EXPLORATION AND ORGANIZATION OF DIGITAL LIBRARY INFORMATION SPACE

Róbert MÓRO, Mária BIELIKOVÁ

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
{moro,bielik}@fiit.stuba.sk

Abstract. When facing a vast information space such as that of a digital library, exploratory search, navigation as well as organization of the documents is always a challenge. We proposed a method of navigation leads to support exploration of the domain by the novice researchers by providing links to related documents within the documents' summaries. In order to provide the researchers with a means of personalized organization of their collections of research articles, we proposed a novel method based on the facet tree. We have evaluated both methods by conducting the user studies in the bookmarking service Annota[1].

1. Introduction

The searches conducted in the digital libraries are usually exploratory in their essence, especially when considering the novice researcher scenario. The goal is not to find a specific document (i.e. a research article), but to learn more about the domain, explore the state-of-the-art, identify the main concepts and relationships between them [5], [9].

The most prevalent approaches used in the digital libraries for their exploration are faceted search or tag cloud navigation. They both provide the users with a global overview of the domain as well as with the means of constructing different partial views. The faceted search and navigation become impractical when there are too many facets or facets value, although this can be partially solved by personalization [8]. Similarly, the contents of the tag cloud can be personalized, e.g. by considering the navigation history of the users [6].

However, the most natural way of navigation seems browsing [7], which does not force users to split their attention between the navigation interface and the search results as does the faceted search or tag cloud navigation.

[1] http://annota.fiit.stuba.sk

The documents found useful by the researchers can be added to their personal collections. There are various personal organizational strategies that can be used, such as piling, filing or structuring [3]. The existing approaches, however, usually support only one strategy and require manual filing of the newly-added resource. In addition, the reorganization of the existing organizational structure is usually very time consuming.

2. Navigation Leads as a Means of Exploration

We proposed a method of digital libraries exploration using the *navigation leads*, which we define as important words extracted from the document and presented in its summary (or abstract) that serve as links to relevant documents [1].

We select the important words to be navigation leads firstly by automatically extracting the terms from the document's summary (abstract) and secondly by computing relevancy score *R(t)* for each of the extracted terms:

$$R(t) = \alpha R_D(t) + \beta R_T(t) + \gamma R_K(t) + \delta R_{TK}(t) \quad (1)$$

where $R_D(t)$ is a relevancy of a term computed based on the document's text, $R_T(t)$ is a relevancy computed based on associated user-added tags, $R_K(t)$ based on keywords added by authors to the document (if present), and $R_{TK}(t)$ based on the top keywords for all the retrieved documents. Linear coefficients α, β, γ, δ are weights of the relevancy components; they can be set empirically based on the desired experimental setup.

The relevancy score thus combines the document's relevancy of the potential navigation lead with its relevancy for the domain, i.e. the whole information space of the documents available in the digital library.

The typical navigation scenario starts with the initial keyword query typed by the user; after the set of results with the navigation leads added to the results' summaries is retrieved, the users can select a lead thus changing their initial query and retrieving a new set of results of related documents.

3. Organization of Resources using a Facet Tree

Typically, when the users (researchers) encounter a relevant document during their navigation session, they can added it to their personal collection of documents for later retrieval or reference.

In order to support automatic filing of the new resources, their fast retrieval and easy reorganization of the collection we proposed a *facet tree method* which allows the users to define hierarchies of the selected facets [2]. The collection is automatically divided into dynamic folders, where each folder represents one facet value; if a new document is added into the collection, it is automatically added into the existing structure based on its metadata.

The facets represent the available metadata types, such as authors, publication year, publication name etc. The important facets are *Keywords* and *Tags*, which allows to organize the collection based on the author-added keywords and user-added tags respectively. In addition, we provide a special *Search* facet that divides the collection into folders based on the user-defined keyword search queries. The documents with the co-occurring metadata values can be clustered, thus providing a better view of the collection.

4. Evaluation

We have evaluated both approaches in the bookmarking service Annota [4] by the means of a user study.

In the first user study we focused on the evaluation of the placement of the navigation leads, comparing the placement in the text of the summaries, under the text and in a cloud of terms. The five participants solved three exploratory search tasks with all of the tested placements in a random order. We measured their time on task and used eye-tracking in order to better understand their navigation behaviour. Additionally, we collected their feedback using questionnaires.

