

1st Workshop on Information Retrieval for Understudied Users (IR4U2)

Maria Soledad Pera¹[0000–0002–2008–9204], Federica Cena²[0000–0003–3481–3360],
Theo Huibers⁴[0000–0002–9837–8639], Monica Landoni⁶[0000–0003–1414–6329],
Noemi Mauro³[0000–0001–8234–3266], and Emiliana Murgia⁵[0000–0002–9728–1771]

¹ Web Information Systems - TU Delft, The Netherlands m.s.pera@tudelft.nl

² University of Turin, Turin, Italy federica.cena@unito.it

³ University of Twente, Enschede, The Netherlands t.w.c.huibers@utwente.nl

⁴ Università della Svizzera italiana, Lugano, Switzerland monica.landoni@usi.ch

⁵ University of Turin, Turin, Italy noemi.mauro@unito.it

⁶ Università di Genova, Italy emiliana.murgia@unige.it

Abstract. Information Retrieval (IR) remains an active, fast-paced area of research. However, most advances in IR have predominantly benefited the so-called “classical” users, e.g., English-speaking adults. We envision IR4U2 as a forum to spotlight efforts that, while sparse, consider *diverse*, and often *understudied*, user groups when designing, developing, assessing, and deploying the IR technologies that directly impact them. The key objectives for IR4U2 are: (1) raise awareness about ongoing efforts focused on IR technologies designed for and used by often understudied user groups, (2) identify challenges and open issues impacting this area of research, (3) ignite discussions to identify common frameworks for future research, and (4) enable cross-fertilization and community-building by sharing lessons learned from research catering to different audiences by researchers and (industry) practitioners across various disciplines.

Keywords: Understudied users · Information Access · Information Retrieval.

1 Motivation

Information retrieval (**IR**) is the dominant form of information access [9]. As such, it is not surprising that technological advances that directly and indirectly foster information access continue to be of interest to researchers and industry practitioners alike. Over the past few decades, developments in search, recommender, and question-answering systems, along with complementary IR tasks such as clustering, filtering, and text processing, have been well studied and documented; so are the efforts that help IR keep up with the fast-paced era of Large Language Models (LLM), Artificial Intelligence (AI), new modalities of interactions, and platforms enabling information access. We see, however, that works at SIGIR, ECIR, RecSys, and other related venues for the most part aid or study the so-called “classical,” e.g., English-speaking adults.

Almost 20 years ago, Anderson et al. [1] discussed the concept of digital democratization of society and its connection to affordable Internet and enhanced information access to users from all backgrounds [4]. From reports from Statista [13], we know that today there are more than 5 billion Internet users *worldwide*. The most popular reasons for using the Internet, among users 16 years old and up, pertain to information access: places and travel, following news and events, finding information, along with education and studying [14]—the last two are also common reasons among younger user groups [3]. Has the IR community, however, genuinely reflected on whether existing IR technologies genuinely support the right to information access of *all* these billions of users? Consider popular search engines that are a portal to a rich and up-to-date set of resources: Do they account for individuals with low literacy levels who might not comprehend the resources retrieved? What about stereotypes that inadvertently might make them into their sets of retrieved results? Are these platforms naturally accessible and ready to prioritize retrieval of resources suited, for example, to those who are visually impaired [15,11,3]? Similar issues affect recommender systems, as recommendation algorithms used on platforms like TikTok or YouTube might inadvertently expose children to disturbing materials with far-reaching consequences [2]. What about AI technologies powering varied IR solutions: Are they inclusive and ensure fair representation of understudied populations?

Although IR works that aim to foster information access among people with diverse abilities and needs exist, they are very sporadic, and target, for instance, older adults, young children and adolescents, individuals afflicted by mental health disorders, users with autism spectrum disorder (ASD), individuals with intellectual disabilities, or those with specific learning needs [19,5,12,10,16,6]. Venues like CHI, UMAP, IDC, and ASSETS have recognized the impact of technology on these understudied populations. In the case of IR, the community has started to look at fairness and bias constructs, bringing awareness to issues of gender and minority representation. Still more is required to get a comprehensive mosaic of how IR technology can and should be shaped to better serve all its users and what that means. To elevate access to information, it is key to account for the environments in which these technologies are deployed and the nuances of individuals using IR technologies. The former is something the IR community is already undertaking, evidenced by endeavors targeting advancements in domain-specific IR, including scholarly works and workshops focused on the design, evaluation, and use of IR systems (and complementary IR technology) in medical [18], legal [17], fashion [8], and financial [7] domains, etc. Still, the wide range of users with different skill sets and needs has been mostly overlooked.

