# 2023 ACM/IEEE Joint **Conference on Digital Libraries** (JCDL 2023)

Santa Fe, New Mexico, USA 26-30 June 2023



**IEEE Catalog Number: CFP23JCD-POD ISBN**:

979-8-3503-9932-5

## Copyright © 2023 by the Institute of Electrical and Electronics Engineers, Inc. All Rights Reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

\*\*\* 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:
 CFP23JCD-POD

 ISBN (Print-On-Demand):
 979-8-3503-9932-5

 ISBN (Online):
 979-8-3503-9931-8

ISSN: 2575-7865

#### 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



## 2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL)

## **JCDL 2023**

### **Table of Contents**

Message from the General Chairxiv
Message from the Program Committee Chairsxx
Organizing Committeexv
Program Committeexvi
Steering Committeexx
Reviewersxxi
Digital Libraries
Evaluating Digital Library Search Systems by Using Formal Process Modelling
ntegrated Digital Library System for Long Documents and Their Elements
SMAuC - The Scientific Multi-Authorship Corpus
A Conceptual Framework for the Design of Digital Repository Interfaces Supporting Digital Content Reuse Assessment

### Scientometrics

Deciphering the Reviewer's Aspectual Perspective: A Joint Multitask Framework for Aspect and Sentiment Extraction from Scholarly Peer Reviews	35
CoCon: A Data Set on Combined Contextualized Research Artifact Use  Tarek Saier (Karlsruhe Institute of Technology, Germany), Youxiang  Dong (Karlsruhe Institute of Technology, Germany), and Michael Färber  (Karlsruhe Institute of Technology, Germany)	47
CitePrompt: Using Prompts to Identify Citation Intent in Scientific Papers  Avishek Lahiri (Indian Association for the Cultivation of Science, India), Debarshi Kumar Sanyal (Indian Association for the Cultivation of Science, India), and Imon Mukherjee (Indian Institute of Information Technology, India)	51
Laying Foundations to Quantify the "Effort of Reproducibility"  Akhil Pandey Akella (Northern Illinois University, USA), David Koop (Northern Illinois University, USA), and Hamed Alhoori (Northern Illinois University, USA)	56
MetaEnhance: Metadata Quality Improvement for Electronic Theses and Dissertations of University Libraries  Muntabir Hasan Choudhury (Old Dominion University, USA), Lamia Salsabil (Old Dominion University, USA), Himarsha R. Jayanetti (Old Dominion University, USA), Jian Wu (Old Dominion University, USA), William A. Ingram (Virginia Tech, USA), and Edward A. Fox (Virginia Tech, USA)	61
unarXive 2022: All arXiv Publications Pre-Processed for NLP, Including Structured Full-Text and Citation Network	66
Web Archiving	
Making Changes in Webpages Discoverable: A Change-Text Search Interface for Web Archives . Lesley Frew (Old Dominion University, USA), Michael L. Nelson (Old Dominion University, USA), and Michele C. Weigle (Old Dominion University, USA)	71
Right HTML, Wrong JSON: Challenges in Replaying Archived Webpages Built with Client-Side Rendering	82
TrendMachine: A Temporal Webpage Resilience Portal	93

End of Term Web Archive Dataset: Longitudinal Web Archive of .GOV and .MIL Domains	8
Less than 4% of Archived Instagram Account Pages for the Disinformation Dozen are  Replayable	2
Haley Bragg (Christopher Newport University, USA), Himarsha R. Jayanetti (Old Dominion University, USA), Michael L. Nelson (Old Dominion University, USA), and Michele C. Weigle (Old Dominion University, USA)	_
Information Retrieval	
Bibliometric Data Fusion for Biomedical Information Retrieval	7
Toward Connecting Speech Acts and Search Actions in Conversational Search Tasks	9
Binding Data Narrations - Corroborating the Plausibility of Scientific Narratives by Open Research Data	2
Knowledge Graphs and Knowledge Organization	
Enriching Simple Keyword Queries for Domain-Aware Narrative Retrieval 14  Hermann Kroll (TU Braunschweig, Germany), Christin Katharina Kreutz (TH Köln - University of Applied Sciences, Germany), Pascal Sackhoff (TU Braunschweig, Germany), and Wolf-Tilo Balke (TU Braunschweig, Germany)	3
SciKGTeX - A LaTeX Package to Semantically Annotate Contributions in Scientific	_
Publications	0
Exploring Multiview Interactive Visualization for Chinese Historical Texts Using Knowledge Aggregation	5
Chengxi Yan (Renmin University of China, China) and Shaojian Li (Renmin University of China, China)	J

