# Lecture Notes in Computer Science

7489

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

#### **Editorial Board**

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich. Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

Panayiotis Zaphiris George Buchanan Edie Rasmussen Fernando Loizides (Eds.)

# Theory and Practice of Digital Libraries

Second International Conference, TPDL 2012 Paphos, Cyprus, September 23-27, 2012 Proceedings



#### Volume Editors

Panayiotis Zaphiris
Fernando Loizides
Cyprus University of Technology
Department of Multimedia and Graphic Arts
3603 Limassol, Cyprus
E-mail: panayiotis.zaphiris@cut.ac.cy, fernando.loizides@gmail.com

George Buchanan City University of London School of Informatics Northampton Square, London, EC1V 0HB, UK E-mail: george.buchanan.1@city.ac.uk

Edie Rasmussen
The University of British Columbia
School of Library, Archival and Information Studies
Irving K. Barber Learning Centre, Vancouver, BC, V6T 1Z1, Canada
E-mail: edie.rasmussen@ubc.ca

ISSN 0302-9743 e-ISSN 1611-3349 ISBN 978-3-642-33289-0 e-ISBN 978-3-642-33290-6 DOI 10.1007/978-3-642-33290-6 Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012946348

CR Subject Classification (1998): H.4, H.2, H.3, H.5, J.1, H.2.8

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, 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.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

### **Preface**

We are delighted to present the proceeding of the 16th European Conference on Research and Advanced Technology for Digital Libraries, running for the second year as "Theory and Practice in Digital Libraries", which ran in Paphos, Cyprus during September 23–27, 2012.

Following the inaugural ECDL in Pisa in 1997, the conference has become established as a major international venue for leading researchers in the field of digital libraries. The global economic turmoil of the last few years has not dented the continued health of the conference. We continue to receive and accept submissions for publication from countries that have not been represented at TPDL before, as an example our host country of Cyprus is well represented with new contributors.

For TPDL 2012, we focussed on attracting contributions on four major themes:

- Applications and User Experience
- Supporting Discovery
- Digital Humanities
- Research Data

These themes are the same as those of TPDL 2011, and while the core computer-science origins of TPDL continue to mature, the approaches of interdisciplinary methods, and the interests of humanists and research scientists in systematic management of their research materials, are widening the range of challenges addressed by digital library researchers.

Our overseas keynote speakers this year represented two contrasting approaches to supporting information seeking. Mounia Lalmas is well known as a global leader in information retrieval research, and represents the technical origins and ideals of many in the DL community. Mounia's passion for excellence and robust methodology has inspired many of her peers across the world. In contrast, Cathy Marshall is immediately familiar to digital library researchers as a leading light in the more human-centred approaches to DL work, and her work has been repeatedly recognised by best-paper awards at the leading DL conferences.

We are also fortunate to have a local keynote speaker, Andreas Lanitis, whose interests are particularly relevant to the contemporary theme of preserving digital heritage. The problems in that domain have been of increasing pertinence to DLs in recent years, as the transience of digital material and the desire to disseminate ancient heritage combine to seek out the means by which long term, remote access to new and old materials can be sustained.

The programme began with the doctoral consortium and our tutorial programme of five varied tutorials, on themes from building digital libraries through to preservation processes.

Following the main conference, four workshops concluded the programme, including the 11th European Networked Knowledge Organization Systems (NKOS) workshop and the 2nd Semantic Digital Archive workshop.

For TPDL 2012, we ran a single programme committee for all submissions, and furthermore combined the call for short papers and posters into a single call for work-in-progress contributions. We continued to use the two-tier model of TPDL 2011 and ECDL 2010, with a main programme committee supervised by a metareview committee. Each contribution was reviewed independently by three or more members of the main committee, and the subsequent discussion and reflection was moderated by one member of the metareview committee.

The submissions received by the conference totalled 139 contributions (129) full papers and work-in-progress, plus and 10 demonstrations). From this, we selected 23 full papers (17.8%), 19 short paper presentations (15%), 14 poster submissions (11%), and 5 demonstrations. Twenty-six countries contributed works for review, the same as last year's conference in Berlin.

The programme of TPDL 2012 proved particularly challenging for the programme committee to finalise. The support of our metareview committee, continuing the practice of recent years at TPDL to use a second level of review, was critical in balancing the merits of the competing submissions. Our programme committee put in many hours of review and discussion, which greatly assisted us as programme chairs, and ensured that our decisions were robust and wellinformed.

Fabrizio Sebastiani, our workshop chair, and Christos Papatheodorou, serving as tutorial chair, put in particularly praiseworthy efforts to help in the success of the conference. Our other supporting chairs also provided invaluable assistance in finalising the programme.

