

Evaluation of a Personal Information Agent derived from the Context Modelling of Evolving Information Needs

Desmond Elliott
Department of Computing Science
University of Glasgow
Sir Alwyn Williams Building
17 Lilybank Gardens
Glasgow
G12 8QQ
delliott@dcs.gla.ac.uk

Joemon Jose
Department of Computing Science
University of Glasgow
Sir Alwyn Williams Building
17 Lilybank Gardens
Glasgow
G12 8QQ
jj@dcs.gla.ac.uk

ABSTRACT

This paper presents a pilot study of a novel context modelling method for capturing an evolving model of information needs during search tasks to assist users in satisfying their long-term needs. Short-term sessions are clustered using terms extracted from explicitly marked search results to create and update aspects of a user profile. The terms that most significantly define the aspects of a user profile are determined using the Ostensive Model and these aspects are used to create a Personal Information Agent to help users satisfy their long-term needs. A novel evaluation technique is introduced comprising four participants performing two simulated work situation tasks over three search sessions where user questionnaires and log file analysis measure the performance of the context modelling method.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Systems and Software-*Relevance feedback*

General Terms

Experimentation, Human Factors, Measurement

Keywords

Ostensive model, context modelling, personalised retrieval, evaluation techniques.

1. INTRODUCTION

In information retrieval, long-term information needs are typically characterised by users seeking resources on the same topic over multiple search sessions. They differ from short-term needs because the user usually has a personal or professional interest driving their behaviour. The majority of long-term information needs have multiple aspects; for example, somebody following developments in U.S. politics may be interested in the legacy George W. Bush leaves behind and the challenges facing Barack Obama during his presidency.

In traditional information retrieval techniques [1] there is an assumption that users will be able to satisfy their information needs with a single query during single session. Relevance feedback techniques [2] emerged to assist users in satisfying their

needs with multiple queries on the same topic in a single session based on result list interactions. The emerging area of contextual information retrieval seeks to develop techniques to capture and represent information needs that evolve over time.

There are several examples of previous research on contextual information retrieval. Harper and Kelly [3] present an interface that allows users to organise their needs into piles and provides contextual feedback through these piles. Their evaluation combines click-through data and a user questionnaire. Martin and Jose [4] present an interface where users bundle documents and receive contextual feedback based on those bundles. Their evaluation also combines click-through data and user questionnaire responses.

Our motivation is help users satisfy their long-term information needs by modelling their interactions with search results and retrieving new web pages on their behalf using these models. In comparison with the approaches in [3][4], our approach combines the data extracted from explicit relevance feedback across multiple sessions to capture the evolving interests of users. We adapt the evaluation techniques used in [3][4] and perform a small user study with two tasks designed to simulate a long-term information need in participants.

We investigate the following questions in this paper:

- How useful is a retrieval system that helps users construct a profile of their evolving information needs?
- How useful is a personal information agent based on aspects of user profiles captured from explicit interactions?

To evaluate our novel context-modelling technique we will change the method of producing terms the system uses and suggests terms that define aspects of user profiles. The *baseline* system will construct user profiles using terms extracted from search queries issued during previous sessions. The *experimental* system will construct user profiles using terms extracted from explicitly marked documents. The evolving significance these terms contribute to the profile will be calculated using the Ostensive Model.

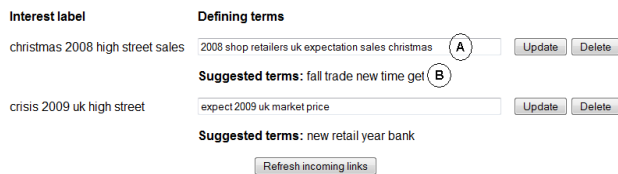


Figure 1: User profile management interface.

The remainder of this paper is structured as follows. Section 2 describes the retrieval system used in our study. Section 3 describes our experimental and baseline context modelling techniques. Section 4 focuses on our evaluation technique and experimental methodology. Section 5 offers some preliminary results and discussion based on an experiment with four participants.

2. SYSTEM DESCRIPTION

Our retrieval system is based on the system originally presented in [5]. It is a web application comprising a separate backend and frontend. The backend consists of components for ad-hoc indexing, session clustering, context modelling, and logging. The frontend consists of interfaces for user profile management, bookmark management, search results, and personal retrieval agent results.

