Workshop on Fairness, Accountability, Confidentiality, Transparency, and Safety in Information Retrieval (FACTS-IR)

Alexandra Olteanu
Microsoft Research
Montreal & New York City
alexandra.olteanu@microsoft.com

Jean Garcia-Gathright
Spotify Research
Boston, MA, USA
jean@spotify.com

Maarten de Rijke
University of Amsterdam
Amsterdam, The Netherlands
derijke@uva.nl

Michael D. Ekstrand
Boise State University
Boise, ID, USA
michaelekstrand@boisestate.edu

ABSTRACT

This workshop explores challenges in responsible information retrieval system development and deployment. The focus is on determining actionable research agendas on five key dimensions of responsible information retrieval: fairness, accountability, confidentiality, transparency, and safety. Rather than just a mini-conference, this workshop is an event during which participants are expected to work. The workshop brings together a diverse set of researchers and practitioners interested in contributing to the development of a technical research agenda for responsible information retrieval.

CCS CONCEPTS

• **Information systems** → **World Wide Web**; **Information retrieval**; • **Human-centered computing**; • **Computing methodologies** → *Artificial intelligence*; *Machine learning*;

KEYWORDS

Responsible information retrieval, fairness, accountability, confidentiality, transparency, safety

ACM Reference Format:
Alexandra Olteanu, Jean Garcia-Gathright, Maarten de Rijke, and Michael D. Ekstrand. 2019. Workshop on Fairness, Accountability, Confidentiality, Transparency, and Safety in Information Retrieval (FACTS-IR). In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '19), July 21–25, 2019, Paris, France.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3331184.3331644

1 INTRODUCTION

Information retrieval (IR) systems and related technologies, such as recommender systems, are responsible for organizing, curating, and promoting most of the information that is being consumed today. Importantly, IR systems are not isolated systems: they reflect the content and interaction data used to develop them and their impact on the environments in which they operate. Indeed, IR systems connect people to information, shaping not only the information consumption patterns, but also the social interactions, affecting who is more visible and when [4, 14].

Recognition of the social and political implications of information retrieval goes back at least two decades [8]. More recent empirical evidence shows, for instance, that there are gaps in access to information across communities, in part due to the information needs of certain communities being less supported than those of others—often the more dominant communities [6, 9, 13, 14]. As with other AI-driven technologies, IR systems are also under the influence of those that design, build, maintain or use them, embedding and amplifying their biases [2, 3, 7, 11]. Failures of IR systems may not always be easily traceable [12] and the extensive use of interaction logs may lead to undesirable leaking of sensitive information [17]. While users are now entitled to explanations of algorithmic decisions in certain parts of the world [5], it is unclear how explanations, evidence-trails and provenance might be communicated to the various user groups and how such communications might change behaviors, and the quality, quantity, and nature of human-computer interaction [10]. Being resilient to manipulation by external parties it is also increasingly critical across a growing number of application scenarios [15].

These fundamental issues concern all aspects of IR system development and deployment. Given the current ubiquitous use of a variety of IR systems, from web search to recommendation platforms to personal assistants, they have potentially wide ranging impact—both positive and negative. We know that people are more likely to trust sources ranked higher in the search or recommendation results, but the recommendation or ranking criteria may rather optimize for user satisfaction, than for providing factual information [16]. For consequential user tasks, such as those related to medical, educational, or financial outcomes, this raises concerns about potential harms and what the right trade-offs might be.

Over the last years, a community has coalesced to address questions of fairness, accountability, transparency, ethics, and justice in machine learning and other computing systems; this workshop aims to give that discussion a home at SIGIR 2019 and provide an opportunity to highlight challenges specific to IR systems.