The success of TPDL 2012 is also due to our various supporters, but in particular the Coalition for Networked Information (CNI).

September 2012

George Buchanan Edie Rasmussen Panaviotis Zaphiris Fernando Loizides

# Organisation

TPDL 2012 was organised by the Cyprus University of Technology, in association with the University of Cyprus and City University London.

## Organising Committee

Conference Chair Panayiotis Zaphiris, Cyprus University of

Technology, Cyprus

Organising Chair Fernando Loizides, Cyprus University of

Technology, Cyprus

Proceedings Chair Fernando Loizides, Cyprus University of

Technology, Cyprus

Publicity Chair Maria Poveda, University of Cyprus, Cyprus

## Programme Committee

Programme Chair George Buchanan, City University London, UK

Edie Rasmussen, University of British

Columbia, Canada

Tutorials Christos Papatheodorou, Ionian University,

Greece

Workshops Fabrizio Sebastiani, ISTI/CNR, Italy
Posters Jennifer Pearson, Swansea University, UK
Panels Rudi Schmeide, TU Darmstadt, Germany

Andri Ioannou, Cyprus University of

Technology, Cyprus

Demonstrations Trond Aalberg, Norwegian University of

Science and Technology, Norway

Doctoral Consortium Stefan Gradmann, Humboldt University Berlin,

Germany

Birger Larsen, Royal School of Library and

Information Science, Denmark

#### Metareview Committee

David Bainbridge University of Waikato, New Zealand

José Borbinha IST, Portugal Donatella Castelli ISTI-CNR, Italy

Gobindha Chowdhury University of Technology, Sydney, Australia

Fabio Crestani University of Lugano, Switzerland

J. Stephen Downie University of Illinois Urbana-Champaign, USA

Schubert Foo Nanyang Technical University, Singapore

#### VIII Organisation

Edward A. Fox Virginia Tech, USA

Ingo Frommholz University of Bedfordshire, UK

Norbert Fuhr University of Duisberg-Essen, Germany Annika Hinze University of Waikato, New Zealand Claus-Peter Klas Fern Universität Hagen, Germany

Carlo Meghini Consiglio Nazionale delle Ricerche, Italy

Michael Nelson Old Dominion University, USA Christos Papatheodorou Ionian University, Greece

Andreas Rauber Vienna University of Technology, Austria

Susan Schreibmnan Trinity College Dublin, Ireland

Hussein Suleman University of Cape Town, South Africa

Elaine Toms Sheffield University, UK Pertti Vakkari Tampere University, Finland

## Programme Committee

Trond Aalberg Norwegian University of Science and

Technology, Norway

Anne Adams Open University, UK

Robert Allen Victoria University Wellington, New Zealand

Simon Attfield Middlesex University, UK

Wolf-Tilo Balke University of Hannover, Germany Alejandro Bia Miguel Hernández University, Spain

Christine Borgman University of California Los Angeles, USA

Leonardo Candela ISTI-CNR, Italy

Les Carr Southampton University, UK
Lillian Cassel Villanova University, USA
Tinni Choudhury Middlesex University, UK

Panos Constantopoulos Athens University of Economics and Business,

Greece

Gregory Crane Tufts University, USA

Pierre Cubaud CNAM, France

Sally Jo Cunningham University of Waikato, New Zealand Theodore Dalamagas National Technical University of Athens,

Greece

Nicola Ferro University of Padua, Italy

Luis Francisco-Revilla University of Texas Austin, USA Nuno Freire The European Library, Portugal

Luanne Freund University of British Columbia, Canada

Pablo de la Fuente University of Valladolid, Spain
Richard Furuta Texas A&M University, USA
C. Lee Giles Pennsylvania State University, USA
Marcos Goncalves Federal University of Minas Gerais, Brazil

Juilo Gonzalo UNED, Spain

Stefan Gradmann Humboldt University Berlin, Germany

Kai Grønbaek Aarhus University, Denmark

Preben Hansen Maureen Henninger Geneva Henry Yannis E. Ioannidis

Max Kaiser Min-Yen Kan

Sarantos Kapidakis

Mike Khoo Martin Klein Alberto Laender

Alberto Laender Carl Lagoze

Birger Larsen

Ronald Larsen Ray Larson

Clare Llewellyn

Fernando Loizides

Andy MacFarlane

Dana McKay Akira Maeda

Stefan Makri Andras Micsik Uma Murthy

Glen Newton

David Nichols Kjetil Nørvag

Pasquale Pagano Rebecca Parker Jennifer Pearson

Stefan Rüger

Robert Sanderson Rudi Schmeide

Heiko Schuldt Michael Seadle

Frank Shipman III Ray Siemens

Beat Signer

Herbert Van De Sompel

Ingeborg Sølvberg

Nicolas Spyratos Shigeo Sugimoto Yin-Leng Theng

Anastasios Tombros Giannis Tsakonas Swedish Institute of Computer Science, Sweden University of Technology Sydney, Australia

