Improving Web Information Retrieval Using Shared Contexts

Andy M. Connor[a,*], Dilip K. Limbu[b], Stephen G. MacDonell[a], Russel Pears[a]

[a]Software Engineering Research Lab, Auckland University of Technology, Auckland, New Zealand
[b]Institute for Infocomm Researchr, Agency for Science, Technology and Research, Singapore

Abstract– **The effective utilisation of a user's context in improving the performance of web search engines is a subject of intense research interest. In particular, much attention has been directed to the enhancement of queries and the provision of more relevant information by taking user context into account. Progress in this field has been limited to date, however, due to ongoing challenges in capturing and representing contextual information. We describe here the development and evaluation of a web-based contextual information retrieval that addresses some of these challenges and makes progress in defining the information required to create contextual profiles. Our system collects and leverages implicit and explicit user data to modify queries with the aim of more accurately reflecting the user's interests. This data is maintained dynamically in each user's contextual profile and utilised to improve the quality of information found during web searches. Where enabled, this data also contributes to the development of a shared contextual knowledge base that can also be used to augment queries. This shared contextual knowledge base is a key aspect of this research. The system has been tested in an observational study that has considered its ability to improve the user's web search experience. This paper presents experimental data to provide evidence of the system's performance, demonstrating that the shared contextual knowledge base extends the functionality associated with the individual contextual profile.**

Keyword: Personalized Web Search, Contextual Search, Relevance Feedback, Query Formulation.

1. Introduction

Mankind has organised information for hundreds of years in order to make it more accessible to others. With the advent of information technology, the process of information retrieval has evolved drastically. Information retrieval is therefore a wide, often loosely defined field. In brief, it involves finding some desired information in a store of information or a database [1]. Web information retrieval is a subset of information retrieval that uses search engines or web directories to facilitate the identification of relevant information. Due to the exponential growth of the WWW, identifying user information needs has been highlighted as one of the most fundamental challenges in the development of Web search engines [2]. In the late 1990s, research began to focus on information retrieval in a given context in an effort to address

*Corresponding author:
Email address: andrew.connor@aut.ac.nz, Ph: +64 99219999

those challenges. Contextual information retrieval has distinct characteristics when compared with other approaches to organizing and retrieving information. This paper describes the development of a novel contextual information retrieval approach that has been tested in a user study involving 30 subjects. The following section provide a brief overview of prior research related to Web search engines in order to highlight the opportunities that are evident in terms of improving query results from the WWW by incorporating an understanding of user context. The remainder of this paper outlines a contextual information retrieval system that has been developed to address some of the challenges encountered in effectively retrieving information from the WWW. A complete description of the system is outside the scope of this paper, so a brief overview of the system is given, along with a descrpition of the the experimental approach used to evaluate the system and some of the results obtained.

2. Searching the WWW

Since its inception, the WWW has continuously grown into one of the largest collections of content in existence. Earlier research reports on the growth of information on the WWW, which is continuing to expand at exponential rates. Each day approximately 60 terabytes of new content is added to the Web's 10 billion or so indexed pages [3]. Given these numbers, it is clear that the complexity of finding relevant information in the Web is increasing rapidly. In fact, "information overload" on the Web is a well recognised problem [4]. Search engines have therefore become an indispensable tool for Web users [5]. Within a few short years, search engines have become part of our daily lives and are a commonly used resource employed to find relevant information on the Web [6, 7].

Current Web search engines are attempting to deal with Internet "information overload" challenges [8]. For instance 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. These features are meant to assist users in finding more relevant information. Search engines have evolved through several generations since their inception and the quality of search has improved dramatically [9]. However, as useful as they are, they are far from perfect. In actual fact, these search engines are faced

with a number of difficult challenges in maintaining and enhancing the quality of their performance [10].

Up until now, search engines have generally used simple user interfaces that provide little support for user interaction during the information seeking process. A common search engine's interface consists of a single text field into which search terms can be entered, and a "search button" that when selected, begins the searching process. Further, the textual content and list based presentation of returned pages can make it difficult for the user to efficiently evaluate the search results [11].

In order to have general appeal, today's search engines are designed in a "one size fits all" manner. The downside of this approach is that most do not provide desired search results that are tailored to any specific individual user [12]. The precision and relevance of search engine results are largely dependent on how the user specified/formulated the search query [13]. Other research has confirmed, however, that queries submitted to search engines by users are short [14, 15] - most are limited to fewer than three key words [16]] - and can be vague with little or no context information provided [14]. Additionally, search engine results are based on simple keyword matches without any concern for the information needs of the user at a particular instance in time [17] 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 examines the indexed Web pages, then filters and returns a list of those documents that contain the specified keyword (i.e. surfing). However, the keyword "surfing", could have completely different meanings depending on the context in which it is used. As a result, the user still must perform most of the relevance filtering [18]. These problems are due to synonymy and polysemy of keywords. Synonymy is when several different terms have the same meaning and polysemy is when a single word has more than one meaning leading to potential ambiguity. The user can include additional search terms that could help to refine the search queries, but it is difficult to select the optimum query terms so that the desired subset of information is retrieved [15]. Some search engines, such as Google or Yahoo, provide a hierarchy of categories to help users to define their search intent. Unfortunately, evidence suggests that 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 [19]. As a result, even the most experienced users have difficulties in finding relevant information from the Web [20].

The need to better target a search to satisfy a user's information needs is well recognised [15]. A critical goal of successful Information Retrieval from the Web is to identify which pages are most relevant to a user's query [6]. However, relevance is typically person-dependent, so personalisation is critical to ongoing search engine development [18]. Many current search engines lack a personalisation mechanism [21, 22] 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. Such results will only be forthcoming if search engines can leverage the user's contextual information, such as the user's behaviour and preferences, to better understand and respond to the underlying intent of the user [5]. In summary, it is clear that today's search engines are faced with a number of difficult challenges, related to the user's information-seeking behaviour. These challenges relate to problems concerned with query formulation, with the user's understanding of the task and with the system's "understanding" of how the user performs that task [11]. Hence, to provide the desired information to the user requires effective methods for identifying the user's task context [23] 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 [5], especially in respect of understanding the underlying intent of the search.

