# Contextual Information Retrieval from the WWW

Dilip Kumar Limbu

A thesis submitted to Auckland University of Technology in partial fulfilment of the degree of Doctor of Philosophy (PhD)

2008

School of Computing and Mathematical Sciences

Primary Supervisor: Dr. Andy M. Connor

# Acknowledgements

This thesis owes much to many people. First and foremost, I am deeply grateful to my supervisor Dr. Andy Connor, for his invaluable suggestions, support, and guidance in my research. He gave me a significant amount of freedom to explore the directions I was most interested in. At the same time, however, he constantly reminded me of the importance of focusing on the essential things and foregoing lower-hanging fruit that carry little benefit in the larger perspective. His direction has helped me to grow as an independent researcher, and for that I am very thankful.

I would also like to acknowledge my profound gratitude to Prof. Stephen G. MacDonell and Dr. Russel Pears for constant support, encouragement, valuable discussions and detailed reviews during the development of this thesis. No question was ever left unanswered, and for that I will always be eternally grateful.

I am deeply grateful for the assistance of Neil Binnie, for his superb statistical analyses guidance and Gordon Grimsey, for his superb IT resource coordination, without which this thesis truly could not have existed. Thank you for being such an excellent and patient teacher and coordinator.

I am grateful for the generosity of the study participants, whose contributions have made this thesis and experience so rewarding. My thanks go to the faculty staff, both administrative and academic, and students, both postgraduate and graduate, at the School of Computing and Mathematical Sciences and the Software Engineering Research Laboratory at the Auckland University of Technology for providing a wonderful community in which to learn and conduct research.

Finally, I would like to add personal thanks to my wife Chi Leng for her love and to my parents for their endless encouragement and support.

This thesis is to my lovely three children; Gayle, Meg and Shaun, and I wish them a happy and healthy future.

# Abstract

Contextual information retrieval (CIR) is a critical technique for today's search engines in terms of facilitating queries and returning relevant information. Despite its importance, little progress has been made in its application, due to the difficulty of capturing and representing contextual information about users. This thesis details the development and evaluation of the contextual SERL search, designed to tackle some of the challenges associated with CIR from the World Wide Web. The contextual SERL search utilises a rich contextual model that exploits implicit and explicit data to modify queries to more accurately reflect the user's interests as well as to continually build the user's contextual profile and a shared contextual knowledge base. These profiles are used to filter results from a standard search engine to improve the relevance of the pages displayed to the user.

The contextual SERL search has been tested in an observational study that has captured both qualitative and quantitative data about the ability of the framework to improve the user's web search experience. A total of 30 subjects, with different levels of search experience, participated in the observational study experiment. The results demonstrate that when the contextual profile and the shared contextual knowledge base are used, the contextual SERL search improves search effectiveness, efficiency and subjective satisfaction. The effectiveness improves as subjects have actually entered fewer queries to reach the target information in comparison to the contemporary search engine. In the case of a particularly complex search task, the efficiency improves as subjects have browsed fewer hits, visited fewer URLs, made fewer clicks and have taken less time to reach the target information when compared to the contemporary search engine. Finally, subjects have expressed a higher degree of satisfaction on the quality of contextual support when using the shared contextual knowledge base in comparison to using their contextual profile. These results suggest that integration of a user's contextual factors and information seeking behaviours are very important for successful development of the CIR framework. It is believed that this framework and other similar projects will help provide the basis for the next generation of contextual information retrieval from the Web.

# Table of Contents

**Acknowledgements** ........ **i**

**Abstract** ........ **ii**

**Table of Contents** ........ **iii**

**Table of Figures** ........ **v**

**Table of Tables** ........ **vii**

**1 Introduction and Outline** ........ **8**
1.1 Introduction ........ 8
1.2 Research Objectives ........ 10
1.3 Overview of CIR Framework and Contributions ........ 12
1.4 Thesis Organisation ........ 13
1.5 Chapter Summary ........ 14

**2 Background and Motivation** ........ **15**
2.1 Introduction ........ 15
2.2 Web Search Engines ........ 15
2.3 Information Seeking Behaviour and Context ........ 21
2.4 Contextual Information Retrieval ........ 24
2.4.1 User Profile Modelling ........ 25
2.4.2 Query Expansion ........ 29
2.4.3 Relevance Feedback ........ 34
2.5 Evaluation ........ 39
2.6 Chapter Summary ........ 41

**3 SERL Search Architecture** ........ **43**
3.1 Introduction ........ 43
3.2 Development Methodology ........ 45
3.3 Architecture Pattern ........ 48
3.4 Architectural Building Blocks ........ 49
3.5 Design and Implementation ........ 56
3.5.1 Presentation Layer ........ 57
3.5.2 Application Layer ........ 64
3.5.3 Database Layer ........ 82
3.6 Chapter Summary ........ 83

**4 Experimental Methodology** ........ **85**
4.1 Introduction ........ 85
4.2 Experimental Phases ........ 86
4.2.1 Component Test (CT) ........ 87
4.2.2 Observational Study (OS) ........ 90
4.3 Subjects ........ 95
4.3.1 Recruitment Process ........ 95
4.4 Tasks ........ 97
4.5 Questionnaires ........ 98
4.5.1 Likert Scales ........ 99
4.5.2 Semantic Differentials ........ 100
4.5.3 Open-ended Questions ........ 101
4.6 Observational Techniques ........ 102
4.6.1 System Log ........ 103
4.6.2 Screen and Audio Recording ........ 103
4.6.3 Think aloud ........ 104
4.7 Experimental Setup and Equipment ........ 104
4.8 Sub-hypotheses ........ 105
4.9 Chapter Summary ........ 108

**5 Experimental Results and Analysis** ........ **109**

5.1 Introduction .......... 109
5.2 Anonymous Questionnaires .......... 110
5.3 Component Test (CT) .......... 112
5.4 Observational Study (OS) .......... 113
5.4.1 Subjects' Characteristics and Search Experience .......... 116
5.4.2 Overall Impressions .......... 118
5.4.3 Find Information Readily (Hypothesis 1) .......... 121
5.4.4 Adaptiveness Support (Hypothesis 2) .......... 127
5.4.5 Recommendation Support (Hypothesis 3) .......... 132
5.4.6 Query Formulation (Hypothesis 4) .......... 137
5.4.7 Interface Support (Hypothesis 5) .......... 139
5.4.8 Search Behaviour .......... 142
5.5 Chapter Summary .......... 143

**6 Discussion & Research Summary .......... 145**
6.1 Introduction .......... 145
6.2 Results Summary .......... 145
6.2.1 Effectiveness and Efficiency .......... 145
6.2.2 Satisfaction .......... 149
6.3 Research Summary .......... 153
6.4 Chapter Summary .......... 156

**7 Future Work & Conclusions .......... 157**
7.1 Introduction .......... 157
7.2 Limitations of the Study .......... 158
7.3 Future Work .......... 159
7.3.1 User Profiling .......... 159
7.3.2 Query Formulation .......... 161
7.3.3 Recommendation and Relevance feedback .......... 162
7.4 Conclusion .......... 163

**8 Appendices .......... 165**
8.1 Appendix A: Experimental Documents .......... 166
8.2 Appendix B: Component Testing .......... 192
8.3 Appendix C: Anonymous Questionnaires Analysis .......... 198
8.4 Appendix D: Normality Test. .......... 209
8.5 Appendix E: Descriptive Statistics and Inferential Analysis .......... 214

**9 References .......... 230**

# Table of Figures

Figure 2.1: Evolution of information retrieval. 16
Figure 2.2: A nested model of the information seeking and information searching research areas, reproduced from Wilson (1999b). 21
Figure 2.3: General model of information seeking and retrieval, reproduced from Järvelin and Ingwersen (2004). 23
Figure 2.4: Classification of behaviours that can be used for implicit feedback from Oard and Kim (2001). 36
Figure 3.1: SDRM methodology and DS research guidelines for CIR. 46
Figure 3.2: Three-tier Architecture. 48
Figure 3.3: Contextual SERL search three-tier architecture. 53
Figure 3.4: UI hierarchy of the contextual SERL search. 58
Figure 3.5: Progress indicator for the contextual SERL search. 59
Figure 3.6: 'No penalty error message' for the contextual SERL search. 59
Figure 3.7: Search results using the contextual SERL search. 59
Figure 3.8: Contextual SERL search home page. 60
Figure 3.9: User registration for the contextual SERL search. 60
Figure 3.10: Search login for the contextual SERL search. 61
Figure 3.11: Contextual SERL search UI. 61
Figure 3.12: Assistance options for the contextual SERL search. 62
Figure 3.13: Updating user profiles in the contextual SERL search. 64
Figure 3.14: Viewing visited URLs in the contextual SERL search. 64
Figure 3.15: Profile collector module functionality. 66
Figure 3.16: Preference collector component functionality. 69
Figure 3.17: Behaviour acquisition functionality. 74
Figure 3.18: Context manager module functionality. 76
Figure 3.19: Knowledge collector functionality. 77
Figure 3.20: Knowledge base query formulation functionality. 78
Figure 3.21: Result analyser process. 80
Figure 3.22: Physical data diagram of the contextual SERL search. 82
Figure 4.1: Overview of component test procedure. 89
Figure 4.2: Overview of observation study procedure. 92
Figure 4.3: Contextual SERL search system log. 103
Figure 4.4: Experimental environment. 105
Figure 5.1: Bar charts of overall reactions regard to the completion of search tasks. 119
Figure 5.2: Bar charts of overall reactions regard to the UI. 120
Figure 5.3: Bar chart for system remembers last search actions. 128
Figure 5.4: Bar chart of system uses the user's last actions to recommend terms. 129
Figure 5.5: Radar chart of system uses the user's last actions to recommend terms. 130
Figure 5.6: The bar chart of system learns the user's interest over the time. 131

Figure 5.7: Bar charts of system uses other users/searchers search actions to recommend terms. ..... 132
Figure 5.8: Bar charts of how well the system communicates. ..... 133
Figure 5.9: Radar chart of how well the system communicates recommendations. ..... 134
Figure 5.10: Bar charts of how well system conveys preferences. ..... 135
Figure 5.11: Radar chart of how well system conveys preferences. ..... 136
Figure 5.12: Bar charts of query formulation strategy. ..... 137
Figure 5.13: Bar charts of trusting the system to choose the search terms. ..... 138
Figure 5.14: Bar charts of relevance of content. ..... 140
Figure 5.15: Radar chart of relevance of content. ..... 140
Figure 5.16: Bar charts of interface guide. ..... 141
Figure 6.1: Comparative median analysis of subjects' satisfaction. ..... 150

# Table of Tables

Table 2.1: Summary of user profile modelling approaches. ..... 26
Table 2.2: Summary of query expansion approaches. ..... 32
Table 3.1: Three-tier architecture benefits. ..... 49
Table 3.2: Decision making factors in choosing three-tier architecture. ..... 56
Table 3.3: Recommendation process. ..... 70
Table 3.4: WSD disambiguation process. ..... 72
Table 3.5: Example of user's preference data. ..... 73
Table 3.6: Keywords extraction process. ..... 75
Table 3.7: Example of user's behaviour data. ..... 75
Table 3.8: Analysing & ranking process. ..... 80
Table 3.9: Similar search queries and associated visited URLs extraction process. ..... 80
Table 3.10: Configurations options for the contextual SERL search. ..... 81
Table 4.1: Number of subject involvement in other studies. ..... 95
Table 4.2: Number of subject involvement in Phase II experiments. ..... 97
Table 4.3: Number of search tasks used in other studies. ..... 97
Table 4.4: Example of simulated work task situations taken from Borlund (2000). ..... 98
Table 4.5: Usage of the observation questionnaires in various phases. ..... 99
Table 4.6: Example Likert scale. ..... 100
Table 4.7: Example set of semantic differentials. ..... 101
Table 4.8: Example of open-ended questions. ..... 102
Table 5.1: Precision summary results of component test. ..... 112
Table 5.2: Summary of search experience and satisfaction of subjects. ..... 118
Table 5.3: Evidence to support experimental hypotheses. ..... 143
Table 6.1: Statistical analysis of Hypothesis 1 for all six search tasks. ..... 146
Table 6.2: Statistical analysis of Hypothesis 1 for search task six. ..... 146
Table 6.3: Summary of various challenges related to the CIR framework. ..... 154

# Chapter 1

# Introduction and Outline

## 1.1 Introduction

The Internet is, in its simplest terms, a huge, searchable database of information reached via a computer (McQuistan, 2000). The Internet makes an enormous amount of information available on the Web leading to *information overload,* the challenge for users then being one of finding *relevant* information (Fan, Gordon, & Pathak, 2004). Information overload on the Internet is not a new phenomenon (Allen & Wilson, 2003), in fact, it is a well recognised problem (Carlson, 2003).

Digital technologies, such as search engines that search the Web (e.g., Google, Yahoo, Lycos, Excite, Infoseek, Alta Vista, and Hotbot) are attempting to deal with the challenges of Internet information overload (Berghel, 1997). Though these search engines have evolved through several generations since their inception in 1994 (Finkelstein et al., 2002), as useful as they are, they are far from perfect. In fact, these search engines are faced with a number of difficult challenges in maintaining and enhancing the quality of their performance (Henzinger, Motwani, & Silverstein, 2002). These challenges are considered in detail in Chapter 2.

In order to address some of these challenges it has been suggested that search engines must leverage the user's contextual information, such as their behaviour and their preferences, especially efforts to understand the underlying intent of the users (Jansen, Booth, & Spink, 2007). As a result, research in contextual information retrieval (CIR) approaches has recently become prominent in the Web information retrieval (Web IR) field and has attracted a significant amount of interest from researchers around the world. However, CIR has distinct characteristics when compared to either general IR or non-contextual information retrieval from the Web. Thus, CIR has been and remains one of the major long-term challenges in IR generally (Allan et al., 2003) due to the difficulty of capturing and representing contextual information about users. As a result, until this point, there has been relatively little published research that explores the usefulness of CIR approaches. There is a compelling need for research in Web IR to provide users with more relevant information more efficiently and in different ways.

This thesis introduces an alternative novel framework for CIR from the World Wide Web (WWW or simply Web) in the context of the shortcomings of existing search engine technology. The framework makes use of three related research themes: user profile modelling, query expansion, and relevance feedback. It utilises a rich contextual user profiling model that exploits a user's implicit and explicit data. Each user's implicit data are gathered from their information seeking behaviour, *such as entered search queries, visited URLs* and *extracted Meta keywords from those visited URLs*. The user's explicit data, *such as alternative term/phrases, Meta keywords or similar phrases, ontology* and *concepts,* are captured from the lexical database, a shared contextual knowledge base and domain-specific ontology (Gruber, 1993)/concepts using a data mining technique and a relevance feedback technique. This data is later used by the framework to modify queries using the Boolean query expansion technique to more accurately reflect the user's interests as well as to build the user's contextual profile and a shared contextual knowledge base. Finally, the approach retrieves personalised or contextual search results from a standard search engine using the modified query.

This study is an interdisciplinary research endeavour incorporating aspects of information science, cognitive science, information retrieval, human-computer interaction, Web mining and machine learning research. In addition, the construction of an alternative CIR framework from the Web seeks to improve the efficiency and effectiveness of search activities, define new ideas, and technical capabilities. As such, this study employs the system development research methodology (Nunamaker, Chen, & Purdin, 1991) and the design-science research guidelines (Hevner, March, Park, & Ram, 2004) to investigate, design, develop and implement a Software Engineering Research Lab Contextual Search system (or simply contextual SERL search) as an alternative CIR framework from the Web.

Most of the notations used and issues raised in this section are addressed in more detail in later Chapters. The remainder of this Chapter sets out the thesis research objectives, describes the contribution of the work to the body of CIR research, and presents an outline of this thesis.

## 1.2 Research Objectives

The three main objectives addressed in this research are:

i. **Research Objective I**: To construct a CIR model that captures both a user's behaviour and preferences as a user's personal contextual profile, and structure this information in such a way as to be able to define a search context that can be refined over time.

ii. **Research Objective II**: To facilitate the collection of multiple users' personal contextual profiles to create a shared understanding of contextual knowledge base.

iii. **Research Objective III**: To use the user's personal contextual profile or together with the shared contextual knowledge base to refine search queries, filter returned results from search engines, and provide user recommendations/suggestions.

These research objectives relate to various CIR fields, especially user profile modelling, query expansion, and relevance feedback. The first and second objectives are related to the user profile modelling, i.e., modelling a user's information seeking behaviour into the Web IR system. The goal is to understand and capture the user's potential *multiple* search interests by monitoring their explicit (i.e., explicit rankings, inputs, and instructions) and implicit (i.e., browsing and typing) data. The third objective is related to the interactive query expansion and relevance feedback techniques. The goal is to provide the user with an expanded set of terms related to their search interests in order to formulate a Boolean search query to improve search results.

This study was both exploratory and confirmatory in nature and so employed a mixed method research approach (Tashakkori & Teddlie, 2003) to evaluate the contextual SERL search. Two evaluation experiments, namely component testing and an observational study, were carried out. The component test experiment was conducted to ensure that each system component of the contextual SERL search performed correctly under all of the conditions that the system could encounter in use and during the observational study experiment. A total of 30 subjects, with different levels of search experience, participated in the observational study experiment. The observational study experiment was undertaken to evaluate the performance of the contextual SERL search and to determine if it meets a range of expectations reflecting overall performance. Five hypotheses were constructed to test the performance of the contextual SERL search along the usability dimensions of effectiveness, efficiency and subjective satisfaction in comparison to the performance of a contemporary search engine. Effectiveness and efficiency were measured by the number of queries, number of clicks, number of hits, number of URLs and length of time to reach target information in each user experiment. Satisfaction was measured from post observation questionnaires responses. The five hypotheses are:

i. Find Information Readily (Hypothesis 1)

   *The contextual SERL search enables subjects to find relevant information more readily than a standard search engine using their personal profile and shared contextual knowledge base.*

ii. Adaptiveness Support (Hypothesis 2)

   *The contextual SERL search adapts to the information needs of the searcher and facilitates effective recommendation of terms.*

iii. Recommendation Support (Hypothesis 3)

*The contextual SERL search eases the conveying preferences process and recommends relevant and useful terms.*

iv. Query Formulation (Hypothesis 4)

*The contextual SERL search facilitates easy, effective and reliable query formulation strategy.*

v. Interface Support (Hypothesis 5)

*The interface support provided by the contextual SERL search facilitates effective information access.*

## 1.3 Overview of CIR Framework and Contributions

This research is application-based, and many of the findings come directly from the implementation and evaluation of the working system, i.e., contextual SERL search. The contextual SERL search performs a number of activities, such as adaptation of a user's information seeking behaviour, recognition of a user's preferences and interests, recommendation of terms, generation of Boolean query and presentation of ranked contextual search results to improve Web IR. Several architectural design and implementation issues, such as scalability, flexibility, performance and robustness, are maximised during the system's development.

The contextual SERL search comprises two main modules: Profile Collector Module (PCM) and Context Manager Module (CMM) to perform the following functions:

a. Gather the user's implicit data, such as entered search queries, visited URLs and corresponding extracted Meta keywords.

b. Capture the user's explicit data, such as alternative terms, Meta keywords or similar phrases and concepts. This data is sourced from a lexical database, a shared contextual knowledge base and domain-specific ontologies.

c. Construct the user's personal contextual profile and a shared contextual knowledge base using data from step 1 and step 2.

d. Modify the user's initial query to more accurately reflect the user's interests.

Each module consists of several components that perform these various functions, with the PCM components providing the core data collection functionality and the CMM components enabling the querying, filtering and recommendation.

The result of this implemented contextual SERL search as the CIR framework from the Web makes two main contributions. First, the framework experimentally demonstrates the construction of an *evolving contextual user profile* and the *shared contextual knowledge base* to define the user's adaptive or dynamic *multiple search contexts,* which can be refined over the time. Second, the CIR framework experimentally demonstrates the *recommendation of alternative terms/concepts* and *formulation of a dynamic Boolean search query* using the user's contextual profile and the shared contextual knowledge base by employing an adaptive data mining technique and relevance feedback approach.

## 1.4 Thesis Organisation

This thesis is structured into seven Chapters followed by references and appendices:

Chapter 1 has presented the research problem, research objectives and hypothesis, justified the need for an alternative CIR framework for the Web, and outlined the main contributions arising from the work undertaken.

Chapter 2 provides the essential background and context for this thesis and provides a fuller justification for the research work described in this thesis.

Chapter 3 describes the design and implementation details of the contextual SERL search in greater detail.

Chapter 4 provides the details of the methodology employed and outlines the hypotheses for the research presented in this thesis.

Chapter 5 describes the summarised results of the anonymous questionnaires and the results of the two experimental phases; component testing and observational study.

Chapter 6 presents the implications of results presented in Chapter 5 and the summary of this research study.

Chapter 7 presents future further research avenues and conclusions based on the contributions made by this thesis.

## 1.5 Chapter Summary

This Chapter has laid the foundations for this thesis. It briefly introduced the research problem, research objectives, research hypothesis, and an alternative CIR framework. A justification for the research problem is outlined, together with an explanation of the research methodology used. The potential contributions of this research to the CIR body of knowledge are outlined. The thesis can therefore proceed on these foundations. The next Chapter examines the CIR related literature relevant to this research problem.

# Chapter 2

# Background and Motivation

## 2.1 Introduction

This Chapter provides the motivation for the research described in this thesis and creates a context within which the work is situated. It begins by discussing existing Web search engines and challenges associated with finding information on the Web. Later sections discuss information seeking behaviour, alternative CIR techniques such as, user profile modelling, query formulation, relevance feedback and results presentation mechanisms designed to resolve these challenges. This chapter also provides the discussion on evaluation of such techniques.

## 2.2 Web Search Engines

Mankind has organised information for hundreds of years in order to make it more accessible to others. With the advent of information technology (IT), the process (or techniques) of information retrieval (IR) - organising and retrieving information - has evolved drastically. The term IR is a wide, often loosely defined term - in brief, it involves finding some desired information in a store of information or a database (Meadow, Boyce, & Kraft, 2000).

Figure 2.1 illustrates the evolution of IR, where Internet/Web IR is a sub-discipline within the general IR arena. Internet/Web IR uses either search engines, which index a portion of documents as a full-text database, or web directories, which classify selected Web documents by subject, to facilitate the identification of relevant information (Yates & Neto, 1999).

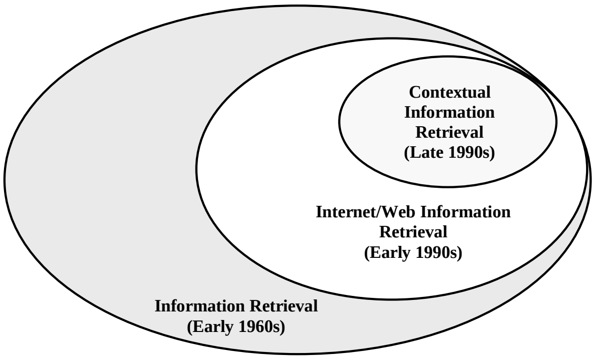


**Figure 2.1: Evolution of information retrieval.**

Due to the exponential growth of the Web since its inception in 1993, identifying user information needs has been highlighted as one of the most fundamental challenges in the development of Web search engines (Huang, Chien, & Oyang, 2003). In the late 1990s, CIR was introduced to address these challenges (section 1.3). However, it has distinct characteristics when compared to either general IR or non-contextual information retrieval from the web. The following paragraphs provide a brief literature review of prior research related to Web search engines in order to identify opportunities that are evident in terms of improving query results from the Web.

Since its inception, the World Wide Web, or simply the Web (Berners-Lee, Cailliau, Luotonen, Nielsen, & Secret, 1994) has continuously grown into one of the largest collections of content in existence. Recent papers (Kobayashi & Takeda, 2000; Levene, 2006) reported on the growth of the Web, which is continuing to expand at exponential rates on a daily basis. In 2004 it was reported that each day approximately 60 terabytes of new content was added to the Web's 10 billion or so indexed pages (Roush, 2004). Given these numbers, it is clear that the complexity of finding relevant information in the Web increases day by day. In fact, "information overload" on the Web is a well recognised problem (Carlson, 2003).

Right from the early days of the Web, search engines have been an indispensable tool for users to find *relevant* information (Jansen et al., 2007; Ruthven, 2003). Until the mid 1990s, search engines were not readily available for the general public; even so, within a few short years, they made themselves part of people's daily lives (Hsieh-Yee, 2001; Sherman & Price, 2001). At present, search engines are one of the most commonly used resources for finding relevant information on the Web (Henzinger, 2004; Morris, Morris, & Venolia, 2008; Sahami, Mittal, Baluja, & Rowley, 2004; Toms & Bartlett, 2001).

Over the years, various search engines have been developed. Upstill, Craswell and Hawkings (2003) presented a detailed development history of early Web search engines. Similarly, Pokorny (2004) reported an overview of the various search-engine architectures, including traditional (or centralised), metasearch, and distributed search. Current Web search engines – such as Google[1], Yahoo[2], Microsoft Live[3] (*formerly MSN Search*), Ask[4] (*formerly Ask Jeeves*), Alta Vista[5], Lycos[6] and so on – are attempting to deal with Internet "information overload" challenges (Berghel, 1997; Chau, Zeng, & Chen, 2001 ; Sahami et al., 2004). To this end, these Web search engines incorporate many features, such as related searches, clustering, find similar, search within, search by language, sort by date, advanced search pages, help pages and so on meant to assist finding the most relevant information (Limbu, Connor, & MacDonell, 2005). These search engines have evolved through several generations since their inception in 1994, and the quality of search has improved dramatically in that time (Finkelstein et al., 2002). However, as useful as they are, they are far from perfect. In fact, these search engines are faced with a number of difficult challenges in maintaining and enhancing the quality of their performance (Henzinger et al., 2002).

---

[1] http://www.google.com
[2] http://www.yahoo.com
[3] http://www.live.com
[4] http://www.ask.com
[5] http://www.altavista.com
[6] http://www.lycos.com

A comprehensive discussion of Web search engine challenges is outside the scope of this thesis. These challenges include effective ranking and optimising algorithms, effective taxonomies, structured quality content, multiple language support, effective search in non-text corpora, and so forth. Rather than addressing these challenges, this thesis confines its goal to study and examine the role of the user, their information seeking behaviour and defining their context as a means to overcome the challenges associated with Web searching. The central focus of this thesis is an adaptation of the user's information seeking behaviour and their context to find relevant information more readily than at present.

Most contemporary search engines are fronted by a simple and yet functional User Interface (UI). The interface provides little support for a user to interact during the information seeking process. For example, a common search engine's UI consists of a single text field – for 20–30 characters – into which search terms can be entered, and a "search button" that when selected, begins the searching process (Rieh & Xie, 2006). Further, the textual content and list based representation of search results makes them difficult to evaluate (Hoeber & Yang, 2006a). According to Toms and Bartlett (2001), the Web search UI relied on an interface design technique - 'form filling' and to date little research has addressed the Web search UI problem. Shneiderman *et al.* (1998) stated that some progress on search UI had been made; however much more remains to be done. They also stated that finding common ground for search interfaces is difficult, but not finding it would be tragic.

Today's search engines are built to serve all users in general, utilising a "one size fits all" approach. They do not provide desired search results to any specific individual user (Allan et al., 2003; Liu, Yu, & Meng, 2004; Zhang, Xu, & Yang, 2001). These search engine's results are largely dependent on the user specified/formulated search query (Taksa, 2005). Research has confirmed that the queries submitted to search engines by users are short (Fonseca, Golgher, Pôssas, Ribeiro-Neto, & Ziviani, 2005; Leake & Scherle, 2001; Sieg, Mobasher, Lytinen, & Burke, 2004). Queries are mostly limited to fewer than three key words (Jansen, Spink, & Saracevic, 2000; Spink, Ozmutlu, Ozmutlu, & Jansen,

2002) and can be vague with little or no context information associated with these queries (Fonseca et al., 2005).

Additionally, search engine’s results are typically based on simple keyword matches without any concern for the information needs of the user at a particular instance in time (Challam, 2004; Keenoy & Levene, 2005) or in a particular context. For example, if a user submits a keyword (e.g., *“surfing”* as a query) to search for information from the Web, the search engine searches through the indexed Web pages, filters and returns a list of those documents that contain the specified keyword (i.e., *surfing*). However, the keyword “surfing“, can have completely different meanings – such as Internet surfing, beach surfing, surfing lesson, surfing shop and so on – depending on the context in which it is used. As a result, the user still must perform most of the relevance filtering to identify useful search results (Aridor, Carmel, Lempel, Soffer, & Maarek, 2000; Tirri, 2003). These problems occur due to synonymy and polysemy of keywords, leading to potential ambiguity. Synonymy is when several different terms have the same meaning and polysemy is when a single word has more than one meaning (Deerwester, T., Furnas, Landauer, & Harshman, 1990). The user can include additional search terms that could help to refine the search queries, but it is difficult for even experienced users to select the optimum query terms so that the desired subset of information is retrieved (Leake & Scherle, 2001). Some search engines, such as Google or Yahoo provide hierarchy of categories to help the users to define their search intent. Unfortunately, users are either too impatient to browse through the hierarchy of categories or they may have difficulties in finding the proper paths leading to suitable categories (Liu et al., 2004). As a result, even the most experienced users find it difficult to find relevant information from the Web (Hölscher & Strube, 2000; O'Hanlon, 1999; Sieg et al., 2004).

The need to better target a search on the information that will satisfy a user’s information needs is well recognised (Leake & Scherle, 2001). A critical goal of successful IR on the Web is to identify which pages are of most relevance to a user’s query (Sahami et al., 2004). On the other hand, relevance is typically person-dependent, so personalisation is critical to the ongoing development of search engines (Tirri, 2003). In this regard today’s search engines are lacking a personalisation mechanism (Levene, 2006; Shen, Tan, & Zhai, 2005) and the capability to ‘understand’ the search query in terms of the information needs of a user at a particular instance in time. This limits their ability to return customised results (Challam, 2004). Jansen *et al.* (2007) suggested that search engines must leverage the user’s contextual information, such as user’s behaviour and their preferences, especially efforts to understand the underlying intent of the users.

In summary, today’s search engines are faced with a number of difficult challenges, related to the user’s information-seeking behaviour. This has several dimensions in terms of query formulation, the user’s understanding of the task, and the system’s “understanding” of how the user performs that task (Toms & Bartlett, 2001). Hence, to provide the desired information to the user requires effective methods for identifying the user’s task context (Bauer & Leake, 2003) and using this information in the search engine to query, filter and return relevant information. In addition, in order for search engines to continue to improve, they must leverage an increased knowledge of a user’s behaviour (Jansen et al., 2007), especially in the respect of “understanding” the underlying intent of the search. Thus, further detailed research in the following areas is required to address the challenges that have been discussed.

a. Understanding and bridging the semantic gap between what a user wants and what he/she gets. For example, the user wants ‘java’ as in ‘java Island Indonesia’, but gets ‘java programming’.

b. Understanding and modelling the user’s information seeking behaviour and their preferences in the Web search.

While a number of contextual factors can influence a user’s information-seeking behaviour, a general review of these is outside the scope of this thesis. In the next section, some of the most important contextual factors in relation to this thesis are discussed.

## 2.3 Information Seeking Behaviour and Context

Information seeking behaviour has been a traditional focus of library and information science studies. More recent work has focused attention on information seeking on the Web in our day-to-day life. Figure 2.2 shows the nested model of the information seeking and information searching research areas. According to Wilson (2000), information-seeking behaviour is the "purposive seeking for information as a consequence of a need to satisfy some goal". Wilson (2000) further advocated that in the course of seeking, a user may interact with traditional information systems *(such as a newspaper or a library),* or with computer-based systems (*such as Web search tools).* Similarly, Case (2002) defined information seeking as a conscious effort to acquire information in response to a need or gap in user's knowledge. Both, Wilson (2000) and Case (2002) suggested information seeking behaviour is a part of information behaviour (Case, 2002; Wilson, 1997). Additionally, Wilson (1999b) suggested information search behaviour is "the 'micro-level' of behaviour employed by the searcher in interacting with information systems of all kinds", which is one aspect of information seeking behaviour.

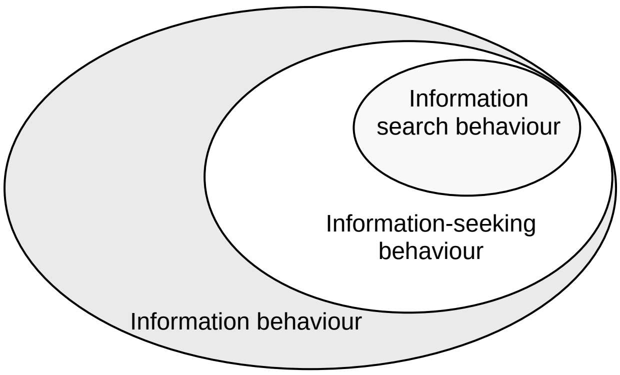


**Figure 2.2: A nested model of the information seeking and information searching research areas, reproduced from Wilson (1999b).**

Information behaviour research has burgeoned over the past few decades. The highlights of research on information behaviour include (but are by no means limited to) Ellis's (1989) behavioural model of information searching strategies, Kuhlthau's (1991) information search process, Wilson's (1997) problem-solving model, Dervin's (1999) sense-making and Hjørland's (2002) domain analysis approach. Wilson (1999b) reviewed most of these models and discovered that these models complement each other instead of compete. Similarly, Järvelin and Wilson (2003), reviewed some of these models more recently and suggested that without detailed analysis of their components, such models provide little or no suggestion of causative factors in information seeking and retrieval (IS&R) phenomena and, consequently, they do not directly suggest hypotheses to be tested. However, these theoretical models have contributed greatly to the understanding of the role of “context” in general on information seeking and searching.

The role of context has been recognised in key models of information behaviour (Dervin, 1999; Wilson, 2000) as well as in interactive information retrieval (Ingwersen, 1996; Järvelin & Ingwersen, 2004; Spink, 1997). As a matter of fact, context has been a widely researched topic not only in information science but also in numerous other disciplines (Talja, Keso, & Pietiläinen, 1999). Large numbers of studies have focused on information seeking in context over the past 20 years (Vakkari, Savolainen, & Dervin, 1997). In the course of these studies (Case, 2002), a great number of contextual factors that play a role in information behaviour have been identified.

Gaslikova (1999) described context of information seeking as a function of many different parameters, such as the time and place of appearance of information need, the time for information seeking, types of participants of the seeking process, for example, their demographic, social, professional, educational and behavioural characteristics, the purpose of information seeking, the processes and situations of information seeking, and many others. Similarly, Kari and Savolainen (2003) reviewed several earlier empirical internet searching studies, which took into account some contextual factors, and they discovered that research of this kind has typically been limited to information seeking, a task, a situation and/or the person, that is, to fairly immediate surroundings of the individual’s Web interaction. More recently, Ingwersen and Belkin (2004) suggested IR in context means one of two things; a) IR that is context-aware, or able to elicit the implicit knowledge of the participatory searcher environment related to the context into an explicit representation and b) the specification models for context and its features that can be used to focus an IR task according to such a context. Figure 2.3 presents a cognitive framework that brings together the field of information seeking and information retrieval, proposed by Ingwersen and Järvelin (2004).

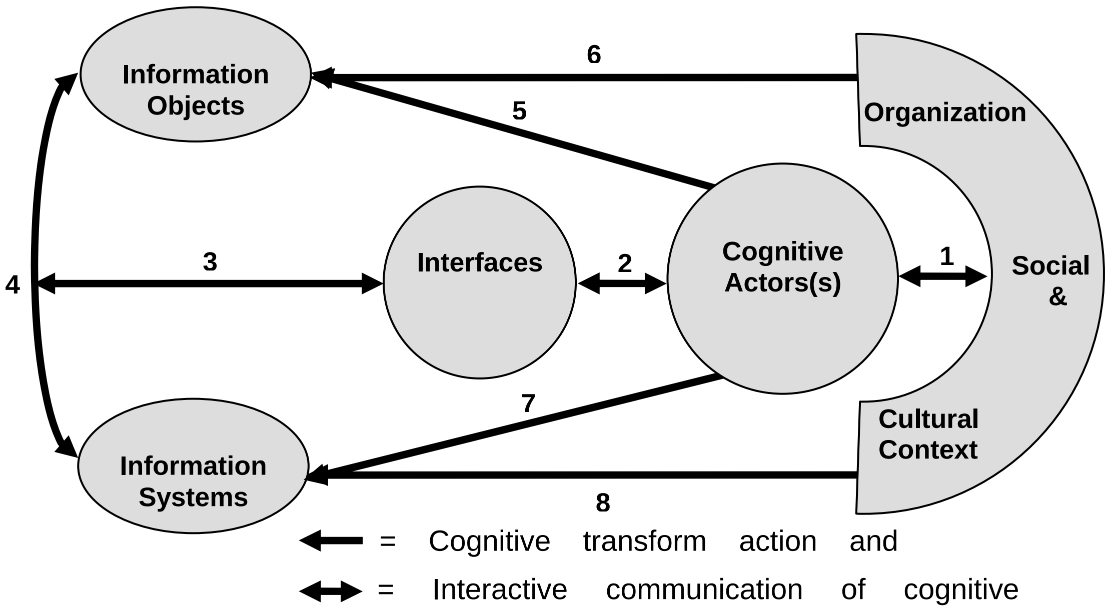


**Figure 2.3: General model of information seeking and retrieval, reproduced from Järvelin and Ingwersen (2004).**

The cognitive framework mirrors cognitive actors, such as information seekers, sitting within several kinds of contexts; social, organizational and cultural affiliations, and interacting with information objects, information systems through interfaces. The framework provides a common 'language' for researchers in this field and made clear the variables and issues that were being discussed (Butterworth & Perkins, 2006).

Despite context being central to most theoretical approaches to information seeking, little has been written about the context at a meaningful level (Johnson, 2003). This may be due to various reasons, such as there is no commonly agreed definition for the notion of "context" (Rey, Coutaz, & Crowley, 2004). The notion of "context" is widely interpreted, or possibly abused, in IR (Finkelstein et al., 2002) and researchers are rediscovering the notion of "context" to serve distinct purposes from different perspectives (Kim, Suh, & Keedong, 2007; Prekop & Burnett, 2003; Rey et al., 2004; Schilit, Adams, & Want, 1994). However, the potential use of context has certainly been under explored, even though it is already present in established Web IR techniques, such as relevance feedback (Jones & Brown, 2004).

Therefore, exploring contextual factors and modelling information seeking behaviour in interactive Web IR is imperative. In this thesis, information seeking in context refers to a user's explicit and implicit input with respect to the entered queries, clicked URLs, browsed sites, the domain, the search situation, and the task. The primary goal of incorporating such context is to increase the relevance of results, although other outcomes, such as effectiveness, efficiency, and subjective satisfaction, could potentially be affected as well.

## 2.4 Contextual Information Retrieval

One of the underlying goals of CIR is to acquire a user's information seeking behaviour, such as search activities and responses, and incorporate this knowledge into a search system. This is intended to deliver more effective, efficient and personalised interactions and a more appropriate retrieval strategy for each user through a system tailored to their preferences.

Allan *et al.* (2003) defined "contextual retrieval" as an information retrieval process that combines search technologies, knowledge about a query and user context into a single framework in order to provide the most appropriate answer for a user's information need. This study adopts this general definition of "contextual retrieval" into a conceptual framework for the study. Large numbers of Web IR systems or approaches have utilised the notion of "user context" to some extent, referring to a diverse range of ideas from recommendation systems, domain specific search engines, personalisation systems, personal assistants and so on. Each has the aim of highly ranking the results that are most relevant to the user. The features that distinguish these approaches are the kinds of information about the user's context that are used, the level of interaction with the user *(explicit or implicit collection of data),* where the information is stored *(client-side or server-side),* the algorithm used to incorporate the information about the user into the search and how information is presented to the user.

Despite its growing importance and the development of all these approaches, there remains no comprehensive model to describe the contextual retrieval or CIR process (Wen, Lao, & Ma, 2004) due to the difficulty of capturing and representing knowledge about users, tasks, and context in a general Web search environment (Allan et al., 2003). In effect, CIR has been and remains one of the major long-term challenges in IR generally (Allan et al., 2003). Previously, researchers working in the area of CIR have focused their efforts on three themes of direct relevance to this thesis: user profile modelling, query expansion, and relevance feedback. These themes have some characteristics in common, but also many differences. Therefore, it is pertinent to present these themes in relation to CIR in the following sections without detailing the algorithms or making any extensive comparison between them.

### 2.4.1 User Profile Modelling

User modelling (UM) is a mature field and has a history in computing date back well over 20 years (Anderson, 1984). The UM technique has been recognised as an important and useful feature in IR systems to estimate a user's true intention and demand (Motomura, Yoshida, & Fujimoto, 2000). One of the underlying goals of these techniques is to gather user-specific information, such as a user's context, during interaction with the user and to then predict their responses, thereby creating a tailored system that can provide more effective, efficient and personalised interactions (Durrani, 1997; Kelly, 2004). For example, content personalisation in interactive Web IR allows for more efficient tailoring of content to users by taking into account predefined roles or proper user profiles. Several user modelling techniques have been reported in many active research fields, such as, dialog systems (Kobsa & Wahlster, 1989), and information retrieval systems (Allan, 1996; Joachims, Freitag, & Mitchell, 1997; Liu et al., 2004; Pazzani, Muramatsu, & Billsus, 1996; Pitkow, Schutze, Cass, Cooley, & et al., 2002; Rhodes & Starner, 1996). Brusilovsky (2001) provided an overview of the user modelling history and development.

Monitoring and capturing a user's search activity over time, and to infer their interests and preferences as their behaviour changes over time is a great challenge. Many Web IR systems have explored various user modelling approaches to improve the personalisation of the Web experience for each user with respect to the user's context (Kraft, Maghoul, & Chang, 2005; Shen et al., 2005; Teevan, Dumais, & Horvitz, 2005). Table 2.1 presents a summary of the main user profile modelling approaches to date.

**Table 2.1: Summary of user profile modelling approaches.**

| Study | Context(s) | |
|---|---|---|
| | User's Behaviour | User's Preferences |
| Letizia (Lieberman, 1995) | ✓ | |
| Remembrance Agent (Rhodes & Starner, 1996) | ✓ | |
| Syskill & Webert (Pazzani et al., 1996) | | ✓ |
| WebWatcher (Joachims et al., 1997) | | ✓ |
| InfoFinder (Krulwich & Burkey, 1997) | | ✓ |
| ABIS (Amati, D'Aloisi, Giannini, & Ubaldini, 1997) | | ✓ |
| WebMate (Chen & Sycara, 1998) | | ✓ |
| Let's Browse (Lieberman, Dyke, & Vivacqua, 1999) | ✓ | |
| Watson (Budzik & Hammond, 1999) | ✓ | |
| SUITOR (Maglio, Barrett, Campbell, & Selker, 2000) | ✓ | |
| PRISM (Leake & Scherle, 2001) | ✓ | |
| WAIR (Zhang & Seo, 2001) | ✓ | |
| WordSieve (Bauer & Leake, 2001) | ✓ | |
| WAWA (Rad & Shavlik, 2003) | | ✓ |
| Web Personalisation (Hofgesang, 2007) | ✓ | |
| User Conceptual Index (Sendhilkumar & Geetha, 2008) | ✓ | |

A review of the user modelling approaches listed in Table 2.1 reveals that all of these approaches attempt to automatically filter web pages on behalf of the user, based on their monitored contextual user profile. Some of these approaches, such as Letizia (Lieberman, 1995), Remembrance Agent (Rhodes & Starner, 1996), Let's Browse (Lieberman et al., 1999), Watson (Budzik & Hammond, 1999), Suitor (Maglio et al., 2000), PRISM (Leake & Scherle, 2001), WAIR (Zhang & Seo, 2001), WordSieve (Bauer & Leake, 2001), Web Personalisation (Hofgesang, 2007) and User Conceptual Index (Sendhilkumar & Geetha, 2008) have focused on identifying and characterising user tasks and search strategies. These approaches monitor the user's behaviour to construct a contextual profile and use this information to return or recommend Web pages or documents which might be relevant to the user's search context. As such, these approaches can be resource-intensive as they require extensive task analysis and classification of users' search behaviours. In addition, these approaches may miss critical information relating to the dynamic nature of the information-seeking process as they extract information from only the current document that the user is viewing or browsing. Besides, these approaches only suggest those documents previously visited by the user and as such do not consider the dynamic nature of the Internet.

Other systems, such as Syskill & Webert (Pazzani et al., 1996), InfoFinder (Krulwich & Burkey, 1997), WebWatcher (Joachims et al., 1997), ABIS (Amati et al., 1997), WebMate (Chen & Sycara, 1998), and WAWA (Rad & Shavlik, 2003) have focused more heavily on the identification and representation of the user's topic, knowledge and document preferences. These approaches utilise user preferences using a relevance feedback (RF) technique (section 2.4.3.) to construct a contextual profile and use this information to construct search queries and dispatches to special purpose search engines (e.g., Lycos) to retrieve pages that might match the user's interest or are relevant to the user's search context. These approaches typically require users to spend time selecting and ranking documents or to specify keywords. For example, WebWatcher (Joachims et al., 1997) and WAWA (Rad & Shavlik, 2003), both require users to answer a series of questions (or instructions) about their general interests. Requiring users to identify keywords that characterise their information needs constrains users and their needs in a variety of ways. According to Belkin (2000), users are often unable to clearly articulate and describe their information needs.

All of the above-mentioned approaches utilise either user's behaviour or user preferences to construct a contextual profile. However, none of the approaches utilise both. None of above approaches accommodate a single user who has multiple, unrelated information needs. There is little discussion of how these multiple needs might be integrated into a single contextual user profile or of how the representation of these various different needs might affect retrieval. Furthermore, a user's information needs can change over time and few systems address this issue. None of above approaches discusses how to harmonise a user's dynamic interests and context to the content of information sources. Except for InfoFinder (Krulwich & Burkey, 1997), none of above approaches discuss Boolean query expansions (section 2.4.2.) using any form of user contextual profile. Finally, apart from WebMate (Chen & Sycara, 1998), they do not seem to have the capability of sharing users' contextual profile information with other users, thereby potentially leading to failure when the user requires access to novel information, or information outside their original context. The shared contextual profile or collaborative filtering (CF) is based on the collective profiles of a number of users and can be used to assist users with similar interests (Pennock & Horvitz, 1999). This approach has been used in personalised search by Sugiyama, Hatano, & Yoshikawa (2004) and Sun, Zeng, Liu, Lu, & Chen (2005). Despite their success, the CF-based search systems have two major limitations. Firstly, the CF techniques have poor performance (or “cold start” problem) when the amount of historical information for each user and their interests or information seeking data is limited. Secondly, as the number of users grows, the number of comparisons in many dimensions seriously influences the system scalability and performance.

The above discussion demonstrates that a number of user profile modelling approaches have been developed in the research community to assist users in different tasks. However, many of these approaches have faced and continue to face a number of difficult challenges, in particular;

a. How to acquire, maintain and represent accurate information about a user’s multiple interests with minimal intervention?

b. How to use this acquired information about the user to deliver personalised search results?

c. How to use the acquired information about various users to build a knowledge base for large communities or groups?

The above discussion also demonstrates that users usually have their own preferences that determine their search intents. Likewise, the users’ Web search behaviours may infer their search intents. It is very useful for systems to make use of such preferences and search behaviours information to define the user’s search interests as well as to build their contextual user profile. Besides, the user’s interests are likely to change dramatically in a short period of time. The system must be able to adapt accordingly. As such, an effective user profile modelling approach involves two important challenges: accurately identifying the user’s search context and organising the user’s search preferences and behaviours information in such a way that matches the particular context.

In this thesis a new user profiling modelling approach is introduced to address some of the challenges discussed above. The approach introduces three distinct techniques to acquire and maintain a user's contextual profile. First, the approach utilises the Lexical database (i.e., WorldNet), the domain specific Ontology (e.g., computer science and travel) and Meta keywords - which are extracted from browsed Websites - technique to capture the user's preferences using the relevance feedback technique. Second, the approach implements a contextual parameter called "search context" to define the user's search multiple search interests/intents as well as to facilitate the building of adaptive user profiles and the shared contextual knowledge base. Third, the approach implements a hybrid contextual user profiling technique to capture a user's adaptive search behaviour by monitoring and capturing their explicit and implicit data.

In the next section, query expansion, the individual's primary point of access in interactive Web IR systems, is described. This continues to be the main approach to information seeking on the Web.

### 2.4.2 Query Expansion

Techniques such as query expansion have been proposed to better satisfy users information needs (Lin, Wang, & Chen, 2006). One of the underlying goals of query expansion techniques is to improve the way that search engines cope with user queries (Balfe & Smyth, 2004). Efthimiadis (1996) identified two query expansion or formulation stages;

i. The *initial query formulation stage*, in which, the user first constructs the search strategy and submits it to the system and;

ii. The *query reformulation stage* in which the initial query is adjusted manually or with the assistance of a system.

According to French, Brown and Kim (1997) the query reformulation process can be even more difficult than the initial formulation since users often experience difficulty in incorporating information from (and knowledge of) previously retrieved documents into their queries. Furthermore, formulating effective queries in search engines can be challenging for some, given that the

Web is used by a diverse population with varying levels of information search skills or expertise (Aguillo, 2000).

In general, most query expansion approaches expand the original search query by adding additional new/related terms to it. Most of these related terms are acquired and added to an existing query either by the user – interactive query expansion (IQE) – or by the retrieval system – automatic query expansion (AQE), in order to increase the accuracy of the search. Comparative user studies of IQE versus AQE (Beaulieu, 1997; Koenemann & Belkin, 1996) have produced inconclusive findings regarding the relative merits. Ruthven (2003) suggested that one reason for this discrepancy in findings is that the design of the interface, search tasks and experimental methodology can affect the uptake and effectiveness of query expansion techniques.

The IQE approach utilises relevance feedback (RF) techniques (section 2.4.3) that require a user to identify which documents are relevant and then select the terms that lead to a query that best distinguishes relevant from irrelevant documents or terms. Ruthven (2003) suggested that, in general, IQE can be reliable but involves extra work for the user and its success depends on three factors: the relevant documents containing appropriate terms to add to the query, effective display of relevant terms for user selection, and the user identifying good relevant documents. Studies of IQE effectiveness have shown that it can be worthwhile but users may often make poor term selections (Belkin et al., 2000; Nanas, Uren, & Roeck, 2003). This may be due to various reasons, such as lack of user’s domain knowledge, difficulty in translating the information needed into appropriate query terms, synonymy (vocabulary mismatch due to use of synonymous terms) and polysemy (similar words with different meanings when used in different contexts) (Deerwester et al., 1990). Additionally, users may not be able to provide enough terms to accurately or exhaustively cover the scope of their information needs. Besides, poor design of an interface of an IR system may cause the process of RF for query expansion to appear more difficult than it really is (Kelly & Teevan, 2003). The difficulty in interfacing with a system might lead the users away from information retrieval systems that use explicit relevance feedback. This would mean that the learning curve for such users to understand the processes could be long and may lead to an under appreciation of such techniques. According to Jansen (2005), users seldom make use of advanced features available in an IR system even though it helps in improving the results.

AQE (also known as probabilistic query expansion) adopts an alternative approach usually based on calculating co-occurrences of terms in documents and selecting terms that are most related to initial query terms. Excellent reviews of early AQE approaches have been conducted by Cui, Wen, Nie, and Ma (2002) and Xu & Croft (2000). The most significant advantage of the AQE approach is that it automatically identifies terms and is often preferred from relevant resources, such as subset of the initial retrieval results, query logs, most commonly retrieved documents and so on. Most AQE approaches can be categorised into two classes – global analysis and local analysis. Both seek to construct a matrix or thesaurus of relations between the terms, in a global or local sense, and use it to expand the original query with the terms considered best related.

Global analysis techniques, such as term clustering, latent semantics indexing (Deerwester et al., 1990) and Phrase finder (Jing & Croft, 1994) require corpus-wide statistics, such as statistics of co-occurrences of pairs of terms, which results in a similarity matrix among terms. To expand a query, terms those are most similar to the query terms are identified and added. Although, these techniques are relatively robust, they consume a considerable amount of computing resource. In addition, these techniques only focus on the document side and do not take into account the query side; they cannot address the term mismatch problem well.

Different from global analysis, local analysis approaches such as local clustering (Attar & Fraenkel, 1977) and local context analysis (Xu & Croft, 2000) use only the top ranked documents retrieved by the original query. Typically, expansion terms are extracted from those documents and this process treats local feedback as a special case of relevance feedback where the top ranked documents were assumed to be relevant (Xu & Croft, 1996). However, the local analysis might not be effective if the top-n relevant documents do not contain appropriate terms to add to the query. Table 2.2 presents summary of various query expansion approaches.

**Table 2.2: Summary of query expansion approaches.**

| Approaches | Study |
|---|---|
| Ontology (Gruber, 1993) or concept based | •RUBIC (Klink, Hust, Junker, & Dengel, 2002)<br>•ARCH (Sieg et al., 2004)<br>•Concept-based interactive query expansion (Fonseca et al., 2005)<br>•Y!Q (Kraft et al., 2005)<br>•WebSifter II (Kerschberg, Kim, & Scime, 2001) |
| Thesauri based | •WordNet-like lexical networks (Bonino, Corno, & Pescarmona, 2005)<br>•Geographical Information Retrieval (Buscaldi, Rosso, & Arnal, 2005)<br>•Word sense disambiguation in queries (Liu, Yu, & Meng, 2005) |
| Search history or query log | •Probabilistic query expansion (Cui et al., 2002)<br>•Query expansion using associated queries (Billerbeck, Scholer, Williams, & Zobel, 2003)<br>•Adaptive Web search (Sugiyama et al., 2004) |
| Fuzzy rules | •Fuzzy information retrieval (Horng, Chen, Chang, & Lee, 2005)<br>•Document retrieval based on fuzzy rules (Lin et al., 2006) |
| Collaborative filtering | •Collaborative filtering (Herlocker, Konstan, Borchers, & Riedl, 1999)<br>•Eigentaste (Goldberg, Roeder, Gupta, & Perkins, 2000) |

A review of the query expansion approaches listed in Table 2.2 reveals that they all expand the original search query by adding additional new/related terms to it. The thesauri and concept based approaches are attractive and promising as they use either thesauri or domain specific concept hierarchies to assist users in formulating an effective search query. However, these approaches are time intensive for the user and can be limited by handcrafted vocabularies or concepts. The search history and query log approaches are useful, as log data can be used to train parameters to obtain the query expansion terms, based on selected terms from past user search history or queries that are associated with documents in the collection. Nevertheless, these approaches can be resource-intensive if there is too much log data available. If there is limited or no log data, they can be completely ineffective. Similarly, the fuzzy rules and collaborative filtering approaches suffer from both sets of limitations just described.

To summarise, the main challenges impeding query expansion techniques are;

a. Which terms should be included in the query expansion?

b. How should these terms be ranked or selected?

c. Which levels of query reformulation should be automatic, interactive or manual?

The above discussion also demonstrates that various query expansion techniques, such as word-sense disambiguation, concept-based, and search history logs, are used to find the desired set of query terms to improve the user information seeking experience in the Web. It is very useful for systems to combine these techniques that may give more control to the user. As it is the user who decides the criteria for relevance in a search and makes better decisions on which terms are likely to be useful (Koenemann & Belkin, 1996). However, effective query formulation is possible only when the users are already familiar with the interface that integrates various query expansion techniques. As such, an effective query formulation approach involves two important challenges: to acquire effective query expansion terms and formulating effective Boolean queries.

In this thesis a new query expansion technique is introduced to address some of the challenges discussed above as well as to improve the user information seeking experience in the Web. The approach introduces two distinct techniques to acquire and formulate an effective Boolean query expansion. First, the approach utilises the Lexical database (i.e., WorldNet), the domain specific Ontology (e.g., computer science and travel) and Meta keywords techniques to acquire query expansion terms using the relevance feedback technique. Second, based on acquired query expansion terms, the approach formulates an effective Boolean query to improve the effectiveness of Boolean retrieval performance.

