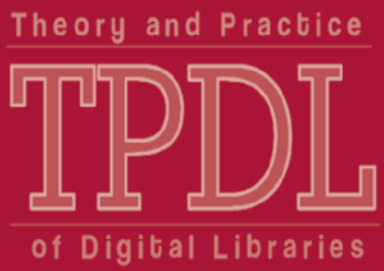
Gianmaria Silvello · Oscar Corcho ·
Paolo Manghi · Giorgio Maria Di Nunzio ·
Koraljka Golub · Nicola Ferro ·
Antonella Poggi (Eds.)

LNCS 13541

Linking Theory and Practice of Digital Libraries

26th International Conference on Theory and Practice of Digital Libraries, TPDL 2022
Padua, Italy, September 20–23, 2022, Proceedings



Springer

Lecture Notes in Computer Science 13541

Founding Editors

Gerhard Goos
Karlsruhe Institute of Technology, Karlsruhe, Germany

Juris Hartmanis
Cornell University, Ithaca, NY, USA

Editorial Board Members

Elisa Bertino
Purdue University, West Lafayette, IN, USA

Wen Gao
Peking University, Beijing, China

Bernhard Steffen
TU Dortmund University, Dortmund, Germany

Moti Yung
Columbia University, New York, NY, USA

More information about this series at https://link.springer.com/bookseries/558

Gianmaria Silvello · Oscar Corcho ·
Paolo Manghi · Giorgio Maria Di Nunzio ·
Koraljka Golub · Nicola Ferro ·
Antonella Poggi (Eds.)

Linking Theory and Practice of Digital Libraries

26th International Conference on Theory and Practice of Digital Libraries, TPDL 2022
Padua, Italy, September 20–23, 2022
Proceedings

Springer

Editors
Gianmaria Silvello
University of Padua
Padua, Italy

Paolo Manghi
CNR-ISTI – National Research Council
Pisa, Italy

Koraljka Golub
Linnaeus University
Växjö, Sweden

Antonella Poggi
Sapienza University of Rome
Rome, Italy

Oscar Corcho
Universidad Politécnica de Madrid
Madrid, Spain

Giorgio Maria Di Nunzio
University of Padua
Padua, Italy

Nicola Ferro
University of Padua
Padua, Italy

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-031-16801-7 ISBN 978-3-031-16802-4 (eBook)
https://doi.org/10.1007/978-3-031-16802-4

© Springer Nature Switzerland AG 2022
Chapters "Should We Preserve the World's Software History, And Can We?", "Extracting Funder Information from Scientific Papers - Experiences with Question Answering", "Implementation Issues for a Highly Structured Research Report" and "B!SON: A Tool for Open Access Journal Recommendation" are licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/). For further details see license information in the chapters.
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

These proceedings contain the full papers, short papers, and accelerating innovation papers selected for presentation at the 26th International Conference on Theory and Practice of Digital Libraries (TPDL 2022)[1]. The event was organized by the Information Management Systems (IMS) research group[2] of the Department of Information Engineering[3] of the University of Padua[4], Italy. The conference was held from September 20 to September 23, 2022, in Padua, Italy.

Over the years TPDL has established an important international forum focused on digital libraries and associated technical, practical, and social issues. TPDL encompasses the many meanings of the term digital libraries, including new forms of information institutions; operational information systems with all manner of digital content; new means of selecting, collecting, organizing, and distributing digital content; and theoretical models of information media, including document genres and electronic publishing. Digital libraries may be viewed as a new form of information institution or as an extension of the services libraries currently provide.

TPDL historically approached digital libraries embracing the field at large, and also recognizing three key areas of interest that can be synthesized as scholarly communication (e.g., research data, research software, digital experiments, digital libraries), e-science/computationally-intense research (e.g., scientific workflows, virtual research environments, reproducibility) and library, archive, and information science (e.g., governance, policies, open access, open science).

