well as assisting users in discovering books, participants also felt that recommendations could make discovery more efficient by allowing searchers to draw on the previous experience and expertise of other users:

"I suppose those kind of recommendations are actually kind of sharing information, because you're seeing things that other students have used. It could save time knowing what other people did." (P20)

"There can be so much stuff in the catalogue that it takes ages to find what you want and recommendations are maybe a way of quickly getting to relevant stuff." (P18)

While levels of enthusiasm for recommendations in the catalogue varied, only three participants felt that recommendations should not be added, and all were postgraduates. Two participants failed to see any benefits since they were confident in their ability to find what they needed using the search function, while another felt that they would struggle to trust recommendations based on other students' behaviour, since there was nothing to guarantee that the quality or level of student whose behaviour the recommendations were based on:

"I know a lot about my subject now but it's also something which undergraduates do, so if I got recommended things they'd probably be because loads of undergraduates had borrowed them which isn't much good for me."

"There's also a chance they could lead you astray. If you weren't 100% sure about a

235

subject you might follow a recommendation thinking it was going to be really good, but it might turn out to be terrible and you'd have wasted your time. I'd rather just make my own choices about what to get." (P9)

Some participants also spoke of how they thought recommendations should be presented. Including book covers was mentioned by several students, perhaps influenced by the presentation of recommendations in Amazon. The number of recommendations was also mentioned, with students generally feeling that a small number of relevant recommendations would be preferable:

> *"They need to be limited, not too many so they don't get in the way, and all relevant. Also it should explain why they are relevant, there must be an explanation of why they are being recommended." (P17)*

The last point mentioned by participant 17 was echoed by other students, who felt it was important that they understand why certain titles were being recommended. Few students were able to come up with suggestions as to how this could best be done beyond a broad explanation of how the recommender system itself worked.

7.9.5 Personalised recommendations in the Library catalogue

Students were also asked specifically whether they welcomed the idea of personalised recommendations in the catalogue. While several students saw some potential for this idea, particularly if the system was designed to recommend recently acquired item on topics deemed to be of interest to them. Others saw little use for them. Many students mentioned the fact their modules changed several times each year, and so recommendations based on their borrowing or viewing habits would soon be out of date – a severe example of the recommender system plasticity problem:

> "I'm not sure how it could work for academic literature. My courses change so often that I don't see how it could keep up." (P7)

> "The whole point of those recommendations is that they need to be absolutely perfect, otherwise I think what's the point? I'm changing subjects every few months so the recommendations would be out of date." (P36)

Several participants also felt that there was something inappropriate about the library collecting and using student records in this way. This was generally linked to issues of privacy, and a sense that the library should somehow be above the practices of commercial sites:

"To me it has a slightly sinister undertone. I don't really expect the library of all places to be collecting information about me." (P26)

"I think those recommendations are spam. I hate the Amazon ones, those emails they send me. The library really shouldn't be doing it." (P28)

7.9.6 Characteristics of a good recommendation

The final interview questions asked participants to reflect on what makes a good book recommendation, and the code book relating to participant responses to this question is found below. One point made by a large number of participants was that the items being

---

Section of the code book relating to what makes a good recommendation

8. What makes a good recommendation
   a. Characteristic of Recommendation
      i. Correct academic level
      ii. Description or summary provided
      iii. Popular
      iv. Recently acquired
      v. Recently published
      vi. Subject area
         1. Both broad and narrow subject areas
         2. Broad subject areas
         3. Narrow subject area
      vii. Transparent

---

recommended should be at an appropriate academic level. This was no doubt influenced by participants' recent experience with Amazon, but some responses indicated that this was likely to be an issue in an academic library system too:

"Also the academic rigour of the item is important. I know there are books in the Sheffield catalogue which are for undergraduates, and I don't really want to be recommended them. So the perfect recommendation would be something that is at exactly the right level for me." (P16)

"I think the level of the recommendation is important, like how difficult it is. I want stuff that's going to help me as a first year, not that I need to have a PhD to understand." (P14)

Another commonly mentioned factor was topic diversity, although opinion was divided as to what this should be. For some students, a good recommendation is one that is extremely

closely related to the topic of the item whose page it is recommended from. This was felt to be most appropriate in situations described elsewhere in this chapter, when the student is seeking a replacement for an item out on loan, or using an anchor item as a starting point for discovery. Other participants felt the opposite; that recommendations should be deliberately broad in subject scope, and serve as a way for users to establish connections between diverse subject areas. To further complicate matters a final group of participants described some combination of these views, suggesting that a good set of recommendations would encompass items both similar and diverse in topic. The following quotations illustrate the extent to which participants' views contradicted each another:

> "A recommendation should be something completely related to whatever I'm doing, on exactly the right subject. That's the most important thing I think, that it fits exactly with the subject I'm searching for." (32)

> "Good recommendations will be thought provoking. I don't want them to be pigeonholed to very specific subjects. I want recommendations to introduce me to cross-domain things outside my knowledge that I wouldn't find if I searched." (26)

> "I guess a good recommendation can be something really specific, like exactly what I'm trying to find but just a different book, but I suppose it could also be something really different too, if it's the right thing that could be the most useful type of recommendation." (4)

Other responses were pragmatic, with some suggesting that a good recommendation might make the user aware of new items added to the corpus, or that had been published relatively recently, while others stated that good recommendations were often of popular items. These arguments seemed to be driven by a desire not to miss out on items that others might have found useful. The final answers to this question dealt with presentational issues, with users believing that a good recommendation is one that is transparent, and that allows users to quickly judge the potential relevance of the item:

> "When you get a recommendation you need to know what it's about so you can tell whether it's useful or not. I'd also want to know why it had been recommended. Like if a friend recommended me a book to read I wouldn't just read it, I'd ask him why he thought it was good, why he was recommending it to me. So a good recommendation needs to have some explanation." (P24)

7.10 Discussion

As with previous chapters, while the bulk of discussion surrounding the results of the Phase 4 study takes place in Chapter 8, it is helpful to briefly summarise the key discussion points to emerge from a review of the user study results.

7.10.1 Performance

A key consideration in interpreting the results of the study is the extent to which differences in performance between the systems can be ascribed to the presence of recommendations in Amazon, as opposed to other difference in system functionality and interface. The exit interviews made clear that participants identified strengths and weaknesses of both systems beyond the presence of recommendations, including the thoroughness of the book description, the availability of the look inside feature, and the volume and ranking of search results. Whilst the study design limited the extent to which these other variables could be controlled for, the post-task questionnaires do indicate that recommendations were seen as more useful than any other feature barring search functionality. This strongly implies that the presence of recommendations was a critical factor influencing task performance and the perceived effectiveness of the systems.

In terms of the number of books found, there was a general trend towards slightly more books being found in Amazon than WorldCat.org, with the exception being the performance of PG students in Task 1. If the definition of performance is expanded to include the efficiency with which items are found, then the results again indicate that the presence of recommendations has a positive impact on performance. Participants generally required fewer system interactions to find items, and found items slightly more quickly when using Amazon (again with the exception of the PG T1 performance). An obvious explanation for these exceptions is the relatively high level of subject expertise held by PG participants. In comparison to undergraduates, PG students are likely to have more demanding relevance criteria, more narrowly defined topic areas, and be less likely to encounter novel items during their search.

Previous research has found a propensity for users to adapt to poor systems, and "make the best" of what functionality is on offer (Smith & Kantor, 2008). The results of this study support this view to a certain degree: whilst participants generally found fewer books on WorldCat.org, most were still able to find a significant number of relevant items. However, users are only likely to expend effort adapting to bad systems if they lack alternatives. In

the case of real-world topic searches this is manifestly not the case, since students have a range of other options available to them beyond the catalogue. If the presence of recommendations does indeed improve the efficiency of the information search process then this is likely to lessen the likelihood of users abandoning library catalogues in favour of Google.

7.10.2   Recommendations and perceived usability

The results of the post-task questionnaires show users found the tasks easier to complete on Amazon than on WorldCat.org. Clearly these judgements are potentially influenced by a myriad range of factors, some of which were identified in the post-study interviews; Amazon's superior search ranking, the availability of full descriptions, and the Look inside feature. However the study results also clearly show that recommendations were a key feature used by a large proportion of participants in completing the task, and offered an efficient means of discovering relevant material. The ease of use judgments are also shown to negatively correlate with the number of actions required to find items on the two system, and while this does not establish causality, it provides some evidence of a link between the presence of recommendations and a system being considered easy to use. Results of the post-task questionnaires also indicate that participants considered recommendations to be a particularly useful feature of Amazon. Given the effectiveness of Amazon in reducing the requirement to reformulate queries and view additional pages of search results, it is reasonable to conclude that the presence of recommendations is a factor in the perceived usability of the system.

7.10.3   Information Search Behaviour

Evaluating search behaviour is a complex proposition, especially given the levels of variance in individual searching strategies. Nonetheless we might identify two key ways in which the presence of recommendations affects information search behaviour. The transition probability diagrams clearly highlight the additional routes to discovery that recommendations afford, and the extent to which participants used them. A natural extension of this is to identify the stage within the session that recommendations are adopted. Analysis of the temporal distribution of recommendation clicks within the session tended towards the second half for Task 1, and the first half for Task 2. Relating this data to the transition diagrams, we might suggest that for a typical academic topic search (Task 1),

users were most likely to test the system's search functionality before choosing to view recommendations. This interpretation is supported by the frequent mention in exit interviews of frustration with "dead-ends" in WorldCat, which echo results previously found by Fast & Campbell (2004). It is likely that item level recommendations are utilized as a means of continuing exploration within the system at a point when a dead-end has been met. It is also interesting to note that the transition diagrams indicate that once users have engaged with a recommendation, they are likely to stay within a recommendation paradigm rather than return to a query strategy.

7.10.4 Influence of Domain Knowledge Recommendation Use

Analysis of the influence of domain knowledge on recommendation impact is founded on two assumptions: (1) for Task 1, we assume postgraduates to have greater domain knowledge than undergraduate students, and (2) we assume all participants to have less domain knowledge for Task 2 than for Task 1. A comparison of UG and PG performance for Task 1, and a comparison of the performance of all participants between Tasks 1 and 2. In both cases results indicate a negative correlation between level of domain knowledge and the impact of recommendations in terms of the number of recommendations viewed and the perceived relevance of the recommendations. The results make it clear that when undertaking a task with little or no domain knowledge, recommendations are an extremely effective means of aiding resource discovery. This is perhaps attributable to the difficulties users face constructing effective queries for topics with which they are not familiar. Recommendations offer a low-effort means of connecting users with potentially relevant items, and to some extent remove the barriers to discovery that search only systems can erect. For Task 2, once users had found one item relevant to the topic they were able to follow chains of recommendations, thereby exploring the collection without constant recourse to query reformulation.

The disparities between UG and PG use of recommendations for Task 1, whilst not as great as the disparity between overall Task 1 and 2 performance, were also notable. Two complementary explanations for the apparent reluctance of some PG participants to use recommendations for T1 are proposed. One is that their advanced domain knowledge allowed PG participants to formulate effective queries, meaning they had little need to turn to recommendations as a means of finding additional items. The other explanation rests on the frequent mention in the exit-interviews of trust as a key factor in assessing the value of

recommendations. The academic premise of Task 1 encouraged participants to think critically about the relevance of items being viewed. PG students, with their detailed knowledge of the subject, perhaps felt greater trust in their own abilities to find and evaluate items than that of a recommender system. These arguments do not necessarily suggest that recommendations offer no potential utility to domain experts. Instead it is perhaps an indication that recommender systems need to better adapt to expert users, both in terms of the diversity of items recommended and the level of trust they inspire.

8. DISCUSSION

8.1 Introduction

This chapter presents a discussion of the findings from each of the four phases of research, integrating results to reach a detailed understanding of WorldCat.org. As outlined in section 3.7, a number of methods of integration were employed, based on Bazeley & Kemp's metaphors for integrative analysis (2011). Thus complementary approaches were used to combine and enrich the data from each phase, and generative approaches to identify key strands emerging from the combined results.

The chapter is organised in three sections. First it addresses the findings relating to the users and use of WorldCat.org, before focussing on how the research has informed understanding of the potential role of recommendations within the system. Finally, a conceptual design for a WorldCat.org recommender system is proposed.

8.2 Understanding WorldCat.org

8.2.1   Classifying users

Fran Miksa, in her essay charting the evolution of notions of information users and use, concludes that "the idea of information users and use remains rather mysterious in its overall sense—rather like the images we see while driving in a fog" (2009: 362). For Miksa, notions of categorising users of information are inherently problematic, since both information and human behaviour are by their nature chaotic, complex and context driven. This is not to say of course that there is not value in attempting to understand the users of a system, since even incomplete or tangled information can be beneficial. Rather it reminds us that any picture to emerge from such analysis is likely to be fuzzy, and must be viewed as an approximation of a complex and dynamic whole rather than an exact representation – a blurred frame from a film, rather than the film itself.

Researchers have long noted the heterogeneity of library catalogue users, and the multitude of ways in which users can be classified (Borgman 1996, Connaway & Dickey 2010). Reflecting on the results from each phase of this study, it becomes apparent that the findings can be applied to multiple dimensions of classification. It should be noted here that this list is not intended to be exhaustive – clearly there are numerous other ways to frame user differences. Rather it summarises the

dimensions which emerge from the study as significant, and for which meaningful conclusions can be drawn. These dimensions are of course interrelated, and any complete understanding of the system's users must necessarily begin to address the nature and extent of these interrelationships. The dimensions differ too in whether they are categorical or spectral. There follows an explanation and discussion of each dimension which reflects on the conclusions that can be drawn from the data collected during this study.

Table 8-1: Dimensions for classifying WorldCat.org's user WorldCat.org user-base

Dimension	Description
Geography	Users can be classified by their geographical location
Occupation	Users can be classified by their occupation or profession
Referrer	Users can be classified by the how they arrived at WorldCat.org (e.g. search engine, library link).
Task	Users can be classified by the nature of the task they are seeking to complete on the system, or the information need they are seeking to address
Engagement	Users can be classified by the extent to which they choose to engage with the system.

8.2.1.1 Geography

Numerous studies have shown that cultural factors effect interactions with systems, including general search behaviour (Zoe & DiMartino, 2000), query re-formulation (Jesper et al. 2013), and information seeking behaviour (Ford et al. 2001). It is therefore instructive to review findings relating to the geographic distribution of WorldCat.org users. Table 8-2 compares the results of the two phases of research applicable to this question; the transaction log analysis and the pop-up survey.

Table 8-2: Geographical location of users - results from TLA and Survey

	Transaction Log Analysis		Pop-up survey	
	Country	% of total traffic	Country	% of total survey responses
1	United States	44.8%	United States	49.9%
2	China	5.3%	Canada	4.8%
3	Canada	5.2%	China	4.7%
4	United Kingdom	3.7%	Germany	3.7%
5	Germany	3.2%	United Kingdom	2.7%
6	France	2.3%	Australia	2.6%
7	India	1.8%	Brazil	2.0%
8	Italy	1.7%	India	1.9%
9	Indonesia	1.7%	Mexico	1.7%
10	Spain	1.5%	Italy	1.7%
11	Netherlands	1.5%	Netherlands	1.3%
12	Mexico	1.3%	France	1.0%
13	Australia	1.3%	Spain	0.8%
14	Brazil	1.3%	Belgium	0.7%
15	Poland	1.2%	Sweden	0.7%
16	Japan	0.9%	New Zealand	0.7%
17	Malaysia	0.9%	Russian Federation	0.7%
18	Korea, Republic of	0.7%	Switzerland	0.7%
19	Russian Federation	0.7%	South Africa	0.6%
20	Singapore	0.7%	Columbia	0.6%

Naturally the results of the log analysis represent the most robust measure of geographical distribution, since they are calculated from a very large and complete sample of traffic. It transpires that the geographical spread of survey respondents is very similar to that observed in the logs; 13 countries appear in both top-20 lists, and both lists show a large proportion of users coming from the US. That the spread of survey respondents appears so similar serves to partially validate the survey findings, at least to the extent that the

respondent population can be shown to generally represent the geographic distribution of the total user population.

As a whole, this study finds that WorldCat.org can justifiably be called a global service. 240 countries are represented in the log data, and while North American traffic accounts for a large percentage of traffic, the long-tail of other countries represent around half of all users coming to the site. It is interesting to note that several focus group participants identified the global scope of WorldCat's holdings as a key strength of the system, and revealed how they used the system to search for material held around the world. There is a sense then that the global reach of WorldCat is at least partly driven by the size of its database. The transaction log analysis also revealed that there were differences between countries in how users were likely to be referred to the system. The North American user population was found to represent over 80% of all traffic originating at the WorldCat.org homepage, while traffic from other countries was most likely to originate from search engine referrals. We also note that location is linked to occupation, with librarians representing a much greater proportion of users from North America than from other countries, where student users were found to constitute the majority of the user population.

8.2.1.2 Occupation

Another key dimension against which to assess the make-up of the Worldcat.org user population is by occupation. Occupation was a key distinguishing feature of the OCLC WorldCat personas that drove the selection of the focus group participant populations, and it is clear that certain occupation groups represent key classes of WorldCat.org user. The survey results indicate three primary user groups (librarians, students and academics), and it is encouraging to note that the focus groups included respondents from each of these groups. They also match the key user groups found in the small amount of literature available on WorldCat.org users (e.g. Calhoun 2009). The focus group data suggest that these different groups have distinct reasons for using the system, and offer different perspectives on the strengths and weaknesses of its content and functionality. Naturally occupation can also be seen to relate closely to the type of task being undertaken on the system, with the usage scenarios described by librarian focus group participants graphically illustrating how umbrella terms such a librarian hide a multitude of distinct usage types.

 A weakness of this project perhaps lies in the relative paucity of data relating to the use of the system by groups other than librarians, students and academics. While rare bookseller

were able to provide a unique view of the system during the focus groups, they are unlikely to be representative of the views and experiences of users from other professions, a group which constituted almost a quarter of survey respondents. It is also important to note that 13% of survey respondents were using the system for recreational purposes. In these cases, while occupation may still be a factor influencing the search strategies and information seeking behaviour of users, the types of task being undertaken are likely to be independent of profession.

8.2.1.3 Referrer

Users of the system can also be distinguished through the process by which they arrive at a Worldcat.org page. The analysis conducted on the WorldCat.org logs included the assignment of a referrer type to each session in the log, with results showing that almost half of all sessions originated from a search engine results page, with a further 14% coming from library pages. Sessions starting directly at the WorldCat.org homepage accounted for 5% of traffic, while another 5% of sessions were found to originate from citation services, GoodReads and Wikipedia. The log analysis also revealed differences in behaviour and levels of system interaction between sessions originating from different referrer types, most significantly in the way that users starting directly at the homepage generally spent longer on the system, and were much more likely to execute queries.

It is also interesting to consider results by referrer type in light of focus group comments relating to levels of awareness of WorldCat.org. A theme emerging from the focus groups was that OCLC could do more to publicise and promote the service. One could argue that commencing a session at the WorldCat.org homepage implies some prior knowledge of or interest in the WorldCat.org service. Thus monitoring sessions in this way might serve as a useful measure for OCLC to track public awareness of the system.

8.2.1.4 Task

Both the focus group study and the pop-up survey revealed that users utilise WorldCat.org for a variety of tasks, and to meet a range of information needs. The taxonomies of work- and search-tasks to emerge from this research are discussed in greater detail in section 8.2.2, but it can be noted here that a primary distinction in the latter taxonomy is between unknown-item and known-item search tasks. Analysis of sample sessions from the WorldCat.org logs suggested that around 60% of sessions including a query represent

known-item searches, and around 20% unknown item. This led to a conservative estimate that between 250,000 and 300,000 sessions per month might include unknown-item searches from within WorldCat.

These results can be compared to the data collected in the pop-up survey, which found that 73% of all respondents were engaged in some of known-item search task, and 46% in an unknown-item or informational task (noting that these options were not mutually exclusive). Thus it appears likely that the survey over-sampled users engaged in unknown-item searches. This suggests that the survey results are best interpreted as representative of a select sub-set of WorldCat.org users, most likely those familiar with and invested in the service.