The next section describes relevance feedback, the most commonly used technique to assist in the formulation of effective query statements.

### 2.4.3 Relevance Feedback

Relevance feedback (RF) (Rocchio, 1971) has a history in IR dating back well over 30 years (Spink & Losee, 1996). RF provides a means for automatically reformulating a query to more accurately reflect a user's interests (Allan, 1996). RF has also been employed for user profiling during long-term modelling of a user's persistent interests and preferences (Kelly & Teevan, 2003) and has been researched extensively in interactive settings (Beaulieu, 1997; Belkin et al., 2001; Koenemann & Belkin, 1996). RF can comprise either explicit feedback or implicit feedback. The key difference between these two feedback approaches is in the mechanism used; the explicit feedback approach asks a user for explicit relevance judgments – such as their rankings, inputs, and instructions – whereas the implicit feedback approach attempts to infer the user's information needs through data provided by or from the user in an unobtrusive manner – such as their browsing, reading, and typing behaviour. This feedback information can be used to construct a user's contextual profile – short and long term – and to enable a search engine framework (or any IR system) to query, filter and return relevant information.

Various experiments in the past (Harman, 1992; Salton & Buckley, 1990) have shown the effectiveness of user involvement in the process of RF as a means of improving the results of an information retrieval system. Researchers have found that users exhibit a desire for explicit feedback features and, in particular, term suggestion features (Beaulieu, 1997; Belkin et al., 2000; Hancock-Beaulieu & Walker, 1992). The key research to date includes the development of ARCH (Sieg et al., 2004), WAWA (Rad & Shavlik, 2003), RUBIC (Klink et al., 2002), OntoBroker (Fensel, Decker, Erdmann, & Studer, 1998), Watson (Budzik & Hammond, 1999), Suitor (Maglio et al., 2000), PRISM (Leake & Scherle, 2001), and WordSieve (Bauer & Leake, 2001).

A review of the literature pertaining to these approaches reveals that all use the explicit feedback technique to obtain feedback from a user and attempt to improve the results of Web IR systems as a result. These approaches have the advantage of simplicity. Furthermore, these approaches have an added advantage as they directly seek the user's true reaction, thus potentially minimising the likelihood of wrong (or noisy) interpretation of the user’s interests or information needs.

However, obtaining explicit feedback from users would likely be problematic in many IR applications. The explicit feedback technique has three serious drawbacks. The first is that a requirement to provide explicit feedback increases the cognitive load on the user. For example, the process of identifying relevant documents from the results set of an information retrieval system and providing relevance judgments requires additional effort on the part of the searcher (Jansen & Spink, 2003; Vinay, Wood, Milic-Frayling, & Cox, 2005). This extra effort works against one of the principal benefits of the approach, the reduced cognitive load that results from an information space more closely aligned with the user's perspective. The second is that users may lack understanding of the purpose of relevance feedback techniques and the process of indicating which information is relevant may be unfamiliar to them, and is adjunct to the activity of locating relevant information. For example, if the user is just beginning to learn about a specific topic, the user may lack the vocabulary to describe the topic or provide effective judgement on topics. This could complicate the process for users in using a system to obtain relevant documents using such techniques (Jansen, 2005). The third is that the explicit feedback technique can have complex or extensive contextual knowledge acquisition processes, as users may be required to read and rate a large number of documents in order to build the contextual user model. One can imagine that the extra time required to rate each article might prevent users from viewing additional documents of interest. Ultimately, the cognitive burden placed on the users by requiring them to rate a large number of documents is problematic. It is likely to be clear to users that they will have to spend additional time engaged in an activity that does not appear to immediately address their primary needs; it is not evident how this will help. Unless the benefit is obvious to users, documents are likely to be un-read and un-rated, and models will not be constructed. These difficulties motivate further study of implicit feedback techniques.

Implicit feedback techniques have been used to retrieve, filter and recommend a variety of items, such as hyperlinks, terms, Web documents, Internet news articles, movies, books, television programs, jobs and stocks (Kelly & Teevan, 2003) Implicit feedback can be inferred from user behaviour without any additional work on the part of the user. Some of the behaviours that have been most extensively investigated as sources of implicit feedback include reading time, saving, printing, selecting and referencing. As shown in Figure 2.4, Oard and Kim (2001) classified observable feedback behaviours according to two axes; Behaviour Category (Examine, Retain, Reference and Annotate) and Minimum Scope (Segment, Object and Class).

**Minimum Scope**

| Behavior Category | Segment | Object | Class |
|---|---|---|---|
| **Examine** | View<br>Listen | Select | |
| **Retain** | Print | Bookmark<br>Save<br>Delete<br>Purchase | Subscribe |
| **Reference** | Copy-and-paste<br>Quote | Forward<br>Reply<br>Link<br>Cite | |
| **Annotate** | Mark up | Rate<br>Publish | Organize |

**Figure 2.4: Classification of behaviours that can be used for implicit feedback from Oard and Kim (2001).**

Kelly and Teevan (2003) examined various implicit feedback techniques and further expanded this framework by adding an additional behavioural category called "Create", which describes those behaviours the user engages in when creating original information. Kelly and Teevan also suggested that more research needs to be conducted on understanding what observable behaviours mean and how they change with respect to contextual factors to allow for the effective use of implicit feedback.

Various experiments have shown the usefulness of reading time as an observable behaviour (Kim, Oard, & Romanik, 2000; Morita & Shinoda, 1994). Claypool, Le *et al.* (2001) examined the utility of other implicit means of obtaining feedback from users, such as mouse clicks, scrolling and time spent on a page. A number of approaches have been proposed that use implicit feedback techniques to enable a system to present relevant documents to users. Despite their long history in IR research, these implicit approaches have not been successfully implemented in Web-based information retrieval (Croft, Cronen-Townsend, & Lavrenko, 2001). The key research to date includes (but is by no means limited to) the development of UCAIR (Shen et al., 2005), Adaptive web search (Sugiyama et al., 2004), Top ranking sentences approach (White, Ruthven, & Jose, 2002), Reinforcement agent (Seo & Zhang, 2000), WAIR (Zhang & Seo, 2001), Watson (Budzik & Hammond, 1999), and InfoFinder (Krulwich & Burkey, 1997).

These approaches use various techniques to gather and interpret information about a user on a completely implicit basis or without requiring additional effort on the part of the user. This contextual information is gathered when the user performs search tasks, such as querying, browsing and using results. In contrast to explicit feedback techniques, such implicit approaches have the advantage that information can be collected at much lower cost, in much larger quantities, and without additional burden on the user of the retrieval system (Kelly & Teevan, 2003; Morita & Shinoda, 1994).

However, the feedback technique, the implicit feedback technique has three serious drawbacks. The first and major concern is user privacy. Systems utilising implicit feedback techniques, such as WAIR (Zhang & Seo, 2001), Watson (Budzik & Hammond, 1999) and InfoFinder, collect information on the behaviour of users, such as past queries, reading time and printing behaviour. Users might perceive such a feature in a system to be a breach of their privacy rights. The second is that implicit feedback techniques are resource intensive as the system needs to continuously monitor and extract information from the user’s information-seeking behaviour. Such systems also need to analyse potentially large history logs while performing recommendation and information filtration. In addition, there may be an overhead in configuring and initialising the system in order to ‘seed’ its effective operation. In fact explicit relevance feedback may be required (e.g., providing some relevant parameters) building order to initially populate user profiles. Users might find it difficult or obstructive to perform such operations when they do not completely understand the benefits of using such a system. In this case, users might be left with no option but to use explicit relevance feedback mechanisms to improve search results. These systems can therefore be less effective in early stages of use due to insufficient user profile data and may be resource intensive in later stages due to large volumes of profile data. It is also possible that if users were for some reason to turn off the implicit feedback mechanism in an information retrieval system (as it is effectively invisible to them), this might negatively impact the performance of the system. Finally, information gathered from the user’s information-seeking behaviours using implicit feedback techniques may be difficult to interpret or could be potentially noisy.

To summarise, extensive research has shown that relevance feedback in query expansion helps to improve the quality of Web search results. The main challenges relating to current RF mechanisms are;

a. How to capture a user's information seeking behaviour and their preferences and structure this information in such a way as to be able to define a search context that can be refined over time?

b. How to help the users form communities of interest while respecting their personal privacy?

c. How to develop algorithms that combine multiple types of information to compute recommendations?

In this thesis, it is proposed that implicit measures can be combined with explicit ratings to obtain a more accurate and complete representation of a user’s interests. This approach also facilitates the construction of adaptive user profiles and the shared contextual knowledge base. This approach may address some of these challenges and potentially improve the user information seeking experience in the Web. In addition, the user privacy issue can be addressed by having anonymous history traces and by aggregating user information seeking behaviours.

The effectiveness of CIR needs to be evaluated. The next section discusses issues in the evaluation of such systems and techniques.

## 2.5 Evaluation

According to Gao, Murugesan and Lo (2004), the main objective of evaluation of IR is to identify advantages and disadvantages of a retrieval strategy and seek methods to improve retrieval results. Over time, numerous evaluation approaches, such as test collection approaches, search log analysis, human experimentation in the lab, and naturalistic observation have been developed. Thomas and Hawking (2006) reviewed some of these established approaches. Each approach has its advantages and limitations and it is difficult (and at times inappropriate) to mix different approaches (Saracevic, 1995). Since the evaluation method was not the central focus of this study (that is, the study set out to deliver a novel Web IR approach, not a novel evaluation method), the dominant approach (i.e., test collection) and more recent alternatives are presented in the following paragraphs.

Cleverdon's (1960) Canfield model is the dominant approach used in the evaluation of IR and interactive information retrieval systems (IIR) (Borlund, 2003) and notably taken up by the TREC[7], CLEF[8], INEX[9], and NT-CIR[10] conferences (Thomas & Hawking, 2006). The model uses a system driven approach and is based on a principle of text collection analysis; a collection of documents, a collection of queries and a collection of relevance assessments. The model most often uses relevance-based measures of recall and precision. These measures may not be sufficient to develop a holistic view on what factors make an effective search system (Borlund, 2003; Su, 1992). Also, the model

---

[7] Text Retrieval Conference, http://trec.nist.gov/

[8] Cross-Language Evaluation Forum, http://www.clef-campaign.org/

[9] Initiative for the Evaluation of XML Retrieval, http://inex.is.informatik.uni-duisburg.de/

[10] NII Test Collection for IR Systems, http://research.nii.ac.jp/ntcir/outline/prop-en.html

suffers from limitations due to restricted assumptions on the cognitive and behaviour features (Ellis, 1996). Furthermore, applying this model to contextual IR systems, or those which include personal corpora, raises particular privacy concerns (Thomas & Hawking, 2006). Hence, various researchers (Borlund, 2003; Borlund & Ingwersen, 1997; Cosijn & Ingwersen, 2000; Harter, 1996; Saracevic, 1995) have advocated a demand for alternative approaches to the performance evaluation of IIR systems. Borlund (2003) proposes an aggregated model for the evaluation of IIR systems and/or information seeking behaviour that provides a framework for the collection and analysis of IR interaction data.

Since the 1980s, there has been a shift towards a ‘person-centered’ approach, rather than a ‘system-centered’ approach to evaluation (Wilson, 2000). Saracevic (1995) suggested that the majority of IR evaluation approaches had used the system approach and often neglected to recognise the people using the IR systems, and that this was limiting progress in IIR. The shift from a ‘system-centered’ approach to a ‘person-centered’ approach also incorporated a shift in research methods, from quantitative methods to qualitative methods. Martzoukou (2005) reviewed Web information seeking research method and found limited comprehensive studies, inconsistencies in method, and a lack of homogeneity in research foci. Martzoukou concluded that the effect of social and cultural elements has not been extensively investigated and qualitative and quantitative methods are needed to produce a comprehensive view of Web information seeking. In addition, several previous studies (Freund & Toms, 2002; Hölscher & Strube, 2000; Kari & Savolainen, 2003; Kellar, 2006; Sandra, 2000; Spink & Jansen, 2004; Spink & Ozmultu, 2002; Wilson, 1999a) have been entirely exploratory. This is due to the still emergent nature of knowledge regarding the Web as a tool for information seeking. Thus studies investigating IR have employed qualitative methods (Harper & Kelly, 2006; Jansen et al., 2007), quantitative methods (Finkelstein et al., 2002; Spink, Wolfram, Jansen, & Saracevic, 2001) or a combination of both (Choo, Detlor, & Turnbull, 2000; Martin & Jose, 2004).

Saracevic (1995) suggested (rightly) that the choice of evaluation approach depends on the intent of the evaluation and in many ways it defines the type of results obtained. However, there is little discussion of how the IR evaluation models are being used to evaluate CIR systems. There are challenges in evaluating CIR using traditional experimental methods, as the use of assigned tasks and standard recall/precision metrics will not be an effective measure of improvement in higher-order relevance. As such, this thesis adopts one of Borlund's (2003) experimental components; simulated work task situations that are used to create search scenarios, and use a broad range of other evaluation methods including metrics for recall and precision, efficiency, user satisfaction, usability and quality output (Järvelin & Ingwersen, 2004) to evaluate the contextual search framework (described in Chapter 3). The concept of simulated work task situations is derived from Ingwersen's cognitive communication models (Ingwersen, 1992, 1996) and the application of the work task concept by Byström and Järvelin (1995) to information problem solving and information seeking processes. As this study incorporates both exploratory and confirmatory elements it utilises the mixed methods research approach (Tashakkori & Teddlie, 2003).

## 2.6 Chapter Summary

This chapter has described the motivation behind the work presented in this thesis. Today's Web search engines suffer from a number of problems that make the CIR approach an appealing alternative that should help users to facilitate better queries that will return the most relevant information. Researchers working in the area of CIR to date have focused their efforts on three major themes: user profile modelling, query expansion, and relevance feedback.

Current user profile modelling techniques face a number of challenges, namely; 1) how to acquire, maintain and represent accurate information about a user's multiple interests with minimal intervention, 2) how to use this acquired information about the user to deliver personalised search results, and 3) how to build and use these acquired information about various users as knowledge based in large communities or groups. Similarly, current query expansion techniques also face a number of challenges, such as 1) which terms should be included in the query expansion, 2) how should these terms be ranked or selected, and 3) which levels of query reformulation should be automatic, interactive or manual. Finally, current RF techniques also have their own challenges; 1) how to capture a user's information seeking behaviour and their preferences and structure this information in such a way as to be able to define a search context that can be refined over time, 2) how to help the users form communities of interest while respecting their personal privacy, and 3) how to develop algorithms that combine multiple types of information to compute recommendations.

In the next chapter, the detailed design and implementation of the contextual SERL search, as an alternative CIR framework to address many of the issues presented in this chapter, is presented.

# Chapter 3

# SERL Search Architecture

## 3.1 Introduction

Chapter 2 identified a number of issues related to Contextual Information Retrieval (CIR) and current approaches. This chapter describes the design and implementation of Software Engineering Research Lab Contextual Search system (or simply contextual SERL search) in the course of this research to address those issues presented in Chapter 2.

The design and implementation of the contextual SERL search involved constructing an alternative CIR framework that seeks to improve the efficiency and effectiveness of search, define new ideas, and technical capabilities. As such, the system development research methodology (SDRM) and the design-science (DS) research guidelines are employed to investigate, design, develop and implement the contextual SERL search. The contextual SERL search was designed to implement several system requirements, such as adaptation of a user's information seeking behaviour, recognition of a user's preferences and interests, recommendation of terms, generation of Boolean query and presentation of ranked contextual search results to address the three main research objectives motivating the work. These requirements imposed the integration of specialised third party modules, such as Jena[11] and WordNet[12] and heterogeneous software, such as Google SOAP Search API[13]. As such, the contextual SERL search deployed the popular three-tier architecture, which offers many benefits, such as increase flexibility and reusability, increased scalability and improved performance. Various design and implementation considerations are made during the development of the contextual SERL search.

The three main research objectives of the contextual SERL search are;

i. **Research Objective I** – To construct a CIR model that captures both a user's behaviour and preferences as a user's personal contextual profile and structure this information in such a way as to be able to define a search context that can be refined over time.

ii. **Research Objective II** – To facilitate the collection of multiple users' personal contextual profiles to create a shared understanding of contextual knowledge base.

iii. **Research Objective III** – To use the user's personal contextual profile or together with the shared contextual knowledge base to refine search queries, filter returned results from search engines, and provide user recommendations/suggestions.

This chapter describes the development methodology followed by the architectural design and implementation details of the contextual SERL search, including the assumptions made and current limitations associated with the system. It provides sufficient detail to act as both a user guide and documentation for modifying or re-implementing the contextual SERL search.

---

[11] Jena – A Semantic Web Framework for Java, http://jena.sourceforge.net/index.html

[12] WordNet®- lexical database of English, http://wordnet.princeton.edu/

[13] Google SOAP Search API (Beta), http://code.google.com/apis/soapsearch/

## 3.2 Development Methodology

Since this research sets out to explore an alternative CIR framework which has not been examined previously, the SDRM (Nunamaker et al., 1991) and the DS research guidelines (Hevner et al., 2004) are employed to investigate, design, develop and implement the contextual SERL search. The SDRM methodology holds that systems development is an evolutionary process and one that has been used extensively in research in information systems development. Similarly, DS seeks to create design-based innovations – in the software systems context it builds ideas, practices, technical capabilities and products through the effective and efficient analysis, design, implementation, management, and use of such systems.

The SDRM methodology consists of five iterative phases, where each phase consists of various research activities. Four additional phases were added to the SDRM methodology; they are literature review, study of existing technology, data analysis and documentation. Similarly, the DS research guidelines consist of seven well-defined research practices, which essentially complement the iterative phases of the SDRM methodology.

Figure 3.1 shows the diagrammatic form of the SDRM methodology, the four additional related phases and the DS research guidelines as applied in the development of the CIR framework. The four additional related phases have been highlighted in Figure 3. 1.

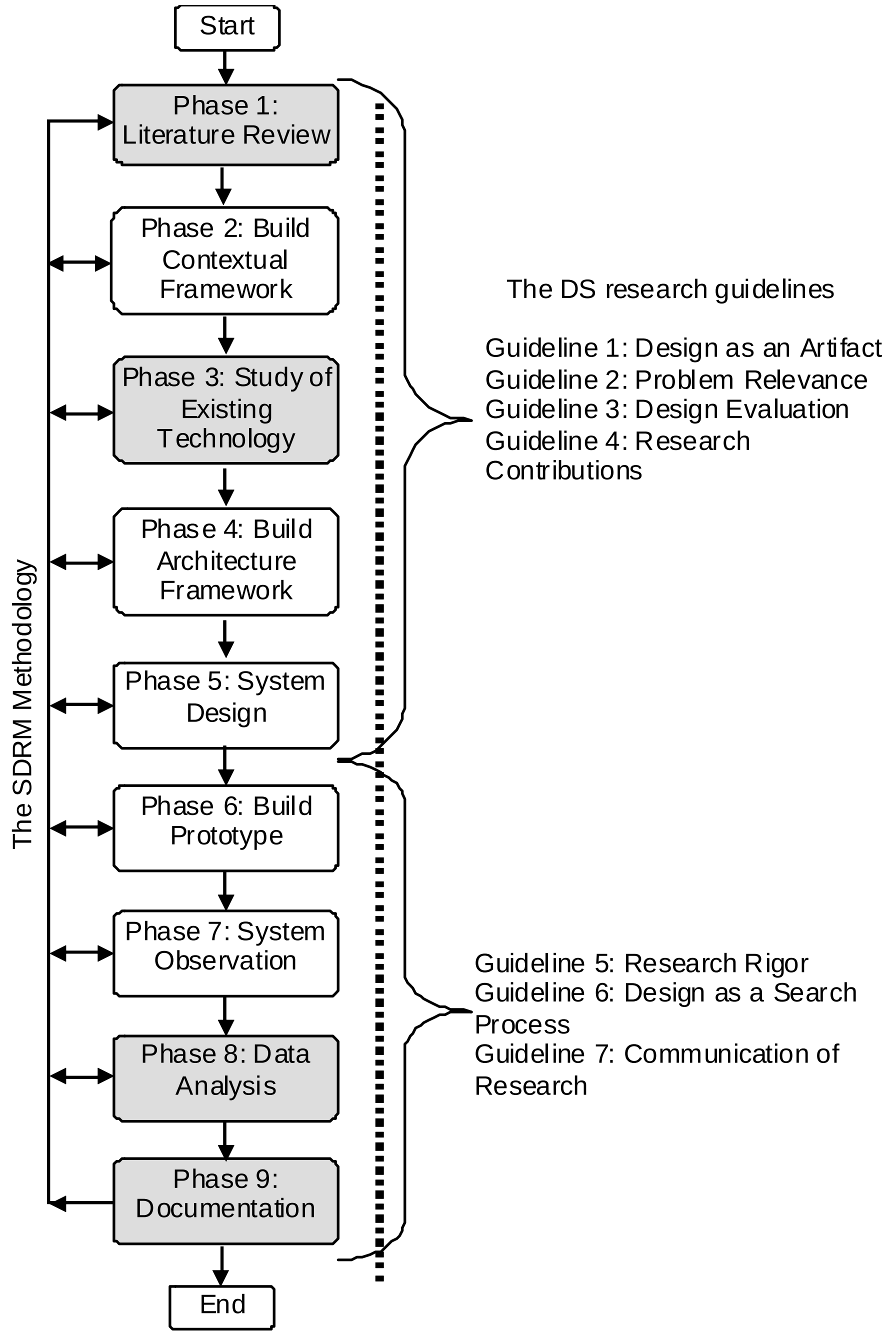


**Figure 3.1: SDRM methodology and DS research guidelines for CIR.**

The outcomes of the in-depth review of existing CIR related literature (phase 1) were described in Chapter 2. This phase identified the existing research status, their challenges, future directions and research questions/objectives. Phase 2 saw the building of the alternative CIR conceptual framework that provides a very high-level system overview of processes and basic components. The development of the framework is informed by three primary research objectives as stated in section 3.1. Phase 3 explored and evaluated existing CIR related technologies (e.g., tools and techniques) and identified the potential use of these technologies in the proposed conceptual framework. Phase 4 defined the proposed contextual SERL search architecture as the CIR framework (section 3.4) that provides the system component details, their relationships and their basic functionalities. The popular three-tier architecture is deployed in the contextual SERL search. The three-tier architecture aims to solve a number of recurring design and development work easily and efficiently. Phase 5 results in the detailed design of the whole proposed contextual SERL search (such as the detailed functionality of each component, their independencies, database design and so on). Several architectural design and implementation issues, such as scalability, flexibility, performance, and robustness were considered during the detailed design of the contextual SERL search. The detailed design and implementation of the contextual SERL search's three-tier architecture will be discussed in details in section 3.5.

Subsequently, in phase 6, the proof of concept of the contextual SERL search (described in this chapter) was developed. This phase also assessed the feasibility, reliability and performance of the contextual SERL search. Phase 7 observed and evaluated the performance of the contextual SERL search under experimental conditions with real users (described in Chapter 4). Phase 8 analysed the resulting evaluation data (described in Chapter 5) and summarised useful information, developed conclusions and outlined directions for future work (described in Chapter 6) as it was recognised that this research area is potentially huge and not all desirable features may feasibly be implemented in the bounded timescale of a PhD. Finally, the research ended with phase 9 that documented all of the processes and findings for the above-mentioned phases.

The design and development of the contextual SERL search conformed to the seven DS research guidelines. As such, for phases 1 – 4, the research study produced a viable artefact and developed technology-based solutions that are potentially important or relevant to business problems. The research study rigorously demonstrated the utility, quality, and efficacy of a design artefact via well-executed evaluation methods. The research study provided clear and verifiable contributions in the areas of the design artefact and design foundations. Similarly, for phases 5 - 9, the research study relied upon the application of rigorous methods and search for the best, or optimal, design, as this is often intractable for realistic information systems problems. Finally, the research study is presented effectively for both technology-oriented and management-oriented audiences.

Use of the SDRM methodology and the DS research guidelines helped to avoid some of the pitfalls of linear processes (e.g., the waterfall model and the incremental model), and facilitated the development of the contextual SERL search and the demonstration of its technical feasibility.

## 3.3 Architecture Pattern

Figure 3.2 shows the three-tier architecture. As the name suggests, the architecture is composed of three layers; the presentation layer (user interface), the application layer (application or middle) and the database layer (persistence).

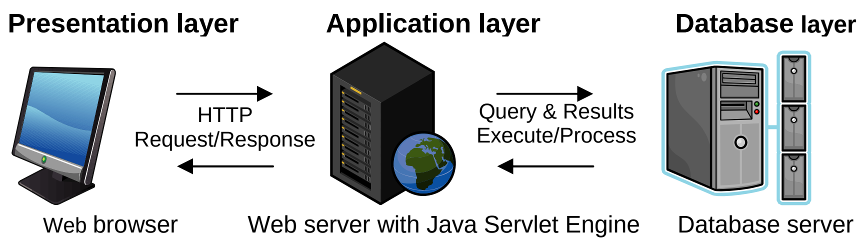


**Figure 3.2: Three-tier Architecture.**

The presentation layer contains presentation logic and is responsible for the delivery and formatting of information to the application layer for further processing or display. The application layer contains functional process logic and is responsible for application processes. The database layer contains the data storage logic and is responsible for writing and retrieving data to and from the database. The three-tier architecture provides several benefits over traditional client-server applications as shown in table 3.1.

**Table 3.1: Three-tier architecture benefits.**

| Factor | Description |
|---|---|
| Scalability | Allows distribution of application components across multiple servers thus making the system much more scalable. |
| Flexibility | Separates the business logic of an application from its presentation logic and makes the application much more flexible to change. |
| Performance | Any performance requirements on the application can be addressed by running either on a single machine or multiple separate machines. |
| Reusability | Separates the application into multiple layers and makes it easier to implement re-usable components. |

As a result of these benefits, a variety of research work, such as scientific simulation environments (Holmes et al., 2002), hurricane occurrence simulation (Chen et al., 2003), and space life support systems (Schreckenghost, Bonasso, Kortenkamp, & Ryan, 1998), has successfully adopted the three-tier architecture to expedite and automate the development of web-based applications. The following section discusses the contextual SERL search's three-tier architecture in more detail.

## 3.4 Architectural Building Blocks

Figure 3.3 shows the contextual SERL search's three-tier architecture. The architecture separates the contextual SERL search's system activities into three layers; Tier 1: presentation layer, Tier 2: application layer and Tier 3: database layer. Several terminologies are used to describe the contextual SERL search's system activities. They include;

**Search context**: The representation of a user's search intent. The search context provides an abstract or a common template to describe or define a user's dynamic search interests (e.g., office related search or home related search). The search context is explicitly defined by the user using one of the Contextual SERL search components, which are updated over time to reflect changes in their search interests. The search context information is later used by the Contextual SERL search to accomplish its various functions (e.g., facilitates the recommendation process to recommend relevant query expansion terms). For example, the user may define the context as ‘office’ and perform a search to find office or work related information on the Internet. This context information is stored in the user’s contextual profile and if shared, it can be used to recommend relevant terms/phrases/concepts and visited URLs to other users with the similar context. With respect to users’ multiple search interests, if a user submits a keyword (e.g., “surfing”) to the contextual SERL search, the system uses the user’s profiles, including a preconfigured search context (e.g., “office” or “home”) and shared profiles (if this is enabled by the user) to recommend related query expansion terms (e.g., for “office”, computer science related terms, and for “home”, travel related terms) that potentially reflect the user’s search intent.

**Ontology:** An explicit specification of a conceptualisation, where conceptualization refers to possible worlds (Gruber, 1993). The ontology (e.g., computer science[14] and travel[15]) incorporates logical relationships and membership rules between hierarchies of concepts. These concepts are used in the contextual SERL search to capture a user's search interests, to facilitate the construction of the user’s contextual profile, and to formulate a Boolean query. These concepts are explicitly selected by the user using one of the Contextual SERL search components. These concepts are stored and later used by the Contextual SERL search to accomplish its various functions (e.g., formulates a Boolean query).

**Meta keywords**: The brief and concise list of important keywords from the visited Website. These keywords are extracted from either the Meta keyword tag (i.e., <meta name="keywords" content="meta keywords tag metadata indexing search engines index meta data elements">) or the TITLE element (i.e., <title>Name or title of the web site</title>) of the visited Web site. The Meta keyword tag consists of keywords that describe/relevant to the Website.

---

[14] The computer science ontology is based on the ‘Top-Level Categories for the ACM Taxonomy’ downloaded from http://cse.unl.edu/~scotth/SWont/acmCCS.owl

[15] The travel ontology is downloaded from http://protege.cim3.net/file/pub/ontologies/travel/travel.owl

The TITLE element defines the title of the document/Web site. For example, the Meta keyword tag and the TITLE element of the IBM Website are;

<meta name="keywords" content="meta keywords tag metadata indexing search engines index meta data elements">.
<title>IBM - United States</title>

Table 3.6 describes on how these Meta keywords are extracted from the browsed Websites in more detail. Similar to Ontology, these Meta keywords are used in the contextual SERL search to capture a user's search interests, to facilitate the construction of the user’s contextual profile, and to formulate a Boolean query. These keywords are explicitly selected by the user using one of the Contextual SERL search components and are later used by the Contextual SERL search to accomplish its various functions (e.g., formulates a Boolean query).

**Contextual User Profile** (*U_profile*): The descriptions of a user’s search interests. The *U_profile* consists of both the user’s Internet search behaviour information, such as previously issued queries, previously visited URLs, extracted Meta keywords and the user’s preferences information, such as the user’s search contexts, selected disambiguated terms (Table 3.4 describes the disambiguation process in more detail), selected Meta keywords, and selected category or concepts. These *U_profile* information is later used by the Contextual SERL search to refine search queries (Knowledge Base Query Formulator section describes the query formulation process in more detail), filter returned results from search engines (section 3.5.2.3 describes the result filtration process in more detail), and provide user recommendations/suggestions (Table 3.3 describes the recommendation process in more detail).

**Contextual Knowledge base (***Kb_profile***):** The collection and organisation of multiple users’ profiles. Similar to the *U_profile*, the *Kb_profile* information is later used by the Contextual SERL search to refine search queries, filter returned results from search engines, and provide user recommendations/suggestions.

**Query:** A search query that comprises of one or more keywords.

**Boolean query:** The Boolean query is the query that is formulated by the contextual SERL search using one or more selected query expansion term(s), such as selected disambiguated terms, selected Meta keywords, and selected category or concepts (Knowledge Base Query Formulator section describes the query formulation process in more detail).

As shown in Figure 3.3, the top and bottom layers are the presentation layer and database layer respectively. The top layer is where users interact with the contextual SERL search. The layer is responsible for acquiring data from the users and rendering data back to the users. The bottom layer is the database layer where the contextual SERL search database connection and manipulation tasks are performed. The layer is responsible for writing and retrieving data to and from the database, and returning the results to the application layer. The layer is concerned with persistent data and connected to a MySQL[16] database.

The middle layer is the application layer where the contextual SERL search's actual business/application functions are performed. The application layer is the core layer of the contextual SERL search as it links the two other layers: presentation layer and database layer. For example, the layer processes requests from the presentation layer (e.g., a user registration) and sends instructions to the database layer to store or retrieve a piece of data (e.g., the registration data).

[16] MySQL, http://www.mysql.com/

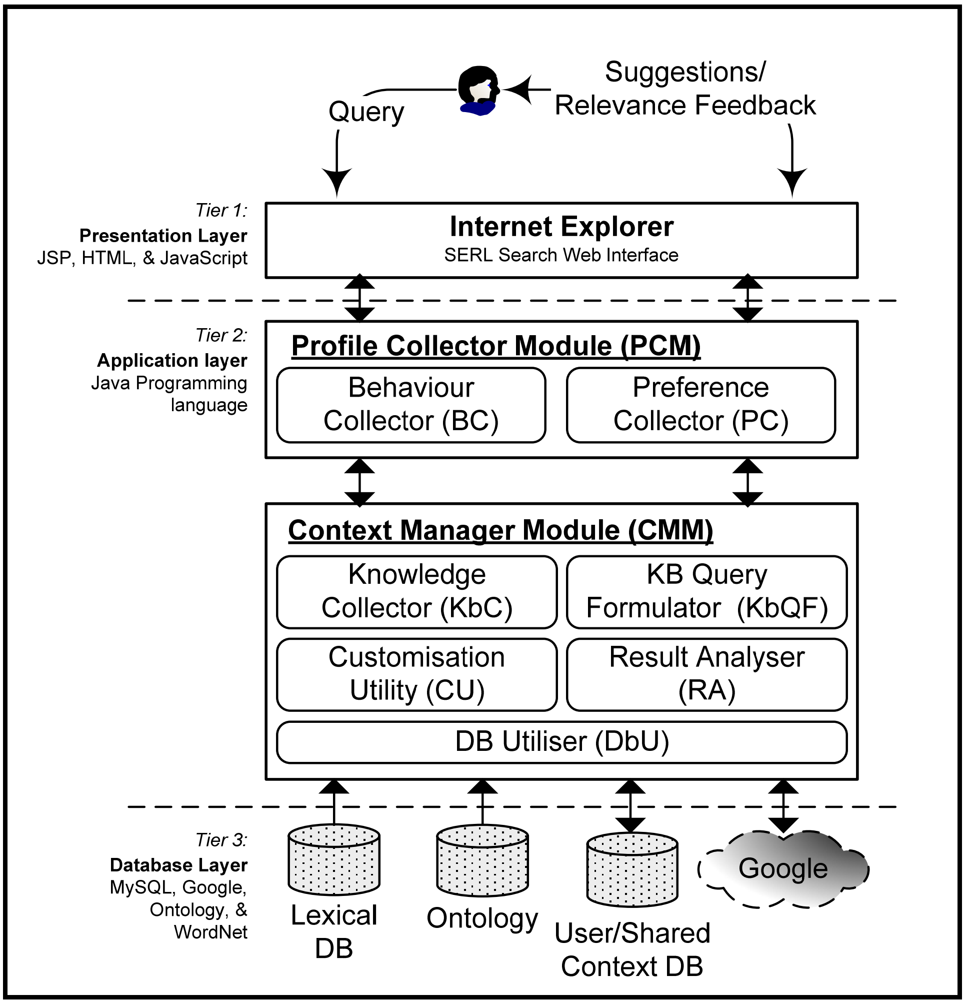


**Figure 3.3: Contextual SERL search three-tier architecture.**

The application layer comprises two main modules: Profile Collector Module (PCM) and Context Manager Module (CMM), each of which contains a number of modules that are anticipated to improve a user's web searching experience. Each of these components can be considered as a candidate solution for a final architecture. Each component has been designed by reflecting on research outcomes of other researchers. The purpose of the research can then be considered as testing the contribution of each of these components to the overall user searching experience. The SDRM is adopted because it promotes iterative refinement of both the individual components and the manner in which they inter-relate. This refinement is seamless because the three-tier architecture allows the application layer to be easily changed, provided the interfaces to the presentation and persistence layers are not changed.

The PCM is implemented to address ***research objective I***. Many Web IR systems have explored various user modelling approaches to address similar research objectives. However, as described in Chapter 2, section 2.4.1, many of these approaches have faced a number of difficult challenges with resulting limitations. In contrast, the PCM introduces three distinct techniques to acquire and maintain a user's contextual profile and structure this information in such a way as to be able to define a search context that can be refined over time. These three techniques are grouped together in the PCM because their function facilitates the core data collection functionality.

First, the PCM integrates not one, but three different techniques, i.e., the Lexical database, the domain specific Ontology and Meta keywords to reflect the user's interests as well as to capture their preferences using the relevance feedback technique. The PCM constantly acquires and maintains all these data with minimal intervention to represent accurate information about the user's multiple interests.

Second, the PCM implements an additional contextual parameter called "search context" (i.e., "office" and "home") that defines the user's multiple search interests as well as facilitates the construction of adaptive user profiles and the shared contextual knowledge base.

Third, unlike most profiling/modelling approaches, the PCM implements a hybrid contextual user profiling approach to capture a user's adaptive search behaviour by monitoring and capturing their explicit data (i.e., user's search context, selected terms/Meta keywords/concepts) and implicit data (i.e., entered search queries, visited URLs, extracted Meta keywords form visited URLs) to obtain a more accurate representation of users' interests.

The PCM consists of two specialised components: Preference Collector (PC) and Behaviour Collector (BC). The main objective of the PC component is to a capture user's preferences (i.e., explicit data) and at the same time to expand a simple keyword query into a more effective query in order to improve the results of that query. The PC utilises a Lexical database, a domain specific Ontology and Meta keywords technique to capture the user's preferences. Similarly, the main objective of the BC component is to monitors and captures the user's daily Internet search activities (i.e., implicit data) as a user's behaviour.

The CMM is implemented to address ***research objective II*** and ***III***. Many user profiling and query expansion techniques have been studied in the field of Web information retrieval system, to address the similar research objectives. However, as described in Chapter 2, section 2.4.1 - 2.4.3, many of these approaches have faced a number of difficult challenges with resulting limitations. In contrast, the CMM introduces two distinct techniques to address the ***research objective II*** and ***III***.

First, the CMM facilitates the construction of the *U_profile* and the collection of multiple users' personal contextual profiles to create a shared understanding of *Kb_profile*.

Second, the CMM formulates a Boolean search query by employing adaptive data mining technique to learn the user's specific information needs while employing the relevance feedback approach to support iterative development of the search query by suggesting alternative terms/Meta keywords/concepts for Boolean query formulation. These two techniques are grouped together in the CMM because their function facilitates the querying, filtering and recommendation functionality.

The CMM consists of four specialised components: Knowledge Collector (KbC), KB Query Formulator (KbQF), Customisation Utility (CU), and Result Analyser (RA). The main objective of the KbC component is to gather a user's information seeking data (i.e., user's behaviour and preferences data), which are stored in the personal *U_profile* and to update the *Kb_profile*. The main objective of the KbQF component is to expand a simple keyword query into a contextual Boolean query in order to improve the results of that query. The main objective of the RA is to analyse and rank the search results based on a user's personal contextual profile and a shared contextual knowledge base. The CU and DbU main objectives are to provide configuration options and to support essential database operations of persistent storage for the contextual SERL search respectively. The later sections describe these components in details.
All three layers are inter-related and all are required to perform the contextual SERL search functions correctly. The combination of various techniques and

their functionalities, provided by the contextual SERL search distinguishes this research work from previous work.

## 3.5 Design and Implementation

Three-tier architecture and WWW technology was chosen to design the contextual SERL search. The choice of the three-tier architecture was governed by several factors as shown in Table 3.2.

**Table 3.2: Decision making factors in choosing three-tier architecture.**

| Factor | Description |
|---|---|
| Scalability | Any heavier demands on the contextual SERL search (e.g., large numbers of client-request processing) can be addressed by adding additional capacities/resources at the appropriate layer without having to fundamentally change the system design or architecture. |
| Flexibility | Any platform changes on the contextual SERL search can be made easily at any particular tier. Any changes on one of its layers does not affect the others layers unless a complete modification in the table design or in the API (Application Programming Interface) is made. For example, a change in the database layer (e.g., MySQL to Oracle database) does not affect the client presentation layer. |
| Performance | Any performance requirements on the contextual SERL search can be addressed by running either on a single machine or multiple separate machines. |
| Reusability | The contextual SERL search actual system functions are implemented into various components and are re-usable components. |

Similarly, the choice of various design and development technologies, such as JSP, HTML, JavaScript, Java, Apache Tomcat and MySQL to build the contextual SERL search was governed by two main reasons. Firstly, Java was chosen as the coding language, because it is an object-orientated language and portable (i.e., provides cross-platform support). Java is a pure object-oriented language and platform independent programming language. Java promotes greater extensibility and maintainability in programming and does not rely on any special features of any single platform and runs on any machine equipped with a JVM[17], regardless of its processor or operating system. Since the

[17] Java Virtual Machine (JVM), http://www.java.com/en/download/index.jsp

prototype was developed in Java, naturally, Apache Tomcat became the choice of application server. Tomcat is designed to be powerful and flexible (i.e., it supports Java and can work on a wide variety of platforms). Tomcat implements the servlet and the JSP specifications from Sun Microsystems, providing an environment for Java code to run in cooperation with a web server. Similarly, MySQL, the world's most popular open source database was chosen because of its consistent fast performance, high reliability and ease of use (*Why MySQL?*, 2008). The second reason for the selection of these technologies was the researcher’s familiarity and comfort in using them.

### 3.5.1 Presentation Layer

The contextual SERL search’s presentation layer contains the actual user interface (UI) for interaction with the user. This layer is responsible for presenting information in a manner suitable for the application or users dealing with the information. The layer also accommodates other functionality, such as forms validation, managing visual layouts, styles, and the navigation of the system. The layer is implemented using JavaServer Pages[18] (JSP) as web scripting language, Hypertext Mark-up Language[19] (HTML) as content presentation and JavaScript[20] as script language. JSP was used to implement the view, generating the dynamic content of the contextual SERL search. Similarly, JavaScript was used to both validate the user's responses and gather them into a consistent and standardised format.

Figure 3.4 shows the overall UI hierarchy of the contextual SERL search. Instead of inventing a new interface, the contextual SERL search’s UIs were inherited from other search engines, and were designed to be accessible, easy to understand and easy to navigate. For example, a simple, reasonably long search field (i.e., 30 characters) is placed on every page of the contextual SERL search. Site names and navigation links to other related UIs are included in each UI of the contextual SERL search.

---

[18] JavaServer Pages (JSP), http://java.sun.com/products/jsp/

[19] Hypertext Mark up Language, http://www.w3.org/MarkUp/MarkUp.html

[20] JavaScript, http://en.wikipedia.org/wiki/JavaScript

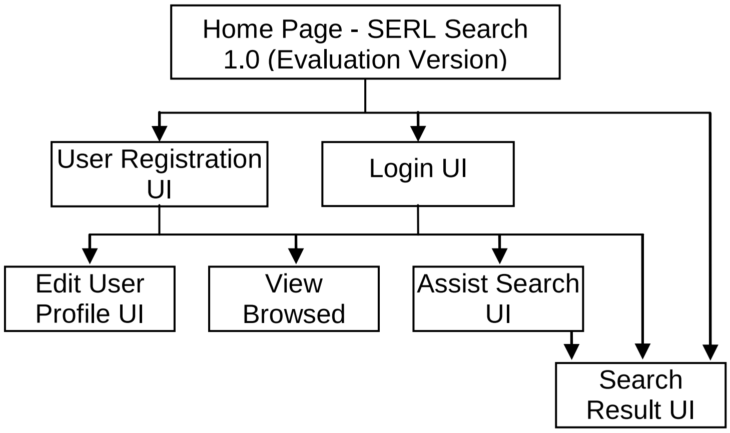


**Figure 3.4: UI hierarchy of the contextual SERL search.**

Marchionini (1995) suggested "a grand challenge for interface designers is to create new features that take advantage of the unique characteristics of the electronic medium". Understanding users' expectations and designing usable, effective, and pleasurable UIs were crucial in the contextual SERL search UI design. As such, Shneiderman's (1998) 'eight golden rules of interface design' and Berry and Browne's (2005) interface design guidelines were employed to elicit the requirements of usability, user experience and interactive UIs for the contextual SERL search. Most of the contextual SERL search's UIs were developed using three types of font and four standard colours. Various UI design considerations, such as the 'progress indicator' (Figure 3.5) and 'no penalty error message' (Figure 3.6) were employed as required. Instead of jargon and technical terms, only simple words, labels and icons were used to explain the search process. As suggested by Shneiderman and Plaisant (2004), the search result UI (Figure 3.7) is designed in such a way that it tells the user how the Contextual SERL search process formulated the Boolean query. For example, as shown in the search result UI (Figure 3.7), the search query is "Java Surf" (A) and the formulated Boolean query is "java (island Indonesia) AND surf OR surf travel company west java" (B). In this context, the search query is entered into the contextual SERL search by a user to satisfy his/her information needs. The search query may include one or more keywords. The formulated Boolean query is the query that is formulated by the contextual SERL search using one or more selected query expansion term(s). The later section (I.e., Knowledge Base Query Formulator) describes the query formulation technique in details.

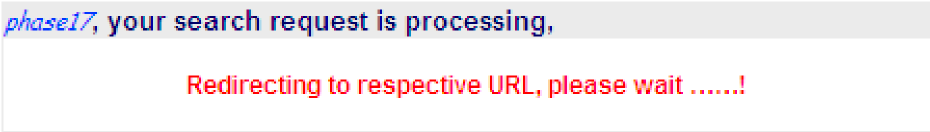


**Figure 3.5: Progress indicator for the contextual SERL search.**

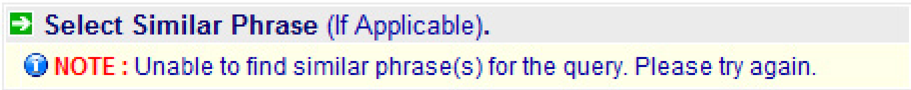


**Figure 3.6: 'No penalty error message' for the contextual SERL search.**

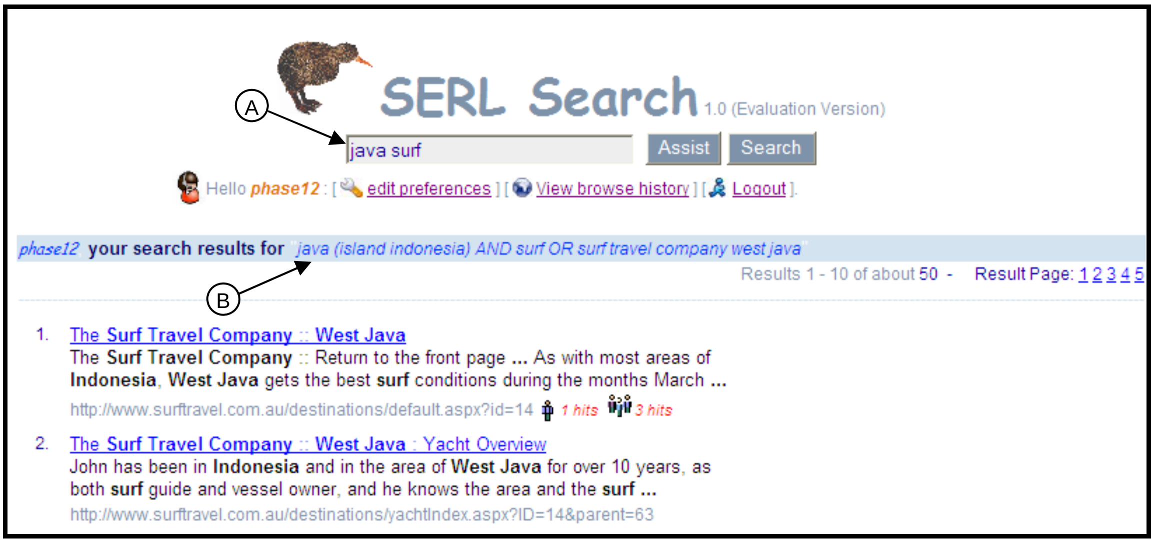


**Figure 3.7: Search results using the contextual SERL search.**

The contextual SERL search home page (Figure 3.8) is the entry page of the contextual SERL Search, where a user has the choice of starting with either a *basic search* or *contextual search*. The differences between the *basic search UI* and *contextual search* UI are briefly highlighted here.

**Basic search** – The basic search UI (Figure 3.8) includes a simple search box and a search button for a user to begin search task(s). The user does not need to *register* or *login* before he/she can perform any search tasks. The basic search is similar to the Google basic search that searches indexed documents for specified keywords and returns a ranked list of the documents where the keywords were found.

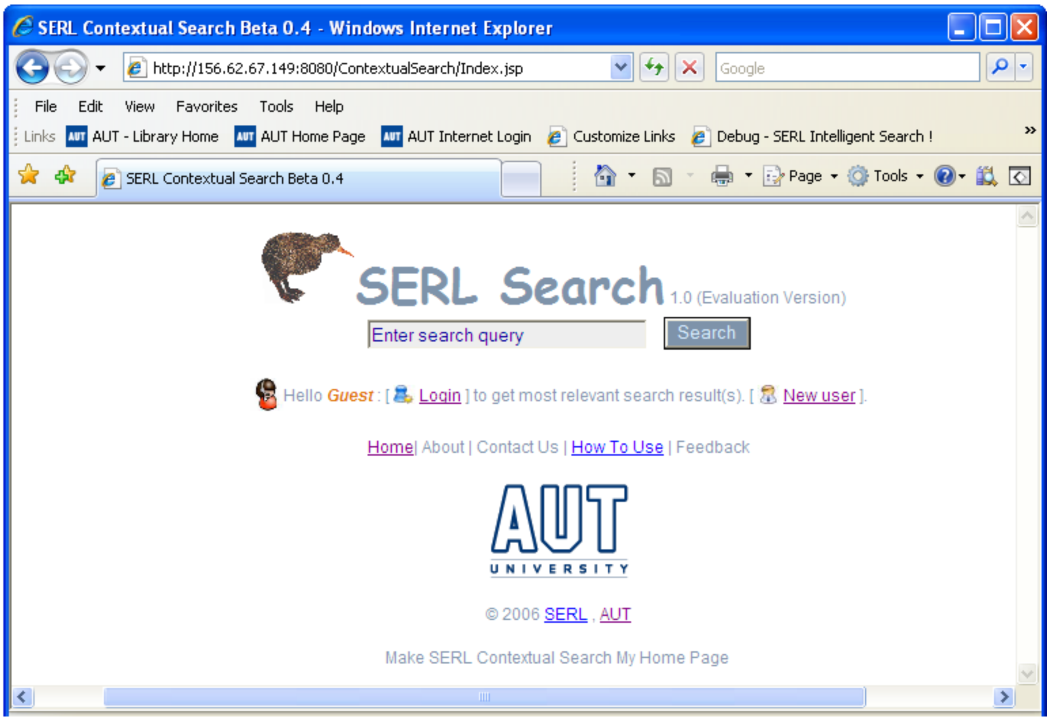


**Figure 3.8: Contextual SERL search home page.**

**Contextual search** – The searcher must initially *register* as a new user employing the user registration UI (Figure 3.9) or *login* as a registered user via the login UI (Figure 3.10) before they may use the contextual SERL search's contextual search feature.

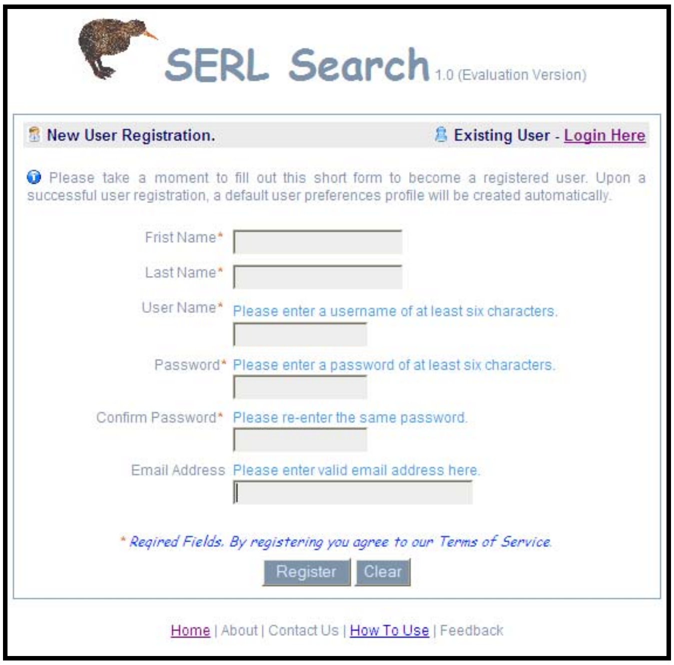


**Figure 3.9: User registration for the contextual SERL search.**


**Figure 3.10: Search login for the contextual SERL search.**

Successful user registration or login will redirect the user to the contextual SERL search UI (Figure 3.11), which includes a simple search box and two buttons – *Assis*t and *Search.* The search button performs a standard search without assistance on the entered "query" and presents search results to the user. In contrast, the assist button is designed to assist the user by presenting a range of advanced search features (Figure 3.12) to refine/clarify user's information needs.


**Figure 3.11: Contextual SERL search UI.**

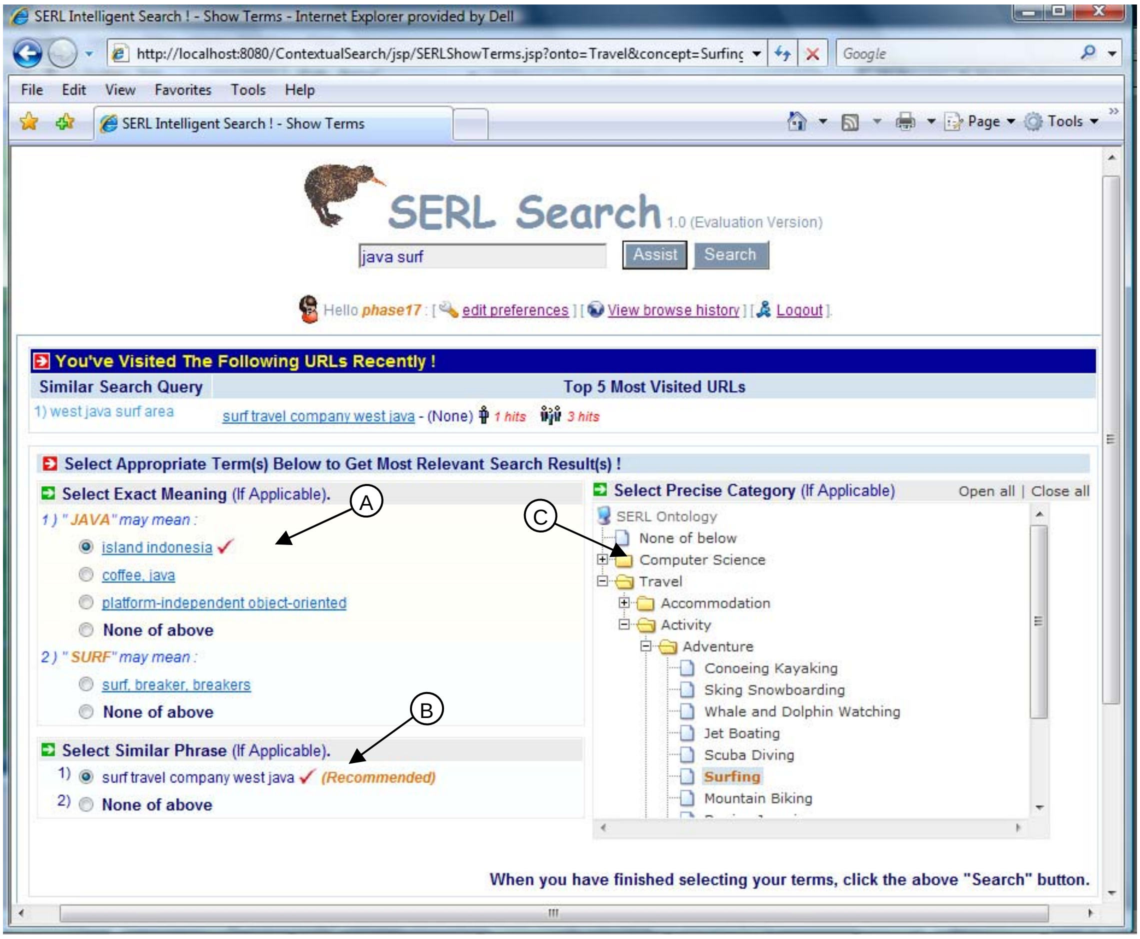


**Figure 3.12: Assistance options for the contextual SERL search.**

The user may refine their query by selecting one or more of the three advanced options as defined in section 3.4:

i. **Exact Meaning** (i.e., Disambiguated Terms) – This option allows the user to select an exact meaning of the entered query. This option uses the lexical database (i.e., WordNet) to disambiguate vague search keywords by promoting disambiguated terms. It also recommends disambiguated terms based upon usage patterns observed (i.e., using either the user's contextual profile or both the user's contextual profile and the shared knowledge base, described in section 3.5.2.1). For example, as shown in the contextual SERL search UI (Figure 3.12, A), the exact Meaning (or dissimulated terms) for the "Java" search keyword are; a) island Indonesia, b) coffee, java, and c) platform-independent object-oriented. The recommended disambiguated term (denoted by  ) for the "Java" keyword is "island Indonesia". Similarly, the exact meaning (or dissimulated term) for the “surf” search keyword is the “surf, breakers, breakers”. The selected exact meaning terms (or disambiguated terms) are used by the Knowledge Base Query Formulator process to formulate a Boolean query.

ii. **Similar Phrase** (i.e. Meta Keywords) – This option allows the user to select a phrase similar to the entered query. This option uses the Meta keywords knowledge to prompt and recommend contextually similar phrases based upon usage patterns observed. For example, as shown in the contextual SERL search UI (Figure 3.12, B); the similar phrase (or Meta Keywords) and recommended similar phrase (denoted by ✓) for the “Java surf” search query is the “surf travel company java”. The selected similar phrase (or Meta keyword) is used by the Knowledge Base Query Formulator process to formulate a Boolean query.

iii. **Precise Category** (i.e., Concepts) – This option allows the user to select a precise category for the entered query. This option uses ontology knowledge to prompt and recommends related concepts based upon usage patterns observed. For example, as shown in the contextual SERL search UI (Figure 3.12, C); the precise category (or concept) for the “Java surf” search query is the “Surfing”. The selected precise category (or concept) is used by the Knowledge Base Query Formulator process to formulate a Boolean query.

