László Kovács
Norbert Fuhr
Carlo Meghini (Eds.)

LNCS 4675

Research and Advanced Technology for Digital Libraries

11th European Conference, ECDL 2007
Budapest, Hungary, September 2007
Proceedings

Springer

László Kovács Norbert Fuhr
Carlo Meghini (Eds.)

Research and Advanced Technology for Digital Libraries

11th European Conference, ECDL 2007
Budapest, Hungary, September 16-21, 2007
Proceedings

Springer

Volume Editors

László Kovács
Hungarian Academy of Sciences
Computer and Automation Research Institute
Department of Distributed Systems
H-1111 Budapest, XI. Lágymányosi u. 11., Hungary
E-mail: laszlo.kovacs@sztaki.hu

Norbert Fuhr
University of Duisburg-Essen
Information Systems Department of Computational and Cognitive Sciences
47048 Duisburg, Germany
E-mail: norbert.fuhr@uni-due.de

Carlo Meghini
Consiglio Nazionale delle Ricerche
Istituto di Scienza e Tecnologie della Informazione
56124 Pisa, Italy
E-mail: meghini@isti.cnr.it

Library of Congress Control Number: 2007934833

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

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

ISSN 0302-9743
ISBN-10 3-540-74850-4 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-74850-2 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12121076 06/3180 5 4 3 2 1 0

Lecture Notes in Computer Science 4675

Preface

We are proud to present the proceedings of the 11th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2007) which, following Pisa (1997), Heraklion (1998), Paris (1999), Lisbon (2000), Darmstadt (2001), Rome (2002), Trondheim (2003), Bath (2004), Vienna (2005) and Alicante (2006), took place on September 16-21, 2007 in Budapest, Hungary. Over the last 11 years, ECDL has created a strong interdisciplinary community of researchers and practitioners in the field of digital libraries, and has formed a substantial body of scholarly publications contained in the conference proceedings.

ECDL 2007 featured separate calls for paper and poster submissions, resulting in 119 full papers and 34 posters being submitted to the conference. All papers were subject to an in-depth peer-review process; three reviews per submission were produced by a Program Committee of 69 members from 27 countries. In total 36 of 119 full paper submissions were accepted at the Program Committee meeting for presentation at the conference and publication in the proceedings with Springer, resulting in an acceptance rate of 30%. Also, 24 poster/demo submissions and another 15 papers from the full paper submissions were accepted for poster presentation and publication in the proceedings volume.

ECDL 2007 was devoted to discussions about hot issues and applications and primarily provided a forum to reinforce the collaboration of researchers and practicioners.The main conference consisted of 12 technical sessions and a poster /demo session on the following topics: Ontologies, Digital Libraries and the Web, Models, Multimedia and Multilingual Digital Libraries, Grid and Peer-to-Peer, Preservation, User Interfaces, Document Linking, Information Retrieval, Personal Information Management, New DL Applications and User Studies.

The conference featured two panels, which addressed timely and important topics, namely, experiences of DL projects in synergy with the European Commission's initiatives in the panel "On the Move Towards the European Digital Library: BRICKS, TEL, MICHAEL and DELOS converging experiences" chaired by Massimo Bertoncini and the special challenges of Digital Library research and development in the host region of the conference in the panel "Digital Libraries in Central and Eastern Europe: Infrastructure Challenges for the New Europe" chaired by Christine Borgman.

The keynote talk by Seamus Ross (Humanities Computing and Information Management, University of Glasgow) addressed the questions of digital preservation, while the keynote talk by Arne Solvberg (Dept. of Computer and Information Science, Norwegian University of Science and Technology) focused on the challenges of WiFi-Trondheim – an experiment in providing Broadband Everywhere for All.

The preceding tutorials provided further in-depth looks at areas of current interest, including “Thesauri and Ontologies in DLs by Dagobert Soegrel, Introduction to DLs” by Ed Fox, “Approaches for Large Scale Digital Library Infrastructures” by Thomas Risse, and “Building DLs On-Demand by Sharing Content, Services and Computing Resources” by Donatella Castelli.

