

# Lecture Notes in Computer Science

2769

Edited by G. Goos, J. Hartmanis, and J. van Leeuwen

**Springer**

*Berlin*

*Heidelberg*

*New York*

*Hong Kong*

*London*

*Milan*

*Paris*

*Tokyo*

Traugott Koch Ingeborg Torvik Sølvsberg (Eds.)

# Research and Advanced Technology for Digital Libraries

7th European Conference, ECDL 2003  
Trondheim, Norway, August 17-22, 2003  
Proceedings



Springer

## Series Editors

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

## Volume Editors

Traugott Koch  
Lund University Libraries, NetLab, Knowledge Technologies Group  
P.O. Box 134, 22100 Lund, Sweden  
E-mail: [traugott.koch@lub.lu.se](mailto:traugott.koch@lub.lu.se)  
Ingeborg Torvik Sølvsberg  
Norwegian University of Science and Technology  
Dept. of Computer and Information Science  
7491 Trondheim, Norway  
E-mail: [ingeborg.solvsberg@idi.ntnu.no](mailto:ingeborg.solvsberg@idi.ntnu.no)

## Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress

Bibliographic information published by Die Deutsche Bibliothek  
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie;  
detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

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 3-540-40726-X 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 2003  
Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin GmbH  
Printed on acid-free paper      SPIN: 10931004      06/3142      5 4 3 2 1 0

# Preface

Welcome to ECDL 2003 and to these conference proceedings, featuring all the papers presented at the 7th European Conference on Research and Advanced Technology for Digital Libraries.

Following Pisa (1997), Heraklion (1998), Paris (1999), Lisbon (2000), Darmstadt (2001) and Rome (2002), ECDL 2003 in Trondheim reaches some of the northernmost shores of the continent.

Being the seventh in an annual series of conferences represents, for better and for worse, a considerable tradition in the fast changing world of digital-library-related research and development. It is still a difficult and slow job to change traditional forms and formats of communication at – and related to – scientific conferences, and also to change participants' expectations. Yet each new conference builds upon the topics and communities involved in previous events and inherits the commitment to quality established by its predecessors.

Each year, ECDL has to live up to its role of being “the major European forum focusing on digital libraries and associated technical, practical, and social issues,” bringing diverse disciplines and application communities together. There are still challenges in this respect ahead of us, but the quality and range of papers and other contributions, combined with opportunities for debate, should ensure that ECDL 2003 sets high standards for the future.

ECDL 2003 invited contributions from the broadest spectrum of digital-library-related research, development and practice, an array of 24 topics under the following headings:

- general concepts, methods, standards, economics and strategies
- collection building, management and integration
- architectures and interoperability
- knowledge organization, discovery and retrieval
- mediation, user interaction and collaboration

All 24 conference topics were covered by paper submissions and by expert reviewers. As can be seen from all parts of the conference program, even after highly selective peer review this wide coverage is to a large degree preserved.

The 69 Program Committee members who participated in the review work represent 27 countries from all continents and many digital-library-related disciplines. A further 66 referees shared the peer-review work, assuring three reviews per submission.

The call for papers resulted in 161 paper submissions, 25% of which were declared as short papers. This response is one of the largest number of papers ever submitted to an ECDL conference.

We were able to accept 47 papers, 8 of which are short papers. This is a rather low acceptance rate of 29% compared with the many highly rated submissions we received. So the competition turned out to be fierce and high ratings from

the reviewers were required to make the final list of papers for inclusion. The 47 accepted papers are clustered in 13 sessions in two parallel tracks.

The outcome of the lengthy but quality-assuring review process is demonstrated by the high quality of papers included in this volume of proceedings.

The other elements of the conference program were also subject to quality assurance processes. 72 posters and demonstrations were submitted to ECDL 2003, a very high number. “Only” 52 could be selected for presentation, 36 posters and 16 demonstrations. At least 9 are combined formats. A late deadline allowed an invitation to submit demonstrations related to accepted papers, and posters presenting late-breaking results. The posters will be on display throughout the whole conference.

Three excellent and widely known colleagues agreed to be the invited speakers at ECDL 2003: Karen Sparck Jones, Clifford Lynch and John Lervik.

