Leveraging Social Networks in Navigation Recommendation

Tomáš KRAMÁR*

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

Finding a relevant document based on few keywords is often difficult. Many keywords are ambiguous, their meaning varies from context to context and from person to person. Some words are ambiguous by nature, e.g., a coach might be a bus or a person, other words became ambiguous only after being adopted for a particular purpose. There are also words whose meaning depends on the person who is using them; clearly, architecture means different things to a processor designer than to an architect. Based on the previous observations, we might conclude that using short queries is not a good idea. Unfortunately, this is how we search.

The order of documents provided by the search engine depends on the adopted relevance function; the most widely used search engine today – Google – uses a PageRank relevance function: the more links to a document, the more likely it is to appear at the top positions. This ordering is however not always compatible with user's information needs. We tackle the problem by implicitly inferring the context and modifying the user's query to include it [1].

The overview of the process is depicted in Fig. 1. The user requests a page via proxy (step 1) configured in her browser. Proxy requests the page from the target server (step 2) and extracts the characteristic document features (step 4) – a vector of document keywords, tags from delicious.com and ODP category. Based on user's activity and the extracted features a *social network* is built (step 5), where a weight of an edge denotes a similarity of two users connected by this edge. The stronger is their relationship, the more similar interests they have and the higher is the weight of the edge connecting them. Next, a *community detection algorithm* is run (step 6), to partition the network into clusters of similar users (based on the stream of their activity). The algorithm is designed to take advantage of the weighted relations in the graph and produces overlapping communities, i.e., a user may belong to multiple communities at one time.

* Supervisor: Ing. Michal Barla, Institute of Informatics and Software Engineering

Spring 2010 PeWe Workshop, April 18, 2010, pp. 47–48.

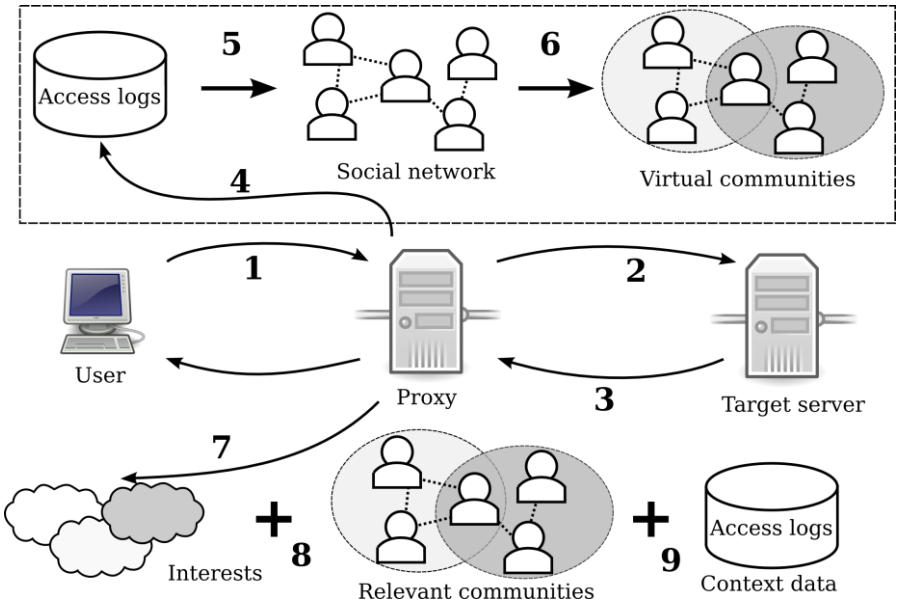

Figure 1. Overview of the query expansion process

In order to identify the *search context*, we need to capture user's current interest (step 7), which in our case is a set of documents features the user is currently interested in. When we detect that a search has been initiated, the current interest helps us to determine all relevant (i.e., sharing at least one feature) communities (step 8). The top *n* matching communities are then considered as the *search context* and passed to the final stage of query expansion.

We use two approaches to infer new keywords, each using the data provided by the members of the communities (step 9). *Query stream analysis* follows a simple observation of how we do our searches. When a search query does not return relevant documents, it is redefined. The redefinition continues unless the user finds the information or gives up. We take all queries issued by users from the *search context* and search for query streams where at least one query matches the user's query. The last query is extracted from each successful query stream and used to enrich the original query. A *keyword co-occurrence analysis* is based on analyzing which additional keywords frequently occur with the words from the query in the documents viewed by the users from the current *search context*. The original query is enriched with the top *n* co-occurring keywords.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 028-025STU-4/2010.

References

[1] Tomáš Kramár, Michal Barla and Mária Bieliková. 2010. Disambiguating search by leveraging the social network context based on the stream of user's activity. *In User Modeling, Adaptation and Personalization - UMAP 2010, to appear.*