

# Report on the 2nd Workshop on Bridging the Gap between Information Science, Information Retrieval and Data Science (BIRDS 2021)

Ingo Frommholz

University of Wolverhampton, UK  
*ifrommholz@acm.org*

Haiming Liu

University of Bedfordshire, Luton, UK  
*haiming.liu@beds.ac.uk*

Massimo Melucci

University of Padua, Italy  
*massimo.melucci@unipd.it*

## Abstract

The aim of the BIRDS workshop (Bridging the Gap between Information Science, Information Retrieval and Data Science) is to bring together the Data Science, Information Retrieval, Information Science and HCI communities. BIRDS 2021 is the second in a series of workshops and was held in conjunction with CHIIR 2021, following a successful event at SIGIR 2020. It consisted of a selection of accepted papers and invited talks. This article reports on BIRDS 2021, provides a discussion of topics, tasks, approaches and data sets and an outline of future directions for interdisciplinary work.

## 1 Introduction

Can Data Science, Information Science, Information Retrieval and Human-Computer Interaction get together and learn from each other? While research is often conducted in silos, i.e. within a research community, the aim of the BIRDS (Bridging the Gap between Information Science, Information Retrieval and Data Science) workshop is to bring together these different communities. The idea was born from the observations the workshop organisers made as part of the Information Retrieval (IR) community, where over the last decades user- and system-oriented approaches started to meld [Ingwersen and Järvelin, 2005]. With the emergence of more data-driven methods, in particular, in the era of deep and machine learning with all its potential biases and the need for transparency, as well as the data scientists' aim to explore, find, combine and make sense of all sorts of heterogeneous internal and external data (be it textual or multimedia, unstructured data, data streams or structured database entries), one idea is to broaden the scope of classical IR and its user- and system-oriented methods, often rooted in Information Science and Human-Computer Interaction, to broader Data Science concepts.

BIRDS 2021 was held in conjunction with CHIIR and attracted an estimated number of around 30 delegates, attending on and off, from different time zones. Like many other conferences and

---

workshops these days during the pandemic, it was held online only using Zoom. More information can be found on the BIRDS 2021 Web site<sup>1</sup>. Selected presentations were recorded and can be viewed on the BIRDS 2021 YouTube playlist<sup>2</sup>.

## 2 Accepted Papers and Invited Talks

BIRDS 2021 at CHIIR is the second in a series of workshops, after BIRDS 2020 at SIGIR [Frommholz et al., 2020]. It consisted of invited talks as well as peer-reviewed, accepted papers [Frommholz et al., 2021].

The following six papers were accepted:

- Ghadeer Abuoda, Chad Hendrix and Stuart Campo: *Automatic Tag Recommendation for the UN Humanitarian Data Exchange* [Abuoda et al., 2021];
- Nicholas Collis and Ingo Frommholz: *AQUACOLD: A Novel Crowdsourced Linked Data Question Answering System* [Collis and Frommholz, 2021];
- Mahmoud Artemi and Haiming Liu: *A User Study on User’s Attention for an Interactive Content-based Image Search System* [Artemi and Liu, 2021];
- Morshed Adnan, Michael Alexander Kaufmann and Matthias Hemmje: *Social Media Mining to Study Social User Groups by Visualizing Tweet Clusters using Word2Vec, PCA and K-Means* [Adnan et al., 2021];
- Stefan Wagenpfeil et al.: *Query Construction and Result Representation based on Graph Codes* [Wagenpfeil et al., 2021]; and
- Thoralf Reis et al.: *Towards Modeling AI-based User Empowerment for Visual Big Data Analysis* [Reis et al., 2021].

The invited talks were as follows:

- *User Discovery and Exploration in Future Digital Libraries* by Ed Fox, Virginia Tech, USA;
- *Searching, fast and slow* by Tony Russell-Rose, 2Dsearch and Goldsmiths, University of London, UK;
- *Data Science and Information Access for Social Research on Technoscientific Issues in the Media* by Emanuele Di Buccio, University of Padua, Italy;
- *Understanding and solving the complex IIR challenges of searching enterprise content* by Martin White, Intranet Focus Ltd and University of Sheffield, UK;
- *Exploiting clinical data to build patients trajectories* by Lorraine Goeuriot, Univ. Grenoble Alpes, France;

---

<sup>1</sup><https://birds-ws.github.io/birds2021/index.html>

<sup>2</sup>[https://www.youtube.com/playlist?list=PLI13W1gRqqf1v\\_gEucOYT049daq16IaQ2](https://www.youtube.com/playlist?list=PLI13W1gRqqf1v_gEucOYT049daq16IaQ2)

- 
- *Querying by Example Using Bootstrapped Explainable Text Categorization in Emergent Knowledge-Domains* by Tobias Eljasik-Swoboda, Fernuniversität Hagen, Germany, and
  - *Design of use case diagrams, personas and GUIs based on the CRISP4BigData process / Conceptual Design and Implementation of a graphical user interface for CRISP4BigData* by Kevin Berwind, Fernuniversität Hagen, Germany.