We argue that to understand how and whether IR technology can truly serve a diverse population it is critical to reflect on different use cases and take a human-driven approach that puts the different needs of real users at the center. Doing so requires multidisciplinary discussion and mutual understanding among the broad and sometimes conflicting perspectives if solutions will, in turn, fit the real-world setting and respond to the complex (information) requirements of a wide range of individuals. IR4U2 is meant to serve as a forum to bring together

researchers and practitioners who have consistently allocated efforts to advance knowledge in this area but have seen their contributions accepted at non-core IR venues. By co-locating IR4U2 at ECIR we intend to make understudied groups more visible among the IR community.

2 Workshop Overview

The **goals** for IR4U2 are to: (1) Build a community involving multidisciplinary Ph.D. students, academics, and industry practitioners aiming to advance understanding of core IR technology and its impact on a broad range of understudied users; (2) Take inventory of the different kinds of use cases, research work, ongoing efforts, and existing resources; (3) Identify open challenges and opportunities for future research directions; and (4) Outline a general approach to facilitate research endeavors driven by and at the service of specific user groups.

We expect contributions that will help construct a snapshot of the works in this area. For this, we will elicit two types of submissions: **peer-reviewed** contributions (research papers and position/vision papers), as well as **editorially-reviewed** contributions (to enable potential attendees to articulate their views on the topics of the workshop, share already-published works, bring awareness to ongoing European projects in this area). **Topics** of interest include: (i) User modeling to enable IR and recommendation technologies tailored to understudied populations, (ii) Data collection and benchmark development of IR catering to understudied populations, (iii) IR applications targeting understudied populations, (iv) UI/UX for search, recommender, and question-answering systems for understudied populations, (v) IR-related technology (clustering, classification, text processing, text complexity) and their impact on understudied populations, (vi) Different perspectives of evaluation, (vii) Design of the user interaction with IR systems, and (viii) Ethical issues associated with IR technologies for understudied users; e.g., their right to be represented, acknowledged, have access, and be served by existing and to-be-developed systems. This is not an exhaustive list, as we are also interested in contributions discussing challenges inherent to designing IR technology at the service of understudied user groups—from the need for multidisciplinary, multi-stakeholder collaborations to how to build datasets.

We will host a **highly participatory, full-day, in-person** workshop, which will include a panel where we will invite participants championing different understudied users and discuss the importance of adopting a user-driven approach and integrating a wide range of perspectives to advance IR in a meaningful manner; brief presentations of accepted contributions; small group discussions; and joint discussions to merge findings from work in small groups that will result in plans for future editions of the IR4U2 workshop. A complete program can be found on the workshop website: <https://ir4u2workshop.wixsite.com/ir4u2>.

3 Organizers

Sole Pera is an Associate Professor at the Web Information Systems group (EEMCS) at TU Delft. Sole’s research focuses on IR, with a special emphasis on enhancing information access for typically underserved user groups. She serves as (S)PC for conferences including SIGIR, UMAP, CHIIR, RecSys, and ECIR. She was General Chair for RecSys ’18 and Program Chair for UMAP ’23. Among others, she co-organized the Workshop on Educational Recommender Systems, the ComplexRec workshop, and 6 editions of the KidRec workshop.

Federica Cena is an Associate Professor at the Computer Science Department of the University of Turin. She works on the intersection of Artificial Intelligence and Human-Computer Interaction. Her recent research has studied the implications of the Internet of Things for user modeling and personalization, with a special focus on assistive applications for cognitive disabilities and frailty.

Theo Huibers has been researching information retrieval and human media interaction for over 30 years. Since 2002, he has been a professor in Human Media Interaction & Computer Science at the University of Twente and co-founder of Wizenoze, an international eTech company founded in 2013.

Monica Landoni is a titular professor at the faculty of Informatics at Università della Svizzera Italiana. She is vice chair of the ACM IDC Conference Steering Committee and an active member of EUGAIN, the European Network For Gender Balance in Informatics. She has worked on national and European projects investigating how technology can support children when searching, writing, and reading for education and pleasure. While doing that, she has happily designed and conducted many collaborative design sessions in formal and informal settings, carefully taking into account the needs, requests, roles, and points of view of real users, often from understudied communities.

Noemi Mauro is an Assistant Professor at the Computer Science Department of the University of Torino. Her research interests concern user modeling, recommender systems, cultural heritage, information filtering, and information visualization. She won the best paper award at UMAP ’20 with the paper “Personalized Recommendation of PoIs to People with Autism”. She is a PC member of the top conferences in her research areas and a reviewer for several related journals. She co-edited the special issue “Intelligent Systems for People with Diverse Cognitive Abilities” in the Human-computer Interaction journal. She is an Editorial Board Member of the UMUAI journal.