TPDL 2022 received a total of 107 submissions in three categories: 51 full papers, 50 short papers, and 6 accelerating innovation papers.

Full papers present high quality, original research of relevance to the TPDL community. Submissions should detail their methods and techniques in sufficient detail to enable replication and reuse. The accepted papers are published in these conference proceedings and were presented as long conference talks.

Short papers present high quality, original research or tools or applications that are of relevance to the TPDL community. The accepted papers are published in these conference proceedings and were presented as short conference talks.

Finally, the newly introduced "accelerating innovation" papers aim to speed up the transfer and exchange of new ideas and late-breaking results in order to accelerate and drive innovation. The accepted papers were presented in a gong show-style and with a poster at the conference.

The geographical distribution of the authors of the submitted papers was as follows: 81.3% were from Europe (38.5% from Germany, 18% from Italy, and 15% from Greece),

[1] http://tpdl2022.dei.unipd.it/.
[2] http://ims.dei.unipd.it/.
[3] http://www.dei.unipd.it/.
[4] http://www.unipd.it/en/.

8.3% from Asia, 8% from North and South America, 1.2% from Africa, and 0.6% from Australasia.

All submissions were reviewed by at least three members of an international Program Committee and by one senior meta-reviewer. Of the full papers submitted to the conference, 18 were accepted for oral presentation (35%), nine were accepted as short papers (17.6%), and two as accelerating innovation papers (0.04%). Of the short papers, 18 were accepted for oral presentation (36%) and 10 were accepted as accelerating innovation papers (20%). Finally, TPDL 2022 accepted four out of the six submitted accelerating innovation papers.

We thank all Program Committee members for their time and effort in ensuring the high quality of the TPDL 2022 program.

Additionally, TPDL 2022 hosted two workshops covering a range of digital library topics. These were selected by the workshop committee.

The workshops were

- The 1st International Workshop on Digital Platforms and Resources for Access to Literary Heritage (DIPRAL 2022); and
- The 2nd International Workshop on Archives and Linked Data (Linked Archives 2022).

We would like to thank our invited speakers for their contributions to the program: Roberto Di Cosmo (Inria - National Institute for Research in Digital Science and Technology, France) and Georgia Koutrika (Athena Research Center, Greece). Short descriptions of these talks are included in the proceedings.

TPDL 2022 was held under the patronage of the University of Padua, and the Department of Information Engineering.

Finally, we would like to thank our sponsors: ExaMode and the Universidad Politecnica de Madrid (gold level); the Department of Information Engineering of the University of Padua (silver level); and MDPI and Springer (bronze level). The conference was supported by the Coalition for Networked Information.

August 2022

Gianmaria Silvello
Oscar Corcho
Paolo Manghi
Giorgio Maria Di Nunzio
Koraljka Golub
Nicola Ferro
Antonella Poggi

Organization

General Chair

Gianmaria Silvello	University of Padua, Italy

Program Chairs

Oscar Corcho	Universidad Politcnica de Madrid, Spain
Paolo Manghi	ISTI-CNR, Italy

Short Program Chairs

Giorgio Maria Di Nunzio	University of Padua, Italy
Koraljka Golub	Linnaeus University, Sweden

Accelerating Innovation Program Chairs

Nicola Ferro	University of Padua, Italy
Antonella Poggi	Sapienza University of Rome, Italy

Workshop Chair

Leonardo Candela	ISTI-CNR, Italy

Doctoral Consortium Chairs

Gerd Berget	Oslo Metropolitan University, Norway
Trond Aalberg	Norwegian University of Science and Technology, Norway

Organization Chair

Fabio Giachelle	University of Padua, Italy

Senior Program Committee

Trond Aalborg	Norwegian University of Science and Technology, Norway
Maristella Agosti	University of Padua, Italy