It is interesting to compare the proportions of known- and unknown-item searches observed in the WorldCat.org logs with other research into search-tasks undertaken in library catalogues. Much of this work is qualitative, and therefore focuses on the development of task-type taxonomies rather than quantifying actual usage rates (e.g. Hert, 1996). In the only purely quantitative study available, Larsonn (1991) analysed transaction logs and found that around half of queries submitted to the Melvyl catalogue could be considered known-item. Slone (2000) includes counts of users engaged in different types of tasks in her qualitative study of public library catalogue use, noting that 20 of 35 participants were engaged in unknown-item tasks, 8 in known-item tasks, and the remainder in area searches. While these figures represent a much greater proportion of unknown-item searching than other studies, this is potentially a consequence of the methodology, which asked library patrons already engaged in searching at a library terminal to participate. We should also note the very low sample size, and the fact that this research is conducted in a public rather than academic library setting. Of the most direct relevance, a survey conducted by Goodale & Clough (2012) as part of their study into use of the SEARCH25 union catalogue found that 85% of users conducted known-item searches "often" or "very often", compared with 59% who regularly conducted topic searches. While this survey data does not directly translate to usage figures, it does suggest that users are frequently using the system for unknown-item searches. The figures derived for WorldCat.org in this study, then, seem to show that the proportion of unknown-item searches is much lower than those found in an academic or public library catalogue, or a regional union catalogue. Referring to the focus group study conducted as part of this project offers some explanation for this. As one participants said, "*I'd purposely use WC if*

I'd exhausted other major resources" (Nottingham Historian). Several participants described looking for resources on a topic first using their institutional catalogue, then a local or national union catalogue, before accessing WorldCat. It is reasonable to imagine that a large number of such unknown-item search-tasks are resolved at the institutional or local level, resulting in a lower number of such queries being executed in WorldCat.org.

8.2.1.5 Engagement

Perhaps the most striking finding from the transaction log analysis study was the number of sessions consisting of no further engagement with the system after arriving at the site. In total more than 6 million of the 15,799,727 sessions in the log involved no query (39.7%), and no viewing of additional pages. A final way for us to characterise different users of WorldCat.org is therefore related to the extent to which they engage with the site after arrival. The notion of characterising catalogue users this way is suggested in Cooper's log study of the University of California's online catalogue (2001). Cooper distinguishes between users who do not execute a query (but who may view other pages on the site) and those who do, terming the former "tourists" and the latter "real sessions". His resulting analysis identified 15% of all sessions as representing tourist traffic.[10] Jones et al (2000), while not using the term "tourist", report that 21.5% of visits to the New Zealand Digital Library did not include a query. Comparing these figures to the WorldCat.org data is problematic, since the WorldCat.org figure of 39.7% represents users who submitted no queries **and** clicked nowhere else on the site, while the figures from Cooper and Jones et al. do not include this second requirement. Since the figure from WorldCat is already almost twice that of either other study, there is no doubt that WorldCat.org sees a higher proportion of tourist traffic than either the University of California catalogue or the Digital Library of New Zealand.

The likely explanation for this difference lies in the fact that such a large proportion of WorldCat.org users are arriving from a search engine page: essentially they have already executed a query, and their visit to WorldCat.org, at least initially, represents the viewing of a result. While the studies by Jones et al. and Cooper can assume that a tourist's visit to the site does not result in the discovery of a resource, and is most likely evaluating the usefulness of the site, the same is not true within WorldCat.org. Whilst it is still helpful to

---

[10] Coopers paper reports this figure as 11%, but includes spider traffic in his calculation. The figure of 15% represents the percentage of tourist sessions from all human sessions.

distinguish between users who engage in further searches within the system and those who do not, WorldCat.org tourists may still be engaged in wider known- or unknown- item search tasks.

8.2.2   Work-Tasks and Search-Tasks

An understanding of the work- and search-tasks driving users to engage with WorldCat.org is obtained primarily through analysis of the data generated during the phase one focus groups. While the number of participants represented a relatively large sample of student and librarian users, it must be acknowledged that other populations with potentially relevant input were not investigated. Several participants described their use of the system for leisure purposes, allowing for the generation of a category of Leisure related work-tasks. Participants also included rare book sellers, who were able to describe their professional reasons for using the site, but it is clear that their needs are highly specialised, and unlikely to represent use cases for a host of other professions identified as users by the phase 3 survey. Thus the emergent work- and search-task taxonomies are necessarily incomplete; while they represent a robust representation of student and librarian needs, there is undoubted potential for expansion to include work-tasks specific to other parts of the user population.

Table 8-3: Work-Tasks Identified During Focus Groups, and Application to Non-Union Library Catalogues

Level 1	Level 2	Applicable to Institutional or Public Library
Academic	Essay / Assignment	Yes
	Research	Yes
Leisure	Hobbies	Yes
	Reading for Pleasure	Yes
Professional	Acquisitions / Collection Development	No
	Cataloging	No
	Inter-Library Loan	No
	Instruction / Training	n/a
	Reading-List Development	Yes
	Valuation	No

The work-tasks that emerged from the focus group study are presented in Table 8-3, and include tasks relating to areas as diverse as hobbies, cataloguing, and writing essays. It is instructive to assess which of these work-tasks might reasonably be undertaken to some

degree on a local institutional or public library catalogue. Of the ten work-tasks listed, one is by its nature specific to WorldCat.org (Instruction and Training). Of the remainder, five represent work tasks that could not be completed using the catalogue of a single institution. This suggests that WorldCat.org plays a vital role assisting users in tasks beyond the scope of their local catalogues.

Moving to the emergent taxonomy of search tasks, we can expand slightly on the hierarchy presented in Chapter 4. Since the transaction log analysis demonstrated a relatively high proportion of users utilising the "Cite/Export" function, the "Citation" task can be added to the sub-set of known-item searches (see Table 8-4).

Table 8-4: WorldCat.org Taxonomy of Search-Tasks

Level 1	Level 2	Level 3
Institutional Information	Location	
	Policies	
	Specialisations	
Known-item	Bibliographic Details	
	Editions	
	Format	
	Location	
	Holdings	
	Citation	
Unknown-item	Related	Author
		Manifestation
		Similar item
	Topic	Completeness
		Monitoring
		Multiple items
		Single item

There is very little literature against which to benchmark these findings. While Goodale & Clough's four use scenarios of the Search25 catalogue (2012) are all represented by this taxonomy, Slone's notion of an Area search (2000) is not included, since it is only applicable in circumstances when the user is searching a catalogue with the intention of determining

the location of an item within the physical library. In general the taxonomy provides a more detailed breakdown of the "Known-item" and "Discovery" purposes identified by Calhoun et al. (2009).

It is instructive to note that there is no simple mapping between the work-task and search task taxonomies. While some work-tasks imply a single specific search-task (for example Valuation → Holdings, Cataloguing → Bibliographic Details, ILL → Locations), in many cases the work-tasks might require any combination of several search-tasks (for example Essay / Assignment → [Single item, Multiple Items, Completeness, Author, Format, Location, Citation]). Similarly, many of the search-tasks might be undertaken as part of a number of work-tasks (for example [Reading for Pleasure, Hobbies, Essay / Assignment, Research, Reading-list Development] → Related Author). An obvious consequence of this is that WorldCat.org must be adaptable enough to meet the needs of users engaged in a variety of search tasks, which in turn form part of a broad range of work-tasks.

8.2.2.1.1   Known- and Unknown-Item Search Tasks

Section 2.4.3 discussed some of the philosophical complexities of the term known-item as identified by Lee et al (2007). While this study adopted a broad definition of the term, the descriptions by focus group participants of occasions when they engaged in known-item searching raise questions about the applicability of some other definitions. Some such definitions include the requirement that the searcher know either the author or title of a book (Lancaster, 1991), yet one participant described a search when the author wasn't known:

> "I knew of a book that I wanted, but couldn't remember the title or the author. I could remember the editor of the book series though, so with a bit of creative searching I could find it in WorldCat and remind myself of the title." (Northeastern Librarian).

It is difficult to justify describing this task as anything other than a known item. Similarly Dahlström & Gunnarsson's definition (2000), requires that the title and author are explicitly stated somewhere in the document, a fact that would perhaps not necessarily be the case for the librarian who described his searches for rare 18th Century surgeons' log books. It therefore appears justified to adopt our widened definition of a known-item search.

Some focus group participants alluded to the fact that a number of known-item searches might be conducted in close sequence, particularly librarians describing their work cataloguing a set of new acquisitions. Although not mentioned in the focus groups, the log analysis showing the frequency of use of the citation service on WorldCat.org would suggest that students may undertake similar sessions of repeated known-item searching. Since the log analysis also confirmed the prevalence of multiple known item-searching tasks with the sessions, it is reasonable to view this phenomenon as an exemplar of Toms's characterisation of search-tasks as discrete sub-tasks within a broader work-task (2011).

Results from both the log and user study support the idea that known-item searches are generally successfully executed. The known-item search that formed the first part of the second user task in the phase four study was completed by all users in an average of less than a minute, with a large majority of participants (88%) requiring a single query. At no stage was any feature other than search or advanced search used by any participant. Analysis of the transaction log sample sessions that were classified as known-item showed a mean of 1.48 queries per unique known-item search task, and 1.19 record page views per query, both suggesting users rarely encountered difficulties locating known-items in the system. This also supports contentions in the literature that this type of look-up task such is well suited to the query-response paradigm (Marchionini 2006), and that modern information retrieval systems are generally able to support such tasks very effectively. With this in mind it is interesting to note that a common theme of many focus group discussions was dissatisfaction with the WorldCat.org search results ranking, with some users specifically citing examples when they had struggled to locate a known-item. One explanation for this apparent contradiction (that known-item searches should be shown to be generally successful, but that users should make them a point of system complaint) lies in the extent to which users expect library services to match their experiences on the wider web (Connaway, 2007). Web searchers have become so accustomed to the hugely effective ranking of search engines that encounters with less optimal systems are perceived with disproportionate negativity (Fast & Campbell 2005).

Unknown-item searches represented the other form of search-task described by users. While WorldCat.org appeared to be considered by focus group participants as a particularly useful resource for identifying rare and obscure material, or ensuring the completeness of a topic search, several users also spoke of using it for typical unknown-item tasks, such as finding resources for an essay or assignment. Although log analysis suggests only a small

minority of total visitors to the site are engaged in unknown-item searches, that number is much higher when viewed as a proportion of sessions that interact with the system once arriving at the site.

It was a noticeable feature of the focus group comments on strengths and weaknesses of the system that little positive was said about the presence of features intended to aid unknown-item resource discovery. Many of the features of next-generation library catalogues are intended to aid unknown-item search tasks, including facets, links to related content, and tags (Kules et al., 2009; White & Roth, 2009). Findings from this study indicate that use of all these features on WorldCat.org is relatively low. Analysis of the sample log sessions indicate that facets were used in only 10% of unknown item searches, and even in those cases no more than two facets were used. The phase four user study revealed similarly scarce instances of interaction with the feature. While facets may offer support for exploratory search in some contexts, this study found no clear evidence to support the notion that their use significantly reduced query formulation, or positively affected resource identification, as claimed elsewhere in the literature (e.g. Kules & Shneiderman, 2008). Similarly the use of hyperlinks to related subjects and authors was used rarely, with observations from the user study suggesting that users frequently failed to notice the feature. The use of tags was found to be even lower, although this is perhaps as likely to be a function of the scarcity of tags across the WorldCat corpus as evidence against the utility of tags themselves.

While next-generation features are often cited as means of aiding resource discovery, in practice users facing difficulties completing an unknown-item task are likely to focus their efforts on the search function. The literature suggests that users struggling to find information will adopt more diverse queries, make greater use of advanced search functionality, and spend more time evaluating the search results (Aula, 2010). These strategies were regularly demonstrated by users participating in the phase four user study. While the eventual number of books found using these strategies was comparable with the quantity found during Amazon iterations of the tasks, analysis of the perceived usefulness of the various features, and overall ease of use, suggested that Amazon was much more positively viewed. The implication here is that in offering users an alternative to advanced search or refinement strategies, recommendations represent an effective strategy to cope with a challenging searching episode.

8.3 Recommendations and WorldCat

The results of this study validate findings elsewhere in the literature that users support the addition of recommendations to the library catalogue (Craven et al. 2010; Connaway, 2007). Data from the focus groups, pop-survey and user study all provide explicit support for the potential implementation of some form of recommendations. In considering the tasks that recommendations might support, it is instructive to review the taxonomy of search-tasks that emerged primarily from phase one of this project. Based on discussion of recommender functionality and use in the focus groups and phase 4 user study interview, and the existing capabilities and functionality of recommender systems technology, it is possible to determine whether recommendations are likely to effectively support each task (see Table 8-5). We note that no known-item search-tasks are likely to be aided by the presence of recommendations. Although it is clearly possible to imagine a scenario where a

Table 8-5: WorldCat.org Search-Tasks and Likely Recommender Utility

Level 1	Level 2	Level 3	Recommendations useful?
Institutional Information	Location		No
	Policies		No
	Specialisations		No
Known-item	Bibliographic Details		No
	Editions		No
	Format		No
	Location		No
	Holdings		No
	Citation		No
Unknown-item	Related	Author	No
		Manifestation	No
		Similar item	Yes
	Topic	Completeness	Yes
		Monitoring	Yes
		Multiple items	Yes
		Single item	Yes

user seeking a known-item clicks in error or confusion on an incorrect search result, and subsequently discovers the required item as a recommendation on the incorrect record page, such an event could hardly be seen to represent a core purpose of recommendations in the catalogue. Rather, recommendations are perceived by users as supporting unknown-item searches of almost all types.

8.3.1 Recommendations and Information Seeking in the Catalogue

The user study conducted as the fourth phase of this research project was intended to reveal the effect of recommendations on user performance and behaviour in the completion of typical unknown-item searches. We recall that the results showed a small difference in the number of items found between the system, with most users performing slightly better on Amazon, and little difference in the time taken for the participant to complete the task to their satisfaction. We did however observe that the Amazon facilitated the more efficient discovery of resources, requiring fewer system interactions to find each item, a measure that negatively correlated with perception of system ease of use. Comparisons of transition probability matrices and diagrams revealed that while tasks completed within WorldCat.org were located entirely within the query-response paradigm, the presence of recommendations in Amazon offered users an alternative mode of navigating the system, and that once interacting with recommendation lists users were likely to explore recommendations for some time.

It is helpful here to interpret the results in the context of the information-seeking and search theories discussed in section 2.4.1. We begin by examining the concept of browsing, and note that the catalyst for a browsing episode is an encounter with some form of visual cue (Bates, 2007). Recommendations, in the form of Amazon's row of book-cover thumbnails, serve as arresting stimuli, enticing the user to depart from the cycle of querying and instead . That recommendations are not always used immediately, or at all for some users, can be partially explained by the role of a range of contextual factors that influence the propensity of a user to engage in browsing (Marchionini & Schneiderman, 1988). Viewing recommendations as a means of efficiently browsing linked items in a collection also relates closely to aspects of Bates' berry-picking model (1989). Accepting that an information need is not static, but evolving in light of information-encounters, the type of browsing supported by recommendations meets Bates' demand for systems to offer functionality to support discovery beyond the query-response paradigm.

The results of the phase four study can also be understood in the context of Information Foraging Theory (Pirolli & Card 1999). A key concept for this theory is the notion of the *information patch*; the cluster of potentially useful resources represented at a macro level by a particular information system, and at the micro level by a set of search results (or by extension any other visible grouping of documents in a collection). Within the theory, the effort required to navigate between information patches is a key factor influencing the time spent evaluating and processing the information encountered within it. A key tenet of Information Foraging Theory is that a system should wherever possible seek to reduce the time and effort required to locate and access these patches. Applied to the user study, as the time and effort involved in reformulating queries to generate novel results increases, the attractiveness of recommendations as an efficient means of navigating to new patches also increases. The thumbnail images representing recommendation serve as information scent, "proximal cues" to new information patches available for exploration. Since each new record page visited has its own recommendations (and therefore constitutes its own patch), the chaining behaviour shown in the transition diagrams that represents users navigating from recommendation set to recommendation can be interpreted as users travelling the most efficient pathway between fertile patches. Since the user is constantly assessing the profitability (in terms of reward, time and effort) of exploring new patches, any decision to revert from recommendation viewing to querying the system (and vice versa) is likely to be driven by a perceived exhaustion of recommendations options, or the discovery of alternative search terms to use.

We can also recognise elements of O'Day & Jeffries' concept of information orienteering (1993) in the way in which participants in the study seek to locate a fruitful information space within the collection, although rather than using broad searches to identify the appropriate system context for the information need, users were observed to use recommendations. This combination of search and browsing behaviour might be said to epitomise the concept of an exploratory search. Exploratory search in this sense occurs when an information seeker is unable to achieve their goal through the traditional IR process model. The result is that the user employs "a combination of searching and browsing behaviour to navigate through (and to) information" (White & Roth, 2009: 10). This behaviour essentially represents a coping mechanism, with the user processing feedback offered by the system, both in terms of the content it offers access to and the search results it generates, to iteratively generate more effective queries. The process

described in the post-session interview by some participants, and observed in the aggregate representation of behaviour provided by the transition diagrams, whereby the browsing of recommendations led to a better understanding of the task and a subsequent ability to use more effective keywords, represents precisely this model of exploratory search.

The inclusion in the study of both postgraduate and undergraduate students, and the differing nature of the two tasks, allowed for some further analysis of the effect of domain knowledge on the interaction with and use of recommendations. The results indicated that recommendations proved a particularly effective means of aiding resource discovery in circumstances when domain knowledge was low. We also noted that for tasks where user domain knowledge was high, recommendation use tended towards the latter parts of the session, while for tasks where the user had little domain knowledge recommendations were utilised form the very beginning of the session. This again conforms to current understanding of the nature of exploratory search, whereby browsing activity is undertaken as a means of locating a context for relevant information content, with focused searching following once (White & Roth, 2009). The fact that a number of postgraduates made no use of recommendations at all is also consistent with existing understanding of the influence of domain knowledge on search behaviour, with experts more likely to trust their own domain knowledge to execute queries than novices (Kang & Fu, 2010). The suggestion that this confidence may be justified, and that postgraduate searchers were in fact better able to better find resources via more effective queries, is also supported by the literature (Wildemuth 2003). These findings also support research in the recommender systems literature which finds that domain knowledge influences the assessment of, and level of interaction with, recommendations (Knijnenburg et al., 2012). While studies in this area are generally focussed on personalised recommendations, the themes to emerge are clearly similar; recommendations are often used at the start of a task when domain knowledge is lowest (Castagnos et al., 2010), and experts are less likely to use recommendations (Hu & Pu, 2010).

8.3.2   User Preferences for a Recommender System

At a number of stages during the research project data was collected regarding users' preferences for the characteristics of a recommender system within WorldCat.org. One significant area that participants addressed in the focus groups and user study interviews

was the potential for personalised recommendation. In general most users were unenthusiastic, with many students expressing doubts that such a system could be of use given the rapidly changing information needs that accompany the progression to different modules each semester. In practice, this issue represents a form of a known recommender system problem, namely plasticity, which can be mitigated through the introduction of a temporal element to the recommendation ranking and the utilisation of a hybrid recommender algorithm that incorporates content-based elements (Burke, 2007). Other participants felt that they would not expect or value recommendations from a service like WorldCat.org, or that the recommendations would be somehow intrusive. These perspectives contrasted clearly with a general enthusiasm towards the idea of item-level recommendations. Amazon was frequently referenced as an example of how these recommendations might be implemented, echoing findings in the literature suggesting that Amazon stands as an exemplar for library catalogue design (Hartley & Booth 2006).

Another key aspect of recommendation interaction identified in the literature is trust. Both forms of trust considered relevant to recommender systems research – *system / impersonal trust* and *context-specific interpersonal trust* (Abdul-Rahman & Hailes, 1997) – emerged as key concerns for participants, both in the focus groups and user study interviews. With regards to the former, there was a clear concern from some participants that the system might somehow "lead astray" searchers by recommending items that were not relevant. This is consistent with research suggesting that users are naturally suspicious of automatically generated recommendations (Sinha & Swearingen, 2002). From the perspective of impersonal trust, some users also expressed fears that recommendations based on implicit user feedback (such as circulation records or session viewing data) might be unreliable due to the effect of other students' behaviour. The implication here was that less able or experienced students who select sub-optimal resources to view in the catalogue or loan would adversely affect the quality of recommendations. In both these cases, the extent to which participants worried about the consequences of poor quality recommendations was surprising. O'Donovan & Smyth (2005) argue that user perspectives on recommendations are driven to a significant extent by the potential cost of a bad recommendation, in which case we might assume that the cost associated with a poor quality set of item-level recommendations is minimal. This perhaps underestimates how keenly the loss of time spent reading an irrelevant item, or even the time spent walking to a library to borrow it, is felt. The preference of survey respondents for "recommendations

by experts" is potentially a manifestation of this effect, representing users' wishes for recommendations to be as dependable as possible.

