Information Retrieval Research 2 (2026) 159-176 Submitted 1/26; Revised 06/26; Published 06/26

CASTlog: A Comprehensive Log of Children's Interactions with the Child Adaptive Search Tool

Christine Pinney CHRISTINEPINNEY@U.BOISESTATE.EDU
Boise State University
Boise, Idaho, USA

Casey Kennington CASEYKENNINGTON@U.BOISESTATE.EDU
Boise State University
Boise, Idaho, USA

Katherine Landau Wright KLWRIGHT@OREGON.EDU
University of Oregon
Eugene, Oregon, USA

Maria Soledad Pera M.S.PERA@TUDELFT.NL
TU Delft
Delft, Netherlands

Jerry Alan Fails JERRYFAILS@U.BOISESTATE.EDU
Boise State University
Boise, Idaho, USA

Editor: Johanne Trippas

Abstract

Children are increasingly interacting with search tools and exploring information online, whether it is for academic purposes or just for fun. Young searchers tend to favor search tools designed for adults and, as a result, face unique barriers to information access at all stages of the online search process. There have been many explorations into the obstacles children encounter during information retrieval (IR), but there is a lack of available data that can support further research from multiple perspectives and engage the larger IR community, resulting in stifled progress towards addressing the unique needs of young searchers. To address this lack of data, we introduce `CASTlog` a new dataset that captures the search queries of children ages 6-12 and the user-system interactions as searchers complete various kinds of online information discovery tasks. We also provide statistics of the data and an initial data analysis informed by empirical explorations based on the dataset. `CASTlog` serves as a foundational resource for researchers and practitioners focused on developing search tools that address the information needs of children.

Keywords: Dataset, Children Information Retrieval, Search, Query Analysis, Child-System Interaction

1 Introduction

As the digital information space continues to grow and technology becomes increasingly accessible, children are beginning to search online at younger and younger ages (Azpiazu et al., 2017; Anuyah et al., 2020; Bilal and Gwizdka, 2018). It is known that children largely

©2026 held by the author(s) License: CC-BY 4.0 WWW: https://irrj.org DOI: 10.54195/irrj.26178

prefer mainstream search engines (SE) like Google and Bing over those more specifically designed for them and their unique information needs (Azpiazu et al., 2017; Bilal and Huang, 2019; Danovitch, 2019; Gwizdka and Bilal, 2017), but mainstream SE have been designed and optimized for adults (Anuyah et al., 2020; Bilal, 2012; Bilal and Boehm, 2013; Bilal and Huang, 2019; Landoni et al., 2021), causing children to face many barriers when it comes to information access. Kuhlthau (1991) defines the *information seeking process* (ISP) as 6 sequential stages – initiation, selection, exploration, formulation, collection, and presentation – and outlines the thoughts, feelings, and actions of users that are characteristic of each stage. From initial query formulation to search result navigation and selection, children face distinctive obstacles within each of these stages (Pinney et al., 2024; Allen et al., 2021b; Bilal, 2012; Anuyah et al., 2020; Allen et al., 2022; Milton et al., 2020a; Allen et al., 2023; Collins-Thompson et al., 2011; Rajalakshmi et al., 2020; Azpiazu et al., 2017; Vanderschantz and Hinze, 2021; Gwizdka and Bilal, 2017). Ultimately, across different age ranges, children struggle to search for, retrieve, and utilize information from online resources (Bilal and Huang, 2019; Allen et al., 2022).

Recent research has explored the unique barriers to information access that children often face. In terms of query formulation, Pera et al. (2023) illustrate that children's unique strategies of query construction have significant implications on the resulting search results. Through analysis of logged queries gathered within the classroom context, the authors found that children tend towards natural language queries rather than keyword-based queries, where mainstream SE are better designed for the latter than the former (Pera et al., 2023). These findings reiterate earlier research carried out by Bilal and Gwizdka (2018), where children ages 11-13 were found to query using phrases and/or questions (70%) more often than keywords (30%). These authors also discovered that there are significant variations within this age range in terms of query formulation and reformulation strategies as well as levels of success in completing several types of search tasks, highlighting the need for "human intervention in terms of digital and search task literacy instruction, as well as system intervention to support children's query formulations and reformulations" (Bilal and Gwizdka, 2018, Section 9, p. 1038) during online search.

While these explorations (among many others not explicitly mentioned above (Landoni et al., 2022; Gwizdka and Bilal, 2017; Vanderschantz and Hinze, 2021; Landoni et al., 2019; Yarosh et al., 2018; Pera et al., 2023)) very clearly provide valuable insights regarding children's struggles with information retrieval (IR) and access, they are carried out via user studies with **limited users** (Landoni et al., 2019; Yarosh et al., 2018; Pinney et al., 2024; Gossen et al., 2014; Druin et al., 2010; Fails et al., 2019) or empirical explorations with **big but non-public data samples** (Landoni et al., 2020b, 2022; Bilal and Gwizdka, 2018; Gwizdka and Bilal, 2017; Vanderschantz and Hinze, 2021; Downs et al., 2020b; Anuyah et al., 2020), **non-contextualized data samples** without interaction (Pera et al., 2023), and **synthetic data samples** (Fröbe et al., 2025; Anuyah et al., 2020). A common denominator hindering the advancement of knowledge in terms of better supporting children's information access is the lack of available data that can enable research from multiple perspectives and comparison across solutions (Pera et al., 2025). Data and benchmarks from real children performing information discovery tasks using SE are scarce. This scarcity stifles progress in addressing challenges encountered by this often overlooked user group – namely, young searchers between the ages of 6 to 12 years old – resulting in a lack of engagement by

the wider IR research community and missed opportunities to enable the development of better IR technologies for children (Pera et al., 2025). Children of this age group are of particular interest as they are emergent searchers actively learning to read and develop their vocabularies and decoding skills as well as learning to search and develop their digital literacy (Kennington et al., 2021).