Omar Alonso	Amazon, USA
Amir Aryani	Swinburne University of Technology, Australia
Sören Auer	Leibniz Universität Hannover, Germany
José Borbinha	Universidade de Lisboa, Portugal
George Buchanan	University of Melbourne, Australia
Donatella Castelli	ISTI-CNR, Italy
Fabien Duchateau	Université Claude Bernard Lyon 1 - LIRIS, France
Edward Fox	Virginia Tech, USA
Norbert Fuhr	University of Duisburg-Essen, Germany
Carole Goble	University of Manchester, UK
Mark Michael Hall	The Open University, UK
Martin Klein	Los Alamos National Laboratory, USA
Michael Nelson	Old Dominion University, USA
Christos Papatheodorou	University of Athens, Greece
Gabriella Pasi	University of Milano-Bicocca, Italy
Thomas Risse	University of Frankfurt, Germany
Herbert Van de Sompel	Data Archiving Networked Services, The Netherlands
Ludo Waltman	Leiden University, The Netherlands

Program Committee

Sawood Alam	Internet Archive, USA
Marco Angelini	Sapienza University of Rome, Italy
Carlos Badenes-Olmedo	Universidad Politcnica de Madrid, Spain
Miriam Baglioni	ISTI-CNR, Italy
Wolf-Tilo Balke	TU Braunschweig, Germany
Alessia Bardi	ISTI-CNR, Italy
Carlos Bobed	University of Zaragoza, Spain
José Borbinha	University of Melbourne, Australia
Ricardo Campos	Polytechnic Institute of Tomar, Portugal
Annalina Caputo	Dublin City University, Ireland
Michelangelo Ceci	University of Bari, Italy
Yinlin Chen	Virginia Tech, USA
Marilena Daquino	University of Bologna, Italy
Boris Dobrov	Moscow State University, Russia
Dennis Dosso	University of Padua, Italy
Guglielmo Faggioli	University of Padua, Italy
Stefano Ferilli	University of Bari, Italy
Nuno Freire	INESC-ID, Portugal
Yannis Foufoulas	University of Athens, Greece
Fabio Giachelle	University of Padua, Italy
Jose Manuel Gomez-Perez	expert.ai, Spain

Antoine Isaac	Europeana and VU University Amsterdam, The Netherlands
Adam Jatowt	University of Innsbruck, Austria
Jaap Kamps	University of Amsterdam, The Netherlands
Johannes Kiesel	Bauhaus-Universität Weimar, Germany
Petr Knoth	The Open University, UK
Maurizio Lana	Università del Piemonte Orientale, Italy
Maria Maistro	University of Copenhagen, Denmark
Andrea Mannocci	ISTI-CNR, Italy
Stefano Marchesin	University of Padua, Italy
Reme Melero	Spanish National Research Council, Spain
Eva Méndez	University Carlos III of Madrid, Spain
Nicola Orio	University of Padua, Italy
Francesco Osborne	The Open University, UK
Silvio Peroni	University of Bologna, Italy
Simone Paolo Ponzetto	University of Mannheim, Germany
Edie Rasmussen	University of British Columbia, Canada
Andreas Rauber	Vienna University of Technology, Austria
Cristina Ribeiro	University of Porto, Portugal
Irene Rodrigues	Universidade de Evora, Portugal
Idafen Santana-Perez	Universidad de Las Palmas de Gran Canaria, Spain
Andrea Scharnhorst	Royal Netherlands Academy of Arts and Sciences, The Netherlands
Heiko Schuldt	University of Basel, Switzerland
Giuseppe Serra	University of Udine, Italy
Markus Stocker	German National Library of Science and Technology (TIB), Germany
Shigeo Sugimoto	University of Tsukuba, Japan
Hussein Suleman	University of Cape Town, South Africa
Atsuhiro Takasu	National Institute of Informatics, Japan
Anna Maria Tammaro	University of Parma, Italy
Carla Teixeira Lopes	University of Porto, Portugal
Francesca Tomasi	University of Bologna, Italy
Pertti Vakkari	Tampere University, Finland
Marieke Van Erp	KNAW Humanities Cluster, The Netherlands
Thanasis Vergoulis	Athena Research Center, Greece
Marco Viviani	University of Milano-Bicocca, Italy
Stefanos Vrochidis	Centre for Research and Technology Hellas, Greece
David Walsh	Edge Hill University, UK
Michele Weigle	Old Dominion University, USA

