

ADAPTIVE QUERY FORMULATION IMPROVEMENT

Ladislav MARTINSKY, Pavol NAVRAT

*Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia*

ladislav.martinsky@gmail.com, navrat@fiit.stuba.sk

Abstract. Adequate query formulation for web search is a very important prerequisite for successful retrieval of desired results. One productive way to improve it that is used in most commonly used search engines is query suggestion. These search engines are quite good at suggesting words to general, highly popular queries, but their suggesting capability deteriorates swiftly when more specific queries are involved. This research proposes an approach to suggest words that are derived from actual search results. The approach guarantees that some words will be suggested for any query, not just for most general or popular ones.

1. Motivation and the idea

Interested fellows, who are just a particular kind of computer users, formulate their information needs to search engines by means of a query. Its interpretation is affected by various factors: length, information richness, type (information, navigational, transactional [3, 5] or context. All these properties can be modified or enhanced using statistical methods [6] or by monitoring interested fellow's behavior [2]. Usability can be enhanced by special interaction techniques [4].

Proper interpreting of a query is greatly complicated by the fact that queries are usually quite short so they convey more often than not too little information. For example, when interested fellow submits "operation" query, how does the search engine "know whether s/he means financial, surgical or mathematical operation? One way to approach this problem is via personalization. This process is based on acquiring and storing interested fellows' profiles or analyzing their actual activities.

Rationale of our approach is based on this simple example case. The interested fellow submits a query composed of very few words, realizing this is just a beginning of the search. For a search engine that cannot guarantee perfect answer in one step it is a good starting point. Its client, the interested fellow understands that the searching process will be an iterative one. Our point is that the search engine, including the shape and organiza-

tion of its user interface, should not behave as if it was able to. The first list of results is not the final answer. It is usually too imprecise, too little recalling at best. At worst, the results may completely miss the intention of the interested fellow. The search engine should act as if continuation of their interaction with the interested fellow was the standard procedure. This requires that the search engine produces some information that will keep the interested fellow involved. Our idea is to suggest query enhancements for the interested fellow to choose from. The interested fellow chooses one of them by a very simple act – just a single click – and hence initiates another round of searching for results. The search engine returns another list of results, hopefully more precise, more recalling, containing documents with information that answers the implicit original information request of the interested fellow. If not, another round in this interaction loop takes place.

2. Our approach

Main contribution of our approach is a way how to find words enhancing the actual query automatically and to suggest a few such alternatives to the interested fellow. Words enhancing the actual query are the result of an analysis of the search results by both the interested fellow and the search engine.

Our approach is based on four operations. Data elicitation gathers and stores words or texts that are related to the query from search results. Data processing transforms those extracts from search results into a set of homogenized word objects, where texts are processed to be free of stop words, punctuation etc. and words are transformed by stemming. The most important operation is data evaluation. The intention is to select some 10+ words. They are to be selected from a few hundred words. We do it using an original set of heuristic rules. Higher evaluation should be given to words that are hopefully more relevant to the original intention expressed by the query. Finally, the selected words are to be presented to the interested fellow. Interested fellow clicks at one of the suggested words (Fig. 1).

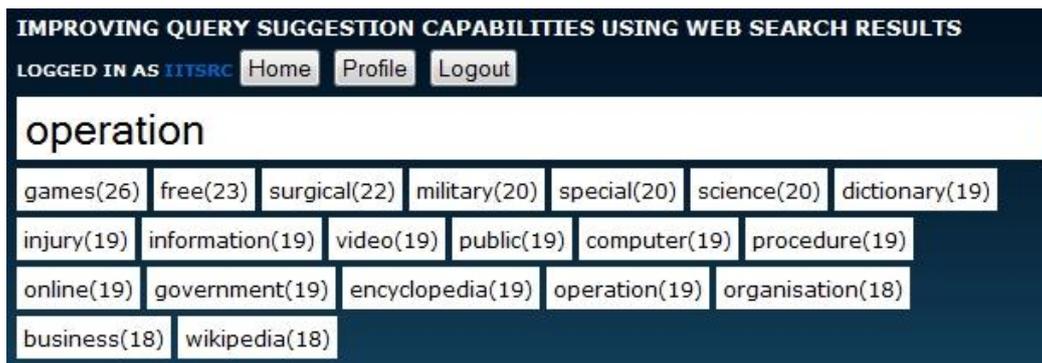


Figure 1. Suggesting selected ordered words to the interested fellow.