The registered user has options to configure the contextual SERL search features (Figure 3.13) and delete or share his/her personal contextual profile (Figure 3.14) to build a shared contextual knowledge base. The configuration option allows the user to specify their search context (i.e., either “office” or “home”), query formulation option (i.e., either automatic query formulation or guided query formulation), search duration option (i.e., either last week, last month or last six months) and use of shared knowledge based option (i.e., either “yes” or “no”). The delete or share option allows the user to delete or share their information seeking behaviour explicitly. Finally, the contextual search returns contextual results as has been previously illustrated in Figure 3.14. Each returned item (i.e., URL) is labelled with the number of hits based on the user’s previous search activities, or those of other users who are sharing their activities in the shared contextual knowledge base. These labels provide feedback on how many times the items have been visited either by the user themselves or by the other users. The various configuration features and contextual SERL search results are discussed in section 3.5.2.4 and 3.5.2.3 respectively.

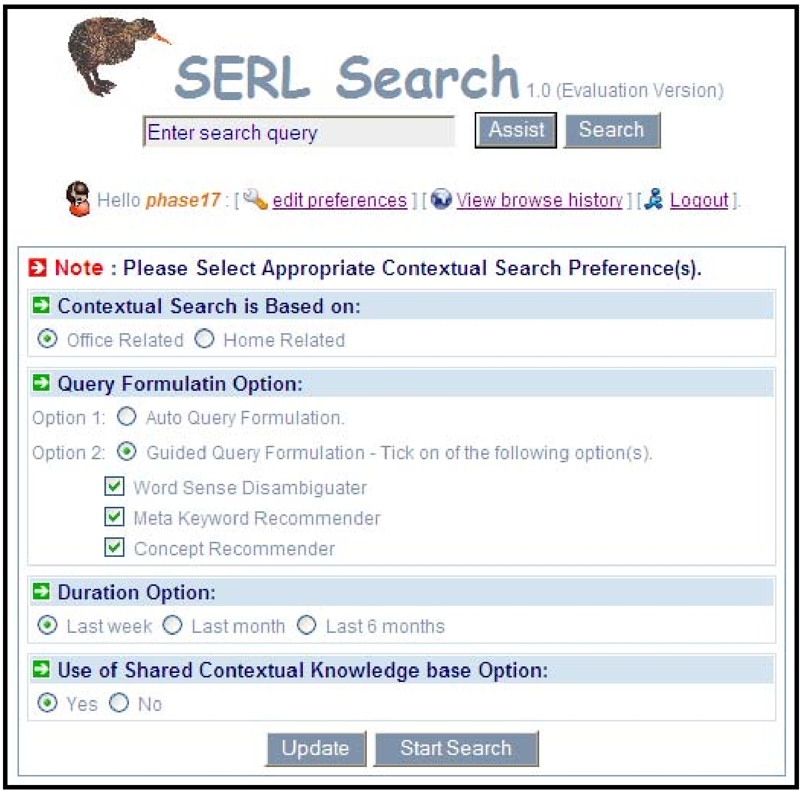


**Figure 3.13: Updating user profiles in the contextual SERL search.**

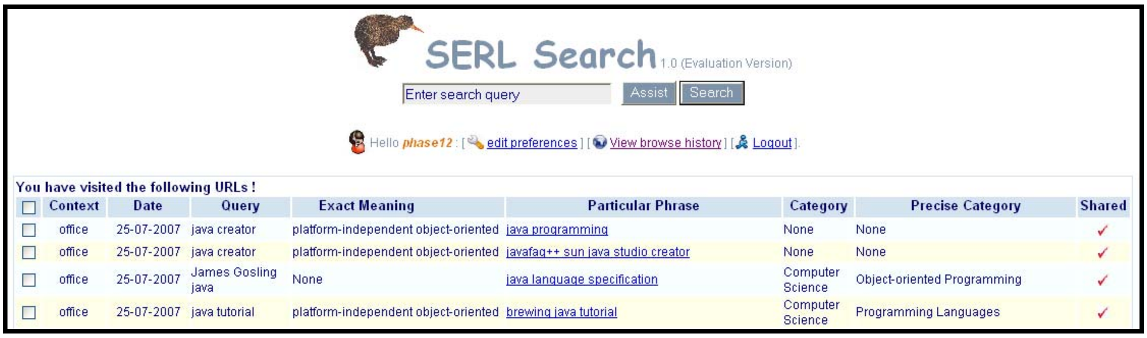


**Figure 3.14: Viewing visited URLs in the contextual SERL search.**

The following section describes how the UI and application logic are integrated to support various contextual SERL search features.

## 3.5.2 Application Layer

The application layer is mainly responsible for handling application logic (e.g., monitoring and capturing users' information seeking behaviour and providing relevance feedback to the users), managing transactions (e.g., storing the users' information seeking behaviour data into the database) and allowing interfaces for interaction with other layers (e.g., retrieving and presenting stored data to the users). The entire application logic is implemented using the Java programming language and deployed as a Web application in an Apache Tomcat servlet container. The application logic is designed and implemented using a component-based development methodology. The methodology has been successfully applied to a wide variety of areas including industrial process control system (Crnkovic & Larsson, 2000), adaptive e-learning (Brusilovsky, 2004), and knowledge discovery in bioinformatics (Etienne, Wachmann, & Zhang, 2006). The methodology reduced several recurring designs and made the contextual SERL search implementation work more efficiently. For example, it was much easier to handle errors/bugs related to specific components that make up a modular system than in a single, large monolithic system. It also improved the testability or the effort required to test if a component performs the intended function. In addition, the methodology allowed building various reusable components and rapid integration with existing pre-built components, such as Jena, WordNet and Google SOAP Search API to meet application requirements of the contextual SERL search.

The layer comprises two main modules: Profile Collector Module (PCM) and Context Manager Module (CTM) to perform the following functions:

i. Gather the user's implicit data, such as *entered search queries*, *visited URLs* and *extracted Meta keywords* from those visited URLs.

ii. Capture the user's explicit data, such as *alternative term/phrases, Meta keywords* or *similar phrases, ontology* and *concepts*. This data is sourced from a lexical database, a shared contextual knowledge base and domain-specific ontology/concepts.

iii. Construct the user's personal contextual profile (*U_profile*) and a shared contextual knowledge base (*Kb_profile*) using data from step i and step ii.

iv. Finally, modify the user's initial query to more accurately reflect the user's interests using steps i, ii and iii.

Each module consists of several components that perform various functions. Some of these functions are illustrated using simple pseudo-code listings and straightforward notations, when necessary. The following sub-sections describe the PCM and CTM module in more detail.

### 3.5.2.1 Profile Collector Module (PCM)

Figure 3.15 illustrates the functionality of the PCM, a hybrid contextual user profiling approach that captures a user's search behaviour by monitoring and capturing their explicit and implicit data. The PCM consists of two specialised components: Preference Collector (PC) and Behaviour Collector (BC). They act as front-end brokers, gathering contextual information from a user. The BC component monitors and captures the user's behaviour implicitly from their search activities. Similarly, The PC component learns a user's specific information needs by capturing their explicit preferences and at the same time recommends terms, phrases and concepts that will be of potential interest to the user. In addition, the PC component interacts with the BC component and Knowledge Collector component to construct an adaptive contextual user profile as well as a shared knowledge base. The PC component also interacts with the Knowledge Base Query Formulator component to formulate a Boolean query.

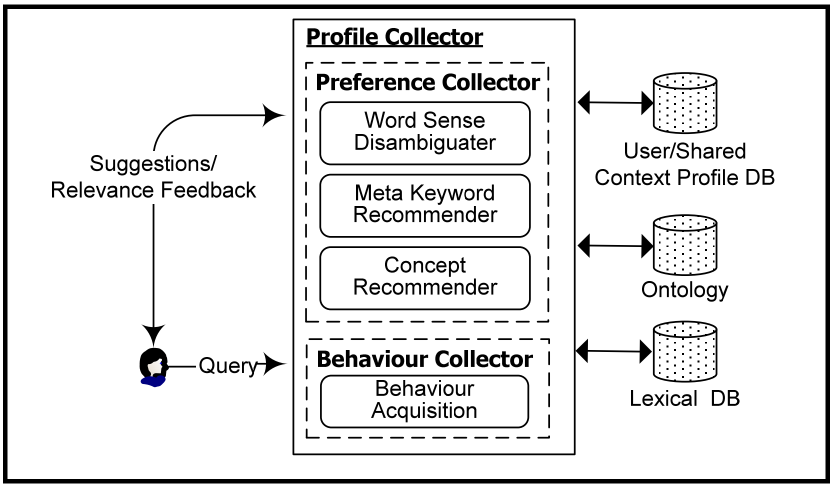


**Figure 3.15: Profile collector module functionality.**

From a user's perspective, the first step in building their personal contextual profile is to obtain a *username* and *password* through the user registration process, which prompts the user to enter minimum information necessary to create their account. The registration process automatically creates their default configuration profiles. For example, default search context is “home”, default query formulation option is “guided query formulation”, default search duration option is “last week”, default use of shared knowledge based option is “no”. However, each user has the option to update their configuration profile on an ongoing basis through the update profile process. In addition, the user has the option to maintain their captured contextual information through the manage profile process, which includes the capability to either share or delete their profile data.

A few assumptions were made during the development of the PCM. First, although users may be reluctant to provide their personal data due to privacy concerns (Markellos, Markellou, Panayiotaki, & Tsakalidis, 2006), we assume that finding relevant information on the Web more readily than at present is a sufficiently compelling reason to persuade users to register and disclose their personal information seeking data with the contextual search system. This assumption may appear too strong. However, Kobsa and Teltzrow (2005) have discovered that users are significantly more willing to share their personal data as long as they can perceive significant benefits from sharing their data. Many Web IR systems have explored various user modelling approaches to improve the Web experience for each user with respect to the user’s context so there appears to be a degree of willingness among uses to share data when sufficient assurances are provided (Kraft et al., 2005; Shen et al., 2005; Teevan et al., 2005).

Second, users’ clicked URLs (or visited URLs) from the returned results are considered to be relevant to the entered query (or reflect their search intent). Various studies (Cui et al., 2002; Wen, Nie, & Zhang, 2001) suggest that the user's choice or clicked URL does suggest a certain degree of relevance. Even if some of the clicked URLs are irrelevant or erroneous, in the contextual SERL search the users have the option to remove these URLs from their profiles so that it more accurately represents their search intent.

Third, we assume that users are willing to provide explicit relevance feedback to reflect/clarify their search intent or information needs. Again, this assumption may appear strong as users are often reluctant to make the extra effort to provide explicit relevant feedback (Kelly & Teevan, 2003). However, various researchers have found that users exhibit a desire for explicit feedback features and, in particular, term suggestion features (Beaulieu, 1997; Belkin et al., 2000; Hancock-Beaulieu & Walker, 1992).

In addition to these assumptions, there are a few known functionality limitations and technical challenges in the PCM that need to be further investigated in future research. First, only two categories of context; ‘home’ and ‘office’, were defined in the contextual SERL search, although the number of categories can be more than two (e.g., personal, family, baby, mother and so on). As a proof of concept, the two categories are considered sufficient for this study. The categories are defined to deal with frequently changing user interests (or multiple interests) that conceptually represent their information needs, uses, or desires while searching information on the Internet.

Second, only two ontologies or domain hierarchies; computer science and travel, were used in the contextual SERL search, though there are no barriers to adding further ontologies (e.g., shopping, music, sports, news and so on). Furthermore, only part of the domain hierarchies’ classes/concepts were used in the PCM. In terms of demonstrating proof of concept, the two categories and part of the domain hierarchies’ classes/concepts should be sufficient.

The following sub-sections describe the PC and BC components of the profile collector module in more detail.

***Preference Collector (PC)***

Figure 3.16 provides a summarised depiction of the functionality of the PC component, consisting of Word Sense Disambiguater (WSD), Meta Keyword Recommender (MKR) and Concept Recommender (CR) processes. The component learns a user’s specific information needs by capturing their explicit preferences and at the same time recommends terms/phrases/concepts that will be of likely interest to the user.

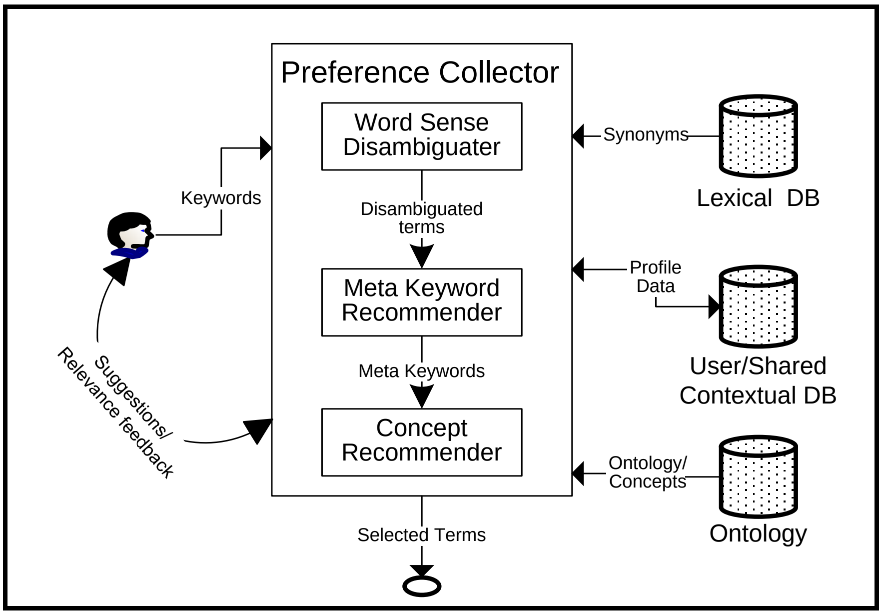


**Figure 3.16: Preference collector component functionality.**

The PC component employs the nearest-neighbour algorithm to learn a user's specific information needs and to provide alternative terms recommendation. The nearest-neighbour approach has previously been applied successfully to classify/recommend relevant information to users including adaptive nearest-neighbour (Ku, Zimmermann, Wang, & Wan, 2005), nearest-neighbour search (Tešic & Manjunath, 2003), and query chains (Radlinski & Joachims, 2005). The component uses a two step 'divide and conquer' algorithm to address the nearest neighbour's scalability problems and poor recommendation issues.

i. Firstly, the algorithm uses the *number of hits*[21] parameter together with other computation parameters, such as search query ($q_0$), user's context ($ct_0$) and dissimulated terms {$d_0$, $d_1$}/Meta keywords or phrases {$m_0$, $m_1$}/concepts {$c_0$, $c_1$}, to cluster a neighbourhood of users that in the past have exhibited similar information seeking behaviour (e.g., entered same type of queries, used same type of context, selected same type of terms, visited same URLs etc.).

ii. Secondly, the nearest-neighbour search is limited within the neighbourhood to analyse and identify a set of top N (N<=5) terms that would be of interest to or liked by the user.

[21] A higher number of hits denotes greater relevance or higher similarity.

Table 3.3 shows the simple nearest-neighbour based recommendation process that identifies a set of *N* terms that would be of likely interest to the user.

**Table 3.3: Recommendation process.**

1. *Begin*
2. ***Input** contextual parameters (e.g., $q_0$, $ct_0$, etc.)*
3. ***Check** profile configuration*
4. *IF configuration set to 'shared'*
5.     ***Retrieve** data from database*
6.     ***Cluster** U_profile and Kb_profile records*
7.     ***Compute** nearest-neighbour on records (Kn)*
8. *ELSE-IF*
9.     ***Retrieve** data from database*
10.     ***Cluster** U_profile records*
11.     ***Compute** nearest-neighbour on records (Kn)*
12. ***Sort** Kn records by no. of hits*
13. ***Extract** Top-K terms from a top Kn records*
14. ***Return** Top-K terms*
15. *End*

The PC component takes the contextual parameters (step 2), such as entered query ($q_0$) and user's context ($ct_0$) to cluster and compute the nearest-neighbour on the *Kb_profile* (if the share feature is enabled) and *U_profile* records (step 5 - 7). It sorts and extracts and returns terms (step 12 - 14), e.g. the disambiguated terms {$d_0$, $d_1$}, from the *Top-K* records by calculating the highest no. of hits.

The PC component employs the relevance feedback approach to support the iterative development of a search query by recommending alternative terms (i.e., disambiguated terms/ Meta Keywords/concepts) for query formulation (Limbu, Pears, Connor, & MacDonell, 2006). However, the effectiveness of learning each user's specific information needs and alternative terms recommendation is directly proportionate to the availability and size of the user's personal contextual profile and shared contextual knowledge base. As such, in this component, the issue of cold start is not fully addressed.

As shown in Figure 3.16, the PC component starts with the WSD process by accepting a user's search query ($q_0$). The process uses WordNet to disambiguate each search keyword. WordNet has been used as a word sense disambiguation tool in queries (Liu et al., 2005), geographical information retrieval systems (Buscaldi et al., 2005), text-to-concept mappings (Bonino et al., 2005) and other such applications. During the disambiguation process, the WSD process corrects original word sense(s) of each search keyword by removing stop words, repeated or similar words, and extracting first three terms as disambiguated terms. For example, the original word senses from WordNet for keyword “java” are;

*i. Java (an island in Indonesia to the south of Borneo; one of the world's most densely populated regions)*

*ii. coffee, java (a beverage consisting of an infusion of ground coffee beans) "he ordered a cup of coffee"*

*iii. Java (a platform-independent object-oriented programming language)*

After the WSD correction process, the extracted disambiguated terms for the keyword “java” are;

*i. island in Indonesia*

*ii. coffee, java*

*iii. platform-independent object-oriented*

Table 3.4 presents the WSD process step by step. For example, the entered search query, $q_0$ (e.g., “java surf”) (step 2), contains two search keywords, $k_0$ (“java”) and $k_1$ (“surf”) (step 3), and both keywords can be disambiguated. Besides, each initial disambiguated term, $d_0$, may contain one or more words $\{w_0, ..., w_n\}$ (e.g., “Java (an island in Indonesia to the south of Borneo; one of the world's most densely populated regions)”) (step 7).

For each initial disambiguated term, the WSD process removes common stop words[22] $\{s_0, ..., s_n\}$ (e.g., a, an, and the etc.), repeated/similar words $\{sw_0, ...,$

[22] Stop words, or stopwords, is the name given to words which are filtered out prior to, or after, processing of natural language data (text). This stop words list is taken from http://www.softexe.com/askw-stop-words-list.html.

$sw_n$}, and similar-to-search-query keywords ($k_0$, $k_1$) from each disambiguated term {$d_0$, $d_1$} (step 8 - 10). In the end, the process extracts only the first three words from each disambiguated term, for example {$w_0$, ..., $w_3$} for each disambiguated terms, and constructs a disambiguated terms vector (i.e., D) (step 11 – 12). Thus, the corrected disambiguated terms for $k_0$ (“java”) and $k_1$ (“surf”) are $d_0$ (“island Indonesia”), $d_1$ (“coffee, java”), $d_2$ (i.e., “platform-independent object-oriented”) and $d_4$ (i.e., “surf, breakers, breakers”).

**Table 3.4: WSD disambiguation process.**

1. Begin
2. ***Input*** *search query (e.g., $q_0$ )*
3. ***Extract*** *search keyword (e.g., $k_0$, $k_1$) from $q_0$*
4. ***Disambiguate*** *search keyword (e.g., $k_0$) into disambiguated term (e.g., $k_0 = \{d_0, d_1, d_2\}$ and $k_1 = \{d_3\}$)*
5. *WHILE disambiguated terms size >= 1 DO*
6.     *IF disambiguated term (e.g., $d_0$ ,$d_1$) exist*
7.         ***Parse*** *$d_0$ into token of words (e.g., $d_0 = \{w_0, w_1\}$)*
8.         ***Remove*** *stop words (e.g., $s_0$, $s_1$) from $w_0$ & $w_1$*
9.         ***Remove*** *repeated/similar words (e.g., $sw_0$, $sw_1$) from $w_0$ & $w_1$*
10.         ***Remove*** *similar query keywords (e.g., $k_0$) from $w_0$ & $w_1$*
11.         ***Extract*** *first three words from refined $d_0$*
12.         ***Construct*** *disambiguated vector D*
13.         ***Return*** *D*
14.     *ELSE-IF*
15.         ***Return*** *‘None’*
16. *END-WHILE*
17. ***Return*** *result*
18. **End**

The WSD process presents only the first two search keywords’ disambiguated terms to the user. For example, if there are two keywords $k_0$ and $k_1$ then the corresponding terms {$d_0$, $d_1$, $d_2$} and {$d_3$} are presented. The main reasons for removing the above information and limiting the presentation to three words per disambiguated term, and to two disambiguated terms, are two-fold. The first was to avoid cluttering the user interface with data that is less relevant to the user. Secondly, it improves the performance of a query by reducing the cost of the query execution process in the server. Various existing query expansion approaches (Bai, Song, Bruza, Nie, & Cao, 2005; Cui et al., 2002; Liu, Jin, & Chai, 2006; White & Marchionini, 2007) have also removed stop words to limit the number of keywords and at the same time to increase the effectiveness of query expansion approaches.

The WSD process then uses a recommendation approach (Table 3.3) to compute the nearest-neighbour in the user's personal contextual profile and the shared contextual knowledge base to recommend disambiguated terms that are relevant to the user's search query ($q_0$). The user has the option to alternatively select more relevant disambiguated terms that better describe the subject of their query. The user selected disambiguated terms $\{d_0, d_3\}$ are later used by the MKR and CR processes. The MKR and CR processes also utilise the recommendation process (Table 3.3) to recommend Meta keywords $\{m_0, ..., m_n\}$ and domain-specific concepts $\{c_0, ..., c_n\}$. However, the user has the option to select one Meta keyword and one domain-specific concept. The CR process uses domain specific ontologies to expand the user's initial search query. Existing systems, such as OntoBroker (Fensel et al., 1998), RUBIC (Klink et al., 2002), and WebSifter II (Kerschberg et al., 2001) have used publicly available ontologies to extract additional query terms/concepts.

Finally, as shown in Table 3.5, the PC component stores all of this information/these parameters as a user's preference data for future use.

**Table 3.5: Example of user's preference data.**

| Query | Context | Date | Disambiguated Terms | Meta Keywords | Concepts |
|---|---|---|---|---|---|
| $q_0$ | $ct_1$ | $dt_1$ | $\{d_0, d_1\}$ | $m_1$ | $c_1$ |
| $q_1$ | $ct_1$ | $dt_1$ | $d_3$ | $m_2$ | $c_2$ |
| $q_2$ | $ct_1$ | $dt_1$ | $d_4$ | $m_3$ | $c_3$ |

For example, for a query $q_0$, the selected disambiguated terms are $\{d_0, d_1\}$, selected Meta keyword is $m_1$ while the selected concept is $c_1$. The context, $ct_1$, is taken from the user's configuration profile [23] and the current date and time, $dt_1$, is taken from the system clock. This data provides a clear indication of a user's search intent, as each entered query together with the search context are closely mapped with various selected terms. As data grows, this data can be used to represent the user's evolving search intents. However, this data can be insufficient or only partial data, if the user ignores to select the query expansion

[23] The user's configuration profile is created when he/she registers with the contextual SERL search.

terms. As such, to obtain better insight into the user's search interests and build the contextual user profile it is not sufficient for the system to only look at the user's preference data, but look at the user's behaviour data.

***Behaviour Collector (BC)***

Figure 3.17 provides a summarised depiction of the functionality of the BC component, centred on a Behaviour Acquisition (BA) process. The BA process monitors and captures a user's daily Internet search activities as a user's behaviour.

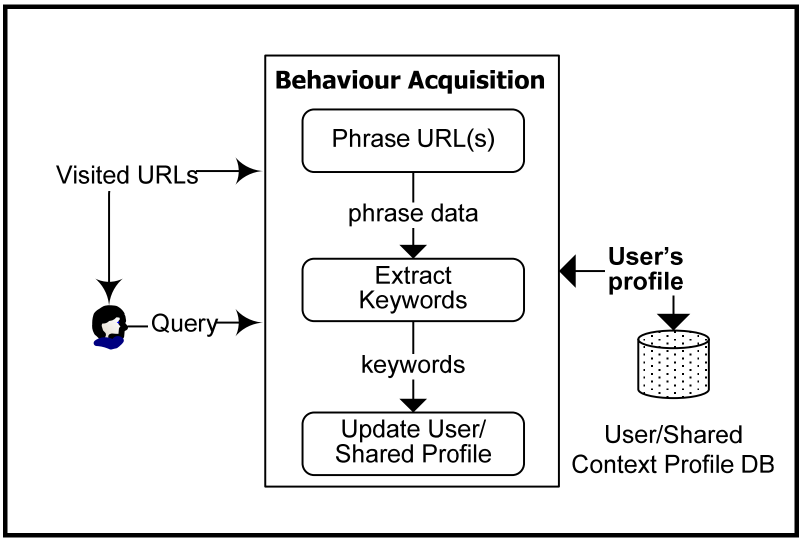


**Figure 3.17: Behaviour acquisition functionality.**

Table 3.6 shows the BA's keyword extraction process that extracts the Meta keywords $\{e_0, ..., e_n\}$ from the visited URLs $\{u_0, ..., u_n\}$.

For example, a user submits a search query, $q_0$, through the contextual SERL search and clicks on the URL(s) $\{u_0, u_1\}$ that are relevant to the query. The BA process computes and extracts a set of Meta keywords $\{e_0, ..., e_n\}$ from each visited URL. The process also removes stop words $\{s_0, ..., s_n\}$, repeated/similar words $\{sw_0, ..., sw_n\}$, and similar-to-search-query keywords $(k_0, k_1)$ from the extracted Meta keywords $\{e_0, ..., e_n\}$. Each Meta keyword may contain one or more words. The process extracts only the first five words from each Meta keyword, for example $\{w_0, ..., w_5\}$ for $e_1$. These Meta keywords are used to construct a Meta keyword vector, M. In addition, the process presents only the first five Meta keywords $\{e_0, ..., e_5\}$ to the user. The main reason for removing all further information, limiting the extraction to five words per Meta keywords and the presentation of the five Meta keywords is to limit the number of keywords/terms in Boolean query and to reduce clutter in the user interface. However, the need for further testing for the effectiveness of number of words in per Meta keywords is acknowledged, this has not been tested practically to any great extent at this stage.

**Table 3.6: Keywords extraction process.**

```
Begin
Parse the visited URL {u_0,… , u_n}
Extract for Meta keywords {e_0,… ,e_n}
IF m_0 exist
    Parse m_0 into token of words (e.g., m_0= {w_0, w_1})
    Remove stop words (e.g., s_0, s_1) from w_0 & w_1
    Remove repeated/similar words (e.g., sw_0, sw_1) from w_0 & w_1
    Remove similar query keywords (e.g., k_0, k_1 ) from w_0 & w_1
    Extract first FIVE words from refined e_0
    Construct Meta keyword vector M
    Return M
  ELSE-IF
    Extract for Title (e.g.,T)
    IF T exist
        Parse T into token of words (e.g., T= {w_0, w_1})
        Remove stop words (e.g., s_0, s_1) from w_0 & w_1
        Remove repeated/similar words (e.g., sw_0, sw_1) from w_0 & w_1
        Remove similar query keywords (e.g., k_0, k_1 ) from w_0 & w_1
        Extract first FIVE words from refined T
        Construct Meta keyword vector M
        Return M
    ELSE-IF
        Return 'none'
Return result
End
```

Finally, as shown in Table 3.7, the BC component stores all this information/these parameters as a user's behaviour data for future use.

**Table 3.7: Example of user's behaviour data.**

| Query | Context | Date | Visited URLs | Meta Keywords |
|---|---|---|---|---|
| $q_0$ | $ct_1$ | $dt_1$ | $u_1$ | $e_1$ |
| $q_0$ | $ct_1$ | $dt_1$ | $u_2$ | $e_2$ |
| $q_1$ | $ct_1$ | $dt_1$ | $u_3$ | $e_3$ |
| $q_1$ | $ct_2$ | $dt_2$ | $u_4$ | $e_4$ |
| $q_1$ | $ct_2$ | $dt_2$ | $u_5$ | $e_5$ |

For example, for a query $q_0$, the context is $ct_1$, the visited URLs are $\{u_1, u_2\}$, and the extracted Meta keyword for $u_1$ is vector $e_1$ and for $u_2$ is $e_2$. The context , $ct_1$

is taken from the user's configuration profile and the current date and time (e.g., $dt_1$) is taken from the system clock.

### 3.5.2.2 Context Manager Module (CMM)

Figure 3.18 illustrates the functionality of the CMM, which performs two main functions. Firstly, it interacts with the PCM to build a user's personal contextual profile and a shared contextual knowledge base, while respecting the user's personal privacy. Secondly, it performs iterative query expansion using the PCM's relevance feedback function (Limbu, Connor, Pears, & MacDonell, 2006). The CMM consists of a number of specialised components: Knowledge Collector (KbC), Customisation Utility (CU), KB Query Formulator (KbQF) and Result Analyser (RA).

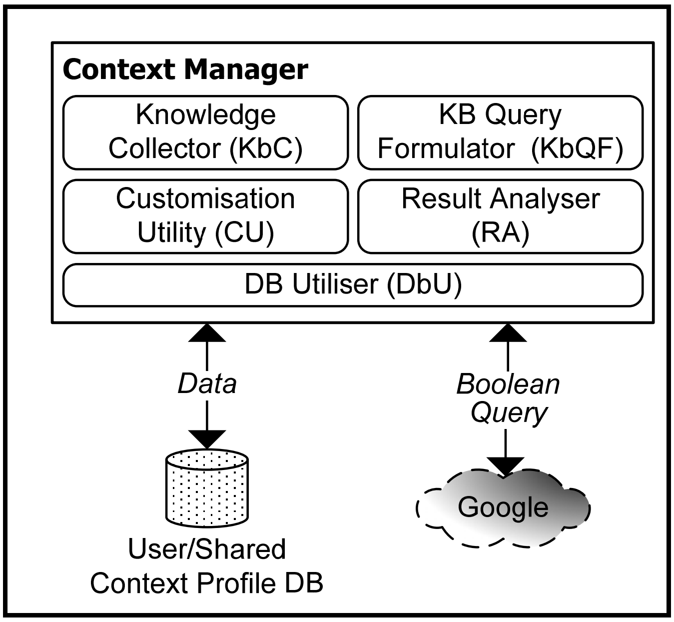


**Figure 3.18: Context manager module functionality.**

There are a few known functionality limitations and technical challenges in the CMM. Firstly, it is known that users are often reluctant to make the extra effort to provide explicit relevant feedback. As a result, building the user's personal contextual profile and the shared contextual knowledge base is a challenge as the system requires a large number of such profiles to train the nearest-neighbour classifier. Secondly, data mining may also present ethical challenges as information on individual users' information seeking behaviour is scrutinised. As such, security and privacy are major issues that warrant separate and extensive consideration. Many users are reluctant to give away personal information either implicitly or explicitly. In both cases, the user loses anonymity and is aware that all of their actions will be recorded and used, perhaps often without consent. As such, privacy violation during the user profiling process may be encountered (Volokh, 2000). Thirdly, the CMM uses a free Google SOAP-based API for accessing Google's index to return search results. There are some limitations associated with the API. There is a limit of one thousand queries per day and a limit of ten results per query. There is also a limit of a maximum number of ten words per query. Finally, though the need for scalability in the CMM is acknowledged, this has not been tested practically to any great extent at this stage. The following sub-sections describe the KbC, KbQF, CU and RA components in more detail.

***Knowledge Collector (KbC)***

Figure 3.19 provides a summarised depiction of the functionality of the KbC component. It is implemented primarily to address ***research objective II*** by facilitating the construction of a personal user profile and collection of multiple users' personal contextual profiles to build a shared understanding of contextual knowledge base.

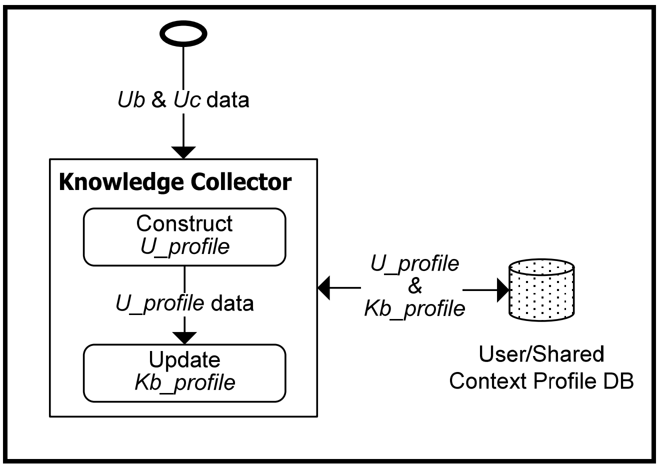


**Figure 3.19: Knowledge collector functionality.**

The KbC interacts with the PCM to gather a user's information seeking data (i.e., user's behaviour and preferences data) which are stored in the personal contextual user profile and which update the shared contextual knowledge base in two ways;

i. It increments/updates the number of hits (n +1, where n is number of hits) by one, in the event that same information seeking data was discovered.

ii. Otherwise, it stores the information seeking data as a new record.

The personal contextual user profile information can be used to present interests and preferences of the user over differing timescales. Both the personal contextual user profile and the shared contextual knowledge base could be used to suggest or recommend disambiguated terms, Meta keywords, ontology and concepts to other users with similar contextual profiles.

### *Knowledge Base Query Formulator (KbQF)*

Figure 3.20 provides a summarised depiction of the functionality of the KbQF component. It is implemented primarily to address ***research objective III*** by expanding a simple keyword query into a contextual Boolean query in order to improve the results of that query.

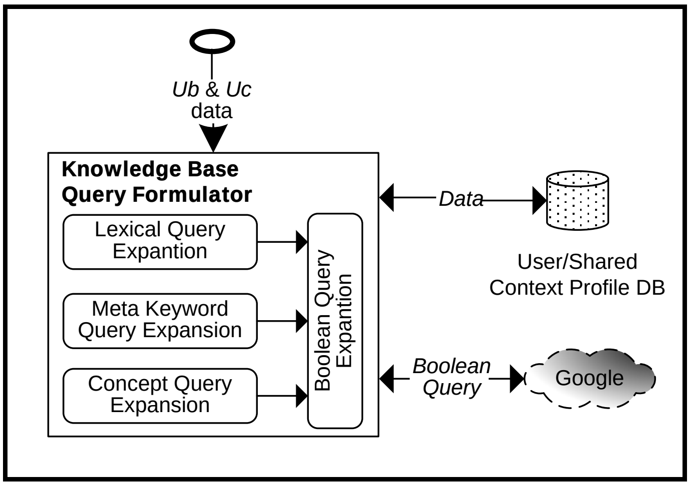


**Figure 3.20: Knowledge base query formulation functionality.**

The KbQF component interacts with the PCM to formulate a contextual Boolean query. It addresses some of the query expansion challenges, especially, the terms to be included in the query expansion and how these terms are ranked or selected, by employing both interactive query expansion (IQE) and automatic query expansion (AQE). For the IQE, the KbQF component collaborates with the PCM's relevance feedback (RF) function to obtain appropriate query expansion terms (i.e., terms, phrases and concepts). The RF technique has been successfully applied to improve the search results quality in a wide variety of problems including interactive text-based image retrieval (Zhang, Chai, & Jin, 2005), adaptive Web search (Sugiyama et al., 2004), content-based music retrieval (Hoashi, Matsumoto, & Inoue, 2003), misuse detection in information retrieval systems (Ma & Goharian, 2005), ARCH (Sieg et al., 2004) and so on. Similarly, for the AQE, the KbQF component automatically identifies expansion terms using the available knowledge about a user's interest from the user's personal contextual profile and shared contextual knowledge base.

Finally, the KbQF component uses these expansion terms to formulate a Boolean search query for submission to a search engine. The KbQF's simple Boolean query expansion technique is as follows: With all query expansion terms;

$$q_m = q_0 \; AND \left(d_1 \; OR \; d_2\right) \; AND \; \left(c_1\right) OR \; \left(m_1\right) \quad (1)$$

With disambiguated terms/phases and selected concept information;

$$q_m = q_0 \; AND \left(d_1 \; OR \; d_2\right) \; AND \; \left(c_1\right) \quad (2)$$

In the above formulae (1 & 2), $q_m$ = modified query; $q_o$ = original query; $d_1$ and $d_2$ = disambiguated term(s); $m_1$ = a selected Meta Keyword; $c_1$ = a selected concept.

Using these simple formulae, the KbQF generates an enhanced (or expanded) Boolean search query. The enhanced query is said to represent the user's search intent more effectively and potentially improves recall and precision(Limbu, Pears et al., 2006). The KbQF adheres to the Google SOAP API limits of ten terms per query by limiting the number of expansion terms. However, the KbQF does not address the challenge of which levels of query reformulation should be automatic, interactive or manual.

### 3.5.2.3 Result Analyser (RA)

The RA component interfaces with the Google SOAP Search API, which has been used in numerous other studies (Curran & Doherty, 2006; Jain, Dahlin, & Tewari, 2005; Koo & Skinner, 2003) to perform general search queries. The RA goes beyond providing or presenting search results from the Google SOAP Search API by performing an on-the-fly analysis and ranking the results based on a user's personal contextual profile and a shared contextual knowledge base. Figure 3.21 and Table 3.8 shows the RA's analysing and ranking process.

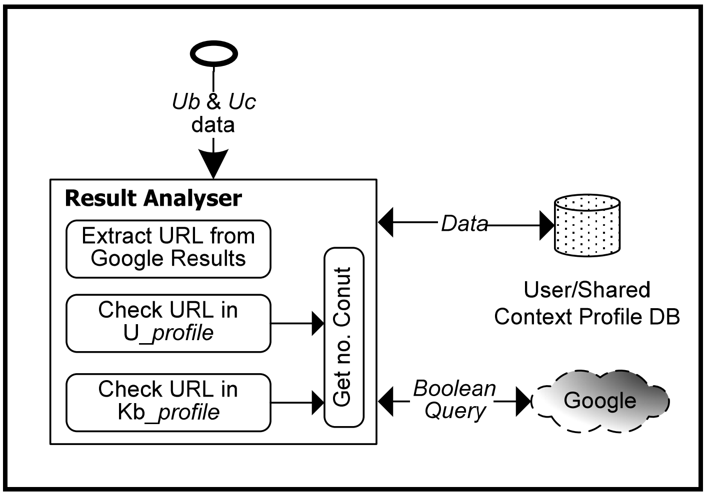


**Figure 3.21: Result analyser process.**

The RA extracts URLs from the Google results, checks if the URL exists in the user's contextual profile or the shared contextual knowledge base, and then returns the number of hits separately if the URL exits in either or both profiles. In this way the user is informed as to how many times the URL has been visited either by them or other users (Figure 3.7).

**Table 3.8: Analysing & ranking process.**

1. **Begin**
2. ***Input*** *Google results (g)*
3. ***Extract*** *URLs (u) from g*
4. ***Perform*** *Table 3.2 step 2 to 11*
5. ***Compare*** *records between Kn and u*
6. *If Kn record = u record*
7. *Retrieve no. of hits*
8. *Rank URL (Rn)*
9. ***Return*** *ranked and non-ranked URLs*
10. **End**

The RA uses the number of hits information to extract *Top-K* search queries similar to current query and associated *Top-K* visited URLs. Table 3.9 shows the RA's similar search queries and associated URLs extraction process.

**Table 3.9: Similar search queries and associated visited URLs extraction process.**

1. **Begin**
2. ***Perform*** *Table 3.2 step 2 to 12*
3. ***Retrieve*** *Top-K similar search queries and associated URLs*
4. ***Return*** *Top-K search queries and ranked URLs*
5. **End**

The RA retrieves and returns similar search queries and associated visited URLs from the Top-K records (step 2 - 3). The RA uses these information to suggest similar search queries and associated visited URLs to the user.

### 3.5.2.4 Customisation Utility (CU)

The CU component is a relatively simple component used for configuring various features of the contextual SERL search. Table 3.10 provides information about the underlying available configuration options for the contextual SERL search.

**Table 3.10: Configurations options for the contextual SERL search.**

| Option(s) | Description |
|---|---|
| Context | This option specifies a user's search context, such as 'office' or 'home'. |
| Query Formulation | This option specifies query formulation alternatives, such as auto query formulation and guided query formulation.<br>• Auto query formulation: This option automatically formulates a Boolean query based on either a user's personal contextual profile or together with a shared contextual knowledge base.<br>• Guided query formulation: The option allows the user to formulate a Boolean query manually by using one or a combination of WSD, MKR and CR features. |
| Duration | This option restricts recommendation terms and URLs to the last week, last month, or last six month periods. |
| Use of Shared Knowledge Base | This option specifies either to use or not to use the shared contextual knowledge base. If it is specified to 'No', the system uses only the personal contextual user profile to recommend terms and URLs. |

### 3.5.2.5 Database Utility (DbU)

The DbU component interfaces between the application layer and the database layer. The DbU implements the four essential database operations of persistent storage: create, read, update and delete (CRUD) to support the application requirements of the contextual SERL search. The DbU separates data persistence components and isolates the contextual SERL search business logic from database dependencies, so that the data source can be changed without requiring modification of the application logic.

### 3.5.3 Database Layer

The contextual SERL search's database layer is responsible for representing and storing information needed for the contextual SERL search. This layer is connected to four different data sources: MySQL, Google, a Lexical database (i.e., WordNet) and a domain specific Ontology (described previously in Preference Collector Section).

Figure 3.22 shows the contextual SERL search's physical database design.

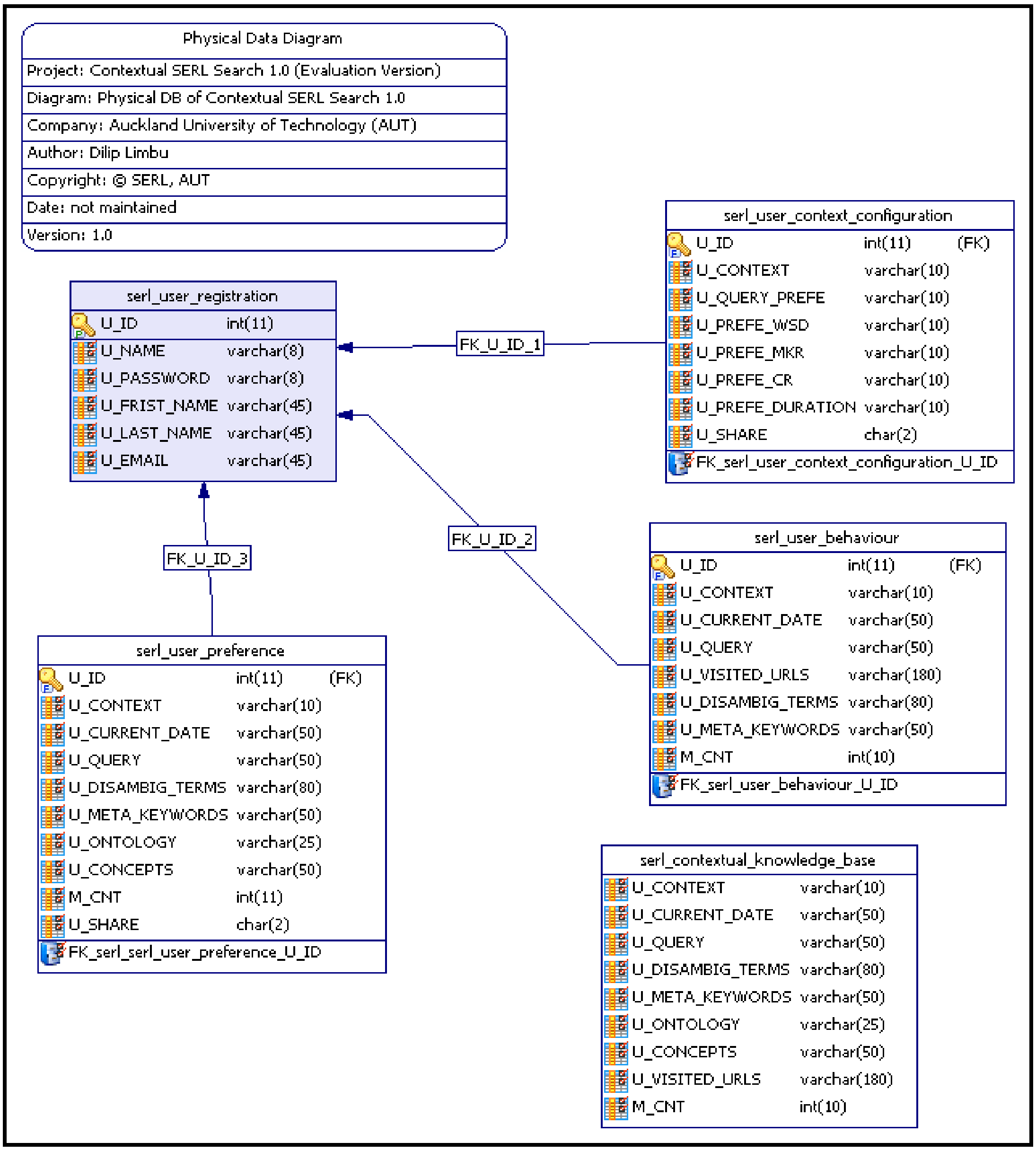


Figure 3.22: Physical data diagram of the contextual SERL search.

The MySQL database is used to store all contextual profile information, including users' registration information, users' contextual profiles, default profile configuration information, and shared contextual knowledge base. The design consists of five tables:

i. *serl_user_registration – stores users' registration information.*
ii. *serl_user_context_configuration – stores users' default profile configuration information.*
iii. *serl_user_behaviour – stores users' behaviour profile information.*
iv. *serl_user_preference – stores users' preference profile information.*
v. *serl_contextual_knowledge_base – stores shared contextual knowledge base information.*

The physical database design represents the tables and the relationships between them. The design specifies the data type used for each column in the table, and determines how tables will be stored in the database.

## 3.6 Chapter Summary

This chapter has presented a framework for CIR from the Web, i.e., the contextual SERL search that is intended to make search results more relevant. The contextual SERL search utilises various approaches/techniques to address some of the many acknowledged challenges that exist in the CIR domain (described in Chapter 2). The SDRM and the DS research guidelines are employed to investigate, design, develop and implement the contextual SERL search. The popular three-tier architecture and various design and implementation considerations are made during the development of the contextual SERL search.

The contextual SERL search centres on the construction of a user's personal contextual profile by combining user behaviour, user preferences and shared contextual knowledge base information. The shared contextual knowledge base can be used to provide user feedback/suggestions and to refine search queries. The framework requires the collection of multiple of users' personal contextual profiles. All of these components are then integrated in a single comprehensive

CIR framework. These features contribute to making this framework open, robust and scalable.

In the next chapter the details of the experimental methods and instruments used to evaluate the contextual SERL search (described in this chapter) are presented.

# Chapter 4

# Experimental Methodology

## 4.1 Introduction

Chapter 3 presented a framework for CIR from the Web, i.e., the contextual SERL search. This chapter describes the experimental methods and instruments used to evaluate whether the contextual SERL search improves a user's information seeking ability to find relevant information from the Web.

According to Voorhees (2002), evaluation (or performance) of a search system may be measured over many different dimensions, such as how well the system can rank documents, economy in the use of computational resources, speed of query processing, or user satisfaction with returned search results. This study is primarily concerned with the effectiveness, the efficiency and the subjective satisfaction of using a contextual user profile and shared contextual profiles, that is, how well a given system can match and retrieve documents that are more useful or relevant to the user's information need using their personal or shared contextual profiles. This is difficult to quantify precisely as it involves more than just assigning some measure to the value of search results. As such, this study utilises both qualitative and quantitative data analysis methodologies, based on simulated work task situations, questionnaires, and observations. Both approaches are informed by earlier theories of cognitive and information-seeking behaviour, but many of the facets involved were undefined at the outset of the research.

The following five hypotheses are drawn to test the performance of the contextual SERL search along the usability dimensions of effectiveness, efficiency and subjective satisfaction in the returned search results.

**Find Information Readily *(Hypothesis 1)***

The contextual SERL search enables subjects to find relevant information ***more readily*** than a standard search engine using their personal profile and shared contextual knowledge base.

**Adaptiveness Support *(Hypothesis 2)***

The contextual SERL search adapts to the information needs of the searcher and facilitates effective recommendation of terms.

**Recommendation Support (*Hypothesis 3)***

The contextual SERL search eases the conveying preferences process and recommends relevant and useful terms.

**Query Formulation (*Hypothesis 4)***

The contextual SERL search facilitates easy, effective and reliable query formulation strategy.

**Interface Support *(Hypothesis 5)***

The interface support provided by the contextual SERL search facilitates effective information access.

These hypotheses examine the subjects' overall information seeking behaviour and their perceptions of the contextual SERL search and the contemporary search engine. This chapter begins by describing the two experimental phases and the experimental methodology including testing instruments and procedures used to evaluate the contextual SERL search.

## 4.2 Experimental Phases

A prototype of a contextual information retrieval (CIR) framework called contextual SERL search was implemented in the course of this research. Two experimental phases; component tests (CT) and observational study (OS), were carried out in order to test whether the contextual SERL search improves a user's information seeking ability to find relevant information from the Web in comparison to a contemporary search engine.

This study employed the mixed methods research approach (Tashakkori & Teddlie, 2003) to accomplish the evaluation goals. A qualitative study was conducted to investigate subjects' information seeking behaviour across typical search scenarios in depth. A quantitative study was conducted to determine the performance of the contextual SERL search system along the usability dimensions of effectiveness, efficiency and subjective satisfaction in the returned search results. Effectiveness was assessed by the accuracy with which subjects formulated queries using the contextual SERL search. Efficiency was assessed by the combination of ease and speed (i.e., the time taken to reach target information) with which subjects could find relevant information. Satisfaction was measured by the elicited subject opinion about the system during the experiment. As such, various testing instruments were used to evaluate the contextual SERL search using specific measurement and analysis techniques.

The CT experimental phase was carried out by the researcher. A group of 30 human subjects then participated in the actual OS experiment to evaluate the capabilities of the contextual SERL search and determine whether it meets required expectations. All participants were experienced users of search engines, searching for information on a daily basis. Appropriate ethical approval was sought from and granted by the Auckland University of Technology Ethics Committee (AUTEC) prior to commencing the OS experiment.

The remainder of the chapter describes the CT, OS experimental phases and the experimental methodology in detail.

### 4.2.1 Component Test (CT)

The objective of the CT experimental phase was to ensure that each system component of the contextual SERL search; namely Word Sense Disambiguation (WSD), Meta Keywords Recommender (MKR) and Concept Recommender (CR), would perform correctly under all of the conditions that the system could encounter in an actual OS experiment, and as much as possible in actual use. Prior to commencing the actual CT experiment on the contextual SERL search, several rounds of initial system component tests were carried out to assess the operation of the user interface, performance, reliability, availability and some functionality (i.e., relevance feedback support, Boolean query formulation support). During these tests, system component functionality defects were discovered and rectified. Significant time was spent on rectifying and fixing these defects and many component functionality improvements were made. Many performance enhancements were made, intended to make the contextual SERL search faster, reliable, effective, efficient and easier to use. The initial component tests not only rectified and fixed defects, but also acted as preparation for the actual CT and OS experiments.

There is no publicly available standard data set and evaluation procedure to evaluate such components as comprised the SERL search. As a result, during the CT experimental phase, a limited range of queries was executed against the system and the precision of the top K retrieved documents was used as the evaluation measure, where

$$Precision = \frac{\text{number of relevant documents retrieved}}{\text{total number of documents retrieved}}$$

The following sub-section describes the CT experimental procedures in detail.

#### 4.2.1.1 Procedure

Figure 4.1 shows the general overview of the CT experimental procedure. The contextual SERL search's components were tested using the following common experimental procedures as described below:

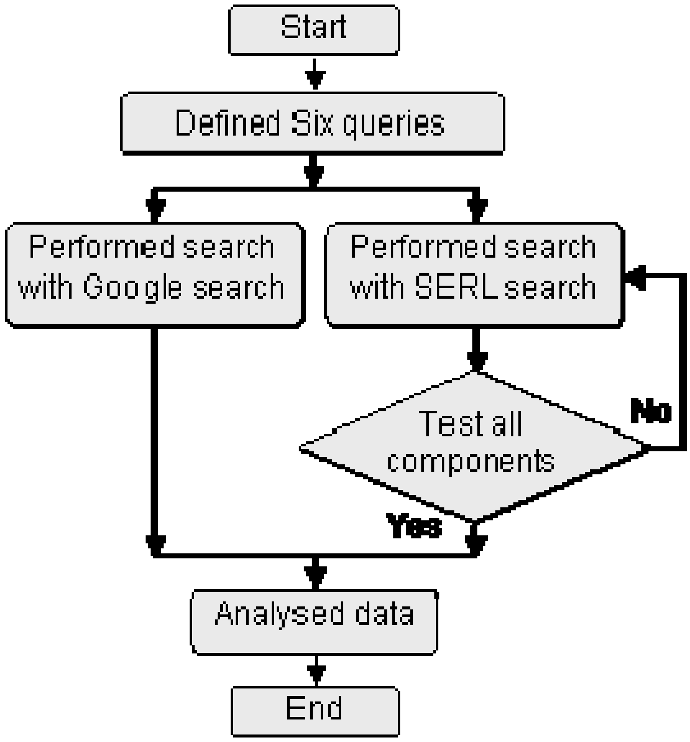


**Figure 4.1: Overview of component test procedure.**

i. Six search queries; three related to computer science domain and three related to travel domain, were defined (Appendix B, page 190).

ii. Each query was submitted to the Google search engine. The first ten hits from the returned search results were manually examined to gather precision data as most people initially select results from the first page when using a search engine (Eastman & Jansen, 2003; Hölscher & Strube, 2000; Jansen et al., 2000).

iii. Similarly, each query was then entered to the contextual SERL search and using one component at a time, an appropriate (or relevant) term was selected to formulate a Boolean query. The formulated Boolean query was then submitted to the contextual SERL search. The first ten hits from the returned search results were manually examined to gather precision data. Each of the three components; WSD, MKR and CR, were used individually, and in turn.

