"Who are you?"
Identifying Young Users from a Single Search Query

Bettencourt, Benjamin; Manzi Muneza, Assoumer Redempta; Green, Michael; Anguiano, Samantha; Fails, Jerry Alan; Kennington, Casey; Wright, Katherine Landau; Pera, Maria Soledad

DOI
10.1145/3563359.3596992

Publication date
2023

Document Version
Final published version

Published in
UMAP 2023 - Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization

“Who are you?”: Identifying Young Users from a Single Search Query

Benjamin Bettencourt
Boise State University
Boise, Idaho, USA
benbettencourt@u.boisestate.edu

Assoumer Redempta Manzi Muneza
Boise State University
Boise, Idaho, USA
assoumerredempta@u.boisestate.edu

Michael Green
Boise State University
Boise, Idaho, USA
michaeljamesgreen86@gmail.com

Samantha Anguiano
Boise State University
Boise, Idaho, USA
samanthaanguiano@u.boisestate.edu

Jerry Alan Fails
Boise State University
Boise, Idaho, USA
jerryfails@boisestate.edu

Casey Kennington
Boise State University
Boise, Idaho, USA
caseykennington@boisestate.edu

Katherine Landau Wright
Boise State University
Boise, Idaho, USA
katherinewright@boisestate.edu

Maria Soledad Pera
Delft University of Technology
Delft, Netherlands
M.S.Pera@TUDelft.nl

ABSTRACT

As an initial step towards enabling the adaptation of (popular, and widely used) web search environments so that they can better serve children and ease their path towards information discovery, we introduce **R**ecognizing **Y**oung **Se**archers (**RYSe**). RYSe leverages lexical, syntactical, spelling/punctuation, and vocabulary features that align with the Concrete Operational stage of development (originally identified by Jean Piaget) in an attempt to identify users that are in this stage. The concrete operational stage is commonly associated with children ages 7-11. Findings emerging from our initial empirical exploration using single queries formulated by children and sample queries from adults showcase the feasibility of relying on different cognitive traits inferred from the short text of a single query to distinguish those that are formulated by younger searchers.

CCS CONCEPTS

• **Social and professional topics** → **Children**; • **Information systems** → *Web searching and information discovery*; Personalization.

KEYWORDS

user modeling, adaptation, web search, children

ACM Reference Format:
Benjamin Bettencourt, Assoumer Redempta Manzi Muneza, Michael Green, Samantha Anguiano, Jerry Alan Fails, Casey Kennington, Katherine Landau Wright, and Maria Soledad Pera. 2023. “Who are you?”: Identifying Young Users from a Single Search Query. In *UMAP ’23 Adjunct: Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization (UMAP ’23 Adjunct), June 26–29, 2023, Limassol, Cyprus.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3563359.3596992

ACM ISBN 978-1-4503-9891-6/23/06…$15.00
https://doi.org/10.1145/3563359.3596992

1 INTRODUCTION

Turning to search engines (**SEs**) for inquiry tasks – no matter the topic – is an integral part of the daily routines of individuals. Consider then a scenario where a searcher wants to gather some general facts about puppies. To initiate the information-seeking journey, the user crafts a simple keyword query to submit to any of the main, popular SEs. For example, the user submits the search query ‘*puppy*’ to Google. As illustrated in Figure 1, the resulting **S**earch **E**ngine **R**esult **P**age (**SERP**) prioritizes the *Top Stories* of the day with four articles containing the word *puppy* including one from CBS Sports titled “Valentina Shevchenko compares Erin Blanchfield to a barking puppy...” which is not relevant to the user’s information need. Now imagine that this user is an eight-year-old child. In this case, an article describing an upcoming Ultimate Fighting Championship match is not only irrelevant but may also be inappropriate for this young searcher.