3. Contextual Search System

In order to improve upon a users web searching experience, and improve the quality of returned results, a contextual search system has been developed that is based around the concept of a contextual profile. The contextual search system 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 a suitable and appropriate Boolean query and presentation of ranked contextual search results to improve web information retrieval. Several important architectural design and implementation issues, such as scalability, flexibility, performance and robustness, were maximised during the system's development. Figure 1 shows a high level view of the system architecture. The contextual search layer is the core of the three-tier architecture and it comprises two main modules: Profile Collector Module (PCM) and Context Manager Module (CMM) to perform the following functions:

1. Gather the user's implicit data, such as entered search queries, visited URLs and corresponding extracted meta keywords.
2. 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.
3. Construct the user's personal contextual profile and a shared contextual knowledge base using data from step 1 and step 2.
4. 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 actions.

3.1. Profile Collector Module (PCM)

The PCM is implemented to capture 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. Many Web IR systems have explored various user modelling approaches to address similar objectives. Figure 2 illustrates the functionality of the PCM, a hybrid contextual user profiling approach that captures a user's adaptive search behaviour by monitoring and capturing their explicit (i.e., explicit rankings, inputs, and instructions) and implicit (i.e., browsing and typing) data. The PCM constantly acquires and maintains these data with minimal intervention to represent
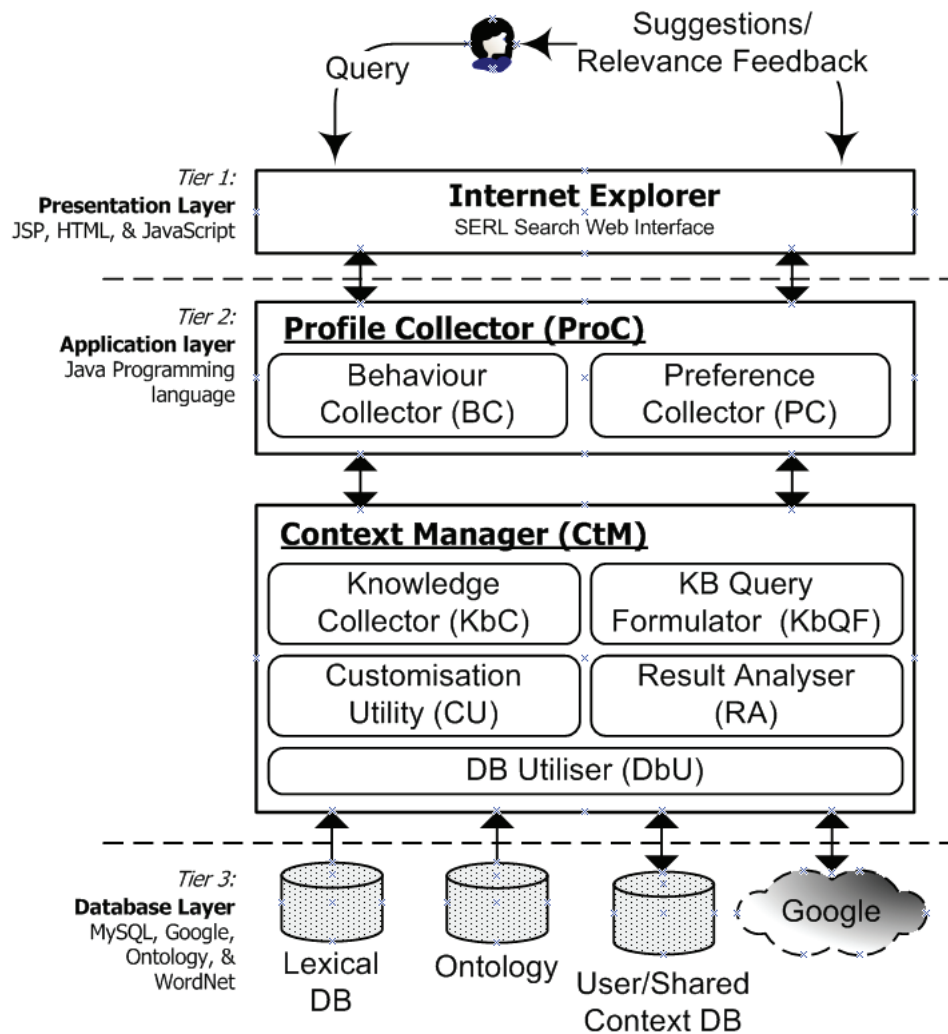
Fig. 1. Contextual search system architecture

accurate information about the user's multiple interests. The PCM consists of two specialised components: Preference Collector (PC) and Behaviour Collector (BC) as shown in Figure 2. Both of these components gather information seeking behaviour from users of the system. The functionality of these components, and the details of assumptions made during their development, are detailed in previous publications [24, 25].

3.2. Context Manager Module (CMM)

The CMM is implemented to facilitate the collection of multiple users' personal contextual profiles, to use the 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. The CMM performs two main functions. Firstly, it interacts with the PCM to build a user's personal contextual profile and a shared contextual knowledge base. Secondly, it performs iterative query expansion using the PCM's relevance feedback function. The CMM consists of five components, as seen in Figure 1. Of these components, three are functional components and two are utility components. The following sections describe the three main functional components in more detail.

3.2.1. Knowledge Collector (KbC)

The functionality of the KbC component is displayed in Figure 3. The KbC facilitates the construction of a personal user profile and collection of multiple users' personal contextual profiles to build the shared contextual knowledge base. The KbC interacts with the PCM to gather user behaviour and preferences data which are stored in the personal contextual user profile and which update the shared contextual knowledge base. 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 are potentially used to suggest or recommend disambiguated terms, meta keywords, ontology and concepts related to the current context.