The concept of serendipity was also discussed with participants in the focus groups and user study interviews, exploring both the capacity for serendipity in WorldCat.org, and the extent to which the system could encourage it further. As noted elsewhere in the literature, library catalogues tend not to support serendipitous discovery, since these systems are principally designed to support focussed search-tasks (McCay-Peet, 2011). It was unsurprising therefore that few participants in either phase recalled incidents of serendipity within the library catalogue, an exception being a student who described the chance discovery of a useful item via a virtual bookshelf feature. It is perhaps significant that the virtual bookshelf should be the focus of this recollection, since several participants described chance encounters with books on nearby shelves while visiting the physical library.

When asked explicitly to define a good recommendation in the context of a library catalogue, participants offered of variety of complimentary and sometimes contradictory answers. Many of the suggestions and requirements represent known good-practice in recommender systems design. Several users mentioned transparency, and it has been shown that recommendations perceived as transparent (i.e. the user understands why they have been recommended) are rated more highly (Sinha & Swearingen, 2002). The provision of details about the item being recommended is also a commonly identified factor in how the recommendation is received (Tintarev & Masthoff 2007, Drineas et al., 2002). Two additional themes to emerge from the study were temporal – that the system should recommend newly acquired or newly published material – while others related to the presentation ("description should be provided") or transparency of the system. Another preference – for recommendations to be at an appropriate academic level – may have been influenced by users' encounters with non-academic material while searching Amazon during the user-study, but also appeared to be applicable to an online catalogue environment, in the sense that students desired recommendations at an appropriate level for the stage of their degree.

The question of recommendation diversity proved most divisive, with a different participants holding clearly contradictory views on how closely related in topic recommended items should be. Views ranged from a preference for items to narrowly

focus on the topic of the original item, to a desire for recommendations that were broad in subject, with still other participants preferring a combination of both. We can note here that the recommender systems literature suggests that it is valuable to artificially promote diverse recommendation sets, both in order to facilitate serendipity, and because diverse recommendation sets are likely to appeal in some way to a diverse range of users (Knijnenburg et al., 2012).

It is important to recognise that the factors identified in this section are all applicable to library contexts beyond WorldCat.org. The tasks users were asked to complete during the user study represent typical usage cases for institutional systems, and participants were asked to think broadly about the way in which recommendations could be effectively incorporated in to catalogue systems.

8.4 Conceptual Design Specifications for a WorldCat.org Recommender System

As outlined in section 2.5.8.4, a number of libraries and researchers have introduced or experimented with recommendations in the catalogue. While these have generally met with positive feedback, there has been no formal evaluation of such active systems in the peer reviewed literature. As such it remains an open question whether these examples are best serving their users, or whether different approaches to the generation and presentation of recommendations would yield a more effective system.

This section outlines the conceptual design specifications for a recommender system for WorldCat.org. While the design processes found in the recommender systems literature commonly have at their foundation a technical analysis of the data sources available to generate recommendations, an assessment of algorithms that might be employed, and a review of the information architecture to be utilised, there is an increasing move toward understanding user needs and requirement as a first step in the development cycle (Ricci et al. 2011). The following specifications are therefore best thought of as conceptual, in that they seek to address the key themes relating to the potential WorldCat.org recommender emerging from this project. While suggestions are made as to potential sources of data with which to compute recommendations, and techniques for obtaining additional values by which to rank them, no comprehensive analysis of technical issues is undertaken.

1. Item-level recommendations should be the priority

While the creation of a personalised recommender system might offer some use to regular users of WorldCat.org, it seems likely based on the results of this investigation that there are greater benefits to be gained from focusing on item-level recommendations appearing at the record page. Although a number of participants at various stages of the process expressed doubts about the utility of personalised recommendations, the notion of item-level suggestions was welcomed by a large majority. The presence of these recommendations would offer users the avenues for exploration utilised by participants of the user study during their interactions with Amazon, and allow for an alternative to the query-response cycle.

2. A hybrid recommendation algorithm should be used

It seems clear from the review of recommender systems literature that some form of collaborative filtering algorithm will provide the most effective form of recommendations. CF systems have been shown to produce more diverse recommendations (Burke 2007), and offer the opportunity to leverage the vast user population of WorldCat.org as a source of implicit feedback. While the scope of this project did not extend to an evaluation of potential data sources, the review of research into and practical implementations of recommender systems in library catalogues reveal two principle methods of implicit feedback. Systems such a BibTip (Monnich & Spiering, 2005) have demonstrated the potential for record-page recommendations to be generated based on the co-occurrences of item views within sessions, while systems such as the SALT and Melvyl recommender projects (Rigby 2011 & 2012; Whitney 2006) and the University of Huddersfield's recommender system demonstrate the potential of using circulation records at a data source.

3. Recommendations should be at the work level

A number of focus group participants commented on the presence of multiple editions within sets of search results. Given the extent to which the sample log sessions revealed multiple views of different manifestations within single sessions, it seems likely that any recommender system based at least partly on session viewing co-occurrences would face the issue of multiple manifestations of the same work appearing as recommendations. Any recommender system must therefore operate at the work level, collating editions either at

the data preparation stage, or during the dynamic rendering of the recommendations themselves. The presence of the "View all editions" feature on WorldCat.org, and the functionality of the Kindred Works project, suggest the system has the capabilities to roll-up records to a work level.

4. User preferences for recommendation characteristics should be incorporated

Preferences for the characteristics of recommendations can be grouped in to three categories; 1) time-based (recommending newly published or acquired items); 2) topic-based (providing more or less topic diversity), and; 3) level-based (ensuring recommendations are at an appropriate academic level). Preference one appears to be relatively simple to implement, and indeed is a feature of some existing library catalogues. It can be argued that users specifically seeking current material can do so through the use of search-results sorting functionality, and results refinement. It appears therefore that the incorporation of preferences two and three offer the greatest utility to users of a recommender system. Both preferences however are problematic: since the system is non-personalised, there can be no simple way of determining the academic level of the user, while the results of this study show that opinions on appropriate topic diversity vary between users.

A potential solution lies in the implementation of an interactive system that allows the user to refine a recommendation set according to his or her preferences. Recommender systems involving the explicit real-time gathering of users' preferences have a relatively long history within Knowledge-based recommender systems research (see for example Resnick & Varian, 1997), and generally require the user to express their preferences using sliders or other interface features. The system proposed here might filter and re-rank recommendations according to the two defined characteristics. The Kindred Works recommender system demonstrates that levels of topical similarity can be calculated using subject heading and classification numbers, or indeed the degree of difference between items' Conspectus code (or other classification scheme). Attempting to estimate the academic level of a work is more complex, but might be guided by existing research into automatically estimating audience level conducted on the WorldCat corpus (O'Neil et al. 2008). Their method utilises the 'library type' classification assigned to all contributing libraries. This classification includes designations for Public, School, Academic, Research and National libraries. O'Neil et al.'s algorithm computes a score based on the proportion

of libraries of different types that hold a particular item. It is proposed here that a similar method might be used to determine academic level for recommendation purposes.

5. The interface should simulate the library shelf, with book-cover thumbnails.

The final aspect of the proposed design relates to the interface. The inspiration here is the idea of the virtual bookshelf, with an interface consisting of thumbnail book-covers arranged in a grid (see Figure 8-1). Grid set displays have been shown to encourage interactions with recommendations (Chen & Tsoi, 2011), and it is intended that the visual nature of the content links encourage the episodes of browsing and exploratory search behaviour. Rather than include sliders to determine the academic level or topic similarity, it is proposed that the recommendations form a matrix ordered by the two variables. Users would then be able to explore in the direction that represents their preferred recommendations (e.g. up for more closely related in topic, down for less closely related).
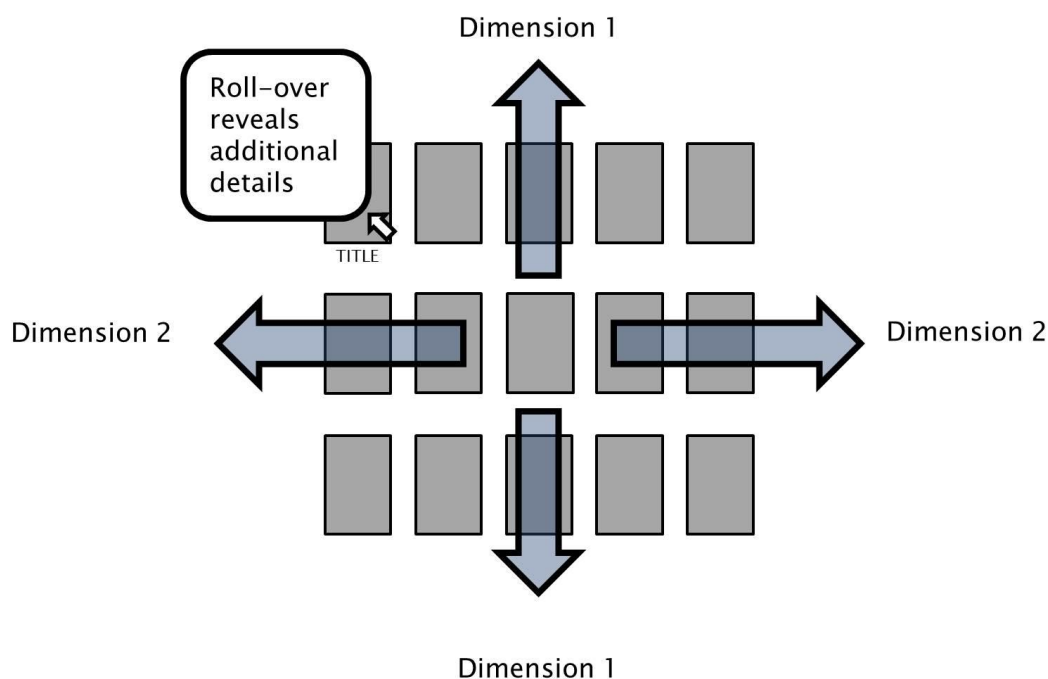
Figure 8-1: Conceptual drawing of proposed Recommender Systems Interface

This resulting system essentially represents a hybrid browsing tool (of a type epitomised by virtual bookshelf features), and recommender system. While sharing many features with

264

existing catalogue recommender systems, the novel aspect of this proposed design is in its incorporation of an interactive element intended to allow a user to explore recommendations according to their preferences for academic level and topic diversity.

9. CONCLUSION

9.1 Addressing the Research Questions

9.1.1   RQ1
Who is using WorldCat.org?

It is clear from the findings of the first three phases of this research project that WorldCat.org is used by a large and diverse user population. Coding of the phase two survey data resulted in the assignment of 20 unique codes to classify types of occupations found in the respondent group, with professions as diverse as gardeners, actors and accountants represented. It seems clear that the two largest single groups of users are librarians and students, with academics also constituting a significant proportion of the whole. Analysis of the log files during the phase 3 study also revealed the diversity of geographic locations from which users access the site. While the majority of traffic originates from North America, many thousands of sessions were found to originate from countries in all continents. Thus while the typical user might be a US librarian or student, it is clear that WorldCat.org must cater to a vast range of cultural and linguistic needs.

Integration of the four strands also revealed other ways of classifying the user population. Analysing the proportion of users originating from different types of referrer provides one way of characterising users. Search engine referrals account for almost half of all traffic, and sessions from these users are generally found to be short, with little or no further interaction with the system. Smaller proportions of sessions arrive from citation services, as well as Wikipedia and GoodReads. These users appear recognisably different in their system interactions to users who start their session at the WorldCat.org homepage, or who are referred by library services.

9.1.2   RQ2
For what purposes are users accessing WorldCat.org?

The phase 1 focus groups enabled the formulation of taxonomies of work- and search-tasks undertaken by the user groups under investigation. Qualitative content analysis of the focus groups transcripts resulted in the emergence of three categories of work-task; Academic, Leisure and Professional. Within each category a number of differing tasks were identified. Academic work-tasks relate to the production of an essay or assignment, and the

266

undertaking of research. Leisure work-tasks relate to the pursuit of a hobby or personal research, and reading for pleasure, while Professional work- relate predominantly to the activities of professional librarianship, including acquisitions and collection development, cataloguing, and instruction and training, and facilitating inter-library loan requests. It should be noted that this list of work-tasks is not intended to be exhaustive; although it provides a relatively robust representation of the motivations of librarian, student and academic users, it does not represent the many other professions who use WorldCat.org.

The taxonomy of search-tasks was also organised into three categories; the search for institutional information, known-item and unknown-item searches. These latter terms were defined for the purposes of this project as follows:

Known-item search: an interaction with the system wherein the searcher is seeking to locate in the catalogue the record of a specific item, about which some bibliographic data is known.

Unknown-item search: an interaction with the system where the searcher is seeking to locate in the catalogue one or more items that offer some potential utility, without knowing the specific items in advance.

The division of tasks into these two classes follows ideas with a long history in library science, and were able to inform the manual coding of sample sessions of WorlCat.org log data. These revealed that around 20% of users engaged in a search-task on the system were undertaking an unknown-item search, while over 60% were completing one or more known-item tasks. We therefore find that supporting known-item search tasks (ascertaining the bibliographic details of an item, determining or identifying manifestations of a work, finding alternative formats for an item, and identifying libraries that hold an item) represents a core required functionality of the system for a high proportion of users. Unknown-item searches (identifying related works by the same author, or on a similar topic, and discovering items on a particular topic), whilst representing a smaller proportion of the searches undertaken on the system, are nevertheless carried out by a large number of users. Many of these search-tasks, for example searching for completeness, or monitoring new publications on a topic, are tasks that WorldCat is uniquely placed to support.

9.1.3 RQ3

When might a recommender system support users of WorldCat.org?

Results from all four phases of the research project have served to inform the understanding of how recommendations can support WorldCat.org's users. An analysis of the search-tasks that emerged from phase 1 of this project, in conjunction with the findings of the phase 4 user study show that most unknown-item search-tasks would be supported by the inclusion of recommendations in the catalogue. Such tasks represent occasions when the user is likely to access a record page, and be open to the suggestion of relevant related content. Occasions when users are seeking multiple unknown-items, such as to identify resources for an essay or assignment, offer the greatest potential benefit for a recommender service. We also find that recommendations are shown to improve users' perception of system ease of use, and to be particularly useful as a means of exploring content when low levels of domain knowledge mean query formulation is problematic.

9.1.4 RQ4

What effect does the presence of recommendations have on information search behaviour in the library catalogue?

The phase four user study addressed this question in detail, finding that while the effect on task performance is limited, the presence of recommendations does radically alter some users' information search behaviour. The production of transition probability matrices reveals the likelihood of participants engaging in cycles of behaviour, most clearly undertaking a change from query-response cycles to the browsing of recommendation links. Relating results to important theories of information seeking and information search behaviour indicates that recommendations play an important role in supporting user activity beyond the query-response paradigm, particularly in terms of stimulating browsing episodes. In the context of Information Foraging Theory, recommendations can be seen as instances of information scent, offering links to new patches of information to be evaluated and consumed. Recommendations also appear to encourage exploratory search behaviours, and offer a low-effort alternative to methods of undertaking problematic searches, such as the use of advanced search.

9.1.5   RQ5

What recommendation characteristics would be most useful to users of library catalogues?

A number of findings emerged from the project to inform the question of optimal recommendation characteristics. The phase four user study essentially confirmed the effectiveness of Amazon item-level recommendations in supporting unknown-item search, and many of the characteristics of those recommendations should be included in any design of a system for a library catalogue. The post-user study interviews, combined with the phase one focus groups, provided rich qualitative data relating to participants' characterisation of a good recommendation. Three main notions of potentially useful recommendation characteristics emerged. It was suggested that recommendations of newly published or newly acquired material would be of particular use to some users, while others spoke of the need within a large corpus to access material at an appropriate academic level. Finally a variety of perspectives on the optimal level of topic diversity were presented, with users varying in the extent to which they preferred highly focused or topically diverse material, or mixtures of the two.

9.2 Limitations

A number of limitations to this project are acknowledged. As discussed elsewhere, the phase one focus groups, while undertaken with undoubtedly key user groups, do not represent a complete sample of all WorldCat.org users. As such the key findings can only be said to apply to the groups investigated. Similarly, the phase 2 survey has been shown to suffer from non-response bias, meaning that the results represent only the perspectives and behaviour of the responding subset of users. It is also recognised that the wording of some questions on the survey was potentially unclear, meaning that results relating to use of the system are problematic to interpret. While the phase three log study used a robust methodology, it is a weakness of this project that more was not done to better understand the traffic arriving at WorldCat.org from search engine referrals. It is also noted that the study included no analysis at the query level. While there is reasonable justification for investigating the logs at the session level, some query level analysis might have allowed for easier comparison with other studies. Finally the phase four user study, while utilising a

relatively large number of participants for an IIR study, still represents a comparatively small quantity of data, collected in laboratory conditions using simulated tasks.

9.3 Future Work

A number of areas for future research were identified in the course of this project. Query level analysis of the WorldCat.org logs would reveal more about the nature of users' search behaviour, and also help with analysis of the intents and search-tasks of users arriving from search engines and other referrers. Work might also be done to compute transition matrices of sequential actions, for comparison with phase four data and other similar log studies. Analysis of item co-occurrences would also determine whether this data could serve as implicit feedback for a recommender system.

Further qualitative research with other WorldCat.org user groups would serve to expand the existing taxonomy of work and search tasks. Additional work might also be done to evaluate a library catalogue recommender system in context.

Finally the development and evaluation of a prototype recommender system following the specifications presented in this study would provide great insight into the potential for recommender systems to enhance institutional library catalogues.

REFERENCES

Abdul-Rahman, A. & Hailes, S. (1997). A distributed trust model. Paper presented at the 1997 workshop on New security paradigms. 23-26 September, 1997, Langdale, UK.

Adamopoulos, P., & Tuzhilin, A. (2011). *On Unexpectedness in Recommender Systems: Or How to Expect the Unexpected*. Paper presented at the Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011). Chicago, IL, USA: ACM Press.

Adomavicious, G. & Tuzhilin, A. (2008).Context-Aware Recommender Systems. Paper presented at the 2008 ACM conference on Recommender systems.23-25 October 2008, Lausanne, Switzerland. New York: ACM.

Adomavicius, G. & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *Transactions on Knowledge and Data Engineering*, 17 (6), 734-749.

Adomavicius, G., & Kwon, Y. (2009). Toward more diverse recommendations: Item re-ranking methods for recommender systems. Paper presented at the 19th workshop on Information Technology and Systems (WITS'09). Phoenix, AR, USA.

Akeroyd, J. (1990). Information Seeking in Online Catalogues. *Journal of Documentation, 46(1),* 33-52.

Allen, B. (1989). Recall cues in known-item retrieval. *Journal of the American Society for Information Science, 40(*4), 246–252.

Alonso, O., Baeza-Yates, R. (2011). Design and Implementation of Relevance Assessments Using Crowdsourcing. *Advances in Information Retrieval Lecture Notes in Computer Science, Volume 6611*, 153-164.

Antell, K., & Huang, J. (2008). Subject searching success: transaction logs, patron perceptions, and implications for library instruction. *Reference & User Services Quarterly, 48*(1), 68-76.

Antelman, K., Lynema, E., & Pace, A. K. (2006). Toward a Twenty-First Century Library Catalog, *Information Technology and Libraries,* 25 (3), 128-139.

Arcolio, A. & Davidson, S. (2009). *WorldCat Local at the University of California: Usability Testings, Round Two, Fall 2009: Findings and Recommendations from the UC-OCLC Fall 2009 Assessment Project Summary Report*. California, US: University of California.

Aula, A., Khan, R. M., & Guan, Z. (2010). *How does search behavior change as search becomes more difficult?* Paper presented *at SIGCHI Conference on Human Factors in Computing Systems (CHI '10),* 35-44. New York, NY, USA: ACM Press. DOI=10.1145/1753326.1753333

Avancini, H., Candela, L., Straccia, U. (2007). Recommenders in a personalized, collaborative digital library environment. *Journal of Intelligent Information Systems*, 28 (3), 253-283.

Babbie, Earl (2001). *The practice of social research* (9th ed.). Belmont: Wadsworth.

Baez, M., Birukou, A., Casati, F., & Marchese, M. (2010).Addressing Information Overload in the Scientific Community. *Internet Computing*, 14 (6), 31 – 38.