## 3 Discussion

### 3.1 Summary and Reflection of Topics, Data, Tasks and Approaches

The invited talks and papers that were presented in the workshop addressed topics related to methods and systems used by the end-users to access and explore complex *data* such as those within enterprises, social, humanitarian, libraries, news and health. The various supported *tasks* consisted of tag recommendation for data set search, question answering, image and multimedia retrieval, social media mining, visual data analysis, enterprise and professional search, data discovery and exploration, making use of medical data, content management and supporting research activities in social sciences. Different *approaches* were discussed to support the respective task at hand on the given data. [Abuoda et al. \[2021\]](#) utilise tf-idf, WordNet and cosine similarity in their tag recommendation pipeline. Different query construction techniques are explored, e.g., graph algorithms [[Wagenpfeil et al., 2021](#)] or crowdsourcing [[Collis and Frommholz, 2021](#)] in conjunction with SPARQL queries. [Artemi and Liu \[2021\]](#) study user engagement and attention in interactive image search to improve interactive image search interfaces. Clustering as an unsupervised machine learning technique as well as word embeddings are utilised to create explainable and semantically labelled tweet clusters and their visualization [[Adnan et al., 2021](#)]. Visual big data analysis systems should empower users by supporting their intellectual capabilities. To this end, [Reis et al. \[2021\]](#) suggest a system architecture for user-centred visual data analysis and AI-based user empowerment.

As we see from the above summary, the BIRDS audience was exposed to a diverse number of different topics, tasks, data sets and approaches, which is one of the goals of BIRDS. As a CHIIR workshop, we hope delegates, from which many are likely rooted in user-centred information retrieval and HCI, were able to learn about the “data science” side of things, the different forms and flavours of heterogeneous data going beyond the traditional notion of ‘documents’, the different information needs arising and the techniques that support data exploration in such an environment. Many of our invited speakers are well established in the Information Retrieval community but extended the horizon beyond classical IR problems, by focusing on discovery and exploration in data-rich environments like digital libraries (Fox), on specific kinds of interactive search such as professional and enterprise search (Russell-Rose and White), supporting data analysis in specific communities (Di Buccio), exploiting different forms of clinical data in the medical domain (Goeriot), applying explainable text categorization in different scenarios (Eljasik-Swoboda) and designing user-oriented systems and user interfaces for data analysis (Berwind). We think these are nice examples of interdisciplinary approaches that take findings from the four target communities, IR, IS, DS and HCI, into account.

---

It should be noted that as diverse as the different techniques and topics presented at BIRDS were, they all share a common goal, to provide better data and information systems for exploration and information-seeking tasks. Some themes, such as the need for explainable AI, are ubiquitous across all fields to provide human-friendly technological tools to users.

### 3.2 Takeaway Messages

As mentioned, the BIRDS audience was exposed to research and results coming from related but different disciplines to the one of its host conference, CHIIR. The talks also made clear that researchers target similar problems, but of course approaching them from a different angle, given their background. However, the problems that were discussed at the BIRDS workshops would benefit from a more holistic viewpoint, either by means of collaboration or through specifically trained researchers that feel at home in different disciplines. In a way similar to the development within Information Retrieval, where the ‘system-oriented’ and the ‘user-oriented’ fractions, which are rooted in Computer Science and Information Science, respectively, are getting closer together, we hope and think to have made a contribution to a similar movement on a cross-disciplinary scale. In fact, the idea of bringing different communities together for a common goal (providing user-friendly data and information systems) is not new and has been the fabric for instance of the Digital Libraries community for many years. The research reported at BIRDS 2021 as well as that reported at the previous edition of the workshop shows that the researchers working in one area, e.g. IR, IS, HCI or DS, might be aware of the respective other areas but are lacking their specialist knowledge. For example, a researcher who is oriented to work with users may know that some user survey data analysis would be necessary, but she likely ignores the details of the whole process of data analysis, which is part of the expertise of a data scientist. Similarly, an IR scientist with a strong background in data analysis which is for instance deployment in evaluation might ignore the subtleties of HCI for which a specific preparation would be advisable.

