

# Preface: 4th Joint Workshop on BIRNDL at SIGIR 2019

Muthu Kumar Chandrasekaran<sup>1</sup> and Philipp Mayr<sup>2</sup>

<sup>1</sup> SRI International, Menlo Park, CA, USA,  
cmkumar087@gmail.com

<sup>2</sup> GESIS – Leibniz-Institute for the Social Sciences, Cologne, Germany,  
philipp.mayr@gesis.org

## 1 Introduction

The deluge of scholarly publication poses a challenge for scholars to find relevant research and policy makers to seek in-depth information and understand research impact. Information retrieval (IR), natural language processing (NLP) and bibliometrics could enhance scholarly search, retrieval and user experience, but their use in digital libraries is not widespread. To address this gap, we organise the 4<sup>th</sup> Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL)<sup>3</sup> and the 5<sup>th</sup> CL-SciSumm Shared Task<sup>4</sup> co-located with the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019).

Over the past several years and at premier conferences, BIRNDL [1–3], together with its parent workshops, has established itself as the primary interdisciplinary venue for cross-pollination of NLP, IR and DL research.

## 2 Overview of the papers

This year, 11 full papers were submitted to the workshop, 5 of which were accepted as full papers for presentation and inclusion in the proceedings. 13 short papers were submitted to the workshop, 4 of which were accepted as short papers for presentation and inclusion in the proceedings. In addition, 5 poster papers for poster presentation were also accepted.

The CL-SciSumm Shared Task had 17 registrations of which 9 teams submitted their systems for evaluation. The results and analysis of the evaluation along with the system summaries are presented in the overview paper titled: “Overview and Results: CL-SciSumm Shared Task 2019” and included in the proceedings. The system description papers are included in the proceedings.

The workshop featured two keynote talks, one invited talk and various full and short paper sessions. The following section briefly lists the keynotes and sessions.

<sup>3</sup> <http://wing.comp.nus.edu.sg/~birndl2019/>

<sup>4</sup> <http://wing.comp.nus.edu.sg/~cl-scisumm2019/>

## 2.1 Keynotes

- Alex D. Wade, Ivana Williams: *Personalized Feed/Query-formulation, Predictive Impact, and Ranking*.
- Bonnie Webber: *Discourse Processing for Text Analysis: Recent successes, current challenges*.

## 2.2 Invited Paper

- Guy Lev, Michal Shmueli-Scheuer, Jonathan Herzig, Achiya Jerbi, David Konopnicki: *TalkSumm: A Dataset and Scalable Annotation Method for Scientific Paper Summarization Based on Conference Talks* [4].  
This paper will appear in *Proceedings of ACL 2019* and is not included in this proceedings.

## 2.3 Research papers

### Full papers

- Katarina Boland, Frank Krüger: *Distant supervision for silver label generation of software mentions in social scientific publications*
- Na Pang, Li Qian, Weimin Lyu, Jin-Dong Yang: *Transfer Learning for Scientific Data Chain Extraction in Small Chemical Corpus with joint BERT-CRF Model*
- Christin Katharina Kreutz, Premtim Sahitaj, Ralf Schenkel: *Revaluating Semantometrics from Computer Science Publications*
- Suzan Verberne, Ioannis Chios, Jian Wang: *Extracting and matching patent in-text references to scientific publications*
- Michael Soprano, Kevin Roitero, Stefano Mizzaro: *HITS Hits Readersourcing: Validating Peer Review Alternatives Using Network Analysis*

### Short papers

- Jason Portenoy, Jevin D. West: *Supervised Learning for Automated Literature Review*
- Arlene J Casey, Bonnie Webber, Dorota Glowacka: *Can Models of Author Intention Support Quality Assessment of Content?*
- Andres Carvallo, Denis Parra: *Comparing Word Embeddings for Document Screening based on Active Learning*
- Philipp Scharpf, Moritz Schubotz, Howard S. Cohl, Bela Gipp: *Towards Formula Concept Discovery and Recognition*

## Poster papers

- Barbara Plank, Reinard van Dalen: *CiteTracked: A Longitudinal Dataset of Peer Reviews and Citations*
- Chifumi Nishioka, Michael Färber: *Evaluating the Availability of Open Citation Data*
- André Greiner-Petter, Terry Ruas, Moritz Schubotz, Akiko Aizawa, William Grosky, Bela Gipp: *Why Machines Cannot Learn Mathematics, Yet*
- Marc Bertin, Pierre Jonin, Frédéric Armetta, Iana Atanassova: *Identifying the conceptual space of citation contexts using coreferences*
- Rajesh Piriyani, Wolfgang Otto, Philipp Mayr, Vivek Kumar Singh: *Analysing author name mentions in citation contexts of highly cited publications*

