D3.2 Demonstrator of query assistance functionality

File name	PuppyIR-D3.2-Query-Assistance-Demonstrator-v1.0.doc
Author(s)	TUD, KUL, UT, UG
Work package/task	WP3/ Task 3.2
Document status	final
Version	1.0
Contractual delivery date	01/06/2010
Confidentiality	Public
Keywords	Query assistance, online demonstration
Abstract	The report describes the functionalities offered by the query assistance demonstrator. A description of the underlying data set is additionally given.

Table of Contents

Executive Summary

The report describes the query assistance demonstrator with its functionalities and necessary steps to achieve them. Query assistance is provided by means of:

- A child-friendly user-adaptive interface with personalized news channels
- Query suggestions based on query logs and community expertise
- Visual query suggestions
- Content filtering to remove unsuitable resources

1. Introduction

The aim of the work on D3.2 Query assistance functionality, is to devise a proof of concept and demonstrator of the use of query assistance for children's web search. Recent research on the web search behaviour of children has identified query formulation to be one of the biggest impediments for completing search tasks [2, 3]. Due their low level of literacy finding the proper query terms, relevant synonyms and their correct spelling is very challenging especially for younger children. Offering appropriate query suggestions that are also presented in a child-friendly way is a mandatory step to overcome these obstacles.

Interface design is another central aspect in the creation of web search facilities for children. Current web search engines typically do not consider the specific needs of children [2].

Finally, ensuring that only age-appropriate resources are shown to the user has a high priority in order to prevent children from being exposed to documents that might be either too hard to understand, topically unsuitable or in general not entertaining. To this aim filtering functionality has been designed.

This document outlines the various aspects of the query assistance functionality designed (Section 2), the data sets used for testing and evaluation purposes (Section 3), as well as the functionality realized thusfar and the improvements envisaged for the next stages of the project (Section 4).

2. Functionality overview

In this section we will give a brief overview of the different functionalities combined in this demonstrator to overcome children's typical problems in information search scenarios.

2.1   GUI

The graphical user interface is based on an interaction-based information filtering system for children [4]. The system tries to reduce the need for explicit query-based user input by exploiting user preference information over time.

Figure 1: Query assistance demonstrator GUI

While offering a standard search interface it also presents recent information on topics that the user searched for previously. Instead of repeatedly having to undergo the difficult and potentially error-prone process of typing keywords in order to find new information about her or his favourite singer, the system will feature a dedicated channel of relevant news. Depending on the relative frequency with which a topic was searched for the personal relevance of the topic changes and is expressed for example in the ordering and font size of new information arriving. This way the

user's attention can be directed towards those topics that are most relevant for her or him. There is also the possibility to manually add channels for given topics of interest.

Responsible partner: UG

2.2   Suggestions

As an additional approach to ease the task of handling keyword search interfaces our demonstrator system offers various means of query suggestions to augment and complete user-provided keywords.

2.2.1 Community-based suggestions

The first category of keyword suggestions is based on the knowledge of large user communities on the Internet. The particular example at which we employ community knowledge is the social bookmarking website Delicious [5]. To exploit the Delicious users' judgements we issue the original user query and select a number of relevant document URLs. Afterwards the system looks up these URLs' Delicious tags and tries to find coherent tag clusters of related concepts. Frequent terms from these clusters are finally suggested to the user as extensions to her or his original query.

Responsible partner: UT

2.2.2 Log-based suggestions

As an additional source of suggested query terms we make use of the extensive log files of large-scale web search engines offered for example by Google. They allow us to query an API to find the most frequent user-issued queries extending our original query.

Responsible partner: KUL

2.2.3 Visual suggestions

The intention behind using query suggestions was to provide relevant keywords which the child might not be aware of either due to little experience with keyword-based search interfaces or because the word itself is not known. In the latter case a purely textual suggestion will hardly be helpful as the child will not know the suggested terms' meanings. In order to solve this problem our demonstrator automatically retrieves images for each suggested query [10]. This step enables the child to select an image that shows what she or he meant rather then having to guess the meaning of potentially unknown words. The images are obtained from the Flickr photo-sharing web site.

