Evaluating an Associative Browsing Model by Simulation

Jinyoung Kim, W. Bruce Croft, David A. Smith
Department of Computer Science
University of Massachusetts Amherst
{jykim,croft,dasmith}@cs.umass.edu

ABSTRACT

In our previous work [5], we proposed an associative browsing model of personal information in which users can navigate through the space of personal items. In another recent work [8], we evaluated the model based on user study.

In this paper, we evaluate the associative browsing model we proposed in the context of known-item finding task. We built a model of user which is parameterized to simulate various aspect of user, and performed experiments to evaluate the associative browsing model under various conditions.

We find that user's level of knowledge and other characteristics affect their known-item finding behavior. The results also confirmed our earlier user-based evaluation, showing that the associative browsing model can play a complementary role in known-item finding.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: [Information Search and Retrieval]

General Terms

Algorithms

Keywords

Associative Browsing, Known-item Finding, Simulated Evaluation

1. INTRODUCTION

Recently, keyword search has become a standard feature for many platforms. Although it can greatly ease the task of finding personal information, there are many cases in which the user's initial search attempt fails. Previously [5], we proposed an associative browsing model of personal information. Our work improves on previously suggested models of associative browsing in that we proposed more general measures of association (e.g. textual similarity and co-occurrence), whereas previous models defined links only between a limited set of items.

We also introduced a learning framework for ranking suggestions for browsing in more recent work [8], and evaluated the associative browsing model in the context of known-item finding, which is the most common task in personal information access [4]. We performed a game-based user study in which participants were asked to find a set of target documents by combining keyword search and associative browsing. The study reveals that the participants often choose to use associative browsing with high chances.

In this paper, we performed experiments with a simulated model of a user. The model is parameterized to simulate various aspects of user behavior in known-item finding. Experimental results using this model suggest that associative browsing can help a user to find the target item when keyword search returns only marginally relevant results effectively (with about 40% success rate) and efficiently (within 2 browsing steps). We also found how a user's level of knowledge and browsing behavior has a subtle relationship with the effectiveness and efficiency of one's interaction with the system.

2. RELATED WORK

Simulated evaluation has been done in many related tasks, including known-item search [1] [7] and associative browsing [9] [10]. Among others, Smucker et al. [10] and Lin et al. [9] recently proposed the use of simulation in the context of associative browsing. While the associative browsing component of user model proposed in this work is similar to their simulation technique, our evaluation is based on the task of known-item finding, which enables modeling user's knowledge based on the target item. The idea of modeling the user's interaction with the system using some kind of probabilistic model was also proposed in [3]. Our user model is conceptually similar to this work, although we implemented and evaluated this idea in a different domain.

3. ASSOCIATIVE BROWSING MODEL

In this section, we briefly introduce the associative browsing model we proposed earlier [5] with some simplification. On a high level, our associative browsing model is composed of documents and the *associations* between them. Given the collection of documents, we allow the user to browse between documents by clicking on the ranked list of related documents. The details of ranking suggestions for browsing are provided in our recent work [8].

HCIR'11, Mountain View, CA, USA

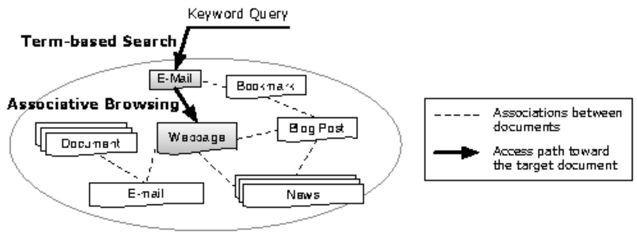

Figure 1: An illustration of how the suggested associative browsing model can be used for known-item finding.

Among many potential use cases for this associative browsing model, we focus on the known-item finding scenario, since it is the most common task in personal information access and the well-defined structure allows us to use the evaluation methods introduced in Section 4. Here we provide an example on how associative browsing can be combined with keyword search for known-item finding.

Imagine a user who is trying to find a webpage she has seen. Further assume that she cannot come up with a good keyword for search, yet she remembers a related email. Using our model, as shown in Figure 1, the user can first use keyword search to find a relevant email, and then browse into the target document (webpage). Here, dotted lines represent the associations between documents. Directed lines denote how a user can access the target webpage by using keyword search and associative browsing.

4. SIMULATED USER MODEL

In our recent work [8], we adopted a game-based user study where we asked people to perform known-item finding tasks using both search and browsing capabilities. We ran two rounds of game-style user studies in which participants were asked to perform given tasks in a competitive environment. Using the data from the user study, we analyzed the user's behavior in finding known-items, and evaluated the algorithms for ranking suggestions for browsing.