Short Program Committee

Hamed Alhoori	Northern Illinois University, USA
Robert Allen	Independent, USA
Vangelis Banos	Aristotle University of Thessaloniki, Greece
Gerd Berget	Oslo Metropolitan University, Norway
José Borbinha	Universidade de Lisboa, Portugal
Maria Manuel Borges	University of Coimbra, Portugal
Vittore Casarosa	ISTI-CNR, Italy
Songphan Choemprayong	Chulalongkorn University, Thailand
Mickal Coustaty	Laboratoire L3i - La Rochelle Université, France
Theodore Dalamagas	Athena Research Center, Greece
Tahereh Dehdarirad	Chalmers University of Technology, Sweden
Boris Dobrov	Moscow State University, Russia
Shyamala Doraisamy	Universiti Putra Malaysia, Malaysia
Fabien Duchateau	Université Claude Bernard Lyon 1 - LIRIS, France
Ralph Ewerth	L3S Research Center, Leibniz Universität Hannover, Germany
Edward Fox	Virginia Tech, USA
Nuno Freire	INESC-ID, Portugal
Maria Gaede	Humboldt-Universität zu Berlin, Germany
Manolis Gergatsoulis	Ionian University, Greece
C. Lee Giles	Pennsylvania State University, USA
Matthias Hagen	Martin-Luther-Universität Halle-Wittenberg, Germany
Mark Hall	The Open University, UK
Morgan Harvey	University of Sheffield, UK
Andreas Henrich	University of Bamberg, Germany
Nikos Housos	University of Patras, Greece
Antoine Isaac	Europeana and VU Amsterdam, The Netherlands
Robert Jäschke	Humboldt-Universität zu Berlin, Germany
Roman Kern	Graz University of Technology, Austria
Martin Klein	Los Alamos National Laboratory, USA
Petr Knoth	The Open University, UK
Stefanos Kollias	National Technical University of Athens, Greece
Marijn Koolen	Royal Netherlands Academy of Arts and Sciences, The Netherlands
Laszlo Kovacs	Institute for Computer Science and Control (SZTAKI), Hungary
Suzanne Little	Dublin City University, Ireland
Ying-Hsang Liu	Oslo Metropolitan University, Norway
Clifford Lynch	CNI, USA
Elena Maceviciute	University of Boras, Sweden

Yannis Manolopoulos	Open University of Cyprus, Cyprus
Bruno Martins	INESC-ID and University of Lisbon, Portugal
Philipp Mayr	GESIS, Germany
Robert H. McDonald	University of Colorado Boulder, USA
Dana Mckay	RMIT University, Australia
Tanja Merun	University of Ljubljana, Slovenia
Andrs Micsik	Institute for Computer Science and Control (SZTAKI), Hungary
Agnieszka Mykowiecka	Institute of Computer Science, Polish Academy of Sciences, Poland
Michael Nelson	Old Dominion University, USA
Erich Neuhold	University of Vienna, Austria
David Nichols	University of Waikato, New Zealand
Kjetil Nørvåg	Norwegian University of Science and Technology, Norway
Christos Papatheodorou	National and Kapodistrian University of Athens, Greece
Vivien Petras	Humboldt-Universität zu Berlin, Germany
Dimitris Plexousakis	FORTH-ICS, Greece
Edie Rasmussen	University of British Columbia, Canada
Cristina Ribeiro	University of Porto, Portugal
Thomas Risse	University of Frankfurt, Germany
Irene Rodrigues	Universidade de Evora, Portugal
Heiko Schuldt	University of Basel, Switzerland
Michalis Sfakakis	Ionian University, Greece
Marc Spaniol	Université de Caen Normandie, France
Cyrille Suire	Université Paris-Saclay and UVSQ, France
Chrisa Tsinaraki	European Commission - Joint Research Center, Italy
Douglas Tudhope	University of South Wales, UK
Yannis Tzitzikas	University of Crete and FORTH-ICS, Greece
Pertti Vakkari	Tampere University, Finland
Stefanos Vrochidis	ITI-CERTH, Greece
Michele Weigle	Old Dominion University, USA
Maja Žumer	University of Ljubljana, Slovenia