Bailey, P., Craswell, N., Soboroff, I., Thomas, P., de Vries, A. P., & Yilmaz, E. (2008). Relevance Assessment: Are Judges Exchangeable and Does it Matter? Paper presented at SIGIR'08. Singapore.

Ballard, T., & Blaine, A. (2011). User search-limiting behavior in online catalogs: comparing classic catalog use to search behavior in next-generation catalogs. *New Library World 112*(5/6), 261-273.

Ballard, T., & Blaine, A. (2011). User search-limiting behavior in online catalogs: Comparing classic catalog use to search behavior in next-generation catalogs, *New Library World*, 112 (5), 261 – 273.

Bates, M. J. (2007). What is browsing—really? A model drawing from behavioural science research.*Information Research*, 12(4), pp. 330.

Bazeley, P., & Kemp, L. (2012). Mosaics, Triangles, and DNA: Metaphors for Integrated Analysis in Mixed Methods Research. *Journal of Mixed Methods Research, 6* (1), 55-72.

Beaulieu, M. & Borgman, C. (1996). A New Era for OPAC Research: Introduction to Special Topic Issue on Current Research in Online Public Access Systems, *Journal of the American Society for Information Science*, 47 (7), 491-492.

Belkin, N. J., Cool, D., Kelly, S. J. L., Park, S. Y., Perez-Carballo, J., & Sikora. C. (2001). Iterative exploration, design and evaluation of support for query reformulation in interactive information retrieval. *Information Processing and Management, 37*(3), 403-434. DOI=10.1016/S0306-4573(00)00055-8

Berendt, B., Mobasher, B., Nakagawa, M., & Spiliopoulou, M. (2002). The impact of site structure and user environment on session reconstruction in web usage analysis. Paper presented at WEBKDD 2002, Mining Web Data for Discovering Usage Patterns and Profiles, Vol, 2703, 159-179. Edmonton, Canada: Springer Berlin/Heudelberg.

Berendt, B., Mobasher, B., Spiliopoulou, M., & Wiltshire, J. (2001). Measuring the Accuracy of Sessionizers for Web Usage Analysis. Paper presented at the Workshop on Web Mining First SIAM International Conference on Data Mining, 7-14. Chicago, IL, USA.

Berg, B. L., & Lune, H. (2011). *Qualitative Research Methods for the Social Sciences.* England: Pearson; Harlow.

Berger, W. K. & Hines, R. W. (1984). What Does the User Really Want? The Library User Survey Project at Duke University, *Journal of Academic Librarianship*, 20 (5-6), 306-309.

Bertot, J. C., Berube, K., Devereaux, P., Dhakal, K., Powers, S., & Ray, J. (2012). Assessing the Usability of WorldCat Local: Findings and Considerations, *The Library Quarterly*, 82 (2), 207-221.

Biesta, G. (2010). Pragmatism and the Philosophical foundations of mixed methods research. *SAGE Handbook of Mixed Methods in Social & Behavioural Research*, 95-118. Thousand Oaks: SAGE.

Birks, M., & Mills, J. (2011). *Grounded theory: A practical guide*. London: Sage.

Björneborn, L. (2008). Serendipity dimensions and users' information behaviour in the physical library interface. *Information Research, 13*(4), 370.

Blecic, D., Bangalore, N. S., Dorsch, J. L., Henderson, C. L., Koenig, M. H., & Weller, A. C. (1998). Using

Bollen, D.,Knijnenburg, B. P., Willemsen, M. C., & Graus, M. (2010). *Understanding Choice Overload in Recommender Systems.* Paper presented at RecSys2010. Barcelona, Spain: ACM Press.

Bollen, J., & Van de Sompel, H. (2006).An Architecture for the Aggregation and Analysis of Scholarly Usage Data. Paper presented at the 6th ACM/IEEE-CS Joint Conference on Digital Libraries, 11-15 June, 2006, Chapel Hill, NC, USA.

Bonhard, P. &Sasse, M. A. (2006). 'Knowing me, knowing you' — using profiles and social networking to improve recommender systems. *BT Technology Journal*, 24 (3), 84-98.

Borgman, C. (1986). Why are Online Catalogs Hard to Use? Lessons Learned from Information Retrieval Studies, *Journal of the American Society for Information Science*, 37 (6), 387-400.

Borgman, C. (1996). Why Are Online Catalogs Still Hard to Use? *Journal of the American Society for Information Science*, 47 (7), 493-503.

Borlund, P. (2003). The Concept of Relevance in IR. *Journal of the American Society for Information Science and Technology, 54*(10). 913–925.

Bostandjiev, S., O'Donovan, J., & Höllerer, T. (2012). *TasteWeights: A Visual Interactive Hybrid Recommender System*. RecSys'12. Dublin, Ireland.

Brantley, P. (2008). Architectures for Collaboration: Roles and Expectations for Digital Libraries. *EDUCAUSE Review,* 43 (2).

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Re-search in Psychology, 3(2), 77-101.

Breeding, M. (2010). The State of the Art in Library Discovery, *Computers in Libraries*, 30 (1), 31–34.

Breeding, M. (2010). The State of the Art in Library Discovery. *Computers in Libraries, 30*(1), 31-34.

Broder, A. (2002). A Taxonomy of Web Search. *SIGIR Forum, 36*(2).

Bryman, A. (2008). Why do researchers integrate/combine/mesh/blend/mix/merge/fuse quantitative and qualitative research. In M. M. Bergman (Ed.), Advances in mixed methods research, 87-100. London, England: SAGE.

Buckland, M. (I992). *Redesigning library services: A manifesto*. Chicago: American Library Association.

Bulmer, M. (2004). *Questionnaires, V.1*. Thousand Oaks, CA: Sage Publications.

Burke, R. (1999). Knowledge-based recommender systems. *Encyclopaedia of Library and Information Systems*, 69 (32).

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331-370.

Burke, R. (2007) Hybrid Web Recommender Systems In: Brusilovsky, P., Kobsa, A., &Nejdl, W. (Eds.). *The Adaptive Web*, pp. 377-408. Berlin: Springer-Verlag.

Butters, G., Craven, J., & Johnson, F. (2009). *Usability Study of the Copac service.* Centre for Research in Library and Information Management.

Byrne, M. D., John, B.E., Wehrle, N. S., & Crow, D. C. (1999). *The Tangled Web We Wove: A Taskonomy of WWW Use*. Paper presented at the Human Factors in Computing Systems: CHI 99, 544–551. New York: Addison Wesley.

Byström, K., & Hansen, P. (2002). *Work tasks as unit for analysis in information seeking and retrieval studies*. Paper presented at the Fourth International Conference on Conceptions of Library and Information Science: Emerging Frameworks and Methods, 239-252. Seattle, WA, USA.

Bystrom, K., & Hansen, P. (2005). Conceptual frameworks for tasks in information studies. *Journal of the American Society for Information Science and Technology, 56*(10), 1050-1061.

Calhoun, K. (2006). *The Changing Nature of the Catalog and its Integration with Other Discovery Tools.* Library of Congress.

Callinan , J. E. (2005). Information-seeking behaviour of undergraduate biology students: a comparative analysis of first year and final year students in University College Dublin. *Library Review, 54*(2), 86-99.

Case, D. (2012). *Looking for Information: A Survey of Research on Information Seeking, Needs and Behavior (Library & Information.* Bingley, UK: Emerald Group Publishing Ltd.

Castagnos, S., & Pu, P. (2010). Consumer Decision Patterns Through Eye Gaze Analysis. Paper presented at EGIHMI'10. Hong Kong.

Catledge, L. & Pitkow, J. (1995). *Characterizing browsing strategies in the world-wide web*. Paper presented at the Third International World-Wide Web Conference on Technology, tools and applications, Vol 27. New York, NY, USA: Elsevier North-Holland, Inc.

Chan, L. M. (2007). *Cataloging and Classification: An Introduction* (3rd ed.). Lanham: Scarecrow Press.

Chen, C.C. & Chen, A. P. (2006).Using data mining technology to provide a recommendation service in the digital library. *The Electronic Library*, 25 (6), 711 – 724.

Chen, L. &Pu, P. (2007). Preference-based Organization Interfaces: Aiding User Critiques in Recommender Systems. *Lecture Notes in Computer Science*, 45 (11), 77-86.

Chen, L. &Pu, P. (2010). A User-Centric Evaluation Framework of Recommender Systems. Paper presented at the ACM RecSys 2010 Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces (UCERSTI).30 September, 2010, Barcelona, Spain.

Chen, L. S., Hsu, F.H., Chen, M.C., & Hsu, Y.C. (2008).Developing recommender systems with the consideration of product profitability for sellers. *Information Sciences*, 178 (4), 1032-1048.

Chen, L., & Pu, P. (2008). *A Cross-Cultural User Evaluation of Product Recommender Interfaces*. Paper presented at the 2008 ACM conference on Recommender Systems, 75-82. Lausanne, Switzerland: ACM Press. DOI=10.1145/1454008.1454022

Chen, L., & Pu, P. (2010). Eye-Tracking Study of User Behavior in Recommender Interfaces. *User Modeling, Adaptation, and Personalization*. *Lecture Notes in Computer Science*, (6075), 375-380

Chen, L., & Pu, P. (2011), *Users' Eye Gaze Pattern in Organization-based Recommender Interfaces*. Paper presented at IUI 2011. Palo Alto, California, USA.

Chen, L., & Pu, P. Experiments on user experiences with recommender interfaces. *Behaviour & Information Technology, 33*(4), 372-394.

Chen, L., & Tsoi, H. K. (2011). *Users' Decision Behavior in Recommender Interfaces: Impact of Layout Design.* RecSys'11 Workshop on Human Decision Making in Recommender Systems. Chicago, IL, USA.

Chen, Y. (2011). Interface and Interaction Design for Group and Social Recommender Systems, *RecSys'11, 197-204*. Chicago, IL, USA.

Cho, Y. H., Kim, J. K., Kim, S., H. (2002). A personalized recommender system based on web usage mining and decision tree induction. *Expert Systems with Application*, 23 (3), 329-342.

Choo, C.W., Detlor, B., & Turnbull, D. (2000). Information seeking on the Web: An integrated model of browsing and searching . *First Monday, 5*(2).

Cleverdon, C. W. (1991). The significance of the Cranfield tests on index languages. Paper presented at the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, 3–12. Chicago, IL, USA: ACM Press.

Clough, P., Sanderson, M., Tang, J., Gollins, T. & Warner, A. (2013). Examining the Limits of Crowdsourcing for Relevance Assessment, *IEEE Internet Computing, 17*(4), 32-38, doi:10.1109/MIC.2012.95

Cohen, L., & Holliday, M. (1996). *Practical Statistics for Students.* London: Paul Chapman.

Connaway , L. S., & Dickey, T. J. (2010). The Digital Information Seeker: Report of the Findings from

Connaway, L. S. (2007). Mountains, valleys, and pathways: Serials users' needs and steps to meet them. Part I: Identifying serials users' needs: Preliminary analysis of focus group and semi-structured interviews at colleges and universities, *Serials Librarian*, 52 (1/2), 223-236.

Connaway, L. S., & Dickey, T. J. (2010). *The Digital Information Seeker: Report of Findings from Selected OCLC, RIN and JISC User Behaviour Projects*. Dublin, Ohio: OCLC, Inc.

Connaway, L. S., Wilcox, D. J., & Searing, S.E. (1997). Online catalogs from the users' perspective: the use of focus group interviews. *College & Research Libraries, 58*(5), 403-420.

Cooley, R., Mobasher , B., & Srivastava, J. (1999). Data preparation for mining world wide web browsing patterns. *Journal of Know ledge and Information Systems, 1*(1).

Cosley, D., Lam, S. K., Albert, I., Konstan, J. A., & Riedl, J. (2003). *How Recommender Interfaces Affect Users' Opinions*. Paper presented at *CHI 2003.* Ft. Lauderdale, Florida, USA.

Cosley, D., Lam, S. K., Albert, I., Konstan, J. A., & Riedl, J. (2003). Is Seeing Believing? How Recommender Interfaces Affect Users' Opinions. Paper presented at the SIGCHI conference on Human factors in computing systems. 5-10 April, 2003, Ft. Lauderdale, USA. New York: ACM.

Cramer, H., & Evers, V. (2008). The effects of transparency on trust in and acceptance of a content based art recommender. *User Modeling and User-Adapted Interaction, 18*(5), 455-496.

Craven, J., Johnson, F., & Butters, G. (2010). *The usability and functionality of an online catalogue*. Paper presented at the Aslib Proceedings: New Information Perspectives, 62(1), 70 – 84.

Craven, J., Johnson, F., & Butters, G. (2010). The usability and functionality of an online catalogue, *Aslib Proceedings*, 62 (1), 70 – 84.

Creswell, J. W. (2010). Mapping the developing landscape of mixed methods research. *SAGE Handbook of Mixed Methods & Behavioural Research*, 46-69. Thousand Oaks: SAGE.

Cutter, C. A. (1876). *Rules for a printed dictionary catalogue.* Washington: Government Printing Office.

Dahlström, M., & Gunnarsson, M. (2000). Document architecture draws a circle: on document architecture and its relation to library and information science education and research. *Information Research 5* (2).

de Winter, J.C.F. & Dodou, D. (2010). "Five-Point Likert Items: t test versus Mann-Whitney-Wilcoxon". *Practical Assessment, Research & Evaluation*, *15* (11), 1-16.

Del Olmo, F. H., & Gaudioso, E. (2008). Evaluation of recommender systems: A new approach. *Expert Systems with Applications*, 35 (3), 790–804.

Demiriz, A. (2004), Enhancing Product Recommender Systems on Sparse Binary Data. *Data Mining and Knowledge Discovery,* 9(2), 147-170.

Dempsey, L. (1999). The library, the catalogue, the broker: Brokering access to informationin the hybrid library. *New Review of Information Networking, 5(1), 3-26.*

Denscombe, M. (2008). Communities of Practice: A Research Paradigm for the Mixed Methods Approach. Journal of Mixed Methods Research, 2(3), 270-283.

Denton, W. & Coysh, S.J. (2011). Usability testing of VuFind at an academic library, *Library Hi Tech*, 29 (2), 173-185.

Devendorf, L, O'Donovan, J., Höllerer, T. (2012). *TopicLens: An Interactive Recommender System based on Topical and Social Connections*. Paper presented at the 6th ACM Conference on Recommender Systems. RecSys 2012, Dublin, Ireland, Septmeber 13, 2012.

Dresang , E. T. (2005). Access: The Information-Seeking Behavior of Youth in the Digital Environment. *Library Trends, 54*(2), 178-196.

Drineas, P., Kerenidis, I., & Raghavan, P. (2002).Competitive Recommendation Systems. Paper presented at the 34th ACM Symposium on Theory of Computing. 19-21 May 2002, Montreal, Canada. New York: ACM.

Dwyer, C. M., Gossen, E. A., & Martin, L. M. (1991). Known-Item Search Failure in an OPAC. *RQ*, *31*(2), 228-236.

http://www.educause.edu/ero/article/architectures-collaboration-roles-and-expectations-digital-libraries

Emanuel, J. (2009). Next Generation Catalogs: What Do They Do and Why Should We Care? *Reference & User Services Quarterly*, 49 (2), 117-120.

Erdelez, S. (1999), Information Encountering: It's More Than Just Bumping into Information. *Bulletin of the American Society for Information Science and Technology, 25*(3), 26-29.

Fagan, J. C. (2010). Usability Studies of Faceted Browsing: A Literature Review, *Information Technology and Libraries*, 29 (2), 58-66.

Fagan, J. C., Mandernach, M. A., Nelson, C. S., Paulo, J. R., & Saunders, G. (2012). Usability Test Results for a Discovery Tool in an Academic Library, *Information Technology and Libraries*, 31 (1), 83-112.

Farzan, R. & Brusilovsky, P. (2011).Encouraging user participation in a course recommender system: An impact on user behaviour.*Computers in Human Behavior*, 27 (1), 276–284.

Fast, K. V., & Campbell, D. G. (2004). "I Still Like Google": university student perceptions of searching OPACs and the Web. Paper presented in the *American Society for Information Science and Technology 48*(1), 138-146.

Feilzer, M. Y. (2010), Doing Mixed Methods Research Pragmatically: Implications for the Rediscovery of Pragmatism as a Research Paradigm. *Journal of Mixed Methods Research,* 4(6), 6-16.

Fleder, D. M., & Hosanagar, K. (2007). *Recommender systems and their impact on sales diversity*. Paper presented at the 8th ACM conference on Electronic commerce, 192-199. San Diego, California, USA: ACM Press.

Ford, N., Miller, D., Moss, N. (2001). The role of individual differences in Internet searching: an empirical study. *Journal of the American Society for Information Science and Technology, 52*(12), 1049–1066.

Foster, A. (2004). A nonlinear model of information-seeking behavior. *Journal of the American Society for Information Science and Technology, 55* (3), 228–237.

Foster, A., & Ford, N. (2003). Serendipity and information seeking: an empirical study. *Journal of Documentation, 59*(3), 321-340.

Fry, A., & Rich, L. (2011). Usability Testing for e-Resource Discovery: How Students Find and Choose e-Resources Using Library Web Sites, *The Journal of Academic Librarianship,* 37 (5), 386–401.

Gbrich, C. (2007). *Qualitative Data Analysis: An Introduction* (1st Edn). London: Sage Publications.

Ge, M., Delgado-Battenfeld, C., & Jannach, D. (2010). *Beyond Accuracy: Evaluating Recommender Systems by Coverage and Serendipity*. Paper presented at ACM RecSys'10, 257-260. Barcelona, Spain: ACM Press.

Gillham, B. (2000). *The Research Interview.* London: Continuum.

Gisbergen, M. S. V., Most, J. V. D., & Aelen, P. (2007). *Visual attention to online search engine results.* Market Research Agency De Vos & Jansen.

Grady. C., & Lease, M. (2010). Crowdsourcing Document Relevance Assessment with Mechanical Turk. Paper presented at the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, 172–179. Los Angeles, CA, USA.

Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, CA: Jossey-Bass.

Greene, J. C., & Hall, J. N. (2010). Dialetics and Pragmatism: Being of consequence. *SAGE Handbook of Mixed Methods in Social & Behavioural Research*. Thousand Oaks: SAGE.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11,* 255–274.

Greene, J. C., Caracelli, V. J., & Graham, W.F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis, 11*(3), 255-274.

Gretarsson, B., Bostandjiev, S., Höllerer, T., O'Donovan, J, & Smyth, B. (2008). *PeerChooser: Visual Interactive Recommendation*. Paper presented at CHI 2008. Florence, Italy.

Gretzel, U. & Fesenmaier, D. (2005).Persuasion in Recommender Systems.*International Journal of Electronic Commerce*, 11 (2), 81-100.

Griffiths, J. R., & Brophy, P. (2005). Student searching behavior and the web: use of academic resources and Google. *Library Trends, 53*(4), 539-554.

Grose, M. W., & Line, M.B. (1968). On the Construction of White Elephants: Some Fundamental Questions Concerning the Catalogue. *Library Association Record, 70*(1), 2-5.

Group. http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

Hackman, J. R. (1969). Towards understanding the role of task in behavioural research. *Acta Psychological*, *31*, 97-128.

Harden, S. & Harden, R. (1994). *UK Public Libraries* (updated 25th March, 2008), available at http://ds.dial.pipex.com/town/square/ac940/ukpublib.html

Harrits, G. S. (2011). More Than Method?: A Discussion of Paradigm Differences Within Mixed Methods Research. Journal of Mixed Methods Research, 5(2) 150–166.

Hassenzahl, M. (2008). User experience (UX): towards an experiential perspective on product quality. Paper presented at the 20th International Conference of the Association Francophone d'Interaction Homme-Machine, 11-15. Metz, France.

Hayes, C., Massa, Avesani, P., & Cunningham, P. (2002).An on-line evaluation framework for recommender systems. In: *RPEC 2002*.

Workshop on *Recommendation* and *Personalization* in *E-Commerce. 28 May 2002, Malaga, Spain.* http://www.tara.tcd.ie/bitstream/2262/13178/1/TCD-CS-2002-19.pdf

Hayes, C., Massa, P., Avesani, P., & Cunningham, P. (2002). *An on-line evaluation framework for recommender systems*. Paper presented at the Workshop on Recommendation and Personalization in E-Commerce, 50–59. Málaga, Spain.