To address this gap, we introduce `CASTlog`[1], a new dataset constructed from logged data collected as a result of children conducting open-Internet searches, as well as targeted information discovery tasks related to specific lessons aligning with the primary school curriculum. More specifically, `CASTlog` is comprised of 13,309 unique queries from 7,931 unique search sessions carried out by children between the ages of 6 and 12 using the Child Adaptive Search Tool (CAST)[2], a search tool tailored to younger searchers. Co-designed with and for children, CAST features a uniquely simplistic design interface with a child-oriented, rule-based, phonetic spellchecker that provides query suggestions better-aligned with children's linguistic development, empowering them to become better searchers (Downs et al., 2020a). The dataset is rich with detailed information on queries, retrieved results, event data, and user interactions from live search sessions during which children progress through the various stages of online search.

Existing data sources related to children's search behavior are limited in both scope and accessibility. The dataset introduced in (Madrazo Azpiazu et al., 2018), while publicly available, includes only queries without interaction data. The datalog collected as a result of the study presented (Aliannejadi et al., 2021) contains queries and click data, yet remains private and inaccessible. The most recent resource (Fröbe et al., 2025) offers queries accompanied by labeled ground truth (determined by adults); however, it lacks authentic interaction data within an information-seeking process and does not reflect genuine child-generated queries, as these were formulated by adults on child-related topics. In contrast, `CASTlog` is distinctive in that it captures both authentic queries produced by children and their corresponding interaction behaviors. This combination of real queries and real user interactions sets it apart from existing datasets. Furthermore, search logs from commercial SE involving children have not been and are unlikely to ever be made publicly available due to strict policies and regulatory constraints. This further underscores the uniqueness of `CASTlog` particularly once it is released with appropriate ethical safeguards and necessary approvals in place. Overall, `CASTlog` constitutes a novel and valuable resource for research in Children's Information Retrieval, Child–Computer Interaction, and Digital Literacy. Its interactive nature, combined with its accessibility, fine-grained data structure, and educational context, provides a rare opportunity to examine the nuanced search behaviors of a protected and underrepresented user population—an opportunity that neither existing corporate nor academic datasets can adequately support.

In the rest of this paper, we discuss CAST and its role in the creation of `CASTlog`, outline the data collection process, describe how this ensemble dataset was produced, and highlight initial statistics on the collected data within `CASTlog`. We report on the outcome informed by empirical exploration based on this dataset, outline existing limitations, and paint a picture of the interplay between SE and children ages 6 to 12 across the ISP stages. To the best of our knowledge, `CASTlog` is the first of its kind and will be foundational for

1. https://huggingface.co/datasets/BSU-CAST/CASTlog
2. https://cast.boisestate.edu/cast-simple/

researchers and practitioners focused on research and developing algorithms and interfaces that respond to and address the needs of children when seeking information.

2 Child Adaptive Search Tool (CAST)

CAST is a child-centric search tool and user interface aimed at addressing the search and information needs of children ages 6-12 years old (i.e., emergent searchers). This search tool is part of a U.S. National Science Foundation-funded research project spearheaded by an interdisciplinary team and intended to empower young searchers by researching, designing, and developing search tools that improve children's information literacy and searching capabilities (see (Downs et al., 2019b; Allen et al., 2021b) for an overview).

The CAST search tool is powered by Microsoft Bing's SafeSearch. In addition to filtering unsafe and inappropriate search results, CAST presents a simplified and de-cluttered interface without advertisements to eliminate unnecessary distractions for younger searchers. CAST also employs a state-of-the-art spellchecker specifically designed to capture the unique misspellings of children and provide more effective support during query formulation when compared with mainstream SE designed with adults in mind (Downs et al., 2019a, 2020a,b). Although there have been other explorations with CAST in terms of re-ranking mechanisms and SERP design features, it is worth noting that these advances are not currently part of the tool (Allen et al., 2023; Pinney et al., 2024).

The CAST interface also includes a *lesson module*, where searchers are guided by educators via pre-recorded videos and prompts to complete curriculum-based information discovery search tasks. Lessons cover age and grade-appropriate topics (e.g., seasons and their characteristics, identifying elements on a map, and the structure and function of living things), and students' provided answers are logged.

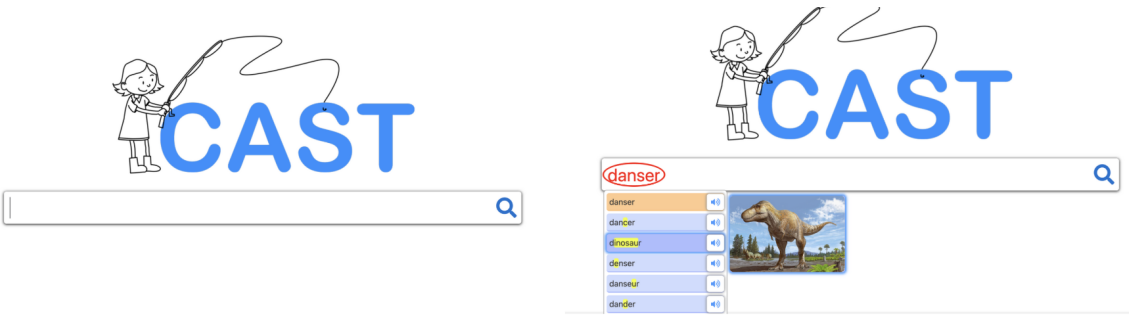

Figure 1: CAST main interface (left) and example spellchecker suggestion (right).