The CT experiment results are presented in Chapter 5. These results provide a quantitative single-value summary of a document ranking relative to a query, i.e., precision. This provides a useful degree of reassurance that the system functions as intended in terms of the functionality and performance of its constituent parts. It does not provide a sufficient evaluation of functionality in use, however. Numerous studies (Bar-Ilan, 1998; Mandl, 2006; Shafi & Rather, 2005) in the literature have explored the applicability of traditional IR evaluation criteria, i.e., precision and recall, on search engine performance. In addition, numerous studies (Eastman & Jansen, 2003; Joshi & Aslandogan, 2006) have used the first ten hits from the returned search results to examine the precision data. However, a measure such as recall is not a valid indicator of users' satisfaction, since the user of an actual system cannot judge recall consistently. Similarly, average precision is derived from recall, and suffers the same problem. Therefore, the observational study experiment was carried out to enable a more comprehensive evaluation from the perspective of the system's intended users. The following section describes the observational study experimental procedure in detail.

### 4.2.2 Observational Study (OS)

The main objective of the OS experiment was to measure the performance of the contextual SERL search system along the usability dimensions of effectiveness, efficiency and subjective satisfaction in the returned results in comparison to a contemporary search engine. Prior to commencing the actual OS experiment, all testing instruments were examined and two pilot studies were carried out on the contextual SERL search.

Six chosen volunteers (mostly postgraduate students at AUT) were asked to carefully examine all testing instruments, such as questionnaires and search tasks, in order to maximise participation response and research outcomes. The main objective was to detect any flaws in the questionnaires and the search tasks prior to their use in the actual experiments. The volunteers were requested to read the questionnaires and make comments. Similarly, they were asked to read each of the search tasks, place themselves in the simulated search scenario, and comment on the clarity and complexity of the search tasks. These comments were informal and as such are not reported in this thesis. However, they did motivate slight changes in the wording of the questionnaires and the search tasks. In addition, these testing instruments were revised several times and various issues related to item validity, reliability, appropriateness, adequacy and sufficiency were addressed.

The first pilot study was carried out with the same six volunteers. They were asked to perform a series of search tasks, and then compare the results achieved using the contextual SERL search with those achieved using their normal search engine. Four out of six users indicated that, for them, the contextual SERL search improved the effectiveness of the search query and improved precision in search results. The other two users stated that the system gave minor improvement in the effectiveness of the search query and the precision in search results. One of the users mentioned that he had obtained more results than the normal search engine. Another user stated that he had difficulties in finding relevant query expansion terms. On the whole, four users judged their searches with the system very successful and the other two considered their searches to have been fairly successful. They were also asked to make comments on the overall User Interface (UI) and usability of the contextual SERL search. Overall, the first pilot study identified some UI and usability issues that motivated changes to the UI and some features on the contextual SERL search.

The second pilot study was carried out with only two subjects (one subject and one of the researcher's supervisors) using the revised testing instruments, enhanced UI and improved features of the contextual SERL search. The outcome of the second pilot study motivated slight changes in the wording of the participant information sheet, observation sheet and also helped to redesign the evaluation procedures such as the sequence of activities, actual timing and the setup of experimental systems. All of the experiments described above were informal and so the results are not reported in this thesis. The above processes, i.e., testing instruments and two pilot studies, not only improved the testing instruments and the experimental system, but also prepared for the actual OS experiment. The following sub-section describes the OS experimental procedures in detail.

#### 4.2.2.1 Procedure

Figure 4.2 shows a general overview of the OS experimental procedure. The phase started with a posting request for participant advertisements, sending out anonymous questionnaires, and recruiting participants.

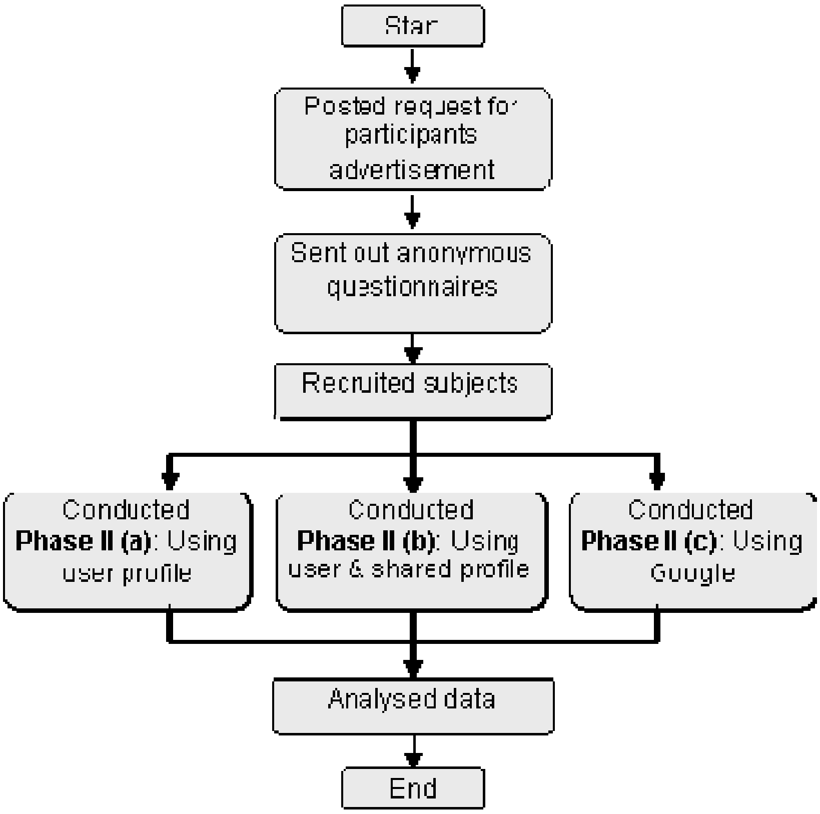


**Figure 4.2: Overview of observation study procedure.**

During the OS (or Phase II) experiment, three sub-phase experiments; Phase II (a), Phase II (b), and Phase II (c) were carried out. Each sub-phase experiment was conducted in a controlled environment where the researcher (as an observer) sat next to the subject's computer for the duration of the experiment. The researcher had little or no influence on the experiment in action, and observed rather than intervened. The time taken for a subject to complete the experiment ranged from between one and one-and-a-half hours.

The Phase II (a) and Phase II (b) experiments were carried out on the contextual SERL search, though their objectives were different. The aim of the Phase II (a) experiment was to determine whether the Contextual SERL search enables subjects to find relevant information, and to do so more readily than a standard search engine, using their personal contextual profiles. Throughout

Phase II (a), subjects performed six search tasks using the Contextual SERL search, and had their search behaviours and preferences captured in order to create their personal contextual profiles as well as a Shared Contextual Knowledge Base (SCKB). However, the SCKB was not accessible to them during their search.

Once Phase II (a) was completed, a second group of subjects repeated the same allotted six search tasks; however these subjects had the SCKB enabled. As such, the aim of the Phase II (b) experiment was to determine whether the Contextual SERL search enables subjects to find relevant information, and to do so more readily than a standard search engine, using their personal contextual profiles and the SCKB. This enabled an assessment of the contribution of the shared profile to (improved) search quality, by comparing the "speed" with which subjects could find data with the first group who did not have access to the shared profile. Throughout the Phase II (a) and Phase II (b) experiments, the following steps were followed;

i. Subject was welcomed and given introductory orientation about the experimentation.

ii. Subject was asked to read through the '**Participant Information Sheet'** and if he/she wishes to take part in the experiments, he/she was asked to sign a consent form.

iii. Subject was asked to complete the '**Entry Questionnaire'**.

iv. Subject was shown a short '**Video Demonstration'** on how to use the contextual SERL search.

v. Subject was given '**Six Search Tasks'** and given an opportunity to clarify any ambiguities. No time limit was set for any of the tasks.

vi. Subject was instructed to "**Talk Aloud**" while he/she engaged in the search tasks, i.e., he/she was asked to describe their actions and reasons for their actions. All their actions and reasons/comments (i.e., audio) were recorded and the researcher took notes. If a subject stopped talking, the observer would remind the subject to continue their narrative by showing a prewritten sign, "Please continue describing your search actions. Thank you."

vii. Subject was informed that his/her search activities were captured using the Camtasia[24] Screen Recorder software so that the time to complete each task could be determined.

viii. Subject was instructed to use the contextual SERL search's assist features if possible.

ix. Subject was instructed to change the search context as necessary – for example for *Topic A* it is 'office' and *Topic B* it is 'home'.

x. Subject was instructed not to enter information directly into the location bar of the browser (i.e., the text field in which URL's were displayed, and could be entered).

xi. Subject was informed that the "**Enter key**" of the keyboard was disabled.

xii. Subject was instructed to revise one of the previously attempted search tasks to check whether the system does as it is supposed to.

xiii. Subject was asked to complete the '**Post Observation Questionnaire**'.

xiv. At the end of the experiment, the subject was asked general questions about their search experiences with the contextual SERL search and the overall evaluation process.

The Phase II (c) experiment was carried out on a contemporary search engine, Google. As such, the experimental procedure of Phase II (c) was slightly different from that of the Phase II (a) and Phase II (b) experiments. The Phase II (c) experiment followed all the experiment procedures of the Phase II (a) experiment except step (iii), step (iv), step (viii), step (ix), step (xi), step (xii), which were related to the contextual SERL search and were thus ignored. The

---

[24] www.camtasia.com

following section describes the recruitment process of experimental subjects in detail.

## 4.3 Subjects

The experimental subjects were mainly staff, postgraduate and undergraduate students at the Auckland University of Technology (AUT). Previous research describing experiments regarding information seeking on the Internet have utilised different numbers of subjects, as shown in Table 4.1.

**Table 4.1: Number of subject involvement in other studies.**

| No. of Subjects | Study |
|---|---|
| 7 | (Liu et al., 2004) |
| 21 | (Kellar, 2006) |
| 22 | (Joachims et al., 2007) |
| 24 | (Harper & Kelly, 2006; White et al., 2002), |
| 30 | (Jansen, 2005) |
| 34 | (Choo et al., 2000) |
| 36 | (White & Marchionini, 2007) |

Since the main goal of this study was to gain a greater understanding of a user's information seeking behaviour and the usability of contextual SERL search in this regard, 30 subjects were to be recruited. Participation in this study was completely voluntary. All subjects had previous experience using Web search engines. The following sub-section describes how volunteers were recruited.

### 4.3.1 Recruitment Process

Subjects were recruited using emails, pigeon-hole drops and on-line advertisement per the ethics code of the School of Computing and Mathematical Sciences, AUT. An email letter was sent in advance to inform (or gain permission from) the staff of the School of Computing and Mathematical Sciences that the 'Anonymous Questionnaire' (Appendix A.1) and the 'Participation Request' (Appendix A.2) will be placed in their pigeon holes. More than eighty sets of both forms were placed in staff pigeon holes. In addition, in order to reach the wider community, an advertisement was placed in AUT's online information notice board with the approval of the appropriate site administrator.

Individuals were requested to complete the Anonymous Questionnaire that asked them about their previous Web searching experience. In addition, willing subjects were requested to indicate their interest and leave their email address if they wished to participate in actual experimentation by completing the Participation Request form. The indication of further interest was included with, but entirely separate from, the Anonymous Questionnaire form. Subjects were requested to return the completed questionnaire form and the filled participation request form either by dropping each in the respective ‘Drop In Box’ placed at the reception (or information counter) of the School of Computing and Mathematical Sciences or posting them ‘SEPARATELY’, i.e., using two separate envelopes; one for the questionnaire form and one for the participation request form, to the address stated on the questionnaire form (in order to retain anonymity in the questionnaire responses).

The above methods were followed up by an email letter sent to potential participants describing the research project and a URL address containing downloadable ‘Anonymous Questionnaires’ and ‘Participation Request’ forms . All these recruitment methods yielded a pool of 42 interested volunteers. All subjects who had indicated a willingness to participate in actual experimentation were invited to participate, up to a maximum of 30 subjects. Selected subjects were sent an email (on a first-come, first-served basis) asking them to respond via email their most convenient date and time. Finally, subjects were sent a confirmation email stating their scheduled experiment date, time and experiment venue. This constitutes the entirety of the recruitment process.

The participating subjects were classified into two groups; a) inexperienced and b) experienced, since there is evidence indicating that experienced and novice searchers conduct their searches differently (Höscher & Strube, 2000). The inexperienced subjects were infrequent searchers or those who had less than or equal to 3 years of Web searching experience. The experienced subjects were frequent searchers and had 4 or more years of Web searching experience. Table 4.2 shows the numbers of inexperienced and experienced subjects involved in the respective experimental phases.

**Table 4.2: Number of subject involvement in Phase II experiments.**

| Phase | System | Inexperienced | Experienced | Total |
|---|---|---|---|---|
| II (a) | Contextual Search | 5 | 5 | 10 |
| II (b) | Contextual Search | 4 | 6 | 10 |
| II (c) | Google | 5 | 5 | 10 |
| Total subjects | | | | 30 |

The next section describes the search tasks given to the experimental subjects.

## 4.4 Tasks

Simulated search tasks are intended to replicate an actual information seeking session and to facilitate realistic interaction in a laboratory setting. The method emphasises that searchers should be given search scenarios that encourage a real information seeking situation. Numerous experiments (Hoeber & Yang, 2006b; White & Marchionini, 2007; White, Ruthven, & Jose, 2005) have used a simulated search tasks approach. Several previous research studies have utilised a range of simulated search tasks, as shown in Table 4.3.

**Table 4.3: Number of search tasks used in other studies.**

| No. of Tasks | Study |
|---|---|
| 2 | (Hoeber & Yang, 2006a; Sihvonen & Vakkari, 2004) |
| 3 | (Tombros, Ruthven, & Jose, 2003; White et al., 2002) |
| 4 | (White, Jose, & Ruthven, 2003) |
| 6 | (White, Bilenko, & Cucerzan, 2007) |

In this study, six simulated search tasks (Appendix A.5) were defined. The tasks were categorised into two domains; a) *computer science domain*, and b) *travel domain*. These domains were selected because they represented a mixture of domains, mitigating the effect of subject's own domain knowledge specialties. White *et al*. (2003) proposed four Web search categories; *fact search, decision search, search for number of items,* and *background search*. Using only the first three Web search categories, each domain was further categorised into three different search tasks categories and carefully worded to ensure that these tasks were as realistic as possible. These search tasks were also designed using realistic search scenarios, known as simulated work task situations, as proposed in Borlund (2000; 2003). Table 4.4 presents an example of a simulated work task situation.

**Table 4.4: Example of simulated work task situations taken from Borlund (2000).**

| Simulated situation |
|---|
| **Simulated work task situation**: After your graduation you will be looking for a job in industry. You want information to help you focus your future job seeking. You know it pays to know the market. You would like to find some information about employment patterns in industry and what kind of qualifications employers will be looking for from future employees.<br><br>**Indicative request**: Find, for instance, something about future employment trends in industry, i.e., areas of growth and decline. |

The simulated work task situations used here served two main purposes: 1) they provided a short 'cover-story' or background information for the search tasks and led to cognitively individual information needing interpretations as in real life; and 2) they included an indicative request that illustrated how a search task might be initiated. Further, the simulated situations positioned the search within a realistic context, and control over the experiment was maintained by using the same tasks for all the subjects.

These simulated search tasks were in the form of questions, but were expressed in a slightly indirect manner, and asked for moderately obscure facts. The tasks required subjects to engage in more involved search behaviour than simply finding each answer as one of the first hits returned by a single query. However, these simulated search tasks were also designed so that the answers would not be so difficult to find that subjects would become frustrated. This balance was achieved through the informal pilot studies already discussed. The use of simulated search tasks and a controlled environment allowed the complexity of the experimental procedure to be managed consistently. The next section describes the questionnaires used in this study.

## 4.5 Questionnaires

Questionnaires are a research instrument consisting of a series of questions and other prompts for the purpose of gathering information from subjects. Various IR experiments (Dumais et al., 2003; Joachims et al., 2007; White & Kelly, 2006; White & Marchionini, 2007) have shown the usefulness of questionnaires to elicit subjects' opinions during the experiment.

In this study, two different questionnaires; a) anonymous questionnaire and b) observation questionnaire, were designed based on the research questions that motivated the study and experiments. A three-page anonymous questionnaire (as a survey instrument) was designed to gather general information regarding users' information-seeking behaviour and their preferences in this respect. An anonymous questionnaire was considered to be more appropriate for this kind of survey as compared to interviews. It enabled responses to be gathered from a relatively large population within an acceptable time frame. It was also believed that the questionnaire would provide an opportunity for respondents to answer at their own leisure and with no sense of a need to conform to pre-conceived expectations (due to the anonymity of the instrument).

In contrast, the observation questionnaires were designed to elicit each subject's opinions in relation to the experiment. The observation questionnaires were split into 'Entry questionnaire' and 'Post-observation questionnaire' and were distributed to experimental subjects at various points during the actual experiment. The 'Entry questionnaire' gathered background, demographic and searching experience information from subjects. The 'Post-observation questionnaire' elicited the tasks and topics that were of current interest, or were expected to be of interest, to the subject during the study. Table 4.5 shows the usage of the observation questionnaires in various experimental phases.

**Table 4.5: Usage of the observation questionnaires in various phases.**

| Phase | System | Entry | Post Observation |
|---|---|---|---|
| IIa | Contextual Search – using user profile | ✓ | ✓ |
| IIb | Contextual Search - using user profile & shared profile | ✓ | ✓ |
| IIc | Google | ✓ | ✓ |

Both questionnaires were typically divided into a series of sections and contained three styles of questions; *Likert scales, Semantic differentials and Open-ended questions*. The following sub-sections explain each style of question in detail.

## 4.5.1 Likert Scales

The Likert scale is a scaling method used to measure attitudes, preferences, and subjective reactions. In our case, the scale measures the extent of agreement or disagreement in response to a set of statements. It can be measured with four, five, six, seven, eight or nine point scales and is widely used in various fields of research. Given their compact size but reasonable granularity, five-point Likert scales are commonly used. Thus, in this study, subjects were asked to express their agreement or disagreement towards a set of statements on a five-point scale. Each degree of response agreement was given a numerical value from one to five. Table 4.6 shows an example of Likert scale taken from the 'Post-observation questionnaire'.

**Table 4.6: Example Likert scale.**

| Section 2: ADAPTIVENESS | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Please tick the numbers which most appropriately reflect your overall impressions of the learning ability of the contextual search. | | | | | | | | |
| 2.1 | The system remembers your last search actions (such as visited URLs, entered query, selected terms etc.) | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | |
| | never | ☐ | ☐ | ☐ | ☐ | ☐ | Always |

The Likert scales are designed to show differentiation among respondents who have a variety of opinions about an attitude object (i.e., anything that the subject may find good or bad), in this case whether or not the system remembers the subject's last search actions (such as visited URLs, entered query, selected terms etc.).

### 4.5.2 Semantic Differentials

An alternative to the Likert scales is semantic differential, which rates or scales two extreme positions or pairs of words of opposite meaning. The rating/scaling gives the researcher a fairly clear idea of intensity of feeling about an attitude object on a set of bipolar adjective scales. Table 4.7 presents a set of three semantic differentials taken from the 'Post-observation questionnaire'.

**Table 4.7: Example set of semantic differentials.**

| Section 3: RECOMMENDATION | | | | | | |
|---|---|---|---|---|---|---|
| Please tick the numbers which most appropriately reflect your overall impressions of the recommendation ability of the contextual search. | | | | | | |
| 3.1 The system communicated its recommendation action in a way that was : | | | | | | |
| | 1 | 2 | 3 | 4 | 5 | |
| unobtrusive | ☐ | ☐ | ☐ | ☐ | ☐ | obtrusive |
| uninformative | ☐ | ☐ | ☐ | ☐ | ☐ | informative |
| timely | ☐ | ☐ | ☐ | ☐ | ☐ | untimely |

In this example, respondents are asked to rate the attitude object of respective experimental questions by placing a tick or a cross in one of the five check boxes on each of the rating scales. In addition, the differential scales (i.e., positive and negative terms) in the experimental questionnaire are reversed in consecutive attitude objects to ensure the subject’s attention does not waver when completing the questionnaires.

### 4.5.3 Open-ended Questions

An open-ended question is designed to encourage a full, meaningful answer using the subject's own knowledge and/or feelings. Many advise against using open-ended questions and advocate using closed questions, due to challenges associated with ensuring relevant answers and with their interpretation. However, open-ended questions can be useful for revealing reasons *why* subjects feel the way they do and giving them a chance to comment freely on aspects of the system, the task or the experiment in general. Table 4.8 shows an example of some open-ended question taken from the ‘Post-observation questionnaire’.

**Table 4.8: Example of open-ended questions.**

| Section 6: COMMENTS AND SUGGESTIONS | |
|---|---|
| 6.1 | List out the positive features of this contextual search<br>__________<br>__________<br>__________<br>__________ |
| 6.2 | List out the negative features of this contextual search<br>__________<br>__________<br>__________<br>__________ |

In this example, respondents are encouraged to list the negative and positive features of the contextual SERL search. These questions helped to gather additional information and increased understanding of each subject's general feeling about the contextual SERL search.

In addition, during the experiment, various observational techniques were used to record each participant's information seeking behaviours at the interfaces to the experimental systems. The next section describes the observational techniques used in the Phase II experiment (section 4.1.2).

## 4.6 Observational Techniques

This section presents the observational techniques, which are a general class of research instrument, used to collect the data presented in Chapter 5. The objective of these techniques is to obtain data directly or indirectly by observing the activity or behaviour of participants in the study. The observation data is gathered in reference to the participant's normal activities with as little interference as possible.

In this study, a wide range of observational techniques were used to identify participants' information seeking behaviour, including a system log, screen capture (or recording), "talk aloud" audio recordings and note taking. Participants were informed that the study was a longitudinal, naturalistic observation of their information seeking behaviour. As such, all of their information seeking behaviours (e.g., use of Boolean query & advanced search features, mouse clicks etc.) and preferences (e.g., selected terms, concepts etc.) were observed or captured using these observational techniques. The following sub-sections describe the observational techniques used in this study.

## 4.6.1 System Log

A system log is a recording of what is happening on the hardware and/or software level while the subject of a study uses a computer. The system log as an observational technique has been used in various previous research (Choo et al., 2000; Dumais et al., 2003; Kelly, 2004).

In this study, the contextual SERL search automatically logged each participant's information seeking behaviours, such as entered query, selected terms, and visited URLs each time they searched for information. Figure 4.3 shows an example of such logged data.

You have visited the following URLs !

| ☐ | Context | Date | Query | Exact Meaning | Particular Phrase | Category | Precise Category | Shared |
|---|---|---|---|---|---|---|---|---|
| ☐ | office | 28-07-2007 | java creator | platform-independent object-oriented | micromail computer books software licences | Computer Science | None | ✓ |
| ☐ | office | 28-07-2007 | name of java creator | person | history java | Computer Science | None | ✓ |
| ☐ | office | 28-07-2007 | java program creator | platform-independent object-oriented | chapter 1 introducing java | Computer Science | None | ✓ |
| ☐ | office | 28-07-2007 | creator of java | person grows | q&a part 2 java creator | None | None | ✓ |

**Figure 4.3: Contextual SERL search system log.**

The logging process did not interfere with any of the participant's natural behaviour; instead, the system unobtrusively monitored and recorded participant interactions with the contextual search. These logged data were later reviewed to verify the observation notes and analysis data.

## 4.6.2 Screen and Audio Recording

The contextual SERL search and Google search are primarily mouse-driven applications naturally aimed at and controlled by end-users. This makes it challenging to log participants' interactions with these systems in real-time via note-taking. Various experiments (Bhavnani et al., 2003; Lazonder, Biemans, & Wopereis, 2000) have shown the usefulness of a screen capture tool to record screen interactions during the experiment.

In this study, Camtasia screen recording software was used to capture the contents of the computer screen as well as audio throughout the experiment for each subject. All this captured data (i.e., screen activities and audio) was saved as an avi file using distinct file names. These avi files were later played back to verify the observation notes and analysis data. It was noticed that screen and audio recordings required a considerable amount of CPU time and slowed down the experiment process (i.e., search process), especially, for those who used ‘think aloud’ protocol extensively. In addition, the generated files by the Camtasia software were large (smallest 43 MB and largest 196 MB) and were transferred to another machine to ensure that sufficient storage was available and to avoid data loss.

### 4.6.3 Think aloud

The think aloud protocol is a method which requires the participants to verbalise their impressions while they solve a problem using a system. Various experiments (Belkin et al., 2000; Koenemann & Belkin, 1996; White et al., 2003) have used the think aloud protocol to get a deeper understanding of users’ cognitive processes in problem-solving or information seeking behaviour.

In this study, participants were asked to perform six search tasks (Appendix A.5) that are goal directed and require effort and concentration. In addition, they were asked to “think aloud” while they engaged in the search. That is, the participants were instructed to articulate what they were thinking and what they were feeling whilst performing their search for the requested information. The commentary was captured using the audio recording feature of the Camtasia screen recorder software. However, those participants whose first language was not English were struggling to exercise the ‘think aloud’ protocol. They were unable to examine effectively the returned results as they were struggling with ‘think-aloud’ protocol. In addition, environmental aspects such as noise rendered some parts of the recordings indistinguishable.

## 4.7 Experimental Setup and Equipment

This section presents the experimental setup (Figure 4.4) and equipment used to collect the data presented in Chapter 5. The experiments (described in section 4.2) were carried out in a controlled environment in a small room. The setup included two standard machines; a) a client machine and b) a server machine, two standard QWERTY keyboards and two two-button optical mice.

The client machine had an Intel(R) Celeron(R) 2.6 GHz processor with 1 GB of RAM and a 50 GB hard disk. The server machine had an Intel(R) Pentium(R) 3.4 GHz processor with 2 GB of RAM and a 100 GB hard disk. Both machines ran the Microsoft Windows XP Professional operating system.

The client machine was connected to two 17 inch flat screen monitors using a VGA monitor cable splitter[25]. The server machine was installed with necessary software; Apache tomcat servlet server, MySQL database, WordNet and most importantly the contextual SERL search web application (section 3.5), and act as normal Web/application server.

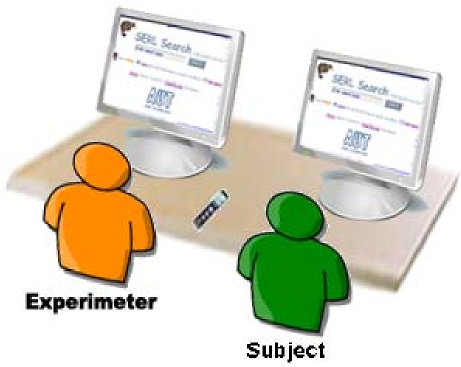


**Figure 4.4: Experimental environment.**

The experimental subject used only the client machine and no other support was offered by the observer. This allowed observation of the subject interaction to be unobtrusive, and also maintained consistency across all of the subjects. As such, the intervention during the experiment was limited only to occasions where technical problems prevented the subject from continuing with their given search tasks.

## 4.8 Sub-hypotheses

The five hypotheses (described in section 4.1) are further divided into a number of sub-hypotheses to make the capture and analysis of data more straightforward. In this section each set of sub-hypotheses is described.

---

[25] The VGA monitor cable splitter provides a fast and easy way to simultaneously connect two monitors to a computer.

#### 4.8.1.1 Find Information Readily (Hypothesis 1)

To test this hypothesis, five performance aspects (representing effectiveness and efficiency) of the experimental system and the contemporary search engine were compared. Each sub-hypothesis is worded as a null-hypothesis indicating that no variation exists between the three user experiments.

Number of queries (Hypothesis 1.1)

*There is no difference in the number of queries entered to reach the target information.*

Number of clicks (Hypothesis 1.2)

*There is no difference in the number of clicks clicked to reach the target information.*

Number of hits (Hypothesis 1.3)

*There is no difference in the number of hits browsed to reach the target information.*

Number of URLs (Hypothesis 1.4)

*There is no difference in the number of URLs visited to reach the target information.*

Length of time (Hypothesis 1.5)

*There is no difference in the length of time taken to reach the target information.*

#### 4.8.1.2 Adaptiveness Support (Hypothesis 2)

To test this hypothesis, four satisfaction aspects of the experimental system were compared. As such, this hypothesis further divided into four sub-hypotheses.

Remember last search actions (Hypothesis 2.1)

*Subjects find that the experimental system remembers their last search actions.*

Use last actions to recommend terms (Hypothesis 2.2)

*Subjects find that the experimental system uses their last search actions to recommend relevant terms.*

Learn interest over time (Hypothesis 2.3)

*Subjects find that the experimental system learns their search interests over time.*

Use of other users' search actions (Hypothesis 2.4)

*Subjects find that the experimental system uses other subjects' search actions to recommend relevant terms.*

#### 4.8.1.3 Recommendation Support (Hypothesis 3)

To test this hypothesis, two satisfaction aspects of the experimental system were compared. As such, this hypothesis was further divided into two sub-hypotheses that are tested in this section.

Recommendation strategy (Hypothesis 3.1)

*Subjects find that the experimental system communicates its recommendations clearly, and in a timely and in an unobtrusive manner.*

Conveying preferences (Hypothesis 3.2)

*Subjects find that the experimental system allows them to convey their preferences easily and in a comfortable manner.*

#### 4.8.1.4 Query Formulation (Hypothesis 4)

To test this hypothesis, two satisfaction aspects of the experimental system were compared. As such, this hypothesis was further divided into two sub-hypotheses.

Query formulation strategy (Hypothesis 4.1)

*Subjects find the number of experimental system query formulation steps just right to formulate an effective search query.*

Trust-worthy (Hypothesis 4.2)

> *Subjects find that the experimental system chooses relevant terms to formulate an effective search query.*

#### 4.8.1.5 Interface Support (Hypothesis 5)

To test this hypothesis, two satisfaction aspects of the experimental system were compared. As such, this hypothesis was further divided into two sub-hypotheses.

Relevance of content (Hypothesis 5.1)

> *Subjects find that the experimental system interface presents useful and effective information.*

Interface guide (Hypothesis 5.2)

> *Subjects find that the experimental system interface guides them to the information they need.*

## 4.9 Chapter Summary

This chapter has presented the experimental methods and Instruments used in this study. Two experimental phases, that were designed to test hypotheses constructed to evaluate the performance of the SERL search engine vis-à-vis Google were described. The subject recruitment process, experimental tasks, experimental procedure and observational techniques have also been detailed. In the next chapter the results of the experiments are presented and analysed.

# Chapter 5

# Experimental Results and Analysis

## 5.1 Introduction

Chapter 4 described the methodology used to define and then collect the data from two experimental phases; component testing and the observational study. This chapter presents the summarised results of the anonymous questionnaires and the results of the two experimental phases. The objective of the component testing was to ensure that the contextual SERL search (described in Chapter 3) performed correctly under typical search conditions. The objective of the observational study was to assess whether the contextual SERL search improved a user's information seeking ability along the dimensions of effectiveness, efficiency and subjective satisfaction in comparison to a contemporary search engine.

As stated in the previous chapter, the component testing was carried out by the researcher. Prior to actual user testing, 42 voluntary anonymous questionnaires were collected before recruiting 30 participants and commencing the observational study. The 30 subjects, with different levels of search experience, participated in the three sub-phase experiments; phase II (a) (i.e., contextual SERL search with user's profile), phase II (b) (i.e., contextual SERL search together with user's profile and shared contextual knowledge base), and phase II (c) (i.e., contemporary search engine). After being grouped into two groups based on prior search experience, subjects were randomly assigned to one of these sub-phase experiments, so that there were ten subjects in each user experiment. Subjects in the phase II (a) and phase II (b) experiments filled in a post-observation questionnaire so that their overall reactions to the experimental systems were captured.

Both qualitative and quantitative research data were collected during the course of the observational study experiments. Descriptive and inferential analyses were carried out to consider the collected data, utilising a computerised statistical analysis program (SPSS 14.0 for Windows). This chapter begins by describing the analysis of the anonymous questionnaires, followed by the summary results of the component testing experimental phase and the detailed results of the observational study experimental phase.

## 5.2 Anonymous Questionnaires

In this section the summary results of the anonymous questionnaire (Appendix C) survey are presented. The objective of this survey was to better understand how general users employ the Web to seek information as part of their daily life. Forty two ($n$ = 42) people responded to the anonymous questionnaire. Respondents were asked to complete five point Likert scales and semantic differentials relating to their information seeking behaviour. The respondents were students and faculty of the Auckland University of Technology, New Zealand.

The key survey findings are listed below:

**Respondent Characteristics**

- 71% respondents were male; and 81% were above 31 years old.
- 55% respondents were academic staff; and most of the respondents held Masters and/or Bachelors degrees.

**Search Experience, Usage and Satisfaction**

- One third of the respondents characterised themselves as expert internet users; and 95% possessed more than 4 years of search experience.

- Half of them frequently use (more than twice a day) web search tools and the other half search the web less often.
- 81% can always find the relevant information they are seeking; and 68% indicated that the time taken is shorter than expected.
- 79% of the respondents use the internet as the first recourse to search and find information.

**Search Behaviour**

- The most popular informal resources used when seeking information are previously bookmarked results and magazines, books, etc.
- 73% of the respondents (30) do not plan a search strategy in advance prior to searching.
- 78% of the respondents (32) use more than one query to resolve an information finding problem, and 50% of the respondents (21) prefer to use advanced search features in seeking relevant information.
- 83% of respondents acquire relevant information in the first 20 hits.
- Two of the most regular semantic differentials in judging the relevancy of search criteria are title and description, which acquired 80% and 78% responses respectively.
- 63% of respondents use those hits that they can use immediately.

**Search Preferences**

- More than three quarters (81%) of the respondents view their search engines usage as a single user, and more than half (56%) of the respondents are not willing to register to search engines.
- On the other hand, 51% of the respondents are willing to share their searching behaviour information anonymously to form a shared knowledge base.
- 100% of the respondents would like a search service/engine to recommend related terms. Among them, 22% of the respondents prefer using their own profiles; 32% using shared profile and 45% using both (i.e., personal profile and shared profile) to recommend related terms.
- However, only half of the respondents (50%) would allow search engines to recommend them as a user with similar interests to other searchers.

A detailed analysis of the responses to the questionnaire is presented in the Appendix C.

## 5.3 Component Test (CT)

The objective of the CT was to ensure that the contextual SERL search (described in Chapter 3) performed correctly under typical search conditions. During the CT, a total of six search queries combining one, two and three keywords, considering that queries are mostly limited to fewer than three key words (Jansen et al., 2000; Spink et al., 2002), across two domains/classifications – travel and computer science – were executed. Each query was submitted to contextual SERL search's three components; Word Sense Disambiguation (WSD), Concept Recommender (CR) and Meta Keywords Recommender (MKR), and to the Google search engine. Though all three components and Google search returned a large number of results, only the first ten results were evaluated for precision to limit the study in view of the fact that most users initially consider the first ten hits of a query. In addition, each query was run on all the three components and Google search on the same day in order to avoid variation that may be caused due to system updating.

Table 5.1 presents the final CT experiment results of the mean precision comparison of Google, WSD, CR and MKR.

**Table 5.1: Precision summary results of component test.**

| No. of Keywords | Google | WSD | CR | MKR |
|---|---|---|---|---|
| 1 | 1.2 | 1.6 | 3.6 | 2.7 |
| 2 | 3.3 | 3.6 | 3.7 | 3.7 |
| 3 | 2.2 | 2.5 | 2.9 | 2.8 |
| Overall | 7.9 | 7.9 | 10.2 | 9.2 |

The precision results under various scenarios show that the CR and MKR components perform better in terms of precision compared to Google, whereas the WSD component (by adding synonymous terms) appears to perform at the same level. The CR component exhibits the best improvement in precision compared to the others. The results of this test show that the WSD does not improve searching performance when used in isolation. One very interesting finding from these results is that the smaller the query, the more it benefits from the contextual SERL search components. On the other hand, the effort involved in creating complex Boolean queries for larger queries does not appear to pay off. Although more data would be required to draw any conclusive comments, it seems that these components' optimised queries could improve the retrieval effectiveness regardless of the performance of initial results. These results also provide a useful degree of reassurance that the system functions as intended in terms of the functionality and performance of its constituent parts. Detailed results of the CT are presented in Appendix B.

## 5.4 Observational Study (OS)

In this section the detailed results of the observational study are presented. During the experimental phase, three user experiments/studies; phase II (a), phase II (b), and phase II (c) were conducted in a controlled environment.

The phase II (a) and phase II (b) experiments were carried out with differing objectives. The aim of the phase II (a) experiment was to determine whether the contextual retrieval system enabled subjects to find relevant information, and to do so more readily than with a standard search engine, using their personal contextual profiles. During the phase II (a) experiment, subjects performed six search tasks using the system, and had their search behaviour and preferences captured in order to create their personal contextual profiles and to populate a Shared Contextual Knowledge Base (SCKB). However, the SCKB was not accessible to them during their search.

Once the phase II (a) experiment was complete, a second group of subjects repeated the same six allotted search tasks; however these subjects had the SCKB enabled. As such, the aim of the phase II (b) experiment was to determine whether the system enabled subjects to find relevant information, and to do so more readily than with a standard search engine, using their personal contextual profiles *and* the SCKB. This enabled the contribution of the shared profile to search quality to be measured, by comparing the "speed" with which subjects could find data with the first group who did not have access to the shared profile. The phase II (c) experiment was carried out on a contemporary search engine, Google. As such, the experimental procedure of phase II (c) was slightly different than that used in the phase II (a) and phase II

(b) experiments. However, the same six search tasks were undertaken to provide a benchmark against which the phase II (a) and phase II (b) results could be compared.

After being grouped on the basis of self-reported search experience, subjects were randomly assigned to one of the three user experiments, so that there were ten subjects in each user experiment. Before the actual user experiment, subjects were given the same general instructions, video demonstration, and filled in an entry questionnaire containing information about their characteristics and search experience. Subjects attempted six search tasks in each of the three user experiments and were able to complete all search tasks successfully. Thereafter, subjects filled in a post observation questionnaire immediately after the experiment had ended to capture overall subjective reactions to the experimental systems.

The results presented in this section are drawn from the observation data and from the entry and post observation questionnaire responses. The observation data, such as the *number of queries, number of clicks, number of hits, number of URLs* and *length of time* taken to reach target information, were extracted from the observation video clips and supported by the observation notes and the system logs. While extracting data, the following procedures were followed to ensure integrity of the data due to the nature of human behaviour;

i. The numbers of queries entered are recorded only for those queries that are entered to the contextual SERL search or the Google search engine. Other queries that are entered to the Web site's built-in search are not recorded as only a minority of subjects used the Web site's built-in search.

ii. The length of time taken to complete a search task is recorded from the time shown in the Windows media player taskbar and determined by recording the time that lapsed between the first entered query for the first search task and the first query entered for the subsequent search task. No other equipment such as a stop watch was used and time was recorded in seconds.

iii. The number of clicks is incremented for all clicks and the enter action/function on Google search is also recorded as a click as the enter key was disabled in the contextual SERL search. However, refreshing the visited page is not recorded as a click.

iv. The numbers of hits browsed are recorded based on the numbers of visited search result page(s).

v. The numbers of URLs visited are recorded from both the visited URLs from the search result page(s) and from the other Web site(s).

The post observation questionnaire responses were extracted from the completed five point Likert scales and semantic differentials, where a higher score (i.e., range 1-5, higher = better) represented stronger agreement with the attitude object. Those randomised semantic differentials (e.g., range 5-1, lower = better) were transformed/reversed on their scores (i.e., range 1-5, higher = better) in order to retain consistency in the analysis and presentation of the results.

Both *parametric* and *nonparametric* statistical methods were used to test statistical significance between the three sets of experimental data. The data gathered from the observational data are interval in nature and parametric methods are thus more appropriate. Graphical and numerical normality tests (Appendix D) were carried out on these interval scale data, to determine whether the data was normally distributed or not. The parametric one-way analysis of variance (ANOVA) and Dunnett's one-tailed tests were performed on those *normally* distributed data to test statistical significance. Since the treatment data sizes were the same, where appropriate, Tukey’s honestly significant difference post hoc tests (i.e., homogeneity subsets detection, pair wise comparison and confidence interval between phases) were used to reduce the likelihood of Type I errors (i.e., rejecting null hypotheses that are true).

During the observation study, it was discovered that one of the six assigned tasks was significantly more challenging than the other five. The task set was “*You are about to depart on a short-tour to Indonesia. Your agenda includes a visit to the west coast of Java, which is renowned for its natural riches, its fertile* *land with some volcanoes and its cultural riches. As your time in Java is limited to a week you would like to save time and find information about any of the west coast's FIVE best leisure and sports activities prior to your departure*". Subjects' answers to this task were often imprecise, and subjects commonly took long periods of time, entered more queries and browsed more hits to discover an answer when compared to the other tasks. There are various Web sites that exist which provide sufficient information to complete the task, but to find them seemed to require a level of prior knowledge of the west coast of Java. Many subjects found difficulty in finding suitable sites, leading to poorly defined answers. Furthermore, there are numerous sites on the Web with incomplete or misleading information on this topic. For this reason, when necessary, further detailed tests were performed for individual search tasks – to assess the effectiveness and efficiency of the contextual SERL search in comparison to a contemporary search engine.

Data gathered from the entry and post observation questionnaires were ordinal in nature (e.g., Likert scales and semantic differentials) and thus nonparametric methods are more appropriate than their parametric equivalents. The nonparametric Kruskal-Wallis tests were performed to test statistical significance. In this study, for all statistical tests, an alpha level of .05 (5% level) was used. The results presented in the following sections are for inter-user experiment comparison.

This section begins by presenting results on the subjects' characteristics, search experience (section 5.4.1) and overall impressions (section 5.4.2) of the experimental systems and the tasks attempted. This is followed by the testing, in turn, of each of the hypotheses outlined, starting with consideration of adaptiveness support, and culminating with the findings on the searching behaviour (section 5.4.8). The next chapter discusses the implications of these findings in more detail.

### 5.4.1 Subjects' Characteristics and Search Experience

This section summarises the subjects' characteristics and search experience. Of the 30 subjects, 73% were male (22) and 27% were female (8). About 30% of the subjects (9) were less than 25 years old. The remaining 70% of the subjects (21) were 26 years or older. About 57% of the subjects (17) were students. The remaining 43% of subjects (13) were a mixture of academic staff, researchers and others. About 80% of the subjects (24) held either master or bachelor degrees. The remaining 17% and 3% of subjects hold other degrees (5) or a PhD (1), respectively.

Subjects were asked to indicate their level of Web search expertise, frequency of Web search usage (at office/home) and years of Web search experience on five point Likert scales and on semantic differentials. In addition, they were asked to indicate their satisfaction with current Web search usage, mainly, whether they find what they are looking for and whether finding relevant information on the Internet has not taken up more time than they expected.

Table 5.2 summarised the subjects' Web search experiences. Key findings include; about 57% of the subjects (17) indicated they were expert or confident about their searching abilities. The remaining 43% of subjects (13) indicated they were novice Web searchers, while 97% of respondents (29) indicated that they have more than four years of Web search experience. The remaining 3% of respondents (1) indicated that they have three years or less of Web search experience. About 53% of subjects (16) indicated that they regularly use Web search for their office or academic tasks. The remaining 43% (13) and 4% (1) of subjects use a Web search engine at least once or twice a day and more than once or twice a month respectively. Similarly, about 53% of subjects (16) indicated they regularly use Web search for their personal tasks. The remaining 30% (9) and 17% (5) of subjects use a Web search engine at least once or twice a day and more than once or twice a month respectively. About 73% of subjects (22) indicated they always find the relevant information for which they are looking although some 57% of respondents (17) indicated that Internet searching had taken up more time than they expected.

**Table 5.2: Summary of search experience and satisfaction of subjects.**

| ***(n=30)*** | | |
|---|---|---|
| Search Expertise | Number | Percentage (%) |
| Novice | 13 | 43.3% |
| Expert | 17 | 56.7 % |
| Years of search experience | | |
| <= 3 years | 1 | 3.3% |
| >= 4 years | 29 | 96.7% |
| Frequency of Web search for work/academic tasks | | |
| Once or twice a month | 1 | 3.3% |
| Once or twice a day | 13 | 43.3% |
| More often | 16 | 53.3% |
| Frequency of Web search for personal tasks | | |
| Once or twice a month | 5 | 16.6% |
| Once or twice a day | 9 | 30% |
| More often | 16 | 53.3% |
| Can find what they are looking for | | |
| Yes | 22 | 73.3% |
| No | 8 | 26.7% |
| Time taken | | |
| Yes | 17 | 56.7% |
| No | 13 | 43.3% |

This finding was potentially a good indicator of experience levels and the usage of Web search. Subjects with more experience and regular Web searcher are more competent searchers and are more confident in their searching skills. It is not surprising that, they feel they are more successful in finding relevant information from the Web. This finding would suggest that the learning curve on using the contextual SERL search would be short.

### 5.4.2 Overall Impressions

This section presents results on the overall impressions formed regarding the experimental systems and the search tasks performed by subjects. The results presented are drawn from the post observation questionnaire data. This analysis is not directly associated with any particular hypotheses but provides interesting insight into the observation study nonetheless. These factors affect the subjects' perceptions of the experimental systems, the search tasks and the observational study as a whole so it is important to consider them in an analysis.

#### 5.4.2.1 Experimental Systems

Subjects were asked to indicate their overall reactions to the experimental systems with regard to the completion of search tasks on four semantic differentials: ‘terrible’/‘wonderful’, ‘frustrating’/‘satisfying’, ‘difficult’/‘easy’, and ‘rigid’/‘flexible’, where a higher score (i.e., range 1-5, higher = better) represented stronger agreement with the attitude object. The bar charts (Figure 5.1) show the graphical representation of subjects’ responses for the three different experimental systems.

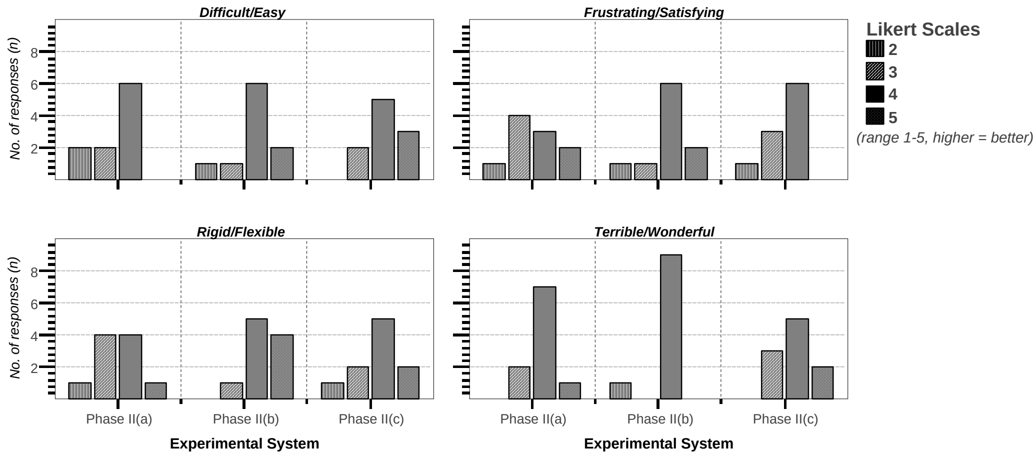


**Figure 5.1: Bar charts of overall reactions regard to the completion of search tasks.**

The bar charts illustrate that a higher number of subjects using the phase II (c) chose the semantic differential and the Likert scale value of ‘five’ (easy *n* = 3 and wonderful *n* = 2) in comparison to the subjects using the phase II (a) (easy *n* = 0 and wonderful *n* = 1) and phase II (b) experimental system (easy *n* = 2 and wonderful *n* = 0). Similarly, the bar charts illustrate that higher numbers of subjects using the phase II (b) chose the semantic differential and the Likert scale value of ‘five’ (flexible *n* = 4) in comparison to the subjects using the phase II (a) (flexible *n* = 1) and phase II (c) experimental system (flexible *n* = 2).

A nonparametric Kruskal-Wallis test was performed to determine whether any significant differences exist between the experimental systems. The results, where wonderful ($X^2$ = .004 and *p* = .998), satisfying ($X^2$ = 1.517 and *p* = .468), easy ($X^2$ = 3.456 and *p* = .178), and flexible ($X^2$ = 4.548 and *p* = .103), established no statistically significant differences between the three experimental systems. A Mann-Whitney U test was applied between the subject groups (i.e., experienced and inexperienced) to determine the significance of any differences. The results, where wonderful (U = 105 and $p$ = .837), satisfying (U = 93 and $p$ = .483), easy (U = 67 and $p$ = .072), and flexible U = 92 and $p$ = .457), established no statistically significant differences between the subject groups.

In addition, subjects were asked to indicate their overall reactions to the experimental systems with regard to the user interface (UI) on three semantic differentials: ‘dull’/‘stimulating’, ‘confusing’/‘clear’ and ‘difficult’/‘easy’. The bar charts (Figure 5.2) show the graphical representation of subjects’ responses for the three different experimental systems.

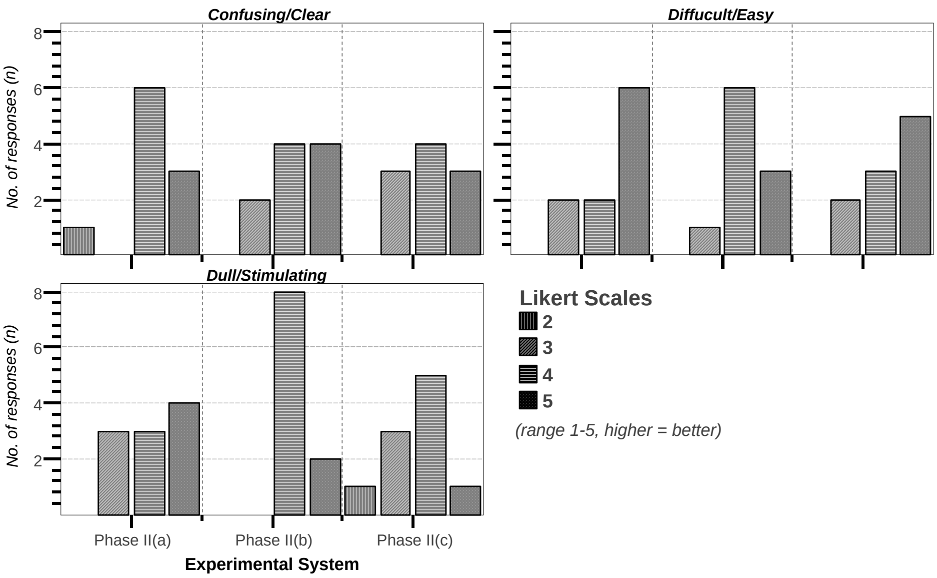


**Figure 5.2: Bar charts of overall reactions regard to the UI.**

The bar charts illustrate that higher numbers of subjects using the phase II (a) chose the semantic differential and the Likert scale value of ‘five’ (easy $n$ = 6 and simulating $n$ = 4) in comparison to the subjects using the phase II (b) (easy $n$ = 3 and simulating $n$ = 2) and phase II (c) experimental system (easy $n$ = 5 and simulating $n$ = 1). Similarly, the bar charts illustrate that higher numbers of subjects using the phase II (b) chose the semantic differential and the Likert scale value of ‘five’ (clear $n$ = 4) in comparison to the subjects using the phase II (a) (clear $n$ = 3) and phase II (c) experimental system (clear $n$ = 3).

A nonparametric Kruskal-Wallis test was performed to determine whether any significance differences exist between the experimental systems. The results, where stimulating ($X^2$= 3.106 and *p* = .212), clear ($X^2$= .375 and *p* = .829), and easy ($X^2$=.669 and *p* = .716), established no statistically significant differences between the three experimental systems. A Mann-Whitney U test was applied between the subject groups (i.e., experienced and inexperienced) to determine the significance of any differences. The results, except for clear (U = 55.50 and *p* = .020), established no statistically significant differences between the subject groups for the stimulating (U = 89.50 and *p* = .385), and easy (U = 73.50 and *p* = .123) scales. The result indicates that those subjects classified as 'experienced' found the user interface of the experimental systems are significantly clearer than the inexperienced group; to a certain extent this validated the subject classification. The results seem to indicate that the three systems are considered equivalent in terms of the subjects' overall experience, implying that the greater complexity in the two versions of the SERL search does not have a negative impact on the user's search experience.

#### 5.4.2.2 Tasks

Subjects were also asked to indicate their overall confidence with regard to completion of tasks on a five point Likert scale (range 1-5, higher = better), with the two scales ranging from 'with difficulty'/'easily' and 'Not at all confident'/'very confident' respectively. A nonparametric Kruskal-Wallis test was performed to determine whether any significance differences exist between the experimental systems. The results, where easily ($X^2$= 4.163 and *p* = .125), and very confident ($X^2$= 2.965 and *p* = .227), established no statistically significant differences between the three experimental systems. A Mann-Whitney U test was applied between the subject groups to determine the significance of any differences. The results established no statistically significant differences for the 'with difficulty'/'easily' scale (U = 78.50 and *p* = .183) but did indicate significant differences for the confidence scale (U = 48.00 and *p* = .008) in the differentials between the subject groups.

### 5.4.3 Find Information Readily (Hypothesis 1)

This section presents results related to the first experimental hypothesis: *the contextual SERL search enables subjects to find relevant information* ***more***

***readily** than a standard search engine using their personal profile and shared contextual knowledge base.* This hypothesis is further divided into five sub-hypotheses to test the performance (i.e., effectiveness and efficiency) aspects of the experimental system relative to the contemporary search engine.

#### 5.4.3.1 Number of Queries (*Hypothesis 1.1*)

**Null hypothesis**

*There is no difference in the number of queries entered to reach the target information.*

$H_0$ : $\mu_{PhaseII\,a} = \mu_{PhaseII\,b} = \mu_{PhaseII\,c}$ *(all expected medians are equal)*

**Alternative hypothesis**

$H_a$: at least one $\leq$

**Measures needed**

*<u>Total number of queries entered</u> to complete all search tasks in each experimental phase.*

The *normality* test results (Appendix D, section D.1, page 209) demonstrated that the total number of queries entered to complete all search tasks is not-normally distributed. This suggests that the most appropriate measure of central tendency is the *median* and most appropriate test is a nonparametric method.

A series of *nonparametric* tests (Kruskal-Wallis test and Mann-Whitney test) were performed (Appendix E, section E.1, page 214 to test hypothesis 1.1. The Kruskal-Wallis test results ($X^2$= 3.056 and $p$ = 0.217) for the total number of queries entered to complete all search tasks established no statistically significant differences between the three experimental systems. However, detailed Kruskal-Wallis test results ($X^2$= 7.009 and $p$ = 0.030) for the total number of queries entered to complete an *individual* search task established *highly* statistically significant differences between the phase II (b) and phase II (c) experimental system for search task six. Search task six was significantly more challenging, as described in the section 5.4. Subjects' answers to this task were often imprecise, and subjects commonly took long periods of time, entered more queries and browsed more hits to discover an answer.

Similarly, the Mann-Whitney U test results (U = 17.00 and *p* = 0.011) established highly statistically significant differences between the phase II (b) and phase II (c) experimental system for search task six. The next chapter discusses the implications of these findings in more detail.

#### 5.4.3.2 Number of Clicks (*Hypothesis 1.2*)

**Null hypothesis**

*There is no difference in the number of clicks clicked to reach the target information.*

$H_0$ : $\mu_{PhaseII\,a} = \mu_{PhaseII\,b} = \mu_{PhaseII\,c}$ *(all expected means are equal)*

**Alternative hypothesis**

$H_a$: *at least one* $\leq$

**Measures needed**

*<u>Total number of clicks clicked</u> to complete all search tasks in each experimental phase.*

The *normality* test results (Appendix D, section D.2, page 210) demonstrated that the total number of clicks clicked to complete all search tasks is normally distributed. A Levene's test for homogeneity of variance ($F$ = 2.555, $p$ = 0.96) did not reveal any significant differences in variance between the three experimental systems. This demonstrates that the data describing the total number of clicks does not violate the normality assumption or the homogeneity-of-variance assumption. This suggests that the most appropriate measure of central tendency is the mean and most appropriate test is a parametric method.