He, D., & Göker, A. (2000). Detecting session boundaries from Web user logs. Paper presented at the BCS-IRSG 22nd Annual Colloquium on Information Retrieval Research, 57-66. Cambridge, UK. DOI=10.1.1.27.1090

He, D., Göker, A., & Harper, D. J. (2002). Combining evidence for automatic Web session identification. *Information Processing & Management, 38*(5), 727–742.

He, Q., Pei, J., Kifer, D., Mitra, P., & Giles, C. L. (2010). Context-aware Citation Recommendation. Paper presented at the 19th international conference on the World Wide Web, 26-30 April, 2010, Raleigh, NC, US.

Head, A. J., & Eisenberg, M. B. (2009). *Lessons Learned: How College Students Seek Information in the Digital Age*. Seattle, US: The Information School, University of Washington.

Heinström, J. (2006). Psychological Factors Behind Incidental Information Acquisition. *Library and Information Science. Res. 28(*4), 579–594.

Herlocker, J. L., & Konstan, J. A. (2001).Content-Independent Task-Focused Recommendation.*IEEE Internet Computing*, 5 (6), 40-47.

Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000).Explaining Collaborative Filtering Recommendations. Paper presented at the 2000 ACM conference on Computer supported cooperative work, 2-6 December, 2000, Philadelphia, USA. New York: ACM.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. (2004).Evaluating Collaborative Filtering Recommender Systems.*ACM Transactions on Information Systems*, 22 (1), 5-53.

Herlocker, J.L., Konstan, J.A., Terveen, K., & Riedl, J.T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transaction on Information Systems, 22*(1), 5-53.

Hert, C. A. (1996). User Goals on an Online Public Access Catalog. *Journal of the American Society for Information Science, 47*(7), 504-518.

Hervas-Drane, A. (2007). Word of Mouth and Recommender Systems: A Theory of the Long Tail. Working paper 07-41, NET Institute, Universitat Pompeu Fabra, Barcelona, Spain

Hijikata, Y., Shimizu, T., and Nishida, S. (2009). *Discovery oriented collaborative filtering for improving user satisfaction*. Paper presented at the 14th international conference on Intelligent user interfaces, 67-76. Sanibel Island, Florida, USA: ACM Press.

Hildreth, C. R. (1995). *Online Catalog Design Models: Are We Moving in the Right Direction?* Washington, D.C.: Council on Library Resources.

Hu, R., & Pu, P. (2011). *Helping Users Perceive Recommendation Diversity*. Paper presented at the workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011/ACM RecSys'11). Chicago, IL, USA: ACM Press.

Huang, Z, Chung, W., Ong, T.H., & Chen, H. (2002). A Graph-based Recommender System for Digital Library. Paper presented at the 2nd ACM/IEEE-CS joint conference on Digital libraries, 14-18 July, 2002, Portland, OR, USA.

Husain, R. & Ansari, M. A. (2006). From Card Catalogue to Web OPACs, *DESIDOC Bulletin of Information Technology* , 26 (2), 41-47.

Iaquinta, L., Gemmis, M., Lops, P., & Semeraro, G. (2008). *Introducing serendipity in a content-based recommender system*. Eighth International Conference on Hybrid Intelligent Systems, 168-174. Barcelona, Spain: IEEE.

Iaquinta, L., Gemmis, M., Lops, P., Semeraro, G., & Molino, P. (2010). *Can a recommender system induce serendipitous encounters?* E-commerce, IN-TECH, 229–246. Vienna.

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation, 52*, 3-50.

Ivankova, N., Creswell, J., & Stick, S. (2006). Using mixed methods in sequential explanatory design: From theory to practice. *Field Methods*, *18*(3), 3-20.

Jamieson, S. (2004). "Likert scales: how to (ab)use them". *Medical Education, 38* (12), 1217-1218.

Jansen, B. J, Spink, A., Blakely, C., & Koshman, S (2007). Defining a Session on Web Search Engines. *Journal of the American Society for Information Science and Technology, 58*(6), 862–871.

Jansen, B. J. (2006). Search log analysis: What it is, what's been done, how to do it. *Library & Information Science Research 28*, 407–432.

Jansen, B. J., & Pooch, U.do Pooch (2001). A review of web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology, 52*(3), 235-246.

Jansen, B. J., & Rieh, S. Y. (2010). The Seventeen Theoretical Constructs of Information Searching and Information Retrieval. *Journal of the American Society for Information Science and Technology 61*(8), 1517–1534.

Jansen, B. J., Booth, D.L., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of Web queries. *Information Processing and Management, 44*. 1251–1266. doi:10.1016/j.ipm.2007.07.015

Jansen, B.J., & Spink, A. (2003). An analysis of Web information seeking and use: Documents retrieved versus documents viewed. Paper presented at the 4th International Conference on Internet Computing, 65–69. Bangalore, India.

Jarvelin, K. & Kekalainen, J. (2000). IR evaluation methods for retrieving highly relevant documents. Paper presented at SIGIR'00, 41-48. Athens, Greece: ACM Press.

Jesper, S., Clough, P., and Hall, M. (2013). Regional Effects on Query Reformulation Patterns. Paper presented at The International Conference on Theory and Practice of Digital Libraries (TPDL 2013), 382-385. Valletta, Malta.

JISC (2009). *JISC national e-books observatory project: Key findings and recommendations.* London, UK; JISC.

Johnson, F. C. & Craven, J. (2010). Beyond Usability: TheStudy of Functionality of the 2.0 Online Catalogue (OPAC), *New Review of Academic Librarianship*, 16 (2), 228-250.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher, 33*(7), 14-26.

Johnson, R. B., & Turner, L. A. (2003). Data collection strategies in mixed methods research. In A.Tashakkori, and C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research,* 297–319. Thousand Oaks, CA: Sage.

Jones, S., Cunningham, S. J., McNab, R., Boddie, S. (2000). A transaction log analysis of a digital library. *International Journal of Digital Libraries, 3*, 152–169.

Kan, M-Y., & Poo, D. C. C. (2005). *Detecting and Supporting Known Item Queries in Online Public Access Catalogs.* Paper presented at ACM/IEEE Joint Conference on Digital Libraries, JCDL. Denver, CO, USA: ACM Press.

Kang , R., & Fu, W. T. (2010). *Exploratory information search by domain experts and novices.* Paper presented at the 15th International Conference on Intelligent User Interfaces (IUI '10), 329-332. New York, NY, USA: ACM Press. DOI=10.1145/1719970.1720023

Kani-Zabihi, E., Ghinea, G., & Chen, S. Y. (2008). User perceptions of online public library catalogues. *International Journal of Information Management, 28*(6), 492–502.

Kapusta, J. (2012). Cut-off time calculation for user session identification by reference length. Paper presented at the 6th International Conference on the Application of Information and Communication Technologies (AICT), 1-6. Tblisi, Georgia: IEEE. DOI=10.1109/ICAICT.2012.6398500

Kellar, M., Watters, C., & Shepherd, M. (2007). A Field Study Characterizing Web-Based Information-Seeking Tasks. *Journal of the American Society for Information Science and Technology, 58*(7), 999–1018.

Kelly, D. (2009). Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends in Information Retrieval, 3*(1-2), 1-224.

Kelly, D., Harperb, D. J., & Landaua, B. (2008). Questionnaire mode effects in interactive information retrieval experiments. *Information Processing & Management, 44*, (1), 122–141.

Kilgour, F. G. (2001). Known-Item Online Searches Employed by Scholars Using Surname plus First, or Last, or First and Last Title Words. *Journal of the American Society for Information Science and Technology,* 52(14), 1203-1209.

Kitalong, K. S., Hoeppner, A., & Scharf, M. (2008). Making Sense of an Academic Library Web Site: Toward a More Usable Interface for University Researchers, *Journal of Web Librarianship*, 2 (2/3), 177–204.

Knijnenburg, B. P., Willemsen , M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User Model User Adap Interact, 22*(4-5), 441–504.

Kodakateri, A., Gauch, S., Luong, H., & Eno, J. (2009).Conceptual Recommender System for CiteSeerx.In: RecSys '09. Paper presented at the Third ACM conference on Recommender systems, 22-25 October, 2009, New York, US.

Kohlbacher, F. (2006) The Use of Qualitative Content Analysis in Case Study Research. *Forum: Qualitative Social Research, 7*(1), Art. 21.

Konstan, J. A., & Riedl, J. (2012). Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction, 22*, 101–123.

Konstan, J. A., McNee, S. M., Ziegler, C. M., Torres, R., Kapoor, N., & Riedl, J. T. (2006). *Lessons on applying automated recommender systems to information-seeking tasks*. Paper presented at *the 21st national conference on Artificial intelligence - Volume 2* (AAAI'06), Vol. 2, 1630-1633. Anaheim, SA, USA: AAAI Press.

Konstan, J. A., McNee, S. M., Ziegler, C. N., Torres, R., Kapoor, N., & Riedl, J. T. (2006). *Lessons on applying automated recommender systems to information-seeking tasks*. Paper presented at the *21st National Conference on Artificial intelligence* (AAAI'06), Vol. 2, 1630-1633. Boston, MA, USA: AAAI Press.

Konstan, J. A., McNee, S. M., Ziegler, C. N., Torres, R., Kapoor, N., & Riedl, J. T. (2006). Lessons on Applying Automated Recommender Systems to Information-Seeking Tasks. Paper presented at the 21st national conference on Artificial intelligence, 16-20 July, 2006, Boston, MA, USA.

Kruschwitz, U., Albakour, M. D., Niu, J., Leveling, J., Nanas, N., Kim, Y., Song, D., Fasli, M., & De Roeck, A. (2010).Moving towards Adaptive Search in Digital Libraries.*Lecture Notes in Computer Science*, 6699, 41-60.

Kuhlthau, C.C. (1991). Inside the search process: Information seeking from the user's perspective". *Journal of the American Society for Information Science, 42*(5), 361-371.

Kuhlthau, C.C. (1993). A principle of uncertainty for information seeking. *Journal of Documentation*, *49*(4), 339-355.

Kuhlthau, C.C. (2005). Kuhlthau's information search process. In: Fisher, K.E., Erdelez, S. & McKechnie, L.E.F. (eds.), *Theories of Information Behavior*, 230-238. Medford: Information Today.

Kules , B., Capra, R., Banta, M., & Sierra, T. (2009). What do exploratory searchers look at in a faceted search interface? Paper presented at the 9th ACM/IEEE-CS joint conference on Digital libraries (JCDL '09), 313-322. New York, USA: ACM Press. DOI=10.1145/1555400.1555452

Kules, B., & Shneiderman, B. (2008). Users can change their web search tactics: Design guidelines for categorized overviews. *Information Processing & Management, 44*(2), 463-484.

Kules, B., Capra, R., Banta, M., & Sierra, T. (2009). *What do exploratory searchers look at in a faceted search interface?* Paper presented at 9th ACM/IEEE-CS joint conference on Digital libraries (JCDL '09), 313-322.New York, NY, USA: ACM Press. DOI=10.1145/1555400.1555452

Lancaster F. W., & Joncich M. J. (1977). *The measurement and evaluation of library services*. Washington: Information Resources Press

Larson, R. (1991). The Decline of Subject Searching: Long-Term Trends and Patterns of Index Use in an Online Catalog. *Journal of the American Society for Information Science and Technology, 42*(3), 197-215.

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research Methods in Human-computer Interaction*. Chichester: John Wiley & Sons.

Leckiea, G. J., & Given, L. M. (2005). Understanding Information-Seeking: The Public Library Context. *Advances in Librarianship, 29*, 1-72.

Lecture Notes in Computer Science Volume 6075, 375-380.

Lee, J. H., Renear, A., & Smith, L. C. (2006). *Known-Item Search: Variations on a Concept*. Paper presented at the 69th ASIS&T Annual Meeting, 43, 1-17. Austin Texas.

Lee, T. Q., Park, Y., & Park, Y-T.(2008) A time-based approach to effective recommender systems using implicit feedback. *Expert Systems with Applications*, 34 (4), 3055–3062.

Lee, U., Liu, Z., & Cho, J. (2005). *Automatic Identification of User Goals in Web Search*. Paper presented at *World Wide Web Conference (WWW 2005).* Chiba, Japan.

Leino, J., & Räihä, K. J. (2007). *Case Amazon: Ratings and Reviews as Part of Recommendations*. Paper presented at RecSys'07: the ACM conference on Recommender Systems. New York, NY, USA: ACM Press.

Lekakos, G. & Giaglis, G. (2006).Improving the prediction accuracy of recommendation algorithms: Approaches anchored on human factors. *Interacting with Computers*, 18 (3), 410-431.

Lewandowski, D. (2010). Using Search Engine Technology to Improve Library Catalogs. *Advances in Librarianship, 32,* 35-54. doi: 10.1108/S0065-2830(2010)0000032005

Li, X. &Murata, T. (2010). Customizing Knowledge-based Recommender System by Tracking Analysis of User Behavior .Paper presented at the 17th International Conference on Industrial Engineering and Engineering Management (IE&EM).29-31 October, 2010, Xiamen, China.

Liao, S. C., Kao, K. F., Liao, I. E., Chen, H. L., & Huang, S. O. (2008). PORE: a personal ontology recommender system for digital libraries. *The Electronic Library*, 27 (3), 496-508.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Newbury Park: Sage.

Linden, G., Smith, B., York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing, 7*(1), 76–80.

Little, G. (2012). Where Are You Going, Where Have You Been? The Evolution of the Academic Library Web Site, *The Journal of Academic Librarianship*, 38 (2), 123–125.

Loeb, S., & Panagos, E. (2011). *Information Filtering and Personalization: Context, Serendipity and Group Profile Effects*. Paper presented at the 8th Annual IEEE Consumer Communications and Networking Conference - Emerging and Innovative Consumer Technologies and Applications, 393-398. Las Vegas, NV, USA. 10.1109/CCNC.2011.5766498

Log Analysis. *Journal of the American Society for Information Science and Technology, 56*(9), 913-934.

Lown, C., Sierra, T., & Boyer, J. (2013). How Users Search the Library from a Single Search Box. *College and Research Libraries*, preprint.

Lubetzky, S. (1986). "Principles of Descriptive Cataloging*"* in *Foundations of cataloging: A sourcebook*

Mai, J. E. (2010). Trusting tags, terms, and recommendations.*Information Research*, 15 (3).http://informationr.net/ir/15-3/colis7/colis705.html

Makri, S., Toms, E.G., McCay-Peet, L., & Blandford, A. (Eds.) (2011). Paper presented at the 1st International Workshop on Encouraging Serendipity in Interactive Systems: 13th IFIP TC13 Conference on Human-Computer Interaction – INTERACT 2011. Lisbon, Portugal: Springer Verlag.

Marcel Chiranov, (2011). Applying pop up survey software to incorporate users' feedback into public library computing service management. *Performance Measurement and Metrics, 12*(1), 50-65. http://dx.doi.org/10.1108/14678041111124298

Marchionini, G. (1989). Information-Seeking Strategies of Novices Using a Full-Text Electronic Encyclopedia. *Journal of the American Society for Information Science, 40*(1), 54-66.

Marchionini, G. (1995). *Information Seeking in Electronic Environments.* New York: Cambridge University Press.

Marchionini, G. (2006). Exploratory search: from finding to understanding. *Communications of the ACM – Supporting exploratory search,* 49(4), 41-46. DOI=10.1145/1121949.1121979

Marchionini, G. (2006). Toward human–computer information retrieval. *Bulletin of the American*

Marchionini, G. & White, R.W. (2008). Find What You Need, Understand What You Find. *Journal of Human-Computer Interaction*, *23*(3), 205–237.

Marchionini, G., & Shneiderman, B. (1988). Finding facts vs. browsing knowledge in hypertext systems. *IEEE Computer*, 21(1), 70-79.

Martin. J. (2008). The information seeking behavior of undergraduate education majors: does library instruction play a role? *Evidence Based Library and Information Practice, 3*(4), 4-17.

Massa, P. &Avesani, P. (2007) Trust-aware Recommender Systems. Paper presented at the 2007 ACM conference on Recommender systems, 19–20 October, 2007, Minneapolis, USA. New York: ACM.

Massa, P. &Bhattacharjee, B. (2004). Using Trust in Recommender Systems: An Experimental Analysis. *Lecture Notes in Computer Science*, 2995, 221-235.

Mat-Hassan, M., & Levene, M. (2005). Associating Search and Navigation Behavior Through

McBirnie, A. (2008). Seeking serendipity: the paradox of control. Paper from Aslib Proceedings: *New Information Perspectives, 60*(6), 600-618.

McCay-Peet, L, (2011). Exploring Serendipity's Precipitating Conditions. Paper presented at the Conference for Human-Computer Interaction - INTERACT 2011. Part IV (LNCS 6949), 398–401. Lisbon, Portugal: Springer Verlag.

McCay-Peet, L., & Toms, E. G. (2011). Measuring the dimensions of serendipity in digital environments. *Information research, 16*(3) 2011.

McGinty, L., & Smyth, B. (2003). *Explaining the user experience of recommender systems on the role of diversity in conversational recommender systems*. Paper presented at the Fifth International Conference on Case- Based Reasoning, 276–290. Trondheim, Norway.

McNee , S. M., Riedl, J., & Konstan, J. A. (2006). Being accurate is not enough: how accuracy metrics have hurt recommender systems. Paper presented at the conference for Human Factors in Computing Systems (CHI EA '06), 1097-1101. New York, USA: ACM Press.

McNee, S. M., Lam, S. K., Guetzlaff, C., Konstan, J. A., &Riedl, J. (2003). Confidence Displays and Training in Recommender Systems. Paper presented at the 9th IFIP TC13 International Conference on Human Computer Interaction.1-5 September, 2003, Zurich, Switzerland.IOS Press.

McNee, S. M., Lam, S. K., Konstan, J. A., & Riedl, J. (2003). *Interfaces for eliciting new user preferences in recommender systems.* Paper presented at *the 9th international conference on User modeling* (UM'03), Peter Brusilovsky, Albert Corbett, and Fiorella de Rosis (Eds.). Springer-Verlag, Berlin, Heidelberg, 178-187.

McNee, S. M., Riedl, J., & Konstan, J. A. (2006). *Being accurate is not enough: how accuracy metrics have hurt recommender systems.* Paper presented at the *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '06), 1097-1101. New York, NY, USA: ACM Press. DOI=10.1145/1125451.1125659

McNee, S. M., Riedl, J., & Konstan, J. A. (2006a). Being Accurate is Not Enough: How Accuracy Metrics have hurt Recommender Systems. Paper presented at the Conference on Human Factors in Computing Systems.24-27 April, *2006,*Quebec, Canada. New York: ACM.

McNee, S. M., Riedl, J., & Konstan, J. A. (2006b). Making Recommendations Better: An Analytic Model for Human-Recommender Interaction. Paper presented at the Conference on Human Factors in Computing Systems.24-27 April, *2006,*Quebec, Canada. New York: ACM.

McNee, S., Kapoor, N., &Konstan, J. A. (2006). Don't Look Stupid: Avoiding Pitfalls when Recommending Research Papers. Paper presented at the 2006 20th anniversary conference on Computer supported cooperative work, 4-8 November, Banff, Canada, pp. 171-180.

McNee, S., Riedl, J., & Konstan, J. (2006). *Making recommendations better: an analytic model for human-recommender interaction*. Paper presented at the 24th International Conference Human factors in computing systems, 1103–1108. Montréal, Canada.

McNee, S.M., Lam, S.K., Konstan, J.A., & Riedl, J. (2003). *Interfaces for eliciting new user preferences in recommender systems*. Paper presented at the 9th International Conference on User Modeling, 178–187. Johnstown, PA, USA.

McSherry, D. (2002). Diversity-Conscious Retrieval. *Advances in Case-Based Reasoning, Lecture Notes in Computer Science*, *2416*, 219-233.

McSherry, D. (2003). *Similarity and compromise*. Paper presented at *the 5th international conference on Case-based reasoning: Research and Development* (ICCBR'03), 291-305. Berlin, Heidelberg: Springer-Verlag.

McSherry, D. (2005). Explanation in Recommender Systems.*Artificial Intelligence Review*, 24 (2), 179–197.

Mendes, L. H., Quinonez-Skinner, J. and Skaggs, D. (2009), Subjecting the catalog to tagging, *Library Hi Tech*, 27 (1), pp.30 – 41.

Merčun, T., &Žumer, M. (2008). New generation of catalogues for the new generation of users: a comparison of six library catalogues. *Program: Electronic Library and Information Systems*, 42 (3), 243 – 261.

Mertens, D. M. (2012). What Comes First? *The Paradigm or the Approach? Journal of Mixed Methods Research, 6*(4) 255–257.