3.2.2. Knowledge Base Query Formulator (KbQF)

The functionality of the KbQF component is displayed in Figure 4. The KbQF expands simple keyword queries into a contextual Boolean query in order to improve the returned results [26]. The KbQF component interacts with the PCM to formulate a contextual Boolean query. It addresses query expansion challenges by employing interactive query expansion (IQE) and automatic
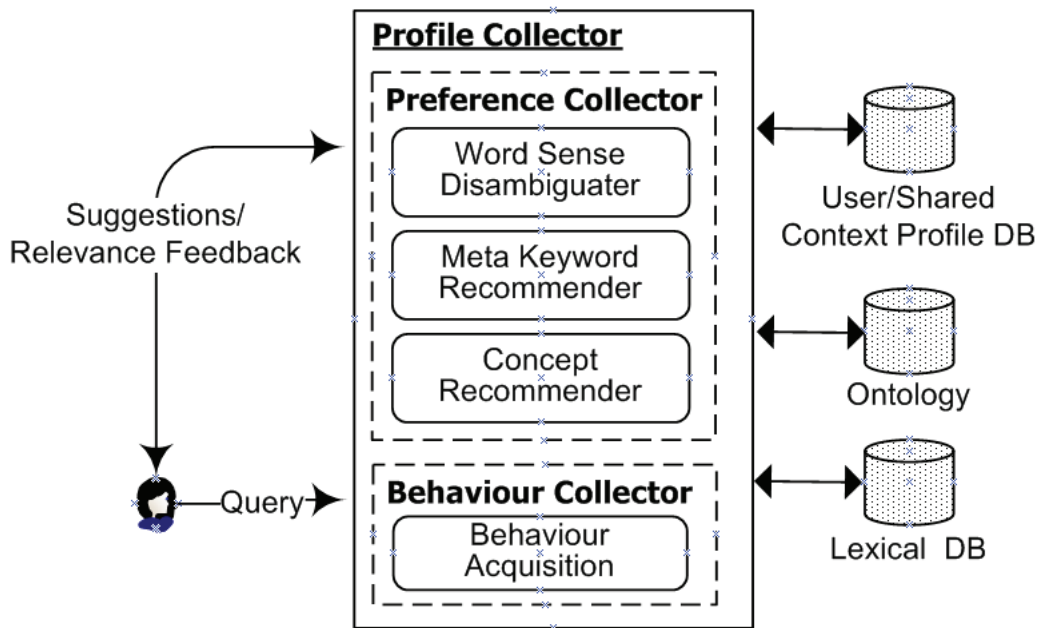
Fig. 2. Profile collector module functionality

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). Finally, the KbQF component uses these expansion terms to formulate a Boolean search query for submission to a search engine.

3.3. Result Analyser (RA)

The RA component interfaces with the Google SOAP Search API. The RA goes beyond providing or presenting search results from Google by performing an on-the-fly analysis and ranking the results based on a user's contextual profile and a shared contextual knowledge base. Figure 5 shows the RA's analysis and ranking process. The RA extracts URLs from the Google results, checks if the URL exists in the user's contextual profile or in the shared contextual knowledge base, and then returns the nunmber of hits in each. In this way the user is informed as to how many times the URL has been visited either by them or by other users with similar search profiles. The rationale behind this system is that the determination of context, the key behind resolving ambiguity in user intent, is too challenging to be resolved by a user's implicit actions alone. We believe that explicit user input is also needed to capture and refine the required information.

3.4. Contextual Profiles

A key aspect of this work is the use of User Contextual Profiles (UCP) that are used to capture an individuals context. This profiles may, at the users discretion, but shared with other users to create a shared contextual knowledge base (SCKB). The primary objective of this paper is to compare performance along two dimensions between having shared or non-shared profiles. This differs from previous work [26] that focused on comparing performance against a contemporary search engine. A User Contextual

Table 1. Example of users preference data

Query	Context	Date	Disambiguated Terms	Metakeywords	Concepts
q0	ct1	dt1	d0, d1	m1	c1
q1	ct1	dt1	d3	m2	c2
q2	ct1	dt1	d4	m3	c3

Profile consists of data collected by the PCM, being a range of explicit preferences as well as implicitly captured behaviour. Each user has the option to specify multiple contexts (e.g. "work", "home" etc) and for each context a typical profile consists of the data shown in Table 1 (preferences) and Table 2 (behaviour). The system employs the nearest-neighbour algorithm to learn a users specific information needs and to provide alternative terms recommendation. Firstly, the algorithm uses the number of hits parameter together with other computation parameters, such as search query (q0), users context (ct1) and disambiguated terms $d1$, $d2$/Meta keywords or phrases $m1, m2$/concepts $c1, c2$, to cluster a neighbourhood of users that in the pasthave 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.).

The preferences data is stored in the profile by query, and for each query made consists of basic data such as the chosen context and the date of the query. In addition, the profile stores any disambiguated terms, previous metakeywords and ontological concepts that the user selects to apply to their search. The word disamiguation and concept recommender process have been described in previous work [26, 27]. Similarly, the behavioural data is associated with a particular query and consists of the URLs themselves and the Metakeywords from those visited URLs. This data is automatically extracted with no user interaction. Using
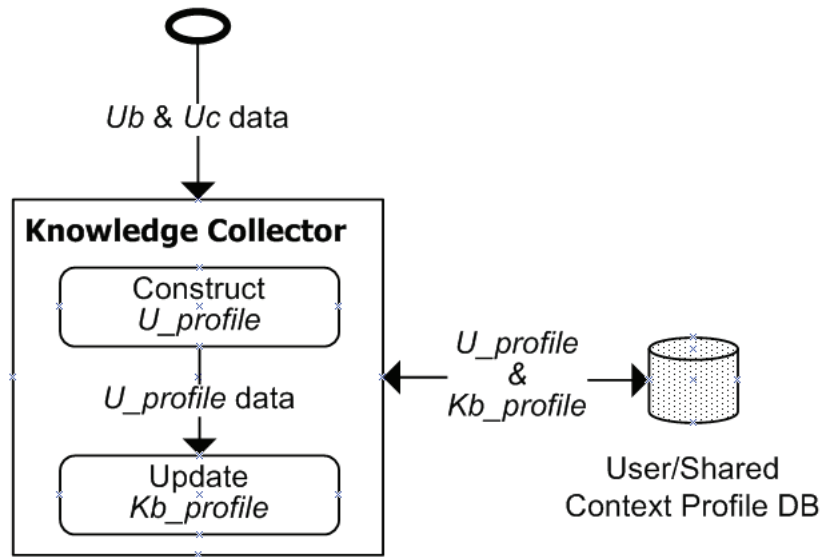
Fig. 3. Knowledge collector functionality