3 Data Collection

Data was collected using CAST over a period of 4 years, as a result of multiple explorations conducted to inform CAST design and deployment. These explorations focused on aspects such as query formulation/reformulation, query spellchecking, readability and ranking of retrieved results, SERP design and navigation, and evaluation of children's search tools. The specific protocols used varied across studies and are documented the associated prior

publications, which we enumerate as follows: Allen et al. (2023, 2021b, 2022); Milton et al. (2019); Pinney et al. (2024); Allen et al. (2021c); Garrett et al. (2021); Kennington et al. (2021); Allen et al. (2021a); Downs et al. (2020a,b); Landoni et al. (2020a); Milton et al. (2020b); Downs et al. (2019a); Fails et al. (2019); Murgia et al. (2019). Collected data was captured and logged from user-system interactions during unique search sessions using CAST.

While the ISP model has traditionally been developed and validated with adults and older students, our protocol design adapts its core stages (e.g., initiation, exploration, formulation, and collection) to align with the cognitive, developmental, and interactional characteristics of children aged 6–12. Different CAST modules were activated depending on the specific study. In controlled studies—encompassing both researcher-designed tasks and CAST lesson-based activities (see Section 4)—children engaged in a range of information-seeking tasks spanning entertainment and educational contexts, including targeted, exploratory, and open-ended search tasks, all using CAST. These studies were conducted in one of three environments: a lab setting (Boise State University), charter middle school classrooms (Boise, ID), or via Zoom (with U.S.-based participants). CAST also supported open Internet use outside formal study conditions, enabling free-search sessions not tied to specific experimental tasks; these interactions, limited to users aged 6–12, are also included in `CASTlog`. Further details regarding search tasks, participant demographics, and recruitment processes can be found in the previously cited research works. All collected data is in English.

4 Dataset

`CASTlog` is organized as tuples of search sessions and queries, further broken down into the individual **events** associated with each query, where events capture the interplay between users and the search system, and help illustrate the ISP process for individual queries within a search session. The dataset is organized in this manner to capture each individual query within a search session along with the many events (e.g., user-system interactions) that occur during ISP as a query is formulated and explored (i.e., keystrokes, query suggestions, retrieved results, and clicks). Figure 2 outlines the structure of the dataset and provides descriptions of each field. Each event type is associated with unique data contained within the *data* column. The *data* column is a key-value pair dictionary. The event types and contained data are:

- **query**: Each query event contains information about the query as well as the retrieved results.
  - *query*: the entered query string submitted to the search tool.
  - *query_len*: the length of the entered query in number of words.
  - *result_titles*: the titles of the top 10 results, presented as a list of strings.
  - *result_snippets*: the snippet of text that each of the top 10 results provide as a preview to result content, presented as a list of strings.
  - *result_urls*: the URLs of the top 10 results, presented as a list of strings.
  - *result_positions*: the position of each of the top 10 results within the presented ranked list of retrieved results, presented as a list of strings.

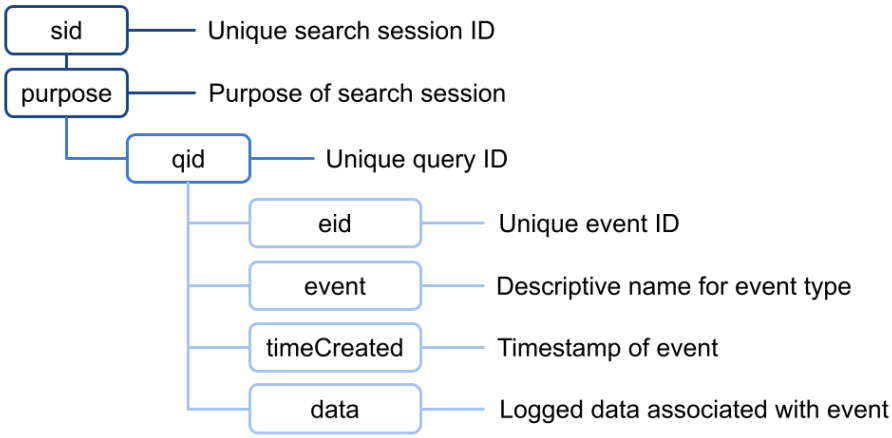

Figure 2: Overview of logged events, associated (meta)data, and their interconnection in `CASTlog`.

  - *result_SA_scores*: the Spache-Allen readability score of each of the top 10 result snippets, presented as a list of integers. The Spache-Allen readability formula, adapted from the original Spache readability formula presented in (Spache, 1953), assesses text difficulty using a calculation that analyzes sentence length and percentage of unfamiliar words (Allen et al., 2022). The formula produces a rounded integer score corresponding to the estimated reading grade-level of the text (i.e., a score of '2' implies a 2nd-grade reading level) and is particularly effective for children's reading materials (Allen et al., 2022). The formula is applied to each result's search snippet.

  Data about the retrieved results maintain order, meaning that the first list item in the *result_titles* list corresponds to the first list item of *result_snippets*, *result_urls*, *result_positions*, and *result_SA_scores* lists, and so on.

- **keystroke**: Each keystroke event contains data about keyboard keys pressed during query formulation.
  - *key*: the growing query string as each individual keyboard key is pressed during query formulation, presented as a string.
  - *pos*: the cursor position after a keyboard key is pressed, presented as an integer.
- **input-parsed**: Any input text in the search bar is periodically parsed for misspelled words.
  - *misspelled_words*: identified misspelled words within the growing query string, presented as a list of strings.
- **popup**: As misspelled words are identified, corresponding suggestion popup events provide spelling suggestions.

  - *misspell*: the misspelled word targeted for spelling suggestions, presented as a string.
  - *suggestions*: up to five suggested words based on the misspelled word, presented as a list of strings.

- **hov-suggestion** and **hov-off-suggestion**: Any hover mouse events are recorded to track mouse positions when spelling suggestions are presented during query formulation.
  - *hovered*: the position of the word within the spelling suggestion list that is hovered on or off, presented as a string.

- **hov-result** and **hov-off-result**: Any hover mouse events are recorded to track mouse positions during SERP navigation and exploration.
  - *hovered*: the position of the result within the result list that is hovered on or off, presented as a string.