In current study, we employed evaluation method based on a simulated model of user. Our goal is to build a reasonable simulation of the actual user behavior for the evaluation of our system for the known-item finding task. The user model is parameterized to simulate various aspects of the user. While we do not argue that experiments based on a simulated user model can substitute for user studies, they provide a valuable means of evaluating the system under various conditions, complementing the evaluation methods where users are involved.

Specifically, we simulate a user who wants to find a known item using the system. Figure 2 shows the diagram of state transitions that are involved in this sequence of interactions between the user and the system. As a starting point, we expect the user to perform a keyword search using the terms he remembers from the document. If the initial search is successful, he can finish the session. Otherwise, he can either reformulate the query or click on one of the top documents to browse into related items. This process continues until he finds the target document or he reaches the limit of his patience.

We divided the model into three components — keyword search, associative browsing, and the transitions between states. The keyword search component models how the user would choose terms for search, and the associative browsing component is responsible for modeling the user's clicks on the ranked list. The state transition part is concerned with the decision made by the user whether to use search or browsing, or whether to continue the current session or terminate. In what follows, we explain each component in detail, focusing on associative browsing.

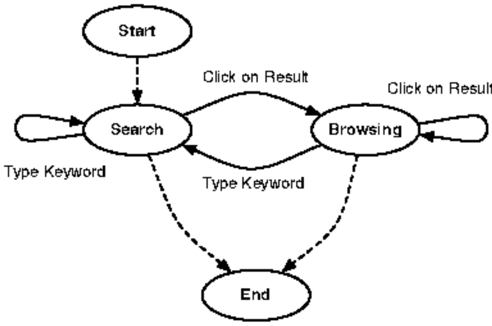

Figure 2: A state transition diagram for suggested probabilistic user model.

4.1 Keyword Search Model

Azzopardi et al. [1] showed that known-item queries can be generated by taking terms from target documents based on some distribution. Kim et al. [6] further refined the model and showed that such known-item queries can be used for experiments in personal document collections. Since we are dealing with known-item finding task in a personal documents collection, we used the query generation model suggested in [6] to get queries targeted for finding a document.

More specifically, given a target document and pre-specified length of query, we choose each query-term from a term distribution P_{term} estimated from the document until we reach the limit in pre-specified length.

4.2 Associative Browsing Model

In our system, when a keyword query returns only marginally relevant results, a user will click on one of top documents to browse related documents to locate the target document. We have several choices in modeling this behavior.

The first choice in modeling browsing is the level of knowledge the user has about the collection and target documents. A more knowledgable user will make a better choice in deciding which document to click on. We introduce three levels of user's knowledge — *random*, *informed* and *oracle*, which correspond to the status of no knowledge, partial knowledge and complete knowledge, respectively.

To implement the level of knowledge in user clicks on the ranked list, we need to evaluate the candidate documents in terms of their value in getting access to the target document. In the known-item finding scenario, where the target document is known and each click leads to a ranked list which may contain the target document at some position, we can use a ranking effectiveness measure (MRR) for each candidate document to evaluate its value in locating the target document.

More specifically, while the *random* user may click on a random position of a ranked list, the *informed* user will

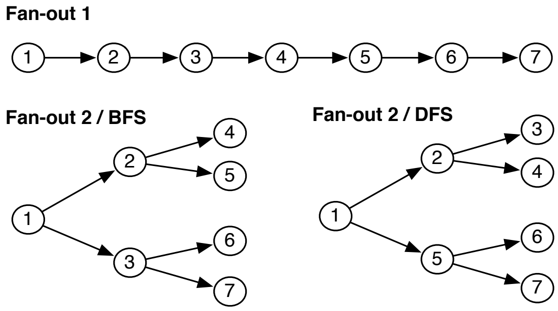

Figure 3: An illustration of two browsing strategies: breadth-first search (BFS) and depth-first search (DFS), both with fan-out of 2. Numbers represent the order of documents in which they are visited.

choose documents from the distribution of candidate documents whose probability corresponds to the estimated value of each candidate document. Finally, the *oracle* user will always click on the document with highest value. The behavior of the *oracle* user is greedy in that the choice is based on what seems the best each moment, and we show in Section 5 that this greedy strategy does not always lead to the highest success.