Table 2. Example of users behaviour data

Query	Context	Date	Visited URLs	Metakeywords
q0	ct1	dt1	u1	e1
q0	ct1	dt1	u2	e2
q1	ct1	dt1	u3	e3
q1	ct2	dt2	u4	e4
q1	ct2	dt2	u5	e5

the system described in this paper, users have the option to share their profile with other users and hence gain access to other data. This option has the potential to provide more effective searching when users have similar contexts. A key aim of this work is to attempt to discover to what extent a shared profile impacts the effectiveness of web information retrieval.

4. Evaluation Methodology

In order to assess our system we designed a research approach comprising simulated work task situations, questionnaires, and observations. Full details of the evaluation methodology are available in previous work [27] so only a brief summary is provided here.

A qualitative study was undertaken to comprehensively investigate subjects' information seeking behaviour for specific scenarios. A quantitative study was conducted concurrently to determine the performance of the system along the usability dimensions of effectiveness and efficiency. In addition, we also collected data (via questionnaires) that reflected users' subjective satisfaction with the returned results. The first experimental phase consisted of testing individual system components and verifying that each contributed to improving web retrieval using quantitative measures.

During the second phase of evaluation, three experiments; II(a), II(b), and II(c) were carried out. The aim of experiment

II(a) was to determine whether the contextual retrieval system enabled subjects to find relevant information when compared to a standard search engine using their personal contextual profiles. In the II(a) experiment subjects performed six search tasks using the system, and had their search behaviours and preferences captured in order to create their personal contextual profiles as well as providing input data for the shared contextual knowledge base. However, this shared resource was not accessible to them during their search.

Once the IIa experiment was completed, a second group of subjects undertook the same six search tasks; however subjects had the shared contextual knowledge base enabled. As such, the aim of II(b) experiment was to determine whether the system enabled subjects to find relevant information more readily than a standard search engine using their personal contextual profiles and the shared contextual knowledge base. This allowed us to assess the ability of the shared profile to improve search efficiency, by comparing the "speed" of finding data with the first group who did not have access to the shared profile. In the II(c) experiment subjects carried out the six search tasks using Google to provide reference results. These results are not covered in this paper, which focuses on the II(a) and II(b) phases.

A total of 30 subjects, with different levels of search experience, participated in the three experiments. Subjects were randomly assigned to one of the experiments, so that there were ten subjects in each, with each group having an equal distribution of experience levels. Before the actual user experiment, subjects were given the same general instructions, watched a video demonstration of the system in use, and filled in an entry questionnaire containing information about their characteristics and search experience. Subjects then attempted the six search tasks, a technique similar to other studies in this area [13]. Thereafter, subjects filled in a post observation questionnaire so that we could learn their overall reactions to the experimental systems.
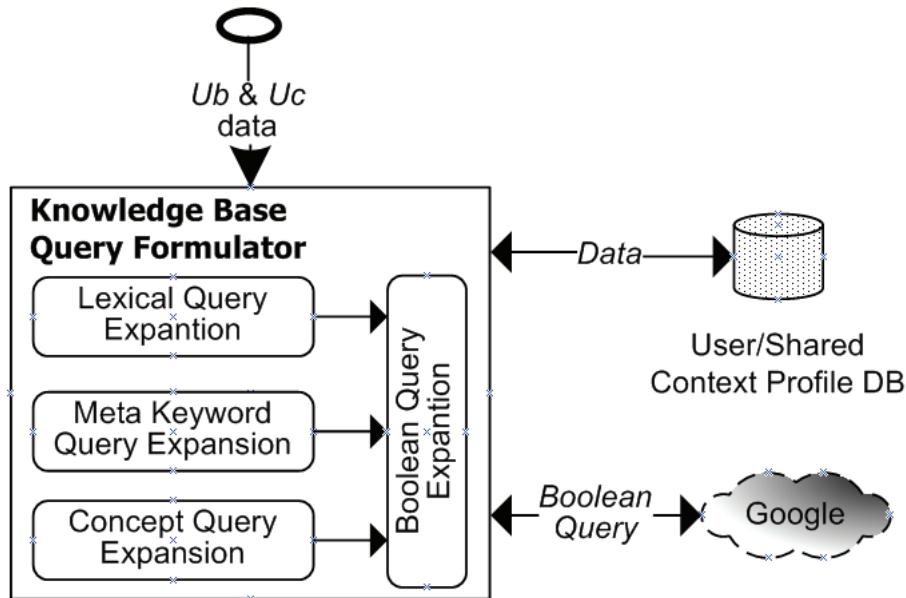
Fig. 4. Knowledge base query formulation functionality

5. Hypotheses and Results

This research investigated the performance of the contextual retrieval system in terms of how it improves web information retrieval. A total of five hypotheses were tested that dealt with the ability to find information readily, the adaptiveness of the system, the level of recommendation, the ability to reformulate queries effectively and the quality of the user interface. These hypotheses examined the subjects' overall information seeking behaviour and their perceptions of the contextual search system. Previous work [25] has reported on the quantative data analysis showing the relative effectiveness of the approach when compared with a contemporary search engine, and so this paper focuses on two further hypotheses that relate to the quantative data gathered during the evaluation with a focus on subjective impressions of the system.

