

2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL 2016)

**Newark, New Jersey, USA
19-23 June 2016**



**IEEE Catalog Number: CFP16JCD-POD
ISBN: 978-1-5090-5254-7**

**Copyright © 2016, Association for Computing Machinery (ACM)
All Rights Reserved**

******This publication is a representation of what appears in the IEEE Digital Libraries. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP16JCD-POD
ISBN (Print-On-Demand):	978-1-5090-5254-7
ISBN (Online):	978-1-4503-4229-2

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

Table of Contents

JCDL 2016 Conference Organization	x
---	---

JCDL 2016 Sponsors & Supporters	xiv
---------------------------------------	-----

Keynote Talks

• Future Digital Libraries: Research and Responsibilities	1
<i>Maria Zemankova (National Science Foundation)</i>	
• The State of Practice and Use of Digital Collections: The Digital Public Library of America as a platform for research	3
<i>Rachel Frick (Digital Public Library of America (DPLA))</i>	
• The Energy of Delusion: The New York Art Resources Consortium (NYARC) & The Digital	5
<i>Stephen Bury (New York Art Resources Consortium (NYARC))</i>	

Paper Session 1: Wikipedia and Newspaper Analysis

• Querylog-based Assessment of Retrievability Bias in a Large Newspaper Corpus	7
<i>Myriam C. Traub, Thaer Samar, Jacco van Ossenburg, Jiyan He (Centrum Wiskunde & Informatica), Arjen de Vries (Radboud University), Lynda Hardman (Centrum Wiskunde & Informatica)</i>	
• Digital History Meets Wikipedia: Analyzing Historical Persons in Wikipedia	17
<i>Adam Jatowt, Daisuke Kawai, Katsumi Tanaka (Kyoto University)</i>	
• Quality Assessment of Wikipedia Articles without Feature Engineering	27
<i>Quang Vinh Dang, Claudia-Lavinia Ignat (Universite de Lorraine & Inria)</i>	
• Glyph Miner: A System for Efficiently Extracting Glyphs from Early Prints in the Context of OCR	31
<i>Benedikt Budig, Thomas C. van Dijk, Felix Kirchner (University Library Würzburg)</i>	

Paper Session 2: Curation and Education

• Enhancing Scholarly Use of Digital Libraries: A Comparative Survey and Review of Bibliographic Metadata Ontologies	35
<i>Jacob Jett (University of Illinois at Urbana-Champaign), Terhi Nurmikko-Fuller (University of Oxford), Timothy W. Cole (University of Illinois at Urbana-Champaign), Kevin Page (University of Oxford), J Stephen Downie (University of Illinois at Urbana-Champaign)</i>	
• Data Curation with a Focus on Reuse	45
<i>Maria Esteva, Sandra Sweat, Robert McLay, Weijia Xu, Sivakumar Kulasekaran (Texas Advanced Computing Center)</i>	
• Unraveling K-12 Standard Alignment; Report on a New Attempt	55
<i>Byron Marshall, René Reitsma (Oregon State University), Carleigh Samson (University of Colorado, Boulder)</i>	
• Research on the Follow-up Actions of College Students' Mobile Search	59
<i>Dan Wu, Shaobo Liang (Wuhan University)</i>	

Paper Session 3: Web Archiving

• Routing Memento Requests Using Binary Classifiers	63
<i>Nicolas J. Bornand, Lyudmila Balakireva, Herbert Van de Sompel (Los Alamos National Laboratory)</i>	
• The Dawn of Today's Popular Domains: A Study of the Archived German Web over 18 Years	73
<i>Helge Holzmann, Wolfgang Nejdl, Avishek Anand (L3S Research Center)</i>	
• ArchiveSpark: Efficient Web Archive Access, Extraction and Derivation	83
<i>Helge Holzmann (L3S Research Center), Vinay Goel (Internet Archive), Avishek Anand (L3S Research Center)</i>	

Paper Session 4: Search

- **Low-cost Semantic Enhancement to Digital Library Metadata and Indexing: Simple Yet Effective Strategies** 93
Annika Hinze, David Bainbridge, Sally Jo Cunningham (*University of Waikato*),
J. Stephen Downie (*University of Illinois*)
- **Desiderata for Exploratory Search Interfaces to Web Archives in Support of Scholarly Activities** 103
Andrew Jackson (*The British Library*), Jimmy Lin, Ian Milligan (*University of Waterloo*),
Nick Ruest (*York University*)
- **Content Selection and Curation for Web Archiving: The Gatekeepers vs. the Masses** 107
Ian Milligan (*University of Waterloo*), Nick Ruest (*York University*), Jimmy Lin (*University of Waterloo*)
- **Towards Better Understanding of Academic Search** 111
Madian Khabsa (*Microsoft Research*), Zhaohui Wu, C. Lee Giles (*The Pennsylvania State University*)
- **Investigating Cluster Stability when Analyzing Transaction Logs** 115
Daniel Grech, Paul Clough (*University of Sheffield*)