Popular SEs are designed to offer access to a broad, up-to-date, set of resources to appeal to as many users as possible. Children having access to such a broad spectrum of resources is not necessarily bad, provided these young users get the relevant and appropriate information that they need. At the same time, existing SEs are not designed to support young searchers [9, 27], even though supporting their information discovery needs is critical. Adapting existing SEs to meet children’s specific needs requires first identifying useful properties of the user. Otherwise, SERPs will include resources that target broader user groups, as we showcased in our aforementioned example. Automatic detection of searchers belonging to a specific user group – specifically children, but also other underserved populations – in the context of web search, would enable adaptation of

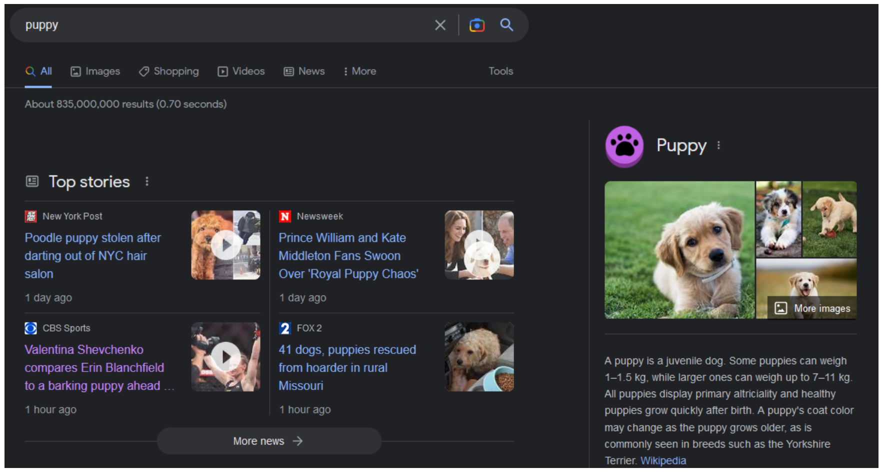

Figure 1: Google SERP retrieved using the search term '*puppy*' (Retrieved February 28th, 2023)

not only the content SEs present but also the extended functionality they can offer to better serve that particular user group. For instance, Landoni et al. [17] found that enhancing a SERP with visual cues, like emojis, led child searchers to click on relevant results more often. Still, this type of adaptation requires recognizing on-the-fly whether a young searcher is behind the keywords prompting a search.

In this paper, we introduce **R**ecognizing **Y**oung **Se**archers (**RYSe**), a strategy for identifying child searchers based on a given search query. To inform the design of this strategy, we turned to Jean Piaget's "Stages of Cognitive Development" which describe the qualitative differences in cognitive development among children at different ages [30]. Piaget's work demonstrated that children are not born with the same cognitive processes as adults and that there are sequential cognitive stages that children go through as they grow. Although a child's thought processes cannot be directly measured, Piaget posited that their current behaviors and language usage reflect their cognitive stage. To control scope, in this initial iteration of our work, we focused on **P**iaget's **C**oncrete **O**perational stage (*PCO*) in which Piaget defines characteristics that are common in children ages 7- to 11-years old.

Children in the *PCO* stage are still developing their ability to engage in abstract thought and understand unobserved processes [30]. As a child progresses through the different stages of cognition, their vocabulary grows and so does their ability to build different types of sentences. Children in *PCO* tend to struggle with abstract ideas, which is reflected in their language use. We hypothesize that this could impact both the content and presentation of their queries.

With this work, we aim to answer the following **RQ**: *Can a typical SE user in the PCO stage of development be identified utilizing their search query?* To do so, we examine child search queries using lexical, syntactical, spelling/punctuation, and vocabulary features that may reveal that a user is in the *PCO* stage of cognitive development. Outcomes from this work offer insights into the feasibility of further exploring user skills and an understanding of their cognitive abilities as a way to enable adaptation – and ultimately personalization – of web search environments and other technologies.

2 RELATED WORK

Varying strategies for automating the detection of users with different characteristics have been proposed and evaluated. Lin [20] utilized sentence length, vocabulary, and emoticons from chat logs created by users (ages 13-59) to help law enforcement identify users' gender and age with a naive Bayes classifier to flag potentially harmful chat exchanges. Weren et al. [31] trained a classifier with three sets of features (Cosine Similarity of conversations, Okapi BM25 Similarity of conversations, and Flesch-Kincaid readability scores) in order to identify users' age and gender by the similarities of both English and Spanish chat log conversations. One limitation of this strategy is that readability scores, like that of Flesch-Kincaid, only utilize surface-level text features (e.g. word count, sentence count, etc.). Tam and Martell [29] focused on users over age 13 and explored word similarity using character n-grams and document metadata (capital letters, tokens, emoticons, length, part-of-speech) among chat logs to identify users' age (13-19, 20s, 30s, 40s, and 50s) using a naive Bayes and support vector machine model. Perhaps the most closely related prior research is that by Duarte et al. [11] who introduced a strategy to identify users by age group, Child (5-12), Teen (13-17), and Mature Teen (18) as well as gender. The proposed strategy relied on supplementing textual features with information unique to SEs, such as sites visited and/or time spent searching. However, this strategy might not be suited for on-the-fly identification as it requires a log of the user's actions. For example, a user would only be classified as a Child if they took a longer time searching and navigated to sites that pertain more to younger users. Overall, the literature discussed thus far supports the importance of examining text samples to detect specific user groups, but none explicitly focuses on young searchers, nor limits their samples to a single, and often very short, query.