Miksa, F. (2009). Information organization and the mysterious information user. *Libraries & the Cultural Record, 44*(3), 343-370.

Millward L. J. (1995). Focus groups. In *Research Methods in Psychology* (Breakwell G.M., Hammond S. & Fife-Schaw C. eds), 274-292. London: Sage.

Mönnich, M. &Spiering, M. (2008).Adding Value to the Library Catalog by Implementing a Recommendation System.*D-Lib Magazine*, *14* (5/6).

Montgomery, A., & Faloutsos, C. (2001). Identifying web browsing trends and patterns. *IEEE Computer, 34*(7), 94–95.

Morgan, D. L. (1993), Qualitative content analysis: a guide to paths not taken. *Qualitative Health Research, 3*, 112–121.

Morgan, D. L. (2007). Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods. *Journal of Mixed Methods Research, 1*(1), 48-76.

Morrison, J.B., Pirolli, P., & Card, S.K. (2001). *A taxonomic analysis of what World Wide Web activities significantly impact people's decisions and actions.* Paper presented at the ACM Conference on Human Factors in Computing Systems, 163–164. Seattle, WA: ACM Press.

Neuman, W. L. (2011). *Social Research Methods: Qualitative and Quantitative Approaches* (7th ed., international ed.). Boston: Pearson.

Nguyen, T. T., Kluver, D., Wang, T. Y., Hui, P. M., Ekstrand, M. D., Willemsen, M. C., & Riedl, J. (2013).

Nicholas, D., Clark, D., Jamali, H. R., & Watkinson, A. (2014). Log Usage Analysis: What it Discloses about Use, Information Seeking and Trustworthiness. *International Journal of Knowledge Content Development & Technology, 4*(1), 23-37.

Nielsen, J. (2000), Why You Only Need to Test with 5 Users. Nielson: Norman

Norris, D. M. (1939). *A history of cataloguing and cataloguing methods 1100-1850: with an introductory survey of ancient times.* London: Grafton &Co.

O'Day, V., & Jeffries, R. (1993). *Orienteering in an information landscape: how information seekers get from here to there*. Paper presented at the ACM SIGCHI Conference on Human Factors in Computing Systems, 438–445. New York, USA: ACM Press. doi:10.1145/169059.169365

O'Donovan, J. & Smyth, B. (2005). Trust in Recommender Systems. In: *IUI '05*. Proceeding s of the 9th International Conference on Intelligent User Interfaces. 19-12 January 2005, San Diego, USA. New York: ACM.

Ochi, P., Rao, S., Takayama, L., &Nass, C. (2010). Predictors of user perceptions of web recommender systems: How the basis for generating experience and search product recommendations affects user responses. *International Journal of Human-Computer Studies*, 68 (8), 472-482.

Ockuly, J. (2003). What Clicks? An Interim Report on Audience Research. Paper presented at the Museums and the Web Annual Conference. Charlotte, NC: Archives & Museum Informatics.

OCLC (2006). *College Students' Perceptions of Libraries and Information Resources.* Dublin, Ohio: OCLC.

Oku, K., & Hattori, F. (2011). *Fusion-based Recommender System for Improving Serendipity*. Paper presented at the Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011). Chicago, IL, USA: ACM Press.

Olson, T.A. (2007). Utility of a faceted catalog for scholarly research. *Library Hi Tech, 25*(4), 550-561.

Onwuegbuzie, A. J., & Leech, N. L. (2005). On Becoming a Pragmatic Researcher: The Importance of Combining Quantitative and Qualitative Research Methodologies. *International Journal of Social Research Methodology, 8(5), 375-387.*

Ozok, A. A., Fan, Q., & Norcio, A. F. (2010). Design guidelines for effective recommender system interfaces based on a usability criteria conceptual model: results from a college student population. *Behavioural Information Technology, 29*(1), 57-83. DOI=10.1080/01449290903004012

Ozok, A.A., Fan, Q., & Norcio, A.F. (2010). Design guidelines for effective recommender system interfaces based on a usability criteria conceptual model: results from a college student population. *Behavioural Information Technology, 29*, 57–83.

Parsons, J., Ralph, P., &Gallager, K. (2004). Using Viewing Time to Infer User Preference in Recommender Systems. Paper presented at the AAAI Workshop on Semantic Web Personalization.25-26 July, 2004, San Jose, USA.

Pathak, B., Garfinkel, R., Gopal, R., Venkatesan, R., & Yin, F. (2010). Empirical Analysis of the Impact of Recommender Systems on Sales. *Journal of Management Information Systems 27*(2), 159-188.

Pattern, D. (2005), Using circ_tran to show borrowing suggestions in HIP, blog, 17 November, available at: http://www.daveyp.com/blog/archives/49

Pattern, D. (2009), The impact of book suggestions / recommendations? blog, 29 March, available at: http://www.daveyp.com/blog/archives/729

Peirce, C.S. (1904). "How to make our ideas clear". In: Hartshorne, C. & Weiss, P. (eds.), *The Collected Papers of Charles Sanders Peirce. Volume 5: Pragmatism and Pragmaticism*. Cambridge: Belknap Press of Harvard University Press.

Perugini, S., Goncalves, M. A., Fox, E. A. (2004). Recommender Systems Research: A Connection-Centric Survey. *Journal of Intelligent Information Systems*, 23 (2), 107–143.

Peters, T. A. (1993). The history and development of transaction log analysis. *Library Hi Tech, 11*(2), 41-66.

Price, B. (2005). *Optimal Recommendation Sets: Covering Uncertainty over User Preferences.* Paper presented at the Twentieth National Conference on Artificial Intelligence.

Philipp, M. (2000). Qualitative content analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research , 1*(2), Art. 20.

Phippen, A., Sheppard, L., & Furnell, S. (2004). A practical evaluation of Web analytics. *Internet Research: Electronic Networking Applications and Policy, 14*, 284–293.

Pickens, J., Golovchinsky, G., Shah, C., Qvarfordt, P., & Back, M. (2008). Algorithmic mediation for collaborative exploratory search. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR '08*, 315. http://doi.org/10.1145/1390334.1390389

Pidgeon, N., & Henwood, K. (1997). *Using grounded theory in psychological research*. In N. Hayes (Ed.), Doing qualitative analysis in psychology, 245-273. Hove, UK: Psychology Press.

Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*. Oxford: Oxford University Press.

Pohl, S., Radlinski, F., & Joachims, T. (2007). Recommending Related Papers Based on Digital Library Access Records. Paper presented at the 7th ACM/IEEE-CS joint conference on Digital libraries, 17-22 June, 2007, Vancouver, Canada.

Pommeranz, A., Broekens, J., Wiggers, P., Brinkman, W. P., & Jonker, C. M.(2012). Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process . *User Modeling and User-Adapted Interaction, 22*, 357–397.

Poulter, A. (2003). Browsing the Virtual Library. *Encyclopaedia of Library and Information Science, 1*, 390-396. DOI: 10.1081/E-ELIS 120008865

Prekopcsák, Z. (2007). *Content organization and discovery: state-of-the-art and new ideas for P2P-Fusion*. Budapest: Budapest University of Technology and Economics.

Price, B., & Messinger, P. R., (2005). *Optimal recommendation sets: covering uncertainty over user preferences*. Paper presented at the 20th National Conference on Artificial Intelligence (AAAI'05), 541–548.

Pu, P., & Chen, L (2007). Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems, 20*, 542–556.

Pu, P., Chen, L., & Hu, R. (2011). *A user-centric evaluation framework for recommender systems*. Paper presented at the fifth ACM conference on Recommender systems (RecSys '11), 157-164. New York, USA: ACM Press.

Rashid, A. M., Karypis, G., & Riedl, J. (2008). Learning Preferences of New Users in Recommender Systems: An Information Theoretic Approach *SIGKDD Explorations*, 10 (2), 90-100.

Rating Support Interfaces to Improve User Experience and Recommender Accuracy. Paper presented at the RecSys'13. Hong Kong, China.

Renda, M. E., & Straccia, U. (2005). A Personalized Collaborative Digital Library Environment: a model and an application. *Information Processing and Management*, 41, 5–21.

Resnick, P. & Varian, H. R. (1997).Recommender Systems. *Communications of the ACM*, 40 (3), 56-58.

Ricci, F., Rokach, L., Shapira, B. (2011).Introduction to the Recommender Systems Handbook.In: Ricci, F., Rokach, L., Shapira, B., & Kantor, P.B. (Eds.), *Recommender Systems Handbook*,pp. 1-29. New York: Springer.

Rice, R.E., McCreadie, M., & Chang, S.L. (2001). *Accessing and browsing information and communication*. Cambridge, MA: MIT Press.

Roberts, R. M. (1989). *Serendipity: Accidental Discoveries in Science*. New York, NY, USA: John Wiley & Sons, Inc.

Rogers, R., & Preston, H. (2009). Usability Analysis for Redesign of a Caribbean Academic Library Web site: A Case Study, *OCLC Systems & Services*, 25 (3), 200–211.

Rose, D. E., & Levinson, D.(2004). *Understanding User Goals in Web Search*. Paper presented at the 13th International World Wide Web Conference (*WWW 2004), 13-19*. New York, USA:ACM Press.

Rosie Jones, R., & Klinkner, K. L. (2008). Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs. Paper presented at CIKM'08, 699-708. Napa Valley, California, USA: ACM Press.

Rowlands, I., Nicholas, D., Williams, P., Huntington, P., & Fieldhouse, M. (2008). The Google generation: the information behaviour of the researcher of the future. *New Information Perspectives 60*(4), 290-310.

Rozanski, H.D., Bollman, G., & Lipman, M. (2001). Seize the occasion! The seven-segment system for online marketing. *Strategy in Business, 24*. Booz Allen & Hamilton.

Ruthven, I. (2008). Interactive information retrieval. *Annual Review of Information Science and Technology, 42*(1), 43-91.

Salton, G. (1984). The Use of Extended Boolean Logic in Information Retrieval, in Paper presented *at the 1984 ACM SIGMOD International Conference on Management of Data* (New York: ACM Pr., 1984), 277.

Salton, G. (1989). *Automatic text processing: the transformation, analysis, and retrieval of information by computer*. Reading, MA: Addison-Wesley.

Saracevic, T. (2011). "Foreword". In: Ruthven, I. & Kelly, D. (eds.), *Interactive Information Seeking, Behaviour and Retrieval*, 15-32. London: Facet

Sarwar, B., Karypis, G., Konstan, J., & Reidl, J. (2001). *Item-based collaborative filtering recommendation algorithms*. Paper presented at the 10th international conference on World Wide Web, 285–295. New York, New York, USA: .ACM Press

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000).Analysis of Recommendation Algorithms for E-Commerce. Paper presented at the 2nd ACM Conference on E-Commerce.17-20 October 2000, Minneapolis, USA. New York: ACM.

Schafer, J. B., Frankowski, D, Herlocker, J., & Sen, S. (2007). Collaborative Filtering Recommender Systems.In: Brusilovsky, P., Kobsa, A., &Nejdl, W. (Eds.), *The Adaptive Web*, pp. 291–324. Berlin: Springer-Verlag.

Schafer, J. B., Konstan, J., & Riedi, J. (2002). *Meta-recommendation Systems: User-controlled Integration of Diverse Recommendations*. Paper presented at CIKM '02 Conference on Information and Knowledge Management, 43-51., McLean, VA, USA: ACM Press.

Schafer, J. B., Konstan, J., & Riedl, J. (1999). Recommender systems in e-commerce. Paper presented at the 1st ACM conference on Electronic commerce (EC '99), 158-166. New York, USA: ACM Press.

Schafer, J. B., Konstan, J., & Riedl, J. (1999). *Recommender Systems in E-Commerce*. Paper presented at the1[st] ACM conference on Electronic Commerce, 158-166. Denver, Colorado.: ACM Press.

Schafer, J. B., Konstan, J., & Riedl, J. (1999). Recommender Systems in E-Commerce.In: *ACM E-Commerce 1999*. Paper presented at the 1st ACM Conference on E-Commerce.3-5 November 1999, Denver, USA. New York: ACM.

Schafer, J. B., Konstan, J., & Riedl, J. (2001). E–Commerce Recommendation Applications.*Data Mining and Knowledge Discovery*, 5 (1/2), 115-153.

Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2005). CROC: A New Evaluation Criterion for Recommender Systems.*Electronic Commerce Research*, 5 (1), 51-74.

Shneiderman, B. & Plaisant, C. (2005). *Designing the User Interface*. Boston, MA: Person/Addison-Wesley.

Schwandt, T. A. (2000). Three epistemological stances for qualitative inquiry. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research,* 189-213. Thousand Oaks CA: Sage.

Selected OCLC, RIN, and JISC User Behaviour Projects. JISC Report.

Sellen, A.J., Murphy, R., & Shaw, K.L. (2002). *How knowledge workers use the Web*. Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, 227–234. Minneapolis, MN: ACM Press.

Shani, G., Heckerman, D., & Brafman, R. I. (2005).An MDP-Based Recommender System.*Journal of Machine Learning Research*, 6, 1265-1295.

Shneiderman, B., Byrd, D., & Croft, W. B. (1997). Clarifying Search: A User-Interface Framework for Text Searches. *D-Lib Magazine, 3*(1).

Sinha, R. & Swearingen, K. (2002).The Role of Transparency in Recommender Systems.In: *CHI 2002*. Paper presented at the Conference of Human Factors in Computing Systems.20-25 April, 2002, Minneapolis, USA. New York: ACM.

Slone, D. J. (2000). Encounters with the OPAC: On-Line Searching in Public Libraries. *Journal of the American Society for Information Science, 51(*8), 757-773.

Smeaton, A. F., & Callan, J. (2005). Personalisation and recommender systems in digital libraries. *International Journal on Digital Libraries, 57*(4), 299-308.

Smeaton, A. F., & Callan, J. (2005).Personalisation and Recommender Systems in Digital Libraries.*International Journal on Digital Libraries*, 57 (4), 299-308.

Smith, C. L., & Kantor, P. B. (2008). *User adaptation: good results from poor systems.* Paper presented at the 31st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '08), 147-154. New York, USA: ACM Press.

Smith, C.L, & Kantor, P. B. (2008). *User adaptation: good results from poor systems*. Paper presented at 31st annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '08), 147-154.

Smyth, B., & McClave, P. (2001). *Similarity vs. Diversity*. ICCBR 2001, LNCS Volume 2080, 347-361. Springer Heidelberg. *Society for Information Science*, *32*(5), 20–22. doi:10.1002/bult.2006.1720320508

Spink, A., & Jansen, B. J. (2004). *Web search: Public searching of the Web*. New York: Kluwer.

Spink, A., Park, M., Jansen, B. J., & Pedersen, J. (2006), Multitasking during web search sessions. *Information Processing and Management, 42*(1), 264–275.

Su, X. & Khoshgoftaar, T. M. (2009). A Survey of Collaborative Filtering Techniques.*Advances in Artificial Intelligence*, 2009, 2-21.

Sutcliffe, A.G., & Ennis, M. (1998). Towards a cognitive theory of information retrieval. *Interacting with Computers, 10*, 321–351.

Svenonius, E. (2000). *The Intellectual Foundation of Information Organization.* Massachusetts: MIT Press.

Svensson, M., Höök, K., & Cöster, R. (2005): Designing and Evaluating Kalas: A Social Navigation System for Food Recipes. *ACM Transactions on Computer-Human Interaction, 12*(3), 374-400.

Swanson, D. R., Vaughan D. K., & Cooper, W. S. (1968). *Requirements study for future catalogs*. Progress report No. 2. Chicago: University of Chicago.

Swanson, T. A., & Green, J. (2011). Why We Are Not Google: Lessons from a Library Web site Usability Study, *The Journal of Academic Librarianship*, 37 (3), 222–229.

Swearingen, K., & Sinha, R. (2001). *Beyond Algorithms: An HCI Perspective on Recommender Systems*. ACM SIGIR. Workshop on Recommender Systems, 13(5-6), 393-408. DOI= 10.1007/978-0-387-85820-3_1

Takacs, G., Pilaszy, I., Nemeth, B., & Tikk, D. (2009).Scalable Collaborative Filtering Approaches for Large Recommender Systems. *Journal of Machine Learning Research*, 10, 623-656.

Tam, Winnie, Cox, Andrew M., & Bussey, A. (2009) Student user preferences for features of next-generation OPACs: a case study of University of Sheffield international students. *Program: Electronic Library and Information Systems*, 43 (4), 349-374.

Tay, A. (2013). *A survey of Library Virtual Shelves - 8 examples* (blog). http://musingsaboutlibrarianship.blogspot.co.uk/2013_06_01_archive.html#.VFeeyvmsV8E

Tintarev, N. & Masthoff, J. (2007).A Survey of Explanations in Recommender Systems.In: *ICDEW '07*. Paper presented at the 2007 IEEE 23rd International Conference on Data Engineering Workshop.17-20 April, 2007, Istanbul, Turkey.

Tintarev, N., & Masthoff, J. (2007). *A Survey of Explanations in Recommender Systems*. Paper presented at the 23rd International Conference on Data Engineering Workshop (ICDEW '07), 801-810. New York, USA: IEEE Computer Society.

Toms, E. G. (2000). *Serendipitous Information Retrieval*. Paper presented at the first DELOS Workshop: Information Seeking, Searching and Querying in Digital Libraries. Zurich, Switzerland.

Toms, E. G. (2000). Understanding and Facilitating the Browsing of Electronic Text. *International Journal of Human Computer Studies, 52*(3), 423–452.

Toms, E. G., & McCay-Peet, L. (2009). Chance Encounters in the Digital Library. *LNCS: Research and Advanced Technology for Digital Libraries, 5714*, 192-202. Berlin; New York: Springer-Verlag.

Torres, R., McNee, S. M., Abel, M., Konstan, J.A., & Riedl, J. (2004). Enhancing Digital Libraries with TechLens+ Paper presented at the JointConferenceon Digital Libraries, 7-9 June 2004, Tuscon, Arizona, US.

Trainor, C. (2009). Open source, crowd source: harnessing the power of the people behind our libraries. *Program: Electronic Library and Information Systems,* 43 (3), 288-298.

transaction log analysis to improve OPAC retrieval results. *College and Research Libraries, 59*(1), 39–50.

Tsuji, K., Kuroo, E., Sato, S., Ikeuchi, U., Ikeuchi, A., Yoshikane, F., & Itsumura, H. (2012). *Use of library loan records for book recommendation*. Paper presented at the 2012 IIAI International Conference on Advanced Applied Informatics (IIAIAAI), 30-35. Los Alamitos, CA, USA: IEEE Computer Society.

University of Minnesota (2009). *Discoverability: Phase 1 Final report*. University of Minnesota Libraries.

Vaidhyanathan, S. (2009). The Googlization of Universities. *The NEA 2009 Almanac of Higher Education*. http://www.nea.org/assets/img/PubAlmanac/ALM_09_06.pdf

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences, 15*(3), 398-405.

Vellino, A. (2010). A comparison between usage-based and citation-based methods for recommending scholarly research articles. *Journal of the American Society for Information Science and Technology 47*(1), 1–2.

Vellino, A. (2010). A Comparison Between Usage-based and Citation-based Methods for Recommending Scholarly Research Articles. Paper presented at the ASIS&T 2010 Annual Meeting, October 22–27, 2010, Pittsburgh, PA, USA.

Vellino, A., & Zeber, D. (2007). A Hybrid, Multi-Dimensional Recommender for Journal Articles in a Scientific Digital Library. Paper presented at the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, 2-5 November, 2007, Silicon Valley, US.

Verbert, K., Parra, D., Brusilovsky, P., & Duval, E. (2013). Visualizing recommendations to support exploration, transparency and controllability. Paper presented at the 2013 International Conference on Intelligent User Interfaces (IUI'13), 351-364. Santa Monica CA, USA: ACM Press.

Victor, P., Cornelis, C., de Cock, M., & da Silva, P. P. (2009). Gradual trust and distrust in recommender systems. *Fuzzy Sets and Systems*, 160 (10), 1367–1382.

Wakeling, S., Clough, P., Sen, B., & Silipigni Connaway, L. (2012). "Readers who borrowed this also borrowed...": Recommender systems in UK libraries. *Library Hi Tech, 30*(1), 134-150.

Wakeling, S., Clough, P., Sen, B. (2014). "Investigating the potential impact of non-personalized recommendations in the OPAC: Amazon vs. WorldCat.org." In Proceedings of the 5th Information Interaction in Context Symposium (IIiX '14). ACM, New York, NY, USA, 96-105.