Paper Session 5: Q&A and Gaming

- **Experimental Evaluation of Affective Embodied Agents in an Information Literacy Game** 119
Yan Ru Guo, Dion Hoe-Lian Goh, Hurizan Bin Hussain Muhamad, Boon Kuang Ong, Zichao Lei
(*Nanyang Technological University*)
- **Evaluating the Quality of Educational Answers in Community Question-Answering** 129
Long T. Le, Chirag Shah (*Rutgers University*), Erik Choi (*Brainly*)
- **Music Information Seeking via Social Q&A: An Analysis of Questions in Music StackExchange Community** 139
Hengyi Fu (*Florida State university*), Yun Fan (*City University of New York*)

Paper Session 6: Publication Mining

- **PDFFigures 2.0: Mining Figures from Research Papers** 143
Christopher Clark, Santosh Divvala (*University of Washington*)
- **Comparing Published Scientific Journal Articles to Their Pre-print Versions** 153
Martin Klein, Peter Broadwell, Sharon E. Farb, Todd Grappone (*University of California Los Angeles*)
- **Extracting Academic Genealogy Trees from the Networked Digital Library of Theses and Dissertations** 163
Wellington Dores, Fabrício Benevenuto, Alberto H. F. Laender (*Universidade Federal de Minas Gerais*)
- **Predicting Medical Subject Headings Based on Abstract Similarity and Citations to MEDLINE Records** 167
Adam K. Kehoe, Vetle I. Torvik (*University of Illinois at Urbana-Champaign*)

Paper Session 7: Recommendation and Prediction

- **Profiling vs. Time vs. Content: What does Matter for Top-k Publication Recommendation based on Twitter Profiles?** 171
Chifumi Nishioka (*Leibniz Information Centre for Economics*),
Ansgar Scherp (*Leibniz Information Centre for Economics (ZBW Kiel) & Kiel University*)
- **Early Prediction of Scholar Popularity** 181
Masoumeh Nezhadbiglari, Marcos André Gonçalves, Jussara M. Almeida
(*Universidade Federal de Minas Gerais*)
- **Evaluating Link-based Recommendations for Wikipedia** 191
Malte Schwarzer, Moritz Schubotz (*TU Berlin*), Norman Meuschke, Corinna Breitingner (*University of Konstanz*),
Volker Markl (*TU Berlin*), Bela Gipp (*University of Konstanz*)

Posters and Demos

- **Context Matters: Towards Extracting a Citation’s Context Using Linguistic Features**201
Daniel Duma, Charles Sutton, Ewan Klein (*University of Edinburgh*)
- **Knowledge Curation Discussions and Activity Dynamics in a Short Lived Social Q&A Community** 203
Hengyi Fu, Besiki Stvilia (*Florida State University*)
- **Avoiding the Drunkard’s Search: Investigating Collection Strategies for Building a Twitter Dataset**205
Clare Llewellyn, Laura Cram, Adrian Favero (*University of Edinburgh*)
- **BIBSURF: Discover Bibliographic Entities by Searching for Units of Interest, Ranking and Filtering**207
Trond Aalberg (*NTNU*), Tanja Merčun, Maja Žumer (*University of Ljubljana*)
- **Mining Advisor-Advisee Relationships in Scholarly Big Data: A Deep Learning Approach**209
Wei Wang, Jiaying Liu, Shuo Yu, Chenxin Zhang, Zhenzhen Xu, Feng Xia (*Dalian University of Technology*)
- **Who are the Rising Stars in Academia?**211
Jun Zhang, Zhaolong Ning, Xiaomei Bai, Wei Wang, Shuo Yu, Feng Xia (*Dalian University of Technology*)
- **Can you learn it? Probably! Developing Learning Analytics Tools in R**213
Giorgio Maria Di Nunzio (*University of Padua*)
- **Characterizing Users Tagging Behavior in Academic Blogs**215
Lei Li, Chengzhi Zhang (*Nanjing University of Science & Technology*)
- **Towards Identifying Potential Research Collaborations from Scientific Research Networks using Scholarly Data**.....217
Yanet Garay, Monika Akbar, Ann Q. Gates (*University of Texas at El Paso*)
- **Personal Video Collection Management Behavior**219
Sally Jo Cunningham, David M. Nichols, Judy Bowen (*University of Waikato*)
- **Knowledge Extraction for Literature Review**221
Tatiana Erekhinskaya, Mithun Balakrishna, Marta Tatu, Steven Werner, Dan Moldovan (*Lymba Corporation*)
- **Question Identification and Classification on an Academic Question Answering Site**223
Bolanle Ojokoh (*Federal University of Technology*), Tobore Igbe (*Rom-Flex Communication Technology*), Ayobami Araoye, Friday Ameh (*Federal University of Technology*)
- **Increasing Datasets Discoverability in an Engineering Data Platform using Keyword Extraction**.....225
Parthasarathy Gopavarapu, Line C. Pouchard, Santiago Pujol (*Purdue University*)
- **Semantic Bookworm: Mining Literary Resources Revisited**.....227
Annika Hinze, Michael Coleman, Sally Jo Cunningham, David Bainbridge (*University of Waikato*)
- **Making Literature Review and Manuscript Writing Tasks Easier for Novice Researchers through Rec4LRW System**229
Aravind Sesagiri Raamkumar, Schubert Foo, Natalie Pang (*Nanyang Technological University*)
- **Can Academic Conferences Promote Research Collaboration?**.....231
Xiaoyan Su, Wei Wang, Shuo Yu, Chenxin Zhang, Teshome Megersa Bekele, Feng Xia (*Dalian University of Technology*)
- **Coagmento 2.0: A System for Capturing Individual and Group Information Seeking Behavior**233
Matthew Mitsui, Chirag Shah (*Rutgers University*)
- **Semantometrics: Towards Fulltext-based Research Evaluation**.....235
Drahomira Herrmannova, Petr Knoth (*N/A*)
- **A Supervised Learning Algorithm for Binary Domain Classification of Web Queries using SERPs**237
Alexander C. Nwala, Michael L. Nelson (*Old Dominion University*)

