

## 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*

Gerhard Woeginger 

*RWTH Aachen, Aachen, Germany*

Moti Yung

*Columbia University, New York, NY, USA*

More information about this subseries at <http://www.springer.com/series/7409>

Gerd Berget · Mark Michael Hall ·  
Daniel Brenn · Sanna Kumpulainen (Eds.)


# Linking Theory and Practice of Digital Libraries


25th International Conference on Theory and Practice  
of Digital Libraries, TPDL 2021  
Virtual Event, September 13–17, 2021  
Proceedings

### *Editors*

Gerd Berget   
OsloMet – Oslo Metropolitan University  
Oslo, Norway

Daniel Brenn  
Martin Luther University Halle-Wittenberg  
Halle, Germany

Mark Michael Hall   
The Open University  
Milton Keynes, UK

Sanna Kumpulainen   
Tampere University  
Tampere, Finland

ISSN 0302-9743

ISSN 1611-3349 (electronic)

Lecture Notes in Computer Science

ISBN 978-3-030-86323-4

ISBN 978-3-030-86324-1 (eBook)

<https://doi.org/10.1007/978-3-030-86324-1>

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

© Springer Nature Switzerland AG 2021

Chapters “A Semantic Search Engine for Historical Handwritten Document Images” and “Self-assessment and Monitoring of CHI Performance in Digital Transformation” 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

We are happy and proud to present this volume of proceedings, which contains the accepted papers presented at the 25th International Conference on Theory and Practice of Digital Libraries (TPDL 2021), which was held online at The Open University, UK, during September 13–17, 2021.

This year TPDL celebrated its 25th edition as a leading scientific forum that brings together both practitioners and researchers in the field of digital libraries and related fields. This edition of TPDL was held under the general theme of “Linking Theory and Practice”, to highlight TPDL’s emphasis on including both researchers and practitioners and on providing a forum for productive exchange between the groups. To further strengthen this, this year’s edition introduced the full-length practitioner paper aimed at allowing the presentation of high-quality, applied work. The interest in and success of this category was clearly shown by its uptake: 27% of the full-length submissions and 30% of the accepted full-length papers were practitioner papers. We hope that this strengthens links between the groups and leads to the inclusion of more applied research in future editions of TPDL.

TPDL 2021 received a total of 53 submissions, divided into 27 research, 10 practitioner, 9 short, 3 poster, and 4 demo papers. The reviewing process was double-blind, with at least three reviews per paper, combined with a meta-review step. Based on the outcome of this process a total of 16 papers were accepted (30%), divided into 7 research (26%), 3 practitioner (30%), 3 short (33%), 1 poster (33%), and 2 demo papers (50%). Additionally, based on the reviewers’ suggestions, we offered 10 papers the opportunity to submit a shortened poster paper, increasing the overall number of accepted papers to 26 (49%). In this respect, the overall acceptance rate is in line with past editions (generally around 45%).

The conference welcomed one keynote speaker: Ulrike Wuttke from the University of Applied Sciences Potsdam, Germany. In keeping with the general theme of “Linking Theory and Practice”, her work on digital humanities methods and tools, research infrastructures, eResearch and research data management, open access and open science, and book and library history perfectly encapsulated the conference’s collaborative and interdisciplinary focus.

In addition to the main conference, the program included a pre-conference workshop on “LinkedArchives – International Workshop on Archives and Linked Data” and a Doctoral Consortium. To enable the conference’s social aspect, making links between researchers and practitioners, between disciplines, and between new and experienced participants, the conference used a technical solution to allow for dynamic, between-session chats and discussions amongst participants.

This conference was only possible because of all the work put in by many people. First, we want to thank all researchers for submitting their work and the members of our Program Committee for their reviews. Thank you also to our Workshop chairs, Dana McKay and Philipp Mayr-Schlegel, and to our Doctoral Consortium chairs,

Morgan Harvey and Helena Francke, for running their tracks so smoothly. Finally, we also express our gratitude to the TPDL Steering Committee, who supported us in experimenting with the online format.