Additional Reviewers

Christian Otto	Aman Ahuja
David Pride	Michalis Mountantonakis
Satvik Chekuri	Eleftherios Kalogeros
Filip Kovacevic	Prashant Chandrasekar
Alexandros Vassiliades	Junaid Ghauri

David Pride
Graziella De Martino
Vivek Mohato
Suchetha Nambanoor Kunnath
Ivan Heibi
Maria Rousi
Paris Koloveas
Theoni Spathi
Patricia Martín-Chozas
Beatriz Esteves
Alexandros Kokkalas
Athanasios Mavropoulos
Cristian Camilo Berro Aroca
Ornella Irrera
Hussam Hallak
Elias Entrup
Tom Alby
Eleftherios Kalogeros
Silvia Piccini
Federica Vezzani
Pavlos Fafalios
Angelo Antonio Salatino

Democratizing Data Access: What if We Could Just Talk to Our Data?

Georgia Koutrika
Athena Research Center, Greece

georgia@athenarc.gr

Abstract. Data is considered the 21st century's most valuable commodity. Analysts exploring data sets for insight, scientists looking for patterns, and consumers looking for information are just a few examples of user groups that need to access and dig into data. However, existing data access and exploration systems are still following two classical paradigms: they are either form-based interfaces that allow users to easily query the data but with limited query capabilities, or they are low-level programmatic tools that allow the users to write their own queries using programming or query languages such as Python or SQL, but they are intended for CS experts.

On the other hand, enabling users to query data in a database using natural language has been considered the holy grail of the database community. In 1974, E. F. Codd was underlining this need: "If we are to satisfy the needs of casual users of databases, we must break the barriers that presently prevent these users from freely employing their native language" [2]. Several natural language interfaces for databases emerged since then; recent approaches are using deep learning techniques with promising results but also important limitations [1, 3].

While these systems are taking us closer to the holy grail of "talking to data", to bridge the gap between data and humans, we envision a system that allows users to dig into data and find answers by conversing and collaborating with the system, as if it were a human. This system would enable interaction using natural language, would understand the data as well as the user intent, would guide the user, and make suggestions, and help the user find answers to questions over data in a more natural way.

These *intelligent data assistants* are much more complex than any "digital assistant" we know of today, which are based on pre-specified dialogues and can answer simple questions. They require the synergy of several technologies and innovation in all these fronts, including natural language interfaces, data exploration, conversational AI, and data management.

The intention of this talk is to give a flavor of an exciting research territory, discuss where we stand today, opportunities and challenges.

Short Bio

Georgia Koutrika is a Research Director at Athena Research Center in Greece. She has more than 15 years of experience in multiple roles at HP Labs, IBM Almaden, and Stanford. Her work emerges at the intersection of data management, natural language processing and deep learning, and focuses on intelligent and interactive data exploration, conversational data systems, and user-driven data management.

Her work has been incorporated in commercial products, described in 14 granted patents and 26 patent applications in the US and worldwide, and published in more than 100 papers in top-tier conferences and journals. Georgia is an ACM Senior Member and IEEE Senior Member.

