

7th International Workshop on Bibliometric-enhanced Information Retrieval (BIR 2018)



Workshop at ECIR 2018, 26 March, 2018

You are invited to participate in the upcoming 7th workshop on **Bibliometric-enhanced Information Retrieval (BIR)**, to be held as part of the 40th European Conference on Information Retrieval (ECIR). <https://www.ecir2018.org/>

Updates

- Workshop Proceedings are available <http://ceur-ws.org/Vol-2080/>
- Keynotes will be given by [Cyril Labbé](#) (Université Grenoble Alpes) and [Ralf Schenkel](#) (University of Trier).

Keynote title of *Cyril Labbé*: "Trends in Gaming Indicators: On Failed Attempts at Deception and their Computerised Detection"

Abstract: Counting articles and citations, analyzing citations and co-authors graphs have become ways to assess researchers and institutions performance. Fairly enough, these measures are becoming targets for institutions and individual researchers thus triggering new behaviors. As a matter of fact, scientometrics and informetrics systems of all kinds have to separate the grain from the chaff. Among others, fields like Information Retrieval, network analysis and natural language processing may offers answers to deal with this kind of problems. Through several emblematic case studies (fake researcher, generated papers, paper mills), we show evidences of attempts to game indicators together with automatic ways to detect them (automatic detection of generated papers, errors detection).



Keynote title of *Ralf Schenkel*: "Integrating and exploiting public metadata sources in a bibliographic information system"

Abstract: Bibliographic information systems need to rely on metadata provided by various sources in various forms and with various quality. The talk gives some insights how the dblp bibliography as an example for such a system is maintained and improved. It shows how metadata can be automatically harvested from publisher websites and how the harvesting process can be steered. It also discusses some open sources of bibliographic metadata and how they can be used to enrich existing bibliographic data, but also how varying their quality is.



Program

Time	Speaker	Title
9:30	Introduction by organizers	Editorial Slides (2,24 MB)
9:40	Cyril Labbé (Keynote)	Trends in Gaming Indicators: On Failed Attempts at Deception and their Computerised Detection Paper Slides (4,61 MB)
10:30	Maria Janina Sarol, Linxi Liu and Jodi Schneider	Testing a Citation and Text-Based Framework for Retrieving Publications for Literature Reviews Paper Slides (4,54 MB)
11:00	Coffee break	
11:30	Anaïs Ollagnier, Sébastien Fournier and Patrice Bellot	BIBLME RecSys: Harnessing Bibliometric Measures for Scholarly Paper Recommender System Paper Slides (2,44 MB)
12:00	André Rattinger, Jean-Marie Le Goff and Christian Guetl	Local Word Embeddings for Query Expansion based on Co-Authorship and Citations Paper Slides (351 kB)
12:30	Mario Cataldi, Luigi Di Caro and Claudio Schifanella	All for One or One for All? Analyzing Collaboration Patterns in Research Environments (Demo) Paper Slides (734 kB)
		Demo-Session 1
13:00-14:30	Lunch break	

Time	Speaker	Title
14:30	Ralf Schenkel (Keynote)	Integrating and exploiting public metadata sources in a bibliographic information system Paper Slides (3,98 MB)
15:30	Marc Bertin and Iana Atanassova	InTeReC : an In-text Reference corpus for applying Natural language processing to Bibliometrics Paper Slides (4,88 MB)
16:00	Coffee break	
16:30	Ameni Kacem Sahraoui and Philipp Mayr	Users are not influenced by high impact and core journals while searching Paper Slides (625 kB)
17:00	Marie-Noëlle Bessagnet	A particular research data: a set of tweets (Demo) Paper Slides (270 kB)
		Demo-Session 2
17:30	Discussions	
18:00	End of the workshop	
18:30	Reception	

Regular papers

- Andre Rattinger, Jean-Marie Le Goff and Christian Guetl:
Local Word Embeddings for Query Expansion based on Co-Authorship and Citations
- Anaïs Ollagnier, Sebastien Fournier and Patrice Bellot:
BIBLME RecSys: Harnessing Bibliometric Measures for a Scholarly Paper Recommender System
- Ameni Kacem Sahraoui and Philipp Mayr:
Users are not influenced by high impact and core journals while searching
- Marc Bertin and Iana Atanassova:
InTeReC : an In-text Reference corpus for applying Natural language processing to Bibliometrics
- Maria Janina Sarol, Linxi Liu and Jodi Schneider:
Testing a Citation and Text-Based Framework for Retrieving Publications for Literature Reviews

