Web Search Query Assistance Functionality for Young Audiences

Carsten Eickhoff[1], Tamara Polajnar[2], Karl Gyllstrom[3], Sergio Duarte Torres[4], and Richard Glassey[2]

[1] Delft University of Technology
c.eickhoff@tudelft.nl
[2] University of Glasgow
{tamara,rjg}@dcs.gla.ac.uk
[3] Katholieke Universiteit Leuven
karl.gyllstrom@cs.kuleuven.be
[4] University of Twente
duartes@cs.utwente.nl

Abstract. The Internet plays an important role in people's daily lives. This is not only true for adults, but also holds for children; however, current web search engines are designed with adult users and their cognitive abilities in mind. Consequently, children face considerable barriers when using these information systems. In this work, we demonstrate the use of query assistance and search moderation techniques as well as appropriate interface design to overcome or mitigate these challenges.

1 Introduction

Today, children are frequently consulting the Internet's search facilities to satisfy their personal information needs. However, the current popular web search engines hardly reflect the specific demands of very young users. Throughout the last 10 years, various studies have identified the typical challenges children face when interacting with search engines. In this work we will summarize their findings and show how targeted means of Information Retrieval and interface design can be used to increase success and enjoyment of young searchers. We focus mainly on children between 8 and 12 years as they already show a sufficient degree of literacy skills to operate and understand textual search interfaces. The following challenges have been observed for members of said age group. **(A)** The first and often most frustrating problem for children is the query formulation step [9,3]. Children typically have a far smaller active vocabulary than adults. Therefore, they often do not know the exact term to describe their information needs, or, if they do, they will not always be able to spell it correctly.

(B) Once the query has been issued the child has to identify those results that are relevant to her or his search interest. Children often struggle with the task of understanding extensive textual result snippets. The cognitive load of interpreting and comparing several multi-sentence snippets including gaps, urls, statistics, etc. appears to be substantial, and forms one of the main challenges at

P. Clough et al. (Eds.): ECIR 2011, LNCS 6611, pp. 776–779, 2011.
© Springer-Verlag Berlin Heidelberg 2011

this point. **(C)** Besides the individual length and degree of detail at which every single result is presented, the overall number of retrieved results per page that children can handle is lower than for adults [8]. This is additionally underlined by the fact that children rarely scroll down the result page. Often they are not even aware of the functionality or at least do not use it intuitively. **(D)** A specific habit that has often been observed for children is a preference for browsing over searching [4]. Where adult users often explore a given topic by iteratively respecifying the search query, children tend to browse through the results of the initial query to find the desired pieces of information. **(E)** There is a large proportion of the content offered on the Internet that is not suitable or not understandable for children. Even factoring out erotic content, which state of the art search engines can reliably detect, many pieces of information are potentially harmful for children. Depending on the specific query the proportion of unsuitable material varies strongly. **(F)** Finally, the Internet and especially the search engine niche have experienced growing influences from the advertisement industry. While adult users rarely fall for the advertisers' tricks such as pop-ups or banners, children are less resistant to such marketing methods and require appropriate protection in this domain [10].

2 Functionality Overview

Keeping the previously introduced challenges for children's web search in mind, we propose a system, based on the PuppyIR framework [2], that allows for easy modular combination of web search services with focus on child audiences. The system is based on a conventional web search engine, augmented by additional pre- and post-processing, to alter both, the issued queries, as well as the returned result lists. In the following we will describe the concrete measures taken to address children's specific needs in web search scenarios. We will refer to the challenge indices from Section 1.

2.1 Faceted Query Expansion

Following the idea of faceted content exploration, we provide a means of expanding queries in such a way that they focus on one specific angle of the search topic. The CollAge system [7], introduced the use of media types commonly consumed by children as facets for exploration. Examples of media types are colouring pages, puzzles, cartoons, or games. In order to make these categories more easily understandable for children, a visual representation is given. This is typically an example of the things to be found in the respective category. Query expansions address children's problems with query formulation (A), by providing easy means of redirecting the query's focus without having to manually reformulate the query, which has been found to be frustrating. Additionally, it allows for a better coverage of browsing strategies (D).