- **clk-suggestion**: Any click mouse events are recorded to track mouse activity when spelling suggestions are presented.
  - *clicked*: the clicked spelling suggestion, presented as a string.
  - *position*: the position of the clicked spelling suggestion within the suggestion list, presented as a string.

- **clk-result**: Any click mouse events are recorded to track mouse activity during SERP navigation and exploration.
  - *position*: the position of the clicked result within the list of results, presented as a string.

- **lesson-start** and **lesson-next**: When the CAST lesson module is utilized, specific data is logged about the lesson and provided answers to prompts. *lesson-start* indicates when a lesson is started, and *lesson-next* indicates when the 'Next' button is pressed to move to a new question within a lesson.
  - *lsid*: a unique identifier for each lesson, similar to *sid* (the search session ID), presented as a string.
  - *answers*: a nested dictionary containing information about provided questions, question types, and entered responses. Each provided question within a lesson is stored as an individual key with an associated dictionary containing more specific information about the question as its value, including:
    * *id*: the question ID (e.g., *q1*, *q2*, etc.).
    * *type*: the question type (openanswer, mcq (multiple choice), or matching). Open answer questions specify a subtype: short or long. Multiple choice and matching questions provide an additional field listing the available choices.
    * *prompt*: the provided question/prompt.

    * *entered*: the entered response.

  Depending on the type of question and the specific lesson or topic, additional fields may be included.
- *topic*: the lesson topic, presented as a string. There two topics for 1st through 4th grade material, and one topic for 5th grade material.
- *grade*: the grade level of the lesson and topic, presented as an integer.

The dataset is annotated with pertinent metadata regarding exploration purposes to support and enable additional analyses and explorations from different perspectives; this metadata is contained within the 'purpose' column of the dataset. This metadata essentially allows for different 'slices' of the dataset to be easily accessed, providing ease of use within the IR community. Search sessions labeled as 'free search' contain all open-Internet data collected that is not associated with a particular lesson-centered search task and/or targeted research experiment. Search sessions labeled as 'research-driven search task' contain data associated with research-driven data collection, where queries and search result navigation are guided and/or directed by researchers for a specific purpose, but not associated with specific lessons. Search sessions labeled as 'research-driven lesson-based search task' contain all data associated with the outlined CAST lessons module and research-driven data collection. Search sessions labeled as 'self-directed lesson-based search task' contain all data associated with CAST lessons module that is not research-driven.

To ensure the dataset was constructed with consistent fields and data types, the logged data required some post-processing, including reshaping/reformatting and removing malformed data (i.e., missing data fields resulting from logging errors). Given the various ethical considerations across explorations, `CASTlog` does not include specific demographic information (age/grade, socioeconomic status, etc.) of users, and we recognize this as a limitation of the dataset.

Table 1 provides an initial overview of the data within `CASTlog` collected from online interactions with CAST during both open-ended use of the tool as well as controlled explorations, including the number of search sessions, unique queries, unique retrieved results, clicked results, and query suggestions. It is important to note that interaction events such as result clicks should not be treated as ground-truth indicators of the identification of relevant resources, successful information seeking, or learning outcomes. Instead, they merely reflect users' selection and navigation behaviors, and equating them with success may lead to misleading inferences from logged data. In the following section, we provide an initial analysis of the data within `CASTlog`.

5 Data Analysis and Future Explorations

Using the stages of ISP as a guide for analysis, we explore and illustrate some key observations emerging from `CASTlog`, discuss the implications for children's information access, and highlight opportunities for future explorations and analyses.

5.1 Initiation and Selection: Query Formulation and Suggestions

At the onset of ISP, searchers recognize the need for information (i.e., the initiation stage), leading them to select or identify a particular search task or topic for exploration (i.e., the

Overall	
Total # of Unique Search Sessions	7,931
Total # of Unique Queries	13,309
Total # of Unique URLs Retrieved	80,729
Total # of Query Suggestions Provided	60,547
Total # of Individual Events	1,632,480
Per Session	
Avg # of Click Events per Session	4.33
Avg # of Result Clicks per Session	3.84
Avg # of Queries per Session	3.55
Avg Session Length (in minutes)	8.87
Queries	
Total # of Non-Lesson Queries	17,308
Total # of Lesson Queries	5,644
Avg Query Length (in words)	3.86
Avg # of Stopwords in Query	1.46
Results	
Total # of Results Clicked	16,183
Avg Position of Results Clicked	2.85
Avg Spache-Allen Readability Score of Result Snippets	7.03
Spellchecker Suggestions	
Total # of Suggestion Hover Events	39,968
Total # of Suggestion Hover-Off Events	34,890
Total # of Suggestion Click Events	4,731
% of Total Suggestions Clicked	7.81
% of Total Suggestions Hovered	66.01
% of Total Suggestions Hovered-Off When Hovered	87.29

Table 1: Summarization of dataset statistics and user interactions.

selection stage) (Kuhlthau, 1991). Within these stages, searchers engage with the search system (i.e., CAST) and begin by formulating an initial query. `CASTlog` contains a total of 22,952 queries (13,309 unique queries), where 5,644 of those queries are related to the CAST lesson modules (i.e., those queries with the phrase 'lesson-based' included within the *purpose* column of the dataset). The average length of queries is 3.86 words, but Figure 3 displays some variation across this measure, revealing a tendency towards longer queries. These findings align with earlier research that reports longer queries resulting from children's query formulation (Duarte Torres et al., 2010; Madrazo Azpiazu et al., 2018). Queries contained an average of 1.46 *stopwords*, or commonly used words that carry little meaning and are often removed during SE optimization (e.g., *the*, *and*, *a/an*, etc.).[3] These findings imply that, on average, 1.5 words of a 4-word query are not keywords, resulting in queries following the pattern of natural language. This reiterates earlier findings that demonstrate a preference

3. Stopwords were identified using the `nltk.corpus.stopwords` module (Bird et al., 2009).

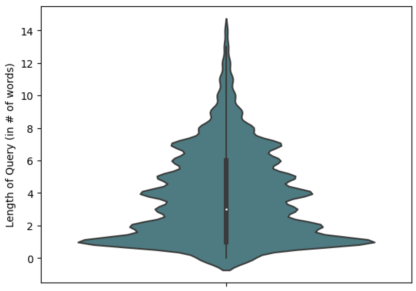

Figure 3: Distributions of query lengths across sessions.