3 RYSE: A STRATEGY FOR IDENTIFYING *PCO*

Given a query, RYSe detects *PCO* child searchers (ages 7-11) by considering a wide range of features inferred from the text that map to the various skills and behaviors attributed to this group. Below, we discuss these features along with their alignment to the

PCO stage; we then describe how they are collectively used for *PCO* identification. The full list of features and their implementation are available at: https://bitbucket.org/bsu-cast/ryse/.

3.1 Query Features Used to Identify *PCO*

Lexical Features: The words in a query are analyzed based on their length and type of words (e.g. noun, verb). Based on Piaget's work, we hypothesized that children in this age group would use less sophisticated lexical features in their queries than *non − PCO* users. We examined three types of lexical features: lexical density, sophistication, and variation. Lexical density is the ratio of lexical words such as nouns, adjectives, verbs (except modal verbs, auxiliary verbs, 'be' and 'have'), and adverbs with an adjectival base, including those that can function as both an adjective and adverb (e.g., 'fast') and those formed by attaching the '-ly' suffix to an adjectival root (e.g., 'particularly') [21] to words in a text. Lexical sophistication is "the proportion of relatively unusual or advanced words in the learner's text" [28]. Lexical variation "refers to the range of a learner's vocabulary as displayed in his or her language use" [21]. Additionally, we studied the lexical characteristics to further examine how keywords differ from natural language queries as types of queries. E.g. the total number of words per query since keywords are often standalone terms or short phrases

Vocabulary Features: We hypothesized that children's vocabulary use would differ from that of adults. Specifically, we hypothesized that *PCO* children would tend to use fewer abstract words. We identified words that children may commonly use in a few ways. First, we used Edu2Vec [3], a mixed strategy that combines (i) the Mikolov et al. [25] skip-gram model, (ii) pre-trained embeddings from Wikipedia and Wikibooks, and (iii) structured knowledge from educational standards to identify relevant terms, topics, and subjects for each grade. As Edu2Vec terms are commonly found in texts from Kindergarten to 6^{th} grade, we expect these terms to be familiar to, and thus used by, children. Consequently, we examine how frequently these terms appear in a query. Additionally, the Age of Acquisition (**AoA**) identifies the average age at which a person generally learns certain words. Therefore, we consider the 50K words in the AoA dataset [15]. Stopwords, domain-specific words, and search operators were also included in this feature set.

Spelling and Punctuation Features: In line with *PCO*, children often make spelling mistakes [13] while writing or typing. Also, children struggle to use some existing search engine query formulation support mechanisms such as query suggestions or auto-correction [1]. We, therefore, investigate the total number of misspelled words in each query and compare those words with a list of typos made by children from the KidSpell dataset [8] to identify which queries contain typos similar to 7-11 year old searchers. In order to capture a variety of child misspellings, we compare the misspelled words with the suggested correction checking for Levenshtein distance.[1] In addition, since young searchers tend to treat search queries as sentences [14], we hypothesize that they likely use capital letters and punctuation. However, capitalization does not matter in web search [7] as well as punctuation. We, therefore, inspect the queries for punctuation that could be seen in regular written sentences.

Syntactical Features: Natural language and keyword queries can differ in syntax. Natural language queries tend to resemble sentences with articles, preposition phrases, etc, while keyword queries are often strings of nouns. The ability to identify the key information in a sentence and select appropriate keywords may be challenging for children. Consequently, we study this difference in the queries, where we compute the ratio of the number of that part of speech to the total number of words (e.g. number of verbs divided by total words).