A series of parametric tests (one-way analysis of variance (ANOVA) and Dunnett's one-tailed test) were performed (Appendix E, section E.2, page 216) to test hypothesis 1.2. The one-way analysis of variance (ANOVA) and Dunnett's one-tailed test results (Anova $p$ = 0.413, Dunnett's one-tailed $p$ = 0.474 for phase II (a) and II (c) and Dunnett's one-tailed $p$ = 0.162 for phase II (b) and II (c)) for the total number of clicks clicked to complete all search tasks established no statistically significant differences between the three experimental systems. Further detailed one-way ANOVA and Dunnett's one-tailed test was performed for the total number of clicks clicked to complete an individual search task. The results (Anova $p$ = 0.041, Dunnett's one-tailed $p$ =

0.011 for phase II (b) and II (c)) established statistically significant differences between the two experimental systems for search task six. To further analyse this difference, Tukey's post hoc test was performed. The results (*p* = 0.032) also established statistically significant differences between the two experimental systems.

In addition, a series of nonparametric tests (Kruskal-Wallis) were performed to guard against the possibility that the assumption of a normal distribution did not hold. The Kruskal-Wallis test results ($X^2$ = 0.954 and *p* = 0.621) for the total numbers of clicks clicked to complete all search tasks established no statistically significant differences either. However, detailed Kruskal-Wallis test results ($X^2$ = 8.987 and *p* = 0.011) for the total number of clicks clicked to complete an individual search task established statistically highly significant differences between the experimental systems for search task six. The results are consistent between parametric and nonparametric tests.

#### 5.4.3.3 Number of Hits (*Hypothesis 1.3*)

**Null hypothesis**

*There is no difference in the number of hits browsed to reach the target information.*

$H_0$ : $\mu_{PhaseII\,a} = \mu_{PhaseII\,b} = \mu_{PhaseII\,c}$ (*all expected means are equal*)

**Alternative hypothesis**

$H_a$: *at least one* ≤

**Measures needed**

*Total number of hits browsed to complete search tasks in each experimental phase.*

The *normality* test results (Appendix D, section D.3, page 210) demonstrated that the total number of hits browsed to complete all search tasks is normally distributed. A Levene’s test for homogeneity of variance (*F* = 1.197, *p* = 0.318) did not reveal any significant differences in variance between the three experimental systems. This demonstrates that the total number of hits browsed data do not violate the normality assumption or the homogeneity-of-variance assumption. This suggests that the most appropriate measure of central tendency is the mean and most appropriate test is a parametric method.

A parametric test (one-way analysis of variance (ANOVA) and Dunnett's one-tailed test) was performed (Appendix E, section E.3, page 220) to test hypothesis 1.3. The one-way analysis of variance (ANOVA) and Dunnett's one-tailed test results (Anova *p* = 0.005 and Dunnett's one-tailed *p* = 0.001 for the phase II (b) and II (c)) established highly statistically significant differences between the experimental systems. A Tukey's post hoc test (*p* = 0.004) also established that the phase II (b) and phase II (c) experimental systems have *highly* significant differences.

In addition, a *nonparametric* test (Kruskal-Wallis) was performed to guard against the possibility that the assumption of a normal distribution did not hold. The Kruskal-Wallis results ($X^2$= 10.448 and *p* = 0.005) established highly statistically significant differences between the experimental systems. The results are consistent between *parametric* and *nonparametric* tests.

#### 5.4.3.4 Number of URLs (*Hypothesis 1.4*)

**Null hypothesis**

*There is no difference in number of URLs visited to reach the target information.*

$H_0$ : $\mu_{PhaseII\,a} = \mu_{PhaseII\,b} = \mu_{PhaseII\,c}$ *(all expected means are equal)*

**Alternative hypothesis**

$H_a$: *at least one* $\leq$

**Measures needed**

*<u>Total number of URLs visited</u> to complete search tasks in three experimental systems.*

The *normality* test results (Appendix D, section D.4, page 211) demonstrated the total number of URLs visited to complete all search tasks is normally distributed. A Levene's test for homogeneity of variance results (*F*= 1.207, *p* = 0.315) established no statistically significant or did not reveal any significant differences in variance between three experimental systems. This demonstrates that the total number of URLs visited treatment data do not violate the normality assumption and the homogeneity-of-variance assumption. This also suggests that the most appropriate measure of central tendency is the mean and most appropriate test is the parametric method.

A series of parametric tests (one-way analysis of variance (ANOVA) and Dunnett's one-tailed test) were performed (Appendix E, section E.4, page 223) to test the hypothesis 1.4. The one-way analysis of variance (ANOVA) and Dunnett's one-tailed tests results (Anova $p$ = 0.004 and Dunnett's one-tailed $p$ = 0.001 for phase II (b) and II (c)) established statistically highly significant differences between experimental systems. A Tukey's post hoc test results ($p$ = 0.003) also established that the phase II (b) and phase II (c) experimental systems have the significant differences.

In addition, a nonparametric test (Kruskal-Wallis) was performed to guard against the possibility that the assumption of normal distribution did not hold. The Kruskal-Wallis results ($X^2$= 10.648 and $p$ = 0.005) established statistically highly significant differences as well. The test results are consistent between parametric and nonparametric tests. The next chapter discusses the implications of these findings in more detail.

### 5.4.3.5 Length of Time (*Hypothesis 1.5*)

**Null hypothesis**

*There is no difference in the length of time taken to reach the target information.*

$H_0$ : $\mu_{PhaseII\,a} = \mu_{PhaseII\,b} = \mu_{PhaseII\,c}$ (*all expected means are equal)*

**Alternative hypothesis**

$H_a$: *at least one* $\leq$

**Measures needed**

*<u>Total time taken (in seconds)</u> to reach the target information in each experimental systems.*

The *normality* test results (Appendix D, section D.5, page 212) demonstrated the total length of time taken to complete all search tasks is normally distributed. A Levene’s test for homogeneity of variance ($F$= 1.445, $p$ = 0.253) did not reveal any significant differences in variance between three experimental systems. This demonstrates that the total length of time taken to complete *all* search tasks data did not violate the normality assumption or the homogeneity-of-variance assumption. This suggests that the most appropriate measure of central tendency is the mean and most appropriate test is a *parametric* method.

A series of *parametric* tests (one-way analysis of variance (ANOVA) and Dunnett's one-tailed test) were performed (Appendix E, section E.5, page 225) to test hypothesis 1.5. The one-way analysis of variance (ANOVA) and Dunnett's one-tailed tests results (Anova $p$ = 0.242, Dunnett's one-tailed: $p$ = 0.559 for phase II (a) and II (c) and $p$ = 0.102 for phase II (b) and II (c)) established *no* statistically significant differences between the three experimental systems.

Further detailed one-way ANOVA and Dunnett's one-tailed test were performed for the total time taken to complete an individual search task. The results (Anova $p$ = 0.011 and Dunnett's one-tailed $p$ = 0.003 for phase II (b) and II (c)) established statistically significant differences between the two experimental systems for search task six. To further analyse this difference, Tukey's post hoc test was performed. The results ($p$ = 0.032) also established statistically significant differences between the two experimental systems.

In addition, a series of nonparametric tests (Kruskal-Wallis) were performed to guard against the possibility that the assumption of a normal distribution did not hold. The Kruskal-Wallis test results ($X^2$= 3.510 and $p$ = 0.173) established no statistically significant differences. However, detailed Kruskal-Wallis test results ($X^2$= 9.177 and $p$ = 0.010) for the total time taken to complete an *individual* search task established highly statistically significant differences between the experimental systems for search task six. The test results are consistent between *parametric* and *nonparametric* tests.

All five sub-hypotheses tests yielded highly similar results and provide some evidence in favour of the first hypothesis.

### 5.4.4 Adaptiveness Support (Hypothesis 2)

This section presents results related to the second experimental hypothesis: *the contextual SERL search adapts to the information needs of the searcher and facilitates effective recommendation of terms.* This hypothesis is further divided into four sub-hypotheses that measure the adaptiveness aspects of the experimental system. The findings presented in this section focus on subjective impressions of the two variants of the contextual SERL search (as adaptiveness is not available in Google search).

### 5.4.4.1 Remember Last Search Actions (*Hypothesis 2.1*)

*Subjects find that the experimental system remembers their last search actions.*

Subjects were asked to complete a five point Likert scale (range 1-5, higher = better) on whether or not the experimental system remembered their last search actions. The bar chart (Figure 5.3) shows the graphical representation of the subjects' responses for the two different experimental systems utilising SERL search phase II (a) and II (b).

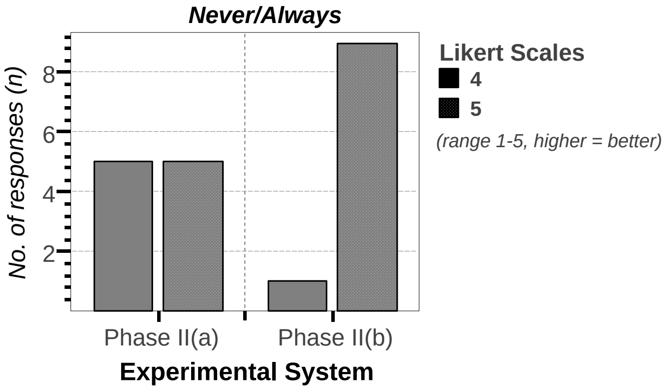


**Figure 5.3: Bar chart for system remembers last search actions.**

The bar chart visually illustrates that a higher number of subjects using the phase II (b) experimental system (always $n$ = 9) chose the Likert scale value of 'five' in comparison to the phase II (a) experimental system (always $n$ = 5). In addition, the phase II (b) had a slightly higher *median* value ($\tilde{X}$ = 5) than the phase II (a) experimental system ($\tilde{X}$ = 4.5). The bar charts and the *median* value results are consistent. However, Mann-Whitney U test results ($U$ = 30.00 and one-tailed $p$ = .143) in regard to the system remembering the last search actions established no statistically significant differences between the two experimental systems.

### 5.4.4.2 Use Last Actions to Recommend Terms (*Hypothesis 2.2*)

*Subjects find that the experimental system uses their last search actions to recommended relevant terms.*

Subjects were asked to indicate their overall reactions to the contextual SERL search system with regard to whether or not the system uses their last search actions to recommend terms on three semantic differentials; ‘irrelevant’/‘relevant’, ‘never’/‘always’, and ‘not useful’/‘useful’. The bar charts (Figure 5.4) show the graphical representation of subjects’ responses for the two different experimental systems. Similarly, the radar chart (Figure 5.5) shows the graphical representation of the *median* values of subjects’ responses for the two different experimental systems.

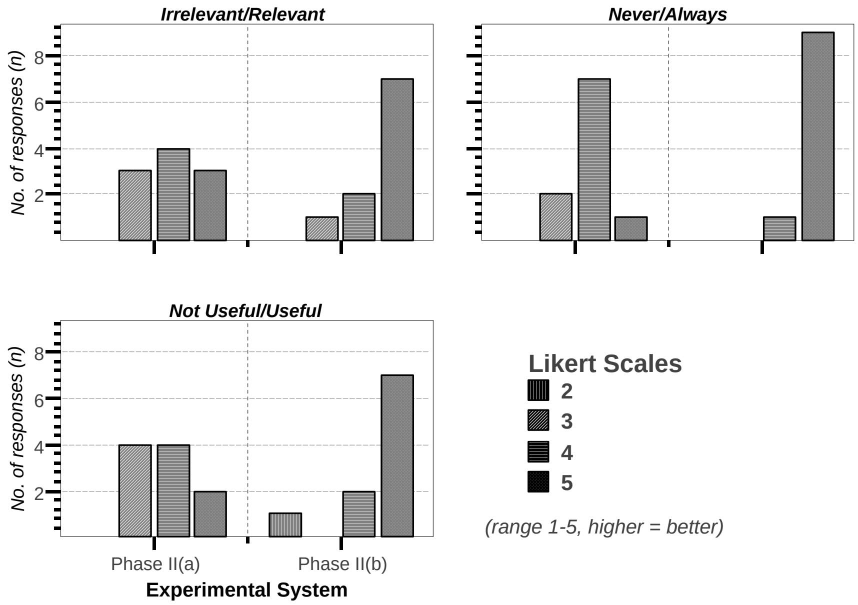


**Figure 5.4: Bar chart of system uses the user’s last actions to recommend terms.**

The bar charts illustrate that a higher numbers of subjects using the phase II (b) experimental system chose the Likert scale value of ‘five’ (relevant $n$ = 7, always $n$ = 9 and useful $n$ = 7) in comparison to the phase II (a) experimental system users (relevant $n$ = 3, always $n$ = 1 and useful $n$ = 2).

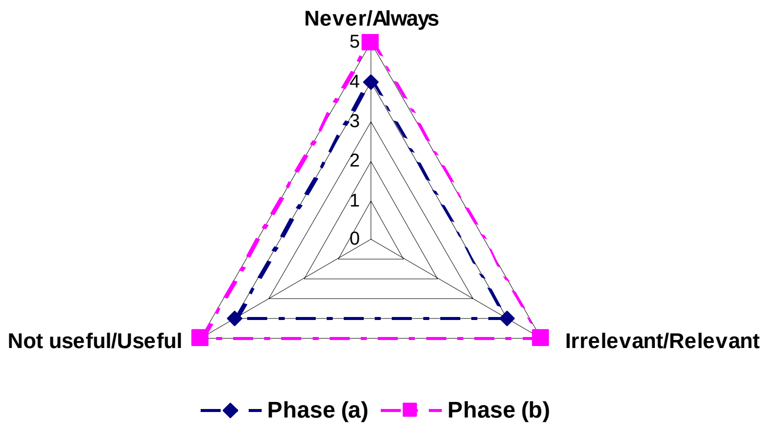


**Figure 5.5: Radar chart of system uses the user's last actions to recommend terms.**

Whilst the *median* radar chart (Figure 5.5) is presenting the same data as the bar chart (Figure 5.4), it is being used to show that the phase II (b) experimental system is consistently scoring points that are near to or at the edge (*higher end of the scale*) while phase II (a) scoring points are close to the centre point (*lower end of the scale*.

Both the bar and radar charts illustrate that the subjects using the phase II (b) experimental system tended to rate the various attributes of performance more positively in comparison to the users of the phase II (a) experimental system. However, Mann-Whitney U test results established that there is *no* statistically significant difference between the two experimental systems for the 'irrelevant'/'relevant' ($U$ = 29.00 and one-tailed $p$ = .123) and 'not useful'/'useful' scale ratings ($U$ = 25.00 and one-tailed $p$ = .063). The same test did indicate that the ratings against the 'never'/'always' scale were significantly different ($U$ = 21.00 and one-tailed $p$ = .001). This result provides some evidence that the experimental systems managed to observe and utilise the user's behaviour (i.e., interests and preferences) to recommend term suggestions to individual users.

#### 5.4.4.3 Learn Interest over Time (*Hypothesis 2.3*)

*Subjects find that the experimental system learns their search interests over time.*

Subjects were asked to complete a five point Likert scale (range 1-5, higher = better) indicating whether the experimental system learnt their interests over time. The bar chart (Figure 5.6) shows the graphical representation of the subjects' responses for the two different experimental systems.

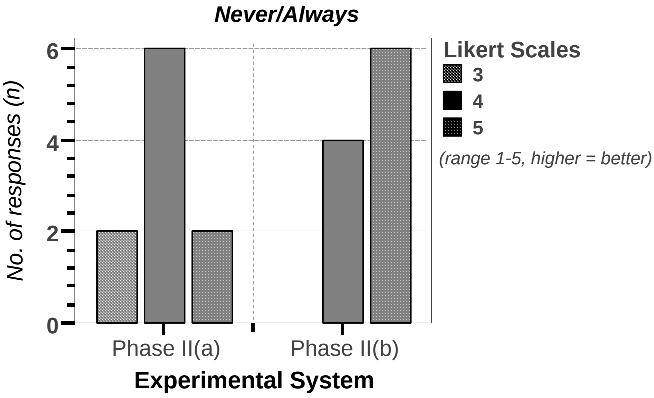


**Figure 5.6: The bar chart of system learns the user's interest over the time.**

The bar chart illustrates that a higher number of subjects using the phase II (b) experimental system chose the Likert scale value of 'five' (always $n$ = 6) in comparison to the phase II (a) experimental system users (always $n$ = 2). Similarly, the phase II (b) had a slightly higher *median* value ($\tilde{X}$ = 5) than the phase II (a) experimental system ($\tilde{X}$ = 4). The bar charts and the *median* value results are consistent. However, Mann-Whitney U test results ($U$ = 26.00 and one-tailed $p$ = .075) established no statistically significant differences between the two experimental systems.

### 5.4.4.4 Use of Other Users' Search Actions (*Hypothesis 2.4*)

*Subjects find that the experimental system uses other subjects' search actions to recommend relevant terms.*

Subjects undertaking the phase II (b) experiment were asked to indicate their overall reactions to the contextual SERL search system utilising OTHER users/searchers search actions to recommend terms, on three semantic differentials; 'irrelevant'/'relevant', 'never'/'always' and 'not useful'/'useful'. The semantic differential scales were responded to by only nine subjects instead of ten. The bar charts (Figure 5.7) show the graphical representation of subjects' responses.

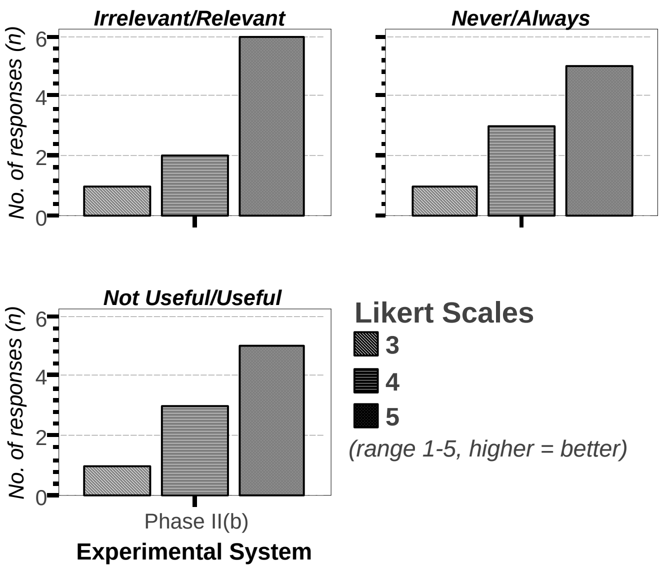


**Figure 5.7: Bar charts of system uses other users/searchers search actions to recommend terms.**

The bar charts illustrates that a higher numbers of subjects chose the semantic differential value of ‘five’ (relevant $n = 6$, always $n = 5$, *and* useful $n = 5$) compared to other semantic differential values. This result provides some evidence that the phase II (b) experimental system managed to observe and utilise the behaviour of other users to recommend term suggestions to individual users.

All four sub-hypotheses tests yielded highly similar results (i.e., subjects rated highly for the adaptiveness supports provided by the experimental system) and provide some evidence in favour of the second hypothesis.

### 5.4.5 Recommendation Support (Hypothesis 3)

This section presents results related to the third experimental hypothesis*: the contextual SERL search eases the conveying preferences process and recommends relevant and useful terms.* This hypothesis is further divided into two sub-hypotheses that measure the recommendation aspects of the experimental system. The findings presented in this section focus on subjective impressions of the two variants of the contextual SERL search.

### 5.4.5.1 Recommendation Strategy (*Hypothesis 3.1*)

> *Subjects find that the experimental system communicates its recommendations clearly, timely and in an unobtrusive manner.*

Subjects were asked to indicate their overall reactions to the contextual SERL search system with regard to the way in which the system communicated its recommendation actions, on four semantic differentials; ‘confusing’/‘clear’, ‘obtrusive’/‘unobtrusive’, ‘uninformative’/‘informative’, and ‘untimely’/‘timely’. The ‘untimely’/‘timely’ semantic differential scale for the phase II (b) experimental system has only nine subjects responded instead of ten. In addition, subjects were asked to complete a five point Likert scale (range 1-5, higher = better) on the clarity of the recommendation terms.

The bar charts (Figure 5.8) show the graphical representation of subjects’ responses for the two different experimental systems. Similarly, the radar chart (Figure 5.9) shows the graphical representation of the *median* value of subjects’ responses for the two different experimental systems.

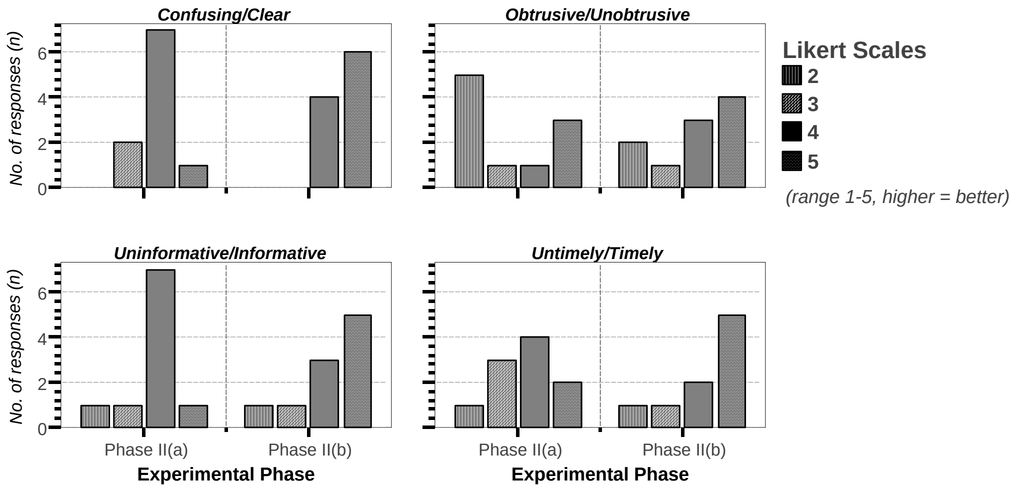


**Figure 5.8: Bar charts of how well the system communicates.**

The bar charts illustrate that a higher numbers of subjects using the phase II (b) chose the semantic differential and the Likert scale value of ‘five’ (clear $n$ = 6, informative $n$ = 5, unobtrusive $n$ = 4, and timely n = 5) in comparison to the subjects using the phase II (a) experimental system (clear $n$ = 1, informative $n$ = 1, unobtrusive $n$ = 3, and timely $n$ = 2).

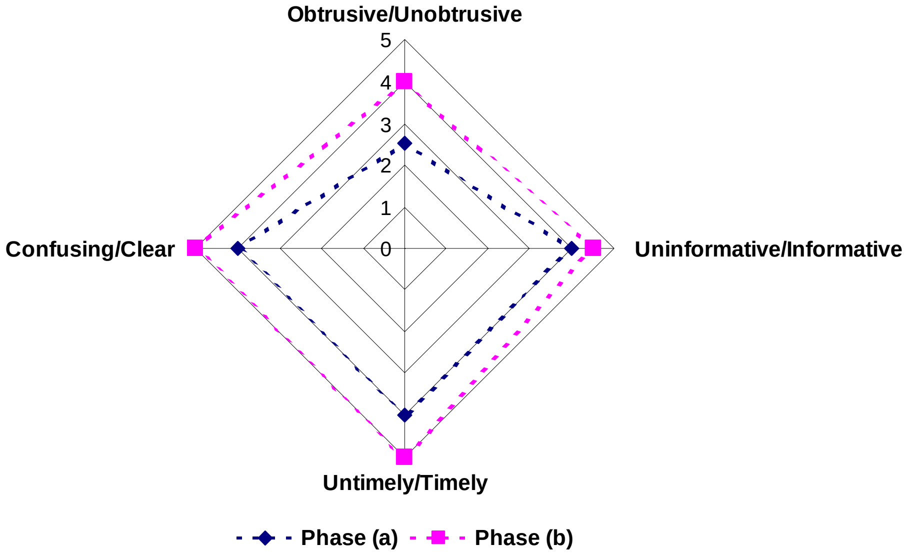


**Figure 5.9: Radar chart of how well the system communicates recommendations.**

The *median* radar chart shows the phase II (b) experimental system scoring points are near to or at the edge (*higher end of the scale*) while most phase II (a) scoring points are close to the centre point (*lower end of the scale*).

Both the bar and radar charts illustrate that the subjects using the phase II (b) experimental system tended to rate the various attributes of performance more positively in comparison to the users of the phase II (a) experimental system. However, Mann-Whitney U tests established that there is no statistically significant differences between the two experimental systems for the 'obtrusive'/'unobtrusive' (*U* = 36.00 and one-tailed *p* = .315), 'uninformative'/'informative' *(U* = 34.00 and one-tailed *p* = .247), and 'untimely/'timely' scales (*U* = 30.00 and one-tailed *p* = .243). The same test did indicate that the responses for the 'confusing'/'clear' scale were significantly different (*U* = 21.00 and one-tailed *p* = .029). The recommendation support provided by the experimental system may have enabled subjects to view clearer or a broader range of query formulation terms and had little problem reformulating their queries. This result provides some evidence that the experimental system evidently communicated its recommendation actions.

### 5.4.5.2 Conveying Preferences (*Hypothesis 3.2*)

> *Subjects find that the experimental system allows them to convey their preferences easily and in a comfortable manner.*

Subjects were asked to indicate their overall reactions to the contextual SERL search system with regard to conveying their preferences on two related questions with five semantic differentials; ‘uncomfortable’/‘comfortable’, ‘not useful’/‘useful’, ‘not in control’/‘in control’, ‘ineffective’/‘effective’, and ‘difficult’/‘easy’. The bar charts (Figure 5.10) show the graphical representation of subjects’ responses for the two different experimental systems. Similarly, the radar chart (Figure 5.11) shows the graphical representation of the *median* value of subjects’ responses for the two different experimental systems.

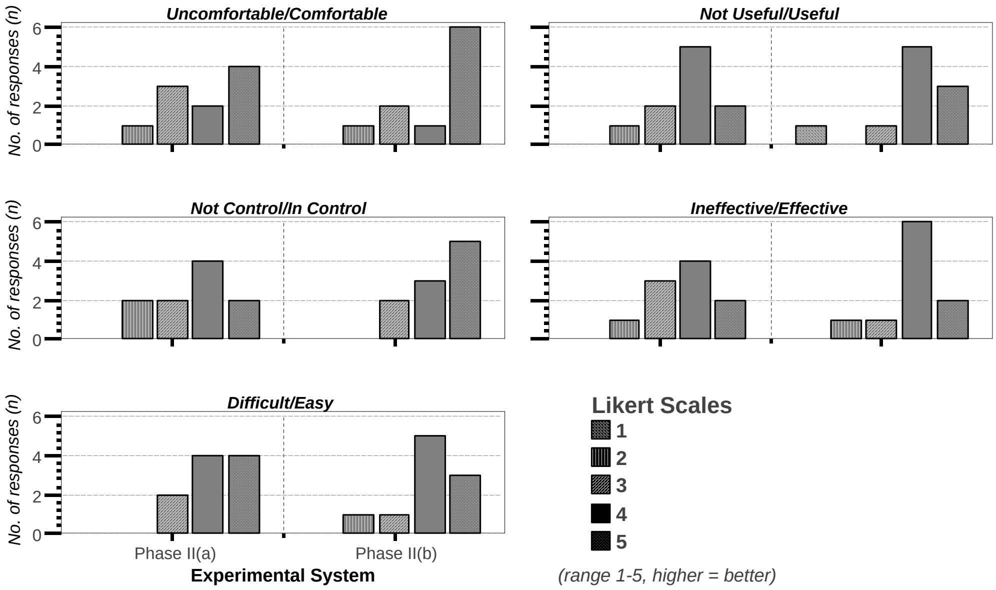


**Figure 5.10: Bar charts of how well system conveys preferences.**

The bar charts illustrate that a higher numbers of subjects using the phase II (b) experimental system (comfortable *n* = 6, useful *n* = 3, in control *n* = 5, effective *n* = 2, and easy *n* = 3) chose the semantic differential value of ‘five’ in comparison to the phase II (a) subjects (comfortable *n* = 4, useful n = *2*, in control *n* = 2, effective *n* = 2, and easy *n* = 4).

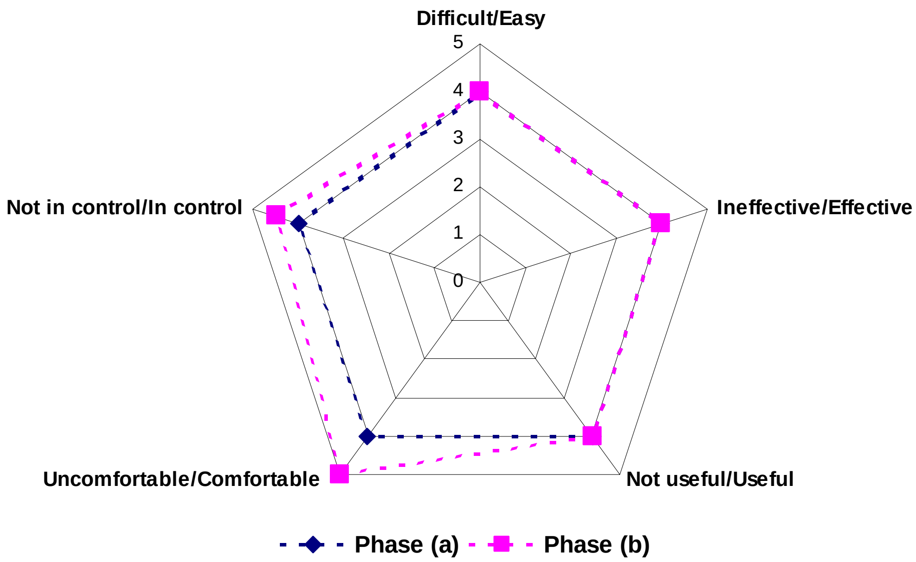


**Figure 5.11: Radar chart of how well system conveys preferences.**

The *median* radar chart shows that both experimental systems scoring points for the ‘difficult’/‘easy’, ‘ineffective’/‘effective’, and ‘not useful’/‘useful’ have the same score and are relatively near to the edge (*higher end of the scale*). The chart also shows that the phase II (b) experimental system scoring points for the ‘not in control’/‘in control’ and ‘uncomfortable’/‘comfortable’ are near to the edge (*higher end of the scale*) in comparison to the phase II (a) experimental system.

The bar charts and the *median* radar chart results are somewhat different. The bar charts illustrate that the subjects using the phase II (b) experimental system indicated in greater numbers that the system allowed them to convey their preferences easily and in a comfortable manner. In contrast, except for the not in control’/‘in control’ and ‘uncomfortable’/‘comfortable’ attributes, the *median* radar chart illustrates that subjects rated both experimental systems fairly comparable. However, Mann-Whitney U test results established no statistically significant differences between the two experimental systems for the comfortable ($U$ = 41.00 and one-tailed $p$ = .529), easy ($U$ = 45.00 and one-tailed $p$ = .739), effective ($U$ = 43.00 and one-tailed $p$ = .631), in control ($U$ = 31.00 and one-tailed $p$ = .165), and useful ($U$ = 43.50 and one-tailed $p$ = .631) scales.

All two sub-hypotheses tests yielded highly similar results (i.e., subjects relatively rated highly for the recommendation supports provided by the experimental system) and provide some evidence in favour of the third hypothesis.

## 5.4.6 Query Formulation (Hypothesis 4)

This section presents results related to the fourth experimental hypothesis*: the contextual SERL search facilitates easy, effective and reliable query formulation strategy*. This hypothesis is further divided into four sub-hypotheses that measure the query formulation aspects of the experimental system. The findings presented in this section focus on subjective impressions of the two variants of the contextual SERL search.

### 5.4.6.1 Query Formulation Strategy (*Hypothesis 4.1*)

*Subjects find the number of experimental system query formulation steps just right to formulate an effective search query.*

Subjects were asked to complete two five point Likert scales (range 1-5, higher = better) on whether the number of steps to formulate a query is just right and whether the system makes formulating a query easy. The bar (Figure 5.12) shows the graphical representation of the subjects' responses for the two different experimental systems.

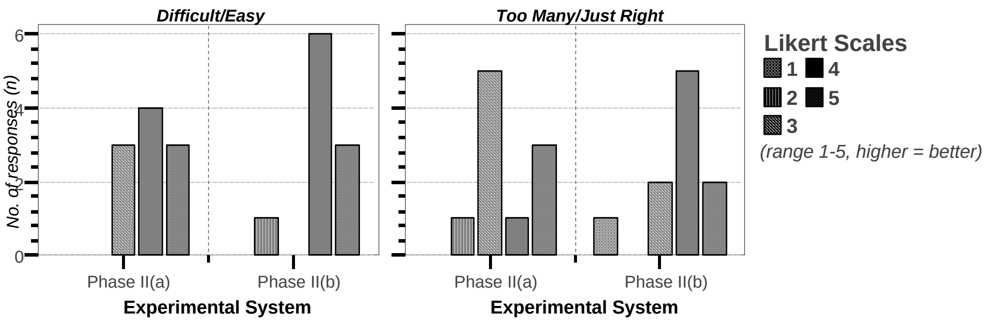


**Figure 5.12: Bar charts of query formulation strategy.**

The bar charts illustrate that a slightly higher number of subjects using the phase II (a) experimental system chose the Likert scale value of 'five' for the "too many/just right" (easy $n$ = 3 and just right $n$ = 3) in comparison to the phase II (b) experimental system users *(*easy $n$ = 3 and just right $n$ = 2). In contrast, the phase II (b) had a slightly higher *median* value ( $\tilde{X}$ = 5) for the "too many/just right" than the phase II (a) experimental system ( $\tilde{X}$ = 4.5). The bar charts and the *median* value results are inconsistent. However, Mann-Whitney U test results established no statistically significant differences between the two experimental systems for the just right ($U$ = 43.50 and one-tailed $p$ = .631) and easy ($U$ = 45.50 and one-tailed $p$ = .684) scales.

### 5.4.6.2 Trust-worthy (*Hypothesis 4.2*)

> *Subjects find that the experimental system chooses relevant terms to formulate an effective search query.*

Subjects were asked to complete two five point Likert scales (range 1-5, higher = better) indicating whether the formulated query was effective and whether they would trust the system to choose the search terms to formulate an effective search query. The bar (Figure 5.13) shows the graphical representation of the subjects' responses on two different experimental systems.

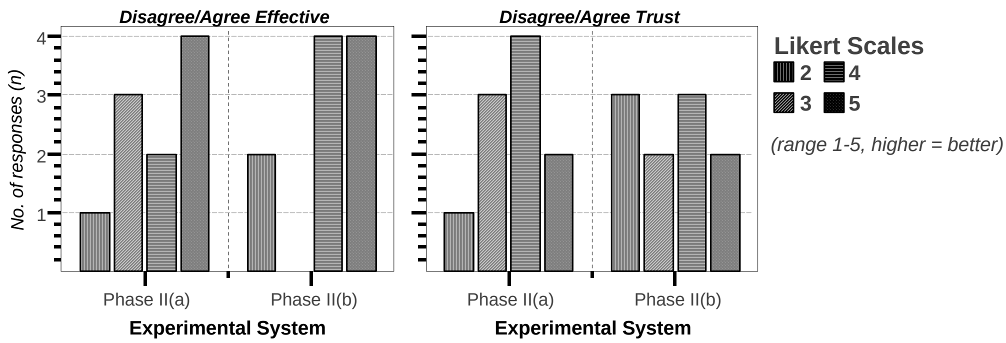


**Figure 5.13: Bar charts of trusting the system to choose the search terms.**

The bar charts illustrate that an equal number of subjects chose the Likert scale value of 'five' for both experimental systems across both attributes (agree effective $n$ = 4 and agree trust $n$ = 2). However, the higher number of subjects using the phase II (b) (agree effective $n$ = 4) chose the Likert scale value of 'four' in comparison to the phase II (a) experimental system (agree effective $n$ = 2). In contrast, the higher number of subjects using the phase II (a) (agree trust $n$ = 4) chose the Likert scale value of 'four' in comparison to the phase II (b) experimental system (agree trust $n$ = 3). Similarly, the phase II (a) had a slightly higher *median* value ($\tilde{X}$ = 4) for the "disagree/agree trust" than the phase II (b) experimental system ($\tilde{X}$ = 3.5). The bar charts and the *median* value results are inconsistent. However, Mann-Whitney U test results established no statistically significant differences between the two experimental systems for the effective

($U$ = 47.00 and one-tailed $p$ = .853) and trust ($U$ = 42.50 and one-tailed $p$ = .579) scales.

The fourth hypothesis's two sub-hypotheses tests yielded similar results (i.e., subjects rated relatively highly for the query formulation supports provided by the experimental system) and provide some evidence in favour of this hypothesis.

### 5.4.7 Interface Support (Hypothesis 5)

This section presents results related to the fifth experimental hypothesis*: the interface support provided by the contextual SERL search facilitates effective information access*. This hypothesis is further divided into two sub-hypotheses that measure the quality of user interface of the experimental system. The findings presented in this section focus on subjective impressions of the two variants of the contextual SERL search.

#### 5.4.7.1 Relevance of Content (*Hypothesis 5.1*)

> *Subjects find that the experimental system interface presents useful and effective information*.

Subjects were asked to indicate their overall reactions to the contextual SERL search system with regard to the information laid out on the results page, on four semantic differentials; 'difficult'/'easy', 'not effective'/'effective', 'frustrating'/'satisfying' and 'not useful'/'useful'. The bar charts (Figure 5.14) show the graphical representation of subjects' responses for the two different experimental systems. Similarly, the radar chart (Figure 5.15) shows the graphical representation of the *median* value of subjects' responses for the two different experimental systems.

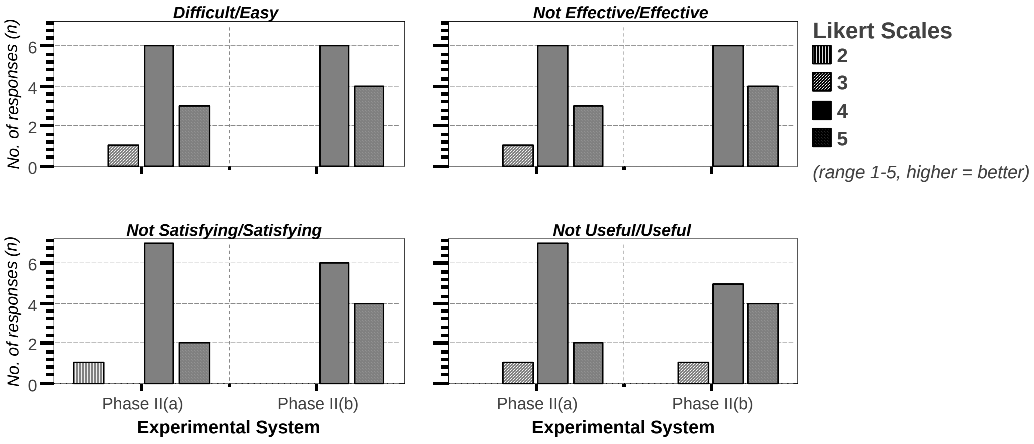


**Figure 5.14: Bar charts of relevance of content.**

The bar charts illustrate that higher numbers of subjects using the phase II (b) experimental system chose the semantic differential value of 'five' (easy *n* = 4, effective *n* = 4, satisfying *n* = 4 and useful *n* = 4) in comparison to the phase II (a) experimental system users (easy n = *3*, effective n = *3*, satisfying n = *2* and useful n = *2*).

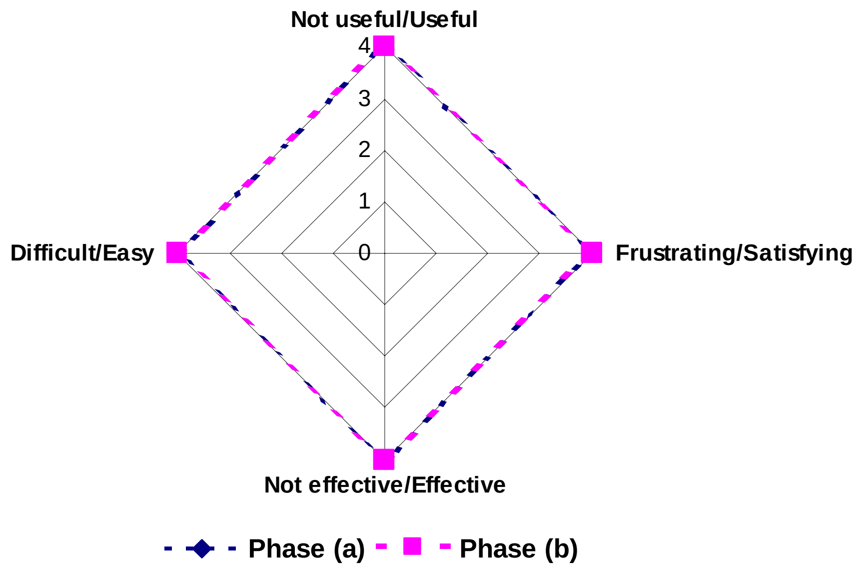


**Figure 5.15: Radar chart of relevance of content.**

The *median* radar chart shows that both experimental systems scoring points for all attributes have same score (*n* = 5) and are on the edge (*higher end of the scale*) of chart.

The bar charts and the *median* radar chart results are somewhat different. The bar charts illustrate that the subjects using the phase II (b) experimental system indicated in greater numbers that the system interface presented useful and effective information. In contrast, the *median* radar chart illustrates that both experimental systems are rated equally by respective users. However, Mann-Whitney U test results established no statistically significant differences between the two experimental systems for the useful ($U$ = 41.00 and one-tailed $p$ = .529), satisfying ($U$ = 37.00 and one-tailed $p$ = .353), easy ($U$ = 42.00 and one-tailed $p$ = .579), and effective ($U$ = 42.00 and one-tailed $p$ = .579) scales.

### 5.4.7.2 Interface guide (*Hypothesis 5.2*)

> *Subjects find that the experimental system interface guides them to the information they need.*

Subjects were asked to complete two five point Likert scales (range 1-5, higher = better) indicating whether the interface guides them to the information they need and whether they managed to find what they are looking for. The bar chart (Figure 5.16) shows the graphical representation of the subjects' responses for the two different experimental systems.

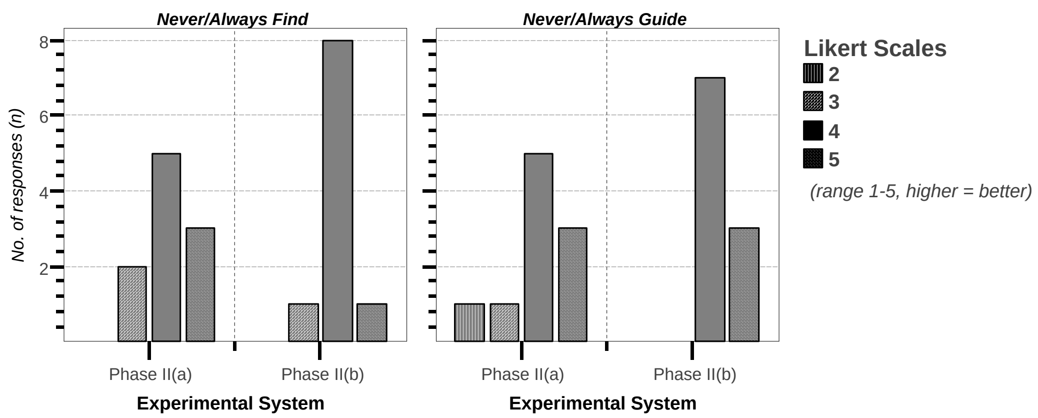


**Figure 5.16: Bar charts of interface guide.**

The bar charts illustrate that an equal number of subjects chose the Likert scale value of 'five' for both experimental systems for the "never/always guide" attribute ($n$ = 3). A slightly higher number of subjects using the phase II (a) experimental system chose the Likert scale value of 'five' for the "never/always find" attribute ($n$ = 3) in comparison to the phase II (b) experimental system users ($n$ = 1). Both experimental systems have the same *median* values ( $\tilde{X}$ =

5). Similarly, Mann-Whitney U test results established no statistically significant differences between the experimental systems for the always guide ($U$ = 43.00 and one-tailed $p$ = *.631)* and always find ($U$ = 45.50 and one-tailed $p$ = .739) scales.

The fifth hypothesis's two sub-hypotheses tests yielded highly similar results (i.e., subjects rated relatively highly for the interface supports provided by the experimental system) and provide some evidence in favour of this hypothesis.

### 5.4.8 Search Behaviour

This section presents a number of observations that characterise subjects' searching behaviour, especially the use of Boolean query and query refinement, formulation of search strategies and making use of more than three search queries, while engaged in search tasks. Whilst this analysis is not necessary to test the hypotheses, the factors may have an impact on subject perceptions. The results presented are drawn from observation video clips and supported by observation notes and the system logs.

During the experiments, it was observed that nearly half of the subjects (14, or 47%) made use of Boolean operators when searching for relevant information in the given six tasks. The remaining subjects did not make use of Boolean operators but they were still able to retrieve accurate information. In addition, nearly 70% of the subjects (21) used more than three search queries in order to retrieve the required information in all three experimental systems. It was notable that most of the subjects used a better choice of keywords to search for the information. All 30 subjects refined their queries with keywords to meet the actual details of the queries. It was observed that most of the subjects were able to get the precise answers after refining all of the given tasks. The overall success in searching may have been influenced by the generally high levels of search experience (more than 4 years of search experience) in many of the subjects – and their background, being computing literate. However, no strategic approach was evident because none of the subjects formulated any search strategy in each of the three phases.

## 5.5 Chapter Summary

This chapter has presented the results drawn from the anonymous questionnaires and two experimental systems: component testing (CT) and observation study (OS). The CT results indicated that the contextual SERL search could correctly perform various functions, such as relevance feedback, Boolean query formulation, and result presentation, under typical search conditions. The results also indicated that the concept recommender (CR) component performed best in improving precision compared to the rest of the contextual SERL search components and Google. The results also indicated that smaller queries might benefit more from the contextual SERL search components. Table 5.3 summarises the observation study results for each of the sub-hypotheses described in this chapter.

**Table 5.3: Evidence to support experimental hypotheses.**

| Hypothesis | Supported | Evidence |
|---|---|---|
| Find Information Readily (*Hypothesis 1*) | | Section 5.4.3 |
| *Number of queries (Hypothesis 1.1)* | ✓ | *Section 5.4.3.1* |
| *Number of clicks (Hypothesis 1.2)* | ✓ | *Section 5.4.3.2* |
| *Number of hits (Hypothesis 1.3)* | ✓ | *Section 5.4.3.3* |
| *Number of URLs (Hypothesis 1.4)* | ✓ | *Section 5.4.3.4* |
| *Length of Time (Hypothesis 1.5)* | ✓ | *Section 5.4.3.5* |
| Adaptiveness Support (Hypothesis 2) | | Section 5.4.4 |
| *Remember last search actions (Hypothesis 2.1)* | ✓ | *Section 5.4.4.1* |
| *Use last actions to recommend terms (Hypothesis 2.2)* | ✓ | *Section 5.4.4.2* |
| *Learn interest over the time (Hypothesis 2.3)* | ✓ | *Section 5.4.4.3* |
| *Use of other user's search actions (Hypothesis 2.4)* | ✓ | *Section 5.4.4.4* |
| Recommendation Support (Hypothesis 3) | | Section 5.4.5 |
| *Recommendation strategy (Hypothesis 3.1)* | ✓ | *Section 5.4.5.1* |
| *Conveying preferences (Hypothesis 3.2)* | ✓ | *Section 5.4.5.2* |
| Query Formulation (Hypothesis 4) | | Section 5.4.6 |
| *Query formulation strategy (Hypothesis 4.1)* | ✓ | *Section 5.4.6.1* |
| *Trust-worthy (Hypothesis 4. 2)* | ✓ | *Section 5.4.6.2* |
| Interface Support (Hypothesis 5) | | Section 5.4.7 |
| *Relevance content (Hypothesis 5.1)* | ✓ | *Section 5.4.7.1* |
| *Interface guide (Hypothesis 5.2)* | ✓ | *Section 5.4.7.2* |

The Hypothesis 1 results are based on performance comparison between the contextual SERL search; i.e., phase II (a) and phase II (b), and the contemporary search engine, i.e., phase II (c). Similarly, the Hypothesis 2 - 5 results are based on subjects' satisfaction on the contextual SERL search's individual profile vs. individual profile together with shared knowledge base profile. All hypotheses are supported by the empirical findings as described in respective sub-sections. The next chapter discusses the implications of the observational study results in more detail.

# Chapter 6

# Discussion & Research Summary

## 6.1 Introduction

The detailed evaluation methodology and the results of the observational study experimental phase are described in Chapter 4 and Chapter 5 respectively. This chapter considers the implications of the observational study research results in the context of this thesis and its overall objectives. This chapter also presents a summary of this research study.

## 6.2 Results Summary

This section discusses the major findings of the study. Potential explanations for the findings that are related to five hypotheses are discussed, along with their implications. Subjects' actual comments on the experimental systems are presented to complement the statistical analysis given in Chapter 5. No alterations are made on subjects' comments so as to avoid giving false or misleading information.

### 6.2.1 Effectiveness and Efficiency

This section presents the implications of the first experimental hypothesis (i.e., find information readily). This hypothesis was to test the performance (i.e., effectiveness and efficiency) aspects of the contextual SERL search relative to the contemporary search engine. Table 6.1 shows a summary of statistical analysis outcomes relating to the subjects' ability to find information readily, which measured the effectiveness and efficiency of the contextual SERL search either with a personal contextual profile or with both a personal contextual profile and a shared contextual knowledge base, in comparison to a contemporary search engine. The results demonstrated that the contextual SERL search impact is significant (*actual p-values bolded*) in terms of the *number of hits browsed* (H 1.3) and *number of URLs visited* (H 1.4) for overall search task completion.

**Table 6.1: Statistical analysis of Hypothesis 1 for all six search tasks.**

| Hypothesis | Parametric | | | Non-Parametric | | Post Hoc |
|---|---|---|---|---|---|---|
| | A | Dtac | Dtbc | $\chi^2$ | KW | THbc |
| H 1.1 | - | - | - | 3.056 | 0.217 | - |
| H 1.2 | 0.413 | 0.474 | 0.162 | 0.954 | 0.621 | - |
| **H 1.3** | **0.005** | 0.163 | **0.001** | 10.448 | **0.005** | **0.004** |
| **H 1.4** | **0.004** | 0.118 | **0.001** | 10.648 | **0.005** | **0.001** |
| H 1.5 | 0.242 | 0.559 | 0.102 | 3.510 | 0.173 | - |

**Legend**
**A = One-way ANOVA**
**KW = Kruskal-Wallis**
**THbc = Tukey's honestly between Phase (b) & Phase (c)**
**Dtac = Dunnett's one-tailed test between Phase (a) & Phase (c)**
**Dtbc = Dunnett's one-tailed test between Phase (b) & Phase (c)**

Further detailed tests of the non-significant hypotheses were performed for *individual* search tasks. The results (Table 6.2) indicated that the contextual search impact is significant in terms of the *number of queries* (H 1.1), *number of clicks* (H 1.2), and *length of time* (H 1.5) for *search task six.* No significant differences were found for the remaining search tasks.

**Table 6.2: Statistical analysis of Hypothesis 1 for search task six.**

| Hypothesis | Parametric | | | Non-Parametric | | | | Post Hoc |
|---|---|---|---|---|---|---|---|---|
| | A | Dtac | Dtbc | $\chi^2$ | KW | U | MU | THbc |
| H 1. 1 | **-** | | | 7.009 | **0.030** | 17.00 | **0.011** | **-** |
| **H 1.2** | 0.041 | 0.145 | **0.011** | 8.987 | **0.011** | **-** | **-** | **0.032** |
| **H 1.5** | 0.011 | 0.056 | **0.003** | 9.177 | **0.010** | **-** | **-** | **0.009** |

**Legend**
**MU = Mann-Whitney U test between Phase (b) & Phase (c)**

The overall finding for the find information readily (Hypothesis 1) is the contextual SERL search delivered either equivalent or improved Web search effectiveness (Table 6. 2), as subjects entered fewer queries to reach the target information in comparison to those using the contemporary search engine. In the case of a particularly complex search task, efficiency (Table 6.1 and 6.2) was improved as subjects browsed fewer hits, visited fewer URLs, clicked fewer clicks, and took less time to reach the target information when compared to the contemporary search engine. These improvements can be directly attributed to three related mechanisms: user profile modelling, query expansion, and relevance feedback that the contextual SERL search utilises. Given that the contextual SERL search incorporates these mechanisms, the fact that there is no additional performance overhead associated with them is a very promising result. The implications of these findings are discussed below.

The effectiveness of the contextual SERL search improvement can be directly attributed to the use of the query expansion and relevance feedback mechanisms. When using the contextual SERL search, subjects are shown three different ambiguity resolution mechanisms (i.e., disambiguater, Meta formulation and concept recommender) to suggest query expansion terms (i.e., disambiguated terms, Meta keywords, and concepts). These terms appeared to reflect subjects' current information needs as subjects progressively entered fewer search keywords manually and utilised the ambiguity resolution mechanisms to refine or reformulate their search query. In particular, subjects utilised the ambiguity resolution mechanisms for more complex tasks and managed to reach targeted information relatively faster in comparison to a contemporary search engine. In a related study, Jansen (2005) has demonstrated that experimental subjects are willing to accept automated assistance during the search process. In addition, Kelly and Fu (2006) have also demonstrated that queries created with term relevance feedback interfaces significantly outperformed corresponding baseline queries. The following are some of the actual comments made by subjects relating to the relevance feedback and query formulation support;

> *"I can retrieve information with a single hit but the user should make sure he gives all the information in the query. Its good that I shouldn't need to use any Boolean or query operators, just a phrase of English is sufficient to retrieve information." [S04]*

> *"Helpful for performing search without a well defined query. Limits the search results so it does not overwhelm the user." [S05]*

Similarly, the efficieness of the contextual SERL search improvement can be directly attributed to the use of the user profile modelling mechanism. When using the contextual SERL search, subjects are shown the recommended query expansion terms, the five top-most similar queries together with the five associated URLs and ranked search results, using either their own contextual personal profiles or together with the shared contextual knowledge base profiles. The provision of this targeted information, such as recommended query expansion terms, similar queries, previously visited URLs and ranked results, was intended to give subjects an additional set of sources from which they could choose their search terms. It may also have given subjects a better sense of their search context and helped them understand how the contextual support (e.g., highlighted terms based on their search context) occurred. In a related study, Teevan *et al.* (2005) have also demonstrated that the user profile built from search-related information and other information about the user can significantly improve the effective and efficient current Web search. In addition, another related study, Balfe and Smyth (2004) have also demonstrated the precision and recall benefits from the collaborative search (i.e., using shared profiles) technique. The following are some of the actual comments received relating to the recommendation and relevance feedback support from different subjects;

> *"Helps users filter some websites by selecting exact meaning. It is wonderful and efficient." [S01]*

> *“Ability to narrow your search quite effectively using suggested terms i.e., Recommended.” [S03]*

> *"Get relevant information so the search engine becomes more specific- Sharing info to the other group; find it very useful." [S09]*

> *“The category feature is very helpful, previous history visible is very useful – My preferences are stored and used later on which is very helpful.” [S15]*

Another key finding of this study is that the search context (i.e., ‘home’ and ‘office’) used in the contextual search not only facilitated the building of adaptive user profiles and the shared contextual knowledge base, but also helped to represent users' multiple search interests. The following are some of the comments made by subjects in relation to the search context;

> *"Very good at avoiding ambiguous results. Netpad creates better Boolean queries i.e., Adding job to Java was easy [S11]".*

> *"It gives you the context of the phrases." [S14]*

> *"Helps me go back to the past search history, recommends the main search areas." [S16]*

> *"Advance search features- easy to use- finds relevant information quickly." [S17]*

> *"Clear display of history sites, recommendation of those sites." [S19]*

Finally, a promising characteristic of the contextual retrieval system is that it does not force subjects to use its contextual features, nor does it interfere radically beyond the scope of their normal search activities. Most subjects commented that the accuracy and speed are the positive features of the contextual SERL search.