Responsible partner: KUL

Figure 2: Envisioned query suggestion functionality

2.3   Filtering

The final aspect of the query assistance demonstrator is a filtering step in which those resources that are not suitable for children are filtered out. Suitability is judged along several different dimensions:

1. Complexity of text
2. Child-friendly page design
3. Appropriateness of page topic
4. Child-friendliness of linked web sites
5. Focus towards a children's audience

In this way we try to only promote those pages that satisfy quality criteria defined by recent surveys [1, 6, 7]. The classification step of web pages is done as described in [9].

Responsible partner: TUD

3.   Data collection

While at the current evaluation version we use the same data set that was employed in [8] we decided to build over the next few months a new web corpus that is suited for PuppyIR but can be used more widely than just the project. This corpus is envisioned to be a subset of the so-called ClueWeb collection developed for TREC2009 (http://boston.lti.cs.cmu.edu/Data/clueweb09/), augmented by fine-grained suitability judgements. In this way we hope to obtain a universally usable corpus with high-quality annotations that will prove to be a valuable resource for further stages of the demonstrator as well as the entire PuppyIR project and the research community.

4.   Current demonstrator state and future steps

This chapter briefly summarises the current status of the described demonstrator functionalities as well as the necessary steps to improve demonstrator quality of future versions.

4.1   GUI

The user interface offers the full range of information personalization and highlighting functionalities demonstrated at ACM SIGIR 2010, and described in [8]. The content filtering and query suggestion components described next are not yet integrated, but will be exploited in the next version of the GUI. Figure 2 shows a prototype of this fully integrated demonstrator.

4.2   Log-based suggestions

At the current state the Google service is not integrated in the interface. Doing so while also ensuring a child-friendly way of displaying the suggestions is envisioned for future versions.

4.3   Community-based suggestions

In analogy to the log-based suggestion approach the community-based method is being offered as a web service. The service is implemented and provided at:

http://130.89.11.231:8080/puppyWebForm/QueryDemo?query=

followed by a given query to be expanded. A future addition to this end is the transformation of the web service's hierarchical output into well-known formats such as OSS or general XML.

4.4   Visual suggestions

The selection of descriptive images for a given query expansion term has been implemented and is going to be embedded in the demonstrator interface in the next step.

4.5   Content filtering

Content filtering can be applied to the data collection to filter out the content statically, or be a part in the weighting of how suitable content is. In the current status of the demonstrator, only static collection filtering is applied.

References

[1] Andrew Large, Jamshid Beheshti, and Tarjin Rahman. Design criteria for children's web portals: The users speak out. *Journal of the American Society for Information Science and Technology*, 53(2):79–94, December 2001.

[2] Hilary Hutchinson, Allison Druin, Benjamin B. Bederson, Kara Reuter, Anne Rose, and Ann C. Weeks. How do i find blue books about dogs? The errors and frustrations of young digital library users. In *Proceedings of the 11th International Conference on Human-Computer Interaction (HCII 2005) (CD-ROM). Mahwah, NJ: Lawrence Erlbaum Associates*, 2005.

[3] M. G. Van Kalsbeek, J. J. De Wit, et al. *Automatic Reformulation of Children's Search Queries.* Technical Report TR-CTIT-10-23, Centre for Telematics and Information Technology, University of Twente, Enschede, 2010.

[4] R. Glassey, D.Elliott, T. Polajnar and L. Azzopardi. Interaction-based Information Filtering for Children. To appear In *Proceedings of the Information Interaction in Context Symposium* (IIiX 2010), August 2010.

[5] Delicious – Social Bookmarking. http://delicious.com/, 2010

[6] Jakob Nielsen. Kids' Corner: Website usability for children. http://www.useit.com/alertbox/children.html , May 2010

[7] Sandra L. Calvert. Children as consumers: Advertising and marketing. *The Future of Children*, 18(1):205–234, 2008.

[8] Leif Azzopardi, Desmond Elliott, Richard Glassey, Tamara Polajnar. Puppy, Go Fetch: Prototyping the PuppyIR Framework. In *SIGIR Demonstrations*, 2010

[9] Carsten Eickhoff, Pavel Serdyukov and Arjen P. de Vries. Finding Web Pages for Children. To appear In *Proceedings of the International Conference on Information and Knowledge Management, Toronto, Canada*, 2010

[10] Zheng-Jun Zha, Linjun Yang, Tao Mei, Meng Wang, Zengfu Wang. Visual Query Suggestion. In *ACM Multimedia, Beijing, China*, 2009