

Springer

Berlin

Heidelberg

New York

Barcelona

Hong Kong

London

Milan

Paris

Tokyo

Maristella Agosti Costantino Thanos (Eds.)

Research and Advanced Technology for Digital Libraries

6th European Conference, ECDL 2002
Rome, Italy, September 16-18, 2002
Proceedings



Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany
Juris Hartmanis, Cornell University, NY, USA
Jan van Leeuwen, Utrecht University, The Netherlands

Volume Editors

Maristella Agosti
Department of Information Engineering, University of Padua
Via Gradenigo 6/a, 35131 Padova, Italy
E-mail: maristella.agosti@unipd.it

Costantino Thanos
Istituto di Scienza e Tecnologie dell' Informazione (ISTI-CNR)
Area della Ricerca CNR di Pisa, Via G. Moruzzi 1, 56124 Pisa, Italy
E-mail: thanos@iei.pi.cnr.it

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Research and advanced technology for digital libraries : 6th European conference ; proceedings / ECDL 2002, Rome, Italy, September 16 - 18, 2002. Maristella Agosti ; Costantino Thanos (ed.). - Berlin ; Heidelberg ; New York ; Hong Kong ; London ; Milan ; Paris ; Tokyo : Springer, 2002 (Lecture notes in computer science ; Vol. 2458)
ISBN 3-540-44178-6

CR Subject Classification (1998): H.3.7, H.2, H.3, H.4.3, H.5, J.7, J.1, I.7

ISSN 0302-9743

ISBN 3-540-44178-6 Springer-Verlag Berlin Heidelberg New York

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-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

<http://www.springer.de>

© Springer-Verlag Berlin Heidelberg 2002
Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Stefan Sossna e.K.
Printed on acid-free paper SPIN: 10871186 06/3142 5 4 3 2 1 0

Preface

ECDL 2002 was the 6th conference in the series of European Conferences on Research and Advanced Technologies for Digital Libraries. Following previous events in Pisa (1997), Heraklion (1998), Paris (1999), Lisbon (2000), and Darmstadt (2001), this year ECDL was held in Rome. ECDL 2002 contributed, together with the previous conferences, to establishing ECDL as the major European forum focusing on digital libraries and associated technical, practical, and social issues. ECDL 2002 continued the tradition already established by the previous conferences in meeting the needs of a large and diverse constituency, which includes researchers, practitioners, educators, policy makers, and users.

The focus of ECDL 2002 was on underlying principles, methods, systems, and tools to build and make available effective digital libraries to end users. Architecture, metadata, collection building, web archiving, web technologies, e-books, OAI applications, preservation, navigation, query languages, audio video retrieval, multimedia-mixed media, user studies and evaluation, humanities, and digital libraries were some of the key issues addressed.

An international Program Committee was set up composed of 61 members, with representatives from 25 countries. A total of 145 paper submissions, 15 poster submissions, and 18 proposals for demos were received. Each paper was evaluated by 3 referees and 42 full papers and 6 short papers of high quality were selected for presentation.

Besides paper presentations, the program also featured two invited talks by Hector Garcia Molina (Stanford University) and Derek Law (Strathclyde University), two panel discussions, and a poster and demo session. Moreover, tutorials and workshops conveniently arranged immediately before and after the conference extended the range of subjects and opportunities for exchange of knowledge and opinions offered by the conference.

We should like here to thank all those institutions and individuals who have made this conference possible: the organizing committee, CNR-IEI, the University of Padua and the Department of Information Engineering of the same university, the DELOS NoE, ESA, ICCU, Discoteca di Stato, Pontifical Gregorian University, the Italian Ministry for Cultural Heritage, the demo and poster chairpersons, the panel chairpersons, the workshop chairman, the tutorial chairpersons, the program committee members, the additional referees, the invited speakers, the tutorialists, the panelists, the poster presenters, the demonstrators, the sponsors, and of course all the participants.

