

Recommendations to Enhance Children Web Searches

Shahrzad Karimi
Department of Computer Science
Boise State University
Boise, ID 83725 USA
shahrzadkarimi@u.boisestate.edu

Maria Soledad Pera
Department of Computer Science
Boise State University
Boise, ID 83725 USA
solepera@boisestate.edu

ABSTRACT

We present the initial design and development of KidsQR, a query recommendation system tailored exclusively for children. KidsQR aids children in their quest for online information by considering children vocabulary, child-friendly phrases, and entities children are familiar with. Initial experiments conducted based on the assessment of parents and elementary school teacher appraisers verify the promising performance of KidsQR.

Keywords

Information retrieval, query recommendation, children

1. INTRODUCTION

Children represent a noticeably increasing group of Internet users [5]. The amount of time that children spend online has significantly increased in the past years [3]. However, finding the appropriate online resources that are interesting from a child's viewpoint is a challenge. Consequently, children looking for online resources are constantly exposed to material that is unsuitable and non-relevant. A crucial requirement of conducting a successful search using search engines is the formulation of a well-defined query [3]. Unfortunately, creating an appropriate query that leads to retrieving relevant information is not an easy task for child users. Previous studies suggest that the relative performance of queries meant to retrieve information for children is poorer than the queries intending to retrieve content for non-children users [4]. The reason lies in children's limited vocabulary, their tendency to use natural language constructions and complex phrasing for expressing their information needs, and their lack of ability to utilize keywords in phrase formulations [5]. A query recommendation system can help these users "formulate elaborated information needs" [3], by providing queries that these users can apply to initiate a search.

Despite the large number of studies conducted in the field of query recommendation, relatively few focus explicitly on the young group of Internet users and their difficulties. Consequently, literature pertaining query recommendation for children is very limited. In fact, most of the existing query recommendation systems are designed based on the information needs of adults [4], which is why they suggest queries that often do not lead to retrieving online resources that "suit the characteristics of content for children" [4]. Existing query

recommendation systems targeting children have taken different approaches, using large-scale query logs, tags, biased random walk methods, and bipartite graphs [4, 5]. These approaches, however, are based on texts that are generated by adults, disregarding informal phrasing based on children writing.

We deem the formulation of keyword queries that children can relate to as the solution to this problem. With that in mind, we have developed KidsQR, a query recommendation system that suggests keyword queries in response to a child-initiated query. Unlike previous works, we will not primarily rely on child-related data produced by adults. Instead, we attempt to consider the patterns of children's informal phrasing and natural language by utilizing texts that have been written by children, to recommend queries that are adequate to initiate the search of content of interest to children, which can lead to a more child-friendly and suitable search experience. KidsQR is unique since it considers child-friendly characteristics to generate query recommendations, including children vocabulary, phrasing patterns, pop-culture, and the popularity of the terms among children. Our intention is to recommend queries that have a closer resemblance to a child's search intent, which results in retrieving suitable documents.

2. METHODOLOGY

In this section we present a brief overview of KidsQR.

Generating Candidates. To identify possible queries to be recommended, i.e., candidate queries, in response to a given user's initial query, KidsQR employs Ubersuggest.org¹. Ubersuggest is a query generation tool that provides hundreds of possible suggestions given an initial user query and offers topical diversity among the suggestions. We have verified that phrases provided by Ubersuggest include terms related to children pop culture, such as the name of cartoon characters.

Analyzing Candidates. To determine the adequacy of the candidates being recommended to the user, i.e. distinguishing child-friendly candidates from the non-child-friendly ones, KidsQR evaluates each candidate query based on a number of child-related characteristics that are applicable to them. In other words, KidsQR considers child-related properties to determine how closely a candidate phrase is related to children's interests, or if the candidate relates to child-friendly content. The properties/characteristics to be observed in quantifying the degree to which a candidate query is likely reflecting children's search intent are described as follows:

- **Vocabulary.** A fundamental step to differentiate child-friendly queries among the candidate ones, is to examine the existence of children's vocabulary terms in each query. We consider children vocabulary lists extracted from children dictionaries and schools' academic vocabulary (such as www.opsu.edu/www/education/BuildAcademicVoc.pdf and

¹ While we used Ubersuggest for development purposes, other tools, such as keywordtool.io, can be considered as well.

kids.wordsmyth.net/we/) and prioritize candidates that include keywords frequently occurring in children pre-defined vocabularies. We do so, since it is anticipated that children will favor queries including keywords they are familiar with. For example, for the queries “color” and “city,” the candidates “coloring pages” and “pig in a city” are preferred over “color spectrum” and “city infrastructure” since “spectrum” and “infrastructure” are not common words among children.

