Looking for the Movie Seven or Sven from the Movie Frozen? A Multi-perspective Strategy for Recommending Queries for Children

Ion Madrazo Azpiazu, Nevena Dragovic, Oghenemaro Anuyah, Maria Soledad Pera
People and Information Research Team (PIReT)
Boise, Idaho
ionmadrazo,nevenadragovic,oghenemaroanuyah,solepera@boisestate.edu

ABSTRACT

Popular search engines are usually tuned to satisfy the information needs of a general audience. As a result, non-traditional, yet active groups of users, such as children, experience challenges composing queries that can lead them to the retrieval of adequate results. To aid young users in formulating keyword queries that can facilitate their information-seeking process, we introduce *ReQuIK*, a multi-perspective query suggestion system for children. *ReQuIK* informs its suggestion process by applying (i) a strategy based on search intent to capture the purpose of a query, (ii) a ranking strategy based on a wide and deep neural network that considers both raw text and traits commonly associated with kid-related queries, (iii) a filtering strategy based on the readability levels of documents potentially retrieved by a query to favor suggestions that trigger the retrieval of documents matching children's reading skills, and (iv) a content-similarity strategy to ensure diversity among suggestions. For assessing the quality of the system, we conducted initial offline and online experiments based on 591 queries written by 97 children, ages 6 to 13. The results of this assessment verified the correctness of *ReQuIK*'s recommendation strategy, the fact that it provides suggestions that appeal to children and *ReQuIK*'s ability to recommend queries that lead to the retrieval of materials with readability levels that correlate with children's reading skills.

CCS CONCEPTS

• **Information systems → Query suggestion**; *Query intent*; • **Social and professional topics → Children**; • **Human-centered computing →** *User studies*;

KEYWORDS

Query suggestions; children; search intent; dataset

ACM Reference Format:
Ion Madrazo Azpiazu, Nevena Dragovic, Oghenemaro Anuyah, Maria Soledad Pera. 2018. Looking for the Movie Seven or Sven from the Movie Frozen? A Multi-perspective Strategy for Recommending Queries for Children. In *Proceedings of 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18)*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3176349.3176379

CHIIR '18, March 11–15, 2018, New Brunswick, NJ, USA

ACM ISBN 978-1-4503-4925-3/18/03...$15.00
https://doi.org/10.1145/3176349.3176379

1 INTRODUCTION

As one of the largest communities that search for online resources, children are introduced to the Web at increasingly young ages [9]. However, popular search tools are not explicitly designed with children in mind nor do their retrieved results explicitly target children. Consequently, many young users struggle in completing successful searches, especially since most search engines do not directly support, or offer weak support, for children's inquiry approaches [16]. As stated in [45], this is an important issue to address since early experiences influence skill development in making proper use of resources for personal and educational growth.

As described in the book Search Engine Society, "Children growing up in the 21^{st} century have only ever known a world in which search engines could be queried, and almost always provide some kind of an answer, even if it may not be the best one" [20]. Even though children, as inexperienced users, struggle with describing their information needs in a concise query [12], they still expect search engines to retrieve relevant information in response to their requirements, or at least suggest better choices for a successful search. As part of their capabilities, search engines often suggest[1] queries to aid users in better defining their information needs. In fact, a recent study conducted by Gossen et al. [17] shows that children pay more attention to suggested queries than adults. Unfortunately, these suggestions are not specifically tailored towards children and thus need improvement [44]. While multiple query suggestion modules have been developed to automatically generate queries that capture users' needs [3, 50], only a small number of them specifically target children. To address this problem, along with a need for more children-related tools, we introduce *ReQuIK* (**Re**commendations based on **Qu**ery **I**ntention for **K**ids), a query suggestion module tailored towards 6-to-13 year old children.

The main goal of *ReQuIK* is to provide query recommendations that explicitly consider diverse and ambiguous users' information needs. Prior to generating recommendations for a given child-initiated query, *ReQuIK* takes advantage of the *search intent module* presented in [9] , which is used to capture the intended meaning of the query. In doing so, *ReQuIK* can deal with long natural language queries or queries that include common patterns children use when searching the web, which are difficult for search engines to process and properly handle. Even when the search intent of a query is identified, it is not enough to trigger the retrieval of suitable materials for each user, since the interests of children can vary depending of their age. To capture a wide range of potential suggestions, *ReQuIK* emulates a popular *query generation* strategy. Thereafter, *ReQuIK* identifies suitable suggestions by using a multi-perspective approach based on raw text analysis and a number of textual traits specifically associated with children content. These traits analyze

[1] *Suggestion* and *recommendation* are used interchangeably in this manuscript.

usage of children words, popular culture terms, entities and diminutives in queries. By applying a multi-perspective approach based on deep learning, the proposed query suggestion module is able to learn distinctive characteristics that portray adults and children queries, and use that knowledge to predict which queries are the most child-friendly among the ones in a candidate set. To guarantee diversity among the recommended queries, *ReQuIK* uses a content similarity strategy that groups together queries that are topically similar and excludes suggestions that refer to the same topic, i.e., queries that would retrieve the same type of resources. We are aware that suggested queries could retrieve resources that children may not easily comprehend due to their high reading difficulty [36]. In order to minimize this situation, *ReQuIK* prioritizes query suggestions that will potentially lead to easier-to-read resources.

Due to the lack of datasets that capture children search activities, we dedicated research efforts to creating one. To do this, we deployed an ad-hoc search framework which interacts with the Bing *api* and facilitates the archival of search sessions. We conducted an experiment with 97 children, ages 6 to 13, who were given research/informational and factual search tasks by their teacher. As a result, we gathered close to 600 queries, which we used for development and evaluation purposes. Thereafter, we verified the validity of *ReQuIK* based on both offline and online experiments, using the aforementioned dataset, which demonstrate that not only does *ReQuIK* suggest queries that are children oriented, but it also leads to resources that are of the adequate reading level.

To the best of our knowledge, *ReQuIK* is the only available system that can be coupled with existing engines to generate query recommendations for children, favoring those that can lead to easier-to-read resources. *ReQuIK* suggests queries that initiate the retrieval of child-related topics and materials, which can lead to improving search engines' performance. The design of the proposed tool explicitly considers different patterns children use while searching the Web to adequately capture the intended meaning of their original queries. For example, if a child submits the query *"elsa"*, *ReQuIK* aims to prioritize query suggestions such as *"elsa coloring papers"* or *"elsa dress up games"* that correlate better with topics of interest to children rather than *"elsa pataky"*, as suggested by Google[2], which is more appealing to mature users. Other contributions of our work include (i) a novel ranking model inspired by a deep-and-wide architecture that, while successfully applied for ranking purposes [7, 49], has never been used in the query suggestions domain, (ii) a strategy to overcome the lack of queries written by children by taking advantage of general purpose children-oriented phrases, and (iii) the aforementioned newly created dataset, which will be made available to the research community[3].

2 RELATED WORK

Creating an appropriate query that leads to retrieving relevant information is challenging for young users. Previous studies state that the performance of a web search engine is poorer when retrieving documents in response to queries targeting information for children than for queries oriented to content of more traditional users [42]. Query recommendations can help children by providing queries that can be used to initiate a search.

