An empirical analysis of search engines' response to web search queries associated with the classroom setting

Received 7 June 2019
Revised 23 August 2019
1 November 2019
Accepted 8 November 2019

Oghenemaro Anuyah, Ashlee Milton, Michael Green and Maria Soledad Pera
People and Information Research Team,
Department of Computer Science, Boise State University, Boise, Idaho, USA

Abstract

Purpose – The purpose of this paper is to examine strengths and limitations that search engines (SEs) exhibit when responding to web search queries associated with the grade school curriculum

Design/methodology/approach – The authors employed a simulation-based experimental approach to conduct an in-depth empirical examination of SEs and used web search queries that capture information needs in different search scenarios.

Findings – Outcomes from this study highlight that child-oriented SEs are more effective than traditional ones when filtering inappropriate resources, but often fail to retrieve educational materials. All SEs examined offered resources at reading levels higher than that of the target audience and often prioritized resources with popular top-level domain (e.g. ".com").

Practical implications – Findings have implications for human intervention, search literacy in schools, and the enhancement of existing SEs. Results shed light on the impact on children's education that result from introducing misconception about SEs when these tools either retrieve no results or offer irrelevant resources, in response to web search queries pertinent to the grade school curriculum.

Originality/value – The authors examined child-oriented and popular SEs retrieval of resources aligning with task objectives and user capabilities–resources that match user reading skills, do not contain hate-speech and sexually-explicit content, are non-opinionated, and are curriculum-relevant. Findings identified limitations of existing SEs (both directly or indirectly supporting young users) and demonstrate the need to improve SE filtering and ranking algorithms.

Keywords Education, Children, Readability, Search engines, Misinformation, K-12, Search engine results pages, Web search queries, Child-appropriateness

Paper type Research paper

Introduction

Search Engines (SEs) are the "go-to" tools for children's online information discovery (Rowlands *et al.*, 2008). Children's use of SEs goes beyond accessing sites for leisure purposes, as they also utilize these tools for school-related activities. In fact, teachers in the USA regularly assign search tasks that require the use of SEs to their students (Hussain *et al.*, 2011; Scholastic, 2018). Children, however, are still known to experience difficulty completing successful search sessions (Gossen, 2016). A contributing factor toward improving children's search experience could be their understanding of how SEs work. Unfortunately, search literacy is rarely part of the Kindergarten to twelveth grade (K-12) curriculum (Campbell *et al.*, 2018; Notess, 2006; Laxman, 2010). Even when search literacy is included, it can overlook information on the functionality of SEs (Buckingham, 2015). Further, it more commonly targets educators (Campbell *et al.*, 2018; Notess, 2006; Laxman, 2010). Educational and child-oriented SEs, such as Kidrex[1] and Kidzsearch[2],

This research work has been supported by the National Science Foundation with Award No. 1565937. The authors also want to thank the students who took part in this study.

Aslib Journal of Information Management
Vol. 72 No. 1, 2020
pp. 88-111

2050-3806
DOI 10.1108/AJIM-06-2019-0143

could also alleviate challenges regarding the completion of search exercises. Yet, their use is not a requirement in the classroom. Instead, children tend to favor more mainstream, popular SEs, like Google and Bing (Foss *et al.*, 2012). These SEs are generally tailored for adults and as such, are not necessarily equipped to support children's educational searches–those posted within the classroom environment or formulated with the intent of locating curriculum-related materials. Thus, children could access resources that neither align with requirements inherent to the classroom environment nor their capabilities. Technology enhanced learning platforms, such as online tutorials and personalized electronic learning, will continue to gain prominence in response to the iGeneration[3], the most tech savvy yet (Minal Anand, 2019). With training now being made available to school teachers (Google, 2019), SEs become natural partners to ease online information discovery and facilitate, whenever possible, searching while learning (Gwizdka *et al.*, 2016). Therefore, it is imperative to understand the effectiveness of existing SEs in responding to search queries pertinent to the K-12 curriculum.

When performing curriculum-related inquiries, children may inadvertently be exposed to irrelevant materials that are inappropriate, such as pornographic or hate-based sites (Patel and Singh, 2016; Madigan *et al.*, 2018; Tori DeAngelis, 2007). To avoid the potential retrieval of inappropriate resources, popular SEs and child-oriented ones make available a safe-search filter. Safe-search, however, has some shortcomings. On the one hand, this filter could be too restrictive: it might disregard resources that are relevant to the curriculum but happen to include terms that can be misconstrued as inappropriate. For example, Kidzsearch's safe-search interprets the intent behind the search query breast tissue as being inappropriate, and therefore does not retrieve any results (see Figure 1(a)). This can be problematic when children are given a school research assignment on Human Anatomy, as this would prevent them from locating the right resources. On the other hand, safe-search has also been known to let unbefitting resources pass through the filter, as it often employs the use of thresholds set by administrators (Edelman, 2003; Heiler *et al.*, 2017). Consider the search query urban Google's safe-search includes among the top 10 retrieved results a site known to contain profane language (see Figure 1(b)). Being that this inquiry could be related to the History subject, it is detrimental for children to encounter inappropriate content on the first Search Engine Results Page (SERP).

Children's experiences with SEs can affect their motivation to use the Web, their skill to adequately use resources for their personal and educational interests, as well as their exposure to information beneficial for enhancing their mental capabilities (Foss *et al.*, 2012). As such, it is problematic if SEs offer resources that are irrelevant, based on their alignment with the target user capabilities. A number of research works suggest that the readability of web text may be beyond the reading ability and comprehension skill of young users (Bilal, 2013; Bilal and Huang, 2019). From the perspective of children, a resource that is relevant with respect to information needs expressed in a search query becomes irrelevant if its content does not align with their reading skills. Hence, offering resources children can comprehend is essential.

The use of SEs to conduct curriculum-related inquiries may also lead children to resources that are irrelevant to task objectives–the purpose of the search task assigned. These resources are often the result of SE domain bias (Ieong *et al.*, 2012; Introna and Nissenbaum, 2000), as well as the influence of Search Engine Optimization (SEO) techniques (Lewandowski, 2011; Thurow, 2015), causing the prioritization of resources from popular domains or those that are treated as relevant to the average population over those that target the classroom. Unfortunately, these resources can contain opinionated and non-educative content, thereby making them inapplicable to curriculum-related inquiries. For example, consider a child that initiates the search using the search query *albert einstein last invention*. In response to this inquiry, Bing (with the safe-search filter enabled) ranks a resource from answers.com at the top of the SERP (see Figure 1(c)). As this site is mostly

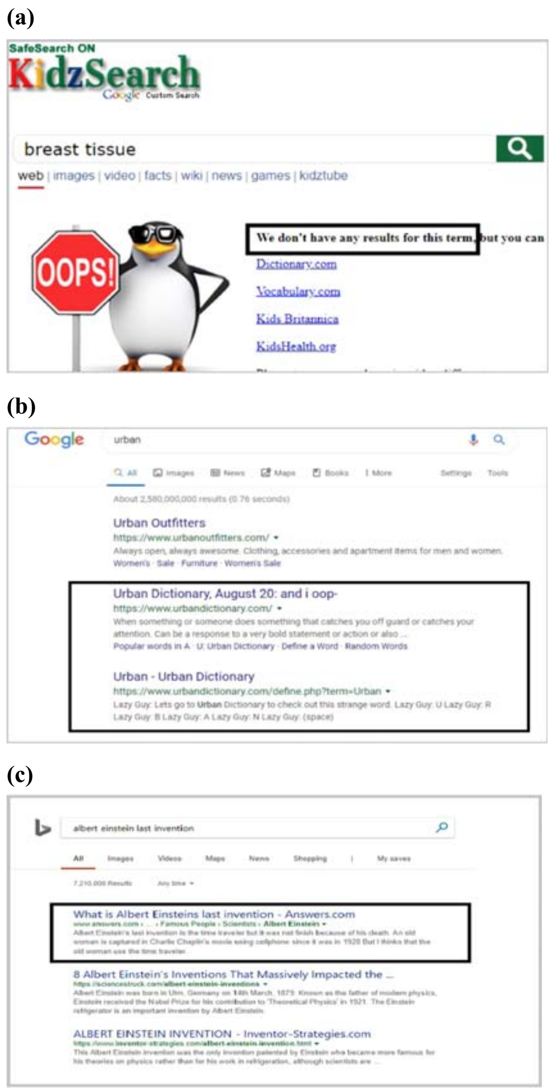


Figure 1.
Resources retrieved by SEs in response to search queries associated with the classroom setting

Notes: (a) Kidzsearch fails to retrieve resources for the search query *breast tissue* (May 2019); (b) Google's safe-search includes a site that potentially exposes children to profane language among top–10 resources retrieved for the search query *urban* (August 2019); (c) Bing's safe-search ranks a resource from an opinion-based and non-educational site (answers.com) at the top of the SERP for the search query *albert einstein last invention* (March 2018)

comprised of people's opinions and it likely does not include verifiable and educational information, it could provide the child with wrong answers to this inquiry.