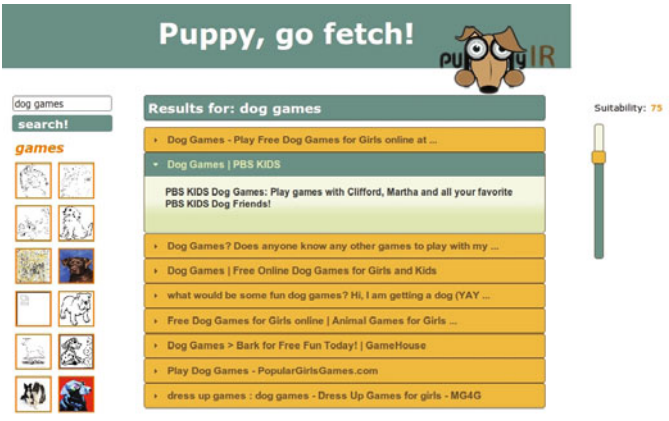


Fig. 1. Query assistance interface

2.2 Community-Created Query Expansion

A second source of query expansion terms is based on human expertise in the form of social bookmarking tags. A given query is issued to a number of independent search engines and for each of them the n top-ranked results are looked up on the social bookmarking platform *Del.icio.us*. The most frequently assigned tags over all retrieved pages are used as query expansion options. In a second step *Wikipedia* and the *ODP* web taxonomy [1] are used to infer high level semantic categories from the tags. These semantic concepts can also be offered as expansion candidates. Similar to faceted query expansion, the community-based approach eases query formulation (A) and, by diversifying the scope of results, enhances the success rate of browsing (D).

2.3 Content Moderation

In order to ensure the suitability of retrieved results for the user's age we offer a result filter, that follows the approach of Eickhoff et al. [5], to automatically estimate each web page's suitability based on a wide range of on-page features. Prominent examples are reading level scores of textual content, language modelling approaches, a high-level syntactical analysis of the way the user is addressed on the page, an estimate of the page's commercial intent as well as an analysis of the page's link neighbourhood. The resulting suitability score ranges from 0 to 1, where 1 indicates a page is definitely suitable for children. A customisable threshold value of this score enables fine-grained system tuning towards the user's personal preferences. In addition to eliminating topically unsuitable results (E), the content moderation step allows the filtering of strongly commercially motivated pages (F).

2.4 Search Interface

A number of potential problems for young users arise from web search interfaces that are mainly designed for adult users. Our interface (see Figure 1) follows the

paradigms detailed by Glassey et al. [6]. In order not to overwhelm the child with the amount of text (B), we only show web page titles that are expandable to show the full original snippet. The number of search results per page is limited as to eliminate the need to scroll (C). A side panel shows possible faceted routes of content exploration. Children can be very sensitive to colourful designs [10]. To improve the user's experience and enjoyment, different graphical styles can be chosen for the interface.

3 Demonstrator Requirements

While the demonstrator can be easily adapted to work based on a closed collection of documents, the query suggestion mechanics, however, rely on on-line services such as Del.icio.us. Therefore, an Internet connection would be required for our demonstration.

Acknowledgement

We would like to express our thanks to Franciska de Jong, Leif Azzopardi, Arjen de Vries and Djoerd Hiemstra for their counsel and advice to this demonstrator. This research is part of the PuppyIR project. It is funded by the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 231507.

References

1. The Open Directory Project (2010), `http://www.dmoz.org/`
2. Azzopardi, L., Glassey, R., Lalmas, M., Polajnar, T., Ruthven, I.: PuppyIR: Designing an Open Source Framework for Interactive Information Services for Children (2009)
3. Borgman, C.L., Hirsh, S.G., Walter, V.A., Gallagher, A.L.: Children's searching behavior on browsing and keyword online catalogs: the Science Library Catalog project. JASIS 46(9) (1995)
4. Druin, A., Foss, E., Hutchinson, H., Golub, E., Hatley, L.: Children's roles using keyword search interfaces at home. In: CHI 2010. ACM, New York (2010)
5. Eickhoff, C., Serdyukov, P., de Vries, A.P.: Web Site Classification on Child Suitability. In: CIKM 2010, Toronto, Canada (2010)
6. Glassey, R., Elliott, D., Polajnar, T., Azzopardi, L.: Interaction-based information filtering for children. In: IIiX 2010. ACM, New York (2010)
7. Gyllstrom, K., Moens, M.F.: A picture is worth a thousand search results: finding child-oriented multimedia results with collAge. In: SIGIR 2010, Geneva, Switzerland (2010)
8. Large, A., Beheshti, J., Rahman, T.: Design criteria for children's Web portals: The users speak out. JASIST 53(2), 79–94 (2002)
9. Moore, P.A., St George, A.: Children as Information Seekers: The Cognitive Demands of Books and Library Systems. School Library Media Quarterly 19(3), 161–168 (1991)
10. Nielsen, J.: Kids corner: Website usability for children. Jakob Nielsens Alertbox (2002)