### 3.3 Future Directions

One direction would be towards greater integration of different expertise at PhD level; for example, a doctoral student who is oriented to work with users should be prepared to deal with the details of the whole process of data analysis; vice versa, an IR student working with evaluation data should be prepared in user cognition and interaction to a larger extent than currently observed.

We also think more work needs to be done to ensure a deeper integration of successful approaches from the different communities. An example might be to introduce the notion of ‘best match’ and rankings, inherent in IR, to data exploration that is more based on classical database ‘exact match’ techniques. How to handle heterogeneous big data, not just in the form of textual or multimedia documents but also other forms such as data streams, and apply ideas such as Information Foraging Theory to guide users in their exploration of such data sets, seems to be uncharted territory.

As the breadth of topics, data, tasks and approaches discussed in the BIRDS workshops shows, there is a huge number of potential interdisciplinary research ideas waiting to be explored. The above examples are therefore in no way meant to be complete or exhaustive. One aim of the BIRDS workshop series is to inspire our audience to look beyond the horizon of their own field

---

and bridge the gap between disciplines that, in the end, share very similar goals on different but not necessarily independent levels.

## 4 Conclusion

In this report, we reported on the BIRDS workshop at CHIIR 2021, whose aim is to bring the different but related communities IS, IR, DS and HCI together. We presented invited talks and accepted papers and discussed their main contribution. We saw a wide range of topics, data sets, tasks and approaches, to foster exploration and information seeking in different settings. Some takeaway messages and future directions were discussed. We think BIRDS 2021 was a success as an event that attempts to be interdisciplinary, but it also shows the long way ahead of us to create true interdisciplinary solutions for the users' benefit. The two BIRDS workshops at SIGIR 2020 and CHIIR 2021 can only be just the beginning of a hopefully wider cross-fertilization of the communities under consideration.

## References

- Ghadeer Abuoda, Chad Hendrix, and Stuart Campo. Automatic tag recommendation for the un humanitarian data exchange. In [Frommholz et al. \[2021\]](#), pages 4–10. URL <http://ceur-ws.org/Vol-2863/#paper-02>.
- Md Morshed Jaman Adnan, Matthias L. Hemmje, and Michael A. Kaufmann. Social media mining to study social user group by visualizing tweet clusters using word2vec, pca and k-means. In [Frommholz et al. \[2021\]](#), pages 40–51. URL <http://ceur-ws.org/Vol-2863/#paper-05>.
- Mahmoud Artemi and Haiming Liu. A user study on user attention for an interactive content-based image search system. In [Frommholz et al. \[2021\]](#), pages 26–39. URL <http://ceur-ws.org/Vol-2863/#paper-04>.
- Nicholas W. Collis and Ingo Frommholz. AQUACOLD: A novel crowdsourced linked data question answering system. In [Frommholz et al. \[2021\]](#), pages 11–25. URL <http://ceur-ws.org/Vol-2863/#paper-03>.
- Ingo Frommholz, Haiming Liu, and Massimo Melucci, editors. *Proceedings of the First Workshop on Bridging the Gap between Information Science, Information Retrieval and Data Science (BIRDS 2020) (BIRDS)*, number 2741 in CEUR Workshop Proceedings, Aachen, 2020. URL <http://ceur-ws.org/Vol-2741/>.
- Ingo Frommholz, Haiming Liu, Massimo Melucci, Nicholas J. Belkin, Gareth J. F. Jones, Noriko Kando, and Gabriella Pasi, editors. *Proceedings of the Second Workshop on Bridging the Gap between Information Science, Information Retrieval and Data Science, and Third Workshop on Evaluation of Personalisation in Information Retrieval (BIRDS/WEPIR)*, number 2863 in CEUR Workshop Proceedings, Aachen, 2021. URL <http://ceur-ws.org/Vol-2863/>.

---

Peter Ingwersen and Kalvero Järvelin. *The turn: integration of information seeking and retrieval in context*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005. ISBN 140203850X.

Thoralf Reis, Sebastian Bruchhaus, Binh Vu, Marco X. Bornschlegl, and Matthias L. Hemmje. Towards modeling ai-based user empowerment for visual big data analysis. In [Frommholz et al. \[2021\]](#), pages 67–75. URL <http://ceur-ws.org/Vol-2863/#paper-07>.

Stefan Wagenpfeil, Felix Engel, Paul McKeivitt, and Matthias Hemmje. Query construction and result presentation based on graph codes. In [Frommholz et al. \[2021\]](#), pages 52–66. URL <http://ceur-ws.org/Vol-2863/#paper-06>.