Another consideration in user modeling is the variations in browsing behavior — how many documents are visited at each time the user sees a ranked list, and in what order. We used a variant of two browsing strategies introduced in Smucker et al. [10] — depth-first strategy and breadth-first strategy.

Figure 3 illustrates three examples of browsing strategies, where each node represents a document, and the number in each dot represents the order in which documents are visited. Since each document corresponds to the ranked list of related documents, it is clear that each arc corresponds to user's click.

In summary, two parameters we use in modeling users' browsing behavior are the fan-out (how many documents the user clicks on a ranked list), and the browsing strategy employed (breadth-first search and depth-first search). Here, higher fan-out means more exploration than exploitation (more clicks per ranked list), while BFS and DFS represents exploration-first strategy and exploitation-first strategy, respectively.

4.3 State Transition Model

The rest of the simulated user model is concerned with the decision made by the user whether to use search or browsing, or whether to continue the current session or terminate. Since our main goal in this work is evaluating the role of browsing as a complement for search, we used a simplifying assumption that users would choose to browse if the initial search is only marginally successful. Although there can be many considerations in modeling this component, we leave them for future work.

5. EVALUATION RESULTS

In this section we present the evaluation results of simulation experiments. The dataset we use (called CS collection) contains 7,984 public emails, webpages, publications and lectures crawled from the computer science department website of the authors. The CS collection was created as a reasonable simulation of personal information, since we performed our previous user-based evaluation [8] with participants from our department, who had some knowledge of these documents.

We now report on the experiments with the simulated user model described in Section 4. As for the parameters of keyword search model, we used the language model of a document for P_{term}, and set the query length to 1.5 on average following the average length of queries in previous studies [2]. For the associative browsing component, we experimented with three levels of user's knowledge (*random*, *informed* and *oracle*), fan-out of 1 to 3, and BFS and DFS browsing strategies.

In modeling the transitions between states, we assumed that the user chooses to browse only when the ranked list returned by keyword search looks marginally relevant, and that there is no transition from browsing back to search. We use the term *marginally relevant* for the case where the target document is located between the rank position of 11 to 50. Whenever the target document is found within top 10 positions, we consider the session as *successful* and finishes the interaction. The session is unsuccessful if the user failed to find the target document within 10 trials.

For each target item, we ran the simulated user model described above. In order to keep the quality of ranking consistent, we used a simple vector space model based on Lucene search engine toolkit[1] for both keyword search and associative browsing (i.e., we used only content similarity feature among features introduced in [8]). Finally, since the simulation involves random generation of user's behavior, we ran all the experiments 10 times and report the average result.

Here are the results from the simulation experiments. Table 2 shows the success ratio of browsing aggregated for three models of user's knowledge and fan-outs. As mentioned above, Success ratio here denotes the portion of sessions where browsing led to *success* among all sessions where initial queries were *marginally relevant*.

In aggregate, the result in the first row of Table 1 shows that browsing was used for around 15% of sessions. Since the algorithm start browsing only when initial search result is marginally relevant, this indicates the quality of keyword query we used — 15% of queries found the target document between the rank of 11 to 50. What's more interesting is that the browsing saved (i.e., led to success) around 42% of those sessions, showing that associative browsing in general is quite effective.

We also compared the results with those from our previous user-based evaluation in Table 1. We found that the success ratios of the simulation study matches with the ratio of successful browsing sessions based on the user study we reported in our previous work, which seems to indicate the assumptions we made in simulation experiments are reasonable.

With respect to different levels of user knowledge, even when the user browses randomly without any knowledge of the target document, the chance for success is almost 30%, showing that associative browsing is an effective alternative to search when the initial query is only marginally relevant. At the same time, however, the success ratio of browsing

[1]http://lucene.apache.org

Table 1: The ratio of the sessions where simulated user model chose to use browsing, and the choice of browsing led to success, compared to our previous user-based evaluation.

Evaluation type	Total	Browsing used	Successful
Simulation	63,260	9,410 (14.8%)	3,957 (42.0%)
User Study [8]	290	42 (14.5%)	15 (35.7%)

was no higher than 50%, showing that there are cases where the target document simply cannot be reached by browsing.