5.1. Hypothesis 1 - Recommendation Support

The results presented here relate to the experimental hypothesis: *The contextual search system eases the process in which users convey their preferences and recommends relevant and useful terms.* This hypothesis is further refined into two sub-hypotheses that measure the satisfaction aspects of the experimental system.

5.1.1. Sub-Hypothesis 1.1 - Recommendation Strategy

Subjects find that the experimental system communicates its recommendations clearly, in a timely and in an unobtrusive manner. This sub-hypothesis was addressed in the post-observation questionnaire through a number of questions. Subjects were asked to indicate their overall reactions to the contextual search system with regard to whether the system communicated its recommendations on three semantic differentials; 'obtrusive'/'unobtrusive', 'uninformative'/'informative', and 'untimely'/'timely'. In addition, subjects were asked to complete

a five point Likert scale (range 1-5, higher = better) on the clarity of the content of recommendation terms.

5.1.2. Sub-Hypothesis 1.2 - Conveying Preferences

Subjects find 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 search system in terms of conveying their preferences, on five semantic differentials; 'difficult'/'easy', 'ineffective'/'effective', 'not useful'/'useful', 'uncomfortable'/'comfortable', and 'not in control'/'in control'.

5.2. Hypothesis 2 - Interface Support

This section presents results related to the experimental hypothesis: *The interface support provided by the contextual 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.

5.2.1. Sub-Hypothesis 2.1 - Relevance of Content

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'.

5.2.2. Sub-Hypothesis 2.2 - Interface Guide

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.
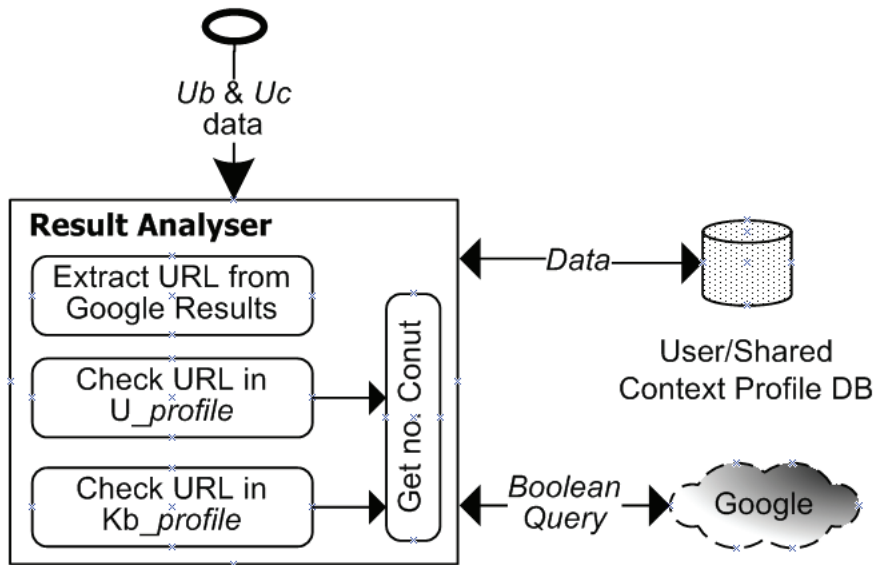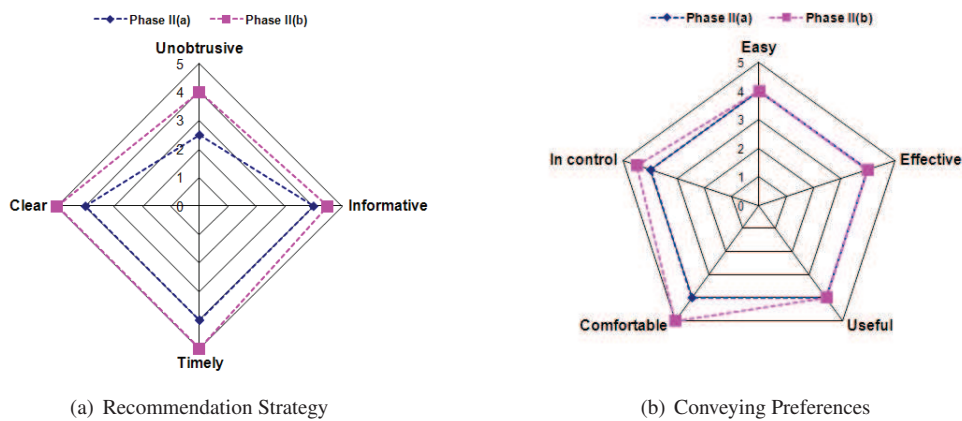
Fig. 5. Result analyser process

(a) Recommendation Strategy   (b) Conveying Preferences

Fig. 6. Usability comparison on median scores
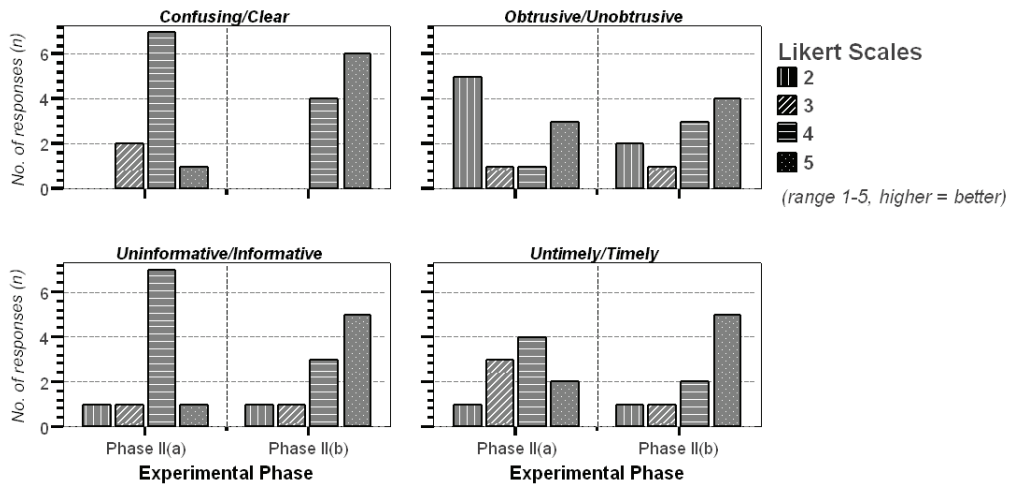
5.3. Results

5.3.1. Recommendation Support

Figure 6 shows the graphical representation of the median value of responses for the various questions relating to the two sub-hypotheses related to the recommendation strategy (Figure 6a) and conveying of preferences (Figure 6b).