2 THEMES, PURPOSE, AND ORGANIZATION

The FACTS-IR workshop covers five key areas of focus, building on the responsible IR agenda articulated in the SWIRL report [1]:

SIGIR '19, July 21–25, 2019, Paris, France

ACM ISBN 978-1-4503-6172-9/19/07.
https://doi.org/10.1145/3331184.3331644

- **Fair IR**: the IR system should avoid discrimination across people and communities. To do so the notion of fairness should be contextual and well grounded in the application setup and domain. Achieving fairness may be further complicated by the multi-stakeholder nature of most IR systems.
- **Accountable IR**: the IR system should be able to justify its recommendations or actions to users and other stakeholders, as well as be reliable at all times. This requires an understanding of the potential harms of using the system and of who is more likely to be affected. It also requires recourse avenues and processes for redress.
- **Confidential IR**: the output or actions of the IR system should not reveal secrets. IR systems often combine extensive behavioral logs to model their users, which if not properly handled can result in unintended leakage of information.
- **Transparent IR**: the IR system should be able to explain to users and other interested stakeholders why and how the suggested results were obtained. Providing proper explanations may require answering who the users and the stakeholders are. More broadly, the IR systems should be able to enable third parties to monitor and probe that the systems behave as expected.
- **Safe IR**: The IR system should be resilient to manipulation by possible adversarial parties, and should not expose the users to undesirable, harmful content.

2.1 Presentations and Agenda Setting

This workshop goal aim is to both provide a venue for work-in-progress and identify gaps in the emerging technical work on responsible IR, including undertheorized and underspecified issues related to each of these five areas of focus and aiming to create actionable technical research agendas for each of them. We supported this with a two-part program consisting of presentations (of both workshop submissions and invited talks) and breakout group discussions, as follows:

- The first part featured presentations, based both on the accepted submissions by our PC members as well as a number of invited talks.
- In the second part, the participants were organized in working groups and were tasked with articulating a research agenda for one of the FACTS-IR topics.

Submissions and Invited Talks. We solicited submissions as both full (8-page) research papers and 2-4 page extended abstracts as position papers. The submissions primarily touched on two of the areas of interest (accountability & transparency) with a majority covering issues related to transparency in textual summarization, to deriving global explanations from local ones through their aggregation, to understanding and explaining predictions from tree-based boosting ensembles, and to making the user bias explicit in fact checking tasks. Other submissions discussed efforts to understand and define fairness metrics in IR systems, including tasks like ranking and recommendations, as well as applications to judicial systems.

Further, to cover the remaining of our focus areas, the workshop features additional short presentations by academic and industry practitioners, which are leaders in the FACTS research areas. The full workshop agenda is available at https://fate-events.github.io/facts-ir/. Accepted papers are not formally published, but are indexed at https://purl.org/mde/facts-ir-papers.

Working Groups and Agenda Setting. In the afternoon, the second part of the workshop includes focused discussions aimed at articulating new research agendas and identifying open problems in the FACTS-IR space, which we organized across a few breakout sessions and working groups. The goal of each breakout session and the working group was the formulation of in-depth, concrete research agendas for each of the five areas, as well as the identification of potential tensions between them. A detailed summary of the outcomes of the discussions in these breakout sessions will be made available after the workshop at https://purl.org/mde/facts-ir-report.

3 COMMITTEES

3.1 Organizing Committee

Alexandra Olteanu (http://www.aolteanu.com), part of the Fairness, Accountability, Transparency, and Ethics (FATE) Group at Microsoft Research Montréal & NYC. Prior to joining the FATE group, she was a Social Good Fellow at the IBM T.J. Watson Research Center, NY. Her work focuses on how data biases and methodological limitations delimit what we can learn from online social traces. She served on the program committees of social media and web conferences (e.g., ICWSM, WWW, WebSci, CIKM, SIGIR), on the steering committee of the ACM Conference on Fairness, Accountability, and Transparency (FAT*), and as the Tutorial Co-chair for ICWSM'18 and FAT*'19.

Jean Garcia-Gathright (https://www.scienceinthenoise.com) is a research scientist at Spotify, where she studies the evaluation of music retrieval and recommendation systems, and the technical and organizational challenges of algorithmic bias mitigation in industry. She served on the program committee for FairUMAP 2019, and as a co-organizer for the 2018 RecSys Challenge.

Maarten de Rijke (http://staff.fnwi.uva.nl/m.derijke) is a University Professor of Artificial Intelligence and Information Retrieval at the University of Amsterdam. He works on different types of technology that connect people to information, both its algorithmic underpinnings, its uses in domains ranging from news and retail to security and well-being, and its broader implications. Maarten is a member of the Royal Dutch Academy of Arts and Sciences (KNAW) and the founding director of the national Innovation Center for Artificial Intelligence. He has previously helped to organize various conferences (CLEF, ECIR, ICTIR, SIGIR, WSDM) and workshops (at CIKM, ECIR, SIGIR, WWW).