September 2021

Gerd Berget  
Mark Michael Hall  
Daniel Brenn  
Sanna Kumpulainen

# Organization

## General Chair

Mark Michael Hall                      The Open University, UK

## Program Committee Chairs

Gerd Berget                      Oslo Metropolitan University, Norway  
Daniel Brenn                      Martin-Luther-Universität Halle-Wittenberg, Germany  
Sanna Kumpulainen              Tampere University, Finland

## Workshop/Tutorial Chairs

Dana McKay                      University of Melbourne, Australia  
Philipp Mayr-Schlegel              GESIS - Leibniz Institute for the Social Sciences,  
   Germany

## Doctoral Consortium Chairs

Morgan Harvey                      Sheffield University, UK  
Helena Francke                      University of Gothenburg, Sweden

## Steering Committee Chair

Trond Aalberg                      OsloMet, Norway

## Steering Committee

Trond Aalberg	OsloMet, Norway
Bolette Ammitzboll	Jurik State and University Library, Denmark
George Buchanan	The University of Melbourne, Australia
Fabien Duchateau	Université Lyon 1, France
Vittore Casarosa	ISTI-CNR, Italy
Milena Dobрева	University of Malta, Malta
Mark Michael Hall	The Open University, UK
Lazaros Iliadis	Democritus University of Thrace, Greece
Jaap Kamps	University of Amsterdam, The Netherlands
Sarantos Kapidakis	Ionian University, Greece
Laszlo Kovacs	MTA SZTAKI, Hungary
Tanja Merčun	University of Ljubljana, Slovenia
Wolfgang Neidl	L3S Research Center, Germany
Yannis Manolopoulos	Aristotle University of Thessaloniki, Greece

Cezary Mazurek	Poznań Supercomputing and Networking Center, Poland
Christos Papatheodorou	Ionian University, Greece
Edie Rasmussen	University of British Columbia, Canada
Andreas Rauber	Technical University of Wien, Austria
Thomas Risse	Goethe University Frankfurt - University Library J. C. Senckenberg, Germany
Heiko Scholdt	University of Basel, Switzerland
Ingeborg Solvberg	Norwegian University of Technology and Science, Norway
Marcin Werla	Qatar National Library, Qatar

## Program Committee

Maristella Agosti	University of Padua, Italy
Hamed Alhoori	Northern Illinois University, USA
Robert Allen	Independent Researcher
Vangelis Banos	Aristotle University of Thessaloniki, Greece
José Borbinha	Universidade de Lisboa, Portugal
Maria Manuel Borges	University of Coimbra, Portugal
George Buchanan	The University of Melbourne, Australia
Ricardo Campos	Ci2 - Polytechnic Institute of Tomar, Portugal
Vittore Casarosa	ISTI-CNR, Italy
Songphan Choemprayong	Culanlongkorn University, Thailand
Mickaël Coustaty	University of La Rochelle, France
Theodore Dalmagas	ATHENA Research Centre, Greece
Tahereh Dehdarirad	Chalmers University of Technology, Sweden
Elena Demidova	Bonn University, Germany
Boris Dobrov	Moscow State University, Russia
Shyamala Doraisamy	University Putra, Malaysia
Antoine Doucet	University of La Rochelle, France
Fabien Duchateau	Université Claude Bernard Lyon 1, France
Ralph Ewerth	Leibniz Universität Hannover, Germany
Edward Fox	Virginia Tech, USA
Nuno Freire	INESC-ID, Portugal
Ingo Frommholz	University of Wolverhampton, UK
Maria Gäde	Humboldt-Universität zu Berlin, Germany
Manolis Gergatsoulis	Ionian University, Greece
C. Lee Giles	Pennsylvania State University, USA
Koraljka Golub	Linnaeus University, Sweden
Sergiu Gordea	Austrian Institute of Technology, Austria
Matthias Hagen	Martin-Luther-Universität Halle-Wittenberg, Germany
Andreas Henrich	University of Bamberg, Germany
Nikos Housos	IRI, Greece
Antoine Isaac	Europeana and VU University Amsterdam, The Netherlands