Presentation of results in this way allows an easy comparison between the two experiments along the multiple dimensions measured by the questionairre. Figure 6 shows that the subjects in experiment II(a) were generally less satisfied with how the system communicated its recommendations than those in experiment II(b), who were utilizing the shared contextual profile. Similarly, there is a generally higher level of satisfaction in subjects in the II(b) experiment in terms of conveying preferences. The obvious exception is the result relating to users' perceptions of usefulness. For this question, the median response was the same for both experiments, but it is important to point out that for both cases the overall impression was positive. Overall, there is a clear

difference between the two experiments, showing that the subjects rated the use of a shared contextual knowledge base more highly than an individual contextual profile alone. This simple analysis based on median responses can be given further visibility by plotting the responses as bar charts, with distinction made between the respondents from the II(a) and II(b) experiments. These results are shown in Figure 7.

For the first sub-hypothesis (Figure 7a), the bar charts clearly illustrate that higher numbers of subjects using the shared contextual knowledge base chose the semantic differential and the Likert scale value of 'five' (clear n = 6, informative n = 5, timely n = 5 and unobtrusive n = 4) in comparison to the experiment II(a) subjects (clear n = 1, informative n = 1, timely n = 2 and unobtrusive n = 3). For the second sub-hypothesis (Figure 7b), the bar charts show that more than half of the subjects in the II(b) experiment - the shared contextual knowledge base - chose the semantic differential value of 'five' (comfortable n = 6, easy n = 3, effective n = 2, in control n = 5, and useful n = 3) in compari-

(a) Recommendation Strategy
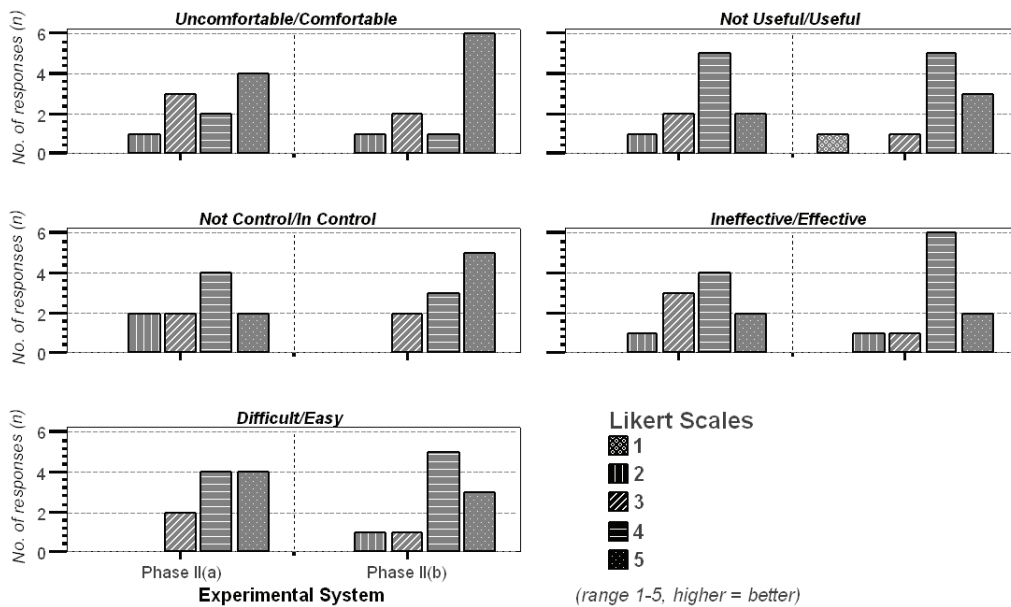
(b) Convey Preferences

Fig. 7. Usability comparison on response volume

son to the II(a) subjects (comfortable n = 4, easy n = 4, effective n = 2, in control n = 2, and useful n = 2).

5.3.2. Interface Support

Figure 8 shows the graphical representation of subjects' responses for the two different experimental system for the various questions relating to the two sub-hypotheses on the ability of the interface to guide the users to relevant information. For this hypothesis, analysis of median results is not included as the results for both experimental systems showed the same outcomes.

Figure 8a illustrates 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 8b 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 experimental system users (n = 1).

6. Conclusion

This paper has presented research regarding the implementation and evaluation of a contextual retrieval system. The system utilises a contextual user profile employing both implicit and explicit data to provide relevant information to users that potentially satisfies their information needs. In our system, this data can be stored as an individual contextual profile and each profile may be shared with other users through a shared contextual knowledge base. The research reported in this paper has has focused on recommendation support. The system incorporates two levels