• How to Identify Specialized Research Communities Related to a Researcher's Changing Interests	239
Hamed Alhoori (<i>Northern Illinois University</i>)	
• Visualizing Published Metadata in Large Aggregations	241
Unmil P. Karadkar, Geoffrey A. Potter, Shengwei Wang (<i>The University of Texas at Austin</i>)	
• MemGator - A Portable Concurrent Memento Aggregator: Cross-Platform CLI and Server Binaries in Go	243
Sawood Alam, Michael L. Nelson (<i>Old Dominion University</i>)	
• Leveraging Tweet Ranking in an Optimization Framework for Tweet Timeline Generation	245
Lili Yao, Feifan Fan, Yansong Feng, Dongyan Zhao (<i>ICST of Peking University</i>)	
• Evaluating Cost of Cloud Execution in a Data Repository	247
Zhiwu Xie, Yinlin Chen, Julie Speer, Tyler Walters (<i>Virginia Polytechnic Institute and State University</i>)	
• Games for Crowdsourcing Mobile Content: An Analysis of Contribution Patterns	249
Dion Hoe-Lian Goh, Ei Pa Pa Pe-Tham, Chei Sian Lee (<i>Nanyang Technological University</i>)	
• A Methodology to Evaluate Triple Confidence and Detect Incorrect Triples in Knowledge Bases	251
Haihua Xie (<i>Peking University, State Key Lab of Digital Publishing Tech & Postdoctoral Workstation of the Zhongguancun Haidian Science Park</i>), Xiaoqing Lu, Zhi Tang (<i>Peking University</i>), Mao Ye (<i>Beijing Founder Apabi Technology Limited</i>)	
• Open Datasets for Evaluating the Interpretation of Bibliographic Records	253
Joffrey Decourselle, Fabien Duchateau (<i>LIRIS, UMR5205, Université Lyon 1</i>), Trond Aalberg (<i>NTNU</i>), Naimdjon Takhirov (<i>Westerdals - Oslo School of Arts, Communication and Technology</i>), Nicolas Lumineau (<i>LIRIS, UMR5205, Université Lyon 1</i>)	
• An Example of Automatic Authority Control	255
Anna Knyazeva (<i>Tomsk Polytechnic University</i>), Oleg Kolobov (<i>Institute of High Current Electronics SB RAS</i>), Igor Turchanovsky (<i>Tomsk Polytechnic University</i>)	
• User Activity Characterization in a Cultural Heritage Digital Library System	257
Cyrille Suire, Axel Jean-Caurant, Vincent Courboulay, Jean-Christophe Burie, Pascal Estraillier (<i>University of La Rochelle</i>)	
• A Mathematical Information Retrieval System Based on RankBoost	259
Ke Yuan, Liangcai Gao, Yuehan Wang, Xiaohan Yi, Zhi Tang (<i>ICST of Peking University</i>)	
• Using Co-authorship Networks for Author Name Disambiguation	261
Fakhri Momeni, Philipp Mayr (<i>GESIS - Leibniz Institute for the Social Sciences</i>)	
• Real-time Filtering on Interest Profiles in Twitter Stream	263
Yue Fei, Chao Lv, Yansong Feng, Dongyan Zhao (<i>Peking University</i>)	
• Preliminary Exploration of the Effect of Time Constraint on Search Interactions on Webpages	265
Chang Liu (<i>Peking University</i>), Tao Xu (<i>State University of New York, University at Albany</i>)	
• Exploiting Network Analysis to Investigate Topic Dynamics in the Digital Library Evaluation Domain	267
Leonidas Papachristopoulos (<i>'Athena' Research Centre</i>), Michalis Sfakakis (<i>Ionian University</i>), Nikos Kleidis (<i>Athens University of Economics & Business</i>), Giannis Tsakonas (<i>University of Patras</i>), Christos Papatheodorou (<i>Ionian University</i>)	
• Inventor Name Disambiguation for a Patent Database Using a Random Forest and DBSCAN	269
Kunho Kim (<i>The Pennsylvania State University</i>), Madian Khabza (<i>Microsoft Research</i>), C. Lee Giles (<i>The Pennsylvania State University</i>)	
• Big Data Processing of School Shooting Archives	271
Mohamed Farag, Pranav Nakate, Edward A. Fox (<i>Virginia Tech</i>)	
• InterPlanetary Wayback: The Permanent Web Archive	273
Sawood Alam, Mat Kelly, Michael L. Nelson (<i>Old Dominion University</i>)	