During the preparation of the proceedings we received a great deal of help from Nicola Ferro and we thank him for this. The work for a conference like this starts at least two years before the conference happens, and during all this time Tarina Ayazi supported us. We thank her for her contribution to the success of the initiative.

Organization

General Chair

Costantino Thanos (National Research Council, Italy)

Program Chair

Maristella Agosti (University of Padua, Italy)

Organization Chair

Pier Giorgio Marchetti (European Space Agency, Italy)

Organization Committee

Vittore Casarosa (National Research Council, Italy)

Marta Giorgi (Pontifical Gregorian University, Italy)

Maria Pia Mariani (Ministry for Cultural Heritage, Italy)

Giovanna Merola (Istituto Centrale per il Catalogo Unico, Italy)

M. Carla Sotgiu (Discoteca di Stato, Italy)

Demo and Poster Chairs

Edward Fox (Virginia Tech, USA)

Mike Papazoglou (University of Tilburg, The Netherlands)

Panel Chairs

Alan Smeaton (Dublin City University, Ireland)

Rachel Heery (UKOLN, UK)

Workshop Chair

Yannis Ioannidis (University of Athens, Greece)

Tutorial Chairs

Hans-Jorg Schek (ETH-Zurich, Switzerland)

Bruce Croft (University of Massachusetts, USA)

US Liaison Chair

Nick Belkin (Rutgers University, USA)

Publicity Chairs

Fillia Makedon (Dartmouth College, USA)

Simonetta Cheli (European Space Agency)

Treasurer

Ettore Ricciardi (National Research Council, Italy)

Program Committee

William Y. Arms (Cornell University, USA)

Daniel E. Atkins (University of Michigan, USA)

Paolo Atzeni (University of Roma 3, Italy)

Ricardo Baeza-Yates (University of Chile)

Thomas Baker (Fraunhofer Gesellschaft, Sankt Augustin, Germany)

Elisa Bertino (University of Milan, Italy)

Klemens Boehm (ETH Zurich, Switzerland)

Jose Borbinha (National Library of Portugal)

Stefano Ceri (Politecnico di Milano, Italy)

Ching-chih Chen (Simmons College, Boston, USA)

Hsinchun Chen (University of Arizona, USA)

Stavros Christodoulakis (University of Crete, Greece)

Panos Costantopoulos (FORTH, Crete, Greece)

Gregory Crane (Tufts University, Medford, USA)

Alberto Del Bimbo (University of Florence, Italy)

Werner A. Deutsch (Austrian Academy of Sciences)

Floriana Esposito (University of Bari, Italy)

Dieter Fellner (Technical University of Braunschweig, Germany)

Norbert Fuhr (University of Dortmund, Germany)

Wendy Hall (University of Southampton, UK)

Jane Hunter (The University of Queensland, Brisbane, Australia)

Peter Ingwersen (Royal School of Library and Information Science,
Copenhagen, Denmark)

Leonid Kalinichenko (Russian Academy of Sciences)

Noriko Kando (National Institute of Informatics, Tokyo, Japan)

Judith Klavans (Columbia University, USA)

Jaana Kekalainen (University of Tampere, Finland)

Martin Kersten (National Research Institute for Mathematics and
Computer Science, The Netherlands)

Traugott Koch (Lund University, Sweden)

Carl Lagoze (Cornell University, USA)