(a) Relevance of Content
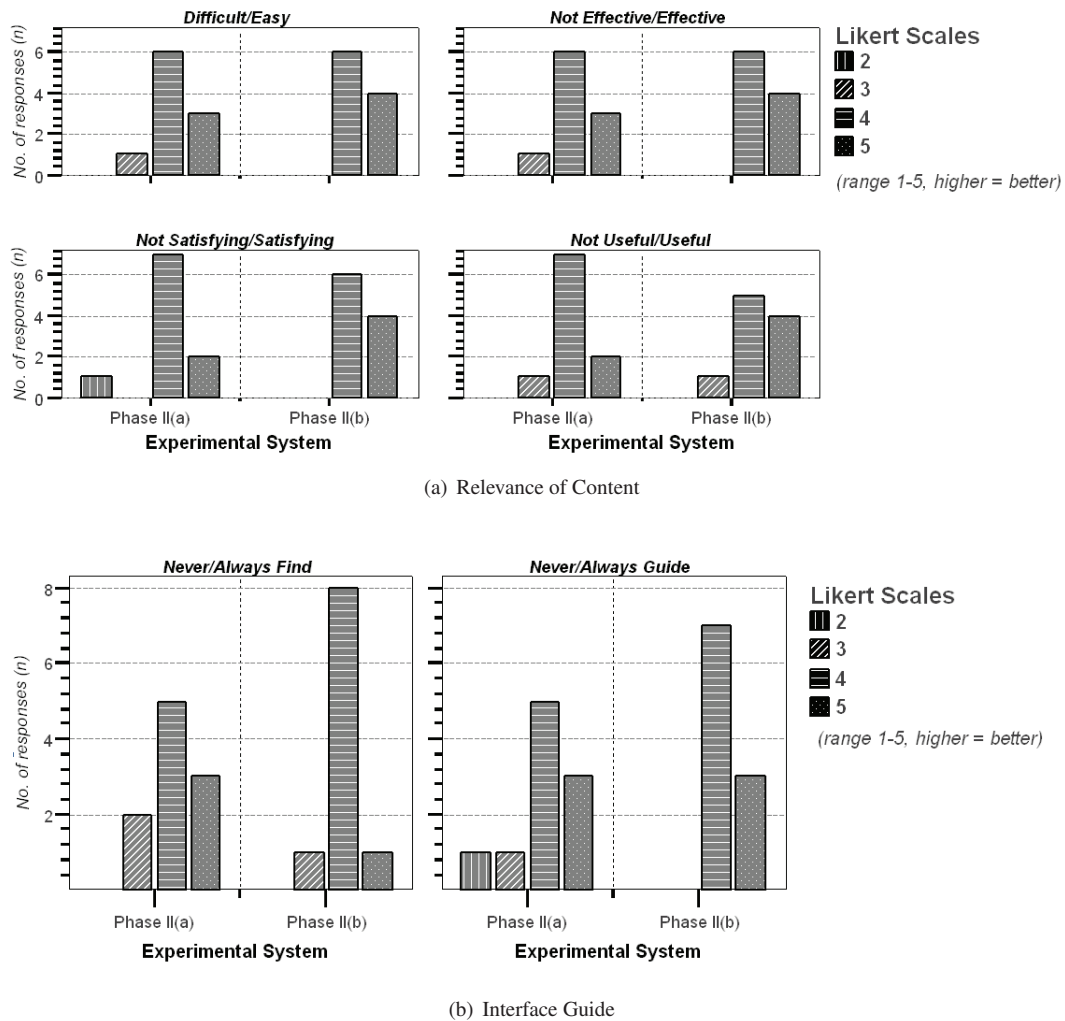

(b) Interface Guide

Fig. 8. Usability comparison on median scores

of recommendation support, namely the suggestion of alternative terms and concepts that can be used to refine a query and the recommendation of relevant pages previously visited by the current or other users when the shared contextual knowledge base is enabled. An observational study has been carried out and analysis of the data indicates that higher user satisfaction is perhaps achieved using this shared contextual knowledge base than when the system is limited to using individual contextual profiles. Further data analysis is required to validate all aspects of the system performance and highlight avenues for future research.

References

[1] C. T. Meadow, B. R. Boyce, and D. H. Kraft, *Text Information Retrieval*. Academic Press, 2000.

[2] C. K. Huang, L. F. Chien, and Y. J. Oyang, "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*, vol. 54, no. 7, pp. 638–649, 2003.

[3] W. Roush, "Search beyond google," *MIT Technology Review*, vol. 107, no. 1, pp. 34–45, 2004.

[4] C. N. Carlson, "Information overload, retrieval strategies and internet user empowerment," in *The Good, the Bad and the Irrelevant (COST 269)*, pp. 169–173, 2003.

[5] B. J. Jansen, D. L. Booth, and A. Spink, "Determining the user intent of web search engine queries," in *WWW '07: Proceedings of the 16th international conference on World Wide Web*, (New York, NY, USA), pp. 1149–1150, ACM, 2007.

[6] M. Sahami, V. Mittal, S. Baluja, and H. Rowley, "The happy searcher: Challenges in web information retrieval," in *Proceedings of the 8th Pacific Rim International Conference on Artificial Intelligence*, 2004.

[7] E. G. Toms and J. C. Bartlett, "An approach to search for the digital library," in *JCDL '01: Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries*, (New York, NY, USA), pp. 341–342, ACM, 2001.

[8] H. Bergel, "Cyberspace 2000: Dealing with information overload," *Communications of the ACM*, vol. 40, no. 2, pp. 19–24, 1997.

[9] L. Finkelstein, E. Gabrilovich, Y. Matias, E. Rivelin, Z. Solan, G. Wolfman, and E. Ruppin, "Placing search in context: The concept revisited," *ACM Transactions on Information Systems*, vol. 20, no. 1, pp. 116–131, 2002.

[10] M. R. Henzinger, R. Motwani, and C. Silverstein, "Challenges in web search engines," *SIGIR Forum*, vol. 36, no. 2, pp. 11–22, 2002.

[11] O. Hoeber and X. D. Yang, "A comparative user study of web search interfaces: Hotmap, concept highlighter, and google," in *WI '06: Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, (Washington, DC, USA), pp. 866–874, IEEE Computer Society, 2006.