She is a member of the VLDB Endowment Board of Trustees, member of the PVLDB Advisory Board, member of the ACM-RAISE Working Group, co-Editor-in-chief for VLDB Journal, PC co-chair for VLDB 2023, co-EiC of Proceedings of VLDB (PVLDB). She has been associate editor in top-tier conferences (such as ACM SIGMOD, VLDB) and journals (VLDB Journal, IEEE TKDE), and she has been in the organizing committee of several conferences including SIGMOD, ICDE, EDBT, among others. She has received a PhD and a diploma in Computer Science from the Department of Informatics and Telecommunications, University of Athens, Greece.

References

1. Affolter, K., Stockinger, K., Bernstein, A.: A comparative survey of recent natural language interfaces for databases. VLDB J. **28**(5), 793–819 (2019). https://doi.org/10.1007/s00778-019-00567-8
2. Codd, E.F.: Seven steps to rendezvous with the casual user. In: Klimbie, J.W., Koffeman, K.L. (eds.) Data Base Management, Proceeding of the IFIP Working Conference Data Base Management, Cargèse, Corsica, France, April 1–5, 1974, pp. 179–200. North-Holland, Jan. (1974)
3. Katsogiannis-Meimarakis, G., Koutrika, G.: A deep dive into deep learning approaches for Text-to-SQL systems. In: Proceedings of the 2021 International Conference on Management of Data, SIGMOD 2021, pp. 2846–2851. Association for Computing Machinery, New York (2021)

Contents

Keynote Talks

Should We Preserve the World's Software History, And Can We? 3
Roberto Di Cosmo

Full Papers

Analyzing the Web: Are Top Websites Lists a Good Choice for Research? 11
Tom Alby and Robert Jäschke

RDFtex: Knowledge Exchange Between LaTeX-Based Research Publications and Scientific Knowledge Graphs 26
Leon Martin and Andreas Henrich

Enriching the Greek National Cultural Aggregator with Key Figures in Greek History and Culture: Challenges, Methodology, Tools and Outputs . . . 39
Haris Georgiadis, Agathi Papanoti, Elena Lagoudi, Georgia Angelaki, Nikos Vasilogamvrakis, Alexia Panagopoulou, and Evi Sachini

Figure and Figure Caption Extraction for Mixed Raster and Vector PDFs: Digitization of Astronomical Literature with OCR Features 52
J. P. Naiman, Peter K. G. Williams, and Alyssa Goodman

FAIROs: Towards FAIR Assessment in Research Objects 68
Esteban González, Alejandro Benítez, and Daniel Garijo

Searching Wartime Photograph Archive for Serious Leisure Purposes 81
Sanna Kumpulainen and Hille Ruotsalainen

Implementation and Evaluation of a Multilingual Search Pilot in the Europeana Digital Library 93
Mónica Marrero and Antoine Isaac

DETEXA: Declarative Extensible Text Exploration and Analysis 107
Yannis Foufoulas, Eleni Zacharia, Harry Dimitropoulos, Natalia Manola, and Yannis Ioannidis

The Way We Cite: Common Metadata Used Across Disciplines for Defining Bibliographic References 120
Erika Alves dos Santos, Silvio Peroni, and Marcos Luiz Mucheroni

Event Notifications in Value-Adding Networks 133
Patrick Hochstenbach, Herbert Van de Sompel, Miel Vander Sande, Ruben Dedecker, and Ruben Verborgh

A Chromium-Based Memento-Aware Web Browser 147
Abby Mabe, Michael L. Nelson, and Michele C. Weigle

Analysis of the Deletions of DOIs: What Factors Undermine Their Persistence and to What Extent? 161
Jiro Kikkawa, Masao Takaku, and Fuyuki Yoshikane

Analysing User Involvement in Open Government Data Initiatives 175
Dagoberto Jose Herrera-Murillo, Abdul Aziz, Javier Nogueras-Iso, and Francisco J. Lopez-Pellicer