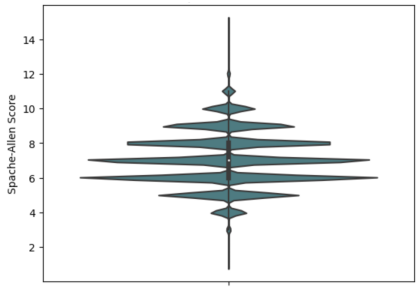

Figure 4: Distribution of Spache-Allen readability scores of retrieved results.

by younger searchers for natural language queries over keyword-based queries (Pera et al., 2023; Bilal and Gwizdka, 2018; Vanderschantz and Hinze, 2021).

`CASTlog` also contains information regarding user engagement with CAST's custom, state-of-the-art phonetic spellchecker tool that offers spelling suggestions while searchers formulate their queries (Downs et al., 2019a, 2020a,b). A total of 60,547 query spellchecking suggestions were provided to searchers during query formulation; of those suggestions provided, about 66% of the suggestions were hovered by users, but only 7.81% of the hovered suggestions were actually clicked on by users. Similar to findings presented by Druin et al. (2009) in their analysis of children's query formulation, these results reveal a lack of engagement with spelling suggestions during query formulation.

5.2 Exploration: Navigation of Search Results

After entering an initial query, searchers encounter a SERP and explore the information and resources provided (i.e., the exploration stage). `CASTlog` contains data about the results retrieved as well as the interactions carried out on the SERP by the searchers. The collected queries resulted in a total of 203,550 retrieved resources (80,729 unique URLs). To provide information on the readability of results, the dataset contains readability scores for each result that was retrieved for a given query. This calculation is applied to each result's search snippet, i.e., the text visible on the SERP extracted from the resource content (Bilal and Huang, 2019), using the Spache-Allen readability formula (Allen et al., 2022). This formula provides a rounded whole number estimating the reading grade level of the target text. On average, the Spache-Allen score for retrieved results is approximately 7, i.e., a 7th grade reading level; see Figure 4. This reiterates previous findings that report inaccessible readability of online resources for younger searchers (Bilal and Huang, 2019; Bilal, 2012; Collins-Thompson et al., 2011; Pinney et al., 2024). This has implications for children's ISP in particular, as unreadable or incomprehensible resources can be considerable barriers for children's information access. If a young searcher cannot read an online resource, they cannot extrapolate the information they are looking for and, thus, cannot complete ISP successfully. Mainstream SE are designed and optimized for adults, so the readability of results is not often taken into account by relevance algorithms employed by SE; this results

in SERPs with top-ranked results containing text that is too complex for younger searchers to comprehend (Bilal and Huang, 2019; Bilal, 2012; Collins-Thompson et al., 2011; Anuyah et al., 2020). Being a search tool tailored to the needs of younger searchers, CAST does employ a relevance algorithm that takes readability into account when ranking results on a SERP. While the returned results do include resources containing text at a more reasonable reading grade level for searchers (as shown in Figure 4), the youngest searchers of the 6-12 year old user group have fewer resources available, highlighting the larger issue of a lack of online resources that are readable for the youngest of online searchers.

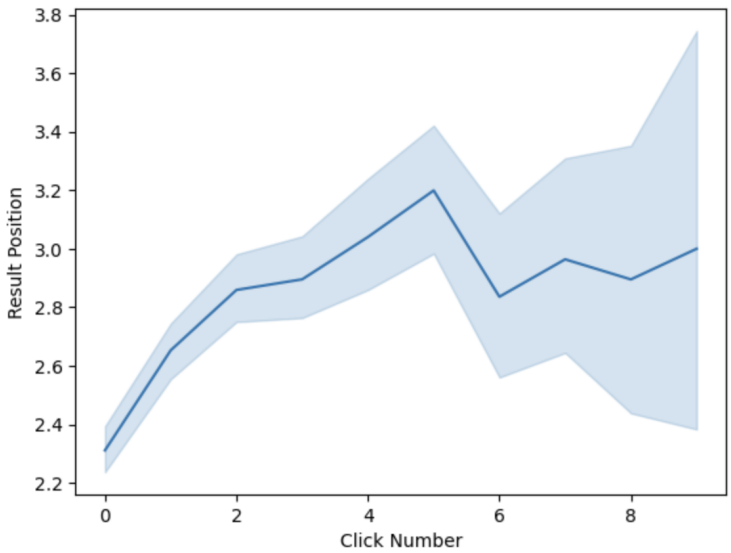

Figure 5: Average linearity of clicked results in sessions with multiple clicked results. Shading indicates the 95% confidence interval of the mean represented by the line.

In terms of user interactions on a SERP, `CASTlog` contains information about the position of clicked results within the retrieved list of results presented on a SERP. Data within `CASTlog` illustrate that a large majority of searchers displayed linear behavior (Landoni et al., 2022), i.e., often sequentially exploring ranked resources from the top result, in terms of clicked results (see Figure 5) and gravitated towards the top-ranked results on a SERP (see Figure 7). Linearity here refers to the relationship between the user's click order of results and the results' positions within the ranked list on the SERP. This reflects earlier findings that report considerable misconceptions by younger searchers about the significance and meaning behind ranking on a SERP, and a tendency towards exploring resources at the top of the presented list of results (Vanderschantz and Hinze, 2021; Bilal, 2012). As aforementioned, mainstream SE often have top-ranked results that are not readable for younger searchers, so their tendency towards exploring SERPs in a top-down, linear fashion increasingly complicates their experience with ISP. We caution researchers against explicitly equating clicked results with the success of a searcher in finding relevant answers, as this measure of success is not recorded in the dataset.

