IR for Children 2000-2020: Where Are We Now?

Theo Huibers*
t.w.c.huibers@utwente.nl
University of Twente
The Netherlands

Emiliana Murgia
emiliana.murgia@unimib.it
Università degli Studi di Milano-Bicocca
Italy

Monica Landoni
monica.landoni@usi.ch
Università della Svizzera Italiana
Switzerland

Maria Soledad Pera
solepera@boisestate.edu
PIReT–Boise State University
USA

ABSTRACT

Over 20 years ago, Information Retrieval (IR) researchers began their quest for sound IR systems for children. The path was not straightforward. Challenges posed by interface design, relevance determination, diverse contexts, ethics, and many more, were taken up and explored from different perspectives. Large projects such as Puppy-IR and the International Children's Digital Library gave this field a certain boost; still, there is neither a sound solution for children in the search area in 2021 nor a roadmap to get there. What is the reason for this? Does the field cry out for specific IR solutions developed on a small scale for very small sub-fields and specific target groups? Are there some significant unforeseen barriers that hinder researchers? What about obstacles natural to areas of study such as this one that require a multidisciplinary approach or involve protected populations? With this workshop, we want to bring together as many key experts as possible from research and industry who focus on IR for children to understand why, unlike other IR areas, this one has not flourished and look for the biggest challenges for the next 10 years. We are not only thinking of traditional researchers and designers, but also of those who develop and use IR systems for fields, such as in music, film, and education, as a way to push past this immobility and look at the problem from new, and perhaps more stimulating, perspectives.

CCS CONCEPTS

• **Social and professional topics** → **Children**; • **Information systems** → Environment-specific retrieval.

KEYWORDS

children, information retrieval, human-media interaction

*Co-founder of Wizenoze.

1 INTRODUCTION

As adults, we wish to help children realize their potential. Fostering children's ability to find and understand information is critical to their development as young adults. The Web presents children with exciting new ways to meet people, learn about different cultures, and take on creative challenges. However, children's access to Web resources is severely impeded by inadequate search tools. Most information retrieval (**IR**) systems are still designed for adults only: they provide resources that, most often than not, do not meet the developmental skills of children, present information in lists that children have difficulty managing, and overlook the fact that children struggle to formulate keyword queries, thus not really enabling children to find readable, relevant and reliable information they are looking for [2, 4, 21]. Worse still, almost all Web search engines expose children to inappropriate content (either because of their age or context of the search task) or filter resources to the point of retrieving none [2].

More than twenty years ago, IR researchers started exploring IR systems specific to children. Yahooligans![1], launched in March of 1996, was one of the first IR commercial large systems with a focus on children. Since then, we have seen several research attempts that from diverse perspectives seek to answer: *how can we help children searching the Web?*. Researchers have identified a range of barriers (summarized in the SWIRL 2012 report [1]) and have proposed a small set of potential solutions (e.g., [5, 8, 11, 17]). Most of these solutions, however, are based on Human-Computer Interaction paradigms and methods, that do not necessarily follow IR standards that are anchored on a system-oriented perspective rather than on a user perspective. While the preliminary exploration of the many challenges related to children and IR, including interfaces, relevance determination, diverse contexts, ethics, and many others, has contributed to a certain extent to the growth of this particular area, Yet, at present, there are no de facto standards nor child-friendly IR systems that children can use to access online information in 2021. IR for children demands specifically targeted solutions for very specific tasks and target groups. With this workshop, we aim to outline the roadmap for future research that enables us to achieve such that. To inspire the discussion, and complement the traditional

[1]https://en.wikipedia.org/wiki/Yahoo!_Kids

system-oriented perspectives with content and context perspectives, we believe it is imperative to also leverage experience from the research community related to search and recommendation systems in very specific domains, such as music (e.g. Spotify Kids[2]), movies (e.g. Netflix Kids[3]), entertainment (e.g., YouTube for kids[4]), and education (e.g. WebforClassrooms[5])

We would like to take advantage of the anniversary year by bringing together various IR researchers and other stakeholders in a workshop on children and IR. We will borrow from SWIRL 2018 [6] the idea of asking our participants to engage in a brief homework and be prepared to answer a positive question - how far have we got since then and where we would like to be in 10 more years? and a negative one - what did stop us from being already there? Together we hope these will trigger discussion in the group and help us better define the scope/remit of this exploration into the future of IR for children.