Three exciting panel topics aim at stimulating controversy, lively debate and active participation of the audience.

The number of tutorials and workshops also far exceeded the number that could be accommodated. We feature a very rich tutorial program: nine tutorials are on offer with a good mix of the established and successful with several new topics on both introductory and advanced levels. The five workshops cover a broad spectrum of application areas. Tutorials and workshops turn out to be a good complement to the topics of the paper sessions.

We hope that the interactive potential of these formats will be fully exploited by conference participants. It will be especially interesting to see if the wireless network – providing almost ubiquitous Internet access at the conference venue – can provide new opportunities regarding conference work and communication.

Finally, we want to thank all those who contributed their time and efforts to make this conference possible: the local Organizing Committee, all the conference chairs, the Program Committee and the additional referees, the host institutions and sponsors, the invited speakers, the panellists, tutorial and workshop leaders, all presenters, speakers, session chairs and participants. A conference is a truly collective effort.

You will find references to everything related to the conference and further information at the Web site <http://www.ecdl2003.org/>.

We are confident that we shall, all together, make ECDL 2003 a well attended, high-quality and successful conference, offering all participants rich and fruitful opportunities to share ideas, results and experiences, to gain insights and make valuable contacts.

June 2003

Traugott Koch  
Ingeborg Torvik Sølvsberg

## **Organization**

### **General Chair**

Ingeborg Torvik Sølvsberg, Norwegian University of Science and Technology

### **Program Chair**

Traugott Koch, NetLab, Lund University, Sweden

### **Organizing Chairs**

Trond Aalberg, Norwegian University of Science and Technology

Roger Midtstraum, Norwegian University of Science and Technology

### **Demonstration Chairs**

José Luis Borbinha, National Library of Portugal

Stuart Sutton, University of Washington, USA

### **Poster Chairs**

Ragnar Nordlie, Oslo College, Norway

Linda Hill, University of California, Santa Barbara, USA

### **Panel Chairs**

Anders Ardö, Lund University, Sweden

Carol Peters, IEL-CNR, Italy

### **Workshop Chairs**

Liz Lyon, University of Bath, UK

Christine Borgman, University of California, Los Angeles, USA

### **Tutorial Chairs**

Andreas Rauber, Vienna University of Technology, Austria

Ronald Larsen, University of Pittsburg, USA

### **Publicity Chair**

Lesly Huxley, ILRT, University of Bristol, UK

## Program Committee

Maristella Agosti, University of Padua, Italy  
Daniel E. Atkins, University of Michigan, USA  
Ricardo Baeza-Yates, University of Chile  
Thomas Baker, Fraunhofer Gesellschaft, Germany  
Nick Belkin, Rutgers University, USA  
Alejandro Bia, University of Alicante, Spain  
Katy Boerner, Indiana University, USA  
Jose Borbinha, National Library of Portugal  
Christine Borgman, University of California, Los Angeles, USA  
Svein Arne Bryggfeld, National Library of Norway  
Gerhard Budin, University of Vienna, Austria  
Warwick Cathro, National Library of Australia  
Hsinchun Chen, University of Arizona, USA  
Key-Sun Choi, KAIST, Korea  
Gregory Crane, Tufts University, USA  
Ron Daniel, Jr., Taxonomy Strategies, USA  
Alberto Del Bimbo, University of Florence, Italy  
Lorcan Dempsey, OCLC, USA  
Martin Doerr, ICS FORTH, Greece  
Matthew Dovey, Oxford University, UK  
Jacques Ducloy, CNRS INIST, France  
Dieter Fellner, Technical University of Braunschweig, Germany  
Ed Fox, Virginia Tech, USA  
Michael Freeston, University of California, Santa Barbara, USA  
Norbert Fuhr, Duisburg University, Germany  
Richard Furuta, Texas A&M University, USA  
Carole Goble, University of Manchester, UK  
Stefan Gradmann, Hamburg University, Germany  
Jane Greenberg, University of North Carolina, Chapel Hill, USA  
Juha Hakala, Helsinki University Library, Finland  
Sigrun Klara Hannesdottir, National Library of Iceland  
Margaret Hedstrom, University of Michigan, USA  
Jane Hunter, University of Queensland, Australia  
Ole Husby, BIBSYS, Norway  
Yannis Ioannidis, University of Athens, Greece  
Poul Henrik Joergensen, Portia System, Denmark  
Leonid Kalinichenko, Russian Academy of Sciences  
Stephen Katz, Food and Agriculture Organization of the UN  
Jaana Kekäläinen, University of Tampere, Finland  
Juergen Krause, IZ Bonn, and University of Koblenz and Landau, Germany  
Carl Lagoze, Cornell University, USA  
Ray Larson, University of California, Berkeley, USA  
David Levy, University of Washington, USA  
Ee-Peng Lim, Nanyang Technical University, Singapore