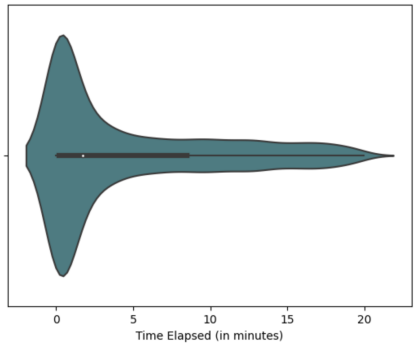

Figure 6: Distribution of time-elapsed during sessions.

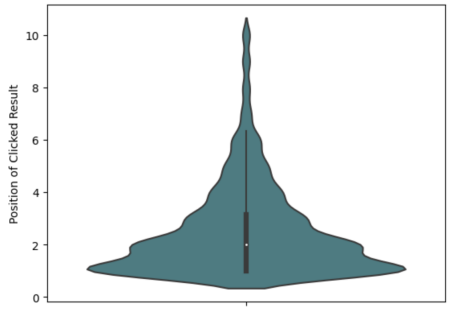

Figure 7: Distribution of positions of clicked results across sessions.

During the exploration stage of ISP, searchers investigate the presented information to extend personal understanding of the search topic, often encountering feelings of doubt and uncertainty, particularly when the information encountered does not align with previously held constructs and/or the searcher isn't able to effectively articulate what information is needed (Kuhlthau, 1991). With the unique barriers that young searchers face throughout ISP, this stage understandably becomes more difficult as they struggle to express search needs precisely during query formulation and successfully navigate the presented information. Some searchers are inclined to abandon the search altogether at this stage due to feelings of discouragement and frustration (Kuhlthau, 1991). This may result in shorter, more brief search sessions, as reflected by the decreased length of individual search sessions recorded within `CASTlog` (see Figure 6).

5.3 Dataset Usage and Ethical Considerations

There is a significant lack of publicly avialable query logs capturing the search behaviors of children in a holistic and ecologically valid manner. What currently exists, offers extremely limited support for empirical exploration and evaluation. In this context, `CASTlog` constitutes a rich, diverse resource capturing nuanced behavior along the key stages of the ISP, making it unique. The breadth and granularity of the logged interactions enable a wide range of research directions. `CASTlog` can be used to inform the understanding of children's information retrieval and the design and evaluation of IR technologies at the core of search (and other information access) systems. With that, we anticipate `CASTlog` supporting advancements related to several research directions that require attention when children are the main stakeholders, including strategies for query formulation and query suggestions, analysis of SERP results that draw attention from children and subsequent SERP navigation behaviors, patterns of interaction with information, and the design and assessment of (re-)ranking algorithms that support children's information discovery.

Moreover, the dataset provides a strong foundation for future extensions, such as the incorporation of ground truth aligned with child-centred notions of relevance, including

readability, educational alignment, and harm minimisation (Chakrabarti et al., 2026). This further strengthens its potential as a benchmark resource tailored to younger users.

While IR researchers may be the most apparent community that would benefit from this resource, there are additional communities that might find the data within `CASTlog` particularly valuable. For instance, the data related to the spellchecker and spellchecking suggestions during query formulation could offer rich insight for educators regarding spelling patterns and behaviors within this age group. The queries and results explored could also provide educators with insight into terminology and language comprehension. Additionally, data regarding mouse click behavior and hovering patterns could be incredibly useful to user interface and child-computer interaction researchers who aim to design and develop interfaces for younger audiences, providing insight into interface limitations and needs.

In light of the vulnerable user group at the center of this work, all ethical and legal considerations were carefully prioritized throughout the development of the dataset and in preparation for dissemination. We mindfully considered compliance with relevant protection laws, data privacy regulations (such as COPPA), and institutional review protocols. The dataset's purpose is strictly to support research advancements, particularly in IR, serving a population that has long been underrepresented and underserved (Pera et al., 2025, 2024). Moving forward, researchers and practitioners must remain aware of the ethical obligations associated with using and manipulating data from child users. They must proactively guard against misuse, assess potential risks, and ensure that any data used does not compromise the rights of this vulnerable population. Responsible use requires a balance between innovation and harm prevention, including scrutiny of research intentions, safeguards, and accountability mechanisms. Ultimately, `CASTlog` should be employed in ways that advance meaningful, positive outcomes for young users in a safe, respectful, and equitable manner.

6 Concluding Remarks

As children have increasingly been engaging with online search tools at younger and younger ages, it has become apparent that mainstream SE are designed with adult searchers in mind and fail to address the unique barriers that children encounter during ISP. The information needs of this particular user group are often overlooked by the IR community, and a lack of data contributes to this issue (Pera et al., 2025). `CASTlog` the first of its kind, is the result of a four-year multidisciplinary project. It serves as a foundation to address this gap and allow a wider array of research and perspectives to be explored to enable the design and development of better search tools for young searchers.

Acknowledgments and Disclosure of Funding

We thank the children who participated in the different studies that ultimately resulted in this data resource. We also appreciate the feedback from reviewers and editors, which helped improve the manuscript. This work was partially funded by NSF Award #1763649.

References

Mohammad Aliannejadi, Monica Landoni, Theo Huibers, Emiliana Murgia, and Maria Soledad Pera. Children's perspective on how emojis help them to recognise relevant results: Do actions speak louder than words? In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, pages 301–305, 2021.