Robert Jäschke	Humboldt-Universität zu Berlin, Germany
Roman Kern	Graz University of Technology, Austria
Heikki Keskustalo	Tampere University, Finland
Martin Klein	Los Alamos National Laboratory, USA
Petr Knoth	The Open University, UK
Stefanos Kollias	Independent Researcher, Greece
Marijn Koolen	Royal Netherlands Academy of Arts and Sciences, The Netherlands
Laura Korkeamäki	Tampere University, Finland
Laszlo Kovacs	Institute for Computer Science and Control, Hungary
Elina Late	Tampere University, Finland
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	University of Lisbon, Portugal
Cezary Mazurek	Poznan Supercomputing and Networking Center, Poland
Robert H. McDonald	University of Colorado Boulder, USA
Tanja Merčun	University of Ljubljana, Slovenia
András Micsik	Institute for Computer Science and Control, Hungary
Jean-Philippe Moreux	Bibliothèque nationale de France, France
Agnieszka Mykowiecka	IPI PAN, Poland
Michael Nelson	Old Dominion University, USA
Erich Neuhold	University of Vienna, Austria
David Nichols	University of Waikato, New Zealand
Kjetil Nørvg	Norwegian University of Science and Technology, Norway
Christos Papatheodorou	Ionian University, Greece
Vivien Petras	Humboldt-Universität zu Berlin, Germany
Nils Pharo	Oslo Metropolitan University, Norway
Dimitris Plexousakis	FORTH, Greece
Edie Rasmussen	University of British Columbia, Canada
Andreas Rauber	Vienna University of Technology, Austria
Christina Ribeiro	University of Porto, Portugal
Thomas Risse	Goethe University Frankfurt, Germany
João Rocha Da Silva	University of Porto, Portugal
Irene Rodrigues	Universidade de Evora, Portugal
Heiko Scholdt	University of Basel, Switzerland
Michalis Sfakakis	Ionian University, Greece
Marc Spaniol	Université de Caen Normandie, France
Shigeo Sugimoto	University of Tsukuba, Japan
Cyrille Suire	Université Paris-Saclay, France
Hussein Suleman	University of Cape Town, South Africa
Atsuhiko Takasu	National Institute of Informatics, Japan

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	Information Technologies Institute, Greece
David Walsh	Edge Hill University, UK
Michele Weigle	Old Dominion University, USA
Iris Xie	University of Wisconsin-Milwaukee, USA
Maja Žumer	University of Ljubljana, Slovenia

## **Additional Reviewers**

Frederik Arnold	Himarsha Jayanetti
Matthew Damigos	Eleftherios Kalogeros
Vasilis Efthymiou	Dimos Ntioudis
Pavlos Fafalios	David Pride
Kritika Garg	Maria Rousi
Yasin Ghafourian	Spyridon Symeonidis

# Keynote Speaker

Ulrike Wuttke 

University of Applied Sciences Potsdam  
<https://ulrikewuttke.wordpress.com/>

Ulrike Wuttke has been the Interim Professor for “Library and Information Technology and Digital Services” at the Department of Information Sciences at the University of Applied Sciences Potsdam (FHP), Germany, since 2020. She received her PhD (Doctor of Literature) from Ghent University in 2012. Her teaching and research interests include digital research infrastructures, research data management, digital humanities, scholarly communication, and open science. She contributes to various (inter-)national networks and working groups in these areas. She has been task leader within the PARTHENOS project (FHP), scientific coordinator of the AGATE-project (Union of the German Academies of Sciences and Humanities), worked for the Humanities Data Center (Göttingen Academy of Sciences and Humanities) and the Göttingen eResearch Alliance (Göttingen State and University Library).

# Contents

## Document and Text Analysis

FETD <sup>2</sup> : A Framework for Enabling Textual Data Denoising via Robust Contextual Embeddings . . . . .	3
<i>Govind, Céline Alec, Jean-Luc Manguin, and Marc Spaniol</i>	
Minimalist Fitted Bayesian Classifier-Based on Likelihood Estimations and Bag-of-Words. . . . .	17
<i>Jean-Rémi Bourguet, Wesley Silva, and Elias de Oliveira</i>	
Inventory and Content Separation in Grammatical Descriptions of Languages of the World. . . . .	29
<i>Harald Hammarström</i>	
An Empirical Study of Span Modeling in Science NER. . . . .	41
<i>Xiaorui Jiang</i>	
Terminology/Keyphrase Extraction for Creation of Book Indexes in Polish . . . . .	49
<i>Małgorzata Marciniak, Agnieszka Mykowiecka, and Piotr Rychlik</i>	
Token-Level Multilingual Epidemic Dataset for Event Extraction . . . . .	55
<i>Stephen Mutuvi, Emanuela Boros, Antoine Doucet, Gaël Lejeune, Adam Jatowt, and Moses Odeo</i>	
A Semantic Search Engine for Historical Handwritten Document Images . . . .	60
<i>Vuong M. Ngo, Gary Munnelly, Fabrizio Orlandi, Peter Crooks, Declan O'Sullivan, and Owen Conlan</i>	
Temporal Analysis of Worldwide War. . . . .	66
<i>Devansh Bajpai and Rishi Ranjan Singh</i>	

