

Editorial Board

Simone Diniz Junqueira Barbosa

*Pontifical Catholic University of Rio de Janeiro (PUC-Rio),
Rio de Janeiro, Brazil*

Phoebe Chen

La Trobe University, Melbourne, Australia

Alfredo Cuzzocrea

ICAR-CNR and University of Calabria, Italy

Xiaoyong Du

Renmin University of China, Beijing, China

Joaquim Filipe

Polytechnic Institute of Setúbal, Portugal

Orhun Kara

TÜBİTAK BİLGEM and Middle East Technical University, Turkey

Tai-hoon Kim

Konkuk University, Chung-ju, Chungbuk, Korea

Igor Kotenko

*St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences, Russia*

Dominik Ślęzak

University of Warsaw and Infobright, Poland

Xiaokang Yang

Shanghai Jiao Tong University, China

Maristella Agosti Floriana Esposito
Stefano Ferilli Nicola Ferro (Eds.)

Digital Libraries and Archives

8th Italian Research Conference, IRCDL 2012
Bari, Italy, February 9-10, 2012
Revised Selected Papers



Springer

Volume Editors

Maristella Agosti
Nicola Ferro
University of Padua
Department of Information Engineering
Via Gradenigo, 6/a
35131 Padua, Italy
E-mail: {agosti, ferro}@dei.unipd.it

Floriana Esposito
Stefano Ferilli
University of Bari
Department of Computer Science
Via E. Orabona, 4
70126 Bari, Italy
E-mail: {esposito, ferilli}@di.uniba.it

ISSN 1865-0929	e-ISSN 1865-0937
ISBN 978-3-642-35833-3	e-ISBN 978-3-642-35834-0
DOI 10.1007/978-3-642-35834-0	
Springer Heidelberg Dordrecht London New York	

Library of Congress Control Number: 2012954629

CR Subject Classification (1998): H.3.3-7, H.5.1, H.5.4, J.1, H.2.8, I.7.4, H.4.m

© Springer-Verlag Berlin Heidelberg 2013

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

The Italian Research Conference on Digital Libraries (IRCDL) is a yearly appointment for Italian researchers, both on the computer science and on the humanities side, interested in digital libraries and related topics. The focus of the eighth conference was on legacy and cultural heritage material. Indeed, digital library systems are becoming increasingly mature and more widely deployed. Not only do they have to grant users effective and personalized access to information, but they also now have to address the need for smoothly processing and including in the DL repositories the available legacy and cultural heritage documents, in addition to born-digital ones. This calls for the ability to deal with compound objects in different media, to provide uniform solutions and methodologies across different cultural heritage institutions, and to take into account preservation, restoration, and curation.

The IRCDL conferences were launched and initially sponsored by DELOS, an EU FP6 Network of Excellence on digital libraries, together with the Department of Information Engineering of the University of Padua. Over the years, IRCDL has become a self-sustainable event supported by the Italian digital library research community.

This volume contains the revised accepted papers selected from among those presented at the 8th Italian Research Conference on Digital Libraries (IRCDL 2012), which was held at the University of Bari ‘Aldo Moro’, Bari, February 9–10, 2012.

The aim of IRCDL is to bring together the Italian research community interested in the diversified methods and techniques that allow the building and operation of digital libraries. A national Program Committee was set up composed of 16 members, with representatives of the most active Italian research groups on digital libraries.

The papers accepted for inclusion in this volume were submitted in an extended version with respect to the papers presented orally. Those papers underwent a new review process and the results of the selection are the papers appearing in this volume. The topics covered are related to the different aspects needed to support information access and interoperability; among those there are:

- legacy documents and cultural heritage
- systems interoperability and data integration
- formal and methodological foundations of DLs
- semantic web and linked data for DLs
- multilingual information access
- DL infrastructures
- metadata creation and management
- search engines for digital library systems

- evaluation and log data
- handling audio/visual and non-traditional objects
- user interfaces and visualization
- DL quality
- policies and copyright issues in DLs
- scientific data curation, citation and scholarly publication
- user behavior and modeling
- preservation and curation.

Taking into consideration that the Italian research community is involved in different relevant projects related at large to the area of digital libraries, it was decided to report in the volume the most recent results from the diverse projects, in order to transfer them to the international community.