Michael D. Ekstrand (https://md.ekstrandom.net) is an Assistant Professor of Computer Science at Boise State University, where he co-directs the People & Information Research Team studying recommender systems and information retrieval from a human-centered perspective. His work on the social impact of recommender systems is funded by an NSF CAREER award (17-51278). He co-founded the FATREC workshop series; served as General Co-chair for RecSys 2018; is a member of the steering committee and senior program

committee for ACM RecSys; and serves the ACM FAT* community on its steering committee, program committee (2017–2018), and as FAT* Network Co-chair.

3.2 Program Committee

The workshop benefited by an extraordinary program committee, tasked with reviewing the workshop submissions, including:

- Ana-Andreea Stoica (Columbia University, US)
- Asia Biega (MPI, Germany)
- Ashudeep Singh (Cornell University, US)
- Avishek Anand (L3S, Germany)
- Carlos Castillo (UPF, Spain)
- Christo Wilson (Northeastern University, US)
- Damiano Spina (RMIT, Australia)
- Daniel Kluver (University of Minnesota, US)
- Diane Kelly (University of Tennessee, US)
- Dong Nguyen (Alan Turing Institute, UK)
- Emilia Gómez (UPF, Spain)
- Emre Kiciman (Microsoft Research, US)
- Faegheh Hasibi (Radboud University Nijmegen, The Netherlands)
- Fernando Diaz (Microsoft Research, Canada)
- Gianluca Demartini (University of Queensland, Australia)
- Hinda Haned (Ahold Delhaize & University of Amsterdam, The Netherlands)
- Ingmar Weber (QCRI, Qatar)
- James Thom (RMIT, Australia)
- Mark D. Smucker (University of Waterloo, Canada)
- Meike Zehlike (TU Berlin, Germany)
- Min Zhang (Tsinghua University, China)
- Pierre-Nicolas Schwab (Solvay Brussels School of Economics and Management, Belgium)
- Rishabh Mehrotra (Spotify, UK)
- Ronald Robertson (Northeastern University, US)
- Solon Barocas (Cornell University & Microsoft Research, US)
- Stefano Balietti (Microsoft Research, US)
- Suzan Verberne (Leiden University, The Netherlands)
- Toshihiro Kamishima (National Institute of Advanced Industrial Science and Technology, Japan)

4 FINAL THOUGHTS

The SIGIR community has the responsibility to care about the broader impact and implications of the systems that we research and the systems that we build in academia and industry. Similar responsibility issues are also being addressed in related fields, with, for instance, the emergence of the community around the ACM Conference on Fairness, Accountability, and Transparency (see https://fatconference.org/), a venue with a cross-disciplinary focus that brings together a diversity of researchers and practitioners interested in fairness, accountability, and transparency in socio-technical systems.

However, there are specific issues in IR stemming from the characteristics of and the reliance on document collections, and the often imprecise nature of search and recommendation tasks. IR has a strong history of using test collections during evaluation. As evaluation tools, test collections also have certain types of bias built-in. For example, the people who construct topics and make relevance assessments arguably are not representative of the larger population. In some cases, they have not been representative of the type of users who are being modeled (e.g., having people who do not read blogs evaluate blogs). Evaluation measures are designed to optimize certain performance criteria and not others, and either implicitly or explicitly have built-in user models. Systems are then tested and tuned within this evaluation framework, further reinforcing and rectifying any existing biases [1]. Safety and privacy issues are also prevalent within most IR applications, as they tend to record vast information about their users and are sometimes prone to manipulation for business or political purposes.

Given the central role that IR technology plays in today's society, it is critical to continue to build a community of researchers and practitioners to characterize and address FACTS-related issues. The agenda setting activities of this workshop were meant to do just that.