## Data Repositories and Archives

Where Did the Web Archive Go? . . . . .	73
<i>Mohamed Aturban, Michael L. Nelson, and Michele C. Weigle</i>	
What's Data Got to Do with It? An Agenda for a New Generation of Digital Libraries . . . . .	85
<i>George Buchanan, Dana McKay, and David Bainbridge</i>	

Semantic Tagging via Entity-Level Analytics: Assessment of Concise Content Tagging . . . . .	97
<i>Amit Kumar and Marc Spaniol</i>	
Automating the Selection of Emulated Rendering Environments for Born-Digital Data-Sets . . . . .	106
<i>Julian Giessler, Rafael Gieschke, Klaus Rechert, and Euan Cochrane</i>	
Colabo.Space - Participatory Platform for Evolving Research and Publishing Workflows . . . . .	112
<i>Sasha Mile Rudan, Sinisha Rudan, Eugenia Kelbert, Andrija Sagic, Lazar Kovacevic, and Matthew Reynolds</i>	
How Can an Archive Be Characterized? . . . . .	118
<i>Marta Faria Araújo and Carla Teixeira Lopes</i>	
Visualizing Copyright-Protected Video Archive Content Through Similarity Search. . . . .	123
<i>Kader Pustu-Iren, Eric Müller-Budack, Sherzod Hakimov, and Ralph Ewerth</i>	
Self-assessment and Monitoring of CHI Performance in Digital Transformation . . . . .	128
<i>Rasa Bocyte, Johan Oomen, and Fred Truyen</i>	
Automatic Translation and Multilingual Cultural Heritage Retrieval: A Case Study with Transcriptions in Europeana . . . . .	133
<i>Mónica Marrero, Antoine Isaac, and Nuno Freire</i>	
<b>Linked Data and Open Data</b>	
Leveraging a Federation of Knowledge Graphs to Improve Faceted Search in Digital Libraries . . . . .	141
<i>Golsa Heidari, Ahmad Ramadan, Markus Stocker, and Sören Auer</i>	
A Comprehensive Extraction of Relevant Real-World-Event Qualifiers for Semantic Search Engines . . . . .	153
<i>Guillaume Bernard, Cyrille Suire, Cyril Faucher, and Antoine Doucet</i>	
Citation Recommendation for Research Papers via Knowledge Graphs . . . . .	165
<i>Arthur Brack, Anett Hoppe, and Ralph Ewerth</i>	
AnnoTag: Concise Content Annotation via LOD Tags derived from Entity-Level Analytics . . . . .	175
<i>Amit Kumar and Marc Spaniol</i>	
SmartReviews: Towards Human- and Machine-Actionable Reviews . . . . .	181
<i>Allard Oelen, Markus Stocker, and Sören Auer</i>	

## User Interfaces and Experience

Comparing Methods for Finding Search Sessions on a Specified Topic: A Double Case Study . . . . .	189
<i>Tessel Bogaard, Aysenur Bilgin, Jan Wielemaker, Laura Hollink, Kees Ribbens, and Jacco van Ossenbruggen</i>	
Clustering and Classifying Users from the National Museums Liverpool Website . . . . .	202
<i>David Walsh, Paul Clough, Mark Michael Hall, Frank Hopfgartner, and Jonathan Foster</i>	
Humanities Scholars and Digital Humanities Projects: Practice Barriers in Tools Usage . . . . .	215
<i>Rui Liu, Dana McKay, and George Buchanan</i>	
Researching Pandemics Through Time: A Covid-19 Inspired Data-Driven Approach to Explore Historical Newspapers . . . . .	227
<i>Mirjam Cuper</i>	
<b>Author Index . . . . .</b>	<b>233</b>