

Towards Profile-Based Document Summarisation for Interactive Search Assistance

Azhar Alhindi
School of Computer Science and Electronic Engineering,
University of Essex, UK
ahalhi@essex.ac.uk

This paper presents an investigation into the utility of profile-based summarisation in the context of site and enterprise search. We employ log analysis to acquire continuously updated profiles to provide profile-based summarisations of search results with the intention of highlighting aspects of the documents that best fit the general user profile. We first introduce the wider context of the research and then focus on a first task-based evaluation using TREC Interactive Track guidelines that compares a search system that uses the outlined profile-based summarisation with two baseline systems to assess whether such summaries could be helpful in a realistic search setting.

Profiling, Summarisation, Enterprise search, Task-based evaluation, Logs

1. MOTIVATION

Finding a specific document on a company Web site or within a university intranet can be non-trivial, because unlike in Web search, such document collections tend to be sparse, in that there might be only a single matching document for a user query, which could be difficult to find. A search for the exam timetable on the University of Essex Web site, for example, will only satisfy the (most common) user need if it results in a specific Excel file that might not be easy to identify when navigating the site. This is a typical problem of enterprise search (Hawking 2011). The growth of document collections on Web sites of universities, companies and other institutions, i.e., collections other than the Web as such, will contribute to this becoming a more and more common problem. Finding solutions that address this problem is considered a challenging task (Hawking 2011).

Generally speaking, there appears to be a trend in assisting a user in the search process, but this could be done in a variety of forms (e.g., apart from query suggestions (Kato et al. 2012), faceted search (Koren et al. 2008) has become very popular in recent years. However, in this process of guiding/assisting the searcher, no dominant paradigm has emerged which would be comparable

to the Google-style paradigm of ad hoc search. This is the (wider) area that we will explore.

One of the trends that play an important role in the exponential growth of document collections is traditional (generic) summarisation (Nenkova and McKeown 2011). However, in this case, the impact of human interest for readers has seldom been considered. Traditional summarisation generates the same summary for different users by utilizing the same methodology without taking into account who is reading. As most of the current summarisation systems generate a uniform version of a summary for one document for all users, most traditional summarisation methods fail to capture user interests during summarisation because they treat their outputs as static and plain texts. However, users need personalisation because they have individual preferences with regard to a particular source document collection; in other words, each user has different perspectives on the same text. So, traditional summarisation methods are to some extent insufficient because, obviously, a universal summary for all users might not always be satisfactory (Yan et al. 2011). As such, personalised text summarisation could be useful in order to present different summaries corresponding to reader interests and preferences (Zhang and Ma 1990). We are interested in personalised summarisation. One of the techniques used to achieve personalisation

is user profiling. User profiles may include the preferences or interests of a single user or a group of users and may also include demographic information (Gauch et al. 2007). Normally, a user profile contains topics of interest to that single user. We are interested in capturing profiles not of single users but groups of users.

Broadly speaking we try to address the following research questions with our work:

1. Can site search and enterprise search benefit from the automated summarisation of results?
2. Will a continuously updated model capturing search and navigation behaviour of a user (or groups of users) be beneficial for the summarisation process?
3. Will such methods result in measurable (quantifiable) benefits such as shorter search sessions, fewer interactions etc?

In this paper we report on initial experiments we have conducted to address the questions.

2. RELATED WORK

Automatic summarisation (Nenkova and McKeown 2011) is a process that creates a shortened version of one or more texts that contains the most important points of the original text, and is both concise and comprehensive. (Hassel 2004) divides summarisation approaches into two main groups: abstractive summaries and extractive summaries. Here, we focus on extractive summarisation.

Some systems generate a summary based on multiple source documents (for example, a cluster of news stories on the same topic), which are known as multi- document summarisation systems (Maybury and Mani 1999), while others can use a single-source document, which are known as single-document summarisation systems. (Radev et al. 2002) use abstraction and extraction methods to apply the text summarisation process to single and multiple documents. The most frequently used methods to build extractive generic summaries use position, thematic words, indicative expressions, text typography, proper nouns and title as described by (Hahn and Mani 2000; Díaz and Gervás 2007). An ongoing issue in this field is that of evaluation; human judgements often have wide variance on what is considered a good summary, thus making the automatic evaluation process particularly difficult (Lin and Hovy 2002).