Demo papers

- Mario Cataldi, Luigi Di Caro and Claudio Schifanella:
All for One or One for All? Analyzing Collaboration Patterns in Research Environments
- Marie-Noëlle Bessagnet:
A particular research data: a set of tweets

Important Dates

- Abstracts: 15 January 2018
- Submissions: 29 January 2018 (extended)
- Notifications: 28 February 2018
- Camera Ready Contributions: **19 March 2018**
- Workshop: **26 March 2018** in Grenoble, France

Introduction

Bibliometric techniques are not yet widely used to enhance retrieval processes in search systems, although they offer value-added effects for users. In this workshop we will explore how statistical modelling of scholarship, such as Bradfordizing or network analysis of coauthorship network, or simple citation graphs, can improve retrieval services for specific communities, as well as for large, cross-domain collections like Mendeley. This workshop aims to raise awareness of the missing link between Information Retrieval (IR) and bibliometrics/scientometrics and to create a common ground for the incorporation of bibliometric-enhanced services into retrieval at the scholarly search engine interface.

See proceedings of the former BIR workshops at ECIR 2014 <<http://ceur-ws.org/Vol-1143/>>, ECIR 2015 <<http://ceur-ws.org/Vol-1344/>>, ECIR 2016 <<http://ceur-ws.org/Vol-1567/>>, ECIR 2017 <<http://ceur-ws.org/Vol-1823/>>, JCDL 2016 <<http://ceur-ws.org/Vol-1610/>> and SIGIR 2017 <<http://ceur-ws.org/Vol-1888/>>.

Aim of the Workshop

In this 7th workshop we aim to engage with the IR community about possible links to bibliometrics and complex network theory which also explores networks of scholarly communication. Bibliometric techniques are not yet widely used to enhance retrieval processes, yet they offer value-added effects for users. Our interests include information retrieval, information seeking, science modelling, network analysis, and natural language processing. The goal is to apply insights from bibliometrics, scientometrics, and informetrics to concrete practical problems of information retrieval and browsing.

Retrieval evaluations have shown that simple text-based retrieval methods scale up well but do not progress. Traditional retrieval has reached a high level in terms of measures like precision and recall, but scientists and scholars still face challenges present since the early days of digital libraries: mismatches between search terms and indexing terms, overload from result sets that are too large and complex, and the drawbacks of text-based relevance rankings. Therefore we will focus on statistical modelling and corresponding visualizations of the evolving science system. Such analyses have revealed not only the fundamental laws of Bradford and Lotka, but also network structures and dynamic mechanisms in scientific production. Statistical models of scholarly activities are increasingly used to evaluate specialties, to forecast and discover research trends, and to shape science policy. Their use as tools in navigating scientific information in search systems is a promising but still relatively new development. We will explore how statistical modelling of scholarship can improve retrieval services for specific communities, as well as for large, cross-domain collections. Some of these techniques are already used in working systems but not well integrated in larger scholarly IR environments.

The availability of new IR test collections that contain citation and bibliographic information like the iSearch collection or the ACL collection could deliver enough ground to interest

(again) the IR community in these kind of bibliographic systems. The long-term research goal is to develop and evaluate new approaches based on informetrics and bibliometrics.

The aim of this workshop is to bring together researchers and practitioners from different domains, such as information retrieval, information seeking, science modelling, bibliometrics, scientometrics, network analysis, natural language processing, digital libraries, and approaches to visualize search and retrieval to move toward a deeper understanding of this research challenge.

