Query by Multiple Example Considering Pseudo-Relevant Feedback

Adam LIESKOVSKÝ*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
adam@lieskovsky.com

Digital libraries aggregate enormous amount of diverse publications and so they are great centralized resources of information for scientists and researchers. With the increase of available information volume, which is prevalent in the space of digital libraries as well as in the other parts of web, comes hand in hand the difficulty of searching and acquiring documents that are most relevant for the user's query. When using these systems, searching itself is not the primary activity of user; on the contrary the user is trying to use and apply gained information later on as a source for his work. Selection among search results requires user's attention, so it is a general effort and tendency that methods and interfaces used on the web are intuitive and effective.

Nowadays the most popular approach of searching is searching by keywords. Novice users often cannot use it the proper way. Prerequisite for achieving satisfying results is accurate formulation of query by identifying correct keywords and certain level of knowledge of the searched domain. One of the alternative approaches that can overcome mentioned problems is *query by example* which has become fairly popular in the domain of multimedia [1, 2], because of the available metadata, that describes the media. We work with a query by multiple examples, where user selects documents as an examples of relevant results which are further used to formulate the query.

In our work, we propose a method of iterative formulation of a query by multiple examples, which uses user's explicit feedback for the query formulation. At the beginning user starts with single keyword search. We take *n* most relevant (top ranked) documents for user's query and apply pseudo relevance feedback [3], which automates the user's evaluation of results and assumes that these documents might be a good source of metadata, for our further query.

Related documents and their similarities are being judged solely from the aspect of available metadata omitting any further full-text analysis. Beside author, author keywords, year, title and publication we have available metadata about references and citations between bookmarked documents. User can explicitly state and select some of the returned documents as positive examples and impact of these documents' metadata on the query will be increased.

---

* Supervisor: Róbert Móro, Institute of Informatics and Software Engineering

We implemented and evaluate our proposed method in Annota[1] (see Figure 1), which is a system for bookmarking and annotating digital documents. Firstly we would like to test our created interface, with the use of an eye tracker in UX lab with a follow up questionnaire. We also intend to evaluate the effect of pseudo relevance feedback on the results of a query.
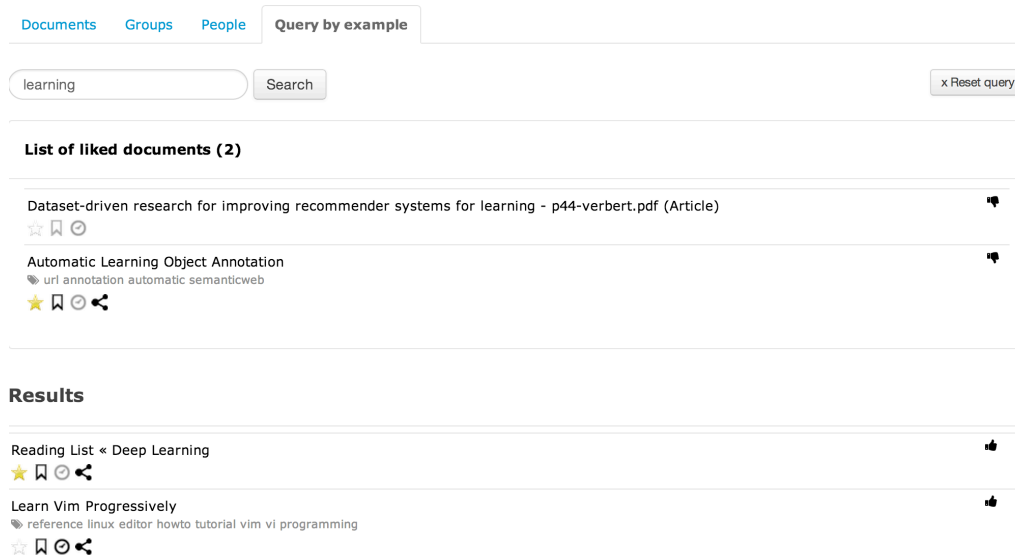
Figure 1. Query by example interface showing explicit feedback in Annota. Documents explicitly selected by a user (by a thumbs-up button) as positive examples are shown in the list of liked documents.

References

[1] Helén, M., Lahti, T.: Query by example methods for audio signals. In: *Proc. of the 7th Nordic Signal Processing Symposium*, (2006), pp. 302–305.

[2] Rasiwasia, N., Vasconcelos, N.: Image retrieval using query by contextual example. In: *MIR '08: Proc. of the 1st ACM Int. Conf. on Multimedia Inf. Retrieval*, (2008), pp. 164–171.

[3] Wu, H., Fang, H.: An incremental approach to efficient pseudo-relevance feedback. In: *SIGIR '13: Proc. of the 36th Int. ACM SIGIR Conf. on Research and Development in Inf. Retrieval*, (2013), pp. 553–562.

---

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