Utilizing History of Navigation in a Digital Library for Term Cloud Personalization

Samuel Molnár, Róbert Móro, Mária Bieliková

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
xmolnars1@stuba.sk, {moro,bielik}@fiit.stuba.sk

Abstract. Navigation by a cloud of terms is useful in various scenarios, e.g. exploratory search or re-finding of documents. We proposed a method of personalized term cloud navigation considering trending words in users' navigation history as a relevant factor for determining their interests. In addition, we recognize a position of a word in a query to have an important role and rank the list of the documents accordingly. We focus on the domain of digital libraries and evaluate our method in Annota[1], bookmarking and annotation web-based system.

1. Introduction

Considering the shear amount of information available, searching for relevant information and navigating in the information space of a digital library can be a challenging task even for a seasoned researcher and more so for novice ones, such as starting master or doctoral students. They can have a hard time formulating keyword queries, because they lack (at least at the beginning) the needed domain overview and knowledge. Therefore, over the past years several novel approaches to navigation were presented. Some of them aim to enrich the traditional search with navigational support [5], others aim to provide new visual means of the information space exploration [3][7] and sense-making, such as tag clouds.

We focus on the latter and introduce a method for cloud navigation which utilizes navigation history as a source of metadata for personalized browsing of information [1]. Existing approaches usually provide no means of personalization of the cloud's content. Alt-

---

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

hough Gwizdka et al. [2] consider navigation history, it is only for the purpose of visualisation of the users' queries in a short period of time. In addition, in most cases they consider only tags as a source of metadata for the content of the cloud [4] exposing their solutions for a cold start problem when recently added documents with no tags cannot be navigated to.

2. Method of history-based term cloud navigation

We personalize the term cloud navigation on the following levels [1]:

1. *Cloud content* – it is enriched by the terms from navigation histories of all users (i.e. their previous queries), thus reflecting the various paths the users usually take when searching for information

2. *Cloud visualization* – the terms in a cloud are differentiated not only by font size which represents their relevance, but also by colour which reflects their freshness (i.e. how far in the past they were last used)

3. *Ranking of the documents* – we assume that the information need of a user evolves with each navigational step (even if slightly), and therefore we prefer documents which are the most relevant to the user's current need represented by the sequence of terms selected from the cloud

The content of the cloud consists not only of tags, but also of the automatically extracted keywords in order to address the aforementioned cold start problem. Since our method is mainly focused on documents within digital libraries' domain, we use the knowledge of their predefined structure to extract keywords only from relevant parts of documents, e.g. an abstract of an article. In the process of cloud creation, we assess relevancy of each term based on how often they were used in previous navigation sessions as follows:

$$R_w(w) = \log(f_q(w, p)) \cdot count(w, D) \tag{1}$$

where $f_q(w, p)$ is a frequency with which the term w occurs in the queries from time period p, D is a list of all documents in the domain that contain term w and *count* is a function determining the number of occurrences of term w in the list of documents D.

In order to rank the documents retrieved after the user selects a term from the cloud, we compute their relevancy to the current query as follows:

$$R_d(d) = \frac{\sum_{w \in Q \cap W_d} \left(\log(pos(w)) + 1 \right)}{|Q|} \tag{2}$$

where Q is a query, *pos(w)* determines a position of term w in Q and W_d is a list of terms (extracted keywords and tags) from document d.

3. Evaluation and conclusions

We evaluated our proposed method in *Annota* [6], a system for bookmarking documents on the Web, annotating of their content and collaborating over it that is being developed at Slovak University of Technology in Bratislava. At the time of the experiments, it was used by more than 100 users and contained about 4700 bookmarked research articles.

We conducted a series of three experiments [1]. The first two were qualitative; we aimed to find out how the users interact with our system, whether they have a tendency to select terms form history during navigation and collect their feedback on the user interface. The last experiment was quantitative, lasting for a period of one week. During that time, 17 users actively navigated among documents using the provided term cloud. In order to evaluate whether the terms from history are of any value to the users, we disabled their colouring so that they would have been indistinguishable from other terms in the cloud.

We found out that around 30% of users' queries consisted of terms from history which is twice as many as if the users selected terms randomly. For more fine-grained evaluation we examined the relationship between the frequency of selection of cloud terms added from the navigation history and the probability of selecting the history terms if the users would have selected the terms during their navigation sessions randomly with respect to the overall number of terms added from history presented in the cloud. By using statistical Wilcoxon signed-rank test, we verified that the observed differences were relevant. This suggests that the terms from history were valuable for users during their navigation sessions.

Although we focused on the scenario of exploratory search of novice researchers, the proposed method could be modified for re-visitation and re-finding scenario as well. In that case we would consider only history of one user instead of history of all users.

References

to other papers publishing the results that are summarized here

[1] 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.

Other references

[2] Gwizdka, J., Studies, L., Brunswick, N., Bakelaar, P.: Tag trails: Navigation with context and history. *Design*, (2009), vol. 69, no. 2, pp. 4579 – 4580.

[3] Rástočný, K.: Navigation based on adaptive views. *Information Sciences and Technologies Bulletin of the ACM Slovakia*, (2011), vol. 3, no. 2, pp. 104–108.

[4] Solskinnsbakk, G., Gulla, J. A.: Contextual search navigation using semantic tag signatures. In: *i-KNOW'11: Proc. of the 11th Int. Conf. on Knowledge Management and Knowledge Technologies*, ACM Press, (2011).

[5] Ševcech, J.: Related documents search using user created annotations. *Information Sciences and Technologies Bulletin of the ACM Slovakia*, (2013), vol. 5, no. 2, pp. 44–47.

[6] Ševcech, J., Bieliková, 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.

[7] 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.