

Lecture Notes in Computer Science

1324

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

Advisory Board: W. Brauer D. Gries J. Stoer

Carol Peters Costantino Thanos (Eds.)

Research and Advanced Technology for Digital Libraries

First European Conference, ECDL'97
Pisa, Italy, September 1-3, 1997
Proceedings



Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany

Juris Hartmanis, Cornell University, NY, USA

Jan van Leeuwen, Utrecht University, The Netherlands

Volume Editors

Carol Peters

Costantino Thanos

Istituto di Elaborazione della Informazione

Consiglio Nazionale delle Ricerche

Via S. Maria, 46, I-56126 Pisa, Italy

E-mail: (c.peters/thanos)@iei.pi.cnr.it

Cataloging-in-Publication data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Research and advanced technology for digital libraries : first European conference ; proceedings / ECDL '97, Pisa, Italy, September 1 - 3, 1997. Carol Peters ; Constantino Thanos (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Budapest ; Hong Kong ; London ; Milan ; Paris ; Santa Clara ; Singapore ; Tokyo : Springer, 1997
(Lecture notes in computer science ; Vol. 1324)
ISBN 3-540-63554-8

CR Subject Classification (1991): H.2, H.3, H.4.3, H.5, I.7.2-3, J.1, J.7

ISSN 0302-9743

ISBN 3-540-63554-8 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 1997
Printed in Germany

Typesetting: Camera-ready by author
SPIN 10545735 06/3142 - 5 4 3 2 1 0 Printed on acid-free paper

Preface

The First European Conference on Digital Libraries is the first in a series of European Conferences on Research and Advanced Technology for Digital Libraries. It is promoted by the European Research Consortium for Informatics and Mathematics (ERCIM) and partially funded by the CEU TMR Programme.

The aim of this series of conferences is to give researchers and professionals from universities, research centres, industry and government the opportunity to meet annually to discuss evolving research issues and applications. In particular the objectives are: to bring together researchers from multiple disciplines relating to the development of digital libraries; to provide an opportunity for these scientists to form a community in Europe specific to digital library development and to enable them to discuss issues and strategies specific to the European context; to assist young researchers in establishing relationships with senior scientists in their areas of interest; to enable review and discussion of research under way in the world on digital libraries; to stimulate researchers, especially young scientists, to explore new areas of interest in digital library development.

From a technical point of view, the European Conference goals are to contribute to the definition of those digital library parameters which especially influence issues of access to, retrieval of and interaction with information; to identify key problems which must be solved to make digital library service an effective reality; to identify a general structure or framework for integrating research and solutions; and to propose and encourage specific, high-priority research directions within such a framework.

An international programme committee was set up for this Conference with representatives from 14 countries. It received 50 full paper and 12 poster submissions plus 15 proposals for demos. Each paper was evaluated by 3 reviewers, and 23 papers of high academic quality were selected for presentation. These papers cover a wide spectrum of topics which includes information retrieval, metadata, multilingual and multimedia information retrieval, user interaction schemes, structured documents and architectures. Poster and demo sessions are also included in the programme. In addition, two invited talks on digital property rights and object database support for digital libraries will be given. The Conference will close looking towards the future with a panel discussion on new initiatives for digital collections.

We should like here to thank all those individuals and institutions who have made this conference possible: IEL-CNR, ERCIM, the CEU TMR Programme, the programme committee members, the additional unknown referees, the invited speakers, the tutorialists, the panelists, the poster presenters and the demonstrators, the sponsors, and of course all the participants.

Organisation

The First European Conference on Research and Advanced Technology for Digital Libraries has been organised in Pisa by the Istituto di Elaborazione della Informazione, Consiglio Nazionale delle Ricerche (IEI-CNR) with the assistance of the Central Office of the European Research Consortium for Informatics and Mathematics (ERCIM).

Programme Chair

Costantino Thanos, IEI-CNR, Italy

Programme Committee

Robert B. Allen, Bellcore
Thomas Baker, Asian Institute of Technology
Nicholas J. Belkin, Rutgers University
William Birmingham, University of Michigan
Bruce Croft, University of Massachusetts
Lorcan Dempsey, University of Bath
Jacques Ducloy, INRIA-Lorraine, France
Edward A. Fox, Virginia Tech University
Norbert Fuhr, Universität Dortmund
Keith Jeffery, RAL-CLRC, UK
Martin Kersten, CWI, The Netherlands
Laszlo Kovacs, MTA-SZTAKI, ASZI, Hungary
Carl Lagoze, Cornell University
Clifford A. Lynch, University of California
A. Desai Narasimhalu, National University of Singapore
Christos Nikolaou, University of Crete and FORTH
Olle Olsson, SICS, Sweden
Andreas Paepcke, Stanford University
Carol Peters, IEI-CNR, Italy
Peter Schäuble, ETHZ, Switzerland
Hans Joerg Schek, ETHZ, Switzerland
Terence R. Smith, University of California Santa Barbara
Ingeborg T. Sølberg, Norwegian University of Science and Technology
Shigeo Sugimoto, University of Library and Information Science
Ulrich Thiel, GMD - IPSI, Germany
C.J. Keith van Rijsbergen, University of Glasgow
Howard Wactlar, Carnegie Mellon University
Stuart Weibel, Online Computer Library Center (OCLC)
Sonia Zillhardt, Bibliothèque Nationale de France

Local Organising Committee

Conference Secretariat:	Tarina Ayazi and Manuela Mennucci
Treasurer and Local Arrangements:	Ettore Ricciardi
Demonstrations:	Stefania Biagioni and Carlo Carlesi