REFERENCES

[1] James Allan, Jaime Arguello, Leif Azzopardi, Peter Bailey, Tim Baldwin, Krisztian Balog, Hannah Bast, Nick Belkin, Klaus Berberich, Bodo von Billerbeck, Jamie Callan, Rob Capra, Mark Carman, Ben Carterette, Charles L. A. Clarke, Kevyn Collins-Thompson, Nick Craswell, W. Bruce Croft, J. Shane Culpepper, Jeff Dalton, Gianluca Demartini, Fernado Diaz, Laura Dietz, Susan Dumais, Carsten Eickhoff, Nicola Ferro, Norbert Fuhr, Shlomo Geva, Claudia Hauff, David Hawking, Hideo Joho, Gareth Jones, Jaap Kamps, Noriko Kando, Diane Kelly, Jaewon Kim, Julia Kiseleva, Yiqun Liu, Xiaolu Lu, Stefano Mizzaro, Alistair Moffat, Jian-Yun Nie, Alexandra Olteanu, Iadh Ounis, Filip Radlinski, Maarten de Rijke, Mark Sanderson, Falk Scholer, Laurianne Sitbon, Mark Smucker, Ian Soboroff, Damiano Spina, Torsten Suel, James Thom, Paul Thomas, Andrew Trotman, Ellen Voorhees, Arjen P. de Vries, Emine Yilmaz, and Guido Zuccon. 2018. Report from the Third Strategic Workshop on Information Retrieval in Lorne (SWIRL 2018). *SIGIR Forum* 52 (June 2018), 34–90.

[2] Ricardo Baeza-Yates. 2018. Bias on the web. *Commun. ACM* 61, 6 (2018).

[3] Solon Barocas and Andrew D Selbst. 2016. Big data's disparate impact. *Cal. L. Rev.* 104 (2016), 671.

[4] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of Attention: Amortizing Individual Fairness in Rankings. In *SIGIR*. ACM, 405–414.

[5] EU. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* L119 (2016), 1–88.

[6] Eric Goldman. 2005. Search Engine Bias and the Demise of Search Engine Utopianism. *Yale JL & Tech.* 8 (2005), 188.

[7] Ben Hutchinson and Margaret Mitchell. 2019. 50 Years of Test (Un) fairness: Lessons for Machine Learning. In *Proc. of FAT**.

[8] Lucas D Introna and Helen Nissenbaum. 2000. Shaping the Web: Why the politics of search engines matters. *The Information Society* 16, 3 (2000), 169–185.

[9] Rishabh Mehrotra, Amit Sharma, Ashton Anderson, Fernando Diaz, Hanna Wallach, and Emine Yilmaz. 2017. Auditing Search Engines for Differential Satisfaction Across Demographics. In *Proc. of WWW*.

[10] Tim Miller, Piers Howe, and Liz Sonenberg. 2017. Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences. *arXiv preprint arXiv:1712.00547* (2017).

[11] Alexandra Olteanu, Carlos Castillo, Fernando Diaz, and Emre Kiciman. 2016. Social data: Biases, methodological pitfalls, and ethical boundaries. *SSRN* (2016). https://ssrn.com/abstract=2886526.

[12] Boris Sharchilev, Yury Ustinovsky, Pavel Serdyukov, and Maarten de Rijke. 2018. Finding influential training samples for gradient boosted decision trees. In *ICML*.

[13] Amanda Spink and Michael Zimmer. 2008. *Web search: Multidisciplinary perspectives*. Vol. 14. Springer Science & Business Media.

[14] Ana-Andreea Stoica, Christopher Riederer, and Augustin Chaintreau. 2018. Algorithmic Glass Ceiling in Social Networks: The effects of social recommendations on network diversity. In *WWW*. 923–932.

[15] Peng Wang, Xianghang Mi, Xiaojing Liao, XiaoFeng Wang, Kan Yuan, Feng Qian, and Raheem Beyah. 2018. Game of Missuggestions: Semantic Analysis of Search-Autocomplete Manipulations. In *Proc. of NDSS*.

[16] Ryen White. 2013. Beliefs and biases in web search. In *Proc. of SIGIR*. 3–12.

[17] Hui Yang, Ian Soboroff, Li Xiong, Charles L.A. Clarke, and Simson L. Garfinkel. 2016. Privacy-Preserving IR 2016: Differential Privacy, Search, and Social Media. In *SIGIR*. ACM, 1247–1248.