We elaborated this idea in several ways. We defined a way to quantify a possible appropriateness of the suggested word. Subsequently, we can order the suggested words. One source for evaluation of appropriateness is the interested fellow's profile. In order to be able to build gradually a profile during a search session, we created a prototype infor-

mation providing agent. The agent submits the query to three common search engines. It analyses results from them. The interested fellow is presented an ordered set of words as suggestions for augmenting the query.

3. Evaluation and conclusions

Our method brings noticeable improvement in offering better answers, but at a cost. The commonly known suggesting search engines are able to make suggestions instantly, our method needs some extra time to allow retrieving search results, analyzing them and producing a list of suggestions. Obviously, this qualitative comparison is slightly misleading, since the common search engines are good in general and frequent queries, whereas we address precisely the complementary kind of queries. The delay is a price for greatly increased completeness of suggestion operation. Our method produces some suggestion for any query, the only limitation being queries for which search engines are not able to return any documents.

We devised a simple experiment to evaluate our method quantitatively. The hypothesis was that (at least some of) the suggested words are indeed relevant to the query viewed as an imperfect expression of interested fellow's intended informational interest. Deciding about relevancy was done by the interested fellow himself/herself. There were 71 interested fellows involved. The results endorse our hypothesis at least partially.

Our approach is different from other approaches in that it often succeeds in suggesting words that refine the meaning of the original query. Let us consider for example an original query "operation". Ordinary search engine suggests "Operation Flashpoint". Our method is able to suggest, among others, also words that help distinguish various alternative meanings of the word, e.g. "military", "financial", "surgical". This could be viewed as a shift towards more sophisticated suggestions.

Our approach has been intentionally geared towards simple solutions. In particular, our current level of personalization is obviously almost trivial comparing with what has already been achieved in that area. We do not employ any persistent interested fellow's model that could be gradually improved. However, even a very simple idea can yield a noticeable improvement in query suggestion, which has been our research goal.

Acknowledgement: This work was partially supported by the Scientific Grant Agency of Slovakia, grant No. VG1/0508/09, by the Agency for Supporting Research and Development supporting the project Tradice, grant No. APVV -0208-10 and it is a partial result of the Research & Development Operational Program for the project PerConIK, ITMS code: 26240220039, co-funded by ERDF.

References

to other papers publishing the results that are summarized here

- [1] Martinský, L., Navrat, P.: Query formulation improved by suggestions resulting from intermediate web search results. *Computing and Information Systems Journal* (2012), vol. 16, no. 1, pp. 56-73.

Other references

- [2] Barla, M.: Towards Social-based User Modeling and Personalization, *Information Sciences and Technologies Bulletin of the ACM Slovakia*, (2011), vol. 3 no. 1, pp. 52-60.
- [3] Cho, J., Lee, U. and Liu, Z. (2005), "Automatic Identification of User Goals in Web Search", in *International World Wide Web Conference*, ISBN:1-59593-046-9: 391-400.
- [4] Kovarova, A. Special Interaction Approaches and their Impact on Usability. *Information Sciences and Technologies Bulletin of the ACM Slovakia*, (2011), vol. 3, no. 3, pp. 14-25.
- [5] Levinson, D. and Rose, D. E. (2004), "Understanding User Goals in Web Search", *International World Wide Web Conference*, ISBN: 1-58113-844-X: 13-19.
- [6] Vanekova, V. (2010), "Preferential Querying for the Semantic Web", *Information Sciences and Technologies Bulletin of the ACM Slovakia*, (2010), vol. 2, no. 2, pp. 137-148.