The above algorithms do not involve interactive mechanisms to capture reader interests, nor do

they utilize user preferences for personalisation in summarisation. They usually are traditional extensions of generic summaries. According to (Díaz and Gervás 2007), experiments have shown that personalised summarisation is important, as the summary sentences match the user's interests. Personalisation techniques include seeking to adapt to individual users, which aims to improve user satisfaction by adapting future interactions and predicting user needs through building models/profiles (Gauch et al. 2007) of user behaviour. One approach of building adaptive community profiles is a biologically inspired model based on ant colony optimisation applied to query logs as an adaptive learning process (Albakour et al. 2011). This approach has been considered in our work to build a domain model/profile that is then applied for summarisation.

A number of personalised summarisation methods have been explored e.g., (Berkovsky et al. 2008; Wang et al. 2007; Park 2008; Zhang et al. 2003). However, (Wang et al. 2007) focused on query-based summarisation for Web pages based on the extraction and ranking. (Park 2008), on the other hand, performed a personalised summary in which the sentences relevant to user interests have been extracted for the query-based and generic summary with regard to a given query based on non-negative matrix factorization (NMF). The potential of personalised summarisation over generic/traditional summaries has already been demonstrated e.g., (Díaz and Gervás 2007), but summarisation of Web documents is typically based on the query rather than a full profile e.g., (Wang et al. 2007; Park 2008). However, such scenarios may not be sufficient enough in that they depend on only the current submitted query, which might not contain much information that accurately describes user interests in the generated summaries. Our specific interest lies in site search and enterprise search, which is different from Web search and has attracted less attention (Hawking 2011). Our approach can effectively and implicitly learn the user profile, which contains user interests, and keep it up-to-date and then generate personalised summaries that reflect those user interests under a wider practical scenario. The benefit of this context is that we can expect a more homogeneous population of searchers who are likely to share interests and information needs. We then apply the acquired profiles to generate summaries that support users who are searching a document collection.

3. LOG-BASED PROFILES

We use query logs to build profiles that represent not an individual user but the entire population of users that access a Web site. The intuition behind this is

that, for example, on a university Web site or in a company's intranet people tend to share information needs and we assume that learning from one user (e.g. a new student trying to find out how to register) might benefit a whole range of future users. We utilise query logs to acquire a profile which is being automatically updated in a continuous learning cycle using an ant colony optimisation (ACO) analogy, as adopted from (Albakour et al. 2011). For the specific experiments, we use the log files collected on an existing site search engine over a period of three years¹ to bootstrap such a model, i.e., our group profile. The idea is that query logs, the structure of logs discussed here, are segmented into sessions and then turned into a graph structure (Kruschwitz et al. 2011). We then apply this profile in the summarisation process. Our search system architecture and some relevant data structure have been discussed in (Alhindi et al. 2013a,b) where more details can be found.

4. INITIAL EXPERIMENT

To explore whether summarisation of documents in a site search context, and profile-based summarisation in particular, offers any measurable benefits we conducted a pilot study (Alhindi et al. 2013a). In that experiment, we simply assessed how users perceive summaries generated using a profile compared with different baselines. We found that a profile-based summarisation process significantly outperformed a centroid-based baseline that did not utilise any profile. We further identified that a query-specific profile gave the best results among a range of profile-based summarisation methods. Hence, we found that there is potential in utilising profiles of either users or groups of users in the summarisation process.

The next step is to investigate whether the results obtained in the pilot study can also be demonstrated in actual search applications.

5. TASK-BASED EVALUATION

Motivated by the findings of the pilot we designed a task-based evaluation that would use summarisation techniques to generate snippets for matching documents returned for a user query, similar to the one presented here (Tombros and Sanderson 1998) with respect to the use of summaries for search of non web content. Based on commonly used standards in task-based evaluations (Kelly 2009; Yuan and Belkin 2010; Diriye et al. 2010), we constructed search tasks from random samples of representative search requests on the domain of

¹More than 1.5 million queries, described in more detail elsewhere (Kruschwitz et al. 2013)

choice and conducted a within-subjects laboratory evaluation to compare three different information retrieval (IR) systems. The evaluation process started with exploring the potential of an adaptive search system by capturing the feedback on it from real users. In line with (Kelly 2009), we chose 18 subjects and each one attempted to complete 6 search tasks which asked them to find documents that are relevant to pre-determined topics. Each subject completed two searches on each system (a within-subjects design). In accordance with the TREC-9 Interactive Track guidelines (Hersh and Over 2001; Hersh 2002), subjects had 10 minutes for each task, and they were asked to complete a number of questionnaires. The questionnaires we used in this study are based on the ones suggested by the TREC-9 Interactive Track guidelines² (using a 5-point Likert scale where appropriate). We used Entry questionnaire, Post-search questionnaire and Exit questionnaire. We discuss the experimental setup in more detail first; then we discuss the results.