[12] J. Allan, J. Aslam, N. Belkin, C. Buckley, J. Callan, B. Croft, S. Dumais, N. Fuhr, D. Harman, D. J. Harper, D. Hiemstra, T. Hofmann, E. Hovy, W. Kraaij, J. Lafferty, V. Lavrenko, D. Lewis, L. Liddy, R. Manmatha, A. McCallum, J. Ponte, J. Prager, D. Radev, P. Resnik, S. Robertson, R. Rosenfeld, S. Roukos, M. Sanderson, R. Schwartz, A. Singhal, A. Smeaton, H. Turtle, E. Voorhees, R. Weischedel, J. Xu, and C.-X. Zhai, "Challenges in information retrieval and language modeling: Report of a workshop held at the center for intelligent information retrieval, university of massachusetts amherst, september 2002," *SIGIR Forum*, vol. 37, no. 1, pp. 31–47, 2003.

[13] I. Taksa, "Predicting the cumulative effect of multiple query formulations," in *Proceedings of the International Symposium on Information Technology: Coding and Computing*, (Las Vegas, Nevada), 2005.

[14] B. M. Fonseca, P. Golgher, B. Pôssas, B. Ribeiro-Neto, and N. Ziviani, "Concept-based interactive query expansion," in *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, (New York, NY, USA), pp. 696–703, ACM, 2005.

[15] D. B. Leake and R. Scherle, "Towards context-based search engine selection," in *IUI '01: Proceedings of the 6th international conference on Intelligent user interfaces*, (New York, NY, USA), pp. 109–112, ACM, 2001.

[16] B. J. Jansen and A. Spink, "An analysis of web information seeking and use: Documents retrieved versus documents viewed," in *Proceedings of the 4th International Conference on Internet Computing*, 2003.

[17] K. Keenoy and M. Levene, *Intelligent Techniques for Web Personalisation*, vol. 3169 of *Lecture Notes in Computer Science*, ch. Personalisation of Web Search. 2005.

[18] H. Tirri, "Search in vain: Challenges for internet search," *Computer*, vol. 36, no. 1, pp. 115–116, 2003.

[19] F. Liu, C. Yu, and W. Meng, "Personalized web search for improving retrieval effectiveness," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 16, pp. 28–40, Jan. 2004.

[20] A. Sieg, B. Mobasher, S. Lytinen, and R. Burke, "Using concept hierarchies to enhance user queries in web-based information retrieval," in *Proceedings of the IASTED International Conference on Artificial Intelligence and Applications*, 2004.

[21] M. Levene, *An Introduction to Search Engines and Web Navigation*. Addison Wesley, 2006.

[22] X. Shen, B. Tan, and C.-X. Zhai, "Implicit user modeling for personalized search," in *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, (New York, NY, USA), pp. 824–831, ACM, 2005.

[23] T. Bauer and D. B. Leake, "Wordsieve: A method for real-time context extraction," in *Proceedings of the Third International and Interdisciplinary Conference on Modeling and Using Context*, 2001.

[24] D. K. Limbu, A. M. Connor, R. Pears, and S. G. MacDonell, "Contextual relevance feedback in web information retrieval," in *Proceedings of the 1st International Symposium on Information Interaction in Context*, 2006.

[25] D. K. Limbu, A. M. Connor, R. Pears, and S. G. MacDonell, "Improving web search using contextual retrieval," in *Proceedings of the Third International Conference on Information Technology: New Generations*, pp. 1329–1334, 2009.

[26] D. K. Limbu, A. M. Connor, R. Pears, and S. G. MacDonell, "Contextual and concept-based interactive query expansion," in *Proceedings of the 19th Annual Conference of the National Advisory Committee on Computing Qualifications*, 2006.

[27] D. K. Limbu, *Contextual Information Retrieval from the WWW*. PhD thesis, Auckland University of Technology, 2009.

Andy Connor is a Senior Lecturer and Associate Director of the Software Engineering Research Laboratory (SERL) at Auckland University of Technology (AUT) in New Zealand. He holds a BEng degree from the University of Manchester and a PhD from Liverpool John Moores University. His primary research interests are in the application of meta-heuristic search techniques in the solution of complex problems and held postdoctoral positions at the Engineering Design Centres at both the University of Bath and the University of Cambridge prior to joining AUT in 2003.

Dilip Limbu received the B.Sc. degree in Computer Science and M.Sc. degree in Information Technology for Manufacture, from Coventry University, UK, in 1997 and 1998, respectively. He received his Ph.D. degree from School of Computing and Mathematical Sciences, Auckland University of Technology (AUT), New Zealand, in 2009. He is currently a Research Fellow in the Department of Human Language Technology, Institute for Infocomm Research (I^2R), Singapore. Prior to this, he was a research engineer in Singapore Institute of Manufacturing Technology (SIMTech), Singapore for four years. His research interests are Web information retrieval, personalisation/user modelling, Web mining/data mining, human-computer/robot-interaction, and robotic intelligence.

Stephen MacDonell is Professor of Software Engineering and Director of the Software Engineering Research Laboratory (SERL) at the Auckland University of Technology (AUT) in New Zealand. During his time at AUT Stephen has also held the roles of Head of the School of Information Technology and Associate Dean (Development). Stephen was awarded BCom (Hons) and MCom degrees from the University of Otago and a PhD from the University of Cambridge. He undertakes research in software metrics and measurement, project planning, estimation and management, software forensics, and the application of empirical analysis methods to software engineering data sets. He is a Member of the IEEE Computer Society and the ACM, and serves on the Editorial Board of Information and Software Technology.

Russel Pears (PhD, Cardiff University UK) is a Senior Lecturer and Programme Leader of the Doctoral Programme at the School of Computing and Mathematical Sciences at the Auckland University of Technology, New Zealand. Currently he is involved in a number of research projects in Data Mining, including mining of classification and association rules in high speed data streams, mining rare association rules, financial time series modelling and prediction, and recommender systems for on line book stores. He teaches in the Data Mining and Data Warehousing areas for undergraduate and postgraduate degrees in Computer Science at the Auckland University of Technology. He has published in a number of refereed international conferences and journals in the areas of Data Mining, Database Management Systems and Data Warehousing.