Webb, A., & Kerne, A. (2008). *The In-Context Slider: A Fluid Interface Component for Visualization and Adjustment of Values while Authoring*. Paper presented at the working conference on Advanced visual interfaces, 91-99. New York, NY, USA: ACM Pres.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences.* Chicago, IL: Rand McNally

Webster, J., Jung, S. & Herlocker, J. (2004), Collaborative Filtering: a new approach to searching digital libraries, *New Review of Information Networking*, Vol. 10 No. 2, pp. 177 – 191.

Westcott, J., Chappell, A., & Lebel, C. (2009).LibraryThing for libraries at Claremont.*Library Hi Tech*, 27 (1), 78-81.

White, R. W., & Morris, D. (2007). *Investigating the querying and browsing behavior of advanced search engine users.* Paper presented at the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 255–262. Pisa, Italy: ACM Press. doi:10.1145/1277741.1277787

White, R. W., & Roth, R. A. (2009). *Exploratory Search: Beyond the Query–Response Paradigm*. San Francisco, CA: Morgan Claypool.

Whitney, C., & Schiff, L. (2006). The Melvyl recommender project: developing library recommendation services. *D-Lib Magazine, 12(12)*.

Whitney, C., &Schiff, L. (2006). The Melvyl Recommender Project: Developing Library Recommendation Services. *D-Lib Magazine*, 12 (12).http://dlib.org/dlib/december06/whitney/12whitney.html

Wildemuth, B. M. (2004). The effects of domain knowledge on search tactic formulation . Journal of the American Society for Information Science and Technology, 55(3), 246–258.

Wildemuth, B. M., & O'Neill A. L. (1995). The "known" in known-item searches: empirical support for user-centered design. *College and Research Libraries, 56*(3), 265–281.

Wilson, M. L., & Schraefel, M. C. (2008). *A longitudinal study of exploratory and keyword search*. Paper presented at the 8th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '08), 52-56. New York, NY, USA: ACM Press. DOI=10.1145/1378889.1378899

Wilson, T. (1999). Models in information behaviour research. *Journal of Documentation, 55*(3), 249-270.

Wilson, T.D., & Walsh, C. (1996). *Information behaviour: an interdisciplinary perspective.* Sheffield: University of Sheffield Department of Information Studies.

Woods, R. (2010). From Federated Search to the Universal Search Solution, *The Serials Librarian*, 58 (2), 141–148.

Woolley, C. M. (2009). Meeting the Mixed Methods Challenge of Integration in a Sociological Study of Structure and Agency. *Journal of Mixed Methods Research, 3*(7).

Xiao, B., & Benbasat, I. (2007). E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact. *MIS Quarterly, 31*(1), 137-209.

Xiao, B., & Benbasat, I. (2007). ECommerce product recommendation agents: use, characteristics, and impact. *Mis Q. 31*, 137–209.

Yager, R. R. (2003). Fuzzy logic methods in recommender systems.*Fuzzy Sets and Systems*, 136 (2), 133–149.

Yang, S. O, & Wagner, K. (2010). Evaluating and comparing discovery tools: how close are we towards next generation catalog? *Library Hi Tech, 28*(4), 690–709.

Yang, S. Q. and Hofmann, M. A. (2011), Next Generation or Current Generation?A Study of the OPACs of 260 Academic Libraries in the United States and Canada, *Library Hi Tech*, 29 (2).

Yang, S. Q. and Wagner, K. (2010), Evaluating and comparing discovery tools: how close are we towards next generation catalog?,*Library Hi Tech*, 28 (4), 690 – 709.

Yee, M. M., & Layne S. S. (1998). *Improving online public access catalogs* Chicago: American Library Association.

Yin, P., Zhang, M., & Li, X. (2007).Recommending Scientific Literatures in a Collaborative Tagging Environment. Paper presented at the 10th international conference on Asian digital libraries, 10-13 December, 2007, Hanoi, Vietnam.

Youngen, G. *Illinois WorldCat Local Usability Study Final Report*. University of Illinois at Urbana-Champaign.

Zhang, J., & Ghorbani, A. A. (2004). The Reconstruction of User Sessions from a Server Log Using Improved Time-oriented Heuristics. Paper presented at the Second Annual Conference on Communication Networks and Services Research. Fredericton, NB, Canada.

Zhang, J., Jones, N., & Pu, P. (2008). *A Visual Interface for Critiquing-based Recommender Systems*. Paper presented at EC'08. Chicago, Illinois, USA.

Zhang, Y. , & Wildemuth, B. M. (2009). *Qualitative analysis of content*. In B. Wildemuth (Ed.), Applications of Social Research Methods to Questions in Information and Library Science, 308-319. Westport, CT: Libraries Unlimited.

Zheng, H., Wang, D., Zhang, Q., Li, H., & Yang, T. (2010). Do Clicks Measure Recommendation Relevancy? An Empirical User Study. Paper presented at the fourth ACM conference on Recommender systems.26-30 September, 2010, Barcelona, Spain. New York: ACM.

Zhou, T., Kuscsik, Z., Liu, J-G., Medo, M., Wakeling, J. R., & Zhang, Y-C. (2010). Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, 107 (10), 4511-4515

Zhou, T., Kuscsika, Z., Liu, J. G., Medoa, M., Wakeling, J. R., & Zhang, Y. C (2010). Solving the apparent diversity-accuracy dilemma of recommender systems. Paper from the *National Academy of Sciences, 107*(10), 4511–4515.

Ziegler, C. N., Lausen, G., & Schmidt-Thieme, L. (2004). *Taxonomy-driven computation of product recommendations*. Paper presented at the Thirteenth ACM conference on Information and knowledge management - CIKM '04, 406. New York, New York, USA: ACM Press.

Ziegler, C. N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). *Improving recommendation lists through topic diversification*. Paper presented at *the 14th international conference on World Wide Web* (WWW '05), 22-32. New York, NY, USA: ACM Press. DOI=10.1145/1060745.1060754

Ziegler, C.N., McNee, S. M., Konstan, J. A., &Lausen, G. (2005).Improving Recommendation Lists Through Topic Diversification. Paper presented at the 14th international conference on World Wide Web, 10-14 May, 2005, Chiba, Japan. New York: ACM.

Zins, A., & Bauernfeind, U. (2005). *Explaining online purchase planning experiences with recommender websites*. Paper presented at the International Conference on Information and Communication Technologies in Tourism, 137–148. Innsbruck, Austria.

Zoe, L. R., DiMartino, D. (2000). Cultural diversity and end-user searching: An analysis by gender and language background. *Research Strategies, 17* (4), 291–305.

APPENDIX 1: Focus Group Interview Questions

1. Tell us about your experiences with WorldCat.org

[PROBES: Do you use WorldCat.org? If not - why not, and what services do you use instead? If you use it, do you use it professionally or is it for leisure tasks? What type of tasks do you typically use it for? What aspects of WorldCat.org functionality do you make use of (e.g. Find a copy in a library, related subjects etc)? Do you have a user profile? Do you look at tags or reviews? Have you ever added a review or tag?]

2. Describe a time when you used WorldCat.org that you considered a success.

[PROBES: What made you go to WorldCat.org? Explain what you did, i.e., what did you search, how did you search etc.? What features of WorldCat.org did you use, e.g., ratings, reviews, recommendations, tags, bibliographic data? Why did you decide / how did you know when to stop searching?]

3. Describe a time when using WorldCat.org was unsuccessful – i.e., you did not get what you wanted.

[PROBES: What made you go to WorldCat.org? Explain what you did (i.e. what did you search, how did you search etc. What features of WorldCat did you use (e.g. ratings, reviews, recommendations, tags, bibliographic data? What made you decide to stop searching? Where did you go to find the information?]

4. Think of a time when you did not find what you were looking for, but did find something else of interest or useful to your work?

[PROBES: How did you find this other source? How else do you think you could have found this other source? What made this source useful? Why do you think the system presented it to you, i.e., why do you think it appeared / was returned?]

5. If you had a magic wand, what would your ideal WorldCat.org provide? How would you go about using it? When? Where? Why?

[PROBES: Try to find out if and under what circumstances/why they would use

WorldCat.org. What changes would you make to WorldCat.org to make it better meet your

needs?]

APPENDIX 2: Breakdown of Focus Groups by Location and Participants

Location	Country	Date	User Group	Participants
National Library of New Zealand	NZ	03.21.2011	Librarians	3
University of Auckland	NZ	03.22.2011	Librarians	6
University of Waikato	NZ	03.25.2011	Librarians	1
National Library of New Zealand	NZ	03.30.2011	Historians	2
National Library of New Zealand	NZ	03.30.2011	Librarians	5
University of Sydney	AUS	04.04.2011	Librarians	4
CAVAL, Melbourne	AUS	04.07.2011	Historians	1
CAVAL, Melbourne	AUS	04.08.2011	Librarians	5
London School of Economics	UK	05.09.2011	Librarians	3
British Library	UK	05.10.2011	Booksellers	9
British Library	UK	05.10.2011	Librarians	9
London School of Economics	UK	05.11.2011	Historians	1
University of Nottingham	UK	05.16.2011	Historians	2
University of Nottingham	UK	05.16.2011	Students	8
University of Nottingham	UK	05.17.2011	Students	9
University of Nottingham	UK	05.17.2011	Librarians	8
Simmons College	US	10.25.2011	Librarians	7
Simmons College	US	10.25.2011	Students	9
Northeastern University	US	10.26.2011	Librarians	13
Northeastern University	US	10.26.2011	Students	12
Simmons College	US	10.27.2011	Students	3
			Total	**120**

APPENDIX 3: Focus Group Code Book

2. Uses
    a. Work-Tasks
        i. Professional
            1. Cataloging
            2. ILL
            3. Acquisitions/Collection Development
            4. Instruction / Training
        ii. Academic
            1. Reading-list development
            2. Research
            3. Essay / Assignment
        iii. Leisure
            1. Reading for pleasure
            2. Personal research
    b. Search-Tasks
        i. Holdings
            1. Ranking
            2. Popularity
            3. Uniqueness
            4. Library specializations
        ii. Publication Trends
        iii. Institutional Information
            1. Address
            2. Policies
        iv. Known item
            1. Reference verification
            2. Provenance
            3. Location
            4. Starting point
            5. Different Editions
            6. Different Media (ebooks)
        v. Unknown item
            1. Related
                a. Author
                b. Version
            2. Subject
                a. Completeness
                b. Serendipity
                c. Niche/specialized items
                d. Teaching
                e. Monitoring

3. System Strengths

    a. Function
        i. Ease of use
        ii. Easy to teach
        iii. Citations
            1. Facebook app

        2. Export list
- iv. User Profile
- v. Reviews
- vi. Interface
- vii. Filters and sorting
- viii. Recommendations
- ix. Mobile Access
- x. Timeline
- xi. "Find more information…"
- b. Content
  - i. TOC
  - ii. Articles
  - iii. OAISTER
  - iv. Full Text
  - v. Foreign Language
  - vi. Obscure Items
  - vii. Obscure Authors
  - viii. Different Editions
  - ix. Global Scope
  - x. Metadata
  - xi. Location
  - xii. Different Formats

4. System Challenges/Difficulties/ Drawbacks
   - a. Function
     - i. Social media
       1. Reviews
       2. Tags
       3. Profile
       4. Lists
     - ii. Search
       1. Ranking
          - a. Lack of explainability
          - b. Poor
       2. Inaccurate results
       3. Too many items retrieved
     - iii. Lacks personalization
     - iv. Interface
     - v. Difficulties Logging-In
     - vi. Inaccurate location in display
     - vii. Problems with Citation Function
     - viii. Dead links
     - ix. Recommendations
     - x. API
       1. Floods federated search
       2. Difficult to use
     - xi. Displays multiple editions instead of FRBR work level display
     - xii. Performance issues

   - b. Content
     - i. Inaccurate metadata

          ii. Inaccurate holdings
          iii. Duplicate records
          iv. Music
                1. Lack of content
          v. Controlled vocabulary not reflective of user needs
          vi. Lack of Primary Sources
          vii. Provenance
     c. Marketing
          i. No knowledge of existence of service
          ii. Lack of understanding of range of services
          iii. Lack of understanding of membership terms, privileges, and contributions

5. Suggested Improvements
     a. Function
          i. Customization
                1. Improve Interface
                2. Ability to select multiple favourite libraries
                3. Ability to turn off and on features
                4. Multiple user levels
                5. Automatic geographic recognition
                6. Advanced Search
                7. Notification of New Items
                8. Facilitate sharing / collaboration
          ii. WorldCat Local functionality
          iii. Better mobile access
          iv. Spelling Variations
                1. "Did you mean…?"
                2. Display results for spelling variations
          v. Search
                1. Highlight search terms
                2. Improve relevance ranking
                3. Transparent ranking
                4. Faceted search display
                5. Sort results by:
                      a. Date
                      b. Publisher
                      c. Format
                      d. Country
                6. Personalisation
                7. Save Search
          vi. Visualization
                1. Map holdings
          vii. Links to related information
                1. Amazon
                2. Antiquarian Booksellers
                3. Publisher pages
                4. Author pages, videos, etc.
                5. Authoritative reviews
          viii. Full ILL capabilities
          ix. Easier to use API

x.   Better Hyperlinking
                                    a.   Links to WC from other catalogs
                                    b.   Links from WC to other OPAC record pages
                   xi.   Improved Accessibility (for disabled)
                  xii.   Work level display (FRBR)
                 xiii.   Popularity metrics (e.g. times item has been viewed)
                 xiv.   Browse by Publisher
                  xv.   Recommendations
                                    1.   Distracting
                                    2.   Not Scholarly
                                    3.
                 xvi.   Improved foreign language handling
                xvii.   Clearer indication of item format
        b.   Content
                    i.   Add supplementary information
                                    1.   Book covers
                                    2.   Summary
                                    3.   Authoritative reviews
                                    4.   Author videos discussing publication
                                    5.   Information about Illustrations
                                    6.   Language of resource
                                    7.   Student reviews
                                    8.   Ratings
                   ii.   Merge duplicate records
                  iii.   Primary Sources
                  iv.   Full text / ebooks
                   v.   Newspapers
                  vi.   Holdings
                                    1.   Up-to-date
                                    2.   Universal
                                    3.   Item availability
                                    4.   Accuracy
                 vii.   Metadata
                                    1.   More accurate
                                    2.   More comprehensive
                viii.   Inform whether item still in print
                  ix.   Greater content granularity
                   x.   Theses

        c.   Marketing
                    i.   OCLC representative visits
                   ii.   Training podcasts
                  iii.   Links from OPAC to WorldCat, if no local item
                  iv.   Encourage word-of-mouth advocacy
                   v.   Clarify membership terms, privileges and contributions

APPENDIX 4: Pop-Up Survey Questions

1. Gender: (Select one answer only)
    i.   Female
    ii.   Male
2. Age (Select one answer only)
    i.   Under 18
    ii.   18-25
    iii.   26-35
    iv.   36-49
    v.   50+
3. In which country do you live?

Dropdown list

4. In which State do you live?

Dropdown list

5. What is your main purpose for using WorldCat.org today? (Select one answer only)
    i.   Educational
    ii.   Professional
    iii.   Recreational
6. What is your occupation? (Select one answer only)
    i.   Undergraduate student
    ii.   Graduate/Post-Graduate student
    iii.   Faculty/Researcher
    iv.   Librarian
    v.   Other (please specify)
7. What material are you looking for? (Select all that apply)
    i.   Audio
    ii.   Book
    iii.   Journal article
    iv.   Video
    v.   Other (please specify)
8. How important is it to you to access full versions of the items you are looking for online
(e.g. ebook or ejournal text)? (Select one answer only)
    i.   Very important
    ii.   Important
    iii.   Somewhat important
    iv.   Not important
9. Where do you normally go to search for online versions of text? (Select all that apply)
    i.   College / University Library website or databases
    ii.   Public Library website or databases
    iii.   Google Books
    iv.   Google Scholar

    v.   Online retailer (e.g. Amazon)

    vi.   Search engine (Google, Yahoo!, Bing, etc.)

   vii.   Wikipedia

 viii.   Other (please specify)

10. What is your reason for using / visiting this site? (Select all that apply)

    i.   To find information about a particular topic (e.g. to find information about cats or World War II)

    ii.   To search for one or more unknown books or resources about a particular topic (e.g. to find a books about cats or World War II that you weren't previously aware of)

   iii.   To find a location (either in a library or online) of a specific known item (e.g. to see whether a local library has a particular item that you are already aware of)

   iv.   To find more information about a specific known item (e.g. to check the publication date or author of an item that you are already aware of).

    v.   Other (please specify)

11. Have you ever created a personal WorldCat.org account? (Select one answer only)

    i.   Yes

    ii.   No

   iii.   Don't know

12. Have you ever used any of the following to find or evaluate an item on this site? (Select all that apply)

    i.   Lists (A way for you to group items you have found on WorldCat.org)

    ii.   Ratings (User-generated star ratings from 1 to 5)

   iii.   Reviews (Written evaluations of items either by users of WorldCat.org or from editorial sources)

   iv.   Tags (Keywords or terms attached to an item by users)

    v.   None of the above

13. Have you ever added any of the following to this site? (Select all that apply)

    i.   Lists (A way for you to group items you have found on WorldCat.org)

    ii.   Ratings (User-generated star ratings from 1 to 5)

   iii.   Reviews (Written evaluations of items either by users of WorldCat.org or from editorial sources)

   iv.   Tags (Keywords or terms attached to an item by users)

    v.   None of the above

14. What type of recommendations (i.e. suggestions of related items) would be useful on this site? (Select all that apply)

    i.   Recommendations based on expert opinion

    ii.   Recommendations based on items with similar content or authors

   iii.   Recommendations based on library circulation data

   iv.   Recommendations based on the most viewed record pages

    v.   Recommendations based on the ratings of other users

   vi.   None – I have no use for recommendations

   vii.   Other (please specify)

15. What type of reviews would be useful on this site? (Select all that apply)

    i.   Reviews by experts (e.g. academics or critics)

    ii.   Reviews by other users

   iii.   None – I have no use for reviews

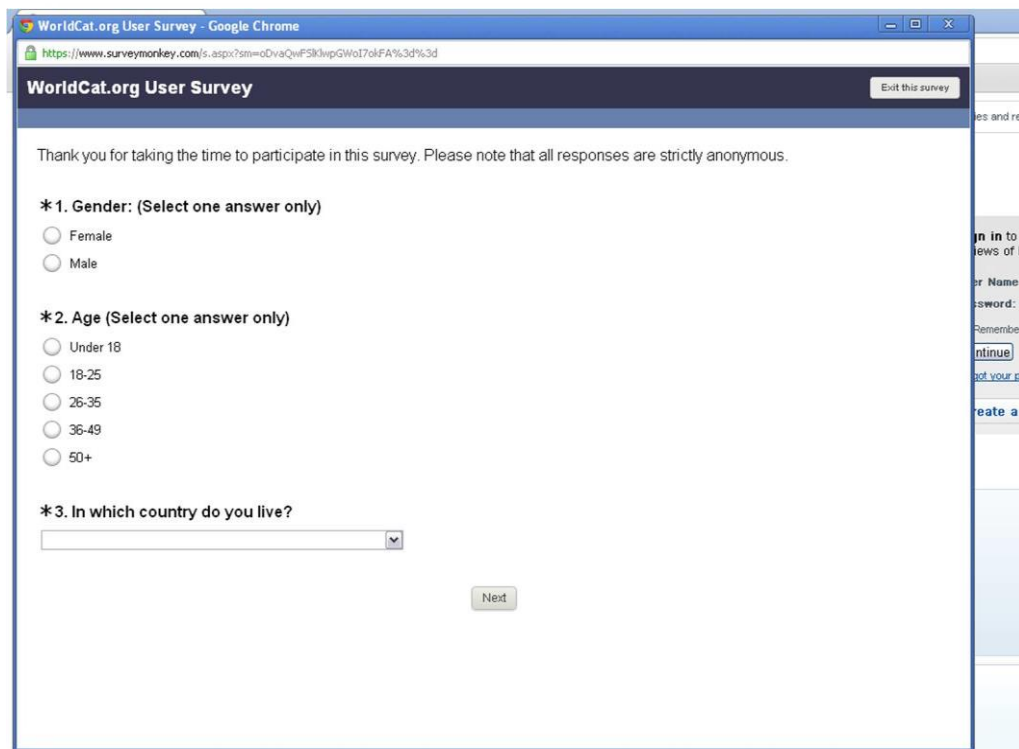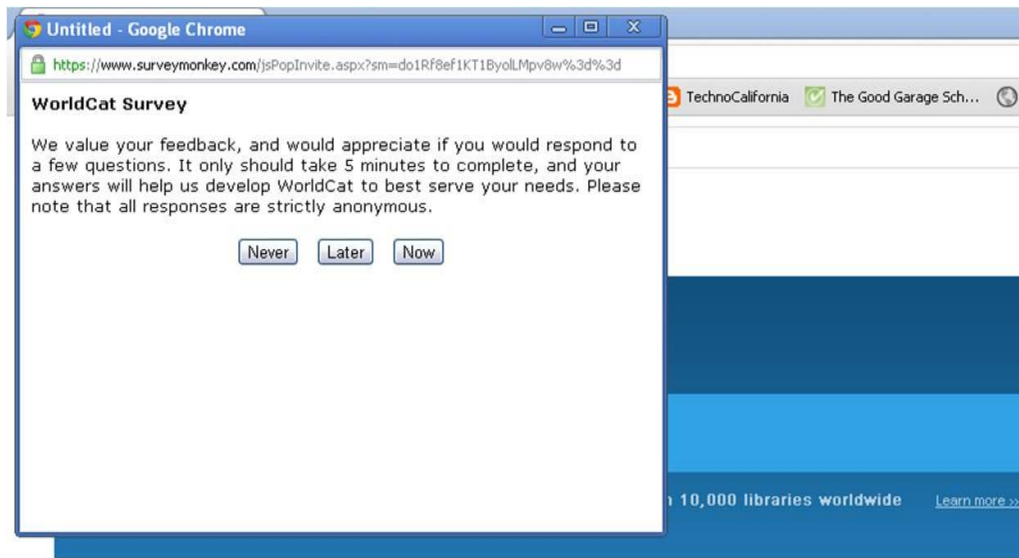iv.   Other (please specify)