Current research on children's query suggestions is limited, with a simple query to ACM Digital Library for "children query suggestions" or "children query recommendations" retrieving five distinct research works from among the top-20 results. Existing research includes the one conducted by Duarte et al. [12], who rely on a bipartite graph constructed using tags and URLs to suggest children queries. The authors enhanced their proposed strategy, as discussed in [42], by considering topical and language modeling features, such as a topic-sensitive Page Rank and a children-related vocabulary distribution. Besides examining tags assigned at Delicious to retrieve web pages, Eickhof et al. [13] consider high-level semantic categories (inferred from Wikipedia and the DMOZ.org taxonomy) associated with tags, and treat them as expansion terms. The aforementioned approaches, however, rely primarily on tags to make their suggestions, which can be poorly defined due to the lack of quality control on user tags which can be inherently noisy. Furthermore, these tags are often provided by adults, instead of children, which explains why the vocabulary used to describe online resources for children might not correlate with the terms used by children. The work conducted by Vidinli and Ozcar [46] focuses on suggesting queries in within an educational search environment. The proposed strategy analyzes a number of features to determine the most suitable queries, among the candidates, that should be recommended, given a child-generated query. Unfortunately, the majority of the features are computed as a result of query-log examination, which is a constraint as query logs generated by K-12 students are rarely accessible.

To the best of our knowledge, the studies done in [24, 38, 47] are the only ones that do not use tags or query logs for generating children query suggestions. Instead, the authors in [38] use bigrams extracted from websites that contain text generated by children and Simple.Wikipedia.org, a collection of documents written for users whose second language is English. As opposed to the strategy in [38], which depends upon a pre-defined set of topical categories, *ReQuIK* relies on a dynamic clustering to ensure diversity of recommended queries. The module presented in [47] creates query suggestions that are semantically different but conceptually similar to a child-initiated query. To do so, the authors consider result set overlap generated by pairs of queries (a given query and a possible suggestion) and term overlap between the queries, and prioritize suggestions including n-grams in Simple Wikipedia or include terms in a pre-defined children vocabulary.

Even if with a different purpose, the work of Eickhoff et al. [14] is similar ours, given that they also aim at distinguishing between children and non children content. In their work, the authors use aesthetic features of websites as discriminators of children-related content, features that are not useful for classifying queries.

Similar to the approach in [24], *ReQuIK* emulates Ubersuggest's query generation strategy to create a set of queries to recommend. However, while the query suggestion strategy discussed in [24] depends on a regression model that combines multiple features, such as children vocabulary, phrasing patterns, popular culture terms related to children, and the popularity of the terms among children, *ReQuIK* adopts a multi-perspective suggestion approach that considers a different and larger set of traits to infer if a query is child-related, as well as content of the query itself. What distinguishes *ReQuIK* the most, among its counterparts, is its ability to simultaneously combine text pattern analysis as well as varied query traits to identify suitable child-related query suggestions.

[2] As verified on May 19, 2017.

[3] The dataset can be found in https://doi.org/10.18122/B2WQ5T.

Algorithm 1 Generating Query Suggestions

```
Input: A query Q written by a child, a trained ranking
model RM, wordId dictionary WD, number of suggestions to
generate k
candidates = empty set
scoredCandidates = empty list
suggestions = empty list
Q' = searchIntent(Q)
candidates = generateCandidates(Q')
for each candidate CQ in candidates do
    features = generateFeatures(CQ)
    wordIds = getWordIDs(CQ,WD)
    score = calculateScore(wordIds, features, RM)
    scoredCandidates = scoredCandidates + <CQ, score>
end for
sort(scoredCandidates)
for each candidate <CQ,score> in scoredCandidates do
    if suggestions is empty then
        suggestions = CQ
    else
        if CQ is not similar to any query in suggestions and
readability(CQ)<8 then
            suggestions = suggestions + CQ
        end if
    end if
    if |suggestions| ≥ k then
        break
    end if
end for
```

3 REQUIK

In this section, we describe the design of *ReQuIK* (see pseudocode in Algorithm 1). Along with the description of each strategy used in *ReQuIK*, we provide a step-by-step example (denoted **R.I.A**, ReQuIK in action) using Q_E: *"I want the trol song"*, a query written by a child, which is also part of the sample introduced in Section 4.1.1. This running example aims to further showcase the practical application of *ReQuIK*. Note that for Q_E, Google neither offers possible query suggestions nor leads to the retrieval of resources that explicitly target younger audiences, as illustrated in Figure 1, which further aids our case in advocating for the existence of query recommendation strategies solely for children.

3.1 Search Intent

As described by Bilal et al. [5], to adequately serve children, search engines must address the fact that children are seldom successful in formulating succinct queries. In fact, researchers have observed that children tend to use long (natural language) queries, as opposed to keyword queries, when performing online searches [10]. Unfortunately, the longer the query, the less likely a web search engine is to retrieve relevant resources in response to it, which can be very frustrating for young users [10]. Furthermore, children tend to misspell words and use writing patterns that differ from those of adults. For example, children can include the word "amazzzzing" instead of "amazing" in a query to emphasize something is really amazing. To best satisfy children needs, *ReQuIK* relies on *QuIK*, the search intent module for children presented in [9], which addresses common patterns detected in children writing, including: use of diminutives, exaggerated and trendy terms as well as higher percentage of misspellings when compared to adult users. In doing so, it transforms an initial query Q into a simplified keyword query Q' that (i) captures the information need meant to be expressed by a child and (ii) can be adequately processed by search engines.

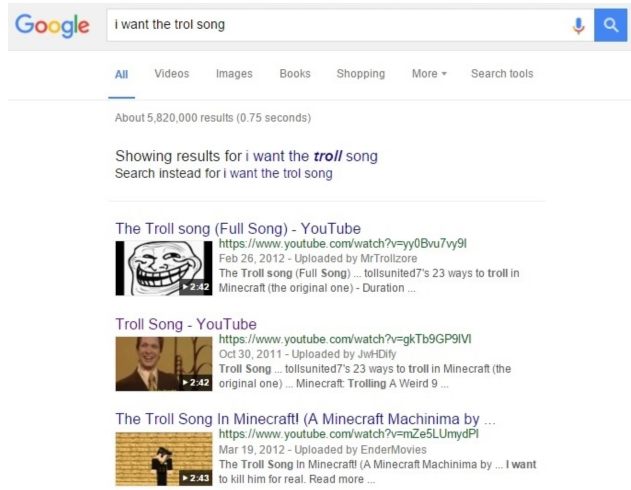

Figure 1: Screenshot of documents retrieved by Google for the query *"I want the trol song"*

R.I.A. The search intent module employed by *ReQuIK* transforms Q_E into Q'_E "troll song", which suitably captures the intended meaning of Q_E, i.e., troll song from the popular movie Frozen. The search intent module solves two problems observed in Q_E: (i) removing terms that are superfluous for capturing the meaning of the query, i.e. *"I want the"*, and (ii) fixing the spelling error on *"trol"*.

3.2 Candidate Generation

Having identified the information need of a young user expressed in a query Q and created a shorter, more concise query Q', *ReQuIK* generates a set of candidate queries, i.e., queries that could possibly be suggested to a user, by emulating the algorithm of Ubersuggest[4] [1], a popular query suggestion tool based on Google auto complete API. The advantage of adopting such a strategy is that it quickly finds keywords based on what users search for on the Internet, creating multiple possible queries [4]. Bypassing the use of static query logs or probabilistic models allows *ReQuIK* to offer up-to-date candidate queries, since the aforementioned auto-complete strategy is constantly updated by online search trends.