### Digital Humanities and Teaching

Reading Lists Systems' Pedagogical Features: A Comparative Analysis	169
A King's Counsel: A Network(ed) Approach: Digitizing the Privy Council Registers of Henry VIII, 1540-1547	179
Caitlin Burge (University of Luxembourg, Luxembourg)	
Yes but Can ChatGPT Identify Entities in Historical Documents?  Carlos-Emiliano González-Gallardo (University of La Rochelle, L3i, La Rochelle, France), Emanuela Boros (EPFL, Digital Humanities Laboratory, Lausanne, Switzerland), Nancy Girdhar (University of La Rochelle, L3i, La Rochelle, France), Ahmed Hamdi (University of La Rochelle, L3i, La Rochelle, France), Jose G. Moreno (University of Toulouse, IRIT, Toulouse, France), and Antoine Doucet (University of La Rochelle, L3i, La Rochelle, France)	184
FastCat Catalogues: Interactive Entity-Based Exploratory Analysis of Archival Documents	190
MINE - A Text Analysis Service for Digital Humanities Scientists  Triet Ho Anh Doan (Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen, Germany), Sven Bingert (Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen, Germany), and Ramin Yahyapour (Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen, Germany)	195
AI / ML / Entity Extraction	
Mining the History Sections of Wikipedia Articles on Science and Technology Wolfgang Kircheis (Leipzig University and ScaDS.AI, Germany), Marion Schmidt (Deutsche Zentrum für Hochschul- und Wissenschaftsforschung, Germany), Arno Simons (Technische Universität Berlin, Germany), Benno Stein (Bauhaus-Universität Weimar, Germany), and Martin Potthast (Leipzig University and ScaDS.AI, Germany)	200
Efficient Ultrafine Typing of Named Entities	205
Mining Semantic Relations in Data References to Understand the Roles of Research Data in Academic Literature	215
Michigan, USA), and Libby Hemphill (Inter-university Consortium for Political and Social Research, University of Michigan, USA)	

Soft-Search: Two Datasets to Study the Identification and Production of Research Software  Eva Maxfield Brown (University of Washington Information School, USA),  Lindsey Schwartz (University of Washington Information School, USA),  Richard Lewei Huang (University of Washington Information School,  USA), and Nicholas Weber (University of Washington Information School,  USA)	228
Extreme Classification for Answer Type Prediction in Question Answering	232
Zero-Shot Entailment of Leaderboards for Empirical AI Research	237
Posters / Demos	
Composable Ledgers for Distributed Synchronic Web Archiving	242
Drug Target Extraction from Biomedical Articles Based on a Two-Stage Cascading Framework  Xuesi Li (National Science Library Chinese Academy of Sciences, China), Liangping Ding (National Science Library Chinese Academy of Sciences, China), and Zhixiong Zhang (National Science Library Chinese Academy of Sciences, China)	245
History-Related Content Recommendation in Social Networks Based on User's Interest Estimation  Kouki Yoshida (Takushoku University, Japan), Adam Jatowt (University of Innsbruck, Austria), and Yasunobu Sumikawa (Takushoku University, Japan)	247
How to Narrow the Last Mile of Cultural Services: An Innovative Way on Public Culture Cloud Platforms  Jingzhu Wei (Sun Yat-sen University, China), Lele Zhang (Sun Yat-sen University, China), and Tongrui Zhang (Sun Yat-sen University, China)	249

PADME-SoSci: A Platform for Analytics and Distributed Machine Learning for the Social	
Sciences	251
Zeyd Boukhers (Fraunhofer Institute for Applied Information Technology	
FIT, Germany; Faculty of Medicine and University Hospital Cologne, University of Cologne, Germany), Arnim Bleier (GESIS – Leibniz	
Institute for the Social Sciences, Germany), Yeliz Ucer Yediel	
(Fraunhofer Institute for Applied Information Technology FIT, Germany;	
RWTH Aachen University, Germany), Mio Hienstorfer-Heitmann (GESIS –	
Leibniz Institute for the Social Sciences, Germany), Mehrshad	
Jaberansary (Faculty of Medicine and University Hospital Cologne,	
University of Cologne, Germany), Sascha Welten (RWTH Aachen	
University, Germany), Adamantios Koumpis (Faculty of Medicine and	
University Hospital Cologne, RWTH Aachen University, Germany), and Oya	
Beyan (Fraunhofer Institute for Applied Information Technology FIT,	
Germany; Faculty of Medicine and University Hospital Cologne,	
University of Cologne, Germany)	
Assessing the Accessibility of Web Archives	253
Mohan Sunkara (Old Dominion University, USA), Sandeep Kalari (Old	
Dominion University, USA), Sampath Jayaratna (Old Dominion University,	
USA), and Vikas Ashok (Old Dominion University, USA)	
Maximizing Equitable Reach and Accessibility of ETDs	256
William A. Ingram (Virginia Tech, USA), Jian Wu (Old Dominion	
University, USA), and Edward A. Fox (Virginia Tech, USA)	
What's in the News? Towards Identification of Bias by Commission, Omission, and Source	
Selection (COSS)	258
Anastasia Zhukova (University of Göttingen, Germany), Terry Ruas	
(University of Göttingen, Germany), Felix Hamborg (Heidelberg Academy	
of Sciences and Humanities, Germany), Karsten Donnay (University of	
Zurich, Switzerland), and Bela Gipp (University of Göttingen, Germany)	
Bibliography Counselor: A Citation Recommendation Tool	260
Debanjan Dutta (Ramakrishna Mission Vivekananda Educational and	
Research Institute, India), Dipasree Pal (Indian Statistical	
Institute, India), Dwaipayan Roy (Indian Institute of Science	
Education and Research, India), and Mandar Mitra (Indian Statistical	
Institute, India)	
The Impact of the Message Framing and Location of Short Video Cover Titles on Users: The	
Case of Health Short Videos on Physical Activities	263
Chengxi Zhou (Wuhan University, China) and Huiyi Tian (Wuhan	
University, China)	
The Memento Tracer Toolset for Human-Guided Focused Crawling of Dynamic Web	265
Lyudmila Balakireva (Los Alamos National Laboratory, USA), Emily	
Escamilla (Old Dominion University, USA), Talya Cooper (New York	
University, USA), Michael L. Nelson (Old Dominion University, USA),	
and Michele C. Weigle (Old Dominion University, USA)	
Touch the Collections: A 3D UI Prototype in DSpace-7	267
Le Yang (University of Oregon, USA) and Zhongda Zhang (University of	
Oklahoma, USA)	