2 WHY NOW AND WHY AT SIGIR?

In 2012, IR researchers gathered for the 2nd Strategic Workshop on Information Retrieval in Lorne (SWIRL) to discuss "long-range challenges and opportunities within the field" [1]. One of the challenges specifically mentioned was "breaking information access barriers and making search facilities available to everyone" [1], with a particular focus on children. Contemporary to SWIRL, the "Towards Accessible Search Systems" workshop [20] was held over ten years ago in Switzerland and explicitly focused on one question: how to make search engines accessible to different types of users. The workshop was triggered by the PuppyIR research project that examined child-friendly IR systems from an international perspective [10, 15]. The basis for this research was the article about Yahooligans published around 2000 [3].

Now, more than 20 years later, intriguing questions remain: where are we now, in terms of research progress, in connection to this important topic? what has research from various disciplines brought us? what is on the agenda for the next decade?. The ACM SIGIR conference brings together researchers worldwide to discuss search technology, but relatively little attention is given to children as a specific target group. Child-focused IR is more frequently seen at conferences where the target group is central, e.g., International Design Conference (IDC), or a particular theme of design like Conference on Human Factors in Computing Systems (CHI).

2.1 Related Workshops

As previously stated, over the past 2 decades there have been a few projects, in addition to workshops co-located with SIGIR and other reputable venues, that have directly or indirectly discussed children's use of IR systems.

Among projects from the early 2000s, we find the International Children's Digital Library (ICDL) [14], a 5-year project focused on the design and evaluation of new interface technologies to enable children's information access, retrieval, and use [9]. The most noteworthy project is perhaps PuppyIR [7][6] –a large international

study conducted between 2009 and 2012 that aimed to help children utilize the Internet safely and successfully. The goal of PuppyIR was to provide system designers with a set of components that would enable them to integrate their information systems with a child-friendly approach. PuppyIR also resulted in new interaction paradigms that allow children to speak for themselves and get information presented intuitively. Further, it leads to academic information services as well as tools for the evaluation of IR systems for children. Researchers developed a multi-layered architecture allowing researchers to choose from different interaction paradigms and information services to build their system. Both projects set the research foundations for children and IR systems. Unfortunately, while research on children's interfaces remains advancing, albeit slowly, core IR algorithms supporting children's interactions with IR systems have not been consistently discussed at SIGIR or similar venues. There is a need for a boost, a push to rekindle and revive the interest in this area while understanding why in the meantime it has been neglected.

Among workshops co-located with ACM SIGIR, we find *Towards Accessible Search Systems*[7] [20] (SIGIR '20) and *Search as Learning* workshop [13] (SIGIR '16). Both shined a light on barriers and areas of study that demand in-depth explorations pertaining to young searchers. Workshop outcomes offer groundwork knowledge for the IR community to tackle in future research endeavors. Unfortunately, neither workshop explicitly focused on children; the main focus of workshop contributions and associated discussions were on the constraints and open challenges of adapting search systems for individuals with specific needs (e.g., children, older adults, people with visual or motor impairments, and people with intellectual disabilities or low literacy) [20] and supporting users in their learning whilst interacting with Web content [13], respectively.

Other workshop worth mentioning is the International and Interdisciplinary Perspectives on Children & Recommender and Information Retrieval Systems (KidRec)[8] [16], co-located with ACM RecSys in its 1st edition and ACM IDC thereafter. To the best of our knowledge, KidRec is the only workshop explicitly focused on children. It has attracted researchers and industry practitioners from human-computer interaction, IR, and education who have outlined a framework establishing the requirements of "good" search and recommendation systems for children. Still, core concepts and challenges inherent to the IR community (e.g., how to consider non-topical aspects of relevance, lack of search literacy, and how it affects young searchers interacting with IR systems, or lack of benchmarks for evaluations) remain overlooked.