R.I.A. Submitting Q'_E to the candidate generator creates more than 200 candidate suggestions, including queries related to children, such as *"troll song from dora"* and *"troll song frozen"* as well as queries that seem intended for more mature users, such as *"troll song no copyright"* and *"troll song hitler"*.

3.3 Ranking model

Not all the candidate queries generated in Section 3.2 are necessarily targeted towards children' needs, reading abilities, and interests. Consequently, to identify suitable suggestions among the candidates, *ReQuIK* takes advantage of a novel model (see Figure 2) that we created by combining two architectures, a *deep model* and a

[4]Ubersuggest queries Google autocomplete API multiple times with the initial query followed by each letter of the alphabet in order to retrieve multiple query candidates.

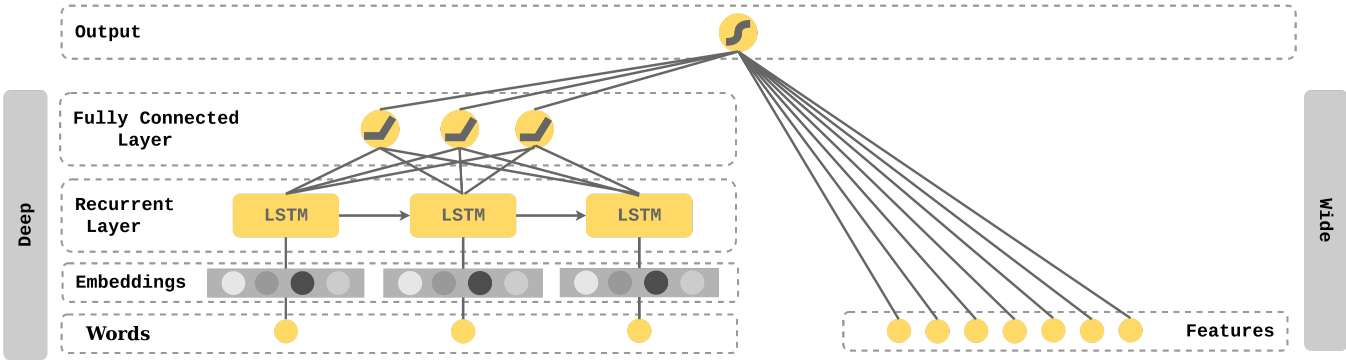

Figure 2: Ranking Model Architecture

wide model, inspired by the app recommendation model recently developed by Cheng et al. [7]. We discuss below insights and benefits of each architecture when applied to child-related tasks, as well as how we combine them into a single model for ranking purposes.

3.3.1 Wide Model: Learning what makes a query child-like.

The wide model incorporates a set of manually-created features that are meant to capture traits often observed on child-related queries. These features are a result of an extensive analysis of children-related sentences (sampled from online sources for children). The wide model is composed of a vector $x_{feats} \in \mathbb{R}^f$, consisting of f features. Those features along with their computations are shown in Table 1, while their descriptions are provided below. Due to space constraints, Table 1 only includes the final set of features considered by *ReQuIK*. However, it is worth mentioning that we conducted an empirical analysis to identify the best subset of features to be used as a part of the ranking process. For example, features based on readability levels calculated by readability formulas such as Gunning FOG [2] and Dale-Chall [6] were overlooked in favor of the features based on an enhanced version of Spache [39] readability formula, which more adequately captures the level of difficulty of web resources.

Trendy Terms: Queries generated by young users contain a large amount of children-related trendy terms, such as names of movies, popular singers, computer games etc. By examining the presence of such terms in a candidate query Q, *ReQuIK* is able to determine if Q is appealing to children or not. (See Table 1-A.)

Entities: Based on a conducted analysis on queries generated by children, we identified a considerable use of entities (e.g., person, location, organization, etc.). For this reason, *ReQuIK* examines the presence of entities as one of the criteria to help decide how likely Q is related to children's interest. (See Table 1-B.)

Children Dictionary: In developing *ReQuIK*, we created a children dictionary using text collected from child-related websites. The use of such a collection of terms is of crucial importance since it best describes appropriate terms children use. Based on this, we concluded that young users use a narrower and unique vocabulary when expressing their needs. Therefore, we analyze the average frequency of terms in Q that are included in our children dictionary to enable *ReQuIK* to decide if the Q is tailored towards adults or children. Note that, this dictionary will be made available to the research community on this project's website upon the publication of this manuscript. (See Table 1-C.)

Flesch-Kincaid: Another criteria that targets a simplicity of Q is based on the well-known readability formula, i.e., Flesch-Kincaid [15], which provides the grade level of Q and enables *ReQuIK* to decide if a young user can comprehend Q. (See Table 1-D.)

Enhanced Spache: As a complement to the Flesch-Kincaid readability assessment, *ReQuIK* uses a feature based on an enhanced Spache formula. As an enhancement of a regular Spache [39] to obtain a greater accuracy, we expanded existing Spache dictionary of common words with children dictionary previously created. The combination of two dictionaries was necessary for considering more reliable and up-to-date trends of kids' vocabulary. This criteria increases the level of detection of words used by children in Q and enables *ReQuIK* to differentiate adult and child-related candidate queries. (See Table 1-E.)

Difficult Terms: *ReQuIK* uses a criteria generated based on the frequency of non-children [5] terms in Q to further separate adult and child-like queries. (See Table 1-F.)

3.3.2 Deep Model: Learning from text.

Manual analysis of queries can lead to identifying distinctive features such as the ones mentioned in Section 3.3.1, however, some patterns can be impossible to detect upon simple observation or empirical analysis. Deep learning enables learning directly from raw data, so that new, unexpected features can be inferred automatically allowing the model to grasp patterns that humans are unable to find. The deep neural network is composed of one input layer and 3 hidden layers. The output layer is shared with the wide model and is therefore described in Section 3.3.3. Each of the other layers are described as follows:

Input Layer. The input of the neural network is represented as $x_{words} \in \mathbb{Z}^k$ where x_{words} is a vector k identifies representing a sequence of words. After analyzing the distribution of length of the queries we gathered, we fixed k at 15, a sufficient length to capture 95% of queries in our sample.

Embedding Layer. The embedding layer's role is to convert each word identifier into a dense representation that will capture the semantics of the word. For doing so we define an embedding function Q : wordId $\rightarrow \mathbb{R}^\alpha$, where α is the embedding size, that converts a word id to an embedding of length α. This function is based on a lookup table $S \in \mathbb{R}^{v \times \alpha}$, where v represents the vocabulary size. The embedding function Q returns a row from S

[5] We treat as "non-children" terms that do not appear or have low frequency on our children dictionary.