In summary, these results provide some evidence to suggest that when the contextual profile and the shared contextual knowledge base are used, the contextual SERL search improves subjects' overall ability to find information readily.

### 6.2.2 Satisfaction

This section presents the implications of the four hypotheses (2 – 5). These implications are focus on subjective impressions of the two variants of the contextual SERL search; phase II (a) experimental system using the personal contextual profile and phase II (b) experimental system using the personal contextual profile and the shared contextual knowledge base.

The radar chart (Figure 6.1) shows the graphical representation of the total *median* value of subjects' responses for the four hypotheses for the two different experimental systems. These results are based on subjects' satisfaction on the contextual SERL search's phase II (a) vs. phase II (b) experimental systems.

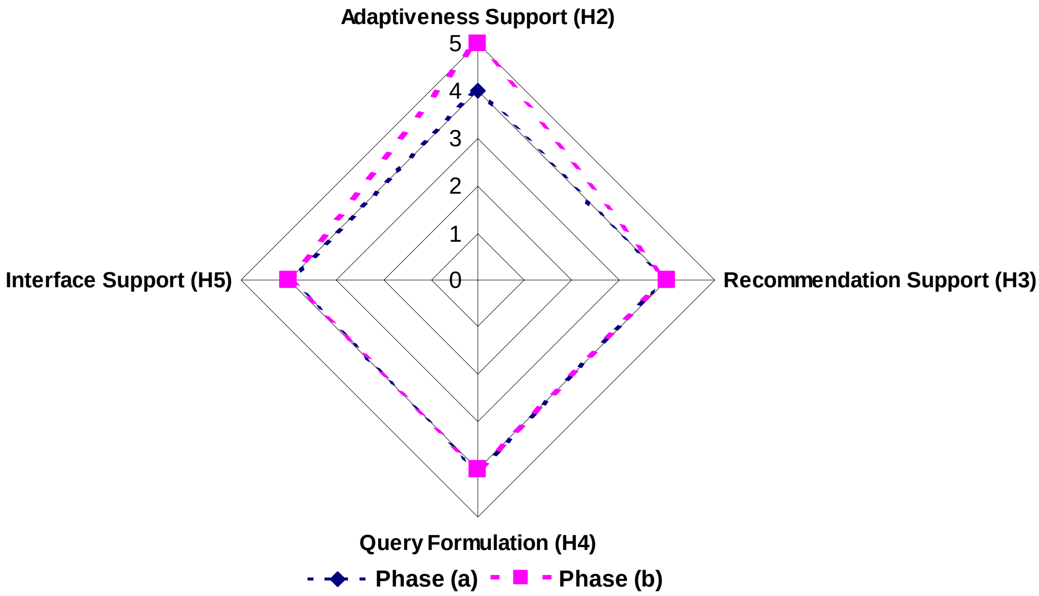


**Figure 6.1: Comparative median analysis of subjects' satisfaction.**

The *median* radar chart shows that both experimental systems scoring points for the recommendation support (Hypothesis H3), query formulation support (hypothesis H4) and interface support (Hypothesis H5) have same score and are relatively near to the edge (*higher end of the scale*). The chart also shows that the phase II (b) experimental system scoring point for the adaptive support (Hypothesis H2) is near to the edge (*higher end of the scale*) in comparison to the phase II (a) experimental system. The implications of these findings are discussed below.

In general, subjects felt that both experimental systems' contextual features indeed assisted them with their searching. It is contended here that this is due to the experimental systems adaptive capabilities, recommendation support, query formulation support, adaptive support, and effective user interfaces that ease completion of search tasks and subsequently shorten the time to complete search tasks.

For adaptive support (Hypothesis H2), subjects felt that both experimental systems remembered their last search actions, recommended terms using these last search actions and learnt their search interests over time. Particularly, subjects using the phase II (b) experimental system expressed generally higher degrees of satisfaction with the quality of adaptive support in comparison to the phase II (a) subjects. This may be because the phase II (b) system used both the user profiles and the contextual shared knowledge base to refine search queries, filter returned results from the search engine, and provide user recommendations/suggestions. As a result, the phase II (b) system utilises other subjects' profiles as a knowledge base to help existing or new subjects. The following are some of the comments made by those who used the phase II (b) experimental system regarding the use of shared profiles from different subjects;

> *"Making the search easier to find information. I can see results from my group." [S02]*

> *"It recommends the best and suited previous search results. It is a knowledge based system which keeps on warning and helps the new user from others experience." [S17]*

> *"Past recommendation from other users. Exact meaning." [S12]*

For recommendation support (Hypothesis H3), subjects felt that both experimental systems communicated their recommendations clearly, and in a timely and unobtrusive manner. The contextual SERL search gathered subjects' information seeking behaviour in an unobtrusively manner and captured their preferences in a comfortable manner. The contextual SERL search also used this information to recommend relevant information, such as similar queries, previously visited URLs and suggested query expansion terms, to subjects in a timely manner. In particular, subjects using the phase II (b) experimental system expressed generally higher degrees of satisfaction with the clarity of recommendation support in comparison to the phase II (a) subjects. This may be because the phase II (b) system used both the user profiles and the contextual shared knowledge base to provide additional assistance support (i.e., remembers last search actions, recommends dissimulated terms/Meta Keywords/concepts effectively, and conveys preferences in a comfortable manner).

For query formulation support (Hypothesis H4), subjects felt that the number of query formulation steps was just right to formulate an effective search query for both experimental systems. This could be due to the interactive query expansion strategy offered by the contextual SERL search. When using the contextual SERL search, subjects are shown three different ambiguity resolution mechanisms accompanied by suggested/recommended terms for query expansion. Subjects had the option to refine their query by selecting one or more of the different ambiguity resolution mechanisms. Subjects were delighted to see their initial query being expanded with relevant terms (or increase the length of initial query). Subjects also perceived that the expanded query returned relevant results. In a related study, Belkin *et al.* (2002) has established that experimental subjects can be more satisfied with search results if they submit longer queries to the search system. However, in circumstances where subjects/searchers may be unfamiliar with the topic of the search, they may be unable to produce longer queries (Kelly & Cool, 2002).

For interface support (Hypothesis H4), subjects felt that both experimental systems' interface presented useful and effective information. Subjects also felt that the experimental systems interface guided them to the information they need. This could be due to many of the interface design decisions (e.g., eight golden rules of interface design) made for the contextual SERL search described in Chapter 3 that arose from volunteers' (and subjects') comments during two Pilot tests (described in Chapter 4). During these Pilot tests, volunteers (Pilot test 1) and subjects (Pilot test 2) evaluated the contextual SERL search and made various suggestions for user interface improvements, including use a simple, reasonably long search field and add it to every page of the site, use simple words to explain the feature or process, focus on usability, and provide simple error messages. These comments were considered with great interest and influenced the development of the systems used in this experiment.

In summary, these findings provide some evidence to suggest that subjects expressed a higher degree of satisfaction regarding the quality of contextual support (i.e., adaptiveness, recommendation, query formulation and interface) provided by the contextual SERL search. Especially, subjects expressed a higher degree of satisfaction regarding the quality of adaptive and recommendation support provided by the contextual SERL search when the contextual profile and the shared contextual knowledge base are used. However, the statistical analysis results (especially for the query formulation and interface support) established no statistically significant differences between the two SERL experimental systems. This was due at least in part to the number of subjects in each group being restricted to 10. As a result the only types of statistical evaluation that were appropriate in most cases were non-parametric tests which are well known to have less power than their parametric counterparts (i.e. a tendency to signal no significant difference between groups when one actually exists), especially at the low sample size that was used (Bathke, Harrar, & Madden, 2008; Schulman, 2006; Wang, Li, & Stoica, 2005).

## 6.3 Research Summary

In Chapter 1, the motivation for this research, the thesis objectives, an alternative framework for contextual information retrieval (CIR) from the Web, the thesis contributions and the organisation of the thesis were presented. Chapter 2 provided a summary of the related work in the field of CIR from the Web, which informed this study and created the context within which the work is situated. From this study of the related work, and particularly the work on user profile modelling, query expansion, and relevance feedback, various challenges were identified as shown in Table 6.4. These findings also reinforced the need for research in CIR for Web IR that provides users with more relevant information more efficiently and in different ways.

In Chapter 3, the implementation of an alternative CIR framework, the contextual SERL search, is presented as a means of addressing the various challenges of CIR from the Web. The contextual SERL search utilises a rich contextual model that exploits a user's implicit and explicit data to build a user's contextual profile. The contextual SERL search also builds a shared contextual knowledge base by consolidating various users' contextual profiles. These mechanisms addressed the user profile modelling related challenges.

**Table 6.3: Summary of various challenges related to the CIR framework.**

| CIR Related Approaches | Challenges |
|---|---|
| User profile modelling | a. How to acquire, maintain and represent accurate information about a user's multiple interests with minimal intervention.<br>b. How to use this acquired information about the user to deliver personalised search results?<br>c. How to use the acquired information about various users to build a knowledge base for large communities or groups? |
| Query expansion | a. Which terms should be included in the query expansion?<br>b. How should these terms be ranked or selected?<br>c. Which levels of query reformulation should be automatic, interactive or manual? |
| Relevance Feedback | a. How to capture a user's information seeking behaviour and their preferences and structure this information in such a way as to be able to define a search context that can be refined over time?<br>b. How to help the users form communities of interest while respecting their personal privacy?<br>c. How to develop algorithms that combine multiple types of information to compute recommendations? |

The CIR framework employs an adaptive data mining technique to learn each user's specific information needs and employs a relevance feedback approach to support the iterative development of a search query by suggesting alternative terms/metakeywords/concepts for Boolean query formulation. This addressed the relevance feedback related challenges. Finally, the framework builds a Boolean query enriched with additional search terms which can then be submitted to a search engine. This addressed the query expansion related challenges.

This study employed the system development research methodology (SDRM) (Nunamaker et al., 1991) and adhered to the design-science (Edwards & Bruce, 2002) research guidelines (Hevner et al., 2004) to investigate, design, develop and implement the contextual SERL search. The contextual SERL search implemented several system activities, such as adaptation of a user's information seeking behaviour, recognition of a user's preferences and interests, recommendation of terms, generation of Boolean queries and presentation of ranked contextual search results, to address three main research objectives (section 3.1). As such, the contextual SERL search architecture was carefully implemented using three tier architecture and a component based approach, which respects the design and implementation issues of scalability, flexibility, performance, and robustness. For example, any heavier demands on the contextual SERL search (e.g., large numbers of client-request processing) can be addressed by adding additional capacities/resources at the appropriate layer without having to fundamentally change the system design or architecture. Similarly, any platform changes on the contextual SERL search can be made easily at any particular tier. Any changes on one of its layers does not affect the others layers unless a complete modification in the table design or in the API (Application Programming Interface) is made.

In Chapter 4, the methodology employed to evaluate the contextual SERL search and the hypotheses for the research are described. Two experiments; component test and observational study, were carried out to ensure that the contextual SERL search performed correctly under typical search conditions and to assess whether the system improves a user's information seeking ability. The latter was considered along the dimensions of effectiveness, efficiency and satisfaction in finding relevant information from the Web, in comparison to a contemporary search engine. The component test experiment was carried out by the researcher. A group of 30 human subjects participated in the actual observational study experiment and evaluated the capability of the contextual SERL search. The participants were all regular users of the Internet, searching for information very often.

The observational study experiment was conducted in a controlled environment in a small room. Entry questionnaires were administered before the experiment. A post-search questionnaire was filled out after the search, and an exit interview after the experiment was conducted. The subjects were required to think-aloud during the experiments. The whole session of each experiment was recorded.

Chapter 5 presented the results from the three different experiments: phase II (a) using the personal contextual profile, phase II (b) using the personal contextual profile and the shared contextual knowledge base, and phase II (c) using a contemporary search engine. This chapter (Chapter 6) discussed the implications of the observational study experimental phase results. Chapter 7 presents future further research opportunities and conclusions based on the contributions made by this thesis.

## 6.4 Chapter Summary

This chapter has presented the implications of the observational study results in the context of the objectives of the thesis and briefly presents a summary of this research study. The ramifications of this work are notable and warrant further investigation. The final chapter outlines potential avenues for such investigation in future work and draws conclusions based on the contributions made by this thesis.

# Chapter 7

# Future Work & Conclusions

## 7.1 Introduction

This thesis has presented the design, implementation and evaluation of the contextual SERL search as a contextual information retrieval (CIR) framework from the Web, which makes two main contributions. First, the contextual SERL search constructs an *evolving contextual user profile* and *a shared contextual knowledge base* to define a user's adaptive or dynamic *multiple search contexts,* which can be refined over the time. Second, the contextual SERL search *recommends alternative terms/concepts* and *formulates a dynamic Boolean search query* using the user's contextual profile and/or the shared contextual knowledge based by employing an adaptive data mining technique and relevance feedback approach. An analysis of the observational study data has shown that the contextual SERL search improves both search effectiveness and efficiency and subjective satisfaction when compared to the performance of a contemporary search engine. Likewise, many research avenues have emerged for the research described to be taken further. This chapter presents limitations of this study, future further research avenues, and then closes with conclusions based on the contributions made by this thesis.

## 7.2 Limitations of the Study

This study is preliminary and exploratory in nature and has some limitations with respect to the analysis and data that may affect the accuracy of the results. First, the observation study experiment is undertaken exclusively in AUT within a population – mainly staff, postgraduate (i.e., master's degree and Ph.D degree) and undergraduate (i.e., bachelor's degree) students – engaged in very similar activities, such as teaching, researching or studying in similar fields, and most are from the same school or department. Participants belonging to such groups that are more likely to be experienced Internet users (more than 3 years of total online experience) and who use the Internet more frequently (more than twice a day) than general Internet users (less than one year of experience). In this study, 29 out of the 30 participants were experienced Internet users. For that reason, the results obtained from the experiments may not be representative of general searchers and should not be generalised beyond this study. In addition, relatively small numbers of participants are participated in the observation study experiment. Such small sample size has a greater probability that the observation just happened to be particularly good or particularly bad. Therefore it is harder to find significant relationships from the data, as statistical tests normally require a larger sample size to justify that the effect did not just happen by chance alone. However, their collective experimental data is extensive and their views are relevant.

Second, the experiments in this study were carefully conducted in a laboratory experimental environment using the simulated search tasks. Great care has been taken to control the situational variables that can impart upon a user's search experience. In such situations, the cognitive load placed on users may increase significantly and the extent to which they can concentrate on the task in hand may decrease. Further, the simulated search tasks may not represent the actual Web search. If the experiments are carried out in real search environment with real search tasks, the results may be different to those seen in this study. It is, therefore, difficult to make any concrete statements on the results obtained from the experiments.

Third, the prototype evaluated in this study is a preliminary prototype. There are a few assumptions and known functionality limitations and technical challenges in the implemented prototype. For example, it has been assumed that finding relevant information on the Web more readily than at present is sufficiently compelling reason for users to register and disclose their personal information seeking data to the contextual SERL search. Similarly, as a functionality limitation, only two categories of context; ‘home’ and ‘office’ and two ontologies or domain hierarchies; computer science and travel, are defined in the contextual SERL search, though both the context and domain hierarchies can be more than two (e.g., context can be personal, family, baby, mother and domain hierarchies can be shopping, music, sports, and news). Finally, as a technical challenge, though the scalability of issues are addressed technically, but has not been tested practically in any greater extent at this stage. As such, to attain useful results, the prototype need to be developed in certain standard that users would like to use i.e. it has to be a fully functional piece of software that offers improvement on the systems ordinarily available to participants. However, developing a research prototype to this standard is beyond the resources of this research. Therefore, the results of the research is constrained by these parameters and the conclusions drawn are indicative, the size and nature of the experiments, together with the results still renders this to be an important piece of research upon which further hypotheses can be based. To investigate the effect of the contextual SERL search in the future, the system needs to be further developed, and further experiments with a larger and more diverse pool of subjects need to be conducted, if possible in real-life search scenarios.

## 7.3 Future Work

This study has delivered a new CIR framework for Web information retrieval. Whilst this research has gone some way to addressing the challenges associated with contextual information retrieval, there is still potential for future improvements in understanding and capturing the underlying intent of web searchers. This section describes some of the main ongoing opportunities and challenges that have been identified in this research.

### 7.3.1 User Profiling

The personalisation parameters and techniques explored in this study represent only a small subset of the space of parameterisations and techniques. This is because the study focuses on a user's explicit and implicit data with respect to the entered queries, clicked URLs, browsed websites, the domain (i.e., lexical database and classification hierarchy), the search situation, and the task context.

Future work should investigate and incorporate more complex and comprehensive personalisation parameters that could reflect the user's multiple search interests. This might include more complete domain knowledge classifications, full lexical databases, and other contextual parameters e.g., eye-tracking, user-highlighted text, Web browser events, histories drawn from proxy storage, bookmarks, current or recent software application usage, the topics of content at the current or recent focus of attention, and time spent on website. Several previous researchers have successfully applied other personalisation parameters, such as documents and email the user has read and created (Teevan et al., 2005) and time spent on each URL (Gauch, Chaffee, & Pretschner, 2003; Hofgesang, 2007; Sendhilkumar & Geetha, 2008) to create user's profiles able to produce moderate improvements when applied to search results. In addition, the approach should incorporate intelligent and less-resource intensive processing techniques, such as agent technology and client side computation, and other techniques that could improve query formulation, Meta keywords extraction from browsed Websites, query term weighting and term recommendation that could assist the user to reach the target information more readily. For example, in this study, users explicitly define their search context (e.g., 'home' or 'office'). Future work should where possible infer the context information implicitly from the interactions between the system (e.g., 'home pc' or 'office pc') and the user.

Finally with respect to user profiling, privacy concerns continue to pose major challenges to such work, and particularly the shared profile approach. The collection of users' information seeking data is also subject to legal regulations in many countries and states. Future work should more fully address privacy, legal and security requirements (e.g., confidentiality, integrity and non repudiation).

### 7.3.2 Query Formulation

The results of the experiments presented in Chapter 5 suggested that the query formulation support provided by the contextual SERL search was fruitful. However, this study raises some important issues, which warrant further investigation. To focus attention on these issues, the following subjects' comments are useful;

*"A little half complex. Make the step simple." [S08].*

*"Sometimes it cannot find any terms... " [S11]*

*"Need to think about putting the most important key word in front of others so the system can give better help." [S11]*

*"At time the query is misleading, the -ve in the system." [S13]*

These comments suggest that the query formulation steps were rather complex and the domain-specific terms were either incomplete or (intentionally) limited to the "travel" or "computer science" domain. Future work should incorporate real time query expansion and instant query modification techniques to ease the query reformulation steps. The real time approach should present related expansion terms as a list very shortly after the searcher finishes typing the first term of their query, and should be updated after each term is typed. However, searchers should be permitted to either select a term or ignore the suggestions, and complete their query. Previous research results (White & Marchionini, 2007) show that offering real time query expansion leads to better quality initial queries, more engagement in the search, and an increase in the uptake of query expansion. In addition, future work should also incorporate comprehensive domain-specific resources, such as the Open Directory Project (www.dmoz.org).

Other issues, such as the consideration of user characteristics including cognitive and behavioural factors, were not totally integrated into the contextual SERL search. During the user studies, various searching behavioural differences among subjects were discovered, such as the use of Boolean operators, search operators, wildcards, advanced features, and search strategies (e.g., determining needed keywords). This indicates that search engine users are a heterogeneous collection and may need to be catered to differently. Future work should investigate and incorporate these factors which lead users to reformulate their queries effectively. Finally, future work should explore other enhanced Boolean query algorithms, which potentially comprise both effective and efficient factors of information retrieval.

### 7.3.3 Recommendation and Relevance feedback

The results of the experiments presented in Chapter 5 suggested that the recommendation support provided by the contextual SERL search was effective. However, there is scope for further advances to be made. This could be informed by the following subjects' comments;

> *"More rigid and restricted compared to context free search. Insists on recommending pages that have been visited regardless whether they are relevant or not, which leads to clustering."[S05]*

> *"A new user a little bit confused with the usability. Too many recommendations chosen, user needs to try each one to check the best results."[S12]*

These comments suggest that there are a number of design and research challenges in the realm of quality recommendation and relevance feedback. It indicates that exposing *all* recommended parameters can be of little or no help to users. In addition, not all visited URLs or browsed Websites are relevant to a particular context. There is a need for future work in developing hybrid positive/negative relevance feedback (explicit and implicit) mechanisms to address these challenges.

In this study, the nearest neighbour approach is employed to recommend query expansion terms and visited URLs using either a user's profile or together with the shared contextual knowledge base. Further work should investigate other more sophisticated machine learning techniques (e.g., Naïve Bayesian classification methods) together with other existing profiling parameters (such as ‘search context’) to improve the quality of recommendation. Future work should also investigate how to collect indications from the user (e.g., explicitly and implicitly rating) of what resources are relevant and/or valuable to a particular information context. This may lead to other challenges, such as who to trust, what to trust, how to measure the accuracy/meaningful rating from users and so on, especially when users share their profiles anonymously. For example, a malicious user may share multiple profiles under false identities designed to promote or demote the recommendation of a particular query. Such actions degrade the objectivity and accuracy of a recommender system, and could cause frustration for its users. Hence, future research should focus on simple yet effective recommendation techniques, such as significance weighting (Herlocker et al., 1999) and trust weighting (O'Donovan & Smyth, 2006), which not only infer users’ multiple search intents, but also allow them to perform their search task without any difficulties.

## 7.4 Conclusion

This thesis has presented research regarding the implementation and evaluation of the contextual SERL search, designed to tackle some of the challenges associated with contextual information retrieval from the Web. The system utilises a rich contextual model that exploits implicit and explicit data to modify queries to more accurately reflect the user’s interests as well as to continually build the user’s contextual profile and a shared contextual knowledge base. These profiles are used to filter results from a standard search engine to improve the relevance of the pages displayed to the user. This system has been tested in an observational study that has captured both qualitative and quantitative data about the ability of the system to improve the user’s web search experience.

The overall finding of the observational study is that the contextual SERL search delivered either equivalent or improved Web search effectiveness, as subjects actually entered fewer queries to reach the target information in comparison to the contemporary search engine. In the case of a particularly complex search task, efficiency was improved as subjects browsed fewer hits, visited fewer URLs, made fewer clicks and took less time to reach the target information when compared to the contemporary search engine. Given that the contextual SERL search incorporates additional capabilities (in terms of word sense disambiguation, term recommendation and so on) the fact that there is no additional performance overhead is a promising result. Finally, subjects expressed a higher degree of satisfaction on the quality of contextual support provided by the contextual SERL search when used with their contextual profile and shared contextual knowledge base. These results suggest that integration of a user’s information seeking behaviours is important in the successful development of the CIR framework.

However, this study is just one step in this direction. Due to the complex nature of CIR framework study, it is impossible to consider and incorporate all the factors that could have an impact on the effectiveness, efficiency and subjects’ satisfaction on the contextual SERL search. The results of this study serve as a partial view of the phenomenon. More research needs to be done in order to validate or invalidate these findings, using larger samples, and if possible in a real-life scenario. This study can be built on in several directions, including integration of complex and comprehensive personalisation contextual parameters, real time query expansion, quality recommendation and relevance feedback.

This thesis has contributed to a better understanding of how information seeking behaviours, such as entered query, selected terms, and visited URLs each time they searched for information, can be used in the development and maintenance of a user’s personal contextual profile and a shared contextual knowledge base. Furthermore, it has contributed to a better understating of how the user’s personal contextual profile and/or the shared contextual knowledge base can be used to refine search queries, filter returned results from search engines, and provide user recommendations/suggestions. The main achievement of this thesis is the development of a CIR framework from the Web that incorporates the above functionality. It is believed that this framework and other similar projects will help provide the basis for the next generation of contextual information retrieval from the Web.

# Appendices

**Appendix A:** *Experimental Documents*

**Appendix B:** *Component Testing*

**Appendix C:** *Anonymous Questionnaires Analysis*

**Appendix D:** *Normality Test.*

**Appendix E:** Descriptive Statistics and Inferential Analysis

## 8.1 Appendix A: Experimental Documents

This Appendix includes the documentation used in the experiment described in Chapter 4 of this thesis. These include:

A.1. Anonymous Questionnaire

A.2. Participation Request Form

A.3. Participant Information Sheet

A.4. Consent Form

A.5. Search Tasks and Task Answers/Notes

A.6. ‘Entry’ and ‘Post-observation’ Questionnaires

  A.6.1*. Used for Contextual SERL Search*

  A.6.2. *Used for Contemporary Search*

A.7. Observation Form

Appendix A.1: Anonymous Questionnaire

# ANONYMOUS QUESTIONNAIRE

**Web Information Seeking Behaviour Investigation**

We are currently undertaking research that will lead to better web searching, enabling users to find relevant information more readily than at present. To help us do this, we need to have an up-to-date understanding of who ‘users’ are, and how and why they search for information using the web.

The aim of this questionnaire is to gather data about your web information seeking behaviour. The data will be used in a PhD project (*i.e. Contextual Information Retrieval from the WWW*) at the Software Engineering Research Lab (*SERL*) in the School of Computing and Mathematical Sciences, Auckland University of Technology (*AUT*), New Zealand. You must be at least 18 years old to participate in this research study. Completion of the questionnaire indicates consent to participate in this study. All information, which is collected about you, will be kept strictly confidential.

This project has been approved by the Auckland University of Technology Ethics Committee on 17/05/2007; AUTEC’s Ethics Application number is 07/12. If you have any questions regarding your rights as a participant, please contact Dr Robert Wellington, Auckland University of Technology Ethics Committee (AUTEC) faculty representative, by email at RWelling@aut.ac.nz or by telephone on 921 9999 at extension 5432.

If you have any questions about the questionnaire itself, please feel free to call us or email us, the researcher or the project supervisor, using the contact details below.

This information sheet is yours to keep.

Sincerely,

*Researcher*
Dilip Limbu
(email: dlimbu@aut.ac.nz or tel: +64 9 921 9999 extn 8953)

*Project Supervisor*
Dr. Andy Connor
(email: andrew.connor@aut.ac.nz or tel: +64 9 921 9999 extn 5211)

# ANONYMOUS QUESTIONNAIRE

## Web Information Seeking Behaviour Investigation

AUCKLAND UNIVERSITY OF TECHNOLOGY
TE WĀNANGA ARONUI O TAMAKI MAKAU RAU

Please CIRCLE/TICK/WRITE your answers as appropriate.

## Section 1: PERSONAL DETAILS

1.1 Please indicate your AGE RANGE:

☐ 18-25 ☐ 31- 40 ☐ > 50

☐ 26-30 ☐ 41- 50

1.2 Please Indicate your GENDER

☐ Male ☐ Female

1.3 What is your area of WORK?

☐ student ☐ research

☐ academia ☐ other: please state ___________________

1.4 What is the highest EDUCATIONAL QUALIFICATION that you have?

☐ PhD ☐ Bachelors degree

☐ Masters degree ☐ other: please state ___________________

## Section 2: SEARCH EXPERIENCE

2.1 How experienced are you in:

| | novice 1 | 2 | 3 | 4 | expert 5 |
|---|---|---|---|---|---|
| • searching using World Wide Web search services (*e.g., Alta Vista, Google, Excite, Yahoo, HotBot, and WebCrawler*)?............................................. | ☐ | ☐ | ☐ | ☐ | ☐ |
| • searching with other search services that are not mentioned above, please specify? | | | | | |
| a. ___________________ ............................. | ☐ | ☐ | ☐ | ☐ | ☐ |
| b. ___________________ ............................. | ☐ | ☐ | ☐ | ☐ | ☐ |
| c. ___________________ ............................. | ☐ | ☐ | ☐ | ☐ | ☐ |

## Section 2: SEARCH EXPERIENCE *(cont.)*

2.2 How often do you use a search service/engine for:

| | Once or twice a year | Once or twice a month | Once or twice a week | Once or twice a day | More often |
|---|---|---|---|---|---|
| ▪ work/academic tasks?..... | ☐ | ☐ | ☐ | ☐ | ☐ |
| ▪ personal tasks?............. | ☐ | ☐ | ☐ | ☐ | ☐ |

2.3 Please indicate your years of Web searching experience:

☐ 1-2 years ☐ 3 years ☐ > 4 years

☐ 2 years ☐ 4 years

2.4 When you search the Internet, you can usually find what you're looking for

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| rarely | ☐ | ☐ | ☐ | ☐ | ☐ | often |

2.5 Finding relevant information on the Internet has taken up more of your time than you expected

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| rarely | ☐ | ☐ | ☐ | ☐ | ☐ | often |

2.6 Please list your top five favourite search services/engine(s) starting with your all-time favourite :

a. ____________________

b. ____________________

c. ____________________

d. ____________________

e. ____________________

## Section 3: SEARCH BEHAVIOUR

3.1 The first resource you use for finding information is the Internet:

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| rarely | ☐ | ☐ | ☐ | ☐ | ☐ | often |

## Section 3: SEARCH BEHAVIOUR *(cont.)*

3.2 In acquiring relevant information on the Internet you make use of informal resources such as :

| | rarely 1 | 2 | 3 | 4 | often 5 |
|---|---|---|---|---|---|
| • reference URLs from friends, family, and colleagues... | ☐ | ☐ | ☐ | ☐ | ☐ |
| • previously bookmarked URLs ................................ | ☐ | ☐ | ☐ | ☐ | ☐ |
| • media such as computer magazines, books etc .... | ☐ | ☐ | ☐ | ☐ | ☐ |
| • search history from a browser ........................... | ☐ | ☐ | ☐ | ☐ | ☐ |
| • others a. ____________________....................... | ☐ | ☐ | ☐ | ☐ | ☐ |
| b. ____________________....................... | ☐ | ☐ | ☐ | ☐ | ☐ |
| c. ____________________....................... | ☐ | ☐ | ☐ | ☐ | ☐ |

3.3 You plan your search strategy in advance prior to searching the Internet

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| rarely | ☐ | ☐ | ☐ | ☐ | ☐ | often |

3.4 In resolving information problems you make use of more than one query

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| rarely | ☐ | ☐ | ☐ | ☐ | ☐ | often |

3.5 In acquiring relevant information on the Internet you make use of search features such as :

| | rarely 1 | 2 | 3 | 4 | often 5 |
|---|---|---|---|---|---|
| • Boolean query............................................. | ☐ | ☐ | ☐ | ☐ | ☐ |
| • advanced search features (*e.g. search within results, find similar, sort by date, date range*) ................................. | ☐ | ☐ | ☐ | ☐ | ☐ |
| • a directory or classification ................................ | ☐ | ☐ | ☐ | ☐ | ☐ |
| • others a. ____________________..................... | ☐ | ☐ | ☐ | ☐ | ☐ |
| b. ____________________..................... | ☐ | ☐ | ☐ | ☐ | ☐ |
| c. ____________________..................... | ☐ | ☐ | ☐ | ☐ | ☐ |

3.6 In acquiring relevant information from the returned results you generally browse through the first

| 10 hits | 20 hits | 30 hits | 40 hits | 50 hits or more |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

## Section 3: SEARCH BEHAVIOUR (*cont.*)

3.7 State the criteria you use in judging relevance on returned search results, prior to browsing :

| | rarely 1 | 2 | 3 | 4 | often 5 |
|---|---|---|---|---|---|
| • title ............................................................. | ☐ | ☐ | ☐ | ☐ | ☐ |
| • descriptions.......................................................... | ☐ | ☐ | ☐ | ☐ | ☐ |
| • highlighted words ................................................ | ☐ | ☐ | ☐ | ☐ | ☐ |
| • reading URL itself ................................................ | ☐ | ☐ | ☐ | ☐ | ☐ |
| • search engine's rankings .................................... | | | | | |
| • others ................................................................ | ☐ | ☐ | ☐ | ☐ | ☐ |
| • others a. ____________________..................... | ☐ | ☐ | ☐ | ☐ | ☐ |
| b. ____________________..................... | ☐ | ☐ | ☐ | ☐ | ☐ |
| c. ____________________..................... | ☐ | ☐ | ☐ | ☐ | ☐ |

3.8 In relevance judgments you select just those hits which you can use immediately

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| rarely | ☐ | ☐ | ☐ | ☐ | ☐ | often |

## Section 4: SEARCH PREFERENCE

4.1 How do you view your usage of search engines?

☐ as a single user
☐ both
☐ as a member of a Community/Group

4.2 Would you be willing to register yourself to a search service/engine that adapts and stores your searching behaviour (*as a user search profile*) and possibly uses this to return relevant information from the Internet (*for your future search tasks*)?

☐ yes ☐ no

4.3 Would you like to share your searching behaviour information to form a community/group/knowledgebase of searching behaviours?

☐ yes – If YES, please state, you would like to do so (mark **ONE** only):

☐ i) ANONYMOUSLY

☐ ii) KNOWN

☐ no

## Section 4: SEARCH PREFERENCE (*cont.*)

4.4 Would you like a search service/engine to suggest or recommend related terms (*such as thesaurus terms, related concepts*) to help you to better define your search context or interest?

☐ yes - IF YES, such terms should be extracted in the first instance from (mark **ONE** only):

☐ i) your own profile

☐ ii) the community/group/knowledgebase

☐ ii) using both

☐ no

4.5 Would you allow the search service/engine to recommend you as a "user with similar interests" to other users or communities?

☐ yes ☐ no

4.6 In your opinion, what makes a search service/engine useful?

________________________________________

________________________________________

________________________________________

Kindly return a completed copy of this questionnaire by July 17th 2007 by either dropping it in the Anonymous Questionnaire Drop-In-Box placed at the reception of School of Computing and Mathematical Sciences or by PLACING IT IN A SEPARATE ENVELOPE and posting it to the address below;

| | | |
|---|---|---|
| Dilip Limbu | | Dilip Limbu |
| Internal Mail No D-75 | *or* | Software Engineering Research Lab, AUT |
| | | Room WT406 |
| | | Private Bag 92006 |
| | | Auckland 1142 |
| | | New Zealand |

Thank you for participation and **your co-operation**!

Appendix A.2: Participant Request Form

# PARTICIPATION REQUEST FORM

**Web Information Seeking Behaviour Investigation**

## PARTICIPATION REQUEST FORM

**In addition to collecting data through the questionnaire, we will also be conducting short (30 minute) searching tests or actual experiments with users, following some predefined search scenarios and using new search engine.**

**The experiments will commence from July 24th to Aug 7th 2007 and will be carried out during normal office hours (*09.00 hours to 17:00 hours*). The experiments will take place at AUT's WT building (*cnr of Wakefield Street and Rutland Street*). If you are willing to take part in this experimentation, please leave your email address below and return this FORM. Thank you.**

Your email address :

________________________________________

Kindly return a completed copy of this participation request form by **July 17th 2007** either by dropping it in the Participation Request Drop-In-Box placed at the reception of School of Computing and Mathematical Sciences or by PLACING IT IN A SEPARATE ENVELOPE and posting it to the address below;

| | | |
|---|---|---|
| Dilip Limbu<br>Internal Mail No D-75 | *or* | Dilip Limbu<br>Software Engineering Research Lab, AUT<br>Room WT406<br>Private Bag 92006<br>Auckland 1142<br>New Zealand |

Thank you for participation and **your co-operation**!

Appendix A.3: Participant Information Sheet

# Participant Information Sheet

### Date Information Sheet Produced:

11 July 2007

### Project Title

Contextual Information Retrieval from the World Wide Web (WWW)

### An Invitation

I am a PhD student at Auckland University of Technology (*AUT*) and I am currently undertaking research study that will lead to better web searching, enabling users to find relevant information more quickly than at present.

You are being invited to take part in this research study. Your participation in this study is completely voluntary and you can withdraw from this study at any time, without explanation. You also have the right to withdraw retrospectively any consent given, and to require that any data gathered on you be destroyed. A decision not to participate will not affect your grades or work performances in any way.

The results of this research study will be used for my Ph.D. research. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take as much time as you would like to read this information carefully.

### What is the purpose of this research?

The aim of this research study is to investigate, prototype and evaluate an alternative framework for contextual information retrieval (*CIR*) from the World Wide Web (*WWW*). The CIR framework aims to improve query results (*or make search results more relevant*), enabling users to find relevant information more quickly than at present. We cannot determine the value of the prototyped CIR framework or contextual search unless we ask those people who are likely to be using them, which is why we need to run experiments like these. Please remember that it is the prototyped CIR framework being evaluated and not you.

### How was I chosen for this invitation?

You were chosen, along with 30 others, because you've indicated your interest in participating in the experimentation in “ANONYMOUS QUESTIONNAIRE” and you work or study at the Auckland University of Technology (*AUT*) or the AUT Technology Park.

### What will happen in this research?

You will perform six search tasks and complete a questionnaire about using either the prototyped CIR system or the Google search. The questionnaire will ask how you felt during each search. All of your information seeking behaviours (*e.g., use of Boolean query & advanced search features, mouse clicks etc*.) and preferences (*e.g. selected terms, concepts etc.*) will be either observed or captured. You are encouraged to talk aloud while you are engaged in the search tasks, i.e., you need to describe your actions and reasons for your actions. All your actions and reasons/comments will be recorded or I will take notes if you so prefer. You will have the option to review, edit, or erase the recording. Please ask questions if you need to and please let me know when you are finished each task. You may be asked some questions about the tasks and systems at the end of the experiment.

### What are the discomforts and risks?

There are no known risks in participating.

### What are the benefits?

There are no direct benefits to you for participating in this research study.

### How will my privacy be protected?

Any data collected in this research study will remain confidential. Any analysis results that would be published will be anonymous.

### What opportunity do I have to consider this invitation?

Since this research study involves spending an hour, the researcher understands that it requires time for you to consider and get back. The researcher will appreciate if you could send a reply within two weeks of the receipt of this form.

### How do I agree to participate in this research?

If you decide to take part in this research study, you will be asked to sign a consent form on the day of study.

### Will I receive feedback on the results of this research?

You will not receive any feedback on the results of this research study as it is the Contextual search system being evaluated, not you. However, if you wish to receive a summary sheet of the experimental findings or publications, the results likely to be published in late 2007 and will be available online at http://serl.aut.ac.nz/.

### What do I do if I have concerns about this research?

Any concerns regarding the nature of this project should be notified in the first instance to the Project Supervisor, Dr. Andy Connor, andrew.connor@aut.ac.nz and Tel: +64 9 921 9999 extn 5211.

Concerns regarding the conduct of the research should be notified to the Executive

Secretary, AUTEC, Madeline Banda, *madeline.banda@aut.ac.nz* , 921 9999 ext 8044.

## Whom do I contact for further information about this research?

### *Researcher Contact Details:*

Dilip Limbu, email: dlimbu@aut.ac.nz , tel: +64 9 921 9999 extn 8953, address : Software Engineering Research Lab, Private Bag 92006, Auckland 1020, New Zealand, Email: serl@aut.ac.nz

### *Project Supervisor Contact Details:*

Dr. Andy Connor, email: andrew.connor@aut.ac.nz, tel: +64 9 921 9999 extn 5211, address: Software Engineering Research Lab, Private Bag 92006, Auckland 1020, New Zealand, Email: serl@aut.ac.nz

**Approved by the Auckland University of Technology Ethics Committee on 17/05/2007, AUTEC Reference number 07/12.**

Appendix A.4: Consent Form

# Consent Form

| | |
|---|---|
| *Project title:* | ***Contextual Information Retrieval from the WWW*** |
| *Project Supervisor:* | ***Dr. Andy Connor*** |
| *Researcher:* | ***Dilip Limbu*** |

- ○ I have read and understood the information provided about this research project in the Information Sheet dated / /2007.
- ○ I have had an opportunity to ask questions and to have them answered.
- ○ I understand that I may withdraw myself or any information that I have provided for this project at any time prior to completion of data collection, without being disadvantaged in any way.
- ○ I agree to take part in this research.
- ○ I wish to receive a copy of the report from the research (please tick one):

*Yes ○ No ○*

Participant's signature: ..........................................................................................

Participant's name: ..........................................................................................

Participant's Contact Details (if appropriate):

………………………………………………………………………………………..

………………………………………………………………………………………..

………………………………………………………………………………………..

………………………………………………………………………………………..

Date: / /2007

***Approved by the Auckland University of Technology Ethics Committee on 17/05/2007, AUTEC Reference number 07/12.***

*Note: The Participant should retain a copy of this form.*

Department: *Computing and Mathematical Sciences*

Appendix A.5: Search Tasks

# Search Tasks

## Title of Project:

Contextual Information Retrieval from the World Wide Web (WWW)

AUCKLAND UNIVERSITY OF TECHNOLOGY
TE WĀNANGA ARONUI O TAMAKI MAKAU RAU

## Name of Researcher:

Dilip Limbu

Please take time to perform all six search tasks give below.

| **Topic A : Computer Science** |
|---|
| **Note: If you are evaluating the Contextual SERL search, please check and change the contextual search context to “office”.**<br><br>**Task 1:** You have just finished reading a copy of a Java technology article. The article informs you that Java is an object-oriented programming language developed by Sun Microsystems in the early 1990s. The language derives much of its syntax from C and C++ but has a simpler object model and fewer low-level facilities. You decide to find out who is the creator or designer of the Java programming? |
| **Task 2:** You are currently working as a part time Web developer for the Software Engineering Research Lab. Your duties include developing a Website that contains various programming materials such as external tutorials via Website addresses (or URLs), downloadable slides, and development tools. Your supervisor has asked you to add THREE external Website addresses that contain an introduction on Java. You decide to find and bookmark them in your browser for later use. |
| **Task 3:** Assume that you've just graduated with a degree in computer science and you are actively looking for a job. A number of friends have advised you that a Java programmer or software engineer job would be suitable for you as you’ve good analytical and Java programming skills. At present you are unaware of these jobs’ roles and responsibilities. You would like to find a job description and activities on these two jobs that may help you to make the best career decision possible. |

**Topic B : Travel**

**Note: If you are evaluating the Contextual SERL search, please check and change the contextual search context to "Home".**

**Task 4:** Recently, you've met a friend from Java, Indonesia. She told you that Java is the world's 13th largest island and there are 124 million people living there. Out of curiosity, you decide to find out the unit of currency used on the island of Java and its exchange rate in US dollar.

**Task 5:** Indonesia has some of the very best surf areas in the world, and their locations are spectacular and exotic. These include Sumatra, Java, Bali, Lombok, Sumbawa, Flores, and Sumba. A close friend of yours is planning to go on a surfing vacation in Java island, Indonesia. He has recently moved house and does not have a phone/Internet connection installed. As a result he has asked you to look for the name of THREE main surf areas in west Java, so that he could use these names to find further information.

**Task 6:** You are about to depart on a short-tour to Indonesia. Your agenda includes a visit to the west coast of Java, which is renowned for its natural riches, its fertile land with some volcanoes and its cultural riches. As your time in the Java is limited to a week you would like to save time and find information about the west coast's any FIVE best leisure and sports activities prior to your departure.

Appendix A.5: Task Answers/Notes

# Task Answers/Notes

**Title of Project:**

Contextual Information Retrieval from the World Wide Web (WWW)

**Name of Researcher:**

Dilip Limbu

Please write your answers or any notes in the space provided below. If you require more paper, please ask the experimenter.

Task 1:

Task 2:

Task 3:

Task 4:

Task 5:

Task 6:

# Appendix A.6

## Appendix A.6.1: Used for Contextual SERL Search

## ENTRY QUESTIONNAIRE

AUCKLAND UNIVERSITY OF TECHNOLOGY
TE WĀNANGA ARONUI O TAMAKI MAKAU RAU

### Contextual Search Experimentation

☐ **Phase IIA** ☐ **Phase IIB**

**ID :** ***phase***

Please answer all the questions below by ticking/typing the appropriate box or commenting.

**Section 1: PERSONAL DETAILS**

1.1 Please indicate your AGE RANGE:

☐ 18-25 ☐ 31- 40 ☐ > 51

☐ 26-30 ☐ 41- 50

1.2 Please Indicate your GENDER

☐ Male ☐ Female

1.3 What is your area of WORK?

☐ student ☐ research

☐ academia ☐ other: please state ______________________

1.4 What is the highest EDUCATIONAL QUALIFICATION that you have?

☐ PhD ☐ Bachelors degree

☐ Masters degree ☐ other: please state __________________

## Section 2: SEARCH EXPERIENCE

2.1 How experienced are you in:

| | novice 1 | 2 | 3 | 4 | expert 5 |
|---|---|---|---|---|---|
| • searching using World Wide Web search services (*e.g., Alta Vista, Google, Excite, Yahoo, HotBot, and WebCrawler*)? ................ | ☐ | ☐ | ☐ | ☐ | ☐ |

| | novice 1 | 2 | 3 | 4 | expert 5 |
|---|---|---|---|---|---|
| • searching with other search services that are not mentioned above, please specify? | | | | | |
| a. ____________________ .............................. | ☐ | ☐ | ☐ | ☐ | ☐ |
| b. ____________________ .............................. | ☐ | ☐ | ☐ | ☐ | ☐ |
| c. ____________________ .............................. | ☐ | ☐ | ☐ | ☐ | ☐ |

2.2 How often do you use a search service/engine for:

| | Once or twice a year | Once or twice a month | Once or twice a week | Once or twice a day | More often |
|---|---|---|---|---|---|
| ▪ work/academic tasks?......... | ☐ | ☐ | ☐ | ☐ | ☐ |
| ▪ personal tasks?................. | ☐ | ☐ | ☐ | ☐ | ☐ |

2.3 Please indicate your years of Web searching experience:

☐ < 1 year ☐ 3 years ☐ > 4 years

☐ 2 years ☐ 4 years

2.4 When I search the Internet, I can usually find what I'm looking for

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| rarely | ☐ | ☐ | ☐ | ☐ | ☐ | often |

2.5 Finding relevant information on the Internet has taken up more of my time than I expected

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| rarely | ☐ | ☐ | ☐ | ☐ | ☐ | often |

2.6 Please list your top five favourite search services/engine(s):

a. ____________________

b. ____________________

c. ____________________

d. ____________________

e. ____________________

# POST- OBSERVATION QUESTIONNAIRE

## Contextual Search Experimentation

**☐ Phase 2A ☐ Phase 2B**

To evaluate the contextual search system, we now ask you to answer some questions about it and your search in general. Completion of the questionnaire indicates consent to participate in this study.

Please answer all the questions below by ticking/typing the appropriate box or commenting.

## Section 1: OVERALL IMPRESSIONS

Please tick the numbers which most appropriately reflect your overall impressions of the performance of the contextual search with regards to satisfaction, completion of tasks and appeal.

1.1 Search tasks can be completed

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| with difficulty | ☐ | ☐ | ☐ | ☐ | ☐ | easily |

1.2 Overall reactions to the contextual search with regards to completion of tasks:

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| terrible | ☐ | ☐ | ☐ | ☐ | ☐ | wonderful |
| frustrating | ☐ | ☐ | ☐ | ☐ | ☐ | satisfying |
| difficult | ☐ | ☐ | ☐ | ☐ | ☐ | easy |
| rigid | ☐ | ☐ | ☐ | ☐ | ☐ | flexible |

1.3 Overall reactions to the contextual search with regards to the user interface :

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| dull | ☐ | ☐ | ☐ | ☐ | ☐ | stimulating |
| confusing | ☐ | ☐ | ☐ | ☐ | ☐ | clear |
| difficult | ☐ | ☐ | ☐ | ☐ | ☐ | easy |

1.4 Overall confidence with regards to completion of tasks

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Not at all confident | ☐ | ☐ | ☐ | ☐ | ☐ | Very Confident |

## Section 2: ADAPTIVENESS

Please tick the numbers which most appropriately reflect your overall impressions of the learning ability of the contextual search.

2.1 The system remembers your last search actions (*such as visited URLs, entered query, selected terms etc.*)

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| never | ☐ | ☐ | ☐ | ☐ | ☐ | always |

2.2 The system uses your last search actions to recommend terms :

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| never | ☐ | ☐ | ☐ | ☐ | ☐ | always |
| irrelevant | ☐ | ☐ | ☐ | ☐ | ☐ | relevant |
| not useful | ☐ | ☐ | ☐ | ☐ | ☐ | useful |

2.3 The system learnt your search interests over the time

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| never | ☐ | ☐ | ☐ | ☐ | ☐ | always |

**ONLY ANSWER QUESTION 2.4 IF YOU ARE PARTICIPATING IN EXPERIMENTATION PHASE 2B.**

2.4 The system uses **OTHER** users/searchers' search actions to recommend terms :

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| never | ☐ | ☐ | ☐ | ☐ | ☐ | always |
| irrelevant | ☐ | ☐ | ☐ | ☐ | ☐ | relevant |
| not useful | ☐ | ☐ | ☐ | ☐ | ☐ | useful |

## Section 3: RECOMMENDATION

Please tick the numbers which most appropriately reflect your overall impressions of the recommendation ability of the contextual search.

3.1 The system communicated its recommendation action in a way that was :

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| unobtrusive | ☐ | ☐ | ☐ | ☐ | ☐ | obtrusive |
| uninformative | ☐ | ☐ | ☐ | ☐ | ☐ | informative |
| timely | ☐ | ☐ | ☐ | ☐ | ☐ | untimely |

## Section 3: RECOMMENDATION (*cont.*)

3.2 The content of the recommendation terms are

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| confusing | ☐ | ☐ | ☐ | ☐ | ☐ | clear |

3.3 You accepted any recommended words because (mark as **MANY** as apply):

1 ☐ they meant the same
2 ☐ they were related to words you had chosen already
3 ☐ you couldn't find better words
4 ☐ they represented new ideas (*i.e. not part of your original request*)
5 ☐ other (*please specify*) ________________________

3.4 How you find conveying your preferences (*e.g. disambiguated terms*) to the system :

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| difficult | ☐ | ☐ | ☐ | ☐ | ☐ | easy |
| effective | ☐ | ☐ | ☐ | ☐ | ☐ | ineffective |
| not useful | ☐ | ☐ | ☐ | ☐ | ☐ | useful |

3.5 When you are conveying your preferences to the system, you feel :

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| comfortable | ☐ | ☐ | ☐ | ☐ | ☐ | uncomfortable |
| not control | ☐ | ☐ | ☐ | ☐ | ☐ | in control |

## Section 4: QUERY FORMULATION

Please tick the numbers which most appropriately reflect your overall impressions of the query formulation ability of the contextual search.

4.1 The number of steps to formulate query is

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| too many | ☐ | ☐ | ☐ | ☐ | ☐ | just right |

4.2 The system makes formulating the query

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| difficult | ☐ | ☐ | ☐ | ☐ | ☐ | easy |

## Section 4: QUERY FORMULATION (*cont.*)

4.3 The formulated query is effective

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| agree | ☐ | ☐ | ☐ | ☐ | ☐ | disagree |

4.4 I would trust the system to choose the search terms to formulate an effective search query

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| agree | ☐ | ☐ | ☐ | ☐ | ☐ | disagree |

## Section 5: DISPLAY RESULTS AND PRECISION

Please tick the numbers which most appropriately reflect your overall impressions of the display of results and precision of the contextual search.

5.1 The information laid out on the results page was :

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| not useful | ☐ | ☐ | ☐ | ☐ | ☐ | useful |
| frustrating | ☐ | ☐ | ☐ | ☐ | ☐ | satisfying |
| difficult | ☐ | ☐ | ☐ | ☐ | ☐ | easy |
| not effective | ☐ | ☐ | ☐ | ☐ | ☐ | effective |

5.2 The interface guides you to the information you need

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| never | ☐ | ☐ | ☐ | ☐ | ☐ | always |

5.3 You find what you looking for

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| never | ☐ | ☐ | ☐ | ☐ | ☐ | always |

## Section 6: COMMENTS AND SUGGESTIONS

6.1 List out the positive features of this contextual search

______________________________________________

______________________________________________

______________________________________________

______________________________________________

6.2 List out the negative features of this contextual search

______________________________________________

______________________________________________

______________________________________________

______________________________________________

6.3 Others

______________________________________________

______________________________________________

______________________________________________

______________________________________________

**This project has been approved by the Auckland University of Technology Ethics Committee on 17/05/2007, AUTEC Reference number 07/12.**

**Thank you** for participating in this evaluation study.

Appendix A.6.2: Used for Contemporary Search

# POST OBSERVATION

AUCKLAND UNIVERSITY OF TECHNOLOGY
TE WĀNANGA ARONUI O TAMAKI MAKAU RAU

## Contextual Search Experimentation

☐ **Phase IIC**

**ID :** ***phase***

Please answer all the questions below by ticking/typing the appropriate box or commenting.

## Section 1: PERSONAL DETAILS

1.1 Please indicate your AGE RANGE:

☐ 18-25 ☐ 31- 40 ☐ > 51

☐ 26-30 ☐ 41- 50

1.2 Please Indicate your GENDER

☐ Male ☐ Female

1.3 What is your area of WORK?

☐ student ☐ research

☐ academia ☐ other: please state ______________________

1.4 What is the highest EDUCATIONAL QUALIFICATION that you have?

☐ PhD ☐ Bachelors degree

☐ Masters degree ☐ other: please state ___________________

## Section 2: SEARCH EXPERIENCE

2.1 How experienced are you in:

| | novice | | | | expert |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| • searching using World Wide Web search services (*e.g., Alta Vista, Google, Excite, Yahoo, HotBot, and WebCrawler*)? ............... | ☐ | ☐ | ☐ | ☐ | ☐ |

| | novice | | | | expert |
|---|---|---|---|---|---|
| • searching with other search services that are not mentioned above, please specify? | 1 | 2 | 3 | 4 | 5 |
| a. __________________ ............................. | ☐ | ☐ | ☐ | ☐ | ☐ |
| b. __________________ ............................. | ☐ | ☐ | ☐ | ☐ | ☐ |
| c. __________________ ............................. | ☐ | ☐ | ☐ | ☐ | ☐ |

## Section 2: SEARCH EXPERIENCE (*cont.*)

2.2 How often do you use a search service/engine for:

| | Once or twice a year | Once or twice a month | Once or twice a week | Once or twice a day | More often |
|---|---|---|---|---|---|
| ▪ work/academic tasks?......... | ☐ | ☐ | ☐ | ☐ | ☐ |
| ▪ personal tasks?................. | ☐ | ☐ | ☐ | ☐ | ☐ |

2.3 Please indicate your years of Web searching experience:

☐ < 1 year ☐ 3 years ☐ > 4 years

☐ 2 years ☐ 4 years

2.4 When I search the Internet, I can usually find what I'm looking for

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| rarely | ☐ | ☐ | ☐ | ☐ | ☐ | often |

2.5 Finding relevant information on the Internet has taken up more of my time than I expected

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| rarely | ☐ | ☐ | ☐ | ☐ | ☐ | often |

2.6 Please list your top five favourite search services/engine(s):

a. ____________________

b. ____________________

c. ____________________

d. ____________________

e. ____________________

## Section 3: OVERALL IMPRESSIONS

Please tick the numbers which most appropriately reflect your overall impressions of the performance of the Google search with regards to satisfaction and completion of tasks.