The Rise of GitHub in Scholarly Publications 187
Emily Escamilla, Martin Klein, Talya Cooper, Vicky Rampin, Michele C. Weigle, and Michael L. Nelson

Whois? Deep Author Name Disambiguation Using Bibliographic Data 201
Zeyd Boukhers and Nagaraj Bahubali Asundi

Investigations on Meta Review Generation from Peer Review Texts Leveraging Relevant Sub-tasks in the Peer Review Pipeline 216
Asheesh Kumar, Tirthankar Ghosal, Saprativa Bhattacharjee, and Asif Ekbal

Improving Accessibility to Arabic ETDs Using Automatic Classification 230
Eman Abdelrahman and Edward Fox

Short Papers

Robots Still Outnumber Humans in Web Archives, But Less Than Before 245
Himarsha R. Jayanetti, Kritika Garg, Sawood Alam, Michael L. Nelson, and Michele C. Weigle

Cui Bono? Cumulative Advantage in Open Access Publishing 260
David Pride, Matteo Cancellieri, and Petr Knoth

Overview Visualizations for Large Digitized Correspondence Collections: A Design Study 266
Laura Swietlicki and Pierre Cubaud

Developing the EOSC-Pillar RDM Training and Support Catalogue 274
Paula Oset Garcia, Lisana Berberi, Leonardo Candela, Inge Van Nieuwerburgh, Emma Lazzeri, and Marie Czuray

Automatic Generation of Coherent Image Galleries in Virtual Reality 282
Simon Peterhans, Loris Sauter, Florian Spiess, and Heiko Schuldt

Extracting Funder Information from Scientific Papers - Experiences with Question Answering 289
Timo Borst, Jonas Mielck, Matthias Nannt, and Wolfgang Riese

CDX Summary: Web Archival Collection Insights 297
Sawood Alam and Mark Graham

"Knock Knock! Who's There?" A Study on Scholarly Repositories' Availability 306
Andrea Mannocci, Miriam Baglioni, and Paolo Manghi

Exploring Research Fields Through Institutional Contributions to Academic Journals 313
Tove Faber Frandsen and Jeppe Nicolaisen

Implementation Issues for a Highly Structured Research Report 320
Robert B. Allen

MINE – Workspace as a Service for Text Analysis 328
Triet Ho Anh Doan, Péter Király, and Sven Bingert

Conducting the Opera: The Evolution of the RDA Work to the Share-VDE Opus and BIBFRAME Hub 335
Ian Bigelow

PH-Remix Prototype: A Non Relational Approach for Exploring AI-Generated Content in Audiovisual Archives 344
Chiara Mannari, Davide Italo Serramazza, and Enrica Salvatori

The SSH Data Citation Service, A Tool to Explore and Collect Citation Metadata 351
Cesare Concordia, Nicolas Larrousse, and Edward Gray

B!SON: A Tool for Open Access Journal Recommendation 357
Elias Entrup, Anita Eppelin, Ralph Ewerth, Josephine Hartwig, Marco Tullney, Michael Wohlgemuth, and Anett Hoppe

Visual Web Archive Quality Assessment 365
Theresa Elstner, Johannes Kiesel, Lars Meyer, Max Martius, Sebastian Schmidt, Benno Stein, and Martin Potthast

The Knowledge Trust: A Proposal for a Blockchain Consortium for Digital Archives 372
Harry Halpin

Enabling Portability and Reusability of Open Science Infrastructures 379
Giuseppe Grieco, Ivan Heibi, Arcangelo Massari, Arianna Moretti, and Silvio Peroni

Simulating User Querying Behavior Using Embedding Space Alignment 386
Saber Zerhoudi and Michael Granitzer

A Closer Look into Collaborative Publishing at Software-Engineering Conferences 395
Rand Alchokr, Jacob Krüger, Yusra Shakeel, Gunter Saake, and Thomas Leich

Design and Evaluation of a Mobile Application for an Italian UNESCO Site: *Padova Urbs Picta* 403
Daniel Zilio and Nicola Orio