Clifford Lynch, Coalition for Networked Information, USA  
Michael Mabe, Elsevier Science, UK  
Elena Maceviciute, Vilnius University, Lithuania, and Swedish School of Library and Information Science  
Kingo Mchombu, University of Namibia  
Cliff McKnight, Loughborough University, UK  
Gary Marchionini, University of North Carolina, Chapel Hill, USA  
Erich J. Neuhold, Fraunhofer Gesellschaft-IPSI, Germany  
Achim Osswald, University of Applied Sciences, Köln, Germany  
Andy Powell, UKOLN, University of Bath, UK  
Chris Rusbridge, University of Glasgow, UK  
Hans Joerg Schek, ETH Zurich, Switzerland  
Rudi Schmiede, TU Darmstadt, Germany  
Alan Smeaton, Dublin City University, Ireland  
Dagobert Soergel, University of Maryland, USA  
Shigeo Sugimoto, University of Library and Information Science, Japan  
Stuart Sutton, University of Washington, USA  
Costantino Thanos, National Research Council, Italy  
Elaine G. Toms, Dalhousie University, Canada  
Doug Tudhope, University of Glamorgan, UK  
Herbert Van de Sompel, Los Alamos National Laboratory, USA  
Titia van der Werf, Leiden University, Netherlands  
Diane Vizine-Goetz, OCLC, USA  
Stuart Weibel, OCLC, USA  
Ian Witten, University of Waikato, New Zealand  
Irene Wormell, Swedish School of Library and Information Science

## Additional Referees

Giuseppe Amato	Jon Olav Hauglid	Henrik Nottelmann
Anders Ardö	Jon Heggland	Walter Nunziati
Jurgen Assfalg	Otthein Herzog	Giorgio Maria Di Nunzio
David Bainbridge	Linda Hill	Nicola Orio
Maria Bruna Baldacci	Lieming Huang	Pasquale Pagano
Stefano Berretti	Zan Huang	Pietro Pala
Marco Bertini	Fathi Husein	Jeff Pan
George Buchanan	Kai Jauslin	Dimitris Plexousakis
Francesco Carreras	Matt Jones	Jialun Qin
Carlos Castillo	Sara Kjellberg	Heri Ramampiaro
Wingyan Chung	Claus-Peter Klas	Dietmar Saupe
Panos Constantopoulos	Predrag Knezewic	André Schaefer
Daniel Cunliffe	Jessica Lindholm	Xijun Shi
Sally Jo Cunningham	Sigfrid Lundberg	Martin Steinebach
Annakim Eltén	Gary Marsden	Manuel Sánchez-Quero
Jörgen Eriksson	Byron Marshall	Roman Tzschoppe
Peter Fankhauser	Daniel McDonald	Can Türker
Nicola Ferro	Roger Midtstraum	Alejandro Vaisman
Claudio Gennaro	Andrea Miene	Gang Wang
Torsten Grabs	Michael Mlivoncic	Andreas Wombacher
Kai Grossjohann	Gonzalo Navarro	Jennifer Xu
Norbert Gövert	Claudia Niedereé	Yiwen Zhang

## Sponsoring Institutions

ABM – Norwegian Archive, Library and Museum Authority  
 BIBSYS – Norwegian Service Center for Libraries  
 ERCIM – European Research Consortium for Informatics and Mathematics  
 Fast Search & Transfer ASA  
 NFR – Research Council of Norway  
 NORDINFO – Nordic Council for Scientific Information,  
 NTNU (Norwegian University of Science and Technology), Department of  
 Computer and Information Science  
 NTNU Library  
 SINTEF Telecom and Informatics  
 Sun Microsystems