3.1 Search tasks can be completed

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| with difficulty | ☐ | ☐ | ☐ | ☐ | ☐ | easily |

## Section 3: OVERALL IMPRESSIONS (*cont.*)

3.2 Overall reactions to the Google search with regards to completion of tasks:

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| terrible | ☐ | ☐ | ☐ | ☐ | ☐ | wonderful |
| frustrating | ☐ | ☐ | ☐ | ☐ | ☐ | satisfying |
| difficult | ☐ | ☐ | ☐ | ☐ | ☐ | easy |
| rigid | ☐ | ☐ | ☐ | ☐ | ☐ | flexible |

3.3 Overall reactions to the Google search with regards to the user interface :

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| dull | ☐ | ☐ | ☐ | ☐ | ☐ | stimulating |
| confusing | ☐ | ☐ | ☐ | ☐ | ☐ | clear |
| difficult | ☐ | ☐ | ☐ | ☐ | ☐ | easy |

3.4 Overall confidence with regards to completion of tasks

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Not at all confident | ☐ | ☐ | ☐ | ☐ | ☐ | Very Confident |

## Section 4: COMMENTS AND SUGGESTIONS

4.1 In your opinion, what makes a search service/engine useful?

____________________

____________________

____________________

____________________

**This project has been approved by the Auckland University of Technology Ethics Committee on 17/05/2007, AUTEC Reference number 07/12.**

**Thank you** for participating in this evaluation study.

## Appendix A.7: Observation Form

# Observation Form

## Observation of Information Seeking Behaviour

☐ **Phase 2A** ☐ **Phase 2B** ☐ **Phase 2C**

**ID :** ***phase 1*** **Date :** / /2007

**Observer's Name :** Dilip Limbu **Location : WT412**

AUCKLAND UNIVERSITY OF TECHNOLOGY
TE WĀNANGA ARONUI O TAMAKI MAKAU RAU

| Class of Observation | Observation | Task 1 | | Task 2 | | Task 3 | | Task 4 | | Task 5 | | Task 6 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Query Syntax** | Use of Boolean operators. | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N |
| | Make use of > 3 search queries. | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N |
| | Query refinement. | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N |
| **User Behaviour** | Formulation of search strategy in advance. | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N |
| | Use of assists/recommended terms options. (*Only for Phase 2A & 2B*) | | | | | | | | | | | | |
| | ▪ Word sense disambiguation | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N |
| | ▪ Meta keyword | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N |
| | ▪ Domain knowledge | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N |
| **Other Characteristics** | Number of queries per task. | | | | | | | | | | | | |
| | Time taken to reach target information. | | | | | | | | | | | | |
| | Number of clicks taken to reach target information. | | | | | | | | | | | | |
| | Number of hits browsed. | | | | | | | | | | | | |
| | Number of URLs visited to reach target information. | | | | | | | | | | | | |
| | Task completed. | Y | N | Y | N | Y | N | Y | N | Y | N | Y | N |

## 8.2 Appendix B: Component Testing

## B.1 Introduction

In this section the summary results of a precision comparison between the three components of the contextual SERL search: Word Sense Disambiguation (WSD), Concept Recommender (CR) and Meta Keywords Recommender (MKR), and Google search are presented as component testing (CT). The objective was to determine the effectiveness of these components search results. During the CT, several Boolean queries were constructed using the three components and executed, and judged (manually) the relevance of the top 10 search results. Similarly, the same queries were submitted to Google search and judged the relevance of the top 10 search results. In the end, using these relevance results, the effectiveness (i.e., precision) was compared between the three components and Google.

### B.1.1 CT Experiment and Results

Table B.1.1 shows the search categories and keywords used during the CT. There were two relevant categories; Computer and Travel, each category containing six search keywords (with combination of one, two and three keywords).

**Table** B.1.1**: classification and search queries.**

| Classification | Search Query |
|---|---|
| **Computer** | Java<br>Java Programming<br>Java Programming Language |
| | Agent<br>Software Agent<br>Java Software Agent |
| **Travel** | Java<br>Java Island<br>Java Island Indonesia |
| | Agent<br>Travel Agent<br>Travel Agent in Java |

Table B.1.2 presents the initial and debugged queries as well as their relevance ranking results from three components of the contextual SERL search and Google search. The first column represents the classification of queries; the second represents no. of keywords, and the third represents the search queries.

The rest of the columns represent the relevance ranking (higher = better) results from three components and Google results. During the test, the three components were debugged as necessary. For example, for search query “java”, the WSD component was debugged once. The initial query ‘java’ was reformulated using the WSD feature to query “java OR java (platform-independent object-oriented)”. However, only 2 relevant results were returned. As such, the WSD query formulation process was debugged by removing the initial “Java OR” keywords from the first reformulated query (i.e., ~~java OR~~ java (platform-independent object-oriented), which returned 6 relevant results. Similarly, all components were debugged in the same ways until the formulation Boolean queries are effective.

**Table B.1.2: Comparison of relevance ranking.**

| Classification | No. of Keywords | Search Query | Relevance Ranking (for *10 hits only*) | | | |
|---|---|---|---|---|---|---|
| | | | Google | WSD | CR | MK |
| **Computer** | One | Java | 5 | 2<br>*java OR java (platform-independent object-oriented)* | 2<br>*java OR Programming Languages* | - |
| | | | | 6<br>*~~java OR~~ java (platform-independent object-oriented)* | 8<br>*java Programming ~~OR~~ Languages* | 9<br>*java programming language* |
| | | Agent | 4 | - | 4<br>*agent OR software* | - |
| | | | | - | 8<br>*agent ~~OR~~ software* | 9<br>*intelligent software agents* |
| | Two | Java Programming | 10 | 6<br>*Java Programming OR Java (platform-independent object-oriented) AND Programming (computer programming)* | 8<br>*Java Programming OR Programming Languages* | - |
| | | | | 9<br>*~~Java Programming OR~~ Java (platform-independent object-oriented) AND Programming (computer programming)* | 8<br>*Java Programming ~~OR~~ Programming Languages* | - |
| | | | | | 10<br>*~~Java Programming~~ java programming languages* | 10<br>*java essentials java programming* |
| | | Software Agent | 8 | 3<br>*Software Agent OR Software (software system)"* | 4<br>*Software Agent OR Software* | - |
| | | | | 2<br>*~~Software Agent OR~~ Software (software system)* | 9<br>*Software Agent ~~OR~~ Software* | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | 8<br>*Software (software system) AND Agent* | 8<br>*software agent* | 10<br>*software agents overview* |
| | Three | Java Programming Language | 7 | 8<br>*Java Programming Language OR Java (platform-independent object-oriented) AND Programming (computer programming)* | 7<br>*Java Programming Language OR Programming Languages* | - |
| | | | | 8<br>*~~Programming Language OR~~ Java (platform-independent object-oriented) AND Programming (computer programming)* | 8<br>*Java Programming Language ~~OR~~ Programming Languages* | - |
| | | | | 9<br>*Java (platform-independent object-oriented) AND Programming (computer programming) AND Language* | 8<br>*java programming language languages* | - |
| | | | | | 10<br>*java programming language* | 10<br>Java Programming Language |
| | | Java Software Agent | 6 | 2<br>*Java Software Agent OR Java (platform-independent object-oriented)* | 1<br>*Java Software Agent OR Software* | - |
| | | | | 0<br>*~~Java Software Agent OR~~ Java (platform-independent object-oriented) AND Software (software system)* | 9<br>*Java Software Agent ~~OR~~ Software* | - |
| | | | | 8<br>*Java (platform-independent object-oriented) AND Software (software system) AND Agent* | 6<br>*java software agent* | 8<br>*intelligent software agents research david* |
| **Travel** | One | java | 1 | 10<br>*java OR java (island indonesia)* | 1<br>*java OR Travel* | - |
| | | | | 10<br>*~~java OR~~ java (island indonesia)* | 10<br>*java ~~OR~~ Travel* | 9<br>*visit indonesia information indonesia* |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | *OR java* |
| | | Agent | 1 | - | 3<br>*Agent OR Travel* | - |
| | | | | - | 10<br>*Agent ~~OR~~ Travel* | |
| | Two | Java Island | 6 | 8<br>*Java Island OR Java (island indonesia)* | 7<br>*Java Island OR Sightseeing* | - |
| | | | | 10<br>*~~Java Island OR~~ Java (island indonesia)* | 10<br>*Java Island ~~OR~~ Sightseeing* | 8<br>*java island pulau jawa* |
| | | Travel Agent | 9 | 9<br>*Travel Agent OR Travel (traveling OR travelling) AND Agent (factor OR broker)"* | 9<br>*Travel Agent OR Sightseeing* | - |
| | | | | *7*<br>*~~Travel Agent OR~~ Travel (traveling OR travelling) AND Agent (factor OR broker)* | 9<br>*Travel Agent ~~OR~~ Sightseeing* | 9<br>*travel agents* |
| | Three | Java Island Indonesia | 7 | 6<br>*Java Island Indonesia OR Java (island indonesia) AND Island (land mass)* | 8<br>*Java Island Indonesia OR Travel* | - |
| | | | | 7<br>*~~Java Island Indonesia OR~~ Java (island indonesia) AND Island (land mass)* | 9<br>*Java Island Indonesia ~~OR~~ Travel* | |
| | | | | 7<br>*Java (island indonesia) AND Island (land mass) AND Indonesia* | - | 7<br>*java island pulau jawa* |
| | | Travel Agent in Java | 2 | 1<br>*Travel Agent in Java OR Travel (locomotion)"* | *1*<br>*Travel Agent in Java OR Travel"* | *3*<br>*yogyakarta tailor tour operator; travel* |

**Table B.1.2** shows the summary results of the precision comparison between three components of the contextual SERL search and Google. In general, the results indicated that the contextual SERL search's CR component has the highest precision ranking followed by MKR component.

**Table B.1.2: Precision results.**

| Classification | No. of Keywords | Search Query | *Precision (First 10 hits)* | | | |
|---|---|---|---|---|---|---|
| | | | Google | WSD | CR | *MKR* |
| Computer | 1 | Java | 0.5 | 0.6 | 0.8 | *0.9* |
| | | Agent | 0.5 | 0.0 | 0.8 | *0.9* |
| | 2 | Java Programming | 1.0 | 0.9 | 1.0 | *1.0* |
| | | Software Agent | 0.8 | 0.8 | 0.8 | *1.0* |
| | 3 | Java Programming Language | 0.7 | 0.9 | 1.0 | *1.0* |
| | | Java Software Agent | 0.6 | 0.8 | 0.9 | *0.8* |
| Travel | 1 | Java | 0.1 | 1.0 | 1.0 | *0.9* |
| | | Agent | 0.1 | 0.0 | 1.0 | *0.0* |
| | 2 | Java Island | 0.6 | 1.0 | 1.0 | *0.8* |
| | | Travel Agent | 0.9 | 0.9 | 0.9 | *0.9* |
| | 3 | Java Island Indonesia | 0.7 | 0.7 | 0.9 | *0.7* |
| | | *Travel Agent in Java* | *0.2* | *0.1* | *0.1* | *0.3* |

## 8.3 Appendix C: Anonymous Questionnaires Analysis

## C.1 Introduction

In this section the summary results of anonymous questionnaires survey are presented. The objective of this survey was to better understand quantitatively how general users utilise Web searches to seek information as part of their daily life.

Of just over 200 survey questionnaires sent, forty two (42) valid responses (i.e. 21%) were received. Respondents were asked to complete a mix of five point Likert scales and semantic differentials (i.e., range 1-5, higher = better) to indicate their information seeking behaviour. All responses scores (i.e., 1 - 5 scores) were transformed to ‘Yes’ (>=4) or ‘No’ (<=3) value in order to retain consistency. The respondents were students, staff and faculty members of Auckland University of Technology (AUT), New Zealand.

### C.1.1 Respondent Characteristics and Search Experience

Table C.1.1 shows the summary of the characteristics of respondents. Of the forty two (42) respondents, 71% were male (30) and 29% were female (12). Less than one quarter (19%) of the respondents (8) were less than 30 years old, while more than three-quarters (81%) of the respondents (34) were 31 years or older. More than half (55%) of the respondents (23) were academic staff. The remaining 45% of respondents (19) were a mixture of students (14), researchers (3) and others (2). About 17% of the respondents (7) hold PhD degree, 45% respondents (19) hold Masters degrees, 26% respondents (11) hold a bachelor degree and 12% (5) respondents hold other degrees.

Table C.1.1: Characteristics of respondents.

| Gender | No. responses | Percentage |
|---|---|---|
| Male | 30 | 71.4% |
| Female | 12 | 28.6% |
| **Age range** | | |
| 18-25 | 6 | 14.3% |
| 26-30 | 2 | 4.8% |
| 31-40 | 14 | 33.3% |
| 41-50 | 10 | 23.8% |
| 51-over | 10 | 23.8% |
| **Area of work** | | |
| Student | 14 | 33.3% |
| Academic | 23 | 54.8% |
| Research | 3 | 7.1% |
| Other | 2 | 4.8% |
| **Educational qualification** | | |
| PhD degree | 7 | 16.7% |
| Master's degree | 19 | 45.2% |
| Bachelor's degree | 11 | 26.1% |
| Other degrees | 5 | 12% |

Table C.1.2 shows the summary of Web search experience of the respondents. More than half (67%) of the respondents (28) indicated that they are expert Web searchers, whilst approximately one third (33%) of respondents (14) indicated that they are novice Web searchers. More than 95% respondents (40) indicated that they have more than four years of Web search experience. The remaining 5% of respondents (2) indicated that they have no more than three years of Web search experience. Half (50%) of the respondents (21) indicated they regularly use Web search service for their office or academic tasks. The remaining 33% (14) and 17% (7) of respondents use Web search service for their office or academic tasks at least once or twice a day and more than once or twice a month, respectively. Likewise, out of 40 respondents, one quarter (25%) of the respondents (10) indicated they regularly use Web searches for their personal tasks. The remaining 32.5% (13) and 42.5% (17) of respondents use Web search service for their personal tasks at least once or twice a day and more than once or twice a month, respectively. More that three quarters (81%) of the respondents (34) indicated they always find the relevant information for which they are looking for. More than one quarter (32%) of the respondents (13) indicated Internet has taken up more time than they expected.

Table C.1.2: Search experience, usage and satisfaction.

| Search expertise | No. responses | Percentage |
|---|---|---|
| Novice | 14 | 33.3% |
| Expert | 28 | 66.7 % |
| **Years of search experience** | | |
| <= 3 years | 2 | 4.8% |
| >= 4 years | 40 | 95.2% |
| **Frequency of Web search for work/academic tasks** | | |
| Once/twice a week/month | 7 | 16.7% |
| Once/twice a day | 14 | 33.3% |
| More often | 21 | 50% |
| **Frequency of Web search for personal tasks (*n=40*)** | | |
| Once/twice a week/month | 17 | 42.5% |
| Once/twice a day | 13 | 32.5% |
| More often | 10 | 25% |
| **Can find what they are looking for** | | |
| Yes | 34 | 81% |
| No | 8 | 19% |
| **Time taken** | | |
| Yes | 13 | 31.7% |
| No | 29 | 68.3% |

A question was asked to respondents to list their all-time top favourite search engines. Table C.1.3 showed that the all-time favourite search engine is Google (42 respondents) and Yahoo (17 respondents) came in second.

Table C.1.3: All-time favourite search engine

| Favourite Search Engines | |
|---|---|
| Google | 42 |
| Yahoo | 17 |
| Alta Vista | 9 |
| Dogpile | 3 |
| MSN | 2 |
| Baidu | 2 |

The following section describes the subjects' information seeking behaviour and their search preferences in details.

## C.1.2 Search Behaviours

This section presents the results of a study of the information seeking behaviour of forty two (42) respondents. Respondents were asked to indicate whether the first resource they use for finding information is the Internet. More than three quarters (79%) of the respondents (33) indicated 'Yes' and less than one quarter (21%) of the respondents (9) indicated 'No' for the first resource they use for finding information is the Internet (Figure C.2.1).

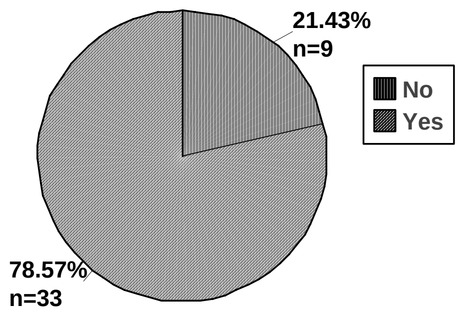


**Figure C.2.1: Internet as first resource for information finding.**

Respondents were asked to indicate on semantic differentials which informal resources, such as the URLs from friends, family and colleagues; the previously bookmarked results; the magazines, books etc; and the search history, they used for finding relevant information on the Internet. It is clear from the analysis (Table C.2.1) that more than one quarter (29%) of the respondents (12) use the URLs from friends, family and colleagues; more than half (55%) of the respondents (23) use the previously bookmarked results; more than one quarter (36%) of the respondents (15) use the magazines, books etc; and 21% of the respondents (9) use the search history browser; indicated 'Yes' as the informal resources for finding relevant information on the Internet.

**Table C.2.1: Informal resources for information finding.**

| Informal Resources | Responses (Yes) | Percentage |
|---|---|---|
| URLs from friends, family, and colleagues | 12 | 28.6% |
| Bookmarked results | 23 | 54.8% |
| Magazines, books etc. | 15 | 35.7% |
| Search history | 9 | 21.4% |

Besides, the respondents indicated that they use numerous other informal sources for finding relevant information on the Internet (Table C.2.2). Among the 42 responses, the popular choices of other informal resources being used to acquiring information on the Internet are search engine and AUT online database (Table C.2.2).

**Table C.2.2:**

| **Other Informal Resources** | |
|---|---|
| search engine | 5 |
| online database | 3 |
| guess the URL | 2 |
| keywords | 2 |
| links on website | 1 |
| software package | 1 |
| URL easy to remember | 1 |

Respondents were asked to indicate whether the respondents make any search strategies in advance prior to searching the Internet. Nearly three quarters (73%) of the respondents (30) indicated 'No' and more than one quarter (27%) of the respondents (9) indicated 'Yes' for the use of search strategies in advance prior searching the Internet (Figure C.2.2).

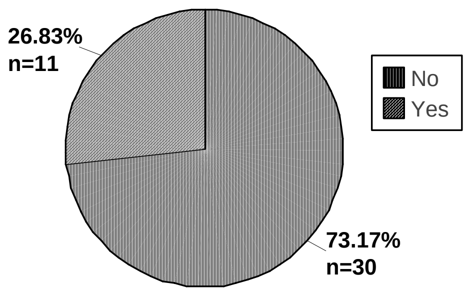


**Figure C.2.2: Search strategy.**

Respondents were asked to indicate whether the respondents make use of more than one query in resolving information finding problem. More than three quarters (78%) of the respondents (32) indicated 'Yes' and less than one quarter (22%) of the respondents (9) indicated 'No' for the use of more than one query in resolving information finding problem (Figure C.2.3).

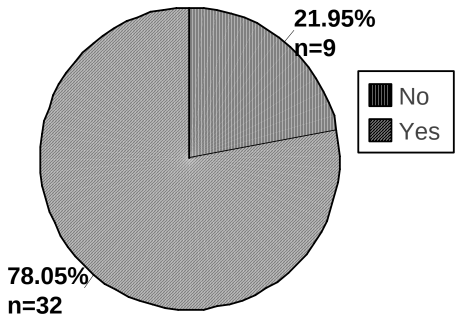


**Figure C.2.3: Use more than one query.**

Respondents were asked to indicate on semantic differentials the type of search features, such as the Boolean query, the advanced search features; and the directory or classification, they use in acquiring relevant information on the Internet. It is cleared from the analysis (Table C.2.3) that more than one quarter (33.3%) of the respondents (14) use the Boolean query; half (50%) of the respondents (21) use the advance search features; less than one quarter (18.4%) of the respondents (7) use the directory or classification as the search features in acquiring relevant information on the Internet.

**Table C.2.3: Search features.**

| | No. responses (Yes) | Percentage |
|---|---|---|
| **Search features** | | |
| Boolean query ***(n = 42)*** | 14 | 33.3% |
| Advance search features ***(n = 42)*** | 21 | 50% |
| Directory or classification ***(n = 38)*** | 7 | 18.4% |

Besides, the respondents indicated that they use numerous other search features in acquiring relevant information on the Internet (Table C.2.4). Three respondents suggested that the keywords, figures, images, skim reading of content and date are the other criteria to be used in judging relevance on returned search results prior to browsing (Table C.2.4).

**Table C.2.4:**

| **Other Criteria to Judge relevant** | |
|---|---|
| Keywords, figure & image | 1 |
| Skim read content | 1 |
| Date | 1 |

Respondents were asked to indicate how many hits they generally browse through to acquire relevant information from the returned results. More than three quarters (83.3%) of the respondents (34) indicated the first 20 hits and less than one quarter (17%) of the respondents (7) indicated the first 10 hits that they generally browse through to acquire relevant information on the Internet (Figure C.2.4).

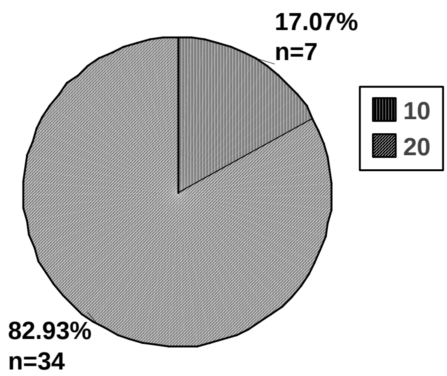


**Figure C.2.4: Number of hits.**

Respondents were asked to indicate on semantic differentials the criteria they use in judging relevance, such as the title; the descriptions; the highlighted words; the URL itself and the search engine's rankings on the returned search results prior browsing. It is clear from the analysis (Table C.2.5) that more than three quarters (80%) of the respondents (32) use the title; 78% of the respondents (32) use the descriptions; 65.8% of the respondents (27) use the highlighted words; 43.6% of the respondents (17) use the URL itself and 42.5% of the respondents (17) use the search engine's rankings, as the criteria in judging relevance on the returned search results prior to browsing.

**Table C.2.5: Criteria used in judging relevance.**

| | No. responses | Percentage |
|---|---|---|
| **Judging relevance** | | |
| Title (n=40) | 32 | 80% |
| Description (n=41) | 32 | 78% |
| Highlighted words (n=41) | 27 | 65.8% |
| URLs (n=39) | 17 | 43.6% |
| Ranking (n=40) | 17 | 42.5% |

Respondents were asked to indicate whether they select those hits that they can use immediately. More than half (63%) of the respondents (24) indicated 'Yes' and less than one quarter (37%) of the respondents (14) indicated 'No' for this question (Figure C.2.5).

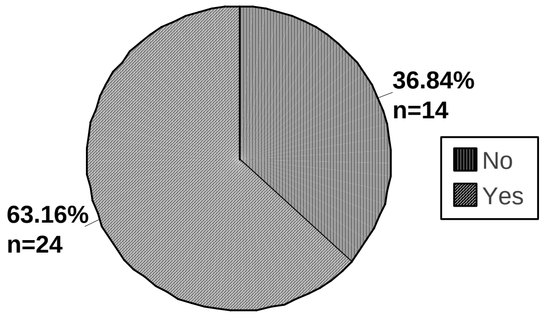


**Figure C.2.5: Immediate use on search results.**

The next section presents the search preferences of forty two (42) respondents.

## C.1.3 Search Preferences

A question was asked to find out how respondents view their usage of search engines; as a single user, as a member of community/group. More than three quarters (81%) of the respondents indicated that they view their search engines usage as the single user; 7% of the respondents (3) view their search engines usage as the member of a group; 12% of the respondents (5) view their search engines usage as both (i.e., the single user and the member of a group) (Figure C.3.1).

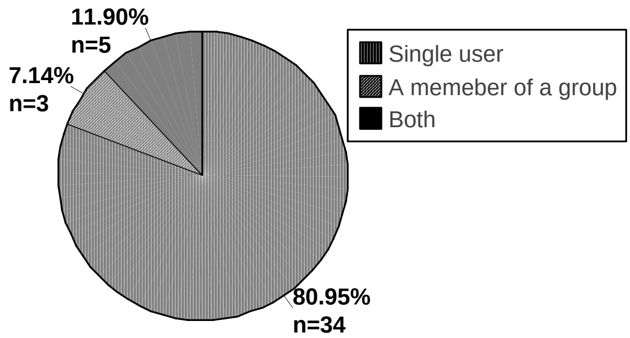


**Figure C.3.1: View of search engines usage.**

A question was asked to find out whether respondents would be willing to register to a search engine that adapts and stores their searching behaviour. Out of 41 respondents, less than half (44%) of the respondents (18) indicated they would be willing to register to a search engine that adapts and stores their searching behaviour. However, more than half (56%) of the respondents (23) indicated they would not be willing to register to a search engine that adapts and stores their searching behaviour (Figure C.3.2).

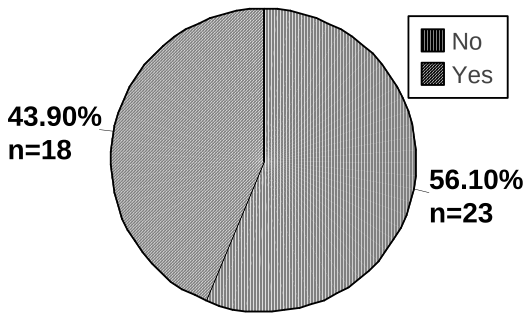


**Figure C.3.2: Willingness to register to a search engine.**

A question was asked to find out whether respondents would like to share their searching behaviour information to form a shared knowledge base. 100% of the respondents (41) indicated that they like to share their searching behaviour information to form a shared knowledge base. Among them (Figure C.3.3), more than half (51%) of the respondents (21) indicated that they like to share their searching behaviour information anonymously. The remaining 49% of the respondents (20) indicated that they like to share their searching behaviour information known.

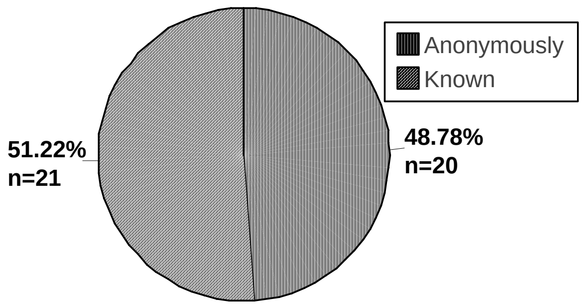


**Figure C.3.3: Willingness to share search behaviour.**

A question was asked to find out whether respondents like a search service/engine to recommend related terms. 100% of the respondents (40) indicated they like a search service/engine to recommend related terms. Among them (Figure C.3.4), 22.5% of the respondents (9) indicated using their own profiles; 32.5% of the respondents (13) indicated using shared profile and about 45% of the respondents (18) indicated using both (i.e., personal profile and shared profile).

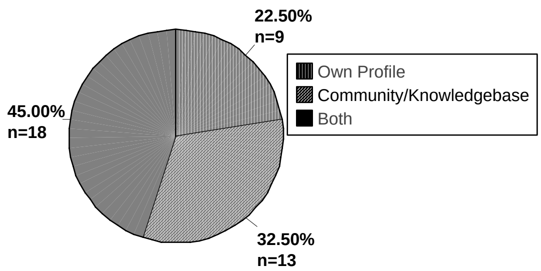


**Figure C.3.4: Recommendation feature.**

A question was asked to find out whether respondents would allow search service/engines to recommend them as a user with similar interests to other searchers. Out of 42 respondents, 50% of the respondents (21) indicated they would allow search service/engines to recommend them as a user with similar interests to other searchers (Table C.3.1).

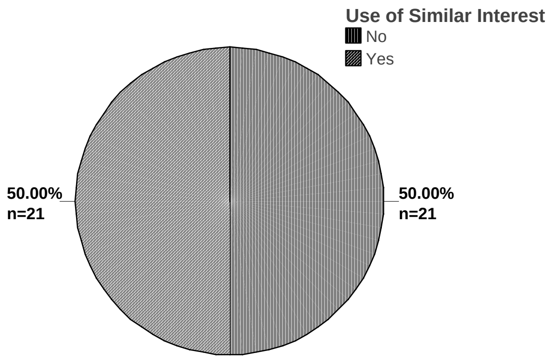


**Figure C.3.5: Recommendations from other similar interest searchers.**

A question was asked to respondents to find out what makes a search service/engine useful. Table C.3.1. presents some of the features that make search engines useful.

**Table C.3.1:**

| Usefulness of Search Engine |
| --- |
| Relevant result, usefulness of retuned sites. information / websites |
| Speed and Time |
| Accuracy, closeness of matching |
| Advanced search , Wide range data sources |
| Function , User Interface |
| The option of sharing your web history to others |
| Its integrity in use of your information. |
| To remember my past relevant pages. |

## 8.4 Appendix D: Normality Test.

This appendix summarises graphical and numerical normality tests that were carried out to determine whether or not treatment data (only for those data with *interval* in nature) were *normally* distributed. The Q-Q and detrended Q-Q plot charts and Shapiro-Wilk normality tests (since the treatment data $n < 50$) were used/performed as a graphical and numerical normality tests respectively.

### D.1 Number of Queries (Hypothesis 1.1)

In the Q-Q plot chart (Figure D.1.1) of total number of queries (or treatment data), the expected normal distribution is the straight (or fitted) line and the line of little circles is the observed values. The Q-Q plot chart indicated that observed values significantly deviate from the fitted line. The detrended Q-Q plot also illustrated a *systematic pattern* of deviation indicating *non-normality* of the observed values.

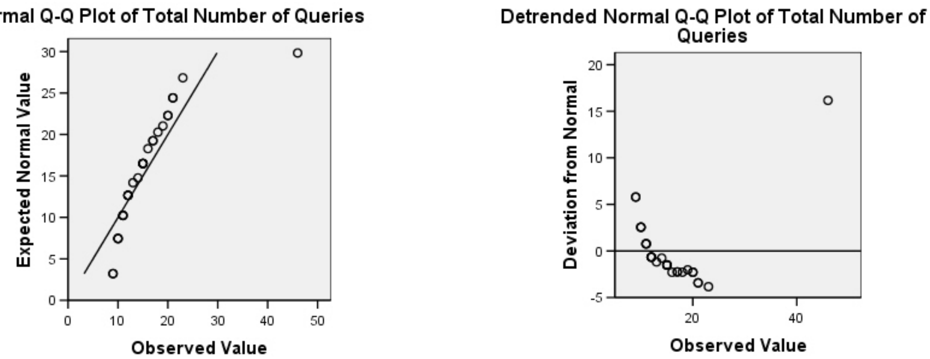


**Figure D.1.1: Q-Q and Detrended Q-Q plot charts of number of queries.**

The results of Shapiro-Wilk normality test (Table D.1.1) also indicated that the treatment data were *not normally* distributed, where phase II (a) ($p$= *.535*), phase II (b) ($p$ = *.061*), and phase II (c) ($p$ = ***.005***).

**Table D.1.1: Total number of queries normality test.**

| | | Kolmogorov-Smirnov [a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|---|
| | Experimental Phase | Statistic | df | Sig. | Statistic | df | Sig. |
| Total Number of Queries | Phase II (a) | .187 | 10 | .200* | .938 | 10 | ***.535*** |
| | Phase II (b) | .198 | 10 | .200* | .852 | 10 | ***.061*** |
| | Phase II (c) | .256 | 10 | .062 | .761 | 10 | ***.005*** |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Based on a consistent result from both graphical and numerical tests, the treatment data concluded as *not* normally distributed.

## D.2 Number of Clicks (Hypothesis 1.2)

The Q-Q plot chart (Figure D.1.2) of total number clicks (or treatment data) indicated *no* significant deviation from the fitted line. The detrended Q-Q plot also illustrated a disorganised pattern of deviation indicating *normality* of the observed values.

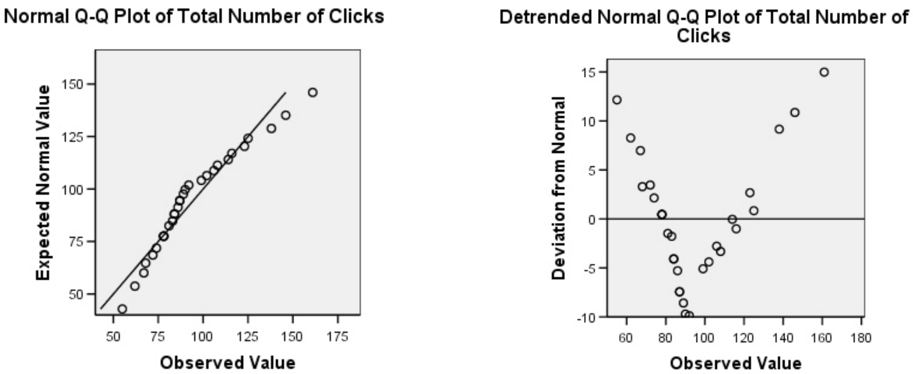


**Figure D.1.2: Q-Q and Detrended Q-Q plot charts of total number of clicks.**

The results of Shapiro-Wilk normality test (Table D.1.2) indicated that the treatment data were *normally* distributed, where phase II (a) ($p = .648$), phase II (b) ($p = .896$), and phase II (c) ($p = .228$).

**Table D.1.2: Total number of clicks normality test.**

| | | Kolmogorov-Smirnov [a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|---|
| | Experimental Phase | Statistic | df | Sig. | Statistic | df | Sig. |
| Total Number of Clicks | Phase II (a) | .149 | 10 | .200* | .948 | 10 | ***.648*** |
| | Phase II (b) | .177 | 10 | .200* | .971 | 10 | ***.896*** |
| | Phase II (c) | .249 | 10 | .080 | .902 | 10 | ***.228*** |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Based on a consistent result from both graphical and numerical tests, the treatment data for total number of queries concluded as *normally* distributed.

## D.3 Number of Hits (Hypothesis 1.3)

The Q-Q plot chart (Figure D.1.3) of total number hits (or treatment data) indicated *no* significant deviation from the fitted line. The detrended Q-Q plot also illustrated a disorganised pattern of deviation indicating *normality* of the observed values.

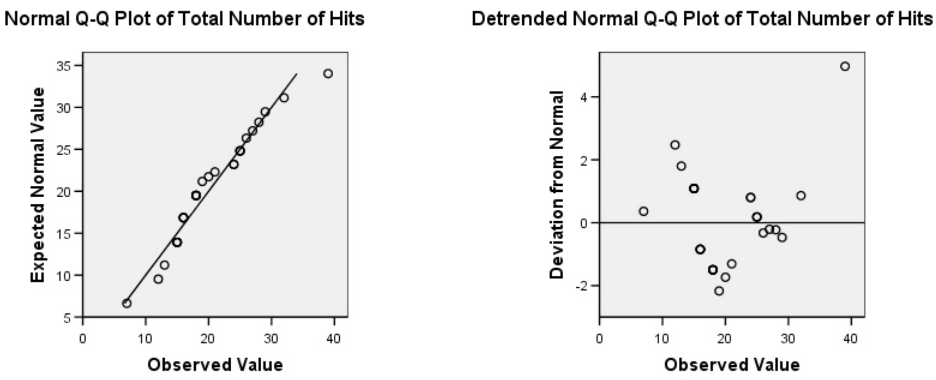


**Figure D.1.3: Q-Q and Detrended Q-Q plot charts of total number of hits.**

The results of Shapiro-Wilk normality test (Table D.1.3) indicated that the treatment data were normally distributed, where phase II (a) ($p = .175$), phase II (b) ($p = .697$), and phase II (c) ($p = .665$).

**Table D.1.3: Total number of hits normality test.**

| | | Kolmogorov-Smirnov [a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|---|
| | Experimental Phase | Statistic | df | Sig. | Statistic | df | Sig. |
| Total Number of Hits | Phase II (a) | .224 | 10 | .169 | .891 | 10 | ***.175*** |
| | Phase II (b) | .179 | 10 | .200* | .952 | 10 | ***.697*** |
| | Phase II (c) | .172 | 10 | .200* | .950 | 10 | ***.665*** |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Based on a consistent result from both graphical and numerical tests, the treatment data concluded as *normally* distributed.

## D.4 Number of URLs (Hypothesis 1.4)

The Q-Q plot chart (Figure D.1.4) of total number URLs (or treatment data) indicated *no* significant deviation from the fitted line. The detrended Q-Q plot also illustrated a disorganised pattern of deviation indicating *normality* of the observed values.

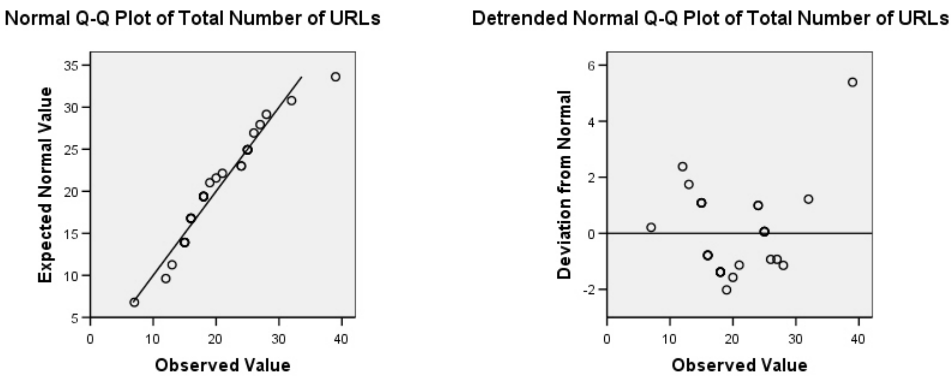


**Figure D.1.4: Q-Q and Detrended Q-Q plot charts of total number of urls.**

The results of Shapiro-Wilk normality test (Table D.1.4) indicated that the treatment data were *normally* distributed, where phase II (a) ($p$ = *.108*), phase II (b) ($p$ = *.697*), and phase II (c) ($p$ = *.665*).

**Table D.1.4: Total number of urls normality test.**

| | | Kolmogorov-Smirnov [a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|---|
| | Experimental Phase | Statistic | df | Sig. | Statistic | df | Sig. |
| Total Number of URLs | Phase II (a) | .222 | 10 | .179 | .873 | 10 | ***.108*** |
| | Phase II (b) | .179 | 10 | .200* | .952 | 10 | ***.697*** |
| | Phase II (c) | .172 | 10 | .200* | .950 | 10 | ***.665*** |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Based on a consistent result from both graphical and numerical tests, the treatment data concluded as *normally* distributed.

## D.5 Length of time (Hypothesis 1.5)

The Q-Q plot chart (Figure D.1.5) of total time taken (or treatment data) indicated *no* significant deviation from the fitted line. The detrended Q-Q plot also illustrated a disorganised pattern of deviation indicating *normality* of the observed values.

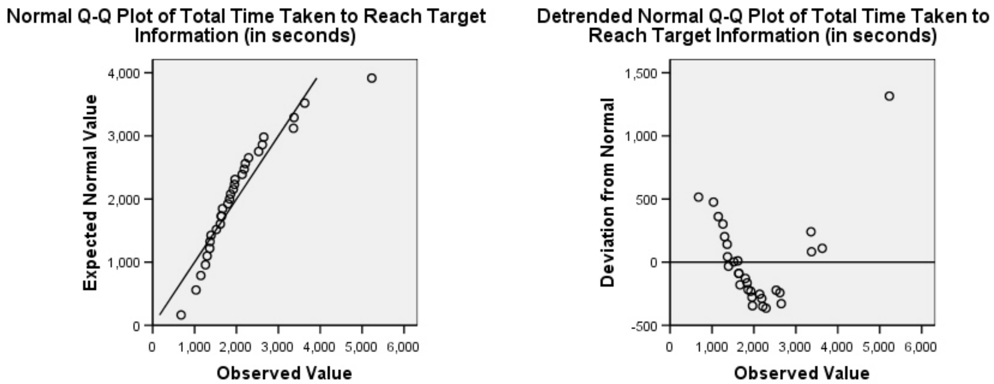


**Figure D.1.5: Q-Q and Detrended Q-Q plot charts of total time take.**

The results of Shapiro-Wilk normality test (Table D.1.5) indicated that the treatment data were *normally* distributed, where phase II (a) ($p$ = *.155*), phase II (b) ($p$ = *.501*), and phase II (c) ($p$ = *.114*).

**Table D.1.5: Total length of time normality test.**

| | | Kolmogorov-Smirnov [a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|---|
| | Experimental Phase | Statistic | df | Sig. | Statistic | df | Sig. |
| Total Time Taken to Reach Target Information (in seconds) | Phase II (a) | .210 | 10 | .200* | .887 | 10 | ***.155*** |
| | Phase II (b) | .202 | 10 | .200* | .935 | 10 | ***.501*** |
| | Phase II (c) | .226 | 10 | .160 | .875 | 10 | ***.114*** |

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Based on a consistent result from both graphical and numerical tests, the treatment data concluded as *normally* distributed.

## 8.5 Appendix E: Descriptive Statistics and Inferential Analysis

This appendix summarises descriptive statistics and inferential analysis to describe the basic features of the data in the three experimental systems; phase II (a), phase II (b), and phase II (c) and to determine if statistically significant differences exist among them.

### E.1 Number of Queries (Hypothesis 1.1)

The descriptive statistic results of the *total* queries entered to complete *all* search tasks in each experimental phase are shown in Table E.1.1. It shows that the phase II (b) and phase II (c) have the lowest *median* ($\tilde{X}$ = *12.50*) and highest *median* ($\tilde{X}$ = *16.50*), respectively.

**Table E.1.1**

| | | | Median | Minimum | Maximum | Std Deviation |
|---|---|---|---|---|---|---|
| Experimental Phase | Phase II (a) | Total Number of Queries | 13.00 | 9.00 | 21.00 | 3.50 |
| | Phase II (b) | Total Number of Queries | 12.50 | 10.00 | 20.00 | 4.03 |
| | Phase II (c) | Total Number of Queries | 16.50 | 9.00 | 46.00 | 10.32 |

A *non-parametric* Kruskal-Wallis test was performed on the *total* number of queries entered to complete all search tasks to determine whether or not any significance differences exist between the three experimental systems. The results ($X^2$= *3.056* and *p* = *0.217*) (Table E.1.2) showed *no* statistically significant differences in the total number of queries between the three experimental systems.

**Table E.1.2**

| | Experimental Phase | N | Mean Rank |
|---|---|---|---|
| Total Number of Queries | Phase II (a) | 10 | 13.35 |
| | Phase II (b) | 10 | 13.70 |
| | Phase II (c) | 10 | 19.45 |
| | Total | 30 | |

**Test Statistics** [a,b]

| | Total Number of Queries |
|---|---|
| Chi-Square | 3.056 |
| df | 2 |
| Asymp. Sig. | .217 |

a. Kruskal Wallis Test

b. Grouping Variable: Experimental Phase

A Kruskal-Wallis test was performed on the total number of queries entered to complete an *individual* search task to determine whether or not any significant differences exist between the three experimental systems. The results showed that except for the search task six ($X^2$ = *7.009* and *p = 0.030*) (Table E.1.3), there is *no* statistically significant differences in the total number of queries between the three experimental systems.

**Table E.1.3**

**Test Statistics [a,b]**

| | Task 1 | Task 2 | Task 3 | Task 4 | Task 5 | Task 6 |
|---|---|---|---|---|---|---|
| Chi-Square | .807 | 1.292 | 3.147 | 1.176 | .143 | 7.009 |
| df | 2 | 2 | 2 | 2 | 2 | 2 |
| Asymp. Sig. | .668 | .524 | .207 | .555 | .931 | .030 |

a. Kruskal Wallis Test

b. Grouping Variable: Experimental Phase

Further, *non-parametric* Mann-Whitney U tests were performed on the search task six to determine whether or not any significance differences exist between any two experimental systems. The results showed that except for the phase II (b) and phase II (c) experimental phase (*U = 17.00* and *p = 0.011*) (Table E.1.4), there is *no* statistically significant differences in the total number of queries between other experimental systems (Table E.1.5).

**Table E.1.4**

| | Experimental Phase | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| Task 6 | Phase II (b) | 10 | 7.20 | 72.00 |
| | Phase II (c) | 10 | 13.80 | 138.00 |
| | Total | 20 | | |

**Test Statistics [b]**

| | Task 6 |
|---|---|
| Mann-Whitney U | 17.000 |
| Wilcoxon W | 72.000 |
| Z | -2.586 |
| Asymp. Sig. (2-tailed) | .010 |
| Exact Sig. [2*(1-tailed Sig.)] | .011[a] |

a. Not corrected for ties.

b. Grouping Variable: Experimental Phase

**Table E.1.5**

| | Experimental Phase | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| Task 6 | Phase II (a) | 10 | 8.25 | 82.50 |
| | Phase II (c) | 10 | 12.75 | 127.50 |
| | Total | 20 | | |

**Test Statistics[b]**

| | Task 6 |
|---|---|
| Mann-Whitney U | 27.500 |
| Wilcoxon W | 82.500 |
| Z | -1.743 |
| Asymp. Sig. (2-tailed) | .081 |
| Exact Sig. [2*(1-tailed Sig.)] | .089[a] |

a. Not corrected for ties.

b. Grouping Variable: Experimental Phase

To summarise, the *non-parametric* Kruskal-Wallis test for the total number of queries entered to complete *all* search tasks established *no* statistically significant differences between the three experimental systems. However, detailed *non-parametric* (Kruskal-Wallis test and Mann-Whitney test) tests for the total number of queries entered to complete an *individual* search task established statistically *highly* significant differences between the phase II (b) and phase II (c) experimental phase for the search task six.

## E.2 Number of Clicks (Hypothesis 1.2)

The descriptive statistic results of the total number of clicks clicked to complete *all* search tasks in each experimental phase are shown in Table E.2.1. It shows that the phase II (b) ($\overline{X}$ = *86.10*) and phase II (c) ($\overline{X}$ = *101.20*) have the lowest mean (i.e., fewer clicks) and highest mean (i.e., more clicks), respectively.

**Table E.2.1**

Total Number of Clicks

| | | | | | 95% Confidence Interval for Mean | | | |
|---|---|---|---|---|---|---|---|---|
| | N | Mean | Std. Deviation | Std. Error | Lower Bound | Upper Bound | Minimum | Maximum |
| Phase II (a) | 10 | 96.0000 | 23.49468 | 7.42967 | 79.1929 | 112.8071 | 62.00 | 146.00 |
| Phase II (b) | 10 | 86.1000 | 18.82935 | 5.95436 | 72.6303 | 99.5697 | 55.00 | 123.00 |
| Phase II (c) | 10 | 101.2000 | 31.97499 | 10.11138 | 78.3265 | 124.0735 | 67.00 | 161.00 |
| Total | 30 | 94.4333 | 25.28279 | 4.61599 | 84.9926 | 103.8741 | 55.00 | 161.00 |

A *parametric* one-way analysis of variance (ANOVA) and Dunnett's one-tailed tests were performed on the *total* number of clicks clicked to complete all search tasks to determine whether or not any significance differences exist between the three experimental systems. The results (Anova *p = 0.413,* Dunnett's one-tailed *p = 0.474* for phase II (a) and II (c), and *p = 0.162* for phase II (b) and II (c)) (Table E.2.2 and Table E.2.3) showed *no* statistically significant differences between the three experimental systems.

**Table E.2.2**

Total Number of Clicks

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 1176.867 | 2 | 588.433 | .915 | .413 |
| Within Groups | 17360.500 | 27 | 642.981 | | |
| Total | 18537.367 | 29 | | | |

**Table E.2.3**

Dependent Variable: Total Number of Clicks

Dunnett t (<control) [a]

| (I) Experimental Phase | (J) Experimental Phase | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | 95% Confidence Interval Upper Bound |
|---|---|---|---|---|---|---|
| Phase II (a) | Phase II (c) | -5.200 | 11.340 | .474 | | 17.45 |
| Phase II (b) | Phase II (c) | -15.100 | 11.340 | .162 | | 7.55 |

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Further, one-way ANOVA and Dunnett's one-tailed tests were performed to test whether the total number of clicks clicked to complete an individual search task has any statistically significant differences between the three experimental systems. The results (Table E.2.4 and Table E.2.5) showed that except for the search task six (Anova *p* = 0.041, Dunnett's one-tailed *p* = 0.011 for phase II (b) and II (c)), there is no statistically significant differences between the three experimental systems.

**Table E.2.4**

| | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Task 1 | Between Groups | 50.867 | 2 | 25.433 | .295 | .747 |
| | Within Groups | 2327.000 | 27 | 86.185 | | |
| | Total | 2377.867 | 29 | | | |
| Task 2 | Between Groups | 176.467 | 2 | 88.233 | .830 | .447 |
| | Within Groups | 2869.700 | 27 | 106.285 | | |
| | Total | 3046.167 | 29 | | | |
| Task 3 | Between Groups | 152.867 | 2 | 76.433 | 1.629 | .215 |
| | Within Groups | 1266.500 | 27 | 46.907 | | |
| | Total | 1419.367 | 29 | | | |
| Task 4 | Between Groups | 115.267 | 2 | 57.633 | 1.842 | .178 |
| | Within Groups | 844.600 | 27 | 31.281 | | |
| | Total | 959.867 | 29 | | | |
| Task 5 | Between Groups | 328.467 | 2 | 164.233 | 1.140 | .335 |
| | Within Groups | 3890.900 | 27 | 144.107 | | |
| | Total | 4219.367 | 29 | | | |
| Task 6 | Between Groups | 1264.867 | 2 | 632.433 | 3.611 | .041 |
| | Within Groups | 4728.600 | 27 | 175.133 | | |
| | Total | 5993.467 | 29 | | | |

**Table E.2.5**

Dunnett t (<control) [a]

| Dependent Variable | (I) Experimental Phase | (J) Experimental Phase | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | 95% Confidence Interval Upper Bound |
|---|---|---|---|---|---|---|---|
| Task 1 | Phase II (a) | Phase II (c) | 2.200 | 4.152 | .843 | | 10.49 |
| | Phase II (b) | Phase II (c) | 3.100 | 4.152 | .892 | | 11.39 |
| Task 2 | Phase II (a) | Phase II (c) | -4.300 | 4.611 | .284 | | 4.91 |
| | Phase II (b) | Phase II (c) | -5.700 | 4.611 | .187 | | 3.51 |
| Task 3 | Phase II (a) | Phase II (c) | -4.400 | 3.063 | .137 | | 1.72 |
| | Phase II (b) | Phase II (c) | -5.100 | 3.063 | .093 | | 1.02 |
| Task 4 | Phase II (a) | Phase II (c) | 1.500 | 2.501 | .860 | | 6.50 |
| | Phase II (b) | Phase II (c) | 4.700 | 2.501 | .992 | | 9.70 |
| Task 5 | Phase II (a) | Phase II (c) | 8.100 | 5.369 | .979 | | 18.82 |
| | Phase II (b) | Phase II (c) | 3.800 | 5.369 | .884 | | 14.52 |
| Task 6 | Phase II (a) | Phase II (c) | -8.300 | 5.918 | .145 | | 3.52 |
| | Phase II (b) | Phase II (c) | -15.900* | 5.918 | .011 | | -4.08 |

*. The mean difference is significant at the .05 level.

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Since the treatment data sizes were equal in the three experimental phases, a Tukey's *post hoc* test was performed for the search task six in order to determine further which experimental systems differ from each other. The results ($p = 0.32$) (Table E.2.6) also showed that the phase II (b) and phase II (c) experimental phase have statistically significant differences.

**Table E.2.6**

Dependent Variable: Task 6
Tukey HSD

| (I) Experimental Phase | (J) Experimental Phase | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | Upper Bound |
|---|---|---|---|---|---|---|
| Phase II (a) | Phase II (a) | | | | | |
| | Phase II (b) | 7.600 | 5.918 | .416 | -7.07 | 22.27 |
| | Phase II (c) | -8.300 | 5.918 | .354 | -22.97 | 6.37 |
| Phase II (b) | Phase II (a) | -7.600 | 5.918 | .416 | -22.27 | 7.07 |
| | Phase II (b) | | | | | |
| | Phase II (c) | -15.900* | 5.918 | .032 | -30.57 | -1.23 |
| Phase II (c) | Phase II (a) | 8.300 | 5.918 | .354 | -6.37 | 22.97 |
| | Phase II (b) | 15.900* | 5.918 | .032 | 1.23 | 30.57 |
| | Phase II (c) | | | | | |

*. The mean difference is significant at the .05 level.

Finally, to guard against the possibility that the assumption of normal distribution did not hold, a *non-parametric* Kruskal-Wallis test was performed on the total numbers of clicks clicked to complete *all* search tasks for each experimental phase. The results ($X^2$ = *0.954* and *p = 0.621*) (Table E.2.7) also showed *no* statistically significant differences.

**Table E.2.7**

**Ranks**

| | Experimental Phase | N | Mean Rank |
|---|---|---|---|
| Total Number of Clicks | Phase II (a) | 10 | 16.35 |
| | Phase II (b) | 10 | 13.30 |
| | Phase II (c) | 10 | 16.85 |
| | Total | 30 | |

**Test Statistics [a,b]**

| | Total Number of Clicks |
|---|---|
| Chi-Square | .954 |
| df | 2 |
| Asymp. Sig. | .621 |

a. Kruskal Wallis Test
b. Grouping Variable: Experimental Phase

Further, detailed Kruskal-Wallis test was performed on the total number of clicks clicked to complete an *individual* search task to determine whether or not any significant differences exist between the three experimental systems. The results showed that except for the search task six ($X^2$ = *8.987* and *p = 0.011*)

(Table E.2.8), there is *no* statistically significant differences in the total number of clicks between the three experimental systems.

**Table E.2.8**

**Test Statistics [a,b]**

| | Task 1 | Task 2 | Task 3 | Task 4 | Task 5 | Task 6 |
|---|---|---|---|---|---|---|
| Chi-Square | 1.173 | 1.118 | 2.506 | 3.891 | 4.780 | 8.987 |
| df | 2 | 2 | 2 | 2 | 2 | 2 |
| Asymp. Sig. | .556 | .572 | .286 | .143 | .092 | .011 |

a. Kruskal Wallis Test

b. Grouping Variable: Experimental Phase

Both *parametric* and *non-parametric* tests results for the total number of clicks clicked to complete *all* search tasks established *no* statistically significant between the three experimental systems. However, both detailed *parametric* and *non-parametric* tests for the total number of clicks clicked to complete the search task six established statistically significant differences between the experimental systems.

## E.3 Number of Hits (Hypothesis 1.3)

The descriptive statistic results of the total number of hits browsed to complete *all* search tasks in each experimental phase are shown in Table E.3.1. It shows that the phase II (b) ($\overline{X}$ = *15.40*) and phase II (c) ($\overline{X}$ = *24.50*) have the lowest mean (i.e., fewer hits) and highest mean (i.e., more hits), respectively.

**Table E.3.1**

Total Number of Hits

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean Lower Bound | 95% Confidence Interval for Mean Upper Bound | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| Phase II (a) | 10 | 21.1000 | 5.21643 | 1.64958 | 17.3684 | 24.8316 | 15.00 | 29.00 |
| Phase II (b) | 10 | 15.4000 | 4.42719 | 1.40000 | 12.2330 | 18.5670 | 7.00 | 24.00 |
| Phase II (c) | 10 | 24.5000 | 7.16860 | 2.26691 | 19.3719 | 29.6281 | 15.00 | 39.00 |
| Total | 30 | 20.3333 | 6.71249 | 1.22553 | 17.8268 | 22.8398 | 7.00 | 39.00 |

A *parametric* one-way analysis of variance (ANOVA) and Dunnett's one-tailed tests were performed on the *total* number of hits browsed to complete all search tasks to determine whether or not any significance differences exist between the three experimental systems. The results (Anova $p = 0.005$ and Dunnett's one-tailed $p = 0.001$ for the phase II (b) and II (c)) (Table E.3.2 and Table E.3.3) showed statistically *highly* significant differences between the experimental systems.

**Table E.3.2**

Total Number of Hits

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 422.867 | 2 | 211.433 | 6.459 | .005 |
| Within Groups | 883.800 | 27 | 32.733 | | |
| Total | 1306.667 | 29 | | | |

**Table E.3.3**

Dependent Variable: Total Number of Hits
Dunnett t (<control) [a]

| (I) Experimental Phase | (J) Experimental Phase | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | Upper Bound |
|---|---|---|---|---|---|---|
| Phase II (a) | Phase II (c) | -3.400 | 2.559 | .163 | | 1.71 |
| Phase II (b) | Phase II (c) | -9.100* | 2.559 | .001 | | -3.99 |

*. The mean difference is significant at the .05 level.