Emiliana Murgia has a degree in Literature and various training courses, including Communication and Management of School and Training Institutions. She transitioned from a career in communication to teaching at the primary school in 1999/2000; actively promoting technology in education, she has been a “digital animator” at school, managing various (inter)national projects, and has collaborated with the University of Milano Bicocca since 2014 on teaching with technologies. In 2018, she joined a multidisciplinary research team focusing on online information access, including AI. Currently on secondment to participate in a National Doctorate program in Learning Sciences and Digital Technologies.

References

1. Anderson, R.H., Bikson, T.K., Law, S.A., Mitchell, B.M.: Universal access to e-mail: feasibility and societal implications (1995)
2. Archie, A.: A u.k. agency has fined tiktok nearly \$16 million for handling of children’s data (Apr 2023), <https://www.npr.org/2023/04/05/1168114842/tik-tok-uk-fine>
3. Azpiazu, I.M., Dragovic, N., Pera, M.S., Fails, J.A.: Online searching and learning: Yum and other search tools for children and teachers. *Information Retrieval Journal* **20**, 524–545 (2017)
4. Baker, P., Potts, A.: ‘why do white people have thin lips?’google and the perpetuation of stereotypes via auto-complete search forms. *Critical discourse studies* **10**(2), 187–204 (2013)
5. Danovitch, J.H.: Growing up with google: How children’s understanding and use of internet-based devices relates to cognitive development. *Human Behavior and Emerging Technologies* **1**(2), 81–90 (2019)
6. Delgado, P., Ávila, V., Fajardo, I., Salmerón, L.: Training young adults with intellectual disability to read critically on the internet. *Journal of Applied Research in Intellectual Disabilities* **32**(3), 666–677 (2019)
7. Feng, F., Luo, C., He, X., Liu, Y., Chua, T.S.: Finir 2020: The first workshop on information retrieval in finance. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. pp. 2451–2454 (2020)
8. Jaradat, S., Dokoochaki, N., Corona Pampin, H.J., Shirvany, R.: Workshop on recommender systems in fashion and retail. In: *Proceedings of the 15th ACM Conference on Recommender Systems*. pp. 810–812 (2021)
9. Manning, C.D.: An introduction to information retrieval. Cambridge university press (2009)
10. Mauro, N., Ardissono, L., Cena, F.: Personalized recommendation of pois to people with autism. In: *Proceedings of the 28th ACM conference on user modeling, adaptation and personalization*. pp. 163–172 (2020)
11. Meyer, G., Wassyng, A., Lawford, M., Sabri, K., Shirani, S.: Literature review of computer tools for the visually impaired: A focus on search engines. *Artificial Intelligence in Healthcare and Medicine* pp. 237–259 (2022)
12. Milton, A., Pera, M.S.: Into the unknown: Exploration of search engines’ responses to users with depression and anxiety. *ACM Transactions on the Web* **17**(4), 1–29 (2023)
13. n.a.: Internet and social media users in the world 2023 (Oct 2023), <https://www.statista.com/statistics/617136/digital-population-worldwide/>
14. n.a.: Reasons for using the internet worldwide by age 2022 (May 2023), <https://www.statista.com/statistics/1387376/internet-using-global-reasons-by-age/>
15. Noble, S.U.: Algorithms of oppression. In: *Algorithms of oppression*. New York university press (2018)
16. Tsiakas, K., Barakova, E., Khan, J.V., Markopoulos, P.: Brainhood: towards an explainable recommendation system for self-regulated cognitive training in children. In: *Proceedings of the 13th ACM International Conference on Pervasive Technologies Related to Assistive Environments*. pp. 1–6 (2020)
17. Verberne, S., Kanoulas, E., Wiggers, G., Piroi, F., de Vries, A.P.: Ecir 2023 workshop: Legal information retrieval. In: *European Conference on Information Retrieval*. pp. 412–419. Springer (2023)

18. White, R.W., Yom-Tov, E., Horvitz, E., Agichtein, E., Hersh, W.: Report on the sigir 2013 workshop on health search and discovery. In: ACM SIGIR Forum. vol. 47, pp. 101–108. ACM New York, NY, USA (2013)
19. Xie, B., Charness, N., Fingerman, K., Kaye, J., Kim, M.T., Khurshid, A.: When going digital becomes a necessity: Ensuring older adults' needs for information, services, and social inclusion during covid-19. In: Older Adults and COVID-19, pp. 181–191. Routledge (2021)