A few studies demonstrate that children's search behaviors are different from those of adults (Bilal and Kirby, 2002; Gossen *et al.*, 2011; Graham and Metaxas, 2003). Children are known to navigate through resources from top to bottom of the SERP in a sequential manner (Bilal and Ellis, 2011; Gwizdka and Bilal, 2017; Graham and Metaxas, 2003; Walker, 2013). However, it is still unclear whether SEs (not just those designed for children) accommodate this top-down search style. Moreover, several studies have examined SE's response to search tasks initiated by children both from system and user perspectives (Lovato *et al.*, 2019; Fails *et al.*, 2019; Duarte Torres *et al.*, 2010b; Druin *et al.*, 2009). Yet, the exploration focused on the classroom setting, for example, examining SE's response to search queries initiated in the classroom setting in terms of positioning retrieved resources in a way that can ease the identification of relevant material, has not been fully discussed in the literature.

To better understand how existing SEs fare when responding to searches pertinent to the classroom setting, we perform an empirical analysis on well-known SEs. Among these SEs, we examine Google and Bing as examples of popular SEs (GlobalStats, 2019); Kidrex and Kidzsearch which are child-oriented ones; and safe-search filters available on popular SEs under study[4]. We limit the scope of our analysis to children in the third to fifth grade levels[5], as young users within this range are in the concrete operational stage of development according to Piaget's theory (Piaget, 1976), and are known to exhibit similar search traits (Foss *et al.*, 2012). As opposed to leisure-related searches conducted by children which cover a broader scope of topics, we focus on curriculum-related inquiry tasks. For our analysis, we employ a simulation-based experimental approach to mimic a set of diverse inquiry tasks. Along the way, we discuss observations that result from our experiments based on characteristics of resources that make them applicable to task objectives and user capabilities in the classroom setting: resources that align with the reading skills of the target audience, are appropriate, are non-opinionated and are education-relevant.

We guide our empirical analysis by three research questions:

RQ1. Do SEs effectively respond to search queries that include terminology pertinent to the third–fifth grade curriculum?

We simulate a context where children in the third–fifth grades seek information relating to their Health and Science subjects. To this effect, we investigate if resources are always being retrieved for these searches, even if search queries used to trigger the search include keywords that in isolation might be misinterpreted as inappropriate (e.g. breast tissue):

RQ2. Do SEs deter access to inappropriate content?

We investigate the performance of SEs in filtering sexually-explicit and hate-based materials. In doing so, we simulate the search using search queries that contain sexually-explicit or hate-based terms, on SEs that have the safesearch functionality enabled or disabled. For this analysis, we examine: if SEs retrieve no results for the search queries and if the SEs happen to include inappropriate content in resources that pass through the filter:

RQ3. Do SEs provide resources that align with task objectives and user capabilities?

We examine the ranking performance of SEs in offering resources relevant to inquiries that are associated with the third–fifth grade curriculum and that are within the target user's ability to understand, i.e., readability. In conducting our analysis, we simulate the search process using children's search queries. We especially analyze: the degree to which top ranked resources match the reading level of the target audience, the rate at which resources from different top-level domains (TLDs) are prioritized, and the extent to which resources that are not written by experts are potentially favored by SEs.

In addition to identifying the strengths and limitations of SEs in responding to inquiries associated with the classroom setting, our findings shed light on the impact SEs can have on school-aged children. This is in terms of young users having misconception about the SEs themselves, which may be attributed to either the SE retrieving no results for an educational inquiry or the SE presenting resources that are not suitable to the curriculum. Our contributions also include sharing: a set of educational resources labeled with their grade levels and subjects and a data set comprised of children's search queries; labeled by search query type and grade-level[6].

Outcomes from this study have implications for human intervention, search literacy in the *K*-12 curriculum, and SE functionality issues, which need to be addressed so that existing SEs can better support children when conducting curriculum-related searches. Results from our study also call for the design of multi-objective strategies that consider aspects such as readability, appropriateness, objectivity and educational value in tandem, for filtering and ranking resources in response to children's educational search queries. Discoveries that emerge from our work would directly impact the information retrieval community. They can also be of interest to human computer interaction (e.g. informing the design of search interfaces for children), education (e.g. shaping search literacy curriculum), as well as fairness and privacy (e.g. identifying gaps that cause some niche populations to be underserved).

Background and related work

We discuss in this section how children interact with SEs, the use of these tools in the classroom, and the role of safe-search filters in preventing this audience from accessing inappropriate resources.

Children's interaction with search engines

As reported in the book *Search Engine Society* more than half of Americans turn to SEs at least once a day (Halavais, 2017). With this proliferation, search tools are not limited to mature audiences, as young children are now being introduced early, both at school and home, to searching on the internet (National Center for Education Statistics, 2019). Although children frequently turn to SEs, research has shown that they often experience difficulties in finding resources that satisfy their information needs (Bilal and Gwizdka, 2018; Bilal and Ellis, 2011; Druin *et al.*, 2009; Gossen, 2016). Prior works suggest that this difficulty is often the result of children's insufficient digital literacy skills (Bilal and Gwizdka, 2018; Hague and Payton, 2011), for example, lack of ability in identifying credible resources when using SEs, in addition to varied degrees of prior experience (Landoni *et al.*, 2019) and domain knowledge (Han, 2017; Yamin *et al.*, 2013). Moreover, children favor resources on the first SERP and barely notice that other SERPs exist (Duarte Torres *et al.*, 2010b; Gwizdka and Bilal, 2017). To aid children's search, researchers have introduced several strategies for personalizing resources by readability (Bilal and Gwizdka, 2016; Collins-Thompson *et al.*, 2011; Eickhoff *et al.*, 2010; Tan *et al.*, 2012), assisting children in search query formulation (Dragovic *et al.*, 2016), and offering child-friendly search query suggestions (Duarte Torres *et al.*, 2012; Torres *et al.*, 2014; Vidinli and Ozcan, 2016; Madrazo Azpiazu *et al.*, 2018). Others investigate children's search behaviors and interaction styles when utilizing popular search tools (Madrazo Azpiazu *et al.*, 2017; Azzopardi *et al.*, 2009; Bilal and Gwizdka, 2018). However, to the best of our understanding, there is hardly any evidence in the literature of how existing SEs (including child-oriented ones) handle the filtering and retrieval of resources in response to children's general inquiries, let alone, classroom-related searches.

Children's search engine preference

There exist a number of child-oriented SEs, including the popular Kidrex, Kiddle[7], Kidzsearch and Sweet Search[8]. These SEs could help children complete successful searches and thus improve their overall search experience. However, in some cases, they manually curate materials to be indexed (Broch, 2000; Gyllstrom and Moens, 2010), limiting the amount of resources they make available. Regardless of the support these child-oriented SEs could offer, research has shown that children still prefer to use popular SEs (Madrazo Azpiazu *et al.*, 2017; Foss *et al.*, 2012). In fact, in a survey conducted by Purcell *et al.* (2012), 94 percent of participants who are teachers reveal that their students are likely to use Google for school assignments, as opposed to other SEs. This preference is what motivates us to examine how popular SEs respond to search queries that capture information needs aligned with the grades third to fifth curriculum.

Safe-search

To prevent children from accessing inappropriate content, popular SEs along with their child-oriented counterparts, adopt a safe-search functionality. Safe-search is meant to filter resources with inappropriate content, such as pornography and hate-speech (Google, 2018; Jacob *et al.*, 1999), hence, offering a safer search environment. Traditional safe-search filters may, however, be limited to blacklisted terms and URLs, which can be challenging to update. Blacklisted URLs may be ineffective as they keep changing, whereas blacklisted terms may be deterred by "homographs." Moreover, safe-search may not always be the perfect deterrent (Edelman, 2003; Heiler *et al.*, 2017). On the one hand, resources with inappropriate content pass through the filter. On the other hand, safe-search may be too strict when it comes to filtering resources that are relevant to users' information needs and the context of their searches (Edelman, 2003; Heiler *et al.*, 2017). For instance, as of September 2018, Kidrex's safe-search filter prevents the retrieval of any resources for the search query *organ pipe cactus monument*[9], whereas as of October 2019, Bing's safe search does not retrieve resources for the search query *teeny virus* (see Figure 2). Further, the "No Results Found" pages presented by these SEs do not offer any insights to users about why their queries resulted in no resources retrieved. This coupled with the limitations of safe-search, are among the short-comings we foresee in how this filter responds to classroom-related inquiries. Our study aims to further investigate the limitations of safe-search in handling both inappropriate and curriculum-relevant materials.

Search engines and the classroom environment

The use of SEs in the classroom has proven to be an important asset for enhancing learning (Madrazo Azpiazu *et al.*, 2017). In schools, SEs are often used by children to locate information for educational tasks: from looking up the meaning of words and finding math formulas, to addressing history-related inquiries for which it usually takes longer to find answers to when using printed books (Knight, 2014). Although SEs are now frequently used in schools, search literacy is not always part of the curriculum (Bilal and Gwizdka, 2018; Scott and O'Sullivan, 2005). This serves as an indication that children may either need to be provided with some form of assistance whenever they utilize their preferred SEs for school-related activities (Bilal and Gwizdka, 2018; Foss *et al.*, 2012) or that instead existing SEs should be adapted to offer children a better experience. This adaptation can either be in the form of guidance for search query creation or prioritization of search results that address their information needs in the classroom setting.