## 2.4 CL-SciSumm

- Chrysoula Zerva, Minh-Quoc Nghiem, Nhung Nguyen and Sophia Ananiadou: *NaCTeM-UoM @ CL-SciSumm-19*
- Shutian Ma, Heng Zhang, Tianxiang Xu, Jin Xu, Shaohu Hu and Chengzhi Zhang: *IR&TM-NJUST @ CLSciSumm-19*
- Lei Li, Yingqi Zhu, Wei Liu, Zuying Huang, Yang Xie, Yinan Liu and Xingyuan Li: *CIST@CLSciSumm-19: Automatic Scientific Paper Summarization With Citances and Facets*
- Yoann Pitarch, Karen Pinel-Sauvagnat, Gilles Hubert, Guillaume Cabanac and Ophélie Fraissier-Vannier: *IRIT-IRIS @ CL-SciSumm 2019: Matching Citances with their Intended Reference Text Spans from the Scientific Literature*
- Bakhtiyar Syed, Vijayasradhi Indurthi, Balaji Vasana Srinivasan and Vasudeva Varma: *Helium @ CL-SciSumm-19: Transfer learning for effective scientific research comprehension*
- Ahmed Ghassan Tawfiq Abura'Ed, Àlex Bravo, Luis Chiruzzo and Horacio Saggion: *LaSTUS/TALN+INCO @ CL-SciSumm 2019*
- Moreno La Quatra, Luca Cagliero and Elena Baralis: *Poli2Sum@CL-SciSumm 2019: identify, classify, and summarize cited text spans by means of ensembles of supervised models*
- Hyonil Kim and Shiyun Ou: *NJU@CL-SciSumm-19: Ranking-based Identification of Cited Text with Deep Learning*
- Aris Fergadis, Dimitris Pappas and Haris Papageorgiou: *ATHENA@CL-SciSumm 2019: Siamese recurrent bi-directional neural network for identifying cited text spans*

## 3 Outlook and further reading

We will continue to organize the BIRNDL and BIR workshops at high profile IR, DL, Scientometric, NLP and CL venues. The combination of research paper presentations, and a shared task like CL-SciSumm with system evaluation has proven to be a successful and agile format. So we will continue with this format.

In line with the BIRNDL 2019 workshop, we invite workshop authors and others to submit extended paper to two independent open calls of special issues.

1. Special issue on “Mining Knowledge from Scientific Data”<sup>5</sup> in the journal *Expert Systems* (Wiley). Guest editors are: Tanmoy Chakraborty, Sumit Bhatia and Cornelia Caragea. Deadline: September 30, 2019
2. Special issue on “Bibliometrics and Information Retrieval”<sup>6</sup> in the journal *Scientometrics* (Springer). Guest editors are: Guillaume Cabanac, Philipp Mayr and Ingo Frommholz. Deadline: September 30, 2019

**Further reading:** A special issue on “Bibliometrics, Information Retrieval and Natural Language Processing in Digital Libraries” appeared in 2018 in the *International Journal on Digital Libraries*, see an overview in [5]. This was a large issue with 13 papers with papers from BIRNDL 2016 as well as new submissions. So, we followed up with another special issue on “Bibliometric-enhanced Information Retrieval and Scientometrics” appeared in *Scientometrics* journal with extended papers from later editions BIR and BIRNDL and new submissions. Please see an overview in [6].

Since 2016 we maintain the “Bibliometric-enhanced-IR Bibliography”<sup>7</sup> that collects scientific papers which appear in collaboration with the BIR/BIRNDL organizers. We invite interested researchers to join this project and contribute related publications.

## Acknowledgments

We would like to thank SRI International for their generous funding of CL-SciSumm '19 and BIRNDL '19. CZI sponsored Alex Wade’s keynote while SRI International sponsored Bonnie Webber. We immensely thank Prof. Dragomir Radev and Michihiro Yasunaga from Yale University for sharing the SciSumm-Net dataset for CL-SciSumm 2019 and co-organising this time. The work by Philipp Mayr was partly funded by Deutsche Forschungsgemeinschaft (DFG) under grant number MA 3964/10-1, the “Establishing Contextual Dataset Retrieval - transferring concepts from document to dataset retrieval (ConDATA)” project.

We are also grateful to the co-organizers of the past BIRNDL workshops - Guillaume Cabanac, Ingo Frommholz, Kokil Jaidka and Dietmar Wolfram, for their continued support and involvement.

## References

1. Mayr, P., Chandrasekaran, M.K., Jaidka, K.: Report on the 3rd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2018). *SIGIR Forum* **52**(2) (2018) 105–110

<sup>5</sup> [http://lcs2.iiitd.edu.in/speIssue\\_Exsy.html](http://lcs2.iiitd.edu.in/speIssue_Exsy.html)

<sup>6</sup> <https://sites.google.com/view/scientometrics-si2019-bir>

<sup>7</sup> [https://github.com/PhilippMayr/Bibliometric-enhanced-IR\\_Bibliography/](https://github.com/PhilippMayr/Bibliometric-enhanced-IR_Bibliography/)

2. Mayr, P., Chandrasekaran, M.K., Jaidka, K.: Report on the Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2017). *SIGIR Forum* **51**(2) (2017) 107–113
3. Cabanac, G., Chandrasekaran, M.K., Frommholz, I., Jaidka, K., Kan, M.Y., Mayr, P., Wolfram, D.: Report on the Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2016). *SIGIR Forum* **50**(2) (2016) 36–43
4. Lev, G., Shmueli-Scheuer, M., Herzig, J., Jerbi, A., Konopnicki, D.: TalkSumm: A Dataset and Scalable Annotation Method for Scientific Paper Summarization Based on Conference Talks. In: arXiv:1906.01351
5. Mayr, P., Frommholz, I., Cabanac, G., Chandrasekaran, M.K., Jaidka, K., Kan, M.Y., Wolfram, D.: Introduction to the Special Issue on Bibliometric-Enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL). *International Journal on Digital Libraries* **19**(2-3) (2018) 107–111
6. Cabanac, G., Frommholz, I., Mayr, P.: Bibliometric-enhanced information retrieval: preface. *Scientometrics* **116**(2) (2018) 1225–1227