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Since the treatment data sizes were equal in the three experimental phases, a Tukey's *post hoc* test was performed in order to determine further which experimental systems differ from each other. The results ($p = 0.004$) (Table E.3.4) also showed that the phase II (b) and phase II (c) experimental systems have *highly* significant differences.

**Table E.3.4**

Dependent Variable: Total Number of Hits

| | (I) Experimental Phase | (J) Experimental Phase | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | Upper Bound |
|---|---|---|---|---|---|---|---|
| Tukey HSD | Phase II (a) | Phase II (a) | | | | | |
| | | Phase II (b) | 5.700 | 2.559 | .084 | -.64 | 12.04 |
| | | Phase II (c) | -3.400 | 2.559 | .392 | -9.74 | 2.94 |
| | Phase II (b) | Phase II (a) | -5.700 | 2.559 | .084 | -12.04 | .64 |
| | | Phase II (b) | | | | | |
| | | Phase II (c) | -9.100* | 2.559 | .004 | -15.44 | -2.76 |
| | Phase II (c) | Phase II (a) | 3.400 | 2.559 | .392 | -2.94 | 9.74 |
| | | Phase II (b) | 9.100* | 2.559 | .004 | 2.76 | 15.44 |
| | | Phase II (c) | | | | | |
| Dunnett t (<control) [a] | Phase II (a) | Phase II (a) | | | | | |
| | | Phase II (b) | | | | | |
| | | Phase II (c) | -3.400 | 2.559 | .163 | | 1.71 |
| | Phase II (b) | Phase II (a) | | | | | |
| | | Phase II (b) | | | | | |
| | | Phase II (c) | -9.100* | 2.559 | .001 | | -3.99 |
| | Phase II (c) | Phase II (a) | | | | | |
| | | Phase II (b) | | | | | |
| | | Phase II (c) | | | | | |

*. The mean difference is significant at the .05 level.

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Finally, to guard against the possibility that the assumption of normal distribution did not hold, a non-parametric Kruskal-Wallis test was performed on the total numbers of hits browsed to complete all search tasks for each experimental phase. The results ($X^2$ = 10.448 and *p* = 0.005) (Table E.3.5) showed statistically highly significant differences as well.

**Table E.3.5**

| | Experimental Phase | N | Mean Rank |
|---|---|---|---|
| Total Number of Hits | Phase II (a) | 10 | 17.35 |
| | Phase II (b) | 10 | 8.45 |
| | Phase II (c) | 10 | 20.70 |
| | Total | 30 | |

**Test Statistics [a,b]**

| | Total Number of Hits |
|---|---|
| Chi-Square | 10.448 |
| df | 2 |
| Asymp. Sig. | .005 |

a. Kruskal Wallis Test

b. Grouping Variable: Experimental Phase

Both *parametric* (one-way ANOVA and Dunnett's one-tailed) and non *non-parametric* (Kruskal-Wallis) tests for the total number of hits browsed to complete *all* search tasks established statistically significant between the three experimental systems.

## E.4 Number of URLs (Hypothesis 1.4)

The descriptive statistic results of total number of URLs visited to complete *all* search tasks in each experimental phase are shown in Table E.4.1. It shows that the phase II (b) ( $\overline{X}$ = *15.40*) and phase II (c) ( $\overline{X}$ = *24.50*) have the lowest mean (i.e., fewer URLs) and highest mean (i.e., more URLs), respectively.

**Table E.4.1**

Total Number of URLs

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound | | |
| Phase II (a) | 10 | 20.7000 | 4.66786 | 1.47611 | 17.3608 | 24.0392 | 15.00 | 27.00 |
| Phase II (b) | 10 | 15.4000 | 4.42719 | 1.40000 | 12.2330 | 18.5670 | 7.00 | 24.00 |
| Phase II (c) | 10 | 24.5000 | 7.16860 | 2.26691 | 19.3719 | 29.6281 | 15.00 | 39.00 |
| Total | 30 | 20.2000 | 6.57267 | 1.20000 | 17.7457 | 22.6543 | 7.00 | 39.00 |

A *parametric* one-way analysis of variance (ANOVA) and Dunnett's one-tailed tests were performed on the *total* number of URLs visited to complete all search tasks to determine whether or not any significance differences exist between the three experimental systems. The results (Anova $p = 0.004$ and Dunnett's one-tailed $p = 0.001$ for phase II (b) and II (c)) (Table E.4.2 and Table E.4.3) showed statistically *highly* significant differences between experimental systems.

**Table E.4.2**

Total Number of URLs

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 417.800 | 2 | 208.900 | 6.755 | .004 |
| Within Groups | 835.000 | 27 | 30.926 | | |
| Total | 1252.800 | 29 | | | |

**Table E.4.3**

Dependent Variable: Total Number of URLs
Dunnett t (<control) [a]

| (I) Experimental Phase | (J) Experimental Phase | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | 95% Confidence Interval Upper Bound |
|---|---|---|---|---|---|---|
| Phase II (a) | Phase II (c) | -3.800 | 2.487 | .118 | | 1.17 |
| Phase II (b) | Phase II (c) | -9.100* | 2.487 | .001 | | -4.13 |

*. The mean difference is significant at the .05 level.

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Since the treatment data sizes were equal in the three experimental phases, a Tukey's *post hoc* test was performed in order to determine further which experimental systems differ from each other. The results (*p = 0.003*) (Table E.4.4) also showed that the phase II (b) and phase II (c) experimental systems have the significant differences.

**Table E.4.4**

Dependent Variable: Total Number of URLs

| | (I) Experimental Phase | (J) Experimental Phase | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | 95% Confidence Interval Upper Bound |
|---|---|---|---|---|---|---|---|
| Tukey HSD | Phase II (a) | Phase II (a) | | | | | |
| | | Phase II (b) | 5.300 | 2.487 | .102 | -.87 | 11.47 |
| | | Phase II (c) | -3.800 | 2.487 | .294 | -9.97 | 2.37 |
| | Phase II (b) | Phase II (a) | -5.300 | 2.487 | .102 | -11.47 | .87 |
| | | Phase II (b) | | | | | |
| | | Phase II (c) | -9.100* | 2.487 | .003 | -15.27 | -2.93 |
| | Phase II (c) | Phase II (a) | 3.800 | 2.487 | .294 | -2.37 | 9.97 |
| | | Phase II (b) | 9.100* | 2.487 | .003 | 2.93 | 15.27 |
| | | Phase II (c) | | | | | |
| Dunnett t (<control) [a] | Phase II (a) | Phase II (a) | | | | | |
| | | Phase II (b) | | | | | |
| | | Phase II (c) | -3.800 | 2.487 | .118 | | 1.17 |
| | Phase II (b) | Phase II (a) | | | | | |
| | | Phase II (b) | | | | | |
| | | Phase II (c) | -9.100* | 2.487 | .001 | | -4.13 |
| | Phase II (c) | Phase II (a) | | | | | |
| | | Phase II (b) | | | | | |
| | | Phase II (c) | | | | | |

*. The mean difference is significant at the .05 level.

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Finally, to guard against the possibility that the assumption of normal distribution did not hold, a *non-parametric* Kruskal-Wallis test was performed to test the total numbers of URLs visited to complete *all* search tasks for each experimental phase. The results ($X^2$ = *10.648* and *p = 0.005*) (Table E.4.5) showed statistically *highly* significant differences as well.

**Table E.4.5**

| | Experimental Phase | N | Mean Rank |
|---|---|---|---|
| Total Number of URLs | Phase II (a) | 10 | 17.15 |
| | Phase II (b) | 10 | 8.45 |
| | Phase II (c) | 10 | 20.90 |
| | Total | 30 | |

**Test Statistics [a,b]**

| | Total Number of URLs |
|---|---|
| Chi-Square | 10.648 |
| df | 2 |
| Asymp. Sig. | .005 |

a. Kruskal Wallis Test

b. Grouping Variable: Experimental Phase

Both *parametric* (one-way ANOVA and Dunnett's one-tailed) and non *non-parametric* (Kruskal-Wallis) tests for the total number of URLs visited to complete *all* search tasks established statistically *highly* significant in the three experimental systems.

## E.5 Length of time (Hypothesis 1.5)

The descriptive statistic results for the total number of time taken to complete *all* search tasks in each experimental phase are shown in Table E.5.1. It shows that the phase II (b) ($\bar{X}$ = *1640.00*) and phase II (c) ($\bar{X}$ = *2291.70*) have the lowest mean (i.e., less time) and highest (i.e., more time) mean, respectively.

**Table E.5.1**

Total Time Taken to Reach Target Information (in seconds)

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound | | |
| Phase II (a) | 10 | 2186.2000 | 791.52171 | 250.30114 | 1619.9795 | 2752.4205 | 1300.00 | 3629.00 |
| Phase II (b) | 10 | 1640.0000 | 510.76326 | 161.51753 | 1274.6220 | 2005.3780 | 1035.00 | 2652.00 |
| Phase II (c) | 10 | 2291.7000 | 1250.44801 | 395.42638 | 1397.1834 | 3186.2166 | 679.00 | 5230.00 |
| Total | 30 | 2039.3000 | 919.26343 | 167.83377 | 1696.0414 | 2382.5586 | 679.00 | 5230.00 |

A *parametric* one-way analysis of variance (ANOVA) and Dunnett's one-tailed tests were performed on the *total* number of time taken to complete all search tasks to determine whether or not any significance differences exist between the three experimental systems. The results (Anova $p$ = *0.242*, Dunnett's one-tailed: $p = 0.559$ for phase II (a) and II (c) and $p = 0.102$ for phase II (b) and II (c)) (Table E.5.2 and Table E.5.3) showed *no* statistically significant differences between the three experimental systems.

**Table E.5.2**

Total Time Taken to Reach Target Information (in seconds)

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 2447258.600 | 2 | 1223629.300 | 1.498 | .242 |
| Within Groups | 22059053.700 | 27 | 817001.989 | | |
| Total | 24506312.300 | 29 | | | |

**Table E.5.3**

Dependent Variable: Total Time Taken to Reach Target Information (in seconds)
Dunnett t (<control) [a]

| | | | | | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| (I) Experimental Phase | (J) Experimental Phase | Mean Difference (I-J) | Std. Error | Sig. | Lower Bound | Upper Bound |
| Phase II (a) | Phase II (c) | -105.500 | 404.228 | .559 | | 701.91 |
| Phase II (b) | Phase II (c) | -651.700 | 404.228 | .102 | | 155.71 |

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Further, a one-way ANOVA and Dunnett's one-tailed tests were performed to test whether the total number of time taken to complete an *individual* search task has any significant differences in the three experimental systems. The results (Table E.5.4 and Table E.5.5) showed that except for the search task six (Anova $p = 0.011$ and Dunnett's one-tailed $p = 0.003$ for phase II (b) and II(c)), there is *no* statistically significant differences between experimental systems.

**Table E.5.4**

| | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| Task 1 | Between Groups | 89637.800 | 2 | 44818.900 | .459 | .636 |
| | Within Groups | 2633636.500 | 27 | 97542.093 | | |
| | Total | 2723274.300 | 29 | | | |
| Task 2 | Between Groups | 9079.467 | 2 | 4539.733 | .122 | .886 |
| | Within Groups | 1005197.500 | 27 | 37229.537 | | |
| | Total | 1014276.967 | 29 | | | |
| Task 3 | Between Groups | 246898.467 | 2 | 123449.233 | 1.809 | .183 |
| | Within Groups | 1842232.500 | 27 | 68230.833 | | |
| | Total | 2089130.967 | 29 | | | |
| Task 4 | Between Groups | 29702.600 | 2 | 14851.300 | .568 | .573 |
| | Within Groups | 705390.600 | 27 | 26125.578 | | |
| | Total | 735093.200 | 29 | | | |
| Task 5 | Between Groups | 181028.867 | 2 | 90514.433 | 1.786 | .187 |
| | Within Groups | 1368400.100 | 27 | 50681.485 | | |
| | Total | 1549428.967 | 29 | | | |
| Task 6 | Between Groups | 727176.067 | 2 | 363588.033 | 5.310 | .011 |
| | Within Groups | 1848869.300 | 27 | 68476.641 | | |
| | Total | 2576045.367 | 29 | | | |

**Table E.5.5**

Dunnett t (<control) [a]

| Dependent Variable | (I) Experimental Phase | (J) Experimental Phase | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | 95% Confidence Interval Upper Bound |
|---|---|---|---|---|---|---|---|
| Task 1 | Phase II (a) | Phase II (c) | 1.300 | 139.673 | .670 | | 280.29 |
| | Phase II (b) | Phase II (c) | -115.300 | 139.673 | .323 | | 163.69 |
| Task 2 | Phase II (a) | Phase II (c) | 12.800 | 86.290 | .723 | | 185.16 |
| | Phase II (b) | Phase II (c) | -28.800 | 86.290 | .527 | | 143.56 |
| Task 3 | Phase II (a) | Phase II (c) | -128.100 | 116.817 | .229 | | 105.23 |
| | Phase II (b) | Phase II (c) | -221.300 | 116.817 | .061 | | 12.03 |
| Task 4 | Phase II (a) | Phase II (c) | 61.000 | 72.285 | .910 | | 205.38 |
| | Phase II (b) | Phase II (c) | 71.300 | 72.285 | .932 | | 215.68 |
| Task 5 | Phase II (a) | Phase II (c) | 174.400 | 100.679 | .988 | | 375.50 |
| | Phase II (b) | Phase II (c) | 21.300 | 100.679 | .746 | | 222.40 |
| Task 6 | Phase II (a) | Phase II (c) | -226.900 | 117.027 | .056 | | 6.85 |
| | Phase II (b) | Phase II (c) | -378.900* | 117.027 | .003 | | -145.15 |

*. The mean difference is significant at the .05 level.

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Since the treatment data sizes were equal in the three experimental phases, a Tukey's *post hoc* test was performed for the search task six in order to determine further which experimental systems differ from each other. The results ($p = 0.009$) (Table E.5.6) also showed that phase II (b) and phase II (c) experimental systems have the significant differences.

**Table E.5.6**

Dependent Variable: Task 6
Tukey HSD

| (I) Experimental Phase | (J) Experimental Phase | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | 95% Confidence Interval Upper Bound |
|---|---|---|---|---|---|---|
| Phase II (a) | Phase II (a) | | | | | |
| | Phase II (b) | 152.000 | 117.027 | .408 | -138.16 | 442.16 |
| | Phase II (c) | -226.900 | 117.027 | .147 | -517.06 | 63.26 |
| Phase II (b) | Phase II (a) | -152.000 | 117.027 | .408 | -442.16 | 138.16 |
| | Phase II (b) | | | | | |
| | Phase II (c) | -378.900* | 117.027 | .009 | -669.06 | -88.74 |
| Phase II (c) | Phase II (a) | 226.900 | 117.027 | .147 | -63.26 | 517.06 |
| | Phase II (b) | 378.900* | 117.027 | .009 | 88.74 | 669.06 |
| | Phase II (c) | | | | | |

*. The mean difference is significant at the .05 level.

Finally, to guard against the possibility that the assumption of normal distribution did not hold, a *non-parametric* Kruskal-Wallis test was performed to test the total number of time taken to complete *all* search tasks for each experimental phase. The results ($X^2$= *3.510* and *p* = *0.173*) (Table E.5.7) showed *no* statistically significant differences.

**Table E.5.7**

| | Experimental Phase | N | Mean Rank |
|---|---|---|---|
| Total Time Taken to Reach Target Information (in seconds) | Phase II (a) | 10 | 17.85 |
| | Phase II (b) | 10 | 11.25 |
| | Phase II (c) | 10 | 17.40 |
| | Total | 30 | |

**Test Statistics [a,b]**

| | Total Time Taken to Reach Target Information (in seconds) |
|---|---|
| Chi-Square | 3.510 |
| df | 2 |
| Asymp. Sig. | .173 |

a. Kruskal Wallis Test
b. Grouping Variable: Experimental Phase

Further, detailed Kruskal-Wallis test was performed on the total number of time taken to complete an *individual* search task to determine whether or not any significant differences exist between the three experimental systems. The results showed that except for the search task six ($X^2$= *9.177* and *p* = *0.010*)

(Table E.5.8), there is *no* statistically significant differences in the total time taken between the three experimental systems.

**Table E.5.8**

**Test Statistics [a,b]**

| | Task 1 | Task 2 | Task 3 | Task 4 | Task 5 | Task 6 |
|---|---|---|---|---|---|---|
| Chi-Square | 1.216 | .034 | 3.414 | 1.985 | 5.024 | 9.177 |
| df | 2 | 2 | 2 | 2 | 2 | 2 |
| Asymp. Sig. | .544 | .983 | .181 | .371 | .081 | .010 |

a. Kruskal Wallis Test

b. Grouping Variable: Experimental Phase

Both *parametric* and non *non-parametric* tests for the total number of time taken complete *all* search tasks established *no* statistically significant between the three experimental systems. However, detailed *parametric* tests for the total number of time taken to complete the search task six established statistically significant differences between the phase II (b) and phase II (c).

# References

Aguillo, I. (2000). A new generation of tools for search, recovery and quality evaluation of World Wide Web medical resources. *Journal of Management in Medicine, 14*(3/4), 240-248.

Allan, J. (1996). *Incremental relevance feedback for information filtering.* Paper presented at the 19th annual international ACM SIGIR conference on Research and development in information retrieval, Zurich, Switzerland.

Allan, J., Harper, D. J., Hiemstra, D., Hofmann, T., Hovy, E., Kraaij, W., et al. (2003). Challenges in information retrieval and language modeling. *ACM SIGIR Forum, 37*(1), 31-47.

Allen, D., & Wilson, T. D. (2003). Information overload: context and causes. *New Review of Information Behaviour Research, 4*(1), 31-44.

Amati, G., D'Aloisi, D., Giannini, V., & Ubaldini, F. (1997). A Framework for filtering news and managing distributed data. *Journal of Universal Computer Science, 3*(8), 1007-1021.

Anderson, J. R. (1984). *Cognitive psychology and intelligent tutoring.* Paper presented at the 6th Annual Cognitive Science Meetings.

Aridor, Y., Carmel, D., Lempel, R., Soffer, A., & Maarek, Y. S. (2000). *Knowledge agents on the Web.* Paper presented at the 4th International Workshop on Cooperative Information Agents IV, The Future of Information Agents in Cyberspace, Boston, USA.

Attar, R., & Fraenkel, A. S. (1977). Local feedback in full-text retrieval systems. *Journal of the ACM (JACM), 24*(3), 397-417.

Bai, J., Song, D., Bruza, P., Nie, J.-Y., & Cao, G. (2005). *Query expansion using term relationships in language models for information retrieval.* Paper presented at the 14th ACM international conference on information and knowledge management Bremen, Germany.

Balfe, E., & Smyth, B. (2004). *Query mining for community based web search.* Paper presented at the 2004 IEEE/WIC/ACM International Conference on Web Intelligence, Beijing, China.

Bar-Ilan, J. (1998). On the overlap, the precision and estimated recall of search engines. A case study of the query “Erdos”. *Scientometrics, 42*(3), 207-228.

Bathke, A. C., Harrar, S. W., & Madden, L. V. (2008). How to compare small multivariate samples using nonparametric tests *Computational Statistics & Data Analysis, 52*(1), 4951-4965.

Bauer, T. L., & Leake, D. B. (2001). *WordSieve: A method for real-time context extraction.* Paper presented at the 3rd International and Interdisciplinary Conference on Modeling and Using Context, Dundee, Scotland.

Bauer, T. L., & Leake, D. B. (2003). Detecting context-differentiating terms using competitive learning. *ACM SIGIR Forum, 37*(2), 4-17.

Beaulieu, M. (1997). Experiments on interfaces to support query expansion. *Journal of Documentation, 53*(1), 8-19.

Belkin, N. J. (2000). Helping people find what they don't know. *Communications of the ACM, 43*(8), 58-61.

Belkin, N. J., Cool, C., Head, J., Jeng, J., Kelly, D., & Lin, S. (2000). *Relevance feedback versus local context analysis as term suggestion devices: Rutgers’ TREC-8 interactive track experience.* Paper presented at the 8th Text Retrieval Conference (TREC 8), Gaithersburg, Maryland.

Belkin, N. J., Cool, C., Kelly, D., Lin, S.-J., Park, S. Y., Perez-Carballo, J., et al. (2001). Iterative exploration, design and evaluation for query reformulation in interactive information retrieval. *Special issue on interactivity at the text retrieval conference (TREC), 37*(3), 403-434.

Belkin, N. J., Kelly, D., Kim, G., Kim, J.-Y., Lee, H.-J., Muresan, G., et al. (2002). *Query length in interactive information retrieval.* Paper presented at the 26th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, Toronto, Canada.

Berghel, H. (1997). Cyberspace 2000: dealing with information overload. *Communications of the ACM, 40*(2), 19-24.

Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. F., & Secret, A. (1994). The world-wide web. *Communications of the ACM, 37*(8), 76-82.

Berry, M. W., & Browne, M. (2005). *Understanding search engines: mathematical modeling and text retrieval* (Second ed.). Philadelphia, PA: Society for Industrial and Applied Mathematics.

Bhavnani, S. K., Christopher, B. K., Johnson, T. M., Little, R. J., Peck, F. A., Schwartz, J. L., et al. (2003). *Strategy hubs: next-generation domain portals with search* *procedures.* Paper presented at the SIGCHI conference on Human factors in computing systems Florida, USA.

Billerbeck, B., Scholer, F., Williams, H. E., & Zobel, J. (2003). *Query expansion using associated queries.* Paper presented at the 12th international conference on Information and knowledge management, New Orleans, LA, USA.

Bonino, D., Corno, F., & Pescarmona, F. (2005). *Automatic learning of text-to-concept mappings exploiting WordNet-like lexical networks.* Paper presented at the 2005 ACM symposium on Applied computing, Santa Fe, New Mexico.

Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation, 56*(1), 71-90.

Borlund, P. (2003). The IIR evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research, 8*(3).

Borlund, P., & Ingwersen, P. (1997). The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation, 53*(3), 225-250.

Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User Adapted Interaction, 11*, 87-110.

Brusilovsky, P. (2004). *KnowledgeTree: a distributed architecture for adaptive e-learning.* Paper presented at the 13th international World Wide Web conference, New York, NY, USA.

Budzik, J., & Hammond, K. (1999). *Watson: anticipating and contextualizing information needs.* Paper presented at the 62nd Annual Meeting of the American Society for Information Science, Washington, USA.

Buscaldi, D., Rosso, P., & Arnal, E. S. (2005, 21-23 September). *A WordNet-based query expansion method for geographical information retrieval.* Paper presented at the Cross-Language Evaluation Forum 2005 WORKSHOP, Vienna, Austria.

Butterworth, R., & Perkins, V. D. (2006). *Using the information seeking and retrieval framework to analyse non-professional information use.* Paper presented at the 1st international conference on Information interaction in context Copenhagen, Denmark.

Byström, K., & Järvelin, K. (1995). Task complexity affects information seeking and use. *Information processing & management, 31*(2), 191-213.

Carlson, C. N. (2003). *Information overload, retrieval strategies and Internet user empowerment.* Paper presented at the The Good, the Bad and the Irrelevant, Helsinki, Finland.

Case, D. O. (2002). *Looking for information: a survey of research on information seeking, needs, and behavior*. San Diego, California, USA: Academic Press.

Challam, V. K. R. (2004). *Contextual information retrieval using ontology-based user profiles.* Unpublished Master's Thesis, University of Kansas.

Chau, M., Zeng, D., & Chen, H. (2001 ). *Personalized Spiders for Web Search and Analysis.* Paper presented at the 1st ACM/IEEE-CS joint conference on Digital libraries Roanoke, Virginia, United States.

Chen, L., & Sycara, K. (1998). *WebMate: a personal agent for browsing and searching.* Paper presented at the International Conference on Autonomous Agents, Minneapolis, Minnesota, USA.

Chen, S.-C., Gulati, S., Hamidt, S., Huang, X., Luo, L., Morisseau-Leroy, N., et al. (2003). *A three-tier system architecture design and development for hurricane occurrence simulation.* Paper presented at the IEEE International Conference on Information Technology: Research and Education (ITRE 2003), Las Vegas, Nevada, USA.

Choo, C. W., Detlor, B., & Turnbull, D. (2000). Information seeking on the web: an integrated model of browsing and searching *First Monday, 5*(2).

Claypool, M., Le, P., Wased, M., & Brown, D. (2001). *Implicit interest indicators.* Paper presented at the 6th international conference on intelligent user interfaces, Santa Fe, New Mexico, USA.

Cleverdon, C. W. (1960). *ASLIB Cranfield research project: report on the first stage of an investigation into the comparative efficiency of indexing systems*: College of Aeronautics, Cranfield.

Cosijn, E., & Ingwersen, P. (2000). Dimensions of relevance. *Information processing & management, 36*(4), 533-550.

Crnkovic, I., & Larsson, M. (2000). *A case study: demands on component-based development.* Paper presented at the 22nd international conference on Software engineering, Limerick, Ireland.

Croft, W. B., Cronen-Townsend, S., & Lavrenko, V. (2001). *Relevance feedback and personalization: a language modeling perspective.* Paper presented at the DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries, Dublin City University, Ireland.

Cui, H., Wen, J.-R., Nie, J.-Y., & Ma, W.-Y. (2002). *Probabilistic query expansion using query logs.* Paper presented at the 11th international conference on World Wide Web, Honolulu, Hawaii, USA.

Curran, K., & Doherty, A. (2006). *Automated broadcast media monitoring using the Google API.* Paper presented at the IEEE Consumer Communications and Networking Conference, Las Vegas, Nevada, USA.

Deerwester, S., T., D. S., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society of Information Science, 41*(6), 391-407.

Dervin, B. (1999). On studying information seeking methodologically: the implications of connecting metatheory to method. *Information processing & management, 35*(6), 727-750.

Dumais, S. T., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., & Robbins, D. C. (2003). *Stuff I've seen: a system for personal information retrieval and re-use.* Paper presented at the 26th annual international ACM SIGIR conference on Research and development in information retrieval Toronto, Canada.

Durrani, Q. S. (1997). *Cognitive modeling: a domain independent user modeling.* Paper presented at the IEEE International conference on System man and cybernetics, Orlando, FL, USA.

Eastman, C. M., & Jansen, B. J. (2003). Coverage, Relevance, and Ranking: The Impact of Query Operators on Web Search Engine Results. *ACM Transactions on Information Systems (TOIS), 21*(4), 383-422.

Edwards, S. L., & Bruce, C. (2002). Reflective internet searching: an action research model. *The Learning Organization, 9*(3/4), 180-188.

Efthimiadis, E. N. (1996). Query expansion. *Annual Review of Information Systems and Technology (ARIST), 31*, 121-187.

Ellis, D. (1989). A behavioral approach to information retrieval design. *Journal of documentation, 45*(3), 318-338.

Ellis, D. (1996). *Progress and problems in information retrieval*. London: Library Association Publishing.

Etienne, J., Wachmann, B., & Zhang, L. (2006). *A component-based framework for knowledge discovery in bioinformatics.* Paper presented at the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, Philadelphia, PA, USA.

Fan, W., Gordon, M. D., & Pathak, P. (2004). Discovery of context-specific ranking functions for effective information retrieval using genetic programming. *IEEE Transactions on Knowledge and Data Engineering, 16*(4), 523-527.

Fensel, D., Decker, S., Erdmann, M., & Studer, R. (1998, April). *Ontobroker: how to make the WWW Intelligent.* Paper presented at the 11th Knowledge Acquisition Workshop, Banff, Alberta, Canada.

Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., et al. (2002). Placing search in context: The concept revisited. *ACM Transactions on Information Systems (TOIS), 20*(1), 116-131.

Fonseca, B. M., Golgher, P., Pôssas, B., Ribeiro-Neto, B., & Ziviani, N. (2005). *Concept-based interactive query expansion.* Paper presented at the 14th ACM international conference on Information and knowledge management, Bremen, Germany.

French, J. C., Brown, D. E., & Kim, N.-H. (1997). A classification approach to Boolean query reformulation. *Journal of the American Society for Information Science, 48*(8), 694-706.

Freund, L., & Toms, E. G. (2002). *A preliminary contextual analysis of the web query process.* Paper presented at the 30th Annual Conference of the Canadian Association for Information Science, Toronto.

Gao, X., Murugesan, S., & Lo, B. (2004). *Multi-dimensional evaluation of information retrieval results.* Paper presented at the 2004 IEEE/WIC/ACM International Conference on Web Intelligence.

Gaslikova, I. (1999). "Information Seeking in Context" and the development of information systems. *Information Research, 5*(1).

Gauch, S., Chaffee, J., & Pretschner, A. (2003). Ontology-based personalized search and browsing. *Web Intelligence and Agent System, 1*(3-4), 219-234.

Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2000). *Eigentaste: a constant time collaborative filtering algorithm*: UCB Electronics Research Laboratory Technical Report.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199-220.

Hancock-Beaulieu, M., & Walker, S. (1992). An evaluation of automatic query expansion in an online library catalogue. *Journal of Documentation, 48*(4), 406-421.

Harman, D. (1992). *Relevance feedback revisited.* Paper presented at the 15th annual international ACM SIGIR conference on Research and development in information retrieval, Copenhagen, Denmark.

Harper, D. J., & Kelly, D. (2006). *Contextual relevance feedback.* Paper presented at the 1st international conference on Information interaction in context Copenhagen, Denmark.

Harter, S. P. (1996). Variations in relevance assessments and the measurement of retrieval effectiveness. *Journal of the American Society for Information Science 47*(1), 37-49.

Henzinger, M. R. (2004). *The past, present and future of web information retrieval.* Paper presented at the 23rd ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, Paris, France.

Henzinger, M. R., Motwani, R., & Silverstein, C. (2002). Challenges in web search engines. *SIGIR Forum, 36*(2), 11-22.

Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J. (1999). *An algorithmic framework for performing collaborative filtering.* Paper presented at the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, Berkeley, California, USA.

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly, 28*(1), 75-105.

Hjørland, B. (2002). Domain analysis in information science: Eleven approaches – traditional as well as innovative. *Journal of Documentation 58*(4), 422-462.

Hoashi, K., Matsumoto, K., & Inoue, N. (2003). *Personalization of user profiles for content-based music retrieval based on relevance feedback.* Paper presented at the 11th ACM international conference on Multimedia, Berkeley, CA, USA.

Hoeber, O., & Yang, X. D. (2006a). *A comparative user study of web search interfaces: HotMap, Concept Highlighter, and Google.* Paper presented at the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, Kowloon, Hong Kong.

Hoeber, O., & Yang, X. D. (2006b). *Interactive web information retrieval using WordBars.* Paper presented at the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, Kowloon, Hong Kong.

Hofgesang, P. I. (2007). *Web personalisation through incremental individual profiling and support-based user segmentation.* Paper presented at the IEEE/WIC/ACM International Conference on Web Intelligence, Silicon Valley, USA.

Holmes, V. P., Kleban, S. D., Miller, D. J., Pavlakos, C., Poore, C. A., Vandewart, R. L., et al. (2002). *An architecture and implementation to support large-scale data access in scientific simulation environments.* Paper presented at the 35th Annual Simulation Symposium, San Diego, California, USA.

Hölscher, C., & Strube, G. (2000). Web search behavior of internet experts and newbies. *Computer Networks, 33*(1-6), 337-346.

Horng, Y.-J., Chen, S.-M., Chang, Y.-C., & Lee, C.-H. (2005). A new method for fuzzy information retrieval based on fuzzy hierarchical clustering and fuzzy inference techniques. *IEEE Transactions on Fuzzy Systems, 13*(2), 216-228.

Höscher, C., & Strube, G. (2000). *Web search behavior of Internet experts and newbies.* Paper presented at the 9th international World Wide Web conference, Amsterdam, The Netherlands.

Hsieh-Yee, I. (2001). Research on web search behavior. *Library & Information Science Research, 23*(2), 167-185.

Huang, C. K., Chien, L. F., & Oyang, Y. J. (2003). Relevant term suggestion in interactive web search based on contextual information in query session logs. *Journal of the American Society for Information Science and Technology (JASIST), 54*(7), 638.

Ingwersen, P. (1992). *Information retrieval interaction*. London, UK: Taylor Graham.

Ingwersen, P. (1996). Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory. *Journal of Documentation, 52*(1), 3-50.

Ingwersen, P., & Belkin, N. (2004). Information retrieval in context : IRiX. *ACM SIGIR Forum, Volume 38*, 50-52.

Jain, N., Dahlin, M., & Tewari, R. (2005). *Using bloom filters to refine web search results.* Paper presented at the 8th International Workshop on the Web and Databases, Baltimore, Maryland.

Jansen, B. J. (2005). Seeking and implementing automated assistance during the search process. *Information processing & management, 41*(4), 909-928.

Jansen, B. J., Booth, D. L., & Spink, A. (2007). *Determining the user intent of web search engine queries.* Paper presented at the 16th international conference on World Wide Web, Banff, Alberta, Canada.

Jansen, B. J., & Spink, A. (2003, 23 – 26 June). *An analysis of web information seeking and use: documents retrieved versus documents viewed.* Paper presented at the 4th International Conference on Internet Computing, Las Vegas, NV.

Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web. *Information processing & management, 36*(2), 207-227.

Järvelin, K., & Ingwersen, P. (2004). Information seeking research needs extension towards tasks and technology. *Information Research, 10*(1).

Järvelin, K., & Wilson, T. D. (2003). On conceptual models for information seeking and retrieval research. *Information Research, 9*(1).

Jing, Y., & Croft, W. B. (1994). *An association thesaurus for information retrieval.* Paper presented at the 4th international conference ‘‘Recherche d’Information Assistee par Ordinateur ’’, New York, US.

Joachims, T., Freitag, D., & Mitchell, T. (1997). *WebWatcher: a tour guide for the world wide web.* Paper presented at the International Joint Conference on Artificial Intelligence, Nagoya, Japan.

Joachims, T., Granka, L., Pan, B., Hembrooke, H., Radlinski, F., & Gay, G. (2007). Evaluating the accuracy of implicit feedback from clicks and query reformulations in Web search. *ACM Transactions on Information Systems (TOIS), 25*(2), Article 7.

Johnson, J. D. (2003). On contexts of information seeking. *Information processing & management, 39*(5), 735-760.

Jones, G. J. F., & Brown, P. J. (2004). *The role of context in information retrieval.* Paper presented at the SIGIR 2004 Information Retrieval in Context (IRiX) Workshop, Sheffield, UK.

Joshi, R. R., & Aslandogan, Y. A. (2006). *Concept-based Web Search using Domain Prediction and Parallel Query Expansion.* Paper presented at the IEEE International Conference on Information Reuse and Integration, Hawaii, USA.

Kari, J., & Savolainen, R. (2003). Towards a contextual model of information seeking on the Web *New Review of Information Behaviour Research, 4*(1), 155-175.

Keenoy, K., & Levene, M. (2005). Personalisation of web search. In *Intelligent Techniques for Web Personalisation* (Vol. 3169, pp. 201-228): Springer.

Kellar, M. (2006). *An examination of user behaviour during web information tasks.* Paper presented at the CHI '06 extended abstracts on Human factors in computing systems Montréal, Québec, Canada.

Kelly, D. (2004). *Understanding implicit feedback and document preference: a naturalistic user study.* Rutgers University, New Brunswick, NJ.

Kelly, D., & Cool, C. (2002). *The effects of topic familiarity on information search behavior.* Paper presented at the 2nd ACM/IEEE-CS joint conference on Digital libraries, Portland, Oregon, USA.

Kelly, D., & Fu, X. (2006). *Elicitation of term relevance feedback: an investigation of term source and context.* Paper presented at the 29th annual international ACM SIGIR conference on Research and development in information retrieval, Seattle, Washington, USA.

Kelly, D., & Teevan, J. (2003). Implicit feedback for inferring user preference: a bibliography. *SIGIR Forum, 32*(2), 18-28.

Kerschberg, L., Kim, W., & Scime, A. (2001). *WebSifter II: a personalizable meta-search agent based on weighted semantic taxonomy tree.* Paper presented at the International Conference on Internet Computing 2001, Las Vegas, Nevada, USA.

Kim, J., Oard, D. W., & Romanik, K. (2000). *Using implicit feedback for user modeling in internet and intranet searching.*: University of Maryland CLIS Technical Report.

Kim, S., Suh, E., & Keedong, Y. (2007). A study of context inference for web-based information systems. *Electronic Commerce Research and Applications, 6* (2), 146-158.

Klink, S., Hust, A., Junker, M., & Dengel, A. (2002, 19-23 August). *Collaborative learning of term-based concepts for automatic query expansion.* Paper presented at the 13th European Conference on Machine Learning, Helsinki, Finland.

Kobayashi, M., & Takeda, K. (2000). Information retrieval on the web. *ACM Computing Surveys (CSUR), 32*(2), 144-173.

Kobsa, A., & Teltzrow, M. (2005). *Contextualized communication of privacy practices and personalization benefits: impacts on users' data sharing and purchase behavior* (Vol. 3424): Springer Berlin / Heidelberg.

Kobsa, A., & Wahlster, W. (1989). *User models in dialog systems*: Springer Verlag Berlin.

Koenemann, J., & Belkin, N. J. (1996). *A case for interaction: a study of interactive information retrieval behavior and effectiveness.* Paper presented at the SIGCHI conference on Human factors in computing systems: common ground Vancouver, British Columbia, Canada.

Koo, M., & Skinner, H. (2003). Improving web searches: case study of quit-smoking web sites for teenagers. *Journal of Medical Internet Research, 5*(4), e28.

Kraft, R., Maghoul, F., & Chang, C. C. (2005). *Y!Q: contextual search at the point of inspiration.* Paper presented at the 14th ACM international conference on Information and knowledge management, Bremen, Germany.

Krulwich, B., & Burkey, C. (1997). The InfoFinder agent: learning user interests through heuristic phrase extraction. *IEEE Expert, 12*(5), 22-27.

Ku, W. S., Zimmermann, R., Wang, H., & Wan, C. N. (2005). *Adaptive nearest neighbor queries in travel time networks.* Paper presented at the 13th annual ACM international workshop on Geographic information systems, Bremen, Germany.

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.

Lazonder, A. W., Biemans, H. J. A., & Wopereis, I. G. J. H. (2000). Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science, 51*(6), 576-581.

Leake, D. B., & Scherle, R. (2001, January 14 -17). *Towards context-based search engine selection.* Paper presented at the International Conference on Intelligent User Interfaces, Santa Fe, New Mexico, USA.

Levene, M. (2006). *An introduction to search engines and web navigation*: Addison Wesley Publishing Company.

Lieberman, H. (1995). *Letizia: an agent that assists web browsing.* Paper presented at the 14th International Joint Conference on Artificial Intelligence Montreal, Quebec, Canada.

Lieberman, H., Dyke, N. V., & Vivacqua, A. (1999). *Let's browse: a collaborative web browsing agent.* Paper presented at the International Conference on Intelligent User Interfaces, Los Angeles, California, USA.

Limbu, D. K., Connor, A., & MacDonell, S. (2005). *A framework for contextual information retrieval from the WWW.* Paper presented at the 14th International Conference on Adaptive Systems and Software Engineering (IASSE05), Canada.

Limbu, D. K., Connor, A., Pears, R., & MacDonell, S. (2006). *Contextual relevance feedback in web information retrieval.* Paper presented at the 1st international conference on Information interaction in context Copenhagen, Denmark.

Limbu, D. K., Pears, R., Connor, A., & MacDonell, S. (2006, 7 - 10 July). *Contextual and concept-based interactive query expansion.* Paper presented at the 19th Annual Conference of the National Advisory Committee on Computing Qualifications, Wellington, New Zealand.

Lin, H. C., Wang, L. H., & Chen, S. M. (2006). Query expansion for document retrieval based on fuzzy rules and user relevance feedback techniques. *Expert Systems with Applications, 31*(2), 397-405.

Liu, F., Yu, C., & Meng, W. (2004). Personalized web search for improving retrieval effectiveness. *IEEE Transactions on Knowledge and Data Engineering, 16*(1), 28-40.

Liu, S., Yu, C., & Meng, W. (2005). *Word sense disambiguation in queries.* Paper presented at the 14th ACM international conference on Information and knowledge management, Bremen, Germany.

Liu, Y., Jin, R., & Chai, J. Y. (2006). A statistical framework for query translation disambiguation. *ACM Transactions on Asian Language Information Processing (TALIP), 5*(4), 360-387.

Ma, L., & Goharian, N. (2005). *Query length impact on misuse detection in information retrieval systems.* Paper presented at the 2005 ACM symposium on Applied computing, Santa Fe, New Mexico.

Maglio, P. P., Barrett, R., Campbell, C. S., & Selker, T. (2000). *SUITOR: an attentive information system.* Paper presented at the International Conference on Intelligent User Interfaces, New Orleans, Louisiana, USA.

Mandl, T. (2006). *Education and evaluation: implementation and evaluation of a quality-based search engine* Paper presented at the 17th conference on Hypertext and hypermedia, Odense, Denmark.

Marchionini, G. (1995). *Information seeking in electronic environments.* (First ed.). New York: Cambridge University Press.

Markellos, K., Markellou, P., Panayiotaki, A., & Tsakalidis, A. (2006). *Semantic web mining for personalized public E-services*: IGI Global.

Martin, I., & Jose, J. M. (2004, April 26-28). *Fetch: A personalised information retrieval tool.* Paper presented at the Recherche d'Information Assistée par Ordinateur, Avignon, France.

Martzoukou, K. (2005). A review of Web information seeking research: considerations of method and foci of interest. *Information Research, 10*(2).

McQuistan, S. (2000). Techniques for current answers: Part 1: Information overload and the internet. *The Journal of Audiovisual Media in Medicine, 23*(3), 124.

Meadow, C. T., Boyce, B. R., & Kraft, D. H. (2000). *Text information retrieval systems* (Second ed.): Academic Press.

Morita, M., & Shinoda, Y. (1994). *Information filtering based on user behavior analysis and best match text retrieval.* Paper presented at the 17th annual international ACM SIGIR conference on Research and development in information retrieval, Dublin, Ireland.

Morris, D., Morris, M. R., & Venolia, G. (2008). *SearchBar: A Search-Centric Web History for Task Resumption and Information Re-finding.* Paper presented at the 26th annual SIGCHI conference on Human factors in computing systems Florence, Italy.

Motomura, Y., Yoshida, K., & Fujimoto, K. (2000). *Generative user models for adaptive information retrieval.* Paper presented at the 2000 IEEE International Conference on Systems, Man, and Cybernetics, Nashville, TN, USA.

Nanas, N., Uren, V., & Roeck, A. D. (2003). *Human interaction: building and applying a concept hierarchy representation of a user profile* Paper presented at the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, Toronto, Canada.

Nunamaker, J. F., Chen, M., & Purdin, T. D. M. (1991). Systems development in information systems research. *Journal of Management Information Systems, 7*(3), 89-106.

O'Donovan, J., & Smyth, B. (2006). *Is trust robust?: an analysis of trust-based recommendation.* Paper presented at the 11th international conference on Intelligent user interfaces Sydney, Australia.

O'Hanlon, N. (1999). Off the shelf & onto the web: web search engines evolve to meet challenges. *Reference & User Services Quarterly, 38*(3), 247-251.

Oard, D. W., & Kim, J. (2001). *Modeling information content using observable behavior.* Paper presented at the 64 Annual Meeting of the American Society for Information Science and Technology, USA.

Pazzani, M., Muramatsu, J., & Billsus, D. (1996). *Syskill & Webert: identifying interesting web sites.* Paper presented at the National Conference on Artificial Intelligence Portland, Oregon, USA.

Pennock, D. M., & Horvitz, E. (1999). *Collaborative filtering by personality diagnosis: a hybrid memory and model-based approach.* Paper presented at the IJCAI Workshop on Machine Learning for Information Filtering, Stockholm, Sweden.

Pitkow, J., Schutze, H., Cass, T., Cooley, R., & et al. (2002). Personalized search. *Communications of the ACM, 45*(9), 50 - 55.

Pokorny, J. (2004). Web searching and information retrieval. *Computing in Science and Engineering (CiSE), 6,* 43-48.

Prekop, P., & Burnett, M. (2003). Activities, context and ubiquitous computing. *Computer Communications, 26*, 1168-1176.

Rad, T. E., & Shavlik, J. (2003). A System for Building Intelligent Agents that Learn to Retrieve and Extract Information. *User Modeling and User - Adapted Interaction, 13*(1-2), 35.

Radlinski, F., & Joachims, T. (2005). *Query chains: learning to rank from implicit feedback.* Paper presented at the 11th ACM SIGKDD international conference on Knowledge discovery in data mining, Chicago, Illinois, USA.

Rey, G., Coutaz, J., & Crowley, J. L. (2004). *The contextor: a computational model for contextual information*.

Rhodes, B. J., & Starner, T. (1996). *Remembrance Agent: a continuously running automated information retrieval system.* Paper presented at the Practical Application Of Intelligent Agents and Multi Agent Technology, London, UK.

Rieh, S. Y., & Xie, H. I. (2006). Analysis of multiple query reformulations on the web: the interactive information retrieval context. *Information processing & management, 42*(3), 751-768.

Rocchio, J. (1971). Relevance feedback in information retrieval. In G. Salton (Ed.), *The SMART Retrieval System: Experiments in Automatic Document Processing* (pp. 313-323): Prentice Hall.

Roush, W. (2004). Search beyond Google. *MIT Technology Review*. Retrieved 06 Aug 2007, from http://www.technologyreview.com/Infotech/13505/.

Ruthven, I. (2003). *Re-examining the potential effectiveness of interactive query expansion.* Paper presented at the 26th annual international ACM SIGIR conference on Research and development in information retrieval, Toronto, Canada.

Sahami, M., Mittal, V., Baluja, S., & Rowley, H. (2004). *The happy searcher: challenges in web information retrieval.* Paper presented at the 8th Pacific Rim International Conference on Artificial Intelligence, Auckland, New Zealand.

Salton, G., & Buckley, C. (1990). Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science, 41*(4), 288-297.

Sandra, E. (2000). *Towards understanding information encountering on the web.* Paper presented at the 63rd Annual Meeting of the American Society for Information Science, Chicago.

Saracevic, T. (1995). *Evaluation of evaluation in information retrieval.* Paper presented at the 18th annual international ACM SIGIR conference on Research and development in information retrieval, Seattle, Washington, USA.

Schilit, B., Adams, N., & Want, R. (1994). *Context-aware computing applications.* Paper presented at the Workshop on Mobile Computing Systems and Applications, Santa Cruz, CA, US.

Schreckenghost, D., Bonasso, P., Kortenkamp, D., & Ryan, D. (1998). *Three tier architecture for controlling space life support systems.* Paper presented at the IEEE International Joint Symposia on Intelligence and Systems, Washington DC, USA.

Schulman, J. (2006). *Managing your Patients' data in the Neonatal and Pediatric ICU: An Introduction to Databases and Statistical Analysis (Kindle Edition)* (First ed.): Wiley.

Sendhilkumar, S., & Geetha, T. V. (2008). *Personalized ontology for web search personalization.* Paper presented at the 1st Bangalore annual Compute conference, Bangalore, India.

Seo, Y.-W., & Zhang, B.-T. (2000). *Learning user's preferences by analyzing Web-browsing behaviors.* Paper presented at the 4th international conference on Autonomous agents, Barcelona, Spain.

Shafi, S. M., & Rather, R. A. (2005). Precision and recall of five search engines for retrieval of scholarly information in the field of biotechnology. *Webology, 2*(2).

Shen, X., Tan, B., & Zhai, C. (2005). *Implicit user modeling for personalized search.* Paper presented at the 14th ACM international conference on information and knowledge management, Bremen, Germany.

Sherman, C., & Price, G. (2001). *The invisible web: uncovering information sources search engines can't see*. USA: Cyber Age Books.

Shneiderman, B. (1998). *Designing the user interface: strategies for effective human-computer interaction.* (Third ed.). Boston, MA, USA: Addison Wesley.

Shneiderman, B., Byrd, D., & Croft, W. B. (1998). Sorting out searching: a user-interface framework for text searches. *Communications of the ACM, 41*(4), 95-98.

Shneiderman, B., & Plaisant, C. (2004). *Designing the user interface: strategies for effective human-computer interaction* (Fourth ed.). Boston, MA: Addison Wesley.

Sieg, A., Mobasher, B., Lytinen, S., & Burke, R. (2004, February). *Using concept hierarchies to enhance user queries in Web-based information retrieval.* Paper presented at the IASTED International Conference on Artificial Intelligence and Applications, Innsbruck, Austria.

Sihvonen, A., & Vakkari, P. (2004). *Subject Knowledge, Thesaurus-assisted Query Expansion and Search Success.* Paper presented at the Recherche d'Information Assistée par Ordinateur (RIAO), Paris, France.

Spink, A. (1997). Study of interactive feedback during mediated information retrieval. *Journal of the American Society for Information Science and Technology (JASIST), 48*(5), 382-394.

Spink, A., & Jansen, B. J. (2004). A study of Web search trends *Webology, 1*(2).

Spink, A., & Losee, R. M. (1996). Feedback in information retrieval. *Annual Review of Information Science and Technology (ARIST), 31*, 33-78.

Spink, A., & Ozmultu, H. C. (2002). Characteristics of question format web queries: an exploratory study. *Information processing & management, 38*(4), 453-471.

Spink, A., Ozmutlu, S., Ozmutlu, H. C., & Jansen, B. J. (2002). U.S. versus European web searching trends. *ACM SIGIR Forum, 36*, 32-38.

Spink, A., Wolfram, D., Jansen, B., & Saracevic, T. (2001). Searching of the web: the public and their queries. *Journal of the American Society for Information Science, 52*(3), 226-234.

Su, L. T. (1992). Evaluation measures for interactive information retrieval *Information processing & management, 28*(4), 503-516.

Sugiyama, K., Hatano, K., & Yoshikawa, M. (2004). *Adaptive web search based on user profile constructed without any effort from users.* Paper presented at the 13th international conference on World Wide Web, New York, USA.

Sun, J.-T., Zeng, H.-J., Liu, H., Lu, Y., & Chen, Z. (2005). *CubeSVD: a novel approach to personalized Web search.* Paper presented at the 14th international conference on World Wide Web Chiba, Japan.

Taksa, I. (2005, 4-6 April). *Predicting the Cumulative Effect of Multiple Query Formulations.* Paper presented at the International Symposium on Information Technology: Coding and Computing, Las Vegas, Nevada.

Talja, S., Keso, H., & Pietiläinen, T. (1999). The production of 'context' in information seeking research: a metatheoretical view. *Information processing & management, 35*(6), 751-763.

Tashakkori, A., & Teddlie, C. (2003). *Handbook of mixed methods in social & behavioral research*. CA, USA: Thousand Oaks, SAGE Publications.

Teevan, J., Dumais, S. T., & Horvitz, E. (2005). *Personalizing search via automated analysis of interests and activities.* Paper presented at the 28th annual international ACM SIGIR conference on Research and development in information retrieval, Salvador, Brazil.

Tešic, J., & Manjunath, B. S. (2003, 16-22). *Nearest neighbor search for relevance feedback.* Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Madison, Wisconsin.

Thomas, P., & Hawking, D. (2006). *Evaluation by comparing result sets in context.* Paper presented at the 15th ACM international conference on Information and knowledge management, Arlington, Virginia, USA.

Tirri, H. (2003). Search in vain, challenges for Internet search. *Computer, 36*(1), 115-116.

Tombros, A., Ruthven, I., & Jose, J. M. (2003). *Searchers' criteria For assessing web pages* Paper presented at the 26th annual international ACM SIGIR conference on Research and development in information retrieval, Toronto, Canada.

Toms, E. G., & Bartlett, J. C. (2001, June 24-28). *An approach to search for the digital library.* Paper presented at the 1st ACM/IEEE-CS joint conference on digital libraries, Roanoke, Virginia, USA.

Upstill, T., Craswell, N., & Hawking, D. (2003). Query-independent evidence in home page finding. *ACM Transactions on Information Systems (TOIS), 21*(3), 286 - 313.

Vakkari, P., Savolainen, R., & Dervin, B. (1997). *Information seeking in context* London: Taylor Graham.

Vinay, V., Wood, K., Milic-Frayling, N., & Cox, I. J. (2005). *Comparing relevance feedback algorithms for web search.* Paper presented at the 14th international conference on World Wide Web, Chiba, Japan.

Volokh, E. (2000). Personalization and privacy. *Communications of the ACM, 43*(8), 84-88.

Voorhees, E. M. (2002). The philosophy of information retrieval evaluation. In *Lecture Notes in Computer Science* (Vol. 2406, pp. 143-170): Springer Berlin / Heidelberg.

Wang, Y., Li, J., & Stoica, P. (2005). *Spectral Analysis of Signals: The Missing Data Case (Synthesis Lectures on Signal Processing)* Morgan & Claypool Publishers.

Wen, J.-R., Lao, N., & Ma, W.-Y. (2004). *Probabilistic model for contextual retrieval.* Paper presented at the Annual ACM Conference on Research and Development in Information Retrieval, UK.

Wen, J.-R., Nie, J.-Y., & Zhang, H.-J. (2001). *Clustering user queries of a search engine.* Paper presented at the 10th international conference on World Wide Web Hong Kong, Hong Kong.

White, R. W., Bilenko, M., & Cucerzan, S. (2007). *Studying the use of popular destinations to enhance web search interaction.* Paper presented at the 30th annual international ACM SIGIR conference on Research and development in information retrieval Amsterdam, The Netherlands.

White, R. W., Jose, J. M., & Ruthven, I. (2003). A task-oriented study on the influencing effects of query-biased summarisation in web searching. *Information processing & management, 39*(5), 707-733.

White, R. W., & Kelly, D. (2006). *A study on the effects of personalization and task information on implicit feedback performance.* Paper presented at the 15th ACM international conference on Information and knowledge management, Arlington, Virginia, USA.

White, R. W., & Marchionini, G. (2007). Examining the effectiveness of real-time query expansion. *Information processing & management, 43*(3), 685-704.

White, R. W., Ruthven, I., & Jose, J. M. (2002). *Finding relevant documents using top ranking sentences: an evaluation of two alternative schemes.* Paper presented at the 25th annual international ACM SIGIR conference on Research and development in information retrieval, Tampere, Finland.

White, R. W., Ruthven, I., & Jose, J. M. (2005). *A study of factors affecting the utility of implicit relevance feedback.* Paper presented at the 28th annual international ACM SIGIR conference on Research and development in information retrieval Salvador, Brazil.

*Why MySQL?* (2008). Retrieved 14/05, 2008, from http://www.mysql.com/why-mysql/.

Wilson, T. D. (1997). Information behaviour: An interdisciplinary perspective. *Information processing & management, 33*(4), 551-572.

Wilson, T. D. (1999a). Exploring models of information behaviour: the 'uncertainty' project. *Information processing & management, 35*(6), 839-849.

Wilson, T. D. (1999b). Models in information behaviour research *Journal of Documentation, 55*(3), 249-270.

Wilson, T. D. (2000). Human information behaviour. *Informing Science, 3*(2), 49-56.

Xu, J., & Croft, W. B. (1996). *Query expansion using local and global document analysis.* Paper presented at the 19th annual international ACM SIGIR conference on Research and development in information retrieval Zurich, Switzerland.

Xu, J., & Croft, W. B. (2000). Improving the effectiveness of information retrieval with local context analysis. *ACM Transactions on Information Systems (TOIS), 18*(1), 79-112.

Yates, R. B., & Neto, B. R. (1999). *Modern information retrieval*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

Zhang, B.-T., & Seo, Y.-W. (2001). Personalized Web document filtering using reinforcement learning. *Applied Artificial Intelligence, 15*(7), 665-685.

Zhang, C., Chai, J. Y., & Jin, R. (2005). *User term feedback in interactive text-based image retrieval.* Paper presented at the 28th annual international ACM SIGIR conference on Research and development in information retrieval, Salvador, Brazil.

Zhang, W., Xu, B., & Yang, H. (2001). *Development of a self-adaptive Web search engine.* Paper presented at the 3rd International Workshop on Web Site Evolution, Florence, Italy.