In the backend, a new session is created when a user issues a query and terminated when the user issues a different query. The Google SOAP Search API¹ is used to retrieve results, which are indexed ad-hoc using *tf-idf weighting* on the extracted text content of each result. The current search session is added to the user profile if one or more results are marked as explicitly relevant during the session and clustered with other search sessions by the terms extracted from the results that define the session.

In the frontend, users can search the Web using a typical search box. Users can click-through to inspect result content and mark search results as explicitly relevant. The different aspects of a user profile can be managed from the profile management interface, shown in Figure 1. For each aspect of their profile, users can view the defining terms (A), view a list of suggested terms derived from the context-modelling component in the backend (B), delete the aspect from their profile, and change the defining terms of the aspect. Search results retrieved by the personal information agent can be viewed and interacted with as shown in Figure 2. For each aspect of the user profile, the defining terms used to retrieve documents on behalf of the user are shown (D) and users can click-through to view the result content and explicitly mark the result as relevant (E).

3. CONTEXT MODELLING METHODS

We use the Ostensive Model [6] to determine the evolving nature of aspects of a user profile in long-term search tasks. The Ostensive Model is a model of developing information needs where the user's needs are described as a set of evolving information objects that exemplify the need. Campbell uses the Ostensive Model in a browse-based system to allow users define their needs through implicit feedback [7]. Upon selecting an object that defines their current need, the system immediately updates the set of objects available to the user.



Figure 2: Personal information agent results.

In a query-based system, we believe that immediately changing the search results would frustrate users. We propose that each set of interactions with respect to a query represent a change in the knowledge state of the user and this change is reflected when a search session terminates. Search query and results will be updated based on this change when the search session resumes.

3.1 Experimental Method

Suppose a user issues a *query* to a retrieval system. This action marks the start of a *search session* from which an aspect of the user's evolving information needs can be captured. Search sessions are represented as date-ordered collections of explicitly relevant search results, from either personal information agent results or from search query results, alongside the terms extracted from the text content of those results using a *tf-idf weighting* on the set of search results returned for that query.

Clusters of search sessions represent an *aspect* of the user's *interests*. An interest represents the history of an aspect of the user's evolving information needs. The relevance of the terms that define interests is calculated using an increasing-uncertainty-with-age discounting function [7]. This type of discounting function promotes the significance of the terms that define interactions in the most recent search sessions.

The five terms with the highest weight extracted from the clustered search sessions are the defining terms of an interest in the user profile. These defining terms represent the evolving needs of the user based on their explicit interactions with search results.

3.2 Baseline Method

The baseline context modelling method differs from the experimental method in how the defining terms for interests are obtained. The defining terms of each interest in the user profile are the five most frequently occurring terms used in queries that define the clustered search sessions. The baseline defining query represents the evolving information need of the user but it does not consider the explicit interactions performed by the user after they have issued a search query.

4. EVALUATION TECHNIQUE

Evaluating a contextual retrieval system presents challenges which are not appropriately covered using traditional measures such as precision and recall [1]. In Web search tasks there is rarely an available set of relevance assessments for a task and traditional search tasks are unsuitable to evaluate a system that attempts to model evolving information needs. We propose to use two simulated work situation tasks [8] to evaluate a system that captures a multiple-aspect user profile because we believe this type of task will allow the user more freedom in their searching.

¹ <http://code.google.com/apis/soapsearch/>

ID	Question
A	The incoming results were useful
B	The system successfully modeled my search sessions
C	The initial defining terms were accurate
D	The suggested defining terms were useful
E	I regularly changed the defining terms of my profile

Table 1: User questionnaire differentials

4.1 Hypotheses

We have the following hypotheses from our research questions:

1. Modelling the interactions of users through explicit feedback will assist in the construction of user profiles that represent multiple-aspect of evolving information needs.
2. A Personal Information Agent based on a multiple-aspect user profile will retrieve useful documents on behalf of users.

4.2 Measurements

To test the first hypothesis, we will measure the number of manual changes performed by each user to their profile. We hypothesise that the experimental system will show a trend of fewer changes to the user profile if the profile accurately represents their needs. From the user questionnaire differentials shown in Table 1, we will use differentials B, C, D, and E. We believe the responses to these questions will help us evaluate participants' impressions of the context-modelling component of the systems.