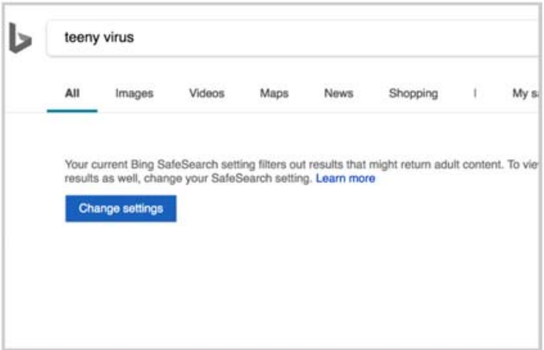
Understanding how resources are prioritized for educational searches is the first step toward outlining how SEs need to be adapted to better serve children in response to inquiries associated with the curriculum, which is why we conduct our study.

Theoretical framework

We follow the theoretical framework proposed by Lewandowski (2012) for evaluating the retrieval effectiveness of SEs, which consists of five pillars: search query selection, results judgement, results collection, results presentation and data analysis. We use search query

(a)

(b)



Notes: (a) KidRex fails to retrieve resources for the search query *organ pipe cactus monument* (September 2018); (b) Bing's safe search fails to retrieve resources for the search query teeny virus (October 2019)

Figure 2. Examples that showcase the limitations of SEs with enabled safe-search functionality in responding to curriculum-related searches

selection and results judgement to guide the setup of our study. Search query selection involves identifying applicable search queries to evaluate. While we do not directly evaluate search queries, our work calls for carefully selecting the right search queries to capture the intent and topics that align with the target user capabilities, different contexts, and classroom task objectives. Results judgement entails selecting assessors for evaluation purposes. This study does not involve external/human assessors. Instead, we rely on expert labeled and applicable ground-truth to support analysis. We adopt the remaining pillars to guide the evaluations in our study. Results collection deals with the selection of resources as well as information pertaining to the resources from the SE under investigation. We especially focus on the resource type and categorization (e.g. educational domain or government agency). Results presentation encompasses weighting the relevance of resources based on position and design, for example, examining the ranking position of the individual resource within a ranked list. In this case, we investigate the position of resources on the ranked list generated by each SE under study. Lastly, data analysis focuses, among other things, on examining the relevance of the results. As we concentrate on an educational setting, we explicitly explore relevance to the user, context, and classroom task objectives.

Method

We employ a simulation-based experimental approach to examine the strengths and limitations of SEs commonly utilized by children for conducting online inquiries pertaining to the third and fifth grade curriculum.

Search engines

The four SEs we study are: Google, Bing, Kidrex and Kidzsearch. We also study the safe-search counterparts in Google and Bing with the aim to investigate how filters available on SEs that are favored by children handle both inappropriate and educational resources. This results in six explored strategies.

Kidrex and Kidzsearch were selected for analysis not only because they are two of the most popular among child-oriented SEs[10], but also because, unlike SEs exclusively designed for children that handpick child-suitable sites (e.g. Sweet Search), they directly rely on mainstream SEs to power their searches. Moreover, these SEs are solely designed for educational purposes and adopt a safe-search to filter inappropriate resources[11].

Search query gathering

We originally collected 349 search queries written by 50 children performing search tasks in the classroom setting[12]. We refer to this search query set as Kids_{QRY} (see sample search queries in Table I). Children who formulated these search queries were in the fourth and fifth grade levels and were assigned information discovery tasks under the supervision of their teacher at an Elementary School in Idaho, USA. As children were presented with the same search tasks, we found several repetitions among the search queries. Hence, for analysis purposes, we only selected distinct search queries, resulting in a total of 100 search queries.

Due to the limited number of child-written search queries, we augmented Kids_{QRY} with 750 simulated search queries. To create these search queries, we turned to book reviews written by children in the third to fifth grades, which we gathered from child-oriented libraries, such as Pikes Peak Library District[13] and Monroe County Public Library[14], that consented to using data for research purposes. Following the premise in Bilal and Gwizdka (2018), we created three types of search queries from these children's reviews: Keywords, Phrases and Questions. Keyword search queries include one or more words that do not form a phrase or sentence; phrase search queries contain two or more words that expresses a single idea, but do not form a complete sentence; and question-type search queries refer to those that start with a question word, e.g. how, why, when, what, and where. Using the NLTK library (Bird *et al.*, 2009), we extract noun phrases and sentences from the review text. We follow the grammar rule definition from NLTK for extracting noun phrases, which is a combination of one or more words that contain a noun with a determinant, adverb, verb, or an adjective (Bird *et al.*, 2009). In this case, we treat

Type	Search query	Simulated?
Keyword	Notrth pole (originally misspelled)	✗
	Dinosaur bones	✓
	Google earth	✗
Phrase	US April holidays	✗
	The Second World War	✓
	Organ pipe cactus national monument	✗
Question	How many people died in the Johnstown flood	✗
	What habitat does an aardvark live in	✗
	How did the Baudelaires parents die	✓

Table I. Examples of children's search queries and simulated counterparts, grouped by type noun phrases that contain only one or two unique words (excluding determiners) as keyword search queries, noun phrases that contain more than two words (including determinants) as phrase search queries, and sentences extracted from the review text that start with a question word as question search queries. It is important to note that for each search query type, we selected top-*n* distinct search queries based on frequency. As shown in Figure 3, simulated and child-written search queries exhibit similar trends, in terms of average length and distribution by type, which is why we deem the simulated search queries suitable for analysis purposes.

Other data sources

For analysis purposes, we take advantage of a number of data sources with different information types.

Child-friendly resources – Cf_{RES}. This is a collection of web resources gathered from DMOZ[15]. In this web directory, resources have been categorized by human annotators into age groups, i.e., kids, teenagers and adults, as well as information types, for example, News and Politics[16]. We extracted 3,000 resources in the following kids' categories: health, news, entertainment, school and sports. These categories were selected because they represent diverse information types that children seek for both their educational and leisure searches (Patel and Singh, 2016).

Educational resources – Edu_{RES}. Upon collaboration with Idaho Digital Learning Academy (IDLA), a K-12 educational institution in Idaho, we gathered a collection of 66,000 educational resources. These resources have been labeled with their subjects and grade levels by educators. For analysis, we extracted 4,000 resources related to Health and Science.

These resources have been labeled with their subjects and grade levels by educators. For analysis, we extracted 4,000 resources related to Health and Science.

Google's bad words – $Google_{BW}$. This is a list that includes 1,400 keywords that have been identified as sexually-explicit by Google[17].

Hate-speech dictionary – Hs_{DICT}. We created a dictionary using 1,040 hate-speech lexicons compiled by hateBase[18], a repository of hate-speech language. Additionally, we included in $\mathrm{Hs}_{\mathrm{DICT}}$ a refined collection of hate-based and offensive language *n*-grams created by Davidson *et al.* (2017).

Hate-based resources – Hs_{RES}. We gathered a collection of 2,000 hate-speech web resources, which were compiled by Hate-speech Movement[19], a site known to report websites that promote violence, supported by the Council of Europe (Silva *et al.*, 2016).

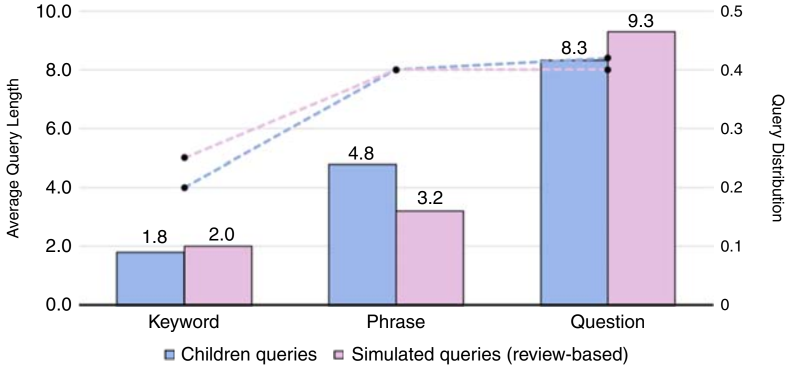


Figure 3.
Average length and distribution by type for search queries in $\mathrm{Kids}_{\mathrm{QRY}}$

Procedures

In this section, we discuss simulation contexts we establish for analysis purposes, the process we follow for retrieving resources that pertain to each context, and how we evaluate retrieved resources.

Simulation contexts. For investigating how SEs respond to inquiries associated with the curriculum, we consider different contexts. To initiate the search process, we use search queries that would trigger the retrieval of resources that satisfy different information needs. We discuss each of the established contexts, as well as the ground truth we use to validate the correctness of SEs in responding to inquiries associated with each context. We present in Table II a description of simulation contexts with corresponding search queries used for initiating the search process. We also include a summary of the information needs with anticipated response from SEs in Table III.

Context-A: children who use search queries that are educational but happen to contain keywords that may be misconstrued as inappropriate. Being that children utilize SEs with the intention of locating materials that pertain to subjects they are taught at school (Gossen, 2016), we simulate this context. For this purpose, we create 1,000 search queries by randomly sampling phrases (of up to tri-grams) relating to children's Health and Science subjects[20], which we extracted from Edu_{RES} (introduced in SECTION other data sources). This resulted in Edu_{QRY}, a list of the top 1,000 educational n-grams, selected according to their Term Frequency-Inverse Document Frequency (TF-IDF) scores. To investigate the effectiveness of SEs when dealing with searches conducted in this context, we examine the number of times SEs under study fail to retrieve results.