Rice University, USA

University of Athens, Greece

Austrian National Library, Austria

National University of Singapore, Singapore

Ionian University, Greece Drexel University, USA

Los Alamos National Laboratory, USA Federal University of Minas Gerais, Brazil

Cornell University, USA

Royal School of Information and Library

Science, Denmark

University of Pittsburgh, USA

University of California, Berkeley, USA

University of Edinburgh, UK

Cyprus University of Technology, Cyprus

City University London, UK

Swinburne University of Technology, Australia

Ritsumeikan University, Japan University College London, UK

Sztaki, Hungary Virgina Tech, USA

Carlton University, Canada

University of Waikato, New Zealand Norwegian University of Science and

Technology, Norway

ISTI-CNR, Italy

Swinburne University of Technology, Australia

Swansea University, UK Open University, UK

Los Alamos National Laboratory, USA

TU Darmstadt, Germany

University of Basel, Switzerland

Humboldt University, Berlin, Germany

Texas A&M University, USA Victoria University, Canada

Vrije Universiteit Brussel, Belgium Los Alamos National Laboratory, USA Norwegian University of Science and

Technology, Norway Université Paris-Sud, France University of Tsukuba, Japan

Nanyang Technical University, Singapore

Queen Mary College London, UK University of Patras, Greece

#### X Organisation

Doug Tudhope Glamorgan University, UK Shalini Urs University of Mysore, India

Nicholas Vanderschantz University of Waikato, New Zealand

Felisa Verdejo UNED, Spain

Stina Westman University of Aalto, Finland Uffe Wiil Syddansk University, Denmark

#### Reviewers

Massimiliano Assante ISTI-CNR, Italy

Vassilis Christophides Foundation for Research and Technology

(FORTH), Greece

Gianpaolo Coro ISTI-CNR, Italy

Sujatha Das Pennsylvania State University, USA

Helen Dodd Swansea University, UK

Martin Doerr Foundation for Research and Technology -

Hellas, Greece

Laurent D'Orazio Clermont Université - Université Blaise Pascal,

France

Giorgos Giannopoulos National Technical University of Athens,

Greece

Madian Khabsa Pennsylvania State University, USA

Ivano Masiero University of Padua, Italy
Simone Peruzzo University of Padua, Italy
Hany Salaheldeen Old Dominion University, USA
Sven Schlarb Austrian National Library, Austria

Gianmaria Silvello University of Padua, Italy

Zhaohui Wu Pennsylvania State University, USA

## Workshops Committee

Fabrizio Sebastiani (Chair) ISTI/CNR, Italy

Nicola Ferro University of Padua, Italy Ed Fox Virginia Tech, USA

Marcos Goncalves Federal University of Minas Gerais, Brazil Vivien Petras Humboldt University Berlin, Germany Andreas Rauber Vienna University of Technology, Austria

#### **Tutorials Committee**

Christos Papatheodorou (chair) Ionian University, Greece

Trond Aalberg Norwegian University of Science and

Technology, Norway

Birger Larsen Royal School of Library and Information

Science, Denmark

Ron Larsen University of Pittsburgh, USA

# **Table of Contents**

# User Behaviour

What Would 'Google' Do? Users' Mental Models of a Digital Library Search Engine	1
An Exploration of ebook Selection Behavior in Academic Library	4.0
Collections	13
Information Seekers' Visual Focus during Time Constraint Document Triage	25
Which Words Do You Remember? Temporal Properties of Language Use in Digital Archives	32
Mobiles & Place	
Toward Mobile-Friendly Libraries: The Status Quo	38
Listen to Tipple: Creating a Mobile Digital Library with Location-Triggered Audio Books	51
Re-finding Physical Documents: Extending a Digital Library into a Human-Centred Workplace	57
Heritage and Sustainability	
User Needs for Enhanced Engagement with Cultural Heritage Collections	64
Christina M. Steiner, Eva-Catherine Hillemann, Micheál Ó Siochrú, and Séamus Lawless	
Digital Library Sustainability and Design Processes	76