3.2 Detecting *PCO*

From a user's search query we derive the features discussed in Section 3.1, and consider them as evidence to determine whether or not the user displays behaviors expected for those in the *PCO* stage of cognitive development. We treat this task as a classification problem and thus use these features as input to a Random Forest Classifier [19]. We selected this classifier given the manner in which it performs feature selection; as discussed in [22] the random forest classifier is noted for its ability to build trees that correlate to specific subsets of features. This feature sub-sampling leads to specialized trees that recognize users based on these subsets of features. Furthermore, the space represented by our feature set is non-linear. Drawing a clear line down the middle of these numbers is a poor approach considering the potential variance of user skill within *PCO* searchers. As such, we have chosen a non-linear classifier. Detailed information regarding training and testing of the chosen classifier is discussed in subsection 4.2.

4 EXPERIMENTAL SET UP

In this section, we describe the data, metrics, and evaluation process undertaken to answer the RQ.

4.1 Search Query Data

In light of concerns regarding children's privacy [12], there are no large, widely available search session records generated by a stereotypical 7-11 year-old searcher. As such, we constructed a dataset[2] using data from various sources, which we summarize in Table 1. This dataset includes: (i) adult queries from the **TREC** data [26], (ii) queries first introduced in the work conducted by Madrazo Azpiazu et al. [23] (henceforth referred to as **Sven**), each indicating whether it was formulated by a child (ages 8 to 11) or adult, and (iii) queries formulated by children ages 7 to 11 years old, which we refer to as **CAST** data[3], that resulted from several user studies conducted to examine children's search behavior [1, 2, 4, 8, 10].

According to data from the US Census Bureau [5], children (ages 3–17) and adults (ages 18+) make up ∼20% and ∼80%, respectively, of the country's internet users. Based on this information, we maintain a 20/80 distribution for children and adults, and therefore randomly sampled 949 *PCO* (child) queries and 3796 non-*PCO* (adult) queries from the dataset for the evaluations that follow.

[1] Levenshtein distance demonstrates the off by one typo meaning insertions, deletions, or substitutions of one character.

[2] We did not directly involve children for data collection for this work. All data was acquired from previous publications.

[3] Queries in the CAST data were collected using the Child Adaptive Search Tool: https://cast.boisestate.edu/.

Table 1: Overview of the data used to develop and analyze RYSe. Final sample is a 1:4 ratio, child:adult.

Type	Data Source	User	Age	Total Queries	Unique Queries	Total Unique	Final Sampled Unique Queries
PCO	CAST [1, 2, 4, 8, 10]	Child	7-11	2571	1729	2023	949
	Sven [23]	Child	8-11	301	294		
non − PCO	Sven [23]	Adult	18+	301	301	3797	3796
	TREC [26]	Adult	18+	5371	3495		

4.2 Evaluation Strategy

To analyze performance, we split the dataset into three sets: 65% training set, 15% development set, and 20% testing set. To evaluate RYSe effectiveness in identifying searchers in the *PCO* stage of cognitive development, we turn to well-known metrics: **Accuracy**, **T**rue **P**ositive **R**ate (**TPR**), and **T**rue **N**egative **R**ate (**TNR**). Accuracy refers to the fraction of the correct predictions. This metric is common for assessing a binary classifier's efficiency [6]. TPR estimates the proportion of which the actual positive observations are correctly predicted, while TNR allows us to study how well the model recognizes users who do not belong to our target group (*non − PCO*).

In order to determine the relative value of the feature sets RYSe utilizes, we conducted an *ablation study* comparing the effectiveness of using each individual feature set as compared to the full feature set utilized in RYSe. As the vocabulary that children use is influenced by their level of language development [30], we grouped language development features (primarily lexical and vocabulary features) as well as features associated with abstract thinking and processes (primarily spelling and syntactical features).

To offer context to the applicability of RYSe, we consider two simple baseline models; a *Majority classifier* which assumes that any new sample belongs to the majority class (in our case is *non − PCO*) which imitates the general SE narrative that assumes all individuals without identification are adults, and a *Text-based classifier* (introduced in [29]) which is based on a Support Vector Machine (with linear kernel, auto kernel coefficient, and the default regularization parameter) that examines tri-grams of texts from chat logs. Given the data available for analysis in the case of RYSe, i.e., search queries, we adapt the original Text-based model and re-trained it using search queries as input text samples instead of chat logs.