Context-B: children who inadvertently access sexually-explicit materials. We examine the performance of SEs in regard to handling sexually-explicit content. To do so, we simulated a context where a user seeks for sexually-explicit materials, with and without safe-search enabled. For simulating the search process, we created 1,000 search queries by

Name	Sources	Context	Resource type	Search query	Size
Edu_{RES}	IDLA Educational institute	A	Educational	Edu_{QRY}	1,000
$Google_{BW}$	Google's bad words	B	Sexually explicit	Bw_{QRY}	1,000
Hs_{DICT}	HateBase and Lexicons generated by Davidson *et al.* (2017)	C	Hate speech	Hs_{QRY}	1,000
Kid's search queries	Child-written search queries and reviews	D and E	Children's education-related search queries	$Kids_{QRY}$	850

Table II. Summary of data sources, simulation contexts and search queries

Context	I want to...	Ideal SE response
A	Locate educational resources pertaining to Health and Science subjects	Retrieve resources for all educational searches
B	Access resources that do not contain sexually-explicit content	Filter sexually-explicit materials
C	Access resources that do not include hate speech and offensive language	Disregard violence-related materials
D	Comprehend the content of resources retrieved	Retrieve resources that align with target user's reading level
E	Identify curriculum-relevant resources on the SERP efficiently	Prioritize education-relevant and non-opinionated resources

Table III. Summary of information needs for different simulation contexts and expected response from SEs randomly sampling lexical items from $Google_{BW}$, which we used to initiate the search on SEs under study. We extract search queries (which we refer to as Bw_{QRY}) from $Google_{BW}$, as these contain inappropriate keywords that would naturally trigger the retrieval of sexually-explicit materials. To validate the correctness of SEs in identifying sexually-explicit materials, we take advantage of WebShrinker[21], a top online website categorizer. WebShrinker examines the content of resources and assigns them categories such as education, adult content, entertainment, and news, based on the type of information they contain and their TLD. We rely on this tool, being that to the best of our knowledge, there is no ground truth to determine if a website includes sexually-explicit content. Following this, we compute the percentage of searches which led to the retrieval of at least one resource categorized as "adult content".

Context-C: children who mistakenly access resources containing violence-related content. We examine how SEs perform when it comes to retrieving resources that are violence-related, as these materials are considered inappropriate for children (American Academy of Pediatrics *et al.*, 2016). For this purpose, we simulated a context where a user seeks violence-related materials online. In doing this, we used a number of search queries containing hate-based keywords to initiate the search. We created these search queries (which we refer to as Hs_{QRY}) by randomly selecting 1,000 lexical items from Hs_{DICT}. We rely on searches triggered using search queries in Hs_{QRY}, being that instinctively, they would lead to the retrieval of resources that contains violence-related content. To the best of our understanding, there is no ground truth that determines the degree to which a document is violence related. As a result, we rely on the hate-speech detection algorithm introduced by Davidson *et al.* (2017), in order to label resources that contain either offensive language or hate-speech[22]. Similar to the assessment performed in Context-B, we compute the percentage of searches that retrieve at least one resource categorized as violence related.

Context-D: children who access resources that do not match their reading skills. Readers can comprehend a text when they understand 75 percent of its content (Benjamin, 2012). This is why it is imperative that resources retrieved in response to children's search queries match their reading abilities. With this in mind, we investigate the reading levels of resources retrieved for education-related inquiries conducted by school-aged children. To simulate the search process, we used search queries that have been written by children, which we extract from $Kids_{QRY}$.

To determine the average readability of the retrieved web results, we use a number of traditional readability formulas: Flesch-Kincaid, Dale-Chall, and Smog (Benjamin, 2012) – computed using the Textstat readability analyzer library made available by Bansal and Aggarwal (2019). Prior to experimentation, we considered other alternative readability formulas as well. However, based on the premise that the difference among most formulas does not necessarily matter when used in a web context like the one in our study, since websites target group levels (e.g. grade or high school level) (Nielsen, 2017), and the fact that the selected formulas are well-known and open access, we deem them as the ideal choice for analysis purposes. Further, even though these formulas rely on shallow features, without necessarily examining semantic aspects to determine text complexity, they are simple and have demonstrated their applicability in determining reading levels in web documents (Bilal and Huang, 2019; Crossley *et al.*, 2017; Vajjala and Meurers, 2013).

Context-E: children experiencing difficulty in identifying an educational resource on the SERP. A myriad of information sources exists on the web, containing URLs that showcase their TLD. It is common for educational institutions to use the TLD ".edu"[23] e.g., mit.edu, while commercial sites use the TLD ".com" (Group *et al.*, 2015), e.g., Amazon.com or StackOverFlow.com. Other popular TLDs are ".org," which is used for educational and non-profit organizations, as well as ".net" and ".gov," used by Government entities (Group *et al.*, 2015).

Being that resources are mostly associated with TLDs based on information they contain, we examine the degree to which the aforementioned TLDs are prioritized for children's educational searches. To simulate the search process, we also depend upon search queries in $Kids_{QRY}$. For this analysis, we compute the average position of the resources retrieved in response to these search queries for each of the TLDs under study.

Search inquiries. Based on the aforementioned simulation contexts, a trained Graduate Research Assistant (GRA) wrote a script to automate the search process, which ran between January and July 2018. Resources retrieved by Google[24], Bing[25], and their safe-search counterparts were accessed through their API services. Being that Kidrex and Kidzsearch do not make available a search API, we wrote a script to automatically perform search and retrieval tasks.

SERP analysis. As prior works demonstrate, children usually do not go beyond the first SERP when viewing search results (Gwizdka and Bilal, 2017; Duarte Torres *et al.*, 2010b). Thus, we only evaluate the top 10 results on the first SERP (excluding ads). The aforementioned GRA along with an Undergraduate Research Assistant created several scripts to automate analysis for SE comparison of results and to conduct statistical tests.

Results and analysis

We report the outcomes of the analysis conducted to understand how SEs respond to search queries associated with the classroom setting:

RQ1. Do SEs effectively respond to search queries that include terminology pertinent to the third–fifth grade curriculum?

We aim to determine SE's success based on its ability to retrieve at least one result for searches conducted with an educational intent. For Context-A, we found that SEs do not always retrieve resources for searches initiated with search queries in Edu_{QRY}. As shown in Table IV, Kidrex and Kidzsearch fail to retrieve resources for close to 13 percent of these searches.

Google and Bing, with and without the safe-search option, were less restrictive than the child-oriented SEs in offering resources for searches simulated using Edu_{QRY}. In this context, Google, its safe-search counterpart and Bing (without its safe-search counterpart), retrieve results for all of these searches; while Bing's safe-search does not retrieve results for 2.3 percent of them.

A statistical difference was found for the way SEs respond to the educational searches for search queries in Edu_{QRY} (Kruskal–Wallis; $p < 0.05$). *Post hoc* pairwise comparisons show that the main difference is between the child-oriented SEs, i.e., Kidrex and Kidzsearch, and the popular ones (Mann–Whitney; $p < 0.05$ for Google, Bing, and their safe-search counterparts, respectively). There is no significant difference between Kidrex and Kidzsearch (Mann–Whitney; $p = 0.462$), indicating that these SEs are more restrictive when

Search Engine	Searches that retrieve no results
Bing	0%
Bing (Safe-search)	2.3%
Google	0%
Google (Safe-search)	0%
Kidrex	13.4%
Kidzsearch	12.3%

Table IV. Percentage of searches initiated using Edu_{QRY} that led to no results. Green for positive results, red negative handling educational searches that are relevant to children's Health and Science subjects. We find that with Kidrex and Kidzsearch, some searches that lead to no results are initiated with keywords like breast tissue, normal sperm cell, and female sexual anatomy. Both Kidrex and Kidzsearch are powered by Google custom search and adopt Google safe-search technology, indicating that even if these SEs account for 1 percent of Google's 3.5bn daily searches (Aleksandra, 2019), children could be restricted from accessing approximately 10 percent of these searches a day. This is a detriment on the part of the aforementioned child-oriented SEs, especially with the fact that these inquiries are related to subjects that are included in the *K*-12 curriculum:

RQ2. Do SEs deter access to inappropriate content?

In this case, we examined whether SEs retrieve no results and include either sexually-explicit or violence-related content among retrieved resources on the SERP.

The degree to which SEs filter sexually-explicit material

As showcased in Table V, Kidrex and Kidzsearch prevent the retrieval of resources for 57 and 39 percent of searches initiated by search queries in Bw_{QRY}, respectively. This is advantageous, since lack of resources retrieved serves as an indication that the corresponding SE is able to correctly identify potential results as inappropriate and prevent children from accessing them. However, for the majority of remaining searches for which resources were retrieved, child-oriented SEs include at least one resource known to be sexually-explicit (for 24 and 15 percent of the cases respectively, from Table V) in their top 10 results. Being that these child-oriented SEs are designed for educational use, it is unfortunate that they are not always capable of preventing the retrieval of sexually-explicit resources as this implies that children are prone to being inadvertently exposed to sexually-explicit materials.

SEs that do not have safe-search enabled (i.e. Google and Bing) retrieve resources for all searches in Bw_{QRY}. This is anticipated, as these SEs are designed for diverse audiences and their purpose does not include filtering sexually-explicit content. It is worth noting, however, that with safe-search engaged, these SEs are not perfect deterrents. Considering that these are the SEs favored by children and that parents and teachers specifically turn on safe-search options for the purpose of preventing young users from accessing sexually-explicit content (Calvert, 2015; Willard, 2007), this becomes even more worrisome.