Feature ID	Feature Name	Description
A	Trendy Terms	$tt(q) = \frac{\sum_{i=1}^{\vert q\vert} t(q_i)}{\vert q\vert}, t(q_i) = \begin{cases} 1, & \text{if } q_i \text{ is recognized as a trendy term} \\ 0, & \text{otherwise} \end{cases}$ To determine if q_i is a child-related trendy term, *ReQuIK* examines its existence in children related pages on well-known websites, such as Amazon.com and CommonSenseMedia.org.
B	Entities	$et(q) = \frac{\sum_{i=1}^{\vert q\vert} e(q_i)}{\vert q\vert}, e(q_i) = \begin{cases} 1, & \text{if } q_i \text{ is recognized as an entity by Stanford NER tool [35]} \\ 0, & \text{otherwise} \end{cases}$ To determine if q_i is an entity, *ReQuIK* uses well-known CoreNLP tools.
C	Children Dictionary	$ct(q) = \frac{\sum_{i=1}^{\vert q\vert} c(q_i)}{\vert q\vert}, c(q_i) = \begin{cases} 1, & \text{if } q_i \text{ is included in a children dictionary} \\ 0, & \text{otherwise} \end{cases}$ We created our own children dictionary, comprised of 100,000 non-stop lemmatized terms, extracted from texts retrieved from a sample of various children-related websites.
D	Flesch-Kincaid	$FK(q) = 0.39\left(\frac{totalwords}{totalsentences}\right) + 11.8\left(\frac{totalsyllables}{totalwords}\right) - 15.59$ In our case *total words* is the total number of words in q_i and *total sentences* is always equal to 1.
E	Enhanced Spache	$ES(q) = (0.121 \times AvgLengthOfq) + (0.082 \times NumberOfUniqueUnfamiliarWordsInq)$ We implemented Enhanced Spache formula by updating the existing dictionary of common words with the children dictionary in C.
F	Difficult terms	$dt(q) = \frac{\sum_{i=1}^{\vert q\vert} d(q_i)}{\vert q\vert}, d(q_i) = \begin{cases} 1, & \text{if } q_i \text{ is recognized as a difficult term} \\ 0, & \text{otherwise} \end{cases}$ *ReQuIK* treats q_i as a difficult term if it is not included in the children dictionary, or Spache dictionary of common words, or trendy terms list in A

Table 1: Criteria description, where q represents a query, q_i is the i^{th} non-stop word, lemmatized term in q, and $|q|$ is length of non-stop-words in q. Each computed criteria score is normalized to fit a 1-5 scale

that corresponds to the provided word id. This function is applied to all the word ids in the input sequence creating a new matrix $H_1 \in \mathbb{R}^{k \times \alpha}$ that will be the input of the next layer. The matrix S is initialized using a random uniform distribution within $[-1, 1]$ and will be trained together with the weights of the neural network.

Recurrent Layer. Recurrent neural networks have been successfully used for processing sequential information [33]. A text document can be seen as a sequence of words, where each word depends on information provided by previous words, making it adequate for a recurrent neural network. The third layer of *ReQuIK*'s ranking deep neural network takes advantage of Long Short Term Memory cells (LSTM) [21], a recurrent cell specially suited for textual documents given its capability to remember long term information. Given a word embedding and a state vector $l_s \in \mathbb{R}^{\beta}$, where β refers to the number of LSTM units, each LSTM cell generates a output $l_{out} \in \mathbb{R}^{\beta}$. These outputs are concatenated to create a vector $h_2 \in \mathbb{R}^{\beta * k}$, which will be the input of the next layer.

Fully Connected Layer. A fully connected layer is one of the most common layers in a neural network. This layer computes a weighted sum over all the outputs of the previous layer. More precisely, given the vector h_2 produced by the LSTM layer, this layer computes the following operation:

$$h_{deep} = \text{relu}(Wh_2 + b) \tag{1}$$

where $h_{deep} \in \mathbb{R}^{\gamma}$ is the output of this layer, $W \in \mathbb{R}^{(\beta * k) \times \gamma}$ is a matrix of weights, $b \in \mathbb{R}^{\gamma}$ a bias vector. γ is a parameter that determines the number of neurons in the layer and *relu* refers to Rectified Linear Unit [32] which corresponds to the activation function that is applied to the result of the weighted sum.

3.3.3 Output Layer: Combining both models. The last layer of the ranking model is the one responsible for combining the aforementioned deep and the wide models. This enables *ReQuIK* to incorporate the benefits of both a wide and a deep model, being capable of learning patterns from text automatically, while also using human crafted features that consider traits related to children queries. For doing this we first concatenate h_{deep} and x_{feats} to create a new vector $h_{comb} \in \mathbb{R}^{\gamma + f}$. Similar to the fully connected layer in the deep model, a weighted sum of all the values is computed, to create an output:

$$y' = \text{sigmoid}(W_{comb} h_c comb + b_{comb}) \tag{2}$$

where $y' \in \mathbb{R}^{c}$ is the prediction of the neural network, $W_{comb} \in \mathbb{R}^{(\gamma + f) \times c}$ is a weight matrix, $b_{comb} \in \mathbb{R}^{\gamma + f}$ a bias vector, c the number of prediction classes (2 in our case) and a sigmoid as activation function. Note that the prediction vector y' is composed of real values, enabling to use the same model for both prediction and as a scoring function for ranking the candidate queries.

3.3.4 Training. To produce relevant predictions, a neural network needs to be trained. This process involves fitting several variables that include, weights, biases and embedding values. For fitting those values a loss function is minimized using input/output pairs from a training dataset. For doing so, we take advantage of Cross Entropy function as a loss function, defined as follows:

$$H_{y'}(y) = -\sum_i y'_i \log(y_i) \tag{3}$$

where y' is the prediction created by the neural network and y is the target ground truth using one-hot encoding. For minimizing the error, we took advantage of the Adaptive Movement Estimation (Adam) [31] optimization technique.

The performance of a neural network is affected by its parameters. To identify the optimal α, β, and γ, we sweep possible values and found that the best combination for this task is $\alpha = 128; \beta = 128; \gamma = 128$. Thus, this is the parameter set used in all the experiments reported in this paper.

R.I.A. Using its ranking strategy, *ReQuIK* assigns a rating to each of the candidate query recommendations generated for to Q'_E. Based on the predicted ratings, queries like *"troll song from dora"* and *"troll song frozen"* are prioritized over queries such as *"troll song no copyright"*, which more likely better capture topics of interest for more mature audiences.

3.4 Readability

We observed that among the list of top-N child-related queries, not all of them lead to the retrieval of documents matching the reading skills of 6 to 13 year old children. This is of high importance for the recommendation process, since children are not able to comprehend resources above their reading capabilities. To determine what queries will most likely retrieve documents with suitable reading levels, *ReQuIK* applies the Flesch-Kincaid formula [15]. Queries that lead to the retrieval of resources associated with readability levels that are greater than readability levels expected for a child are excluded from the list of queries to be suggested[6]. Note that we treat 8 as an appropriate children grade level since it corresponds to reading level of 13 year old user.

Relying on this filtering strategy enables *ReQuIK* to identify suitable suggestions based not only on query content itself, but also on the readability levels of documents that would potentially be retrieved using those queries to initiate the search process.

R.I.A. ReQuIK further filters candidate queries to ensure that recommendations shown to its users most likely trigger the retrieval of documents they can understand. Based on the average readability scores of the top-N documents retrieved in response to *"troll song frozen"* and *"troll song no copyright"*, which are 6.7 and 11.3, respectively, *ReQuIK* retains the former and excludes the latter from the set of possible query recommendations.

