What does good look like?: Report on the 3rd International and Interdisciplinary Perspectives on Children & Recommender and Information Retrieval Systems (KidRec) at IDC 2019

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

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

Maria Soledad Pera
People and Information Research Team
Dept. of Computer Science
Boise State University – Boise, ID, USA
solepera@boisestate.edu

Jerry Alan Fails
Dept. of Computer Science
Boise State University
Boise, ID, USA
jerryfails@boisestate.edu

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

Natalia Kucirkova
University of Stavanger
Stavanger, Norway
natalia.kucirkova@uis.no

Abstract

This short review discusses the outcomes of the 3rd Workshop of the *International and Interdisciplinary Perspectives on Children & Recommender and Information Retrieval Systems (KidRec 2019)*, co-located with the 2019 ACM Interaction Design and Children (IDC) Conference, which took place June 12-15 in Boise, Idaho, USA. The goal for the workshop was to explore the characteristics of a "good" information retrieval system for children. The diversity of attendees – including industry representatives, undergraduate and graduate students, and senior scholars in various areas of computer science – made it possible to continue to build community around this important topic and further discuss and outline the salient concerns and the next steps to promote further exploration in this area.

1   Introduction

KidRec is the Workshop of the International and Interdisciplinary Perspectives on Children & Recommender and Information Retrieval Systems. Building on the complex problems related to information retrieval (IR) systems for children identified during the 1st KidRec workshop (co-located with ACM RecSys in 2017) [8] and the lessons learned from the discussions pertaining to the educational IR systems during the 2nd KidRec (co-located with ACM IDC 2018) [3], we organized the 3rd KidRec (co-located with ACM IDC 2019) to focus on the characteristics of a "good" information retrieval system for children. Good was defined as optimal, effective, and child-appropriate.

Interaction Design and Children (IDC) is a conference that is focused on the target audience of interest for KidRec: children. Further, IDC welcomes attendees who are experts in various areas of interest in computer science, as well as education and social sciences. Due to the multidisciplinary nature of the work typically presented at IDC and the diversity of its participants, we felt that IDC was the ideal venue to continue to nurture dialogue around child-centered information retrieval systems. IDC 2019 was the 18th annual conference and was held in Boise, Idaho, June 2019.

The workshop brought together an interdisciplinary group of researchers in education, child-development and computer science as well as designers interested in issues related to information retrieval (**IR**) systems, e.g., search and recommendation engines, designed for children. The workshop participants discussed existing platforms and apps that incorporate IR technology, as well as ethics, techniques, and privacy and security issues related to educationally-oriented IR systems. The common thread throughout the discussion was how to evaluate what makes an IR system for children "good" when multiple perspectives (both at the algorithm and stakeholder level) needed to coexist.

In the rest of this report, we present the structure of our interactive workshop and discuss the outcomes including the outline of an initial framework that we collectively defined for determining how to quantify and qualify "good" performance of IR systems for children.

2   Workshop Description

As outlined in the workshop proposal [5], we aimed to organize a highly-interactive experience where accepted contributions would drive the conversation on important issues related to the evaluation of information retrieval systems for which children are the target audience.

2.1   Audience of Interest

To foster an inclusive and heterogeneous discussion, we approached researchers and developers with diverse expertise: technology for children; human-computer interaction; information retrieval; assessment; ethics; educators; learning and development; educational games, and e-commerce. By reaching out to members from academia and industry, we anticipated continuing to foster a multidisciplinary perspective that pinpoints challenges when evaluating information retrieval technology for which children are the major stakeholders.

2.2 Accepted Contributions

Six papers were selected to be presented during the workshop [1, 2, 4, 6, 7, 9]. Based on these contributions, we organized two interactive panel sessions. The goal for these sessions was for presenters to highlight key elements of their work and initiate discussion among the attendees. We include below the abstracts for accepted contributions.

2.2.1 Interactive Panel 1

My Name is Sonny, How May I help You Searching for Information? [7]. We describe the set up and procedure followed in an initial study where we observe a small group of primary school children interact with a vocal assistant or a standard GUI interface, while looking for online resources that can help them answer a set of predefined questions. We examine log files, observations, and feedback provided by the young users via an elementary interview at the end of each search section. Results from our analysis prompt us to consider how to evaluate the search process, specially one initiated with a vocal assistant, while considering both correctness of results, children preferences, and context of the search.

