

Exploratory Search Using Automatic Text Summaries

Róbert MÓRO*

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova, 842 16 Bratislava, Slovakia
moro@fiit.stuba.sk*

Nowadays, keyword search is a prevalent search paradigm on the Web. This approach works reasonably well for simple information retrieval tasks such as fact finding. Selecting the relevant links and navigating among the documents can, however, be an uneasy task if the information seeking problem at hand is more complex and requires exploring multiple sources to find relevant information, such as researching a new domain. For these types of searches the term *exploratory search* was coined by Marchionini [1]. Exploratory search is characterized by information seeking problems which are open-ended and start with ill-defined information needs; and by processes which can span over multiple search session and require employing different search strategies [4].

In order to support the exploration of the domain by the users, we propose a method of navigation using automatic text summaries. The summaries consist of sentences conveying the most important information of the document; they can reduce information overload by helping the users to decide whether the document is relevant for them and they should read the whole text or not. For the purpose of summary navigation we applied the method of summarization proposed in [2] utilizing the term frequency rater (using tf-idf) and the location rater which prefers terms in the first and the last sentence of the document. We do not summarize the whole contents of the papers from the ACM Digital Library, only their abstracts, thus constructing short generic summaries indicating what the main idea of the paper is.

The actual method of navigation is based on enriching the individual document summaries with the navigation leads; we enrich the summaries in two ways. Firstly, users can manually select a word or a phrase in a summary; popup appears with the possibility to filter the search results by the current selection. After that an explicit link is created; it is visualized in the summary the next time the user comes across the document and can be used in the future searches to filter the results of the user's query.

* Supervisor: Mária Bielíková, Institute of Informatics and Software Engineering

Moreover, we exploit social relations and recommend potentially valuable navigation leads followed by other users based on the collaborative filtering approach. The user can follow the recommended navigation lead, or reject it, thus providing us with an implicit (in the first case), or explicit feedback (in the second case) which we use in computing similarity between users as well as in evaluating candidate lead words.

We realized our proposed method of navigation in the bookmarking system Annota [3] on the dataset consisting of 566 scientific papers that have been bookmarked by the users at the ACM Digital Library over the period of five months.

In order to find out what words users choose to navigate and explore their differences, we conducted a qualitative experiment with five participants. We described them three situations that motivated their search and presented them with the corresponding search results for the initial query. Their task was to select the words they deemed useful for the further exploration of the topic. We verified the hypothesis that the users would select similar words from the summaries if their work task context was similar and yet could benefit from recommending of the leads followed by others.

In the follow-up questionnaire the participants independently agreed that the main advantage of our approach is the easier and smoother navigation which requires less cognitive load, because they can see the words in their context and instantly refine their query to explore the lead.

We plan to extend our approach by recommending the whole navigation trails using the leads in the summaries that will support different exploratory search strategies. Because it is crucial what information is selected into the document summaries we will experiment with our proposed method of summarization to achieve optimal settings of raters utilizing additional information about documents such as the keywords or highlights added by the users.

Extended version was published in Proc. of the 9th Student Research Conference in Informatics and Information Technologies (IIT.SRC 2013), STU Bratislava, 163-169.

Acknowledgement. This work was partially supported by the Slovak Research and Development Agency under the contract No. APVV-0208-10.

References

- [1] Marchionini, G.: Exploratory search: from finding to understanding. *Communications of the ACM*, (2006), vol. 49, no. 4, pp.41–46.
- [2] Móra, R., Bielíková, M.: Personalized text summarization based on important terms identification. In: *DEXA '12: Proc. of the 23rd Intl. Workshop on Database and Expert Systems Applications*, IEEE CS Press, (2012), pp. 131–135.
- [3] Ševcech, J., Bielíková, M., Burger, R., Barla, M.: Researcher activity tracking in digital library of scientific resources enriched by annotations (in Slovak). In: *WIKT '12: Proc. of 7th Workshop on Intelligent and Knowledge Oriented Technologies*, STU Press, (2012), pp. 197–200.
- [4] White, R.W., Roth, R.A.: Exploratory search: beyond the query-response paradigm, Morgan & Claypool, (2009).