3.5 Diversity

To guarantee that generated suggestions cater to diverse user interests, *ReQuIK* excludes from the set of top-N query recommendations candidate suggestions that are, to a degree, similar to each other. *ReQuIK* applies the Semantic Similarity algorithm developed by Yuhua Li et al. [34], which provides WordNet-based scores that are used to determine if any two suggestions are semantically the same, i.e., would trigger the retrieval of similar resources. In this context, two suggestions are treated as similar if their similarity score is above a threshold. By applying this topical filtering strategy, we select top-k[7] diverse suggestions from the ranked list generated in Section 3.4. For doing so, the first suggestion S_1 is always included in the final set of suggestions. Each subsequent candidate suggestion S_n is compared to the suggestions already in the final set. Thereafter, S_n is included only if it yields similarity scores of at most 0.7[8] with respect to the already-selected suggestions.

R.I.A. In the last step of its process, *ReQuIK* selects not only highest-rated and readability-level suitable queries, but also queries that offer topical diversity and thus target a wide range of users. For example, using its similarity-based filtering, *ReQuIK* treats *"troll song frozen"* and *"troll song frozen movie"* as highly similar and excludes the latter from the set of suggestions to be presented to the users. Lastly, the final set of suggestions generated by *ReQuIK* in response to the initial child-query Q_E includes *"troll song frozen"* and *"troll song from dora"*. These suggestions not only capture different information needs but do so in a keyword fashion that enable search engines to retrieve more relevant and suitable results for 6-13 year olds.

[6] To determine which candidate query cq potentially retrieves resources with reading levels above the expected ones, we compute and average the reading levels of the top-3 resources retrieved in response to cq.

[7] We set k = 4 to emulate the number of suggestions often offered by search engines.

[8] We experimentally set 0.7 as the similarity threshold.

4 EXPERIMENTAL RESULTS

In this section, we detail the results of the offline and online studies conducted to demonstrate the correctness of *ReQuIK*'s methodology and the relevance of its generated query recommendations.

4.1 Evaluation Framework

We discuss below the datasets and query suggestion strategies used for comparison purposes.

4.1.1 Dataset & Other Resources. Obtaining children-related data is not a simple task, especially due to children protection regulations that make sharing this type of data highly restrictive. While other researchers have used queries extracted from existing query logs like the popular AOL [11], the extraction strategy may be limited and not always identify queries written by children, as some writing patterns are common among both children and adult queries. To address this issue, given that neither datasets that can be used to evaluate query recommendations for children nor query logs comprised only of children queries are publicly available, we created our own dataset[9].

Search Environment. In order to construct a dataset, we developed an online search framework that emulates the behavior and appearance of Google, and enables us to gather search sessions[10] of children and archive information such as: query typed, selected query suggestions (if available), clicked URLs, and timestamps. We made the appearance of this framework similar to that of a popular search engine, given children's preference to use well-known engines, as opposed to the counterparts designed for them, e.g., KidzSearch. The search framework contains an initial page that a teacher can configure based on the grade of his/her class.

Gathering Queries. In order to get children to use the search framework, we collaborated with elementary schools in the Idaho (USA) area. We asked K-9 teachers to propose their students information discovery tasks, for which they used our framework to create queries. For determining the type of tasks we followed the strategy used by Gwizdka et al. [19] and included both: research/informational tasks such as *Find information about fire belly toads* or *Find information about tigers* and factual tasks such as *How long do toads live* or *When does summer start.* Each teacher started the class with specific questions, however, children were later allowed to find information about things of their interest. A total of 97 children between the ages of 6 and 13 participated in the study, generating 591 unique queries.

Even though *ReQuIK*'s objective is to recommend children queries, for development and assessment purposes we also required a set of non-children queries. Thus, the resulting dataset, denoted ReQ_{qs}, includes the aforementioned 591 queries labeled *children-queries* and 591 queries randomly selected among the ones in the Yahoo's query log dataset [48] labeled *adult-queries.*

[9] The process of gathering and archiving children queries was supervised by the Institutional Review Board at Boise State University in order to ensure children-related ethical and legal concerns were met.

[10] Information obtained from a child's search session is anonymous.

Other Data Sources. Due to the limited amount of children query-logs in ReQ_{qs}, we gathered ReQ_{corp}, a sample of 1,061,666 sentences for development and training purposes. In creating ReQ_{corp}, we extracted sentences from websites oriented to children, including Dogo [8], Spaghetti [40], Toy Insider [43], Raising Children [37], Kidzvuz [30], Kids-in-mind [23], and Edutaining-kids [28]. We also included in ReQ_{corp} sample sentences from Wikipedia in order to provide negative, i.e., non-children, examples.

4.1.2 Comparison Strategies. In the following sections, we discuss the results of a number of experiments conducted to demonstrate the need for a query suggestion modules specifically designed with young users in mind and showcase the correctness and effectiveness of *ReQuIK*. To give context to such results, we compared the performance of *ReQuIK* with that of a number of baseline and state-of-the-art strategies. We examine (the queries suggested by) *Google*, *Yahoo*, and *Bing*, given that children favor well-known search engines when performing information discovery tasks [5]. Given that we argue for the need of techniques tailored for children, we also consider a number of search engines designed exclusively for children, including: *AskKids*[27], *Kidzsearch*[29], *Ipl2*[22], *KidRex*[26], and *SweetSearch* [41]. We also include in our experiments CQS [38] (discussed in Section 2), which is a state-of-the-art alternative for generating queries tailored to children.

4.2 Usefulness of the Search Intent Module

In addition to the results reported in [9], which demonstrate the performance of the search intent module, we conducted an experiment for measuring the impact this module has for query recommendation. For doing so, we submitted each of the child-written queries in ReQ_{qs} to a number of popular search engines. Thereafter, we determined for which of these queries, the corresponding search engine provided suggestions.

As shown in Table 2, there is a consistent decrease in the number of children queries that can be handled by search engines, in terms of providing suggestions given a child-initiated query. This fact is evidenced in both commercial search engines and state-of-the-art systems such as CQS. Unlike its counterparts, *ReQuIK* is able to provide suggestions for 94% of the queries considered in this study, which is a clear indicator that taking advantage of a search intent module is beneficial not only to enhance the document retrieval process [9] but also the query recommendation process. While in Table 2 we report results on search engines favored by children, it is important to mention that those percentages remain so, even on search engines designed specially for children. This is anticipated, since many of the search tools for children are powered by Google's safe search (e.g., Ask.forkid.co) or do not offer query suggestions (e.g., Cybersleuth-kids.com and Kiddle.co). CQS is only able to generate suggestions for 57% of the children-generated queries in ReQ_{qs}, due to its candidate generation strategy that struggles with identifying suggestions for queries longer than four-grams.

Google	Bing	Yahoo!	Kidsearch	CQS	ReQuIK
46%	36%	65%	76%	57 %	94%

Table 2: Percentage of queries that trigger a recommendation in each compared system

4.3 Effectiveness of the Recommendations

As previously mentioned, there is a lack of datasets comprised of children queries and gathering and sharing children data to create such datasets is not trivial due to time and privacy concerns. Unfortunately, queries on ReQ_{qs} are not sufficient to train a deep neural network (i.e., ranking) model. In order to amend this issue, we train *ReQuIK*'s ranking model using children and non children oriented sentences in ReQ_{corp}. We hypothesize that these sentences are similar enough to a search query, permitting the modeled knowledge to be transferable to the query suggestion context. Therefore, to showcase the correctness of *ReQuIK*'s ranking strategy, we conducted two experiments. The first one measures the performance of the model in predicting whether a sentence is oriented for children or adults, and the second one measures how well this knowledge is transferred to the task of ranking children query suggestions.