We use the development set (15% of the dataset) for hyperparameter tuning of RYSe's Random Forest Classifier. We do so via a grid search, optimizing for TPR. The resulting parameters are: 400 estimators, unbalanced class weight, Gini impurity as the criterion, and no bootstrapping. Both RYSe's and baselines' results are based on 5-fold cross-validation over the training (65%) and testing (20%) sets. When juxtaposing strategies considered in the ablation study as well as comparing baseline models to RYSe, we employ the McNemar test [16] with $p \leq 0.05$ to determine statistical significance; this test was considered because the resulting values reported are dichotomous rather than continuous.

5 RESULTS AND DISCUSSION

The main goal of this work was to identify a user in the *PCO* stage of cognitive development, based on features extracted from a single search query. In this section, we present the findings from the evaluation strategy outlined above.

Table 2 shows that the RYSe strategy results in a fairly high accuracy of 90.9%, a moderate TPR of 64.4%, and a high TNR of 97.6%. Comparing RYSe to its individual feature sets reveals that the *Language Development Related Features*[4] exhibit similar (but slightly lower) measures as compared to RYSe which may indicate a high impact of those particular features. This lends validity to using Jean Piaget's stages of cognitive development as a means to identify users as this feature set seems to most closely model the users' developmental stage. The *Language Development Related Features*, *Lexical Features*, and *Vocabulary Features* did not have a significant difference compared to RYSe in terms of TPR. The language development features, again, seem to be the most significant contributor to the RYSe strategy. The relatively high accuracy for the *Lexical Features* and *Vocabulary Features* seems to indicate that there is, indeed, a notable words difference between *PCO* and *non − PCO* users' queries. The *Spelling and Punctuation Features* had relatively low accuracy (79.5%), compared to the accuracy yielded when considering the remaining feature sets. Although it had the highest TNR (99.0%) among those reported in Table 2, it also had the lowest TPR with 2.4%, where essentially it always classifies as *non − PCO*. These results reveal this feature set is perhaps the least meaningful set when trying to identify users in the *PCO* stage. We attribute this to the fact that typos and misspellings are common among all user groups [18].

As shown in Table 3, RYSe significantly outperformed ($p \leq 0.05$) both the Text-Based and Majority classifiers in terms of Accuracy and TPR. Considering that RYSe also yielded significantly higher TPR, compared to the majority and text-based counterparts, we treat this as an indication of RYSe's ability to identify *PCO* users.

As discussed in the previous sections, with the majority classifier, the new data points are classified based on the majority class, i.e., *non − PCO* in our case. Since our dataset contains 80% and 20% distribution of *PCO* and *non − PCO* observations respectively, this justifies the 79.8% accuracy with 100% TNR for this classifier. The text-based classifier takes into consideration word trigrams and as anticipated, outperforms the majority classifier on all metrics considered. This indicates that the three words sequence in *PCO* and *non − PCO* search queries can be differentiated them rather than assuming that each query belongs to non-PCO without any further analysis as the Majority classifier does. Still, with its low TPR

[4] Language Development Related Features refer to those that are directly related to a child's language development, such as lexical density or the number of words in a query.

Table 2: Feature ablation study results. *Indicates statistical significance of a particular feature set compared to RYSe ($p \leq 0.05$).

Feature Set	Accuracy	TPR	TNR
RYSe	0.909	0.644	0.976
Language Development Related Features	0.904	0.629	0.974
Abstract Process Related Features	0.851	0.531*	0.931*
Lexical Features	0.892	0.613	0.963
Vocabulary Features	0.896	0.617	0.966
Spelling and Punctuation Features	0.795*	0.024*	0.990
Syntactical Features	0.851	0.537*	0.930*

Table 3: RYSe vs. baseline classification models. *Indicates statistical significance of model compared to RYSe ($p \leq 0.05$).

Classifier	Accuracy	TPR	TNR
RYSe	0.909	0.644	0.976
Majority	0.798*	0.000*	1.000
Text-Based	0.838*	0.255*	0.996

(25.5%), we argue that it is not an effective solution for identifying *PCO*, and is not as effective as RYSe.

The task at hand is a challenging one, as the mainstream user group (i.e., *non* − *PCO*) is more prominent in real life and therefore in our dataset. This is evident by the relatively high accuracy of the majority classifier. We posit that the improvement over the majority baseline achieved by the Text-based classifier is a positive signal of the richness of information that can be inferred even from a short query. This is further emphasized by RYSe's overall performance improvement by simultaneously accounting for different perspectives (i.e., feature sets) inferred from short text samples – even very short keyword queries that are not always semantically meaningful. Although the accuracy of RYSe is high, the main purpose of RYSe is to correctly identify users in the *PCO* stage. With a 64.4% TPR, RYSe showcases promise in achieving its goal but also indicates that there is room for improvement.