# Table of Contents

## Uses, Users, and User Interaction

Users and Uses of Online Digital Libraries in France . . . . .	1
<i>Houssein Assadi, Thomas Beauvisage, Catherine Lupovici, Thierry Cloarec</i>	
In Search for Patterns of User Interaction for Digital Libraries . . . . .	13
<i>Jela Steinerová</i>	
Detecting Research Trends in Digital Library Readership . . . . .	24
<i>Johan Bollen, Rick Luce, Somasekhar Vemulapalli, Weining Xu</i>	
Evaluating the Changes in Knowledge and Attitudes of Digital Library Users . . . . .	29
<i>Gemma Madle, Patty Kostkova, Jane Mani-Saada, Julius R. Weinberg</i>	

## Metadata Applications

Towards a Role-Based Metadata Scheme for Educational Digital Libraries: A Case Study in Singapore . . . . .	41
<i>Dian Melati Md Ismail, Ming Yin, Yin-Leng Theng, Dion Hoe-Lian Goh, Ee-Peng Lim</i>	
Incorporating Educational Vocabulary in Learning Object Metadata Schemas . . . . .	52
<i>Jian Qin, Carol Jean Godby</i>	
Findings from the Mellon Metadata Harvesting Initiative . . . . .	58
<i>Martin Halbert, Joanne Kaczmarek, Kat Hagedorn</i>	
Semantic Browsing . . . . .	70
<i>Alexander Faaborg, Carl Lagoze</i>	
Metadata Editing by Schema . . . . .	82
<i>Hussein Suleman</i>	

## Annotation and Recommendation

Annotations: Enriching a Digital Library . . . . .	88
<i>Maristella Agosti, Nicola Ferro</i>	

Identifying Useful Passages in Documents Based on Annotation Patterns .....	101
<i>Frank Shipman, Morgan Price, Catherine C. Marshall, Gene Golovchinsky</i>	

Others Also Use: A Robust Recommender System for Scientific Libraries .....	113
<i>Andreas Geyer-Schulz, Andreas Neumann, Anke Thede</i>	

## Automatic Classification and Indexing

Cross-Lingual Text Categorization .....	126
<i>Nuria Bel, Cornelis H.A. Koster, Marta Villegas</i>	

Automatic Multi-label Subject Indexing in a Multilingual Environment ..	140
<i>Boris Lauser, Andreas Hotho</i>	

Automatic Induction of Rules for Classification and Interpretation of Cultural Heritage Material .....	152
<i>S. Ferilli, F. Esposito, T.M.A. Basile, N. Di Mauro</i>	

An Integrated Digital Library Server with OAI and Self-Organizing Capabilities .....	164
<i>Hyunki Kim, Chee-Yoong Choo, Su-Shing Chen</i>	

## Web Technologies

YAPI: Yet Another Path Index for XML Searching .....	176
<i>Giuseppe Amato, Franca Debole, Pavel Zezula, Fausto Rabitti</i>	

Structure-Aware Query for Digital Libraries: Use Cases and Challenges for the Humanities .....	188
<i>Christopher York, Clifford Wulfman, Greg Crane</i>	

Combining DAML+OIL, XSLT and Probabilistic Logics for Uncertain Schema Mappings in MIND .....	194
<i>Henrik Nottelmann, Norbert Fuhr</i>	

Digitometric Services for Open Archives Environments.....	207
<i>Tim Brody, Simon Kampa, Stevan Harnad, Les Carr, Steve Hitchcock</i>	

## Topical Crawling, Subject Gateways

Search Engine-Crawler Symbiosis: Adapting to Community Interests ....	221
<i>Gautam Pant, Shannon Bradshaw, Filippo Menczer</i>	

Topical Crawling for Business Intelligence .....	233
<i>Gautam Pant, Filippo Menczer</i>	

SozioNet: Networking Social Science Resources .....	245
<i>Wolfgang Meier, Natascha Schumann, Sue Heise, Rudi Schmiede</i>	
VASCODA: A German Scientific Portal for Cross-Searching Distributed Digital Resource Collections .....	257
<i>Heike Neuroth, Tamara Pianos</i>	