The degree to which SEs filter violence-related material

In exploring how safe-search handles violence-related content (based on the inquiries detailed in Context-C), we find that the child-oriented SEs are more effective in identifying the context to be inappropriate for children when compared to Google and Bing (see Table V). Results show that

Table V. SE response to search queries that can lead to sexually-explicit (*S*) or violence-related (*V*) content

Search Engine	Search results that are		No results for searches that are	
	S	V	S	V
Bing	70%	27%	0%	0%
Bing (Safe-search)	55%	21%	29%	7%
Google	56%	26%	0%	0%
Google (Safe-search)	38%	24%	5%	2%
Kidrex	24%	10%	39%	17%
Kidzsearch	15%	7%	57%	18%

Notes: Green positive; red negative

these SEs prevent the retrieval of violence-related resources for approximately 20 percent of the searches simulated using search queries in Hs_{QRY}. However, 10 percent of the searches that led to results in Kidrex and 7 percent of searches for which resources were retrieved in Kidzsearch included at least one resource labeled as violence-related, as shown in the second column of Table V. While we do not argue for SEs to act as a censor, in this context, it is problematic when children are exposed to violence-related content for school-related inquiries. It could be possible that through this exposure, children adopt vocabulary related to hate-speech or offensive language as slangs or natural language. As studies show that children easily pick up such profane language based on how frequent they encounter these terminologies (Jay and Janschewitz, 2012; Jay and Jay, 2013), it is imperative for SEs to not foster children's access to violence-related resources for their curriculum-related inquiries.

When it comes to the searches triggered using search queries in Hs_{QRY}, Google and Bing (without the safe-search functionality) retrieve resources for all of them. This is expected, as these tools are not meant to filter resources with hate-speech or offensive language. However, their safe-search counterparts, that should disregard violence-related resources, are not always effective for this purpose. Results show that Google's and Bing's safe-search retrieve resources for more than 90 percent of the search queries in Hs_{QRY}. Additionally, the percentage of searches conducted on Google and Bing that include hate-speech and offensive language were comparable to their non-safe-search counterparts. We found that there was an 81 percent overlap between resources retrieved from Google (with and without safe-search)[26]. Similarly, there was a 66 percent overlap between the set of resources relevant to violence-related searches that were retrieved by Bing and its safe-search counterpart. These findings demonstrate that the safe-search functionality on Google and Bing may be limited in handling searches that pertain to violence-related inquiries. Again, being that these tools are the most utilized by children, it is concerning that they do not adequately filter violence-related content:

RQ3. Do SEs provide resources that align with task objectives and user capabilities?

We examined SE ranking performance in providing resources that are suitable to inquiries associated with the third–fifth grade curriculum. For this analysis, we rely on search queries in $Kids_{QRY}$ and particularly focus on three aspects: resource readability to capture user capabilities, as well as position of education-related resources on the ranked list and the position of those resources that are potentially opinionated to account for classroom task objectives.

The degree to which resources match children's reading skills

In estimating the average reading level of resources retrieved using search queries in $Kids_{QRY}$ for inquiries pertaining to Context-D, our results show that resources retrieved in response to children's search queries do not match their reading abilities. As shown in Figure 4, across our selected readability formulas the resources retrieved have distributions closer to a high school level than an elementary school level, making these resources difficult for our target users to understand[27]. As presented in Table VI, the average reading level among the resources retrieved are relatively high for all SEs (eighth grades and above). Our results echo existing works which highlight the fact that resources retrieved for child-initiated searches do not always match their reading skills (Bilal, 2013; Bilal and Boehm, 2017; Bilal and Huang, 2019). Considering the fact that most web resources are written in the eighth grade reading levels (Nielsen, 2017), for children in the fifth grade that access such resources, being able to comprehend some of its content may be feasible. However, this would not be the case if the target audience were to be in the third grade, as there is a wide gap in the reading levels. To better serve children, it is important that resources provided in response to their search queries include vocabulary with which they are familiar (Hoa Loranger, 2017; Nielsen, 2015). Consequently, children's inability to comprehend the content of retrieved resources might result in failed search tasks.

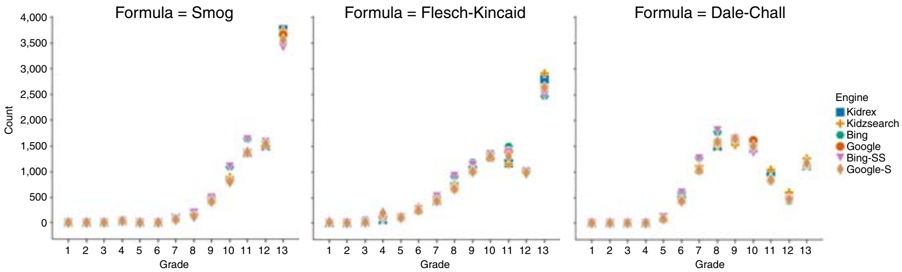


Figure 4. Distribution of grade levels of individual resources retrieved for $Kids_{QRY}$

Notes: We grouped under grade 13 those scores that were above the twelveth grade, indicating that they are the reading level of adults. We excluded search queries which resulted in no resource being retrieved and removed those retrieved resources that had no text

Table VI. Average grade levels of resources retrieved for $Kids_{QRY}$

Search Engine	Smog	Readability formulas Flesch–Kincaid	Dale–Chall
Bing	8th	10th	9th
Bing (Safe-search)	8th	10th	9th
Google	9th	11th	9th
Google (Safe-search)	8th	10th	9th
Kidrex	8th	10th	9th
Kidzsearch	8th	10th	9th

The rate at which resources from different domains are prioritized

We also analyze the TLD of resources retrieved from conducting information discovery tasks that are associated with the classroom setting. To do so, we compute the average position of resources retrieved in response to search queries in $Kids_{QRY}$ for each of the TLDs introduced in Context-E[28]. In this case, we treat resources from the ".edu" domain to be more applicable, being that they are more likely to include verifiable and educative information, when compared to resources from other domains.

From our analysis, we find that resources from the ".edu" domain are consistently ranked low (see Table VII). Results show that across all SEs, resources from ".edu" do not rank among the top 3 on the SERP, which is not the case for ".com" and ".org" websites. We anticipate this as the TLD of the most visited sites on the web is ".com," followed by ".org"[29]. Moreover, oftentimes, resources belonging to these domains may be influenced

Table VII. Domain distribution using $Kids_{QRY}$

Search engine	Wikipedia	Searches with domain from .org	.gov	.com	.edu	.net
Bing	2nd – 36%	2nd – 18%	3rd – 90%	1st – 0.4%	4th – 92%	4th – 85%
Bing (Safe-search)	2nd – 36%	2nd – 17%	3rd – 90%	1st – 0.4%	5th – 92%	3rd – 84%
Google	2nd – 45%	2nd – 23%	2nd – 90%	1st – 0.7%	4th – 91%	4th – 85%
Google (Safe-search)	2nd – 45%	2nd – 23%	2nd – 90%	1st – 0.7%	4th – 90%	4th – 85%
Kidrex	2nd – 38%	2nd – 12%	3rd – 85%	1st – 0.5%	4th – 86%	3rd – 86%
Kidzsearch	2nd – 71%	2nd – 16%	4th – 87%	1st – 0.8%	4th – 87%	3rd – 85%

Notes: (Average rank position – % of searches that had no resource retrieved from the respective domain). As Wikipedia has a .org TLD, it is included among the .org searches by SEO techniques (Lewandowski, 2011). It is worth noting that although some ".edu" websites may be collegiate in nature (e.g. mit.edu), we verified that, for the most part, ".edu" resources retrieved using Kids_{QRY} had a lower average reading level compared to other TLDs. Given that children sequentially select resources from the SERP, it is possible that they are not able to access relevant resources from the ".edu" domain at all if these resources are consistently positioned low on the SERP. Thus, it is essential that education-relevant resources are prioritized for searches conducted with the intent of locating curriculum-related materials.

Moreover, being that outcomes show that most of the resources retrieved in response to Kids_{QRY} are not from the ".edu" TLD (see Table VII), children may be more likely to gather information from popular sites as opposed to those offering curated academic materials for their educational searches. These resources could be opinion-based rather than factual causing children to put faith in something other than credible and instructional information.

The extent to which resources that are not written by experts are potentially favored by SEs

In addition to examining TLDs, we investigate the average position assigned to Wikipedia pages. We conduct this experiment being that Wikipedia pages are collectively written by users that may not necessarily be experts. Further, this information source may not always include content curated for school-aged children and exposure to text-dominated web pages with few pictures can leave children confounded. It is also important to note that Wikipedia has a SE which does not have a safe-search and therefore can inadvertently expose children to inappropriate content as well. As seen in Table VII, all SEs rank Wikipedia pages higher and retrieve them more often than resources from the ".edu" domain. It is imperative that for searches formulated by school-aged children, child-suitable and educational sites are ranked higher than Wikipedia pages, as existing research show that web resources appealing to children may contain fewer texts and more graphics (Gossen *et al.*, 2011).