We would like to thank those institutions and individuals who made the conference and this volume possible:

- the Program Committee members,
- the Steering Committee members,
- the Knowledge Acquisition and Machine Learning Lab (LACAM) of the Department of Computer Science of the University of Bari ‘Aldo Moro’, and the members of the same research group who contributed to the organization of the event, namely Teresa M.A. Basile, Claudio Taranto, and Domenico Redavid, and
- the Department of Information Engineering of the University of Padua.

A Call for Participation for IRCDL 2013 will be circulated, but meanwhile we invite all researchers having research interests in digital libraries to start thinking about possible contributions to next year’s conference.

September 2012

Maristella Agosti
 Floriana Esposito
 Stefano Ferilli
 Nicola Ferro

Organization

IRCDL 2012 was organized by the Knowledge Acquisition and Machine Learning Lab (LACAM) of the Department of Computer Science of the University of Bari ‘Aldo Moro’.

General Chairs

Maristella Agosti	University of Padua
Floriana Esposito	University of Bari ‘Aldo Moro’

Program Chairs

Stefano Ferilli	University of Bari ‘Aldo Moro’
Nicola Ferro	University of Padua

Local Organization Committee

Teresa M.A. Basile	University of Bari ‘Aldo Moro’
Claudio Taranto	University of Bari ‘Aldo Moro’
Domenico Redavid	Artificial Brain S.r.l.

Program Committee

Nicola Barbuti	University of Bari
Marco Bertini	University of Florence
Francesco Buccafurri	University of Reggio Calabria
Leonardo Candela	ISTI CNR, Pisa
Michelangelo Ceci	University of Bari
Nicola Di Mauro	University of Bari
Costantino Grana	University of Modena
Maria Guercio	University of Urbino “Carlo Bo”
Simone Marinai	University of Florence
Carlo Meghini	ISTI CNR, Pisa
Nicola Orio	University of Padua
Fausto Rabitti	ISTI CNR, Pisa
Giuseppe Serra	University of Florence
Anna Maria Tammaro	University of Parma
Erilde Terenzoni	Soprintendente archivistico per il Veneto, Venice
Mario Vento	University of Salerno

IRCDL Steering Committee

Maristella Agosti	University of Padua
Tiziana Catarci	University of Rome “La Sapienza”
Alberto Del Bimbo	University of Florence
Floriana Esposito	University of Bari ‘Aldo Moro’
Carlo Tasso	University of Udine
Costantino Thanos	ISTI CNR, Pisa

Supporting Institutions

IRCDL 2012 benefited from the support of the following organizations:

Department of Computer Science, University of Bari ‘Aldo Moro’

Department of Information Engineering, University of Padua

Centro Interdipartimentale di Logica e Applicazioni, University of Bari ‘Aldo Moro’

University of Bari ‘Aldo Moro’

Ontology-Based Integration of Cultural Heritage Metadata (Invited Talk)

Christos Papatheodorou

Dept. of Archives and Library Sciences – Ionian University
72 Ioannou Theotoki str., 49100 Corfu, Greece
papatheodor@ionio.gr

Managing heterogeneous data is a challenge for cultural heritage institutions, archives, libraries, and museums which usually develop collections with heterogeneous types of material, described by different metadata schemas. For example, the Library of Congress, USA, provides EAD metadata for the archives description, MARC 21 records for the description of a wide variety of material, such as books and photographs, Text Encoding Initiative (TEI) for documenting the text of digital reproductions in the American Memory Collection, etc. The wide use of a number of cultural heritage metadata schemas imposes the development of interoperability techniques that facilitate unified access to cultural resources. One of the widely implemented techniques is the Ontology-Based Integration. Ontologies provide formal specifications of a domain's concepts and their interrelations and act as a mediated schema between heterogeneous sources.

This presentation describes an ontology-based metadata integration architecture. Its components are the mediator, which is based on the CIDOC CRM, the local sources, whose schemas are XML-based metadata schemas, and the mappings between the local sources and the mediator. Four integration scenarios are proposed on this architecture. Moreover a mapping language is demonstrated, called Mapping Description Language (MDL), to define the mappings of the metadata schemas of the local sources to the CIDOC CRM ontology as rules. An application of MDL usage to define the mapping from Encoded Archival Description (EAD) to the CIDOC CRM ontology is exhibited. Finally an algorithm for the transformation of EAD metadata to CIDOC CRM, as well as a query transformation algorithm from XPATH to CIDOC CRM are demonstrated.