2.2 Workshop Theme & Purpose

To foster a discussion among attendees that can help us take inventory of where are we as a community when it comes to IR systems for children, as well as where do we want and should go, we have outlined several topics of interest for the workshop. We are aware that the list is broad and extensive, nevertheless, we see this list as the catalyst motivating researchers and industry practitioners to join the conversation around the important topic of this proposed workshop. In particular, we welcome contributions spotlighting

---

[3]https://www.netflix.com/browse/genre/27346

[4]https://www.youtube.com/kids/

[5]https://www.webforclassrooms.com/

[6]https://liir.cs.kuleuven.be/projects.php?project=175

[7]https://personal.cis.strath.ac.uk/ian.ruthven/accessiblesearch/organisers.html

[8]https://kidrec.github.io/

issues related to IR systems for children–from core IR concepts to exploration of requirements that affect specific use cases. Among topics of interest we include:

- Understanding the effects of domain expertise, age, user experience and cognitive abilities on search goals and results evaluation.
- Non-topical aspects of relevance: text style, readability, appropriateness of language (harassment and explicit content detection), alignment with the context of the search.
- Considering biases and transparency in Web searching.
- Development of test collections for evaluation of classroom-related IR systems.
- Using assistive technologies for interaction with IR systems, e.g. speech recognition querying and browsing.
- Multimedia search: audio, video, images and their impact on the search experience.
- New frameworks to model the searching for learning vs. searching for pleasure paradigms.
- Beyond Cranfield, moving on to online evaluation, task-based, session-based, multi-turn, interactive search.
- One size does not fit all when it comes to children:
  – Collaborative search techniques for assisting users with specific needs (e.g. parents helping children).
  – Potential of search personalization techniques to satisfy users with specific needs.
- Considerations pertaining to IR systems supporting learning:
  – Understanding of search behavior of users with specific classroom-related needs.
  – Understanding of relevance criteria of users with specific classroom-related needs.
  – Search interfaces and result representation for people with specific classroom-related needs.
  – The social side of searching in the classroom and the implications on the design of search tools to be shared by adults (teachers and parents) and children.
  – Ethics in the classroom: defining what is suitable, good and useful when searching in the classroom, who is in charge, and who decides.
  – Web search from a teacher's standpoint.
- Domain-specific use cases for IR systems for children:
  – IR systems for music, podcasts, books, and videos.

3 ORGANIZERS

We have assembled a multidisciplinary team with expertise in IR, human-computer interaction, and digital media, and K-12 education; we also ensured academic and industry diversity.

Theo Huibers has been researching IR and human media interaction for over 30 years. Since 2002, he is a professor in Human Media Interaction & Computer Science at the University of Twente in the Netherlands and co-founder of an international edTech company called Wizenoze, founded in 2013. Huibers co-initiated a European FP7 research project called PuppyIR in 2008. This international project was granted 4.3M euro from the European Commission for a 3-year study (2009-2012) on all aspects (e.g. ethics, data-gathering, technology, and business models) of developing an open-source search and media environment for children.

Monica Landoni is a Senior Researcher at the faculty of Informatics at Università della Svizzera Italiana (USI). She has worked on many national and European projects, including PuppyIR, investigating how technology can support children when searching, writing and reading for education and pleasure. While doing that she has happily survived the design and running of many collaborative design sessions in formal and informal settings. Carefully taking into account the different needs, requests, roles and points of views of parents, teachers, librarians; always putting children first.

Emiliana Murgia is a primary school teacher at the Stoppani Institute in Milan, where she works on developing and experimenting innovative teaching methods with technology. She is also affiliated with the Department of Human Sciences for Education at the University of Milano Bicocca. Emiliana has worked on many national projects investigating how technology can support children so that they can get the best out of their learning experience. Emiliana is also a co-organizer of the KidRec workshop [18].

Maria Soledad Pera is an Associate Professor in the CS Department at Boise State University. Sole's research focuses on IR, in particular, IR for non-traditional users, such as children, educators, and individuals affected by mental health disorders. Her work related to IR applications tailored towards children has been funded by the National Science Foundation. She has served as PC and reviewer for IR conferences and journals, including SIGIR and RecSys; she was General Chair of ACM RecSys 2018. Sole is one of the co-organizers of the KidRec workshop, currently in its 5th edition [18].