To test the second hypothesis, we will measure click-through and explicit marking rates recorded from the interface. We hypothesise that the experimental system will show a trend of higher click-through and higher explicit marking than the baseline system. From the user questionnaire in Table 1, we will use differential A. We believe a trend of support will be present for this question if participants find the results useful.

4.3 Tasks

The tasks associated with traditional test collections such a TREC² and Cranfield [1] are too narrow in scope to evaluate an adaptive retrieval system. These tasks assume users have static and narrow information needs, for example TREC-9 Web Topics 451³: "Provide information on the Bengal cat breed." To address this problem, we have devised two simulated work tasks, of which Task B is shown in Table 2. The purpose of these tasks is to simulate a reasonably complex information need that can be divided into several subtasks and addressed in an order that suits the user's searching style.

4.4 Methodology

Four participants were recruited from within the Computing Science Department, their ages between 18 and 45. They were required to complete two simulated work situation tasks, over three 15 minute sessions for each work task, on three different days and were paid £25 upon completion of the experiment.

Each participant used both the baseline and experimental system during the experiment. The allocation of systems and tasks were

Imagine you are journalist writing an article on the sales performance of shops on the High Street in the United Kingdom over the Christmas 2008 period. Your article will be published in a major newspaper and read by hundreds of thousands of people. You will want to include a short-guide to the current financial crisis to set the context for your article, some well-sourced material on the affect of the crisis on Christmas sales, and what expectations are for the 1st quarter of 2009 on the High Street.
--

Table 2: Task description for simulated work situation task B.

rotated in a Latin square to attempt to minimise learning effects. Participants had no knowledge of whether they were using the experimental or baseline system for either task; the interface for the systems provided labels of System A and System B. With the exception of the labelling of each system, the baseline and experimental system interfaces presented to users were identical.

In response to an entry questionnaire, all participants were educated to at least Master's degree level, were extremely confident using computers to complete everyday tasks, used Web search engines every day, and mostly found what they wanted using Web search engines.

Before commencing the experiment, participants were given a 10-minute training task to familiarise them with the features of the retrieval system and were allowed to ask questions. After completing the 1st session for both tasks, participants were asked to complete an interim questionnaire. Participants were able to complete the second and third sessions unsupervised from their own computer. After the 3rd session for both tasks, participants completed a post-task questionnaire; and finally, the participants completed an exit questionnaire.

5. RESULTS AND DISCUSSION

In this section we present preliminary results from our user study. Figure 3 shows the mean intermediate and post-task responses to the user questionnaire described in Table 1. The responses range from 1-5 on a Likert scale, where 1 represents a negative response and 5 represents a positive response, with the exception of differential E which has a reversed scale.

5.1 Context-modelling Method

The result of the log file analysis, shown in Figure 4, appears to contradict our initial hypothesis. In the experimental system each user made additions to the terms defining aspects of their user profile, however, insufficient data was captured to determine if the additions made by users to their user profile in the experimental system were a result of the term suggestions. Responses to questionnaire differentials B, C, and D in Figure 3 present little difference between the baseline and experimental systems.

5.2 Personal Information Agent

The results of the log file analysis, shown in Figure 5, show that three out of four participants marked more personal information agent results as relevant in the experimental system. For questionnaire differential A, in Figure 3, there is little different in responses at between the baseline and experimental systems.

² <http://trec.nist.gov/>

³ http://trec.nist.gov/data/topics_eng/index.html

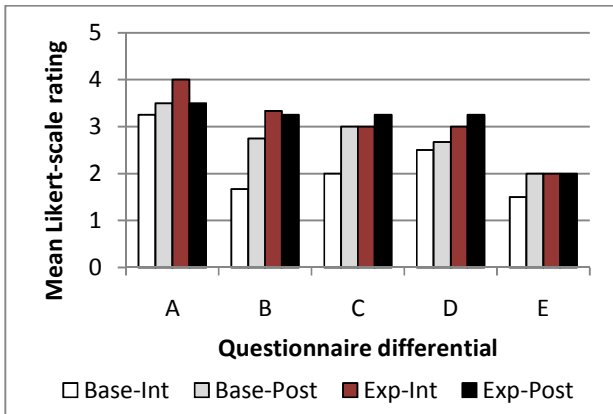


Figure 3: Graph of questionnaire responses and interim stage and post-task stage on differentials A –E from Table 1.