We found out that the tasks were completed 20% faster when using the cloud of terms placement. However, this might have been caused by the fact, that the participants read the abstracts more carefully when the leads were presented within the summaries or under them, as was suggested by our analysis of the eye-tracking data. The participants also appreciated the immediate context of the leads, which was missing when they were presented in the cloud.

In the second user study we focused on the evaluation of the facet tree method of organization of documents compared to the traditional folders [2]. Each of the six participants was supposed to solve four tasks during the study. The first task was to organize the given document collection using the folders and the facet tree, the second to archive a new source in the digital library, the third to locate a resource based on incomplete information about it and the fourth to reorganize the collection.

Based on the user feedback the facet tree method was found superior to the folders in all of the tasks, outperforming the folders in terms of time as well as effort needed to organize and maintain the collection. On the other hand, the most of the participants reorganized the facet tree structure for each task, thus suggesting that a single facet tree is not robust enough to changes. The participants used *Keywords* and *Search* facet the most often. They also used clustering in order to produce larger groups of documents.

5. Conclusions

All of the three navigation leads placements, i.e. within the text of the summaries, under it and in a cloud of terms have their advantages and disadvantages as supported by the finding of our user study. As a compromise, we proposed a presentation of leads under the each document's summary with differentiation of their relevance (e.g. by the used font size). Additionally, in order to preserve their immediate context, the position of the leads within the summary should be highlighted upon mouse hover.

In order to make the facet tree method more robust to changes we proposed a concept of a facet forest, which would allow to maintain various facet trees at the same time, thus providing different views of the collection and eliminating the need to create a new facet tree structure anew for the repeating tasks.

Although we applied the facet tree method only for organization of resources in the digital library, it could be also used in combination with the navigation leads for exploratory search and finding of the new resources.

Acknowledgement: This work was partially supported by the Scientific Grant Agency of the Slovak Republic, grant No. VG1/0675/11 and VG1/0971/11 and by the Slovak Research and Development Agency under the contract No. APVV-0208-10. It was created also with the support of the Research and Development Operational Programme for the project "University Science Park of STU Bratislava", ITMS 26240220084, co-funded by the European Regional Development Fund.

References

to other papers publishing the results that are summarized here

[1] Móro, R.: Navigation in digital libraries aided by summarization (in Slovak). In: *WIKT '13: Proc. of the 8th Workshop on Intelligent and Knowledge Oriented Technologies*. Centre for Inf. Tech., Košice, (2013), pp. 77–80.

[2] Móro, R., Bieliková, M., Burger, R.: Facet tree for personalized web documents organization. In: *WISE '14: Proc. of 15th Int. Conf. on Web Inf. Systems Engineering*, LNCS 8786. Springer, Berlin, (2014), pp. 372–387.

Other references

[3] Henderson, S., Srinivasan, A.: Filing, piling & structuring: Strategies for personal document management. In: *IEEE Proc. of 44th Hawaii Int. Conf. on System Sciences*, IEEE Press, (2011), pp. 1530–1605.

[4] Holub, M., Móro, R., Ševcech, J., Lipták, M., Bieliková, M.: Annota: Towards enriching scientific publications with semantics and user annotations. D-Lib Magazine, (2014), to appear.

[5] Marchionini, G.: Exploratory search: From finding to understanding. *Communications of the ACM*, (2006), vol. 49, pp. 41–46.

[6] Molnár, S., Móro, R., Bieliková, M.: Trending words in digital library for term cloud-based navigation. In: *SMAP'13: Proc. of the 8th Int. Workshop on Semantic and Social Media Adaptation and Personalization*, IEEE CS, (2013), pp. 53–58.

[7] Návrat, P.: Cognitive Traveling in Digital Space: From Keyword Search Through Exploratory Information Seeking. Central European J. of Computer Science, 2(3), 170–182 (2012)

[8] Tvarožek, M.: Exploratory Search in the adaptive social semantic web. *Information Sciences and Technologies Bulletin of the ACM Slovakia*, (2011), vol. 3, no. 1, pp. 42–51.

[9] White, R.W., Roth, R.A.: Exploratory Search: Beyond the Query-Response Paradigm. Morgan & Claypool, (2009).