BIRDS 2020 – Bridging the Gap between Information Science, Information Retrieval and Data Science

Ingo Frommholz[1 [0000−0002−5622−5132]], Haiming Liu[1 [0000−0002−0390−3657]], and Massimo Melucci[2]

[1] University of Bedfordshire, Luton, UK
[2] University of Padua, Italy

1 Introduction

The BIRDS 2020 workshop was a virtual event at SIGIR 2020 as an interdisciplinary workshop for students, practitioners and researchers in Data Science, Information Retrieval and Information Science. BIRDS aimed to foster the cross-fertilization of Information Science (IS), Information Retrieval (IR) and Data Science (DS). Information Science (IS) and Data Science (DS) represent two pillars of a wide range of theories, models and methods for information and data processing and management. Roughly speaking, in the spectrum of Data – Information –Knowledge, DS is 'data-driven' while IS is 'user-driven' and mainly concerned with the user's needs to handle information and acquire knowledge to satisfy a certain task, as it is illustrated in Figure 1. IR, naturally and historically concerned with both the system and user side in the world of heterogeneous big data, can be regarded as a kind of bridge. Based on these considerations, the overarching theme of the BIRDS workshop was to look at how IR, DS and IS can complement each other by applying a more holistic approach to these disciplines that go beyond traditional IR or DS or IS alone.

Due to the COVID-19 situation BIRDS was held as an online event with 2 keynotes, 2 invited talks and several long, short and position papers that were selected after a peer-reviewing process. By offering two main blocks we tried to accommodate different timezones. Further information can be found on the workshop page at `https://birds-ws.github.io/birds2020/index.html` .

2 Papers

2.1 Keynotes

Carlos Castillo presented the first keynote on fairness and transparency in ranking. He first asked whether algorithms can discriminate and looks at different

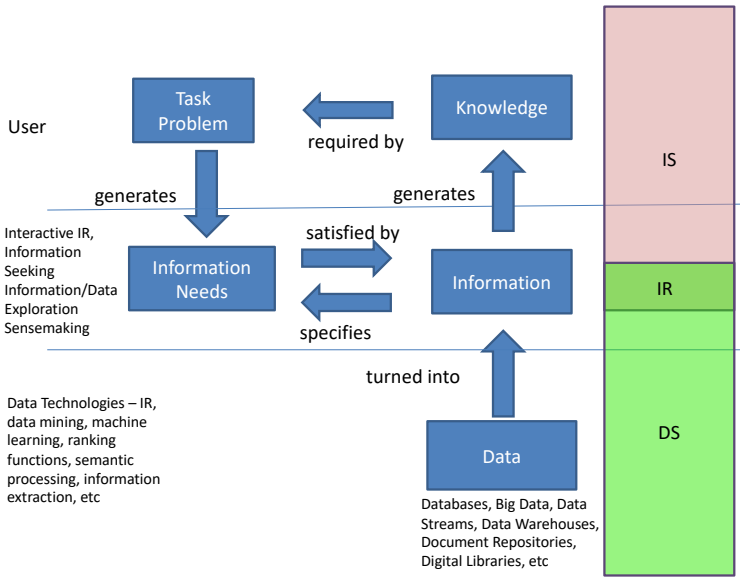

Fig. 1. The (simplified) BIRDS view on Data Science, Information Retrieval and Information Science

forms of biases, discrimination and fairness for searchers and those searched. He then discussed how we can measure fairness in rankings before looking at how we can create fairer rankings and improve ranking transparency.

The second keynote was by Nick Belkin on "Challenges and Opportunities for IS, IR DS in an Era of Information Ubiquity". He remarked that while IS and IR have a long history together, there seems to be less interaction between DS on the one hand and IR and IS, respectively, on the other hand. Apart from DS, IR and IS, Nick brought another important player into the game, Human-Computer Interaction (HCI), asking how IS, IR, DS and HCI can support what he called *Radical Personalisation*.

2.2 Invited Talks

Besides keynotes, BIRDS 2020 had two invited talks. Riccardo Guidotti discussed the lack of transparency in AI and Machine Learning systems and gave an overview of research in eXplainable AI (XAI). Xi (Sunshine) Niu introduced faceted search as an example where IS, IR and DS complement each other.

2.3 Research papers

The following research papers (full, short and position papers) were presented in two block sessions.

In the first session, Amit Kumar Jaiswal, Haiming Liu and Ingo Frommholz discussed how reinforcement learning and the formalism of quantum probabilities can be used to model information seeking based on Information Foraging. Steven Zimmerman, Stefan Herzog, Jon Chamberlain, David Elsweiler and Udo Kruschwitz presented their ideas of a framework for harm prevention in Web search. Kritika Agrawal and Vikram Pudi looked at how to find grand challenges and saturated problems in the scientific literature. The last presentation of the first session was given by Sehrish Sher Khan and Haiming Liu who explored the impact of user information search behaviour by Affective Design.

In the second session, Hong Qing Yu discussed his approach for extracting causal knowledge from UK health web sites to create an AI-enabled healthcare system. Tuomas Ketola and Thomas Roelleke extended the well-known BM25 formula and proposed BM25-FIC as an enhanced BM25F method that combines information-oriented search and parameter estimation. Mahmoud Artemi and Haiming Liu discussed a new CBIR system design based on Vakkari's three-stage model to capture user's feedback at the query formulation stage for content-based image retrieval. In the final presentation, Massimo Melucci looked at Structural Equation Modelling as a methodology to investigate the causal relationships underlying search engines and recommender systems, for instance, to understand when the system produced biased results.

Above sessions were followed by a closing discussion about the overall interdisciplinary topic of BIRDS.

3 Acknowledgement

BIRDS 2020 has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 721321 (QUARTZ).