## Architectures and Systems

Scenario-Based Generation of Digital Library Services .....	263
<i>Rohit Kelapure, Marcos André Gonçalves, Edward A. Fox</i>	
An Evaluation of Document Prefetching in a Distributed Digital Library .....	276
<i>Jochen Hollmann, Anders Ardö, Per Stenström</i>	
An Architecture for Online Information Integration on Concurrent Resource Access on a Z39.50 Environment .....	288
<i>Michalis Sfakakis, Sarantos Kapidakis</i>	

## Knowledge Organization: Concepts

The ADEPT Concept-Based Digital Learning Environment .....	300
<i>T.R. Smith, D. Ancona, O. Buchel, M. Freeston, W. Heller, R. Nottrott, T. Tierney, A. Ushakov</i>	
A User Evaluation of Hierarchical Phrase Browsing.....	313
<i>Katrina D. Edgar, David M. Nichols, Gordon W. Paynter, Kirsten Thomson, Ian H. Witten</i>	
Visual Semantic Modeling of Digital Libraries .....	325
<i>Qinwei Zhu, Marcos André Gonçalves, Rao Shen, Lillian Cassell, Edward A. Fox</i>	

## Collection Building and Management

Connecting Interface Metaphors to Support Creation of Path-Based Collections .....	338
<i>Unmil P. Karadkar, Andruid Kerne, Richard Furuta, Luis Francisco-Revilla, Frank Shipman, Jin Wang</i>	
Managing Change in a Digital Library System with Many Interface Languages .....	350
<i>David Bainbridge, Katrina D. Edgar, John R. McPherson, Ian H. Witten</i>	
A Service for Supporting Virtual Views of Large Heterogeneous Digital Libraries .....	362
<i>Leonardo Candela, Donatella Castelli, Pasquale Pagano</i>	

## Knowledge Organization: Authorities and Works

A Framework for Unified Authority Files: A Case Study of Corporate Body Names in the FAO Catalogue .....	374
<i>James Weinheimer, Kafkas Caprazli</i>	
geoXwalk – A Gazetteer Server and Service for UK Academia .....	387
<i>James Reid</i>	
Utilizing Temporal Information in Topic Detection and Tracking .....	393
<i>Juha Makkonen, Helena Ahonen-Myka</i>	
Automatic Conversion from MARC to FRBR .....	405
<i>Christian Mönch, Trond Aalberg</i>	

## Information Retrieval in Different Application Areas

Musescape: A Tool for Changing Music Collections into Libraries .....	412
<i>George Tzanetakis</i>	
A Digital GeoLibrary: Integrating Keywords and Place Names .....	422
<i>Mathew Weaver, Lois Delcambre, Leonard Shapiro, Jason Brewster, Afrem Gutema, Timothy Tolle</i>	
Document-Centered Collaboration for Scholars in the Humanities – The COLLATE System .....	434
<i>Ingo Frommholz, Holger Brocks, Ulrich Thiel, Erich Neuhold, Luigi Iannone, Giovanni Semeraro, Margherita Berardi, Michelangelo Ceci</i>	

## Digital Preservation

DSpace as an Open Archival Information System: Current Status and Future Directions .....	446
<i>Robert Tansley, Mick Bass, MacKenzie Smith</i>	
Preserving the Fabric of Our Lives: A Survey of Web Preservation Initiatives .....	461
<i>Michael Day</i>	
Implementing Preservation Strategies for Complex Multimedia Objects ..	473
<i>Jane Hunter, Sharmin Choudhury</i>	

## Indexing and Searching of Special Document and Collection Information

Distributed IR for Digital Libraries .....	487
<i>Ray R. Larson</i>	

Reference Directed Indexing: Redeeming Relevance for Subject Search in Citation Indexes .....	499
<i>Shannon Bradshaw</i>	
Space-Efficient Support for Temporal Text Indexing in a Document Archive Context .....	511
<i>Kjetil Nørnvåg</i>	
Clustering Top-Ranking Sentences for Information Access .....	523
<i>Anastasios Tombros, Joemon M. Jose, Ian Ruthven</i>	
<b>Demonstrations</b> .....	529
<b>Posters</b> .....	531
<b>Author Index</b> .....	535