Ronald Larsen (University of Maryland, College Park, USA)
 Clifford Lynch (Coalition for Networked Information, Washington, DC, USA)
 Elizabeth Lyon (UKOLN, University of Bath, UK)
 Gary Marchionini (University of North Carolina at Chapel Hill, USA)
 Herman Maurer (University of Gratz, Austria)
 Massimo Melucci (University of Padua, Italy)
 Marc Nanard (LIRMM, Montpellier, France)
 A. Desai Narasimhalu (National University of Singapore)
 Erich Neuhold (Fraunhofer Gesellschaft - IPSI, Germany)
 Bo Ohrstrom (Danish National Library Authority)
 Andreas Paepcke (Stanford University, USA)
 Michael Papazoglou (Tilburg University, The Netherlands)
 Carol Peters (Italian National Research Council)
 Berthier Ribeiro-Neto (Federal University of Minas Gerais, Brazil)
 Steve Robertson (MS Research, Cambridge, UK)
 Chris Rusbridge (University of Glasgow, UK)
 Rudi Schmiede (Technische Universitat Darmstadt, Germany)
 Fabrizio Sebastiani (Italian National Research Council)
 Timos Sellis (National Technical University of Athens, Greece)
 Maria Sliwinska (Nicholas Copernicus University, Poland)
 Alan Smeaton (Dublin City University, Ireland)
 Terrence R. Smith (University of California, Santa Barbara, USA)
 Ingeborg Solvberg (Norwegian University of Science and Technology, Trondheim, Norway)
 Shigeo Sugimoto (University of Library and Information Science, Tsukuba, Japan)
 Elaine G. Toms (Faculty of Information Studies, University of Toronto, Canada)
 Keith van Rijsbergen (University of Glasgow, UK)
 Howard Wactlar (Carnegie Mellon University, USA)
 Stuart Weibel (Online Computer Library Center, USA)
 Robert Wilensky (University of California, Berkeley, USA)
 Robin Williams (IBM Almaden Research Center, San Jose, USA)
 Ian H. Witten (University of Waikato, New Zealand)
 Jerome Yen (The University of Hong Kong)

Additional Referees

Gianni Amati	Marco Degemmis
Michela Bacchin	Maria de Lourdes da Silveira
Maria Bruna Baldacci	Nicola Di Mauro
Teresa M.A. Basile	Koji Eguchi
Maria Carla Cavagnis Sotgiu	Nicola Fanizzi
Theodore Dalamagas	Dieter Fellner

Stefano Ferilli
Nicola Ferro
Ingo Frommholz
Jette Hyldegaard
Gerald Jäschke
Predrag Knezevic
Marcello L'Abbate
Birger Larsen
Patrick Lehti
Oriana Licchelli
Pasquale Lops
Anne Mahoney

Pier Giorgio Marchetti
Christian Mönch
Jocelyne Nanard
Jeppe Nicolaisen
Pasquale Pagano
Luca Pretto
Thomas Risse
Piet Seiden
Luciano Romero Soares Lima
Erik Thorlund Jepsen
Marco Zens

Table of Contents

Web Archiving

A First Experience in Archiving the French Web	1
<i>S. Abiteboul, G. Cobéna, J. Masanes, G. Sedrati</i>	

Austrian Online Archive Processing: Analyzing Archives of the World Wide Web	16
<i>Andreas Rauber, Andreas Aschenbrenner, Oliver Witvoet</i>	

e-Book

Conversion of eBook Documents Based on Mapping Relations	32
<i>Seung-Kyu Ko, Myoung-Soo Kang, Won-Sung Sohn, Soon-Bum Lim, Yoon-Chul Choy</i>	

Guidelines for Designing Electronic Books	47
<i>Ruth Wilson, Monica Landoni, Forbes Gibb</i>	

Collection Building

Personalized Classification for Keyword-Based Category Profiles	61
<i>Aixin Sun, Ee-Peng Lim, Wee-Keong Ng</i>	

Statistical Analysis of Bibliographic Strings for Constructing an Integrated Document Space	75
<i>Atsuhiko Takasu</i>	

Focused Crawls, Tunneling, and Digital Libraries	91
<i>Donna Bergmark, Carl Lagoze, Alex Sbityakov</i>	

Web Technologies

Goal-Oriented Requirements Specification for Digital Libraries	107
<i>Davide Bolchini, Paolo Paolini</i>	

OntoLog: Temporal Annotation Using Ad Hoc Ontologies and Application Profiles	118
<i>Jon Heggland</i>	