5.1. Experimental Setup

We have developed an integrated Solr-based search system applying a number of different methods for building summaries for search results using the Web site of the University of Essex. One of the methods we have applied is our adaptive summarisation method and would like to test it against two baselines. These three IR systems will be called System A (baseline 1), System B (baseline 2) and System C (the adaptive system). All three systems looked identical to the user, but each one is characterized as follows:

1. *System A* is a copy of a standard search engine that the users usually use to locate their information needs. The query is submitted by the user, our search engine returns results, and the top 100 matches are displayed (10 results per page) using snippets returned by the search engine.
2. *System B* is the same as System A but uses a centroid-based approach (Radev et al. 2004) instead of snippets to summarise the document. This algorithm is designed for traditional (generic) summarisation, and it represents a widely used baseline e.g., (Yan et al. 2011).
3. *System C* is the same as System B but uses the ACO query refinements technique which is profile-based and query-specific instead of using snippets to summarise the document, as described in (Alhindi et al. 2013a).

²<http://www-nlpir.nist.gov/projects/t9i/qforms.htm>

Table 1: An example of different snippets for the same document returned for query “tuition fees”.

System	Snippet/Summary
A	University of Essex :: Postgraduate tuition fees 2011-12 for taught courses Skip to main content University of Essex Search this site: Search Students Staff Study with us Life at Essex International.
B	Graduates of the University of Essex may be eligible for an Alumni Loyalty Discount Please see Additional Notes below.
C	Our Tuition Fee Payment and Liability Policy sets out the University’s regulations regarding payment of fees and our Tuition Fee Deposit Policy offers guidance on who is required to pay fee deposits and the rules for doing so.

Note that, in System B and System C, sometimes we cannot generate a summary for a document for a number of reasons, such as there is no text within the document, the document cannot be parsed (e.g., it is a PDF and not HTML document), or there is no query refinements for the submitted query (as in System C). In this case, we present snippets provided by the search engine (as in System A). Table 1 shows an example of a document extracted from the three systems but with different snippets/summaries (according to the characteristics of each system).

5.2. Protocol and Search Tasks

We followed the procedure adopted in (Craswell et al. 2003) to guide the subjects during the task-based evaluation, which was conducted in an office in a one-on-one setting. Systems and task orders revolved and were counterbalanced. Tasks were assigned to subjects based on a Graeco-Latin square design (Kelly 2009) to avoid task bias and potential learning effects. At the beginning, subjects were asked first to fill in the entry questionnaire. After that, subjects were given 5 minutes’ introduction of the three systems without being told anything about the technology behind them. Then, each subject had to perform 2 search tasks on each system according to the matrix in (Kelly 2009). After each task, subjects were asked to fill in the post-search questionnaire. When they completed both search tasks on one system, they were asked to fill in the post-system questionnaire. Finally, when subjects finished all the search tasks, they had to fill in the exit questionnaire.

We constructed the search tasks in line with the brief review guidelines suggested by (Kules and Capra 2008). Tasks were tailored based on commonly submitted queries in logs of the existing Web site to make the search tasks as realistic as possible (Dignum et al. 2010).

5.3. Subjects

In order to get a good selection of different types of users and to avoid bias in the selection process, we sent an e-mail to the local university mailing list and

Table 2: Average completion time (in seconds).

System	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Total average
A	173	247	190	190	163	260	204
B	154	240	194	187	177	264	203
C	174	237	187	185	170	260	202

selected the first 18 volunteers who replied. They were informed in advance that they would receive £10 for participation. Out of 18 participants, 13 were male and 5 were female. Their ages ranged from 18 to 25 (average age: 20.83). Participants had a variety of backgrounds, including 7 different disciplines (Criminology and Media, Computer Science, Electronic Engineering, Information and Communication Technology, Modern Languages, Banking and Finance and Computer Engineering). Only 4 subjects had taken part in previous studies on online searching.

All subjects declared that they use the Internet on a regular basis. The average time subjects have been doing online searching is 7.38 years (9 of them between 3 and 15 years, but there was also a user who stated 0 years). When asked for their searching behaviour, 15 (or 83%) of the participants selected ‘daily’. Note that our users (who we would consider typical target users of the system) tend to have a lot of experience using Web search systems (mean: 4.94) but little experience using commercial search engines (mean: 2.78).

5.4. Results and Analysis

Completion Time / Number of Turns

Table 2 gives a picture of the average completion time (derived from the logged data) broken down for each task. We measured the time between presenting the search task to the users and the submission of the result. Overall, the average time spent on a search task on System A was 204 seconds, System B was 203 seconds, and System C was 202 seconds, with no statistically significant difference between each pair of them.