4 WORKSHOP PLANS

We propose a highly interactive workshop to foster active discussions on children and IR among industry and academia participants.

4.1 Format

Given the restrictions imposed by the ongoing COVID-19 pandemic, we propose a half-day workshop. By allocating only half a day, we can accommodate participants from around the world, while avoiding the fatigue usually associated with online work gatherings [19].

4.2 Website

We will make available a website (http://www.fab4.science/IR4C/), where we will post relevant information related to the workshop, including important dates, accepted contributions, and schedule.

4.3 Participants & Call for Contributions

We anticipate a call for *vision* papers focusing on past experiences, ongoing projects, and future directions related to IR systems for whom children are the target audience. These papers will be at most 4 pages in length (as per the 1 column ACM template) and be submitted via EasyChair. Accepted contributions, selected through peer-review, for which we will recruit a Program Committee (see Section 4.5), will be published on the workshop website.

Note that we will also accept *informal submissions* (via online form) from practitioners interested in participating.

In addition to sharing CFP via social media posts (e.g., Twitter) and traditional outlets (e.g., SIGIR-List) we will directly reach out to

researchers involved in PuppyIR, as well as industry practitioners potentially interested in this topic.

4.4 Proposed Activities

We will facilitate a highly participatory workshop [12] in which attendees can discuss the landscape of the area under study, identify key elements that contribute to the current status of IR systems for children, and propose future research directions. We will do so via an interactive format involving community building exercises, informal interactions, facilitated group work, and very short presentations of accepted contributions and attendees' backgrounda.

To encourage lively brainstorming we will assign a **pre-workshop** homework to our participants and invite them to answer "how far have we got since then and where we would like to be in 10 more years?" (a positive question to help us better define the remit of this exploration into the future of IR for children) and "what did stop us from being already there?" (a negative one about existing barriers).

The proposed workshop **schedule** (times in CET) as follows:

Welcome & Introductions (16.00-16.15). A brief welcome address, along with an overview of activities planned for the workshop and tools that we will use to encourage interaction, e.g., Miro and Zoom's breakout rooms.

Overview (16.15-16.30). A brief presentation showcasing the motivating factors prompting the need for space for the IR community to discuss children and IR systems.

Lightning Round (16.30-17.00). Showcase of accepted contributions (i.e., vision papers and interest form); presentations will last 5 minutes. These presentations ensure that workshop participants are aware of each others' interests and expertise, but will be short to keep the workshop flow and to allow for focused group work later on.

Discussion (17.00-17.45). Discussions in various groups to identify the most important research questions for the next ten years.

Outcomes (17.45-18.45). Joint discussion to merge findings from small-group work, resulting in a research agenda for the next 10 years for children and IR.

Wrap (18.45-19.30). Final notes emerging from the day's group work. Plans for next steps.

4.5 Potential Program Committee

We will gather a program committee with the experience to examine submissions and value the respective possible contributions regarding informing workshop discussion on understanding the "status" of this particular area of study as well as helping shape future research agenda on this area.

5 EXPECTED WORKSHOP OUTCOMES

A primary goal of the workshop is to reactivate the network of IR people with a (future) focus on children. The various research institutes and large and small tech companies want to jointly define a research agenda for the next ten years, in which all topics are considered in a broad context so that we can discuss how to collaborate going forward. Based on this workshop, we hope there will be a centrifugal force for IR and Children, so that ten years from now we will be able to say, "*Well, we have really moved forward*".

As an immediate outcome, we will summarize findings from this workshop on a report to be submitting to SIGIR Forum, to reach the IR community at large.

REFERENCES

[1] James Allan, Bruce Croft, Alistair Moffat, and Mark Sanderson. 2012. Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop on information retrieval in Lorne. In *ACM SIGIR Forum*, Vol. 46. 2–32.

[2] Oghenemaro Anuyah, Ashlee Milton, Michael Green, and Maria Soledad Pera. 2019. An empirical analysis of search engines' response to web search queries associated with the classroom setting. *Aslib Journal of Information Management* (2019).

[3] Dania Bilal. 2000. Children's use of the Yahooligans! Web search engine: I. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for information Science* 51, 7 (2000), 646–665.