Persistent Identifiers and Research Data	269
TEIMMA: The First Content Reuse Annotator for Text, Images, and Math  Ankit Satpute (FIZ Karlsruhe Leibniz Institute for Information Infrastructure, Germany), André Greiner-Petter (University of Göttingen, Germany), Moritz Schubotz (FIZ Karlsruhe Leibniz Institute for Information Infrastructure, Germany), Norman Meuschke (University of Göttingen, Germany), Akiko Aizawa (National Institute of Informatics, Japan), Olaf Teschke (FIZ Karlsruhe Leibniz Institute for Information Infrastructure, Germany), and Bela Gipp (University of Göttingen, Germany)	271
Neural Ensemble for Organizing Business News-Streams for Corporate Libraries	274
Capturing Stability of Information Needs in Digital Libraries	. 276
Book Size, Book Format: Historical Relationship in the HathiTrust Digital Library Metadata (1500–1799)	. 279
AI Usage Cards: Responsibly Reporting AI-Generated Content  Jan Philip Wahle (University of Göttingen, Germany), Terry Ruas  (University of Göttingen, Germany), Saif M. Mohammad (National  Research Council, Canada), Norman Meuschke (University of Göttingen,  Germany), and Bela Gipp (University of Göttingen, Germany)	282
Multimodal Search on Iconclass using Vision-Language Pre-Trained Models  Cristian Santini (Karlsruhe Institute of Technology, Germany), Etienne Posthumus (FIZ Karlsruhe, Germany), Tabea Tietz (Karlsruhe Institute of Technology, Germany), Mary Ann Tan (Karlsruhe Institute of Technology, Germany), Oleksandra Bruns (Karlsruhe Institute of Technology, Germany), and Harald Sack (Karlsruhe Institute of Technology, Germany)	. 285
The Library of Babel as a Digital Library of Human Mistakes  Lars Quakulinski (RWTH Aachen University, Germany), Adamantios Koumpis (University of Cologne, Germany), and Oya Deniz Beyan (University of Cologne, Germany; Fraunhofer Institute for Applied Information Technology, Germany)	. 288
Measuring the Growth of Ideas in a Title Corpus	291
Discovering Image Usage Online: A Case Study With "Flatten the Curve"	293

Semantification of Space Data - A Feasibility Study	295
Perseus 6.0: Towards a Next Generation Reading Environment for Born-Digital Editions and Corpora	297
Visualizing Relation Between (De) Motivating Topics and Public Stance Toward COVID-19 Vaccine  Ashiqur Rahman (Northern Illinois University, USA) and Hamed Alhoori (Northern Illinois University, USA)	299
A Eurovision Digital Library: Let's Open up with Linked Data	301
Workshops	
JCDL2023 Workshop: Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data	303
JCDL2023 Workshop: Joint Workshop of the 4th Extraction and Evaluation of Knowledge Entities from Scientific Documents (EEKE2023) and the 3rd AI + Informetrics (AII2023)	306
2nd International Workshop on Digital Language Archives	308
Persistent Identifiers & US Federal Agency Policies: An Overview of the Recent Policies, PID Requirements and Timelines  Brian Minihan (ORCID, Inc.), Carly Robinson (U.S. Department of Energy, Office of Science), Amanda French (Research Organization Registry), and Sara Bowman (Crossref)	311

FAIR Data for Large Research Facilities	12
Web Archiving and Digital Libraries (WADL) 2023	14
Tutorials	
The Solid Protocol and Its Potential for Open Science	16
Tutorial: Up and Running with ARK Persistable Identifiers	18
Panels	
Who can Submit an Excellent Review for this Manuscript in the Next 30 Days? – Peer Reviewing in the age of Overload	19
Research Data Without Borders: How the Sections of the NFDI Work on Cross-Cutting Topics  Across Disciplines and Consortia	21
AI and Public Archives: Collaborative Leadership for Responsible Adoption	23
Author Index	25