

2022 ACM/IEEE Joint Conference on Digital Libraries (JCDL 2022)

**Cologne, Germany
20-24 June 2022**



**IEEE Catalog Number: CFP22JCD-POD
ISBN: 978-1-6654-9155-6**

**Copyright © 2022, Association for Computing Machinery (ACM)
All Rights Reserved**

This individual papers in this work are licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>.

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

IEEE Catalog Number:	CFP22JCD-POD
ISBN (Print-On-Demand):	978-1-6654-9155-6
ISBN (Online):	978-1-4503-9345-4

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

INVITED TALKS

Bridging Worlds: Indigenous Knowledge in the Digital World	1
<i>Heike Winschiers-Theophilus</i>	
D-Lib Magazine Pioneered Web-based Scholarly Communication	3
<i>Michael L. Nelson, Herbert Van De Sompel</i>	

SESSION 1: NATURAL LANGUAGE PROCESSING

A Domain-Adaptive Pre-training Approach for Language Bias Detection in News	15
<i>Jan-David Krieger, Timo Spinde, Terry Ruas, Juhi Kulshrestha, Bela Gipp</i>	
X-SCITLDR: Cross-Lingual Extreme Summarization of Scholarly Documents	22
<i>Sotaro Takeshita, Tommaso Green, Niklas Friedrich, Kai Eckert, Simone Paolo Ponzetto</i>	
TinyGenius: Intertwining Natural Language Processing with Microtask Crowdsourcing for Scholarly Knowledge Graph Creation	34
<i>Allard Oelen, Markus Stocker, Sören Auer</i>	
Vision and Natural Language for Metadata Extraction from Scientific PDF Documents: A Multimodal Approach.....	39
<i>Zeyd Boukhers, Azeddine Bouabdallah</i>	

SESSION 2: INFORMATION RETRIEVAL & ACCESS

Specialized Document Embeddings for Aspect-Based Similarity of Research Papers	44
<i>Malte Ostendorff, Till Blume, Terry Ruas, Bela Gipp, Georg Rehm</i>	
Studying Retrievability of Publications and Datasets in an Integrated Retrieval System	56
<i>Dwaipayan Roy, Zeljko Carevic, Philipp Mayr</i>	
What a Publication Tells You – Benefits of Narrative Information Access in Digital Libraries.....	65
<i>Hermann Kroll, Jan Pirklbauer, Florian Plötzky, Wolf-Tilo Balke</i>	

SESSION 3: SEARCH AND RECOMMENDATION

Causal Factorization Machine for Robust Recommendation	73
<i>Yunqi Li, Hanxiong Chen, Juntao Tan, Yongfeng Zhang</i>	
Query-Specific Subtopic Clustering	82
<i>Sumanta Kashyapi, Laura Dietz</i>	
On Modifying Evaluation Measures to Deal with Ties in Ranked Lists	91
<i>Sourav Saha, Dwaipayan Roy, Mandar Mitra</i>	

SESSION 4: WEB ARCHIVES

ABCDEF - the 6 Key Features Behind Scalable, Multi-tenant Web Archive Processing with ARCH: Archive, Big Data, Concurrent, Distributed, Efficient, Flexible	95
<i>Helge Holzmann, Nick Ruest, Jefferson Bailey, Alex Dempsey, Samantha Fritz, Peggy Lee, Ian Milligan</i>	
Investigating Bloom Filters for Web Archives' Holdings.....	106
<i>Martin Klein, Lyudmila Balakireva, Karolina Holub, Draženko Celjak, Ingeborg Rudomino</i>	
StreamingHub: Interactive Stream Analysis Workflows.....	116
<i>Yasith Jayawardana, Vikas G. Ashok, Sampath Jayarathna</i>	

SESSION 5: BIBLIO- /ALTMETRICS

Open-Source Code Repository Attributes Predict Impact of Computer Science Research	126
<i>Prajwal Bhattarai, Mohammed Ghassemi, Tuka Alhanai</i>	
Altmetrics and Citation Counts: An Empirical Analysis of the Computer Science Domain.....	133
<i>Yusra Shakeel, Rand Alchokr, Jacob Krüger, Thomas Leich, Gunter Saake</i>	
Pre-Trained Transformer-Based Citation Context-Aware Citation Network Embeddings	144
<i>Masaya Ohagi, Akiko Aizawa</i>	