Garrett Allen, Brody Downs, Aprajita Shukla, Casey Kennington, Jerry Alan Fails, Katherine Landau Wright, and Maria Soledad Pera. BiGBERT: Classifying educational web resources for kindergarten-12th grades. In *Lecture Notes in Computer Science*, Lecture notes in computer science, pages 176–184. Springer International Publishing, Cham, 2021a.

Garrett Allen, Benjamin L Peterson, Dhanush Kumar Ratakonda, Mostofa Najmus Sakib, Jerry Alan Fails, Casey Kennington, Katherine Landau Wright, and Maria Soledad Pera. Engage!: Co-designing search engine result pages to foster interactions. In *Proceedings of the 20th Annual ACM Interaction Design and Children Conference*, IDC '21, pages 583–587, 2021b. Association for Computing Machinery.

Garrett Allen, Jie Yang, Maria Soledad Pera, and Ujwal Gadiraju. Using conversational artificial intelligence to support children's search in the classroom. *arXiv [cs.HC]*, 2021c.

Garrett Allen, Ashlee Milton, Katherine Landau Wright, Jerry Alan Fails, Casey Kennington, and Maria Soledad Pera. Supercalifragilisticexpialidocious: Why using the "right" readability formula in children's web search matters. In *European Conference on Information Retrieval*, pages 3–18. Springer, 2022.

Garrett Allen, Katherine Landau Wright, Jerry Alan Fails, Casey Kennington, and Maria Soledad Pera. Multi-perspective learning to rank to support children's information seeking in the classroom. In *2023 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 311–317. IEEE, 2023.

Oghenemaro Anuyah, Ashlee Milton, Michael Green, and Maria Soledad Pera. An empirical analysis of search engines' response to web search queries associated with the classroom setting. *Aslib Journal of Information Management*, 72(1):88–111, 2020.

Ion Madrazo Azpiazu, Nevena Dragovic, Maria Soledad Pera, and Jerry Alan Fails. Online searching and learning: YUM and other search tools for children and teachers. *Information Retrieval Journal*, 20(5):524–545, 2017.

Dania Bilal. Ranking, relevance judgment, and precision of information retrieval on children's queries: Evaluation of Google, Yahoo!, Bing, Yahoo! kids, and Ask kids. *Journal of the American Society for Information Science and Technology*, 63(9):1879–1896, 2012.

Dania Bilal and A Reid Boehm. Towards new methodologies for assessing relevance of information retrieval from web search engines on children's queries. *Qualitative and Quantitative Methods in Libraries*, 2(1), 93–100, 2013.

Dania Bilal and Jacek Gwizdka. Children's query types and reformulations in Google search. *Inf. Process. Manag.*, 54(6):1022–1041, 2018.

Dania Bilal and Li-Min Huang. Readability and word complexity of SERPs snippets and web pages on children's search queries: Google vs bing. *Aslib Journal of Information Management*, 71(1), 2019.

Steven Bird, Edward Loper, and Ewan Klein. *Natural Language Processing with Python*. O'Reilly Media Inc., 2009. Stopwords Corpus compiled by Porter et al.

Hrishita Chakrabarti, Diletta Micol Tobia, Garrett Allen, Monica Landoni, and Maria Soledad Pera. It is relevant, but is it useful? A reflection on human-centred evaluation in children information retrieval. In *Proceedings of the 34th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '26, page 392–397, 2026. Association for Computing Machinery. ISBN 9798400723117. doi: 10.1145/3774935.3806167.

Kevyn Collins-Thompson, Paul N Bennett, Ryen W White, Sebastian De La Chica, and David Sontag. Personalizing web search results by reading level. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 403–412, 2011.

Judith H Danovitch. Growing up with Google: How children's understanding and use of internet-based devices relates to cognitive development. *Hum. Behav. Emerg. Technol.*, 1 (2):81–90, 2019.

Brody Downs, Tyler French, Katherine Landau Wright, Maria Soledad Pera, Casey Kennington, and Jerry Alan Fails. Searching for spellcheckers: What kids want, what kids need. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, 2019a. ACM.

Brody Downs, Tyler French, Katherine Landau Wright, Maria Soledad Pera, Casey Kennington, Jerry Alan Fails, and Jerry Alan. Children and search tools: Evaluation remains unclear. In *Proceedings of KidRec '19*, 2019b.

Brody Downs, Oghenemaro Anuyah, Aprajita Shukla, Jerry Alan Fails, Sole Pera, Katherine Wright, and Casey Kennington. KidSpell: A child-oriented, rule-based, phonetic spellchecker. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 6937–6946, 2020a.

Brody Downs, Aprajita Shukla, Mikey Krentz, Maria Soledad Pera, Katherine Landau Wright, Casey Kennington, and Jerry Fails. Guiding the selection of child spellchecker suggestions using audio and visual cues. In *Proceedings of the Interaction Design and Children Conference*, 2020b. ACM.

Allison Druin, Elizabeth Foss, Leshell Hatley, Evan Golub, Mona Leigh Guha, Jerry Fails, and Hilary Hutchinson. How children search the internet with keyword interfaces. In *Proceedings of the 8th International Conference on Interaction Design and Children*, 2009. ACM.

Allison Druin, Elizabeth Foss, Hilary Hutchinson, Evan Golub, and Leshell Hatley. Children's roles using keyword search interfaces at home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010. ACM.

Sergio Duarte Torres, Djoerd Hiemstra, and Pavel Serdyukov. Query log analysis in the context of information retrieval for children. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, 2010. ACM.

Jerry Alan Fails, Maria Soledad Pera, Oghenemaro Anuyah, Casey Kennington, Katherine Landau Wright, and William Bigirimana. 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*, 2019. ACM.

Maik Fröbe, Sophie Charlotte Bartholly, and Matthias Hagen. How child-friendly is web search? An evaluation of relevance vs. harm. In *Lecture Notes in Computer Science*, pages 214–222. Springer Nature Switzerland, 2025.