[4] Dania Bilal and Li-Min Huang. 2019. Readability and word complexity of SERPs snippets and web pages on children's search queries. *Aslib Journal of Information Management* (2019).

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

[6] J Shane Culpepper, Fernando Diaz, and Mark D Smucker. 2018. Research frontiers in information retrieval: Report from the third strategic workshop on information retrieval in lorne (swirl 2018). In *ACM SIGIR Forum*, Vol. 52. 34–90.

[7] F.M.G. de Jong. 2012. PuppyIR – An Open Source Environment to construct Information Services for Children. Final report. Available at: shorturl.at/ixO13.

[8] Nevena Dragovic, Ion Madrazo Azpiazu, and Maria Soledad Pera. 2016. " Is Sven Seven?" A Search Intent Module for Children. In *Proc. of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 885–888.

[9] Allison Druin, Benjamin B Bederson, Ann Weeks, Allison Farber, Jesse Grosjean, Mona Leigh Guha, Juan Pablo Hourcade, Juhyun Lee, Sabrina Liao, Kara Reuter, et al. 2003. *The International Children's Digital Library: Description and analysis of first use.* Technical Report.

[10] Carsten Eickhoff, Pieter Dekker, and Arjen P De Vries. 2012. Supporting children's web search in school environments. In *Proc. of the 4th Information Interaction in Context Symposium*. 129–137.

[11] Jerry Alan Fails, Maria Soledad Pera, Oghenemaro Anuyah, Casey Kennington, Katherine Landau Wright, and William Bigirimana. 2019. Query formulation assistance for kids: What is available, when to help & what kids want. In *Proc. of the 18th ACM International Conference on Interaction Design and Children*. 109–120.

[12] Seeds for Change. 2017. *Facilitating Participatory Workshops. Available at: https:// we.riseup.net/ assets/ 25682/ FacilitatingWorkshops.pdf* .

[13] Jacek Gwizdka, Preben Hansen, Claudia Hauff, Jiyin He, and Noriko Kando. 2016. Search as learning (SAL) workshop 2016. In *Proc. of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 1249–1250.

[14] Hilary Browne Hutchinson, Allison Druin, and Benjamin B Bederson. 2007. Supporting elementary-age children's searching and browsing: Design and evaluation using the international children's digital library. *Journal of the American Society for information Science and Technology* 58, 11 (2007), 1618–1630.

[15] Michel Jansen, Wim Bos, Paul van der Vet, Theo Huibers, and Djoerd Hiemstra. 2010. TeddIR: tangible information retrieval for children. In *Proc. of the 9th international conference on interaction design and children*. 282–285.

[16] Monica Landoni, Jerry Alan Fails, Theo Huibers, Natalia Kucirkova, Emiliana Murgia, and Maria Soledad Pera. 2020. 4th KidRec workshop "what does good look like?" from design, research, and practice to policy. In *Proc. of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts*. 103–110.

[17] Monica Landoni, Davide Matteri, Emiliana Murgia, Theo Huibers, and Maria Soledad Pera. 2019. Sonny, Cerca! evaluating the impact of using a vocal assistant to search at school. In *International Conference of the Cross-Language Evaluation Forum for European Languages*. Springer, 101–113.

[18] Monica Landoni, Maria Soledad Pera, Jerry Alan Fails, Emiliana Murgia, Natalia Kucirkova, and Theo Huibers. 2020. 4 th KidRec–What does Good Look Like: From Design, Research, and Practice to Policy. In *SIGIR forum*, Vol. 54.

[19] Robby Nadler. 2020. Understanding "Zoom fatigue": Theorizing spatial dynamics as third skins in computer-mediated communication. *Computers and Composition* 58 (2020), 102613.

[20] Pavel Serdyukov, Djoerd Hiemstra, and Ian Ruthven. 2011. Towards accessible search systems. In *ACM SIGIR Forum*, Vol. 44. 23–27.

[21] Catherine L Smith. 2017. Investigating the role of semantic priming in query expression: A framework and two experiments. *Journal of the Association for Information Science and Technology* 68, 1 (2017), 168–181.