An XML Log Standard and Tool for Digital Library Logging Analysis	129
<i>Marcos André Gonçalves, Ming Luo, Rao Shen, Mir Farooq Ali, Edward A. Fox</i>	

OAI Applications

Notes from the Interoperability Front: A Progress Report on the Open Archives Initiative	144
<i>Herbert Van de Sompel, Carl Lagoze</i>	
Dynamic Generation of Intelligent Multimedia Presentations through Semantic Inferencing	158
<i>Suzanne Little, Joost Geurts, Jane Hunter</i>	
Technical Report Interchange through Synchronized OAI Caches	176
<i>Xiaoming Liu, Kurt Maly, Mohammad Zubair, Rong Tang, Mohammed Imran Padshah, George Roncaglia, JoAnne Rocker, Michael Nelson, William von Ofenheim, Richard Luce, Jacqueline Stack, Frances Knudson, Beth Goldsmith, Irma Holtkamp, Miriam Blake, Jack Carter, Mariella Di Giacomo, Jerome Nutter, Susan Brown, Ron Montbrand, Sally Landenberger, Kathy Pierson, Vince Duran, Beth Moser</i>	

Case Studies

Functional Requirements for Online Tools to Support Community-Led Collections Building	190
<i>Michael Khoo, Holly Devaul, Tamara Sumner</i>	
A Study on the Evaluation Model for University Libraries in Digital Environments	204
<i>Byeong Heui Kwak, Woochun Jun, Le Gruenwald, Suk-Ki Hong</i>	
Renardus: Following the Fox from Project to Service	218
<i>Lesly Huxley</i>	
From Digital Archive to Digital Library – A Middleware for Earth-Observation Data Management	230
<i>Stephan Kiemle</i>	

Navigation/Query Language

Navigating in Bibliographic Catalogues	238
<i>Trond Aalberg</i>	
Foundations of a Multidimensional Query Language for Digital Libraries	251
<i>Donatella Castelli, Carlo Meghini, Pasquale Pagano</i>	

Audio/Video Retrieval

The TREC2001 Video Track: Information Retrieval on Digital Video Information	266
<i>Alan F. Smeaton, Paul Over, Cash J. Costello, Arjen P. de Vries, David Doermann, Alexander Hauptmann, Mark E. Rorvig, John R. Smith, Lide Wu</i>	

Automated Alignment and Annotation of Audio-Visual Presentations	276
<i>Gareth J.F. Jones, Richard J. Edens</i>	

Architecture I

OpenDLib: A Digital Library Service System	292
<i>Donatella Castelli, Pasquale Pagano</i>	

Prototyping Digital Library Technologies in <i>zetoc</i>	309
<i>Ann Apps, Ross MacIntyre</i>	

Employing Smart Browsers to Support Flexible Information Presentation in Petri Net-Based Digital Libraries	324
<i>Unmil P. Karadkar, Jin-Cheon Na, Richard Furuta</i>	

IR

On the Use of Explanations as Mediating Device for Relevance Feedback	338
<i>Ian Ruthven</i>	

Qualitative Evaluation of Thesaurus-Based Retrieval	346
<i>D. Blocks, C. Binding, D. Cunliffe, D. Tudhope</i>	

Meta-data Extraction and Query Translation. Treatment of Semantic Heterogeneity	362
<i>Robert Strötgen</i>	

Architecture II

MetaDL: A Digital Library of Metadata for Sensitive or Complex Research Data	374
<i>Fillia Makedon, James Ford, Li Shen, Tilmann Steinberg, Andrew Saykin, Heather Wishart, Sarantos Kapidakis</i>	

Importing Documents and Metadata into Digital Libraries: Requirements Analysis and an Extensible Architecture.....	390
<i>Ian H. Witten, David Bainbridge, Gordon Paynter, Stefan Boddie</i>	

The Mellon Fedora Project (Digital Library Architecture Meets XML and Web Services) 406
Sandra Payette, Thornton Staples