Evaluating "Just Right" in EdTech recommendation [9]. Continuing with the recommendation system progress presented at KidRec 2018, Age of Learning product, design, and research leaders are expanding their work on recommendation systems to include conversation across three different products: ABCmouse, Mastering Math, and ReadingIQ. Authors present a description of ways in which a "just right system for one product may differ from that of another product, depending on the learning and experience goals of the particular platform. Products are evaluated in their current states and in light of future plans based on scales of dynamic content presentation, perceptions of a personalized user experience, and availability of content from which children can choose.

Evaluating prediction-based recommenders for kids [4]. In this position paper, we highlight a number of issues that exist with the use of traditional metrics in evaluating recommender systems when children are the target audience. Our focus is on discrepancies that arise as a result of the differing rating behaviour of adult users when compared to children, and how these differences can warrant a reconsideration of existing assessment metrics and their validity in this context.

2.2.2 Interactive Panel 2

Relevance and utility in an educational search environment [6]. Technology is increasingly being used in education. Children are also increasingly using search engines when searching for information on certain themes. The question of what is a good search system to use in education has still not been answered definitively. In this article we explain the steps Wizenoze takes to build and evaluate a good search system. We split our analysis in the way Cooper already proposed in 1971 as the two aspects of relevance: logical relevance and utility.

The need for a comprehensive strategy to evaluate search engine performance in the classroom [1]. Given how ingrained Search Engines (SEs) are in educational environments, it is essential to evaluate their performance in response to inquiries that pertain to the classroom setting. In this position paper, we discuss the limitations of relying solely on traditional Information Retrieval metrics and usability studies. We argue in favor of a new comprehensive strategy for assessment that integrates other aspects (i.e., the search context)

with new and existing measures, in order to better quantify positive and negative factors that influence SEs outcomes targeting children.

Children and search tools: Evaluation remains unclear [2]. As children search the internet for materials, they often turn to search engines that, unfortunately, offer children little support as they formulate queries to initiate the search process or examine resources for relevance. While some solutions have been proposed to address this, inherent to this issue is the need to evaluate the effectiveness of these solutions. We posit that the evaluation of the diverse aspects involved in the search process from query suggestion generation to resource retrieval requires a complex, multi-faceted approach that draws on evaluation methods utilized in human-computer interaction, information retrieval, natural language processing, education, and psychology.

2.3 Discussion, Framework, & Open Challenges

The conversation that arose from the presentation of the accepted contributions led us to the question of what is *perceived* to be correct or "good" versus what it is *actually* correct, when it comes to IR systems for children. This turned out to be a theme that continued to appear throughout the day, as the former is better quantified from a user perspective (usability) whereas the latter is assessed from a system perspective (well-known IR metrics). The evaluation however is further complicated by the fact that children often disagree with parents/teachers (and other major stakeholders) in what is good and engaging for them, when it comes to IR tools that they can turn to to locate resources of interest.

Another insight that emerged from the discussion is the need for suitable resources, especially for the classroom context. In this case, the biggest issue affecting "goodness" of the IR system is the need for *curation*: on the one hand, we want children to take the best advantage possible from online resources available online. On the other hand, curation might prevent access to inappropriate or simply unsuitable content. This is advantageous but it requires more data points (resources), which might be the reason why more commercial IR systems, rather than those freely-accessible, are the ones offering such content.

In the end, relevance and usability are only two of the many facets that can and should be used as guidance for the design, development, and evaluation of IR systems for children. Others include utility for all the stakeholders involved (children, teachers, parents, system hosts, content providers, etc.), external factors (e.g., in an educational context, learning is not always fun and as such the systems must align with curricular expectations if it is meant to be labeled as "good").

Various activities were completed during the day, including: *sticky note activities* to identify challenges, opportunities, and facets to be considered when designing what is good; *group discussions* on open problems in the area; and voting on main perspectives to consider moving forward. From these activities a framework emerged, which we present below.