16. Please add any other comments in the space below

Free text field

17. Please enter an email address below if you are willing to participate in a more detailed follow up questionnaire.

Free text field

APPENDIX 5: Screenshots of Pop-Up Invitation and Survey

APPENDIX 6: Actions Identified in WorldCat.org Logs

GENERAL
Change language from the main page
Provide feedback on a particular search or item from the search results

NAVIGATION
Open WorldCat.org homepage
Open "Search for Lists" page
Open "WorldCat Genres" page
Open "Search" page
Open "Advanced search" page
Open "About WorldCat" page
Open "Feedback" page
Open "Search for a Library" page
Open "Search for Contacts" page
Open "WorldCat Mobile" page

SEARCH FOR ITEMS
Query submitted from WorldCat.org homepage Search Articles tab
Query submitted from WorldCat.org homepage Search Everything
Query submitted from WorldCat.org homepage Search Books
Query submitted from WorldCat.org homepage Search DVDs
Query submitted from WorldCat.org homepage Search CDs
Query submitted from WorldCat.org homepage Search for Library Items
Query submitted from the search results page
Query submitted from "No results" page
Query submitted after clicking on spelling suggestion
Query submitted from elsewhere on site
Query submitted from "Search for Lists" page
Query submitted from "Advanced search" page
Query submitted includes advanced search operators
Query submitted with no search terms from "Advanced search" page

Query submitted from an affiliate
Query submitted from external library
Query submitted from iframe in external webpage

RESULTS PAGE
Sort search results by Author (A-Z)
Sort search results by Date (Newer First)
Sort search results by Date (Older First)
Sort search results by Title (A-Z)
Click "First" link in the search results page
Click "Next" link in the search results page
Click "Prev" link in the search results page
Click number link to a particular page in the search results page
Click "View all editions and formats"
Refine search results by audience
Refine search results by author
Refine search results by content
Refine search results by format
Refine search results by language
Refine search results by topic

Refine search results by year
Remove refine filters from search results

VIEW ITEM
View an item after clicking a result from on the search results page
View an item from another WorldCat page
View an article in a journal through the OpenURL Gateway
View an item after clicking in the results page of another library or resource
View an item from a frame in an external webpage
View a journal after clicking in the results page of another library or resource

From record page, click on a related subject (executes a subject search)
From record page, click on an author name (executes an author search)
From record page, click on "Write a review"
From record page, click on "Add to list"
From record page, click on "Add Tags""
From record page, click on "Cite/Export"
From record page, export citation to reference manager service
From record page, click on "View all editions and formats"
From record page, click on "Find a copy in the library"
From record page, click on link to a library holding the item.
From record page, click "Return to Search Results"
From record page, click "Preview this item"
From record page, click "Find out more about:" (navigates to WorldCat identities)
From record page, click on a single tag
From record page, click on "View all tags for this item"

ACCOUNT
Click "Create an Account"
Click "Sign In"
Click "Forgot your password?"
Click "My Lists"
Click "Modify a list"
Click "My Reviews"
Click "My Saved Searches"
Click "My Tags"
Click "My Watchlist"
Click "My WorldCat"
Click "Edit profile"
Click "Save Profile"
View a specific account holder List
View another user's profile
View a list of items tagged with specific keywords by a particular user

LIBRARY / SYSTEM / USER INFROMATION
View information on a library

APPENDIX 7: Full Latin Square Design for User Study

	Task Order			
Subject	1	2	3	4
S1	2 (2, 2)	1 (2)	1 (1)	2 (1, 1)
S2	2 (1, 2)	1 (1)	2 (2, 1)	1 (2)
S3	1 (2)	2 (2, 2)	2 (1, 1)	1 (1)
S4	1 (1)	2 (1, 2)	1 (2)	2 (2, 1)
S5	2 (1, 1)	1 (2)	1 (1)	2 (2, 2)
S6	2 (2, 1)	1 (1)	2 (1, 2)	1 (2)
S7	1 (2)	2 (1, 1)	2 (2, 2)	1 (1)
S8	1 (1)	2 (2, 1)	1 (2)	2 (1, 2)
S9	2 (2, 2)	2 (1, 1)	1 (1)	1 (2)
S10	1 (2)	1 (1)	2 (2, 1)	2 (1, 2)
S11	2 (1, 1)	2 (2, 2)	1 (2)	1 (1)
S12	1 (1)	1 (2)	2 (1, 2)	2 (2, 1)
S13	1 (2)	2 (1, 1)	1 (1)	2 (2, 2)
S14	2 (2, 1)	1 (1)	1 (2)	2 (1, 2)
S15	2 (1, 1)	1 (2)	2 (2, 2)	1 (1)
S16	1 (1)	1 (2)	2 (1, 2)	2 (2, 1)
S17	2 (1, 1)	2 (2, 2)	1 (1)	1 (2)
S18	1 (2)	1 (1)	2 (1, 2)	2 (2, 1)
S19	2 (2, 2)	2 (1, 1)	1 (2)	1 (1)
S20	1 (1)	1 (2)	2 (2, 1)	2 (1, 2)
S21	1 (2)	2 (2, 2)	1 (1)	2 (1, 1)
S22	2 (1, 2)	1 (1)	1 (2)	2 (2, 1)
S23	2 (2, 2)	1 (2)	2 (1, 1)	1 (1)
S24	1 (1)	2 (1, 2)	2 (2, 1)	1 (2)
S25	2 (2, 2)	1 (2)	1 (1)	2 (1, 1)
S26	2 (1, 2)	1 (1)	2 (2, 1)	1 (2)
S27	1 (2)	2 (2, 2)	2 (1, 1)	1 (1)
S28	1 (1)	2 (1, 2)	1 (2)	2 (2, 1)
S29	2 (1, 1)	1 (2)	1 (1)	2 (2, 2)
S30	1 (1)	2 (2, 1)	2 (1, 2)	1 (2)
S31	1 (2)	2 (1, 1)	2 (2, 2)	1 (1)
S32	1 (1)	2 (2, 1)	1 (2)	2 (1, 2)
S33	2 (2, 2)	2 (1, 1)	1 (1)	1 (2)
S34	1 (2)	1 (1)	2 (2, 1)	2 (1, 2)
S35	2 (1, 1)	2 (2, 2)	1 (2)	1 (1)
S36	1 (1)	1 (2)	2 (1, 2)	2 (2, 1)

Key: x (y, z), where x = Task, y = System, and z = Book

APPENDIX 8: Demographic Questionnaire

1. What is your country of birth?

<Drop Down List>

2. What is your country of residence?

<Drop Down List>

3. What language do you speak at home?

<Drop Down List>

4. Which languages do you use to search the web?

<Drop Down List>

5. Please select your age group:

- 18-25
- 26-35
- 36-45
- 46-55
- 56-65
- Over 65

6. I identify my gender as:

- Female
- Male
- Trans*
- _____

7. Please select completed an in progress educational courses and programmes

- Secondary School
- Further education / College Diploma
- Undergraduate

- Masters
- Doctorate
- Professional (law, medicine etc.)

8. Are you currently and primarily (i.e. more than 50% of the time:

- Employed
- Student
- Unemployed
- Other

9. What degree are you currently studying for? Please state degree type and subject (e.g. BA English Literature; MSc Chemistry)

APPENDIX 9: Pre-Task Questionnaire

1. How often do you use the Star Plus library catalogue for academic purposes?

- Every day
- A few times per week
- A few times per month
- A few time per year
- Never

2. How often do you use Amazon for academic purposes?

- Every day
- A few times per week
- A few times per month
- A few time per year
- Never

3. How often do you use WorldCat.org for academic purposes?

- Every day
- A few times per week
- A few times per month
- A few time per year
- Never

4. How often do you use the *Star plus* library catalogue for leisure purposes?

- Every day
- A few times per week
- A few times per month
- A few time per year
- Never

5. How often do you use Amazon for leisure purposes?

- Every day
- A few times per week
- A few times per month
- A few time per year
- Never

6. How often do you use WorldCat.org for leisure purposes?

- Every day
- A few times per week
- A few times per month
- A few time per year
- Never

7. How easy to use us the Star Plus catalogue?

Never Used

1 (Extremely difficult) → 7 (Extremely easy)

8. How easy to use us the Star Plus catalogue?

Never Used

1 (Extremely difficult) → 7 (Extremely easy)

9. How easy to use us the WorldCat.org catalogue?

Never Used

1 (Extremely difficult) → 7 (Extremely easy)

APPENDIX 10: Post-Task Questionnaire

1. I am satisfied with the number of books I found to complete this task

1 (strongly disagree) → 7 (strongly agree)

2. I am satisfied with the quality of the books I found to complete this task

1 (strongly disagree) → 7 (strongly agree)

3. The system made completing this task easy

1 (strongly disagree) → 7 (strongly agree)

4. I would have found this task easier if I was using a different system

1 (strongly disagree) → 7 (strongly agree)

5. I feel like there were books in the system I couldn't find

1 (strongly disagree) → 7 (strongly agree)

6. How useful were the following features in helping you complete this task?

<list of system features>

Did not notice this feature

1 (not at all useful) → 5 (extremely useful)

APPENDIX 11: User Study Screenshots of System Features

Amazon

Front Page of Amazon
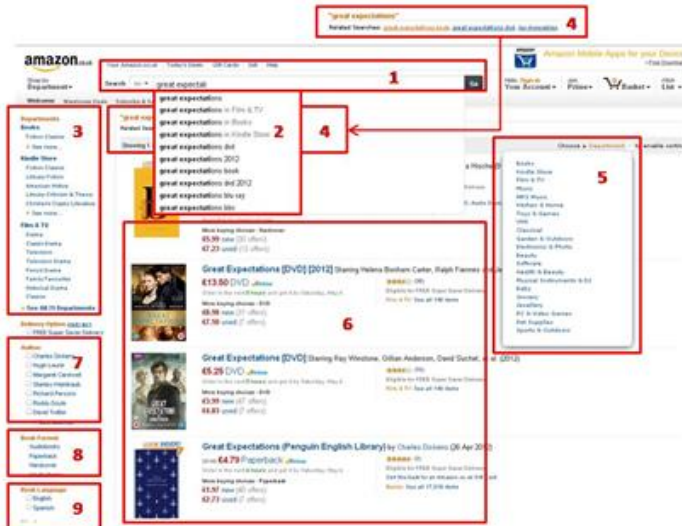
Item level page of Amazon
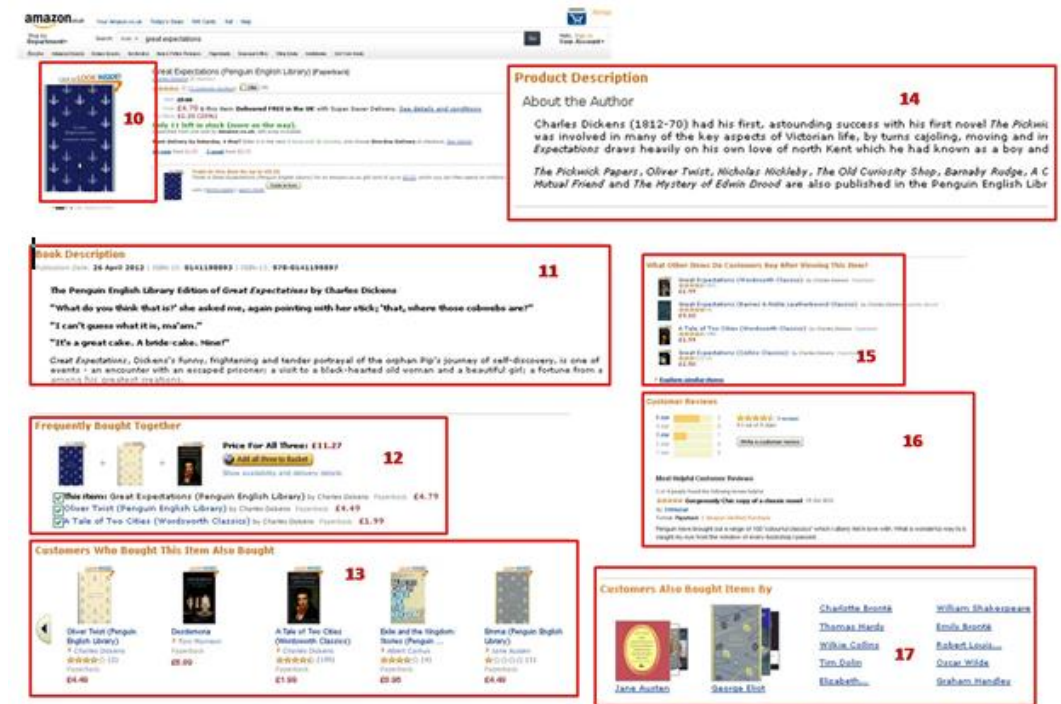
WorldCat.org

Front page of WorldCat.org

Item level page of WorldCat.org

APPENDIX 12: User Study Recruitment Email 1

Subject: Should the Library Catalogue be more like Amazon

We are investigating the functionality of library catalogues, and in particular whether the additional features offered by online retail sites such as Amazon could be usefully employed in Academic library systems.

We are seeking volunteers to participate in a study designed to answer some key questions about the value of certain features of Amazon in a library context.

During the study you will be asked a series of questions about your background, and your current perceptions of a variety of library and online retail systems. You will then be asked to undertake four tasks designed to simulate typical interactions with library catalogues. We will be using screen capture software to record your interactions with the systems while you complete the tasks. Finally there will be a short interview so we can learn more about your experience completing the tasks, and your feelings about searching library systems in general.

The study will take no more than an hour and fifteen minutes, and will take place in the new Usability Lab in the Information School. Participants should meet me at the ground floor reception of the Information School, Regents Court, 211 Portobello Street, S1 4DP.

If you wish to participate, please contact Simon Wakeling at s.wakeling@sheffield.ac.uk. Please include your name, email address and level of study ("Undergraduate, "Masters" or "PhD") in the email.

This study forms part of my PhD research, and has received research ethics approval from the University of Sheffield Research Ethics Committee. If you have any questions please contact me (s.wakeling@sheffield.ac.uk).

Many thanks,

Simon Wakeling

(Supervised by Dr. Paul Clough in the Information School)

APPENDIX 13: User Study Recruitment Email 2

Subject: Earn #10 by participating in Amazon experiment

We are investigating the functionality of library catalogues, and in particular whether the additional features offered by online retail sites such as Amazon could be usefully employed in Academic library systems.

We are seeking volunteers to participate in a study designed to answer some key questions about the value of certain features of Amazon in a library context.

**** YOU WILL BE PAID #10 IN CASH FOR PARTICIPATING IN THE EXPERIMENT ****

During the study you will be asked a series of questions about your background, and your current perceptions of a variety of library and online retail systems. You will then be asked to undertake four tasks designed to simulate typical interactions with library catalogues. We will be using screen capture software to record your interactions with the systems while you complete the tasks. Finally there will be a short interview so we can learn more about your experience completing the tasks, and your feelings about searching library systems in general.

The study will take no more than an hour and fifteen minutes, and will take place in the new Usability Lab in the Information School. Participants should meet me at the ground floor reception of the Information School, Regents Court, 211 Portobello Street, S1 4DP.

If you wish to participate, please contect Simon Wakeling at s.wakeling@sheffield.ac.uk. Please include your name, email address and level of study ("Undergraduate, "Masters" or "PhD") in the email.

This study forms part of my PhD research, and has received research ethics approval from the University of Sheffield Research Ethics Committee. If you have any questions please contact me (s.wakeling@sheffield.ac.uk).

Many thanks,

Simon Wakeling

(Supervised by Dr. Paul Clough in the Information School)

APPENDIX 14: User Study Template Word Document

N.B This document was tailored to the participant's specific task order

Task 1

Module or aspect of PhD research:

Book Titles:

[Copy and paste book titles here]

Task 2

Module or aspect of PhD research:

Book Titles:

[Copy and paste book titles here]

Task 3

Date of publication:

Book Titles:

[Copy and paste book titles here]

Task 4

Date of publication:

Book Titles:

[Copy and paste book titles here]

APPENDIX 15: User Study Task Instructions

N.B. Instructions were tailored for each participant to the task order determined by the Latin Square design.

Task 1

Open the browser with www.amazon.co.uk running.

Using this system, find a range of books that would be useful for your studies in a module you are currently taking. (If you are a PhD student, please search instead for books relating to a particular aspect of your research).

When completing the task, do not feel like you have to find as many books as possible. We are interested in the process, so think carefully about how relevant the books you choose are.

You can copy and paste the books you find into the blank Word Document that is open. Do not worry about formatting or the author's name – the title will be sufficient.

You will have a **maximum** of 10 minutes to complete this task.

When you have completed the task, click NEXT PAGE in the survey software browser.

Task 2

Open the browser with www.amazon.co.uk running.

Now imagine you have been recommended a book by a friend. The book is called **A History of Leeds** by **W.R. Mitchell**.

a) Use the system to find out when the book was published

b) Now imagine that you want to take the book out of the library, but all the copies are out on loan. Use the system to find a range of other books that you could get out instead.

When completing the task, do not feel like you have to find as many books as possible. We are interested in the process, so think carefully about how relevant the books you choose are.

You can copy and paste the books you find into the blank Word Document that is open. Do not worry about formatting or the author's name – the title will be sufficient.

You will have a **maximum** of 10 minutes to complete this task.

When you have completed the task, click NEXT PAGE in the survey software browser.

Task 3

Go to the browser with www.worldcat.org open.

Using this system, find a range of books that would be useful for your studies in a **different** module you are currently taking. (If you are a PhD student, please search instead for books relating to another particular aspect of your research).

When completing the task, do not feel like you have to find as many books as possible. We are interested in the process, so think carefully about how relevant the books you choose are.

You can copy and paste the books you find into the blank Word Document that is open. Do not worry about formatting or the author's name – the title will be sufficient.

You will have a **maximum** of 10 minutes to complete this task.

When you have completed the task, click NEXT PAGE in the survey software browser.

Task 4

Open the browser with www.worldcat.org running.

Now imagine you have been recommended another book by a friend. The book is called **Birmingham: A History of the City and Its People** by **Malcolm Dick**.

a) Use the system to find out when the book was published

b) Now imagine that you want to take the book out of the library, but all the copies are out on loan. Use the system to find a range of other books that you could get out instead.

When completing the task, do not feel like you have to find as many books as possible. We are interested in the process, so think carefully about how relevant the books you choose are.

You can copy and paste the books you find into the blank Word Document that is open. Do not worry about formatting or the author's name – the title will be sufficient.

You will have a **maximum** of 10 minutes to complete this task.

When you have completed the task, click NEXT PAGE in the survey software browser.