Table of Contents

Panel

Experiences and Perspectives in Management for Digital Preservation of Cultural Heritage Resources (Panel)	1
<i>Maristella Agosti</i>	
Where Do Humanities Computing and Digital Libraries Meet?.....	4
<i>Dino Buzzetti</i>	
The ArchiMEDE Project for an Electronically Digitized Archive of Historical Monographs	11
<i>Onofrio Erriquez</i>	
Considerations on the Preservation of Base Digital Data of Cultural Resources	13
<i>Nicola Barbuti</i>	

Papers

Supporting Tabular Data Characterization in a Large Scale Data Infrastructure by Lexical Matching Techniques.....	21
<i>Leonardo Candela, Gianpaolo Coro, and Pasquale Pagano</i>	
Data Interoperability and Curation: The European Film Gateway Experience	33
<i>Michele Artini, Alessia Bardi, Federico Biagini, Franca Debole, Sandro La Bruzzo, Paolo Manghi, Marko Mikulicic, Pasquale Savino, and Franco Zoppi</i>	
Annotating Digital Libraries and Electronic Editions in a Collaborative and Semantic Perspective	45
<i>Michele Barbera, Federico Meschini, Christian Morbidoni, and Francesca Tomasi</i>	
Empowering Archives through Annotations.....	57
<i>Nicola Ferro and Gianmaria Silvello</i>	
Metadata Inference for Description Authoring in a Document Composition Environment	69
<i>Tsuyoshi Sugibuchi, Ly Anh Tuan, and Nicolas Spyratos</i>	
A Multi-layer Digital Library for Mediaeval Legal Manuscripts.....	81
<i>Monica Palmirani and Luca Cervone</i>	

Extracting Keyphrases from Web Pages	93
<i>Felice Ferrara and Carlo Tasso</i>	
Learning to Recognize Critical Cells in Document Tables	105
<i>Nicola Di Mauro, Stefano Ferilli, and Floriana Esposito</i>	
Document Image Understanding through Iterative Transductive Learning	117
<i>Michelangelo Ceci, Corrado Loglisci, Lucrezia Macchia, Donato Malerba, and Luciano Quercia</i>	
A Domain Based Approach to Information Retrieval in Digital Libraries	129
<i>Fulvio Rotella, Stefano Ferilli, and Fabio Leuzzi</i>	
Uncertain (Multi)Graphs for Personalization Services in Digital Libraries	141
<i>Claudio Taranto, Nicola Di Mauro, and Floriana Esposito</i>	
Improving Online Access to Archival Data	153
<i>Vittore Casarosa, Carlo Meghini, and Stanislava Gardasevic</i>	
Quick and Easy Implementation of Approximate Similarity Search with Lucene	163
<i>Giuseppe Amato, Paolo Bolettieri, Claudio Gennaro, and Fausto Rabitti</i>	
Establishing a Digital Library in Wide-Ranging University's Context: The Sapienza Digital Library Experience	172
<i>Angela Di Iorio, Marco Schaerf, and Matteo Bertazzo</i>	
Digital Curators' Education: Professional Identity vs. Convergence of LAM (Libraries, Archives, Museums)	184
<i>Anna Maria Tammara, Melody Madrid, and Vittore Casarosa</i>	
A Contribution for the Dissemination of Cultural Heritage Content to a Wider Public	195
<i>Maristella Agosti, Lucio Benfante, and Nicola Orio</i>	
Engaging the User: Elaboration and Execution of Trials with a Database of Illuminated Images	207
<i>Chiara Ponchia</i>	
Modeling Archives by Means of OAI-ORE	216
<i>Nicola Ferro and Gianmaria Silvello</i>	
Reflecting on the Europeana Data Model	228
<i>Silvio Peroni, Francesca Tomasi, and Fabio Vitali</i>	

The Europeana Linked Open Data Pilot Server	241
<i>Nicola Aloia, Cesare Concordia, and Carlo Meghini</i>	
Managing Authenticity through the Digital Resource Lifecycle	249
<i>Maria Guercio and Silvio Salza</i>	
An Innovative Character Recognition for Ancient Book and Archival Materials: A Segmentation and Self-learning Based Approach	261
<i>Nicola Barbuti and Tommaso Caldarola</i>	
Author Index	271