We also investigated the number of turns, which is the number of steps required to find the answer for a search task (Table 3). A turn can be inputting a query (this turn is considered here), following the link to the next 10 matches or following a hyperlink to open a document. On average, users needed 3.83 turns on System A, 3.68 turns on System B and 3.65 turns on System C. For five out of six tasks the average number of turns taken is shorter on the profile-based system than on any of the two baselines, although overall the differences are not significant.

Questionnaires

Table 3: Average number of turns to complete a task.

System	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Total average
A	3.65	4.80	3.25	3.25	2.66	5.38	3.83
B	2.65	4.72	3.43	3.12	2.98	5.22	3.68
C	3.60	4.40	3.12	3.00	2.80	5.00	3.65

After participants finished each search task, they had to fill in a post-search questionnaire and answer a number of common questions. One question is 'Are you satisfied with your search results?' Overall, users were satisfied with the results returned by the three systems. A pairwise t-test over the average ratings of the tasks on each system indicates that there is no significant difference between each pair of the three systems.

In this post-search questionnaire, after the common questions users were also asked to state whether they were able to complete their search task successfully. For System A, 5 answered with 'No'; for System B, 3 answered with 'No'; and for System C, there was 2 cases altogether. We also looked into the submitted documents during the search session after the user finish a task and judged whether we would consider a document a match for the task (as we already knew the required documents for each task). We found that a large number of submitted documents exactly matched the information request as specified by the task (34 on System C, 33 on System B and 31 on System A). Only 10 of the 108 search tasks did not result in exact matches, and that includes partial matches. There was no significant difference between the three systems in that respect, but there were two particularly difficult tasks: tasks 2 and 6 (clearly reflected by the user satisfaction). Only 12 of the 18 users found a correct document for task 6, and 14 were correctly submitted for task 2. If we look at those two tasks in detail and compare them to the results reported earlier, we find that they have a higher number of turns; on average, users needed 4.64 turns for task 2 and 5.2 turns for task 6 (see Table 3). We also find a higher average completion time for those two tasks (see Table 2). The success rate was comparable across all systems (with only 10 cases of incomplete tasks).

After two search tasks were performed on one system, participants filled in a post-system questionnaire. No statistically significant difference between the three systems can be observed from the overall results regarding learning to use, ease of use and understanding of each system. This is perhaps what we expect, as the three systems look identical and only differ in the snippets/summaries they make.

In the exit questionnaire, users were asked to answer the question 'Which of the three systems did you like the best overall?' There were marginal differences between systems, so most users found

no difference: 5 users preferred System C, 2 users preferred System B, 2 users preferred System A, and 9 found no difference. A large majority of users also judged that there was no difference between the three systems in two other categories: in the ease of use (C: 4 users, B: 2, A: 2, no difference: 10) and in ease of learning to use (C: 3 users, B: 1, A: 2, no difference: 12).

6. CONCLUSION AND FUTURE DIRECTIONS

Our initial experiment suggested that there is certainly potential in using profile-based summarisation in a site search context. In the task-based evaluation we conducted we found that our profile-based summarisation approach was marginally better than the two baselines according to any of the criteria we investigated. This is a good starting point for further work which will aim to exploit the full potential of the approach. It needs to be pointed out that getting significant improvements over strong baselines such as the ones chosen here will not be easy to achieve (note that search engines tend to generate good snippets for queries).

Apart from exploring the full potential of profiles in the given context our future work will focus on multi-document summarisation and the application of the profiles in navigation rather than search, an area where profile-based suggestions of links has already been demonstrated to work well (Saad and Kruschwitz 2013).

REFERENCES

- Albakour, M.-D., U. Kruschwitz, N. Nanas, D. Song, M. Fasli, and A. De Roeck (2011). Exploring ant colony optimisation for adaptive interactive search. In *Proceedings of ICTIR'11*.
- Alhindi, A., U. Kruschwitz, and C. Fox (2013a). A pilot study on using profile-based summarisation for interactive search assistance. In *Proceedings of ECIR'13*.
- Alhindi, A., U. Kruschwitz, and C. Fox (2013b). Site Search Using Profile-Based Document Summarisation. In *Proceedings of DIR'13*.
- Berkovsky, S., T. Baldwin, and I. Zukerman (2008). Aspect-based personalized text summarization. In *Adaptive Hypermedia and Adaptive Web-Based Systems*, pp. 267–270. Springer.
- Craswell, N., D. Hawking, T. Upstill, A. McLean, R. Wilkinson, and M. Wu (2003). TREC11 Web and interactive tracks at CSIRO. In *Proceedings of TREC-12*.