6 CONCLUSION, LIMITATIONS, AND FUTURE WORK

In this paper, examined whether *a typical SE user in the PCO stage of development could be identified by utilizing their search query*. For this we introduced RYSe, a classification strategy that leverages multiple features aligning with the *PCO* stage of development to detect *PCO* searchers from their search queries. With RYSe being able to identify *PCO* users consistently in a naturalistic unbalanced query sample this indicates a promising route in terms of identifying users based on their perceived stage of development. This performance, showcased by the RYSe strategy, is a crucial step towards enabling on-the-fly adaptability of SE to better support a wide range of user groups.

With regards to limitations, RYSe currently considers a single query from a user, performance could potentially be improved by including multiple queries and/or full search sessions. RYSe also only takes behaviors derived from queries into account, however, there are more search behaviors that users exhibit when searching online which could prove useful in the identification of users when searching online (e.g. results clicked, time taken to formulate a query, multiple queries from the same user). Moreover, extending RYSe to account for non-American English Cultures is likely nontrivial; although Jean Piaget's Stages of Cognitive Development may apply just as well to children of other cultures [24], the language features that map to *PCO* might be different. We foresee applications of our proposed strategy being useful not only for adapting SERPs to better serve child searchers but also for the filtering of potentially inappropriate and/or irrelevant content. In the future, we plan to conduct user studies to evaluate the applicability of embedding RYSe in SE to facilitate those applications. Moreover, we plan to extend the design of RYSe to also encapsulate different developmental stages, not just the *PCO* stage, to cater to the needs of users at the various stages.

ACKNOWLEDGMENTS

This work was partially supported by NSF awards 1763649 and 1565937.

REFERENCES

[1] 2019. Query Formulation Assistance for Kids: What is Available, When to Help & What Kids Want. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children (IDC '19)*. Association for Computing Katherine Landau and Bigirimana, William, Boise, ID, USA, 109–120. https://doi.org/10.1145/3311927.3323131

[2] Garrett Allen, Benjamin L Peterson, Dhanush kumar Ratakonda, Mostofa Najmus Sakib, Jerry Alan Fails, Casey Kennington, Katherine Landau Wright, and Maria Soledad Pera. 2021. Engage!: Co-designing Search Engine Result Pages to Foster Interactions. In *Interaction Design and Children (IDC '21)*. Association for Computing Machinery, New York, NY, USA, 583–587. https://doi.org/10.1145/3459990.3465183

[3] Oghenemaro Anuyah, Ion Madrazo Azpiazu, and Maria Soledad Pera. 2019. Using structured knowledge and traditional word embeddings to generate concept representations in the educational domain. In *Companion Proceedings of The 2019 World Wide Web Conference*. 274–282.

[4] Oghenemaro Anuyah, Jerry Alan Fails, and Maria Soledad Pera. 2018. Investigating query formulation assistance for children. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. ACM, 581–586. https://doi.org/10.1145/3202185.3210779

[5] US Census Bureau. 2012. Computer and Internet Access in the United States: 2012. https://www.census.gov/data/tables/2012/demo/computer-internet/