Creation of Textual Versions of Historical Documents from Polish Digital Libraries	89
Adam Dudczak, Miłosz Kmieciak, and Marcin Werla	
Increasing Recall for Text Re-use in Historical Documents to Support Research in the Humanities	95
Marco Büchler, Gregory Crane, Maria Moritz, and Alison Babeu	
Preservation	
PrEV: Preservation Explorer and Vault for Web 2.0 User-Generated	
Content	101
Preserving Scientific Processes from Design to Publications	113
Losing My Revolution: How Many Resources Shared on Social Media Have Been Lost?	125
Automatic Vandalism Detection in Wikipedia with Active Associative Classification	138
Applying Digital Library Technologies to Nuclear Forensics  Electra Sutton, Chloe Reynolds, Fredric C. Gey, and Ray R. Larson	144
Linked Data	
Identifying References to Datasets in Publications	150
Collaborative Tagging of Art Digital Libraries: Who Should Be Tagging? A Case Study	162
A System for Exposing Linguistic Linked Open Data	173
Linking the Parliamentary Record: A New Approach to Metadata for Legislative Proceedings	179

Analysing and Enriching Documents	
A Ground Truth Bleed-Through Document Image Database	185
Identifying "Soft 404" Error Pages: Analyzing the Lexical Signatures of Documents in Distributed Collections	197
User-Defined Semantic Enrichment of Full-Text Documents:  Experiences and Lessons Learned	209
Semantic Document Selection: Historical Research on Collections That Span Multiple Centuries	215
Content and Metadata Quality	
Finding Quality Issues in SKOS Vocabularies	222
On MultiView-Based Meta-learning for Automatic Quality Assessment of Wiki Articles	234
Folksonomy and Ontology	
A Methodology for Folksonomy Evaluation	247
Advanced Automatic Mapping from Flat or Hierarchical Metadata Schemas to a Semantic Web Ontology: Requirements, Languages, Tools	260
Justyna Walkowska and Marcin Werla	
Ontological Formalization of Scientific Experiments Based on Core Scientific Metadata Model	273
Domain Analysis for a Video Game Metadata Schema:  Issues and Challenges  Jin Ha Lee, Joseph T. Tennis, and Rachel Ivy Clarke	280

# Information Retrieval

A Benchmark for Content-Based Retrieval in Bivariate Data	200
Collections	286
Web Search Personalization Using Social Data	298
A Model for Searching Musical Scores by Instrumentation	311
Extending Term Suggestion with Author Names	317
Organising Collections	
Evaluating the Use of Clustering for Automatically Organising Digital Library Collections	323
A Unique Arrangement: Organizing Collections for Digital Libraries, Archives, and Repositories	335
Mix-n-Match: Building Personal Libraries from Web Content	345
Machine Learning in Building a Collection of Computer Science Course Syllabi	357
PubLight: Managing Publications Using a Task-Oriented Approach Matthias Geel, Michael Nebeling, and Moira C. Norrie	363
Extracting and Indexing	
Improved Bibliographic Reference Parsing Based on Repeated Patterns	370
Catching the Drift – Indexing Implicit Knowledge in Chemical Digital Libraries	383
Benjamin Köhncke, Sascha Tönnies, and Wolf-Tilo Balke Using Visual Cues for the Extraction of Web Image Semantic	
Information	396

Poster Papers	
Malleable Finding Aids	402
Improving Retrieval Results with Discipline-Specific Query Expansion	408
An Evaluation System for Digital Libraries	414
Enhancing Digital Libraries and Portals with Canonical Structures for Complex Objects	420
Exploiting the Social and Semantic Web for Guided Web Archiving Thomas Risse, Stefan Dietze, Wim Peters, Katerina Doka, Yannis Stavrakas, and Pierre Senellart	426
Query Expansion of Zero-Hit Subject Searches: Using a Thesaurus in Conjunction with NLP Techniques	433
Towards Digital Repository Interoperability: The Document Indexing and Semantic Tagging Interface for Libraries (DISTIL)	439
Aggregating Content for Europeana: A Workflow to Support Content Providers	445
Diva: A Web-Based High-Resolution Digital Document Viewer  Andrew Hankinson, Wendy Liu, Laurent Pugin, and Ichiro Fujinaga	455
Collaborative Authoring of Walden's Paths	461
Quantitative Analysis of Search Sessions Enhanced by Gaze Tracking with Dynamic Areas of Interest	468
Generating Content for Digital Libraries Using an Interactive Content Management System	474

Table of Contents

XV

## XVI Table of Contents

Enhancing the Curation of Botanical Data Using Text Analysis Tools	480
Clare Llewellyn, Clare Grover, Jon Oberlander, and Elspeth Haston	100
Ranking Distributed Knowledge Repositories	486
Demonstration Papers	
The CMDI MI Search Engine: Access to Language Resources and Tools Using Heterogeneous Metadata Schemas	492
SIARD Archive Browser	496
PATHS – Exploring Digital Cultural Heritage Spaces	500
FrbrVis: An Information Visualization Approach to Presenting FRBR Work Families	504
Metadata Enrichment Services for the Europeana Digital Library	508
Collaboratively Creating a Thematic Repository Using Interactive Table-Top Technology	512
Author Index	517