An IR system for children is "good" if:

- It provides resources that are logically relevant, useful, and foster learning

- It is designed with a user-centered perspective while acknowledging that multiple stakeholder perspectives and needs exist

- It is ethically sound and supports the right of the child

- Users are deeply engaged with the system

In the end, we agree that the framework offers a starting point that researchers and developers both in industry and academia can use to measure the degree to which their systems adequately meet the needs and expectations of the target user under study. Next steps will require identifying metrics that individually and collectively can not only demonstrate the overall "goodness" of a given system but also serve as a standard for comparison with new systems.

3 Next Steps & Future Workshop

The workshop brought together stakeholders from various perspectives. It fostered a rich discussion that resulted in an initial framework that can be utilized, evaluated, revised, and expanded.

We plan to continue the multiple-stakeholder conversations at venues like the next ACM IDC conference, which will take place in London, England, in June 2020. In the meantime, additional information pertaining to KidRec 2019, including a list of accepted papers and the detailed workshop schedule, can be found at the workshop website: `https://kidrec.github.io/`; and you can join the conversion by emailing kidrec-group@boisestate.edu.

4 Acknowledgments

We thank the KidRec 2019 Program Committee members for their prompt and insightful reviews, which directly impacted the level of the workshop discussions. We also thank the Organizing Committee for IDC 2019 – especially the Workshop Chairs Lisa Anthony (University of Florida, USA) and Asimina Vasalou (University College London, UK) – for giving us the opportunity to host this workshop in conjunction with the main IDC 2019 conference.

We also thank the workshop presenters and attendees, who greatly contributed to discussion and continue to help create a foundation for future areas of work and topics of interests upon which this community can continue to grow.

References

[1] O. Anuyah, M. Green, A. Milton, and M. S. Pera. The need for a comprehensive strategy to evaluate search engine performance in the classroom. In 3rd KidRec Workshop co-located with ACM IDC 2019. Available at: `https://kidrec.github.io/papers/KidRec_2019_paper_1.pdf`, 2019.

[2] B. Downs, T. French, K. L. Wright, M. S. Pera, C. Kennington, and J. A. Fails. Children and search tools: Evaluation remains unclear. In 3rd KidRec Workshop co-located with ACM IDC 2019. Available at: `https://kidrec.github.io/papers/KidRec_2019_paper_5.pdf`, 2019.

[3] J. A. Fails, M. S. Pera, and N. Kucirkova. Building community: Report on the 2nd international and interdisciplinary perspectives on children & recommender systems (kidrec) at IDC 2018. *ACM SIGIR Forum*, 52(1):138–144, 2019.

[4] M. Green, O. Anuyah, D. Karsann, and M. S. Pera. Evaluating prediction-based recommenders for kids. In 3rd KidRec Workshop co-located with ACM IDC 2019. Available at: `https://kidrec.github.io/papers/KidRec_2019_paper_2.pdf`, 2019.

[5] T. Huibers, N. Kucirkova, E. Murgia, J. A. Fails, M. Landoni, and M. S. Pera. 3rd kidrec workshop: What does good look like? In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, pages 681–688. ACM, 2019.

[6] T. Huibers and T. Westerveld. Relevance and utility in an educational search environment. In 3rd KidRec Workshop co-located with ACM IDC 2019. Available at: `https://kidrec.github.io/papers/KidRec_2019_paper_4.pdf`, 2019.

[7] M. Landoni, E. Murgia, T. Huibers, and M. S. Pera. My name is Sonny, how may i help you searching for information? In 3rd KidRec Workshop co-located with ACM IDC 2019. Available at: `https://kidrec.github.io/papers/KidRec_2019_paper_3.pdf`, 2019.

[8] M. S. Pera, J. A. Fails, M. Gelsomini, and F. Garzotto. Building community: Report on kidrec workshop on children and recommender systems at recsys 2017. *ACM SIGIR Forum*, 52(1):153–161, 2018.

[9] M. Rothschild, T. Horiuchi, and M. Maxey. Evaluating "just right" in edtech recommendation. In 3rd KidRec Workshop co-located with ACM IDC 2019. Available at: `https://kidrec.github.io/papers/KidRec_2019_paper_6.pdf`, 2019.