computer-use-2012.html.
[6] Davide Chicco and Giuseppe Jurman. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics* 21 (2020), 1–13.
[7] Silviu Cucerzan. 2010. Does Capitalization Matter in Web Search?. In *KDIR*. 302–306.
[8] Brody Downs, Oghenemaro Anuyah, Aprajita Shukla, Jerry Alan Fails, Sole Pera, Katherine Wright, and Casey Kennington. 2020. Kidspell: A child-oriented, rule-based, phonetic spellchecker. In *Proceedings of The 12th Language Resources and Evaluation Conference*. 6937–6946.
[9] Brody Downs, Tyler French, Katherine Landau Wright, Maria Soledad Pera, Casey Kennington, and Jerry Alan Fails. 2019. Children and search tools: Evaluation remains unclear. *KidRec Workshop co-located with ACM IDC 2019* (Jun 2019). https://par.nsf.gov/biblio/10099389-children-search-tools-evaluation-remains-unclear
[10] Brody Downs, Maria Soledad Pera, Katherine Landau Wright, Casey Kennington, and Jerry Alan Fails. 2022. KidSpell: Making a difference in spellchecking for children. *International Journal of Child-Computer Interaction* 32 (Jun 2022), 100373. https://doi.org/10.1016/j.ijcci.2021.100373
[11] Sergio Duarte Torres, Djoerd Hiemstra, and Pavel Serdyukov. 2010. An analysis of queries intended to search information for children. In *Proceedings of the third symposium on Information interaction in context (IIiX '10)*. Association for Computing Machinery, New York, NY, USA, 235–244. https://doi.org/10.1145/1840784.1840819
[12] Michael D. Ekstrand. 2017. *Challenges in Evaluating Recommendations for Children.* https://md.ekstrandom.net/pubs/kidrec-eval
[13] Tatiana Gossen, Julia Hempel, and Andreas Nürnberger. 2013. Find it if you can: usability case study of search engines for young users. *Personal and Ubiquitous Computing* 17 (2013), 1593–1603.
[14] Hyejung Han. 2018. Children's help-seeking behaviors and effects of domain knowledge in using Google and Kids. gov: Query formulation and results evaluation stages. *Library & Information Science Research* 40, 3-4 (2018), 208–218.
[15] Victor Kuperman, Hans Stadthagen-Gonzalez, and Marc Brysbaert. 2012. Age-of-acquisition ratings for 30,000 English words. *Behav. Res. Methods* 44, 4 (2012), 978–990.
[16] Peter A Lachenbruch. 2014. McNemar test. *Wiley StatsRef: Statistics Reference Online* (2014).
[17] Monica Landoni, Mohammad Aliannejadi, Theo Huibers, Emiliana Murgia, and Maria Soledad Pera. 2022. Have a Clue! The Effect of Visual Cues on Children's Search Behavior in the Classroom. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval (CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 310–314. https://doi.org/10.1145/3498366.3505845
[18] Yanen Li. 2020. Query spelling correction. *Query Understanding for Search Engines* (2020), 103–127.
[19] Andy Liaw, Matthew Wiener, et al. 2002. Classification and regression by randomForest. *R news* 2, 3 (2002), 18–22.
[20] Jane Lin. 2007. *Automatic author profiling of online chat logs.* Technical Report. Naval Postgraduate School Monterey CA.
[21] Xiaofei Lu. 2012. The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal* 96, 2 (2012), 190–208.
[22] Ion Madrazo. 2016. *Towards multipurpose readability assessment.* Master's thesis. BOISE STATE UNIVERSITY.
[23] Ion Madrazo Azpiazu, Nevena Dragovic, Oghenemaro Anuyah, and 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 the 2018 Conference on Human Information Interaction & Retrieval (CHIIR '18)*. Association for Computing Machinery, New York, NY, USA, 92–101. https://doi.org/10.1145/3176349.3176379 event-place: New Brunswick, NJ, USA.
[24] James Mangan. 1978. Piaget's Theory and Cultural Differences. The Case for Value-Based Modes of Cognition. *Human Development* 21, 3 (1978), 170–189. https://doi.org/10.1159/000271581 Publisher: Karger Publishers.
[25] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems* 26 (2013).
[26] NIST. 2014. Trec session track 2010-2014. https://trec.nist.gov/data/session.html.
[27] Maria Soledad Pera, Katherin Landau Wright, Casey Kennington, and Jerry Alan Fails. 2023. Children and Information Access: Fostering a Sense of Belonging. In *Joint Proceedings of the IUI 2023 Workshops: HAI-GEN, ITAH, MILC, SHAI, SketchRec, SOCIALIZE (CEUR Workshop Proceedings, Vol. 3359)*, Alison Smith-Renner and Paul Taele (Eds.). CEUR, Sydney, Australia, 254–257. https://ceur-ws.org/Vol-3359/#paper34
[28] John Read. 2000. *Assessing vocabulary.* Cambridge university press.
[29] Jenny Tam and Craig H. Martell. 2009. Age Detection in Chat. In *2009 IEEE International Conference on Semantic Computing*. 33–39. https://doi.org/10.1109/ICSC.2009.37
[30] Diane H. Tracey and Lesley Mandel Morrow. 2017. *Lenses on reading: an introduction to theories and models* (third edition ed.). the Guilford Press, New York, NY.
[31] ER Weren, Viviane P Moreira, and J Oliveira. 2013. Using simple content features for the author profiling task. In *Notebook for PAN at cross-language evaluation forum. Valencia, Spain.*

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009