- **Popularity.** The popularity of terms among children is considered by analyzing term frequency distributions² on children stories, poems, and blog posts. Candidate queries including popular children terms are also given precedence.
- **Phrase-Formulating.** Examining the child-friendliness of individual terms in candidate queries is crucial, but not sufficient in confirming the appropriateness of a candidate query since it does not consider the query phrase as a whole. For example, having the words “bar” —as in “chocolate bar”—and “open” in children vocabulary does not imply that “open bar” is a child-related phrase. We consider stories and poems written for children, as well as texts, blog posts, and online reviews written by children, to determine the appropriateness of the combination of the words, and capture children’s informal phrasing patterns. Candidate queries that have similar patterns to children’s informal phrasing behavior, or are child-appropriate as a phrase, most likely address a child’s search intention, hence, are prioritized.
- **Pop-Culture.** We observed that candidate queries that do not include children vocabulary, or do not literally make sense as a phrase, can still be related to children’s popular culture. KidsQR examines candidate queries in the context of children pop-culture and prioritizes queries including terms related to children’s movies, songs, and toys (extracted from Pixar.com and Allmovie.com, to name a few). For example, “Mary Poppins” and “Mr. Potato Head” are valid candidates since they refer to a movie character and a toy, respectively, even though the former contains “Poppins”, a word not included in children’s vocabulary, and the latter consists of child-related words but does not have a literal meaning as a phrase.

Ranking. KidsQR analyzes each of the candidate queries based on the characteristics mentioned above, and prioritizes candidates that (i) are simple, (ii) refer to children’s topics of interests, (iii) include terms children are familiar with, and (iv) resemble children’s informal phrasing behavior. KidsQR relies on a multiple regression analysis model that simultaneously considers the different contributing factors in determining whether a candidate query is, in fact, child-friendly and generates a single ranking score for each candidate query recommendation. The top-N candidates are presented to the user as the corresponding query recommendations that can help capture his search intent and guide the online search process.

3. INITIAL EXPERIMENTS

As far as we know, a benchmark dataset that specifically addresses queries conducted by children has yet to be developed. Thus, we created our own dataset by conducting a user study and collecting data from 10 appraisers who were either parents of children between the ages of 3 and 12, or elementary school teachers. We presented each appraiser with 8 queries and the corresponding set of query recommendations, comprised of

² Sample sources considered for determining term popularity and phrase suitability include kidsblogclub.com and storybud.org.

randomly-positioned recommendations generated by Google, Bing, and KidsQR. Appraisers were then asked to select the two recommendations that they found most child-friendly for each query and their selections were treated as the gold standard.

Using the created dataset we evaluated KidsQR based on *Mean Reciprocal Rank* (MRR) and *Normalized Discounted Cumulative Gain* (NDCG). We also compared the performance of KidsQR with that of Google and Bing, two well-known search engines that offer query recommendations and that are frequently used by children [1]. As shown in Table 1, KidsQR outperforms both Bing and Google, in terms of MRR and NDCG. The higher NDCG implies that queries useful for children are positioned higher in the ranking of recommended queries by KidsQR. The higher MRR indicates that, on average, users of KidsQR need to scan through less query recommendations before locating a suitable, useful one than users of other systems.

Table 1. Performance analysis for Bing, Google, and KidsQR

System	MRR	NDCG
Google	0.36	0.51
Bing	0.27	0.35
KidsQR	0.7	0.72

4. CONCLUSION

We have developed a query recommendation system, KidsQR, designed specifically to address the challenges of children in query formulation. KidsQR distinguishes the child-friendly query candidates from the non-child-friendly ones by simultaneously considering multiple desired properties on children queries.

We aim to further enhance the initial development of KidsQR so that it can adequately handle informal as well as natural language phrasing which are very common among children. We also intend to further enhance the performance of KidsQR by addressing children pop-culture more comprehensively. We believe the more aspects of children’s pop-culture that we consider, the more closely we can predict a child user’s search intention, i.e. recommend queries that are anticipated to be appealing from a child’s perspective can be generated. Moreover, we will examine children vocabulary and words provided by school vocabulary lists more accurately and consider the age gap among young children, i.e., we will group children by age groups and explicitly consider their reading ability in making query recommendations for children in the respective groups.

5. REFERENCES

- [1] D. Bilal & M. Boehm. Towards New Methodologies for Assessing Relevance of Information Retrieval from Web Search Engines on Children’s Queries. *QJRM*, 1:93-100, 2013.
- [2] S. Duarte Torres, D. Hiemstra, I. Weber and P. Serdyukov. Query Recommendation for Children. In *ACM CIKM*, pp. 2012-2014, 2010.
- [3] S. Duarte Torres and I. Weber. What and How Children Search on the Web. In *ACM CIKM*, pp. 393-402, 2011
- [4] S. Duarte Torres, D. Hiemstra, and P. Serdyukov. An Analysis of Queries Intended to Search Information for Children. In *IIIX*, pp. 235-244, 2010.
- [5] S. Duarte Torres, D. Hiemstra, I. Weber, and P. Serdyukov. Query Recommendation in the Information Domain of Children. *JASIST*, 65(7): 1368-1384, 2014.