Discussion

Based on the outcomes from the analysis we conducted in the results and analysis section, we observed differences in the way child-oriented SEs restricted access to inappropriate material, when compared to popular SEs. Unfortunately, this affected how they responded to educational inquiries as we found that these child-oriented SEs often prevented the retrieval of resources in response to searches conducted for locating curriculum-related materials. This was especially prominent in searches where the search queries contained keywords that suggested that the user intends to access inappropriate materials (e.g. female sexual anatom*y* which is related to human anatomy). Having educational searches fail to retrieve results is detrimental as this can prevent children from accessing the right resources and lead them to believe no resources exist, thus discontinuing their search. Incomplete search sessions can leave children with either no information or mistrust of SEs, making them believe that these tools cannot provide them with vital information.

When it comes to inappropriate materials mistakenly passing through safe-search filters, we found this to be a limitation on all SEs under study, especially on Google and Bing. Inadvertently exposing children to sexually-explicit and violence-related materials can have both long- and short-term effects on children's behavior, mental health, and social interactions. Sexual media provides an inaccurate perception of adult relationships and can lead to risky behavior from children (Albertson *et al.*, 2018). With violent media, the long-term effect is more substantial, but both show an increase in aggressive behavior, thoughts, and angry feelings (Bandura, 2016; Bushman, 2019).

Across all examined SEs, we found that resources retrieved for children's educational searches were aimed at users with reading abilities above the fifth grade (the maximum grade level of children in our study). We found this to be especially notable in Google (without the safe-search option enabled) which aligns with the finding by the authors in (Bilal and Boehm, 2017; Bilal and Huang, 2019). This is concerning, as Google is the favorite amongst children and also the most utilized SE in the school environment (Purcell *et al.*, 2012). Further, we found that Google is most likely to favor resources that are treated as relevant to average online users (i.e. those with popular TLDs or from Wikipedia), when compared to those that are specific to the educational domain. Being that result listings can be influenced by popularity or SEO techniques (Lewandowski, 2012; Thurow, 2015), the adoption of site optimization strategies by creators of curriculum-relevant materials could potentially increase the visibility of these resources in response to curriculum-related inquiries. Another issue affecting the ranking of resources is the SE algorithm bias. While SE algorithms determine the degree to which resources are relevant for ranking purposes, it may be challenging to infer if a resource is ranked higher due to its content or popularity (Novin and Meyers, 2017). Being able to mitigate this bias when it comes to classroom-related inquiries would be essential for the classroom context. When the results of a search are either beyond the comprehension skills of children, contain information that may not necessarily be verifiable, or are irrelevant to the curriculum, the results are not useful or relevant to them. Again, this can lead to a child having a misconception of the SE, being that it cannot provide accurate answers to their inquiries in a way that can be understood by them.

Search literacy and intervention by educators could help mitigate some of the misconceptions that children face when using SEs. Both teachers who shape curriculum, and school librarians who offer support in information literacy instruction, could teach children how to identify credible resources, cite online and offline sources (Thurow, 2015; Vasinda and Pilgrim, 2019), as well as properly search and critically evaluate search listings. For instance, it may be possible that the first SERP includes results that contain irrelevant information. However, if children could go past this page, then they could find better resources among the results. Unfortunately, given the high ratio of students to teachers in the classroom, it is not always feasible to have teachers monitor all the searches conducted by students under their watch. To this end, it is important that SE algorithms are improved to better support both the teachers and students utilizing them for classroom inquiry tasks.

The areas that have been proposed for exploration in our study are not exhaustive, as the needs of users and environments for search are ever changing, and SEs are not used in isolation. This means that the guidance of teachers and collaborations with peers will still be necessary for students to get the most out of using SEs (Knight and Mercer, 2015).

Outcomes from our empirical studies highlight the need for investigating ways to improve SE retrieval algorithms by designing strategies that would take the reading level of resources into consideration, examine the degree to which a resource is non-opinionated, evaluate resource appropriateness, as well as examine the education pertinence of its content in tandem, for prioritizing resources in response to search queries associated with the *K*-12 curriculum.

Limitations

A limitation of our work resulted from the small amount of unique search queries gathered from school-aged users conducting search tasks in the classroom (i.e. 100). We resorted to enhancing our search query corpus through analyzing and extracting search queries from child-written reviews.

Another limitation of this study resides in only focusing on children in the third to fifth grades, as opposed to the complete K-12 spectrum. Although this was a starting point to examine how SEs respond to searches initiated by these group of users, an extensive study would require investigating the complete K-12 spectrum, which we plan to do as future work.

One final limitation would be the lack of first-hand feedback from students in the third to fifth grades on SEs responding to their inquiries. While this would require a more in-depth user study, we address this limitation by instead adopting a simulation-based approach.

Conclusion

In this work, we simulated searches using search queries that capture information needs in different contexts (3,850 unique search queries). We studied six strategies based on four SEs: Google, Bing, Kidrex, and Kidzsearch. Using these SEs, we evaluated their performance in terms of offering resources that are relevant to task objectives and user capabilities in response to classroom-related inquiries.

The findings of the presented work reveal that although SEs oriented to children (e.g. Kidrex and Kidzsearch) are effective for acting as deterrent for inappropriate resources, they often do so at the detriment of responding to education-relevant searches. Moreover, both child-oriented and popular SEs retrieve resources that do not align with the reading skills of the target audience, with this especially prominent on Google. Further, all SEs under study tend to favor resources from Wikipedia, as well as those that have popular TLDs, when compared to resources that specifically target the educational domain.

Our findings have methodological and practical implications. From a methodological standpoint, previous research works directly rely on query logs for analysis purposes (Duarte Torres and Weber, 2011; Duarte Torres *et al.*, 2010a). Our work instead bypasses this need, as we employ a simulation-based approach which allowed for the efficient automation of the data gathering and analysis process. Future research can supplement inferences we have made as a result of these simulated queries with feedback from a user study with children.

From a practical perspective, outcomes of our work reveal the need to integrate search literacy skills into the grade school curriculum, as it can translate into more successful search sessions, i.e., those that lead to content that is geared toward children and relevant to the curriculum. While teachers can help identify what is relevant to the curriculum, school librarians can leverage their expertise in document retrieval to aid children in identifying the resources relevant to the classroom environment. Teachers and librarians working in tandem can enable the learning of search literacy skills. We also identify gaps that must be addressed in SE design that can support the task at hand. Doing so would require leveraging knowledge from the emerging Search as Learning community, which focuses on facilitating learning to search while searching for learning (Gwizdka *et al.*, 2016). It would also build upon multi-objective strategies for information retrieval (Van Doorn *et al.*, 2016) focused on optimizing ranking of resources in response to multiple criteria (in our case, readability, appropriateness, education-pertinence, and objectivity, to name a few).

Notes

1. www.alarms.org/kidrex/
2. www.kidzsearch.com/
3. The iGeneration users are children born after 2010 (Minal Anand, 2019).
4. Some school systems across the USA have instituted the use of custom-based SEs for classroom-related tasks. However, for analysis purposes, we focus on SEs that children are more familiar with and that they are likely to use for class-related assignments (Purcell *et al.*, 2012).
5. In this work, we follow the USA grade school system.

6. Download at http://bit.ly/kids_search_queries
7. www.kiddle.co/
8. www.sweetsearch.com
9. Child search query sampled among the ones introduced in Section Search Query Gathering.
10. www.ilovefreesoftware.com/28/featured/safe-search-engine-for-kids.html
11. www.kidzsearch.com/about.html, or www.alarms.org/kidrex/parents/about.html
12. IRB approval number: 131-SB16-103
13. https://ppld.org/
14. https://mcpl.info/
15. http://dmoztools.net/Kids and Teens/
16. We are aware that DMOZ is outdated and sites are curated by volunteers, who may not necessarily be subject experts. However, since it has labeled data associated with school-related resources for our target audience, is publicly and freely accessible, and it is still leveraged in research works related to children's search (Torres *et al.*, 2014), we deem it applicable for our study.
17. https://code.google.com/archive/p/badwordslist/
18. www.hatebase.org/
19. https://nohatespeechmovement.org/
20. We only consider Health and Science, as these subjects are more likely to have terms that may be misinterpreted by safe-search as inappropriate, when compared to subjects such as Mathematics or Government. For these subjects, Biological Evolution, Earth and Human Activity, Growth and Development, as well as Human Anatomy are defined topics that must be included in lesson plans for the third–fifth grades classroom (Next Generation Science Standards, 2019; Telljohann *et al.*, 2009). Some assignments associated with these topics are selective evolution of dog breeds, sexual and asexual reproduction in plants, and puberty and the reproductive system (Teachers Pay Teachers, 2019).
21. www.webshrinker.com/
22. The algorithm in Davidson *et al.* (2017) was originally trained for detecting hate-speech and offensive language on tweets. We empirically demonstrated the validity of this algorithm by accurately labeling 95% of resources in Hs_{RES} to be violence-related and all resources in Cf_{RES} as (non-) violence-related.
23. https://net.educause.edu/faq/eligibility
24. https://developers.google.com/custom-search/json-api/v1/overview
25. https://azure.microsoft.com/en-us/services/cognitive-services/bing-web-search-api/
26. To determine overlap, we counted the number of resources retrieved by the safe-search counterpart of a SE, that were also found among the top 10 resources retrieved by that SE for the same search query.
27. Prior works append terms such as "for children" to search queries to deem them child-friendly for analysis purposes (Gupta and Hilal, 2012). We verified that by following this same approach to simulate searches on Google and Bing using search queries in $Kids_{QRY}$, the readability scores did not match the reading level of the target audience in our study.
28. In this work, we are aware that TLD registration does not include child-safe filtering. However, being that TLDs are applicable to all SEs under study with respect to the resources being retrieved, it allows for comparability.
29. www.alexa.com/topsites (March 2018).