Allen Garrett, Katherine Landau Wright, Jerry Alan Fails, Kennington Casey, and Maria Soledad Pera. CASTing a net: Supporting teachers with search technology. *arXiv [cs.CY]*, 2021.

Tatiana Gossen, Juliane Höbel, and Andreas Nürnberger. Usability and perception of young users and adults on targeted web search engines. In *Proceedings of the 5th Information Interaction in Context Symposium*, 2014. ACM.

Jacek Gwizdka and Dania Bilal. Analysis of children's queries and click behavior on ranked results and their thought processes in Google. In *Proceedings of the conference on conference human information interaction and retrieval*, pages 377–380. 2017.

Casey Kennington, Jerry Alan Fails, Katherine Landau Wright, and Maria Soledad Pera. Spellchecking for children in web search: A natural language interface case-study. In *Proceedings of the First Workshop on Bridging Human–Computer Interaction and Natural Language Processing*, pages 8–13, 2021.

Carol C Kuhlthau. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 1991.

Monica Landoni, Davide Matteri, Emiliana Murgia, and Maria Soledad Pera. Sonny, cerca! evaluating the impact of using a vocal assistant to search at school. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction, Processings of the 10th International Conference of the CLEF Association, 2019*, pages 101–113. 2019.

Monica Landoni, Jerry Alan Fails, Theo Huibers, Natalia Kucirkova, Emiliana Murgia, and Maria Soledad Pera. 4th KidRec workshop - "what does good look like?": From design, research, and practice to policy. In *Proceedings of the ACM Interaction Design and Children Conference: Extended Abstracts*, 2020a. ACM.

Monica Landoni, Maria Soledad Pera, Emiliana Murgia, and Theo Huibers. Inside out: Exploring the emotional side of search engines in the classroom. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 2020b. ACM.

Monica Landoni, Mohammad Aliannejadi, Theo Huibers, Emiliana Murgia, and Maria Soledad Pera. Right way, right time: Towards a better comprehension of young

students' needs when looking for relevant search results. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*, UMAP '21, pages 256–261, 2021. Association for Computing Machinery.

Monica Landoni, Mohammad Aliannejadi, Theo Huibers, Emiliana Murgia, and Maria Soledad Pera. 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, pages 310–314, 2022. Association for Computing Machinery.

Ion Madrazo Azpiazu, Nevena Dragovic, Oghenemaro Anuyah, and Maria Soledad Pera. 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*, pages 92–101, 2018.

Ashlee Milton, Michael Green, Adam Keener, Joshua Ames, Michael D Ekstrand, and Maria Soledad Pera. StoryTime: eliciting preferences from children for book recommendations. In *Proceedings of the 13th ACM Conference on Recommender Systems*, RecSys '19, pages 544–545, 2019. Association for Computing Machinery.

Ashlee Milton, Oghenemaro Anuya, Lawrence Spear, Katherine Landau Wright, and Maria Soledad Pera. A ranking strategy to promote resources supporting the classroom environment. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 121–128. IEEE, 2020a.

Ashlee Milton, Levesson Batista, Garrett Allen, Siqi Gao, Yiu-Kai D Ng, and Maria Soledad Pera. Don't judge a book by its cover: Exploring book traits children favor. In *Fourteenth ACM Conference on Recommender Systems*, 2020b. ACM.

Emiliana Murgia, M Landoni, T Huibers, J A Fails, and M S Pera. The seven layers of complexity of recommender systems for children in educational contexts. *Proceedings of the Workshop on Recommendation in Complex Scenarios Co-Located with 13th ACM Conference on Recommender Systems*, Vol. 2449. CEUR-WS. pages 5–9, 2019. `http://ceur-ws.org/Vol-2449/paper1.pdf`

Maria Soledad Pera, Emiliana Murgia, Monica Landoni, Theo Huibers, and Mohammad Aliannejadi. Where a little change makes a big difference: A preliminary exploration of children's queries. In *Lecture Notes in Computer Science*, Lecture notes in computer science, pages 522–533. Springer Nature Switzerland, Cham, 2023.

Maria Soledad Pera, Emiliana Murgia, Monica Landoni, Federica Cena, Noemi Mauro, and Theo Huibers. Report on the 1st workshop on information retrieval for understudied users (IR4U2 2024) at ECIR 2024. In *ACM SIGIR Forum*, volume 58, pages 1–5. ACM, 2024.

Maria Soledad Pera, Theo Huibers, Emiliana Murgia, and Monica Landoni. Some things never change: Overcoming persistent challenges in children IR. In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3107–3109, 2025. ACM.

Christine Pinney, Benjamin J Bettencourt, J A Fails, C Kennington, Katherine Landau Wright, and M S Pera. How readability cues affect children's navigation of search engine result pages. *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*, 62–69 2024. doi: 10.1145/3628516.3655818

R Rajalakshmi, Hans Tiwari, Jay Patel, R Rameshkannan, and R Karthik. Bidirectional gru-based attention model for kid-specific url classification. In *Deep Learning Techniques and Optimization Strategies in Big Data Analytics*, pages 78–90. IGI Global, 2020.

George Spache. A new readability formula for primary-grade reading materials. *The Elementary School Journal*, 53(7):410–413, 1953.

Nicholas Vanderschantz and Annika Hinze. Children's query formulation and search result exploration. *International Journal on Digital Libraries*, 22(4):385–410, 2021.

Svetlana Yarosh, Stryker Thompson, Kathleen Watson, Alice Chase, Ashwin Senthilkumar, Ye Yuan, and A. J. Bernheim Brush. Children asking questions: speech interface reformulations and personification preferences. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, IDC '18, page 300–312, 2018. Association for Computing Machinery. ISBN 9781450351522. doi: 10.1145/3202185.3202207.