Evaluation

Hybrid Partition Inverted Files: Experimental Validation 422
Wensi Xi, Ohm Sornil, Ming Luo, Edward A. Fox

Digital Library Evaluation by Analysis of User Retrieval Patterns 432
Johan Bollen, Somasekhar Vemulapalli, Weining Xu

Interactive Search Results 448
Ioannis Papadakis, Ioannis Andreou, Vassileios Chrissikopoulos

Multimedia/Mixed Media

An Investigation of Mixed-Media Information Retrieval 463
Gareth J.F. Jones, Adenike M. Lam-Adesina

Alignment of Performances with Scores Aimed at Content-Based Music Access and Retrieval 479
Nicola Orio

Alternative Surrogates for Video Objects in a Digital Library: Users' Perspectives on Their Relative Usability 493
Barbara M. Wildemuth, Gary Marchionini, Todd Wilkens, Meng Yang, Gary Geisler, Beth Fowler, Anthony Hughes, Xiangming Mu

Word Alignment in Digital Talking Books Using WFSTs 508
António Serralheiro, Diamantino Caseiro, Hugo Meinedo, Isabel Trancoso

Preservation/Classification/User Studies

Migration on Request, a Practical Technique for Preservation 516
Phil Mellor, Paul Wheatley, Derek Sergeant

Information Alert in Distributed Digital Libraries: The Models, Languages, and Architecture of DIAS 527
M. Koubarakis, T. Koutris, C. Tryfonopoulos, P. Raftopoulou

DSpace: An Institutional Repository from the MIT Libraries and Hewlett Packard Laboratories 543
MacKenzie Smith

User Behavior Tendencies on Data Collections in a Digital Library 550
Michalis Sfakakis, Sarantos Kapidakis

Student Comprehension of Classification Applications in a Science Education Digital Library	560
<i>Jane Greenberg, Kristen A. Bullard, M.L. James, Evelyn Daniel, Peter White</i>	

Architecture III

Designing Protocols in Support of Digital Library Componentization	568
<i>Hussein Suleman, Edward A. Fox</i>	
Exploring Small Screen Digital Library Access with the Greenstone Digital Library	583
<i>George Buchanan, Matt Jones, Gary Marsden</i>	
Daffodil: An Integrated Desktop for Supporting High-Level Search Activities in Federated Digital Libraries	597
<i>Norbert Fuhr, Claus-Peter Klas, André Schaefer, Peter Mutschke</i>	

Humanities

Using Human Language Technology for Automatic Annotation and Indexing of Digital Library Content	613
<i>Kalina Bontcheva, Diana Maynard, Hamish Cunningham, Horacio Saggion</i>	
Cultural Heritage Digital Libraries: Needs and Components	626
<i>Gregory Crane</i>	
Visualization of Variants in Textual Collations to Analyze the Evolution of Literary Works in the Cervantes Project	638
<i>Carlos Monroy, Rajiv Kochumman, Richard Furuta, Eduardo Urbina, Eréndira Melgoza, Arpita Goenka</i>	

Demos and Posters

Alinari Online: Access Quality through a Cultural Photographic Heritage Site	654
<i>Andrea de Polo, Sam Minelli</i>	
An Access Control System for Digital Libraries and the Web: The <i>MaX</i> Prototype Demonstration	656
<i>Elisa Bertino, Elena Ferrari, Andrea Perego</i>	
Human Language Technology for Automatic Annotation and Indexing of Digital Library Content	658
<i>Kalina Bontcheva, Hamish Cunningham</i>	

The IntraText Digital Library: XML-Driven Online Library Based on High Accessibility, Lexical Hypertextualization and Scholarly Accuracy in Philological / Textual Notations 659
Nicola Mastidoro

The Meta Project – Automated Digitisation of Books and Journals 660
Günter Mühlberger, Birgit Stehno

COVAX: A Contemporary Culture Virtual Archive in XML 661
Luciana Bordoni

Author Index 663