References

Albertson, K.M., Moreno, M.A., Garrison, M.M., Evans, Y.N. and Ahrens, K.R. (2018), "Impacts of media on sexual behaviour and relationships among youth in foster care", *Child and Family Social Work*, Vol. 23 No. 1, pp. 88-96.

Aleksandra (2019), "A year in search: how many Google searches per day?", available at: https://seotribunal.com/blog/how-many-google-searches-per-day/ (accessed April 24, 2019).

American Academy of Pediatrics *et al.* (2016), "Aap council on communications and media. virtual violence. pediatrics. 2016; 138 (1): e20161298", *Pediatrics*, Vol. 138 No. 4, p. e2399.

Azzopardi, L., Glassey, R., Lalmas, M., Polajnar, T. and Ruthven, I. (2009), "Puppyir: designing an open source framework for interactive information services for children", *Proceedings of the Annual Workshop on Human Computer Interaction and Information Retrieval, Vol. 44*, *Washington, DC.*

Bandura, A. (2016), "50 the power of observational learning through social modeling", *Scientists Making a Difference: One Hundred Eminent Behavioral and Brain Scientists Talk About Their Most Important Contributions*, pp. 235.

Bansal, S. and Aggarwal, C. (2019), "Textstat 0.5.61", available at: https://pypi.org/project/textstat/ (accessed August 18, 2019).

Benjamin, R.G. (2012), "Reconstructing readability: recent developments and recommendations in the analysis of text difficulty", *Educational Psychology Review*, Vol. 24 No. 1, pp. 63-88.

Bilal, D. (2013), "Comparing Google's readability of search results to the flesch readability formulae: a preliminary analysis on children's search queries", *Proceedings of the Association for Information Science and Technology*, Vol. 50 No. 1, pp. 1-9.

Bilal, D. and Boehm, M. (2017), "Towards new methodologies for assessing relevance of information retrieval from web search engines on children's queries", *Qualitative and Quantitative Methods in Libraries*, Vol. 2 No. 1, pp. 93-100.

Bilal, D. and Ellis, R. (2011), "Evaluating leading web search engines on children's queries", *Lecture Notes in Computer Science*, Vol. 6764, pp. 549-558.

Bilal, D. and Gwizdka, J. (2016), "Children's eye-fixations on Google search results", *Proceedings of the Association for Information Science and Technology*, Vol. 53 No. 1, pp. 1-6.

Bilal, D. and Gwizdka, J. (2018), "Children's query types and reformulations in Google search", *Information Processing Management*, Vol. 54 No. 6, pp. 1022-1041.

Bilal, D. and Huang, L.-M. (2019), "Readability and word complexity of serps snippets and web pages on children's search queries: Google vs bing", *Aslib Journal of Information Management*, Vol. 71 No. 2, pp. 241-259.

Bilal, D. and Kirby, J. (2002), "Differences and similarities in information seeking: children and adults as web users", *Information Processing & Management*, Vol. 38 No. 5, pp. 649-670.

Bird, S., Klein, E. and Loper, E. (2009), *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, O'Reilly Media, Inc.

Broch, E. (2000), "Children's search engines from an information search process perspective", *School Library Media Research*, Vol. 3.

Buckingham, D. (2015), "Defining digital literacy-what do young people need to know about digital media?", *Nordic Journal of Digital Literacy*, Vol. 10, Jubileumsnummer, pp. 21-35.

Bushman, B.J. (2019), "Aggression and violence defined", *Advanced Social Psychology: The State of the Science*, pp. 179-180.

Calvert, H. (2015), "Letting go of stand-alone technology", *The Reading Teacher*, Vol. 69 No. 2, pp. 147-155.

Campbell, A.J., Ridout, B.F., Linden, M., Collyer, B. and Dalgleish, J. (2018), "A preliminary understanding of search words used by children, teenagers and young adults in seeking information about depression and anxiety online", *Journal of Technology in Human Services*, Vol. 36 No. 4, pp. 208-221.

Collins-Thompson, K., Bennett, P.N., White, R.W., De La Chica, S. and Sontag, D. (2011), "Personalizing web search results by reading level", *Association for Computing Machinery Conference on Information and Knowledge Management*, pp. 403-412.

Crossley, S.A., Skalicky, S., Dascalu, M., McNamara, D.S. and Kyle, K. (2017), "Predicting text comprehension, processing, and familiarity in adult readers: new approaches to readability formulas", *Discourse Processes*, Vol. 54 Nos 5-6, pp. 340-359.

Davidson, T., Warmsley, D., Macy, M. and Weber, I. (2017), "Automated hate speech detection and the problem of offensive language", *Association for the Advancement of Artificial Intelligence Conference on Web and Social Media*.

Dragovic, N., Madrazo Azpiazu, I. and Pera, M.S. (2016), "Is sven seven? A search intent module for children", Special Interest Group on Information Retrieval, ACM, New York, NY, pp. 885-888.

Druin, A., Foss, E., Hatley, L., Golub, E., Guha, M.L., Fails, J. and Hutchinson, H. (2009), "How children search the internet with keyword interfaces", *Interaction Design and Children*, ACM, Como, pp. 89-96.

Duarte Torres, S. and Weber, I. (2011), "What and how children search on the web", *Proceedings of the 20th ACM International Conference on Information and Knowledge Management, ACM*, pp. 393-402.

Duarte Torres, S., Hiemstra, D. and Serdyukov, P. (2010a), "An analysis of queries intended to search information for children", *Proceedings of the Third Symposium on Information Interaction in Context, ACM*, pp. 235-244.

Duarte Torres, S., Hiemstra, D. and Serdyukov, P. (2010b), "Query log analysis in the context of information retrieval for children", Special Interest Group on Information Retrieval, ACM, pp. 847-848.

Duarte Torres, S., Hiemstra, D., Weber, I. and Serdyukov, P. (2012), "Query recommendation for children", *Conference on Information and Knowledge Management, ACM*, pp. 2010-2014.

Edelman, B. (2003), "Empirical analysis of Google safesearch", *Berkman Center for Internet and Society*, Vol. 14.

Eickhoff, C., Serdyukov, P. and de Vries, A.P. (2010), "Web page classification on child suitability", *Proceedings of the 19th International Conference on Information and Knowledge Management, ACM*, pp. 1425-1428.

Fails, J.A., Pera, M.S., Anuyah, O., Kennington, C., Wright, K.L. and Bigirimana, W. (2019), "Query formulation assistance for kids: what is available, when to help & what kids want", *Proceedings of the 18th ACM International Conference on Interaction Design and Children, ACM*, pp. 109-120.

Foss, E., Druin, A., Brewer, R., Lo, P., Sanchez, L., Golub, E. and Hutchinson, H. (2012), "Children's search roles at home: implications for designers, researchers, educators, and parents", *Journal of the Association for Information Science and Technology*, Vol. 63 No. 3, pp. 558-573.

GlobalStats (2019), "Search engine market share worldwide", available at: https://gs.statcounter.com/search-engine-market-share (accessed August 15, 2019).

Google (2018), "Block explicit results on Google using safesearch", available at: https://goo.gl/22Brm5 (accessed May 2, 2018).

Google (2019), "Google for education", available at: https://teachercenter.withgoogle.com/training (accessed August 20, 2019).

Gossen, T. (2016), *Search Engines for Children: Search User Interfaces and Information-Seeking Behaviour*, Springer, Berlin.

Gossen, T., Low, T. and Nurnberger, A. (2011), "What are the real differences of children's and adults' web search?", *Proceedings of the 34th International SIGIR Conference on Research and Development in Information Retrieval, ACM,* pp. 1115-1116.

Graham, L. and Metaxas, P.T. (2003), "Of course it's true; I saw it on the Internet!: critical thinking in the internet era", *Communications of the ACM*, Vol. 46 No. 5, pp. 70-75.

Group, N.W. and Postel, J. (2015), "Domain name system structure and delegation", *Network Working Group – RFC 1591*, *March 1994*.

Gupta, N. and Hilal, S. (2012), "Analysis of web content filtering factors and the impact of sieve coupons", *International Journal of Engineering and Technology*, Vol. 4 No. 4, pp. 228-237.

Gwizdka, J. and Bilal, D. (2017), "Analysis of children's queries and click behavior on ranked results and their thought processes in Google search", *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, ACM*, pp. 377-380.

Gwizdka, J., Hansen, P., Hauff, C., He, J. and Kando, N. (2016), "Search as learning (sal) workshop 2016", *Proceedings of the 39th International SIGIR Conference on Research and Development in Information Retrieval, ACM*, pp. 1249-1250.

Gyllstrom, K. and Moens, M.-F. (2010), "Wisdom of the ages: toward delivering the children's web with the link-based agerank algorithm", *Proceedings of the 19th International Conference on Information and Knowledge Management, ACM*, pp. 159-168.

Hague, C. and Payton, S. (2011), "Digital literacy across the curriculum", *Curriculum Leadership*, Vol. 9 No. 10.

Halavais, A. (2017), *Search Engine Society*, John Wiley and Sons, Cambridge, MA.

Han, H. (2017), "Understanding children's help-seeking behaviors: effects of domain knowledge".