Ontology-Based Metadata Integration for Oral History Interviews 410
Maria Vrachliotou and Christos Papatheodorou

Making FAIR Practices Accessible and Attractive 417
Lyudmila Balakireva and Fedor Balakirev

Structured References from PDF Articles: Assessing the Tools for Bibliographic Reference Extraction and Parsing 425
Alessia Cioffi and Silvio Peroni

On Dimensions of Plausibility for Narrative Information Access to Digital Libraries 433
Hermann Kroll, Niklas Mainzer, and Wolf-Tilo Balke

Early Experiments on Automatic Annotation of Portuguese Medieval Texts 442
Maria Inës Bico, Jorge Baptista, Fernando Batista, and Esperança Cardeira

Creating Structure in Web Archives with Collections: Different Concepts from Web Archivists 450
Himarsha R. Jayanetti, Shawn M. Jones, Martin Klein, Alex Osbourne, Paul Koerbin, Michael L. Nelson, and Michele C. Weigle

Weighted Altmetric Scores to Facilitate Literature Analyses 459
Yusra Shakeel, Abhisar Bharti, Thomas Leich, and Gunter Saake

Accelerating Innovation Papers

Exploring LSTMs for Simulating Search Sessions in Digital Libraries 469
Sebastian Günther, Paul Göttert, and Matthias Hagen

Got 404s? Crawling and Analyzing an Institution's Web Domain 474
Martin Klein and Lyudmila Balakireva

Automatic Knowledge Extraction from a Digital Library and Collaborative Validation 480
Eleonora Bernasconi, Miguel Ceriani, Massimo Mecella, and Alberto Morvillo

Holistic Graph-Based Representation and AI for Digital Library Management 485
Stefano Ferilli

Optimizing the Digitization Workflow of Heritage Institutions to Increase Quality of Digitized Texts 490
Mirjam Cuper and Sarah D'Huys

Mapping STI Ecosystems via Open Data: Overcoming the Limitations of Conflicting Taxonomies. A Case Study for Climate Change Research in Denmark 495
Nicandro Bovenzi, Nicolau Duran-Silva, Francesco Alessandro Massucci, Francesco Multari, and Josep Pujol-Llatse

BetterPR: A Dataset for Estimating the Constructiveness of Peer Review Comments 500
Prabhat Kumar Bharti, Tirthankar Ghosal, Mayank Agarwal, and Asif Ekbal

BookSampo Fiction Literature Knowledge Graph Revisited: Building a Faceted Search Interface with Seamlessly Integrated Data-Analytic Tools 506
Eero Hyvönen, Annastiina Ahola, and Esko Ikkala

Solutions for Data Sharing and Storage: A Comparative Analysis of Data Repositories 512
Joana Rodrigues and Carla Teixeira Lopes

LOD that Picture: Leveraging Linked Open Data to Enhance Navigation of the Photographic Archive of the Teatro Comunale of Bologna 518
Andrea Schimmenti and Paolo Bonora

MedSeer: A Medical Controversial Information Retrieval System Based on Credible Sources 523
Dina Sayed, Mohamed Noureldin, and Heiko Schuldt

Referency: Harmonizing Citations in Transdisciplinary Scholarly Literature 528
Elisa Bastianello, Alessandro Adamou, and Nikos Minadakis

A Study of the Grammacographical Textual Genre Across Time 533
Silvia Muzzupappa and Giorgio Maria Di Nunzio

A Methodology for the Management of Contact Languages Data. The Case Study of the Jews of Corfu 538
Georgios Vardakis and Giorgio Maria Di Nunzio

SurvAnnT: Facilitating Community-Led Scientific Surveys and Annotations 543
Anargiros Tzerefos, Ilias Kanellos, Serafeim Chatzopoulos, Theodore Dalamagas, and Thanasis Vergoulis

Author Index 549