6. CONCLUSION AND FUTURE WORK

The preliminary results presented in this paper support the utility of a contextual retrieval system that assists users with their long-term information needs. The support is weak, however, and further studies with a refined evaluation design and more users are required to draw conclusions.

We believe the difficulty of designing an evaluation technique for a multi-aspect personalised retrieval system attributed to these results. In the future, we plan to redesign the evaluation technique presented in this paper. Evaluating a retrieval system across multiple search sessions poses several problems, including how to ensure the tasks assigned to users are complex enough to warrant multiple sessions, whether to allow users the freedom to complete the tasks unsupervised, and which measures will appropriately evaluate the performance of such a retrieval system.

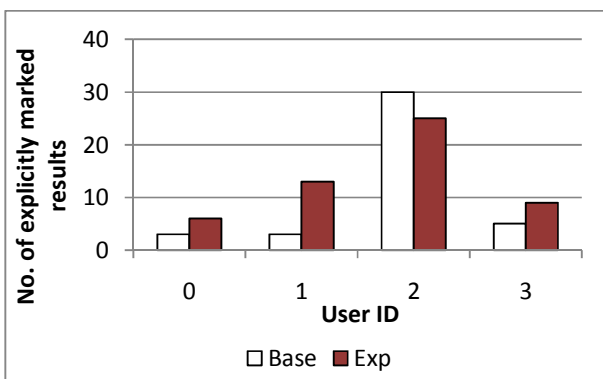


Figure 4: Graph of number of results retrieved by each user's PIA marked as explicitly relevant over the experiment.

7. ACKNOWLEDGEMENTS

This research was supported by the European Commission under contract FP6-033715-MIAUCE. It is the view of the authors but not necessarily the view of the community.

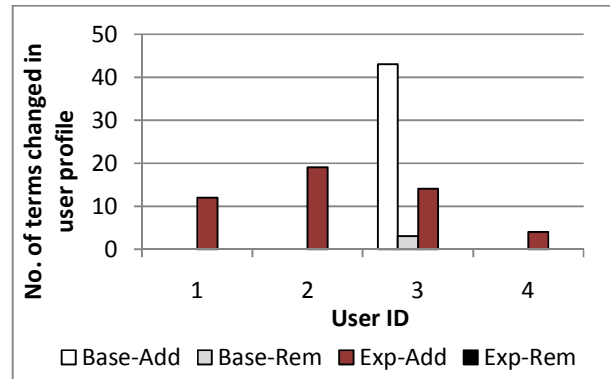


Figure 5: Graph of the number of additions and removals made to the defining terms of aspects of each user's profile.

8. REFERENCES

- [1] Cleverdon, C.W., and Keen, M. 1968. Factors determining the performance of indexing systems, Vol. 1: Design, Vol. 2: Test Results. Aslib Cranfield Research Project, Cranfield, Bedford, England.
- [2] Rocchio, J.J. 1971. Relevance feedback in information retrieval. In *The SMART Retrieval System: Experiments in Automatic Document Processing*, G. Salton, Eds. Prentice-Hall, Englewood Cliffs, NJ, USA, 313-323.
- [3] Harper, D.J., and Kelly, D. 2006. Contextual Relevance Feedback. In *IiX: Proceedings of the 1st International Conference on Information Interaction in Context*, Copenhagen, Denmark, 129-137.
- [4] Martin, I., and Jose, J. 2003. A Personalised Information Retrieval Tool. In *SIGIR '03: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, Canada, 423-424.
- [5] Psarras, I., and Jose, J. 2006. A System for Adaptive Information Retrieval. In *Proceedings of 4th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, Dublin, Ireland, 313-317.
- [6] Campbell, I., and van Rijsbergen, K. 1996. The Ostensive Model of Developing Information Needs. In *Proceedings of CoLIS-2*, Copenhagen, Denmark, 251-263.
- [7] Campbell, I. 2000. Interactive Evaluation of the Ostensive Model using a New Test-Collection of Images with Multiple Relevance Assessments. *Journal of Information Retrieval* 2, 89-114.
- [8] Borlund, P. 2000. Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation* 56, 71-90.