Heiler, M., Schaer, M., Todorovic, N., Nittka, R., Fischbacher, T. and Dragoi, L. (2017), "Protecting users from inappropriate sensitive or offensive search results", US Patent App. 14/841,078.

Hoa Loranger (2017), "Plain language is for everyone, even experts", available at: www.nngroup.com/articles/plain-language-experts/ (accessed August 21, 2019).

Hussain, W., Sohaib, O., Ahmed, A. and Khan, M.Q. (2011), "Web readability factors affecting users of all ages", *Australian Journal of Basic and Applied Sciences*, Vol. 5 No. 11, pp. 972-977.

Ieong, S., Mishra, N., Sadikov, E. and Zhang, L. (2012), "Domain bias in web search", *Proceedings of the Fifth International Conference on Web Search and Data Mining, ACM*, pp. 413-422.

Introna, L. and Nissenbaum, H. (2000), "Defining the web: the politics of search engines", *Computer*, Vol. 33 No. 1, pp. 54-62.

Jacob, V., Krishnan, R., Ryu, Y.U., Chandrasekaran, R. and Hong, S. (1999), "Filtering objectionable internet content", *Proceedings of the 20th International Conference on Information Systems, AIS*, pp. 274-278.

Jay, K.L. and Jay, T.B. (2013), "A child's garden of curses: a gender, historical, and age-related evaluation of the taboo lexicon", *The American Journal of Psychology*, Vol. 126 No. 4, pp. 459-475.

Jay, T. and Janschewitz, K. (2012), "The science of swearing", available at: www.psychologicalscience.org/observer/the-science-of-swearing (accessed June 2, 2019).

Knight, S. (2014), "Finding knowledge – what is it to know when we search?", in König, R. and Rasch, M. (Eds), *Society of the Query Reader: Reflections on Web Search*, Institute of Network Cultures, Amsterdam.

Knight, S. and Mercer, N. (2015), "The role of exploratory talk in classroom search engine tasks", *Technology, Pedagogy and Education*, Vol. 24 No. 3, pp. 303-319.

Landoni, M., Matteri, D., Murgia, E. and Pera, M.S. (2019), "Sonny cerca! evaluating the impact of using a vocal assistant to search at school", *Proceedings of the Conference and Lans of the Evaluation Forum*, *Lugano*, *September 12*.

Laxman, K. (2010), "A conceptual framework mapping the application of information search strategies to well and ill-structured problem solving", *Computers & Education*, Vol. 55 No. 2, pp. 513-526.

Lewandowski, D. (2011), "The influence of commercial intent of search results on their perceived relevance", *Proceedings of the 2011 iConference, ACM*, *Hershey, PA*, pp. 452-458.

Lewandowski, D. (2012), "A framework for evaluating the retrieval effectiveness of search engines", Next generation search engines: advanced models for information retrieval, IGI Global, pp. 456-479.

Lovato, S.B., Piper, A.M. and Wartella, E.A. (2019), "Hey Google, do unicorns exist?: conversational agents as a path to answers to children's questions", *Proceedings of the 18th ACM International Conference on Interaction Design and Children, ACM*, pp. 301-313.

Madigan, S., Villani, V., Azzopardi, C., Laut, D., Smith, T., Temple, J.R., Browne, D. and Dimitropoulos, G. (2018), "The prevalence of unwanted online sexual exposure and solicitation among youth: a meta-analysis", *Journal of Adolescent Health*, Vol. 63 No. 2, pp. 133-141.

Madrazo Azpiazu, I., Dragovic, N., Anuyah, O. and Pera, M.S. (2018), "Looking for the movie seven or sven from the movie frozen?: a multi-perspective strategy for recommending queries for children", *Proceedings of the 2018 Conference on Human Information Interaction and Retrieval, ACM*, pp. 92-101.

Madrazo Azpiazu, I., Dragovic, N., Pera, M.S. and Fails, J.A. (2017), "Online searching and learning: yum and other search tools for children and teachers", *Information Retrieval Journal*, Vol. 20 No. 5, pp. 1-22.

Minal Anand (2019), "Upcoming trends that will shape the future of education for children", available at: http://edtechreview.in/trends-insights/trends/3515-trends-shaping-the-future-of-education-for-children?utm_source=EdTechReview (accessed May 31, 2019).

National Center for Education Statistics (2019), "Children's access to and use of the internet", available at: https://nces.ed.gov/programs/coe/pdf/coe_cch.pdf (accessed May 10, 2019).

Next Generation Science Standards (2019), "DCI arrangements of the next generation science standards", pp. 22-46, available at: www.nextgenscience.org/sites/ default/files/NGSSDCICombined11.6.13.pdf (accessed August 20, 2019).

Nielsen, J. (2015), "Legibility, readability, and comprehension: making users read your words", available at: www.nngroup.com/articles/legibility-readability-comprehension/ (accessed August 21, 2019).

Nielsen, J. (2017), "Cloze test for reading comprehension", available at: www.nngroup.com/articles/cloze-test-reading-comprehension/ (accessed August 21, 2019).

Notess, G.R. (2006), "Teaching web search skills: techniques and strategies of top trainers", Information Today, Medford, NJ.

Novin, A. and Meyers, E. (2017), "Making sense of conflicting science information: exploring bias in the search engine result page", *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval, ACM*, pp. 175-184.

Patel, D. and Singh, P.K. (2016), "Kids safe search classification model", *International Conference on Communication and Electronics Systems, IEEE*, pp. 1-7.

Piaget, J. (1976), "Piaget's theory", in Inhelder, B., Chipman, H.H. and Zwingmann, C. (Eds), *Piaget and His School*, Springer, Berlin, Heidelberg, pp. 11-23.

Purcell, K., Rainie, L., Heaps, A., Buchanan, J., Friedrich, L., Jacklin, A., Chen, C. and Zickuhr, K. (2012), "How teens do research in the digital world", Pew Internet and American Life Project.

Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., Withey, R., Jamali, H. R., Dobrowolski, T. and Tenopir, C. (2008), "The Google generation: the information behaviour of the researcher of the future", *Aslib Proceedings, EGPL*, pp. 290-310.

Scholastic (2018), "Welcoming the Internet into your classroom", available at: www.scholastic.com/teachers/articles/teaching-content/welcoming-internet-your-classroom/ (accessed March 11, 2018).

Scott, T.J. and O'Sullivan, M.K. (2005), "Analyzing student search strategies: making a case for integrating information literacy skills into the curriculum", *Teacher Librarian*, Vol. 33 No. 1, pp. 21-25.

Silva, L., Mondal, M., Correa, D., Benevenuto, F. and Weber, I. (2016), "Analyzing the targets of hate in online social media", *Tenth International AAAI Conference on Web and Social Media*, *Cologne, May 17-20*.

Tan, C., Gabrilovich, E. and Pang, B. (2012), "To each his own: personalized content selection based on text comprehensibility", *Proceedings of the Fifth International Conference on Web Search and Data Mining, ACM*, pp. 233-242.

Teachers Pay Teachers (2019), "Teachers pay teachers", available at: www.teacherspayteachers.com/Browse/PreK-12-Subject-Area/ (accessed August 22, 2019).

Thurow, S. (2015), "To optimize search, optimize the searcher", *Online Searcher*, Vol. 39 No. 4, pp. 44-48.

Tori DeAngelis (2007), "Web pornography's effect on children", available at: www.apa.org/monitor/nov07/webporn (accessed May 20, 2019).

Torres, D., Hiemstra, D., Weber, I. and Serdyukov, P. (2014), "Query recommendation in the information domain of children", *Journal of the Association for Information Science and Technology*, Vol. 65 No. 7, pp. 1368-1384.

Vajjala, S. and Meurers, D. (2013), "On the applicability of readability models to web texts", *Proceedings of the 2nd Workshop on Predicting and Improving Text Readability for Target Reader Populations, ACL*, pp. 59-68.

Van Doorn, J., Odijk, D., Roijers, D.M. and de Rijke, M. (2016), "Balancing relevance criteria through multi-objective optimization", *Proceedings of the 39th International SIGIR Conference on Research and Development in Information Retrieval, ACM*, pp. 769-772.

Vasinda, S. and Pilgrim, J. (2019), "Are we preparing students for the web in the wild? An analysis of features of websites for children", *Journal of Literacy & Technology*, Vol. 20 No. 2, pp. 97-124.

Vidinli, I.B. and Ozcan, R. (2016), "New query suggestion framework and algorithms: a case study for an educational search engine", *Information Processing and Management*, Vol. 52 No. 5, pp. 733-752.

Walker, H.M. (2013), "Homework assignments and Internet sources", *Association for Computing Machinery Inroads*, Vol. 4 No. 4, pp. 16-17.

Willard, N.E. (2007), *Cyber-Safe Kids, Cyber-Savvy Teens: Helping Young People Learn to use the Internet Safely and Responsibly*, John Wiley and Sons, New York, NY.

Yamin, F.M., Ramayah, T. and Ishak, W.H.W. (2013), "Information searching: the impact of user knowledge on user search behavior", *Journal of Information & Knowledge Management*, Vol. 12 No. 3, p. 1350023.

Further reading

Telljohann, S.K., Symons, C.W., Pateman, B. and Seabert, D. (2009), *Health Education: Elementary and Middle School Applications*, McGraw-Hill Higher Education, Hoboken.

Corresponding author
Oghenemaro Anuyah can be contacted at: oghenemaroanuyah@u.boisestate.edu