SESSION 6: INFORMATION EXTRACTION

Mining Mathematical Documents for Question Answering Via Unsupervised Formula Labeling	149
<i>Philipp Scharpf, Moritz Schubotz, Bela Gipp</i>	
Visual Descriptor Extraction from Patent Figure Captions: A Case Study of Data Efficiency Between BiLSTM and Transformer	160
<i>Xin Wei, Jian Wu, Kehinde Ajayi, Diane Oyen</i>	

SESSION 7: SEARCH AND RECOMMENDATION II

Asking for Help in Community Question-Answering: The Goal-Framing Effect of Question Expression on Response Networks.....	165
<i>Qian Wu, Chei Sian Lee, Dion Hoe-Lian Goh</i>	
Recommending Research Papers to Chemists: A Specialized Interface for Chemical Entity Exploration	170
<i>Corinna Breitingner, Kay Herklotz, Tim Flegelskamp, Norman Meuschke</i>	

SESSION 8: USER BEHAVIOR

Leveraging User Interaction Signals and Task State Information in Adaptively Optimizing Usefulness-Oriented Search Sessions.....	174
<i>Jiqun Liu, Chirag Shah</i>	
Reliable Editions from Unreliable Components: Estimating Ebooks from Print Editions Using Profile Hidden Markov Models.....	185
<i>A. B. Riddell</i>	

Information Seeking Within Academic Digital Libraries: A Survey of Graduate Student Search Strategies	190
<i>Orland Hoeber, Dale Storie</i>	

A Reference Dependence Approach to Enhancing Early Prediction of Session Behavior and Satisfaction	195
<i>Tyler Brown, Jiqun Liu</i>	

SESSION 9: SCHOLARLY COMMUNICATION I

Comparing Different Perspectives of Characterizing Interdisciplinarity of Scientific Publications: Author Vs. Publication Perspectives	200
<i>Shan Liu, Hongkan Chen, Yi Bu</i>	

How Does Author Affiliation Affect Preprint Citation Count? Analyzing Citation Bias at the Institution and Country Level.....	210
<i>Chifumi Nishioka, Michael Färber, Tarek Saier</i>	

DeepASPeer: Towards an Aspect-Level Sentiment Controllable Framework for Decision Prediction from Academic Peer Reviews	218
<i>Sandeep Kumar, Hardik Arora, Tirthankar Ghosal, Asif Ekbal</i>	

SESSION 10: SCHOLARLY COMMUNICATION II

Complexities Associated with User-Generated Book Reviews in Digital Libraries: Temporal, Cultural, and Political Case Studies	229
<i>Yuerong Hu, Zoe Leblanc, Jana Diesner, Ted Underwood, Glen Layne-Worthey, J. Stephen Downie</i>	

The Significance and Impact of Winning an Academic Award: A Study of Early Career Academics	241
<i>Jing Ren, Yajie Shi, Adrian Shatte, Xiangjie Kong, Feng Xia</i>	

Between Acceptance and Rejection: Challenges for an Automatic Peer Review Process	252
<i>Gustavo Lúcius Fernandes, Pedro O. S. Vaz-De-Melo</i>	

SESSION 11: CLASSIFICATION

How Many Sources Are Needed? : The Effects of Bibliographic Databases on Systematic Review Outcomes.....	264
<i>Carolin Keller, Tamara Heck, Marc Rittberger</i>	

Cross-Domain Multi-Task Learning for Sequential Sentence Classification in Research Papers.....	275
<i>Arthur Brack, Anett Hoppe, Pascal Buschermöhle, Ralph Ewerth</i>	

A Library Perspective on Nearly-Unsupervised Information Extraction Workflows in Digital Libraries	288
<i>Hermann Kroll, Jan Pirklbauer, Florian Plötzky, Wolf-Tilo Balke</i>	

DEMOS

Integration of Text and Geospatial Search for Hydrographic Datasets Using the Lucene Search Library	299
<i>Matthew Y. R. Yang, Siwen Yang, Jimmy Lin</i>	