Table 2: Success ratio of browsing for marginally relevant queries.

	random	informed	oracle
FO1	0.337	0.401	**0.424**
FO2	0.408	**0.453**	0.441
FO3	**0.442**	0.436	0.426

Table 3: Average length of successful browsing session.

	random	informed	oracle
FO1	1.417	1.391	**1.266**
FO2-BFS	2.186	1.904	**1.661**
FO3-BFS	2.083	1.959	**1.814**
FO1	1.417	1.391	**1.266**
FO2-DFS	2.280	1.928	**1.257**
FO3-DFS	2.327	1.805	**1.323**

An interesting trend in Table 2 is the relationship between the user's level of knowledge and fan-out. We originally expected that the *oracle* user would outperform others at all fan-outs, yet it was found that higher fan-out (more exploration) only hurts *oracle* user whose level of knowledge is very high, yet always make a locally optimal decision.

In contrast, the success ratio of the *random* user increased with higher fan-out, which shows that exploration is valuable only when user's level of knowledge is low. Overall, the *informed* user with a fan-out of two gave the best performance.

We then looked at the efficiency of using browsing for known-item finding, which we measured by the average length of successful browsing sessions — how many clicks it took for the user to find the target document. Here we compared three levels of user's knowledge and two browsing strategies — BFS and DFS for each fan-out.

The result in Table 3 shows that one or two clicks are usually sufficient to get to the target document by browsing. Comparing different levels of user knowledge, the *oracle* user model is always more efficient, followed by *informed* and *random*. We can conclude that the user's level of knowledge has a direct influence on the efficiency of browsing.

Among different browsing behaviors, it is clear that higher fan-out (more exploration) leads to lower efficiency in most cases, as we would expect. A less obvious trend is that the *random* user is more efficient with BFS strategy (exploration first), while the *oracle* user is more efficient with DFS strategy (exploitation first), while. Again, we can infer that higher levels of knowledge makes exploitation more valuable than exploration.

In summary, the simulation experiments show that associative browsing provides an effective (30-40% of success) and efficient (within 1–2 clicks) way of getting to the target document when keyword search is marginally relevant. Comparison of results across different levels of user knowledge and browsing behavior also reveals the influence of various aspects of the user on the value of associative browsing for known-item finding.

6. CONCLUSIONS

In this paper, we evaluated an associative browsing model we proposed using simulated user model in the context of known-item finding. Our evaluation confirmed our earlier user-based evaluation [8], showing that associative browsing provide an effective and efficient alternative when keyword search fails. Comparison of results across different levels of user knowledge and browsing behavior also reveals the influence of various aspects of user on the value of associative browsing for known-item finding. We plan to refine the simulated user model in future work by incorporating more characteristics of user and the system.

7. ACKNOWLEDGEMENTS

This work was supported in part by the Center for Intelligent Information Retrieval, in part by NSF grant #IIS-0707801, and in part by National Science Foundation under NSF grant #IIS-0326249.

8. REFERENCES

[1] L. Azzopardi, M. de Rijke, and K. Balog. Building simulated queries for known-item topics: an analysis using six european languages. In *SIGIR '07*, pages 455–462, New York, NY, USA, 2007. ACM.

[2] S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins. Stuff I've seen: a system for personal information retrieval and re-use. In *SIGIR '03*, pages 72–79, New York, NY, USA, 2003. ACM.

[3] D. Elsweiler, S. Mandl, and L. Azzopardi. Evaluating personal information management behaviour using markov models. In *Wissensmanagement*, pages 280–288, 2009.

[4] D. Elsweiler and I. Ruthven. Towards task-based personal information management evaluations. In *SIGIR '07*, pages 23–30, New York, NY, USA, 2007. ACM.

[5] J. Kim, A. Bakalov, D. A. Smith, and W. B. Croft. Building a semantic representation for personal information. In *In Proceedings of CIKM'2010, Toronto, Ontario, Canada*, 2010.

[6] J. Kim and W. B. Croft. Retreival experiments using pseudo-desktop collections. In *in Proceedings of CIKM'2009, Hong Kong, China*, pages 1297–1306, 2009.

[7] J. Kim and W. B. Croft. Ranking using multiple document types in desktop search. In *In Proceedings of SIGIR '10*, pages 50–57, New York, NY, USA, 2010. ACM.

[8] J. Kim, W. B. Croft, D. A. Smith, and A. Bakalov. Evaluating associative browsing model for personal information. In *To Appear in Proceedings of CIKM'2011, Glasgow, Scotland, UK*, 2010.

[9] J. Lin and M. D. Smucker. How do users find things with pubmed?: towards automatic utility evaluation with user simulations. In *SIGIR '08*, SIGIR '08, pages 19–26, New York, NY, USA, 2008. ACM.

[10] M. D. Smucker and J. Allan. Find-similar: similarity browsing as a search tool. In *SIGIR '06*, pages 461–468, New York, NY, USA, 2006. ACM.