- Díaz, A. and P. Gervás (2007). User-model based personalized summarization. *Information Processing & Management* 43(6), 1715–1734.
- Dignum, S., U. Kruschwitz, M. Fasli, Y. Kim, D. Song, U. Beresi, and A. de Roeck (2010). Incorporating seasonality into search suggestions derived from intranet query logs. In *Proceedings of WI'10*.
- Diriye, A., A. Blandford, and A. Tombros (2010). When is system support effective? In *Proceedings of IliX '10*.
- Gauch, S., M. Speretta, A. Chandramouli, and A. Micarelli (2007). User profiles for personalized information access. *The Adaptive Web*, 54–89.
- Hahn, U. and I. Mani (2000). The challenges of automatic summarization. *Computer*, 29–36.
- Hassel, M. (2004). Evaluation of automatic text summarization. *Licentiate thesis, School of Computer Science and Communication, Royal Institute of Technology (KTH), published*.
- Hawking, D. (2011). Enterprise Search. In R. Baeza-Yates and B. Ribeiro-Neto (Eds.), *Modern Information Retrieval* (2nd ed.), pp. 641–683. Addison-Wesley.
- Hersh, W. (2002). TREC 2002 interactive track report. In *Proc. TREC*.
- Hersh, W. and P. Over (2001). The TREC-9 interactive track report. In *Proc. TREC*.
- Kato, M. P., T. Sakai, and K. Tanaka (2012). Structured query suggestion for specialization and parallel movement: effect on search behaviors. In *Proceedings of WWW '12*.
- Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. In *Foundations and Trends in Information Retrieval*, Volume 3. Now.
- Koren, J., Y. Zhang, and X. Liu (2008). Personalized interactive faceted search. In *Proceedings of WWW '08*.
- Kruschwitz, U., D. Lungley, M.-D. Albakour, and D. Song (2013). Deriving Query Suggestions for Site Search. *JASIST*. Forthcoming.
- Kruschwitz, U., M. Albakour, J. Niu, J. Leveling, N. Nanas, Y. Kim, D. Song, M. Fasli, and A. De Roeck (2011). Moving towards adaptive search in digital libraries. *Advanced Language Technologies for Digital Libraries*, 41–60.
- Kules, B. and R. Capra (2008). Creating exploratory tasks for a faceted search interface. In *Second Workshop on Human-Computer Interaction (HCIR 2008)*.
- Lin, C. and E. Hovy (2002). Manual and automatic evaluation of summaries. In *Proceedings of the ACL-02 Workshop on Automatic Summarization-Volume 4*, pp. 45–51. Association for Computational Linguistics.
- Maybury, M. and I. Mani (1999). Advances in automatic text summarization.
- Nenkova, A. and K. McKeown (2011). Automatic summarization. In *Foundations and Trends in Information Retrieval*, Volume 5, pp. 103–233. Now.
- Park, S. (2008). Personalized summarization agent using non-negative matrix factorization. In *Proceedings of PRICAI '08*.
- Radev, D., E. Hovy, and K. McKeown (2002). Introduction to the special issue on summarization. *Computational linguistics* 28(4), 399–408.
- Radev, D., H. Jing, M. Stys, and D. Tam (2004). Centroid-based summarization of multiple documents. *Information Processing & Management* 40(6), 919–938.
- Saad, S. Z. and U. Kruschwitz (2013). Exploiting Click Logs for Adaptive Intranet Navigation. In *Proceedings of ECIR'13*.
- Tombros, A. and M. Sanderson (1998). Advantages of query biased summaries in information retrieval. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 2–10. ACM.
- Wang, C., F. Jing, L. Zhang, and H.-J. Zhang (2007). Learning query-biased web page summarization. In *Proceedings of CIKM '07*.
- Yan, R., J. Nie, and X. Li (2011). Summarize what you are interested in: An optimization framework for interactive personalized summarization. In *In Proceedings of EMNLP 2011*, pp. 1342–1351. Edinburgh, United Kingdom, July 27-31.
- Yuan, X. and N. Belkin (2010). Investigating information retrieval support techniques for different information-seeking strategies. *JASIST* 61(8), 1543–1563.
- Zhang, H. and Z. C. W.-y. Ma (1990). A Study for Documents Summarization based on Personal Annotation. *Text*, 41–48.
- Zhang, H., Z. C. W.-y. Ma, and Q. Cai (2003). A study for documents summarization based on personal annotation. *Proceedings of the HLT-NAACL 03 on Text summarization workshop - 5*, 41–48.