Workshop Topics

To support the previously described goals the workshop topics include (but are not limited to) the following:

- IR for digital libraries and scientific information portals
- IR for scientific domains, e.g. social sciences, life sciences etc.
- Information Seeking Behaviour
- Bibliometrics, citation analysis and network analysis for IR
- Query expansion and relevance feedback approaches
- Science Modelling (both formal and empirical)
- Task based user modelling, interaction, and personalisation
- (Long-term) Evaluation methods and test collection design
- Collaborative information handling and information sharing
- Classification, categorisation and clustering approaches
- Information extraction (including topic detection, entity and relation extraction)
- Recommendations based on explicit and implicit user feedback
- Recommendation for scholarly papers, reviewers, citations and publication venues
- (Social) Book Search
- Information extraction (including topic detection, entity and relation extraction)

We especially invite descriptions of running projects and ongoing work. Papers that investigate multiple themes directly are especially welcome.

Submission Details

All submissions must be written in English following Springer LNCS author guidelines (6 to 12 pages) and should be submitted as PDF files to EasyChair. All submissions will be

reviewed by at least two independent reviewers. Please be aware of the fact that at least one author per paper needs to register for the workshop and attend the workshop to present the work. In case of no-show the paper (even if accepted) will be deleted from the proceedings AND from the program.

Springer LNCS: <http://www.springer.com/gp/computer-science/lncs/conference-proceedings-guidelines>

EasyChair: <https://easychair.org/conferences/?conf=bir2018>

Workshop proceedings will be deposited online in the CEUR workshop proceedings publication service (ISSN 1613-0073) - This way the proceedings will be permanently available and citable (digital persistent identifiers and long term preservation).

Programme Committee

- Muhammad Kamran Abbasi, University of Bedfordshire, UK
- Karam Abdulahhad, GESIS – Leibniz Institute for the Social Sciences, Germany
- Iana Atanassova, CRIT, Université de Franche-Comté, France
- Joeran Beel, Trinity College Dublin, Ireland
- Patrice Bellot, Aix-Marseille Université - CNRS (LSIS), France
- Marc Bertin, Université Lyon 1, France
- Jose Borbinha, IST / INESC-ID, Portugal
- Cornelia Caragea, Kansas State University, USA
- Zeljko Carevic, GESIS - Leibniz Institute for the Social Sciences, Germany
- Muthu Kumar Chandrasekaran, National University of Singapore, Singapore
- Nicola Ferro, University of Padova, Italy
- Edward Fox, Virginia Polytechnic Institute and State University, USA
- Norbert Fuhr, University of Duisburg-Essen, Germany
- Behnam Ghavimi, GESIS – Leibniz Institute for the Social Sciences, Germany
- C. Lee Giles, The Pennsylvania State University, USA
- Bela Gipp, Universität Konstanz, Germany
- Gilles Hubert, University of Toulouse, France
- Peter Ingwersen, University of Copenhagen, Denmark

- Kokil Jaidka, University of Pennsylvania, USA
- Roman Kern, Know-Center GmbH, Germany
- Petr Knoth, The Open University, UK
- Marijn Koolen, Huygens Institute for the History of the Netherlands, Netherlands
- Rob Koopman, OCLC, The Netherlands
- Cyril Labbé, Grenoble University, France
- Vincent Larivière, EBSI-UdeM, Canada
- Haiming Liu, University of Bedfordshire, UK
- Stasa Milojevic, Indiana University Bloomington, USA
- Peter Mutschke, GESIS – Leibniz Institute for the Social Sciences, Germany
- Horacio Saggion, Universitat Pompeu Fabra, Spain
- Philipp Schaer, TH Cologne, Germany
- Andrea Scharnhorst, DANS-KNAW, The Netherlands
- Ralf Schenkel, University of Trier, Germany
- Vivek Singh, Banaras Hindu University, India
- Henry Small, SciTech Strategies, USA
- Cassidy Sugimoto, Indiana University Bloomington, USA
- Lynda Tamine, University of Toulouse, France
- Ludovic Tanguy, University of Toulouse, France
- Simone Teufel, Cambridge University, UK
- Ulrich Thiel, Fraunhofer IPA-PAMB, Germany
- Dietmar Wolfram, University of Wisconsin-Milwaukee, USA
- Haozhen Zhao, Navigant, USA

PC chairs

- Philipp Mayr, GESIS - Leibniz Institute for the Social Sciences, Germany
- Ingo Frommholz, University of Bedfordshire in Luton, UK
- Guillaume Cabanac, University of Toulouse, France