Enhancing Exploratory Search: Graphs, User Modeling and Search History

Jakub ŠIMKO*

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
kubo.simko@gmail.com

We focus on exploratory search (ES), which is oriented toward helping users during learning and investigative search tasks where the goal is not a single web document or fact [1]. ES fights the "information space invisibility problem", best described with typical example of web search, where a user has to guess search keywords rather than picking a query from a list. ES gives users clues at each step of the search session which are produced based on extraction and analysis of information space semantics and presented with respect to the user's personal interests.

We facilitate exploratory search in several ways. First, we develop a novel approach intended to reduce user effort required to retrieve and/or revisit previously discovered information exploiting web search and navigation history (its state of art was well described by M. Mayer [2]). We collect streams of user search actions and identify user agendas (i.e. groups of actions that form sessions with a common user goal, the principle was implemented in various projects [3]). The semantics of each action (text query, facet restriction, visited document) is represented by a term vector, constructed with the help of lemmatization and term extraction API's (WordNet, OpenCalais). Actions are grouped by cosine similarity of their vectors and the time gaps between them using fuzzy rules. Our early experiments of session identification over the AOL corpus show promising results.

Based on the identified sessions, we construct and persistently store visual trees representing session history (see Figure 1). Trees, visible to user, provide an overview of the current (complex) session and improve orientation among visited results. We also provide users with a History Map – a scrutable graph of semantic terms and web resources, constructed by merging individual session history trees, using the Delicious Taxonomy, and the associated web documents (see Figure 2). The History map has full-text search capability over individual history entries and enables navigation throughout the visualized history graph. We evaluate our approach on the Web via supervised and unsupervised live user experiments.

---

* Supervisor: Michal Tvarožek, Institute of Informatics and Software Engineering

Spring 2010 PeWe Workshop, April 18, 2010, pp. 17–18.

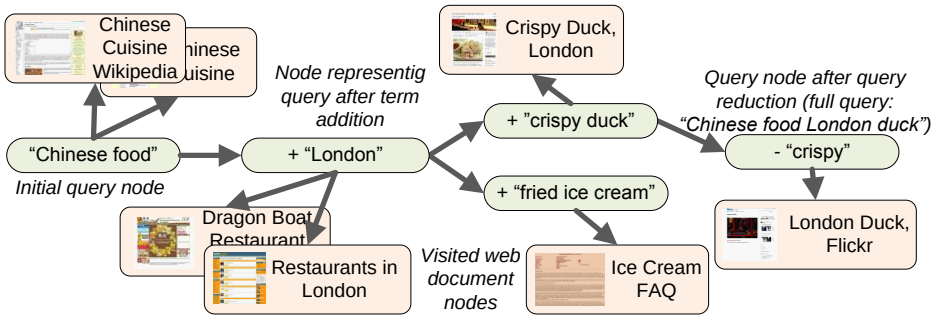

Figure 1. Search History Tree example.

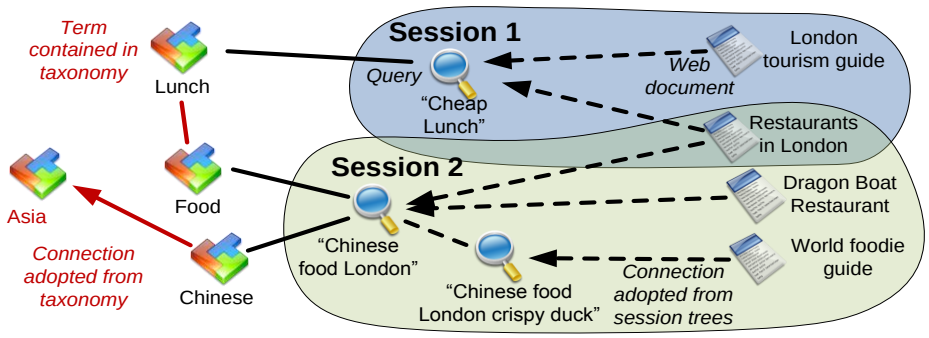

Figure 2. History Map example.

Also, our interest lies in discovering relationships among web documents and terms, that are useful for providing navigation clues. Based on identified search sessions, we map their initial queries to end results and refine suggestions for future occurrences of such queries. We focus on discovering relationships not detected by search engines.

We also discover relations between the terms themselves via a special web search game in which users formulate queries in a specific format to minimize the number of returned results. The query format forces players to use terms that are related together, where multiple occurrence of the same term combinations results in a relationship.

Acknowledgement. This work was partially supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0508/09.

References

[1] G. Marchionini. Exploratory search: from finding to understanding. Communications of the ACM 49 (2006), 41–46.

[2] M. Mayer. Web history tools and revisitation support: A survey of existing approaches and directions. Foundations and Trends in HCI, vol. 2, no. 3, pp. 173–278, 2009.

[3] B. J. Jansen e.al.. Defining a session on web search engines: Research articles. J. Am. Soc. Inf. Sci. Technol., 58(6):862–871, 2007.