- **Improving Similar Document Retrieval Using a Recursive Pseudo Relevance Feedback Strategy**275
Kyle Williams, C. Lee Giles (*The Pennsylvania State University*)
- **Curve Separation for Line Graphs in Scholarly Documents**277
Sagnik Ray Choudhury, Shuting Wang, C. Lee Giles (*The Pennsylvania State University*)

Panels

- **Issues of Dealing with Fluid Data in Digital Libraries**279
Soo-yeon Hwang (*Rutgers University*), Melissa Cragin (*University of Illinois at Urbana-Champaign*), Michael Lesk (*Rutgers University*), Yu-Hung Lin (*Rutgers University Libraries*), Daniel O'Connor (*Rutgers University*)
- **Preserving Born-digital News: Challenges and opportunities for collection and access of born-digital journalism**281
Edward McCain (*University of Missouri*), Martin Klein (*University of California, Los Angeles*), Matthew Weber (*Rutgers, The State University of New Jersey*)

Tutorials

- **Introduction to Digital Libraries**283
Edward A. Fox (*Virginia Tech*)
- **Introduction to the Digital Public Library of America API**285
Unmil P. Karadkar (*The University of Texas at Austin*), Audrey Altman, Mark Breedlove, Mark Matienzo (*Digital Public Library of America*)
- **Information Extraction for Scholarly Digital Libraries**287
Kyle Williams, Jian Wu, Zhaohui Wu, C. Lee Giles (*The Pennsylvania State University*)

Workshop Summaries

- **ACHS'16: First International Workshop on Accessing Cultural Heritage at Scale**289
Paul D. Clough, Paula Goodale (*University of Sheffield*), Maristella Agosti (*University of Padua*), Séamus Lawless (*Trinity College Dublin*)
- **Physical Samples and Digital Libraries**291
Unmil P. Karadkar (*The University of Texas at Austin*), Kerstin Lehnert (*Columbia University*), Chris Lenhardt (*University of North Carolina at Chapel Hill*)
- **WADL 2016: Third International Workshop on Web Archiving and Digital Libraries**293
Edward A. Fox, Zhiwu Xie (*Virginia Tech*), Martin Klein (*University of California Los Angeles*)
- **5th International Workshop on Mining Scientific Publications (WOSP 2016)**295
Petr Knoth (*Mendeley Ltd.*), Drahomira Herrmannova, Lucas Anastasiou, Nancy Pontika (*The Open University*)
- **Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries (BIRNDL 2016)**299
Guillaume Cabanac (*University of Toulouse*), Muthu Kumar Chandrasekaran (*NUS School of Computing*), Ingo Frommholz (*University of Bedfordshire*), Kokil Jaidka (*Adobe Systems Inc.*), Min-Yen Kan (*NUS School of Computing*), Philipp Mayr (*GESIS - Leibniz Institute for the Social Sciences*), Dietmar Wolfram (*University of Wisconsin-Milwaukee*)

- **Author Index**301