	Wide	Deep	Wide and Deep
Accuracy	0.68	0.92	0.94

Table 3: Performance of diverse ranking strategies

For the first task we trained the model described in Section 3.3 using sentences from ReQ_{corp}. Note that sentences from Wikipedia were randomly sampled to make it comparable to that of the children data sources, which resulted in an evenly balanced dataset. We used a 10-cross-fold-validation framework for computing prediction accuracy, which was measured and averaged for each fold. This procedure was applied using the 3 different submodels with the aim of demonstrating the validity of combining both the wide and deep models. As reported in Table 3, the wide and deep model outperforms all its counterparts with a statistically significant improvement using a pairwise t-test with a confidence value $p < 0.05$.

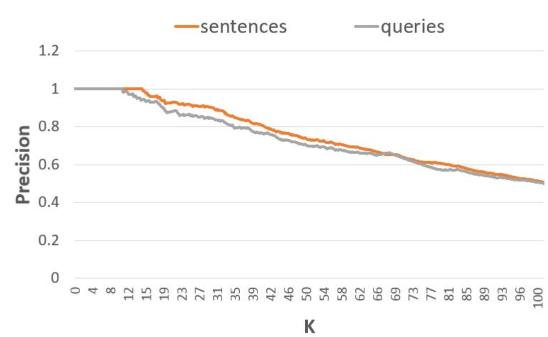

Figure 3: *Precision@K*, **where** K **is defined as the percentage of queries and sentences analyzed. We define** K **as a percentage, as the raw counts of sentences and queries are not comparable otherwise**

In order to demonstrate how this model can be translated to a more real query suggestion ranking environment we conducted an experiment where we ranked both sentences from ReQ_{corp} and queries from ReQ_{qs} using the model trained on sentences of ReQ_{corp}. Figure 3 illustrates the *Precision@K*, where *Precision* is measured as the ratio of children queries/sentences among the K considered and K indicates the proportion of ReQ_{qs} and ReQ_{corp} examined. As shown in the figure, the model achieves similar results for both the sentences and the queries, proving that a model trained over sentences can be translated to the context of ranking queries.

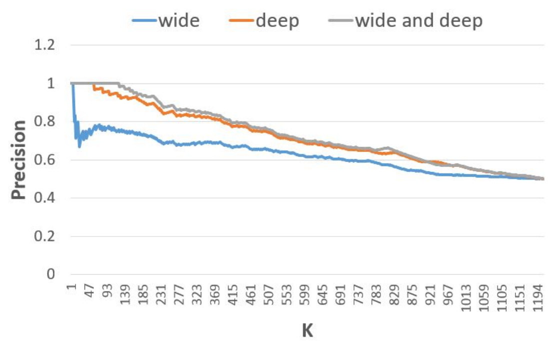

Figure 4: Model assessment based on *Precision@K*, where *K* is the number of queries examined

To further demonstrate the validity of each submodel (i.e., wide, deep, and wide and deep), we conducted an experiment using each of them. Figure 4 illustrates the *Precision@K*, in this case measured as the ratio of children queries among the top-k analyzed, for $K \in$ 1..1194. The wide and deep model outperforms both other models, followed closely by the deep model. The first adult result appears in position 123, 59, and 9, for the combined model, the deep model and the wide model respectively. Even if the suggestions of the wide model might look poor, Figure 4 illustrates that it complements the deep model improving the overall ranking.

It is worth mentioning that upon manual examination, we noticed that most of the queries in ReQ_{qs} labeled as "non-children" and ranked high by *ReQuIK* refer to content that could have been searched by children. For example, *"naruto shimpuden cheats"* ranked in 64^{th} position by the deep and wide model refers to a videogame that could be equally of interest for young or mature audiences, and thus it is treated by *ReQuIK* as a false positive.

4.4 Suitability of Retrieved Documents

We conducted another experiment to verify the suitability of the documents retrieved in response to queries suggested by *ReQuIK* and to highlight the need of readability based filtering. For conducting this experiment, we generated suggestions for the child queries in ReQ_{qs} using popular search engines, children search engines, and *ReQuIK*. We treat as *suitable* documents that have readability levels matching the reading skills expected of children ages 6 to 13.

As children"systematically go through retrieved resources and rarely judge retrieved information sources" [18] we averaged the readability scores computed for the top-3 documents retrieved in response to each of the top-2 query suggestions generated by each evaluated strategy. For measuring readability, we use the well-known Flesch-Kincaid formula [15], which considers various textual features, such as term sentence length.

The results in Table 4 show that documents retrieved by *ReQuIK*'s recommendations are in general easier to read and understand than the ones retrieved by other search engines, even those that explicitly target children, or CQS [38]. This is evidenced by the fact that the average readability score of documents retrieved by recommendations of popular search engines is above 11 in all cases, while for*ReQuIK* is 7.7. This correlates with the grade level of a 13

ReQuIK	Google	Bing	Yahoo!	CQS
7.71	12.46	19.96	11.42	11.82
AskKids	**Ipl2**	**Sweet Search**	**KidRex**	**Kiddle**
13.3	10.9	12.3	12.7	12.83

Table 4: Average readability of top-3 documents retrieved for test query recommendations

year old child, usually in 7th or 8th grade, for which the results have been filtered for, demonstrating the benefits of a readability filtering strategy as part of the query suggestion process.

We are aware that Kiddle does not offer query suggestion, however, being the premier search engine for children [25], it needed to be part of this comparison. For this reason, we analyze the readability level of the top results retrieved by Kiddle for the children queries in ReQ_{qs}. Kiddle achieves an average readability of 12.83 on retrieved document, which is comparable to the one obtained by the engine that powers it, Google. These results provide further evidence for the need for query recommendation tools for children that lead to the retrieval of resources children can read and understand.

4.5 Online Assessment of ReQuIK

To further validate the performance of *ReQuIK*, we conducted an online survey intended to quantify the effectiveness of *ReQuIK* from a user's perspective. For doing so, we included in the survey 10 queries randomly selected among the ones written by children in ReQ_{qs}. These queries were not included among the ones used to train our wide and deep model presented in Section 3.3.3. To ensure diversity among sampled queries, we included unigrams and n-grams queries. We also considered question-type queries, which tend to be used by children. For each sampled query, we generated top-N suggestions using *ReQuIK*, a popular commercial search engine (Google), a popular children search engine (Kidzsearch), and CQS [24], the system discussed in Section 2, which explicitly provides children query suggestions. Note that some of the sampled queries were misspelled, as they are written by children. We purposefully left these queries as they were originally written, in order to be able to consider the effect of misspelled query terms and measure its impact on the suggestions offered by each strategy.

For each sampled query, we selected the top-2 suggestions from among the list of suggestions provided by each strategy, including *ReQuIK*. These suggestions were randomly merged into a list, which was then presented to a group of independent appraisers: teachers. Following Institutional Review Board guidelines, we recruited a cohort of teachers from 5 schools in Boise, Idaho. This gave us a total of 11 teachers that participated in this study. We considered teachers as ideal appraisers for this experiment, given that they know what children are particularly interested in and are trained on the needs of children; making them capable of offering knowledgeable judgments. In the survey, we used a practical scenario, which prompted the teachers to select the best two suggestions for each query that would lead a child to locate child-friendly and interesting web-pages from their view-point. We treated their selections as the *gold standard*.