Sponsoring Institutions

CEU Telematics - TMR

ERCIM - European Research Consortium for Informatics and Mathematics

CNR: National Advisory Committee for Physics

CNR: National Advisory Committee for Information Sciences and Technologies

Table of Contents

Invited Talks

Libraries and Digital Property Rights	1
<i>Mark Stefik and Giuliana Lavendel (Xerox Palo Alto Research Center)</i>	
Object Database Support for Digital Libraries	11
<i>Serge Abiteboul (Stanford University and INRIA-Rocquencourt)</i>	

Supporting User Interaction

Enhancing Community and Collaboration in the Virtual Library	25
<i>Rob Procter, Andy McKinlay, Ana Goldenberg (Edinburgh University), Elisabeth Davenport (Queen Margaret College, Edinburgh) and Peter Burnhill, Sheila Cannell (Edinburgh University Library)</i>	
Comprehension and Object Recognition Capabilities for Presentations of Simultaneous Video Key Frame Surrogates	41
<i>Laura Slaughter, Ben Shneiderman and Gary Marchionini (University of Maryland)</i>	

Metadata

Automating the Construction of Authority Files in Digital Libraries: A Case Study	55
<i>James C. French, Allison L. Powell, John L. Pfaltz (Univ. of Virginia) and Eric Schulman (National Radio Astronomy Observatory, Virginia)</i>	
Using Semantic, Geographical, and Temporal Relationships to Enhance Search and Retrieval in Digital Catalogs	73
<i>Klaus Tochtermann, Wolf-Fritz Riekert, Gerlinde Wiest (FAW - Research Institute for Applied Knowledge Processing, Ulm), Jürgen Seggelke and Birgit Mohaupt-Jahr (Federal Environmental Agency, Berlin)</i>	
Metadata Repositories Using PICS	87
<i>Renato Iannella (DSTC Pty Ltd.)</i>	

Information Retrieval I

Relevance Feedback and Query Expansion for Searching the Web: A Model for Searching a Digital Library	99
<i>Alan F. Smeaton and Francis Crimmins (Dublin City University)</i>	

Text Segmentation by Topic	113
<i>Jay M. Ponte and W. Bruce Croft (University of Massachusetts)</i>	

Scalable Text Retrieval for Large Digital Libraries	127
<i>David Hawking (Australian National University)</i>	

Architectures

Awareness Services for Digital Libraries	147
<i>Arturo Crespo and Hector Garcia-Molina (Stanford University)</i>	

Towards a Common Infrastructure for Large-Scale Distributed Applications	173
<i>Christos Nikolaou, Manolis Marazakis, Dimitris Papadakis, Yiorgos Yeorgiannakis (Foundation for Research and Technology (FORTH)) and Jakka Sairamesh (T. J. Watson Research Center, NY)</i>	

Machine Learning + On-line Libraries = IDL	195
<i>Giovanni Semeraro, Floriana Esposito, Donato Malerba, Nicola Fanizzi and Stefano Ferilli (Università degli Studi di Bari)</i>	

Multilingual Information Retrieval

Building a Multilingual Electronic Text Collection of Folk Tales as a Set of Encapsulated Document Objects: An Approach for Casual Users to Browse Multilingual Documents on the Fly	215
<i>Myriam Dartois, Tetsuo Sakaguchi, Shigeo Sugimoto, Koichi Tabata (University of Library and Information Science, Japan), Akira Maeda (Nara Institute of Science and Technology) and Takehisa Fujita (Kyoritsu Women's University)</i>	

Automated Indexing with Thesaurus Descriptors: A Co-occurrence Based Approach to Multilingual Retrieval	233
<i>Reginald Ferber (German National Research Center for Information Technology (GMD-IPSI))</i>	

Cross-Language Information Retrieval in a Multilingual Legal Domain	253
<i>Páraic Sheridan, Martin Braschler and Peter Schäuble (Swiss Federal Institute of Technology (ETH))</i>	

Structured Documents

The Digital Library and Computational Philology: The BAMBI Project ..	269
<i>Andrea Bozzi (Istituto di Linguistica Computazionale - CNR, Pisa) and Sylvie Calabretto (LISI-INSa de Lyon)</i>	

Multivalent Annotations	287
<i>Thomas A. Phelps and Robert Wilensky (University of California, Berkeley)</i>	

A Semantic Network Approach to Semi-structured Documents Repositories	305
<i>Vassilis Christophides, Martin Dörr and Irene Fundulaki</i>	
<i>(Foundation for Research and Technology (FORTH))</i>	

Information Retrieval II

Modelling the Retrieval of Structured Documents Containing Texts and Images	325
<i>Carlo Meghini, Fabrizio Sebastiani and Umberto Straccia</i>	
<i>(Istituto di Elaborazione dell'Informazione - CNR, Pisa)</i>	

Probabilistic Retrieval of OCR Degraded Text Using N-Grams	345
<i>S. M. Harding, W. B. Croft (University of Massachusetts) and</i>	
<i>C. Weir (Lockheed Martin C2 Systems, PA)</i>	

Case Studies

Deposit for Dutch Electronic Publications: Research and Practice in The Netherlands	361
<i>Trudi C. Noordermeer (National Library of The Netherlands)</i>	

Charging for a Digital Library – The Business Model and the Cost Models of the MeDoc Digital Library	375
<i>Michael Breu (Fakultät für Informatik, München) and Ricarda Weber (Technische Universität München)</i>	

Bibliothèque Nationale de France's Audiovisual System: Digital Audio, Video, and Photo Consultation in a Library	387
<i>Sylvie Mony (Bibliothèque Nationale de France)</i>	

The Electronic Colloquium on Computational Complexity (ECCC): A Digital Library in Use	405
<i>J. Bern, C. Damm and Ch. Meinel (Universität Trier)</i>	

Author Index	423
--------------	-----