SchenQL: A Query Language for Bibliographic Data with Aggregations and Domain-Specific Functions	304
<i>Christin Katharina Kreutz, Martin Blum, Ralf Schenkel</i>	
DevOps Practices in Digital Library Development	309
<i>Yinlin Chen</i>	
TermoPL – a Tool for Extracting and Clustering Domain Related Terms	313
<i>Piotr Rychlik, Malgorzata Marciniak, Agnieszka Mykowiecka</i>	
Collaborative Annotation and Semantic Enrichment of 3D Media: A FOSS Toolchain.....	317
<i>Lozana Rossenova, Zoe Schubert, Richard Vock, Lucia Sohmen, Lukas Günther, Paul Duchesne, Ina Blümel</i>	
The Hurdles of Current Data Citation Practices and the Adding-Value of Providing PIDs Below Study Level.....	322
<i>Janete Saldanha Bach, Claus-Peter Klas, Peter Mutschke</i>	
BIP! Scholar: A Service to Facilitate Fair Researcher Assessment.....	327
<i>Thanasis Vergoulis, Serafeim Chatzopoulos, Kleanthis Vichos, Ilias Kanellos, Andrea Mannocci, Natalia Manola, Paolo Manghi</i>	
Memento Validator: A Toolset for Memento Compliance Testing	332
<i>Bhanuka Mahanama, Lyudmila Balakireva, Sampath Jayarathna, Michael Nelson, Martin Klein</i>	

DATASETS

S2AMP: A High-Coverage Dataset of Scholarly Mentorship Inferred from Publications	335
<i>Shaurya Rohatgi, Doug Downey, Daniel King, Sergej Feldman</i>	
A Prototype Gutenberg-HathiTrust Sentence-level Parallel Corpus for OCR Error Analysis: Pilot Investigations.....	340
<i>Ming Jiang, Ryan C Dubnick, Glen Worthey, Ted Underwood, J. Stephen Downie</i>	
HedgePeer: A Dataset for Uncertainty Detection in Peer Reviews.....	345
<i>Tirthankar Ghosal, Kamal Kaushik Varanasi, Valia Kordoni</i>	

DOCTORAL CONSORTIUM

Opening Scholarly Documents Through Text Analytics.....	350
<i>Bipasha Banerjee</i>	
Thesis Plan: The Effect of Data Science Teaching for non-STEM Students	352
<i>LuiZ Barboza</i>	
Context-Sensitive, Personalized Search at the Point of Care	354
<i>Sameh Frihat</i>	
Synthesizing Digital Libraries and Digital Humanities Perspectives for Illuminating Under- Investigated Complexities Associated with User-generated Book Reviews	356
<i>Yuerong Hu</i>	

Integration of Models for Linked Data in Cultural Heritage and Contributions to the FAIR Principles	358
<i>Inês Koch</i>	

TUTORIALS

Building Digital Library Collections with Greenstone 3 Tutorial	360
<i>David Bainbridge</i>	
Open Refine to Wikibase: A New Data Upload Pipeline	362
<i>Lucia Sohmen, Lozana Rossenova</i>	

WORKSHOPS

JCDL2022 Workshop: Extraction and Evaluation of Knowledge Entities from Scientific Documents (EEKE2022)	364
<i>Chengzhi Zhang, Philipp Mayr, Wei Lu, Yi Zhang</i>	
20th European NKOS Workshop – Networked Knowledge Organization Systems and Services	366
<i>Douglas Tudhope, Claudio Gnoli, Koraljka Golub, Philipp Mayr</i>	
Understanding Literature References in Academic Full Text (ULITE)	368
<i>Anastasiia Iurshina, Muhammad Ahsan Shahid, Tobias Backes, Philipp Mayr, Steffen Staab</i>	
Web Archiving and Digital Libraries (WADL) 2022.....	370
<i>Martin Klein, Mat Kelly, Zhiwu Xie, Edward A. Fox</i>	
2nd Workshop on Digital Infrastructures for Scholarly Content Objects (DISCO'22).....	372
<i>Wolf-Tilo Balke, Hermann Kroll, Yuanxi Fu, Jodi Schneider, Anita De Waard</i>	
E-Only Theses Submission and Preservation Workshop.....	374
<i>Leo Konstantelos, Michelle O'Hara, Sally Bell, Clare Paterson, William J. Nixon</i>	

Author Index