We collected 181 responses for the survey (i.e., 10 teachers selected at most two suggestions for each of the 10 sampled queries), a snapshot of which is shown in Figure 5. For performance analysis we use accuracy of each query suggestion strategy S, computed as the fraction of suggestions selected by a teacher for a given query over the total number of suggestions generated by S. Based on the analysis of the collected responses, which we illustrate in Figure 6, we observe that appraisers favored suggestions created by search engines and recommendation modules that target children, i.e., *ReQuIK* and Kidzsearch, rather than the ones generated by general purpose engines, i.e., Google. We should mention that except for the case of unigram queries, *CQS* is consistently outperformed by the remaining considered strategies. We believe this is caused by the "novelty" of the majority of the queries included in the survey, which refer to contemporary topics of interest to children (such as names and characters in recent Disney movies), when *CQS* suggestions are based on pre-trained probabilistic models that may not account for the probability of occurrence of such query terms.

We observed that for queries that could have been formulated by either children or more mature audiences, Google has an advantage. For example, for the query *"how do seahorses swim"* suggestions offered by Google were preferred almost exclusively over suggestions generated using any of the counterparts in this study. We hypothesize that this is due to the fact that Google's suggestions are based on common query formulations, which in this case leads to better suggestions, as they do not necessarily have to focus on the retrieval of child-related resources. On the other hand, we observed that for other queries, bias towards children content, as in the case of Kidzsearch, *CQS*, and *ReQuIK*, positively affected the suggestion-selection process. For example, for the query *"Elsa"*, the top-2 suggestions generated by using Google were *"Elsa Pataky"* and *"Elsa Hosk"*, which are names of two popular celebrities. The suggestion most likely to be preferred by a child–as per teacher responses–is *"Elsa and Anna"*, which was not included among the top-2 suggestions of Google, but was presented as the top suggestions using *ReQuIK*.

ReQuIK achieved the highest accuracy, as it consistently offered child-friendly suggestions for unigram, bigram and n-gram queries. Based on a paired t-test, the improvement in the *overall* performance of *ReQuIK* with respect to each of its counterparts is statistically significant with $p < 0.05$.

5 SCOPE AND LIMITATIONS

We highlight below limitations we encountered in designing *ReQuIK*; which we could not address due to the project scope.

ReQuIK is a query recommendation system for children. The project is intended to explore the personalization aspect of the query recommendation task for children. Therefore, common tasks performed in traditional query recommendation systems, such as candidate generation or topic filtering do not suppose a novelty by themselves, and thus, are not extensively explored in this paper. We are also aware that both Flesch-Kincaid and Spache are simple readability formulas, which sometimes lack precision, an issue that is usually highlighted when the text is short. This constraint for short documents is common for all readability formulas and to solve it is out of scope for this project.

ReQuIK is a system that is meant to integrate and complement existing search engines, as opposed to function as a standalone tool. Consequently, tasks that are usually performed by the search engine itself, such as document retrieval and ranking, are omitted from the assessment of this project.

Due to project scope, the presented online experiment is meant to gather initial feedback on the quality of *ReQuIK*'s suggestions, as the experiment is based only on a small set of sampled queries. Given the promising results, we will conduct further online assessments.

Finally, user interaction and perceived usability of *ReQuIK*, as well as its effectiveness in terms of offering suggestions that lead to child friendly websites, are beyond the scope of this work, but will be addressed as future work (see Section 6).

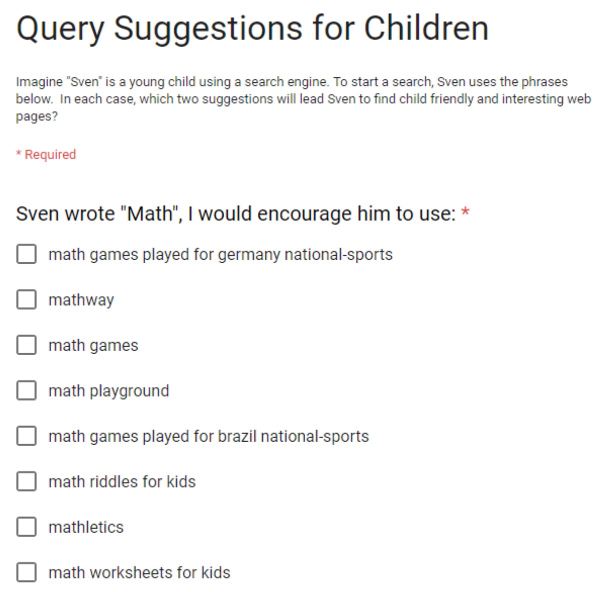

Figure 5: Snapshot of online survey presented to teachers for overall assessment of *ReQuIK* and other strategies

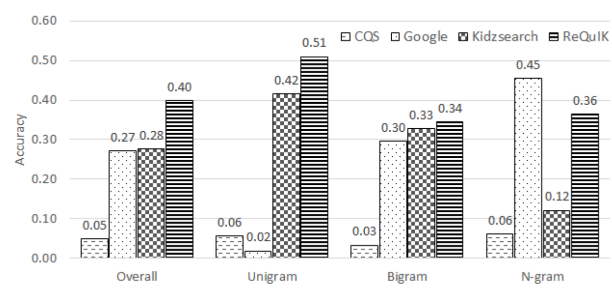

Figure 6: Comparison of query recommendation systems

6 CONCLUSIONS AND FUTURE WORK

In this paper, we presented *ReQuIK* a multi-perspective query recommendation system specifically tailored to facilitate information-seeking tasks for children. *ReQuIK* takes advantage of multiple strategies that inform the process of generating query suggestions and prioritize queries that are of interest to children. For assessing the performance of the system we conducted a study in collaboration with teachers of four different schools in the area of Idaho, where 97 students used our framework to complete factual and research/informational search tasks assigned to them by their teacher. We also conducted a user study to gather feedback from the teachers themselves. Using the set of queries written by children, we conducted a number of empirical analysis and demonstrated the validity of our proposed strategy and its value, in terms of offering up-to-date suggestions aimed at helping children with their information-seeking needs.

The technical contributions associated with our research work include (i) introducing a strategy which takes advantage of a search intent for capturing the purpose for which the query was written, (ii) creating a novel wide and deep neural network which considers both the raw text and traits frequently associated with child-related queries, (iii) employing a strategy to train a model based on children/non-children oriented resources that transfers well to the query suggestion context, (iv) explicitly considering the readability levels of both the query and results retrieved by the candidate suggestions, favoring the ones that leads to the retrieval of content that will be easier to read, (v) using a whole pipeline that considers multiple perspectives for candidate query generation, and (vi) creating a dataset comprised of children and non-children queries that can be used by the research community at large for children-related research.

In the future, we will extend the candidate query suggestion process by also considering phrases extracted from children corpora provided it evolves over time. In addition, we are aware that the range of 6 to 13 years can be too broad, as the interest of a 6 year old child are not the same as that of a 13 year old. Therefore, we plan to extend *ReQuIK* so that it is able to generate recommendation for more specific age ranges. Moreover, we plan to extend the manner in which resource readability is computed, to account for information beyond text, such as the different aesthetics of web pages, which can impact the level of complexity of resources. Finally, while in this paper we evaluated *ReQuIK* (and other query suggestions modules), in terms of offering queries that are suitable for children, in the future, we plan to conduct a user study to verify the degree to which *ReQuIK* leads to the retrieval of children-related resources and facilitates information discovery tasks for children.

ACKNOWLEDGMENTS

This work has been supported in part by the National Science Foundation Award Number 1565937.

REFERENCES

[1] Ubersuggest. http://ubersuggest.io, Accessed: 2017-05-18.
[2] J. Albright, C. de Guzman, P. Acebo, D. Paiva, M. Faulkner, and J. Swanson. Readability of patient education materials: implications for clinical practice. *Applied Nursing Research*, 9(3):139–143, 1996.
[3] A. Anagnostopoulos, L. Becchetti, C. Castillo, and A. Gionis. An optimization framework for query recommendation. In *ACM WSDM*, pages 161–170, 2010.
[4] Backlinco. How to find long tail keywords. http://backlinko.com/long-tail-keywords, Accessed: 2017-05-18.
[5] D. Bilal and R. Ellis. Evaluating leading web search engines on children's queries. In *Human-Computer Interaction. Users and Applications*, pages 549–558. Springer, 2011.
[6] J. S. Chall and E. Dale. *Readability revisited: The new Dale-Chall readability formula*. Brookline Books, 1995.
[7] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. Corrado, W. Chai, M. Ispir, et al. Wide & deep learning for recommender systems. In *Deep Learning for Recommender Systems workshop*, pages 7–10, 2016.
[8] Dogo. News, book, and movie reviews for kids by kids. https://www.dogonews.com, 2017.
[9] N. Dragovic, I. Madrazo Azpiazu, and M. S. Pera. Is sven seven?: A search intent module for children. In *ACM SIGIR*, pages 885–888, 2016.
[10] A. Druin, E. Foss, L. Hatley, E. Golub, M. L. Guha, J. Fails, and H. Hutchinson. How children search the internet with keyword interfaces. In *SIGCHI*, pages 89–96, 2009.
[11] S. Duarte Torres, D. Hiemstra, and P. Serdyukov. An analysis of queries intended to search information for children. In *IIiX*, pages 235–244, New York, NY, USA, 2010. ACM.
[12] S. Duarte Torres, D. Hiemstra, I. Weber, and P. Serdyukov. Query recommendation for children. In *ACM CIKM*, pages 2010–2014, 2012.
[13] C. Eickhoff, T. Polajnar, K. Gyllstrom, S. D. Torres, and R. Glassey. Web search query assistance functionality for young audiences. In *Advances in Information Retrieval*, pages 776–779. Springer, 2011.
[14] C. Eickhoff, P. Serdyukov, and A. P. de Vries. Web page classification on child suitability. In *ACM CIKM*, pages 1425–1428. ACM, 2010.
[15] R. Flesch. A new readability yardstick. *Journal of Applied Psychology*, 32(3):221, 1948.
[16] E. Foss, A. Druin, R. Brewer, P. Lo, L. Sanchez, E. Golub, and H. Hutchinson. Children's search roles at home: Implications for designers, researchers, educators, and parents. *JASIST*, 63(3):558–573, 2012.
[17] T. Gossen, J. Höbel, and A. Nürnberger. A comparative study about children's and adults' perception of targeted web search engines. In *SIGCHI*, pages 1821–1824, 2014.
[18] L. Graham and P. T. Metaxas. Of course it's true; i saw it on the internet!: critical thinking in the internet era. *Communications of the ACM*, 46(5):70–75, 2003.
[19] J. Gwizdka and D. Bilal. Analysis of children's queries and click behavior on ranked results and their thought processes in google search. In *ACM CHIIR*, pages 377–380, 2017.
[20] A. Halavais. *Search engine society*. John Wiley & Sons, 2013.
[21] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
[22] Ilp2. Search engine. http://www.Ilp2.com, 2017.
[23] K. in mind. Data source for kids. https://www.kids-in-mind.com, 2017.
[24] S. Karimi and M. S. Pera. Recommendations to enhance children web searches. *ACM RecSys*, 2015.
[25] L. Keating. Kiddle search engine is the google for kids. http://www.techtimes.com/articles/136639/20160225/kiddle-search-engine-kids.htm, Accessed: 2017-05-18.
[26] Kidrex. Safe search for kids. http://www.KidRex.org, 2017.
[27] A. kids. Search engine. http://www.askkids.com, 2017.
[28] E. Kids. Data source for kids. http://www.edutainingkids.com, 2017.
[29] KidzSearch. Web search engine for kids. http://www.kidzsearch.com, 2017.
[30] Kidzvuz. Data source for kids. https://www.kidzvuz.com, 2017.
[31] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.
[33] S. Lai, L. Xu, K. Liu, and J. Zhao. Recurrent convolutional neural networks for text classification. In *AAAI*, pages 2267–2273, 2015.
[34] Y. Li, D. McLean, Z. A. Bandar, J. D. O'shea, and K. Crockett. Sentence similarity based on semantic nets and corpus statistics. *IEEE TKDE*, 18(8):1138–1150, 2006.
[35] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky. The Stanford CoreNLP natural language processing toolkit. In *ACL System Demonstrations*, pages 55–60, 2014.
[36] J. Palotti, G. Zuccon, and A. Hanbury. The influence of pre-processing on the estimation of readability of web documents. In *ACM CIKM*, pages 1763–1766, 2015.
[37] RaisingChildren. Data source for kids. http://raisingchildren.net.au, 2017.
[38] M. Shaikh, M. S. Pera, and Y.-K. Ng. Suggesting simple and comprehensive queries to elementary-grade children. In *IEEE/WIC/ACM WI-IAT*, volume 1, pages 252–259, 2015.
[39] G. Spache. A new readability formula for primary-grade reading materials. *The Elementary School Journal*, 53(7):410–413, 1953.
[40] Spaghetti. Book reviews by kids for kids. http://www.spaghettibookclub.org, 2017.
[41] Sweetsearch. A search engine for students. http://www.sweetsearch.com, 2017.
[42] S. D. Torres, D. Hiemstra, I. Weber, and P. Serdyukov. Query recommendation in the information domain of children. *JASIST*, 65(7):1368–1384, 2014.
[43] Toyinsider. Data source for kids. http://www.thetoyinsider.com, 2017.
[44] A. Usta, I. S. Altingovde, I. B. Vidinli, R. Ozcan, and Ö. Ulusoy. How k-12 students search for learning?: analysis of an educational search engine log. In *ACM SIGIR*, pages 1151–1154, 2014.
[45] N. Vanderschantz, A. Hinze, and S. J. Cunningham. "sometimes the internet reads the question wrong": Children's search strategies & difficulties. *JASIST*, 51(1):1–10, 2014.
[46] I. B. Vidinli and R. Ozcan. New query suggestion framework and algorithms: A case study for an educational search engine. *Information Processing & Management*, 52(5):733–752, 2016.
[47] A. Wood and Y.-K. Ng. Orthogonal query recommendations for children. In *iiWAS*, pages 298–302, 2016.
[48] Yahoo! Academic Relations. L13 - yahoo! search query tiny sample. http://webscope.sandbox.yahoo.com/catalog.php.
[49] H. Zamani, M. Bendersky, X. Wang, and M. Zhang. Situational context for ranking in personal search. In *WWW*, pages 1531–1540, 2017.
[50] Z. Zhang and O. Nasraoui. Mining search engine query logs for query recommendation. In *WWW*, pages 1039–1040, 2006.