The workshops, held in conjunction with ECDL2007, covered wide areas of interest: CLEF 2007 – Cross-Language Evaluation Forum, Workshop on “Foundations of Digital Libraries”; LADL 2007 – Cross-Media and Personalized Learning Applications on Top of Digital Libraries, Curriculum Development in Digital Libraries: An Exercise in Developing Lesson Plans , Towards a European Repository Ecology: Conceptualizing Interactions Between Networks of Repositories and Services, NKOS -Networked Knowledge Organization Systems and Services, and Libraries in the Digital Age: What If. . . ?

We would like to take the opportunity to thank everybody who made this conference possible, all the conference participants and presenters, who provided an exciting full-week program of high technical quality. We greatly appreciate the contribution of the Program Committee members, who did an outstanding reviewing job under tight time constraints; and we are grateful to all Chairs and members of the Organization Committee, who worked hard to make the best out of the Conference.

Finally, we would also like to thank the conference sponsors and cooperating agencies: the Computer and Automation Research Institute of the Hungarian Academy of Sciences (MTA SZTAKI), the DELOS Network of Excellence on Digital Libraries, and the Hungarian Tourism Office.

September 2007

László Kovács
Norbert Fuhr
Carlo Meghini

Organization

Organization Committee

General Chair

László Kovács	Department of Distributed Systems, Computer and Automation Research Institute, Hungarian Academy of Sciences, Hungary

Program Co-chairs

Norbert Fuhr	Information Systems Faculty of Engineering Sciences, University of Duisburg-Essen, Germany
Carlo Meghini	Consiglio Nazionale delle Ricerche Istituto di Scienza e Tecnologie dell'Informazione, Italy

Workshops Chairs

Maristella Agosti	University of Padua, Italy
Birte Christensen-Dalsgaard	State and University Library, Denmark

Poster and Demo Chairs

Ulrike Steffens	OFFIS, Germany
José Borbinha	DEI/IST/UTL and INESC-ID, Portugal

Tutorials Chair

Rudi Schmiede	Darmstadt University of Technology, Germany

Publicity and Exhibit Chairs

Yuzuru Tanaka for Asia	Meme Media Laboratory, Hokkaido University, Japan
Jane Hunter for Australia	School of ITEE, Australia
Hussein Suleman for Africa	University of Cape Town, South Africa

Panel Chairs

Seamus Ross	University of Glasgow, UK
Edward Fox	Virginia Tech / Dept. of Computer Science, USA

Doctoral Consortium Chairs

Tiziana Catarci	University of Rome 1, Italy
Nicolas Spyratos	Université de Paris-Sud, France

Local Arrangements Chair

Gusztáv Hencsey	Computer and Automation Research Institute, Hungarian Academy of Sciences, Hungary

Program Committee

Hanne Albrechtsen	Institute of Knowledge Sharing, Denmark
Margherita Antona	FORTH, Greece
Tom Baker	State and University Library, Germany
Nicholas Belkin	Rutgers University, USA
Maria Bieliková	Slovak University of Technology in Bratislava, Slovakia
George Buchanan	University of Wales, Swansea
Gerhard Budin	University of Vienna, Austria
Tiziana Catarci	University of Rome 1, Italy
José H. Canós Cerda	Universidad Politecnica de Valencia, Spain
Hsinchun Chen	University of Arizona, Tucson, USA
Anita S.Coleman	University of Arizona, USA
Gregory Crane	Tufts University, USA
Sally Jo Cunningham	University of Waikato, New Zealand
Mário J. Gaspar da Silva	Universidade de Lisboa, Portugal
Pablo de la Fuente	University of Valladolid, Spain
Susanne Dobratz	Humboldt University, Germany
Boris V.Dobrov	Moscow State University, Russia
Jacques Ducloy	CNRS-INIST, France
Lim Ee-Peng	Nanyang Technological University, Singapore
Floriana Esposito	University of Bari, Italy
Schubert Foo	Nanyang Technological University, Singapore
Edward Fox	Virginia Tech, USA
Richard Furuta	Texas A & M University, USA
Stefan Gradmann	University of Hamburg, Computing Center, Germany
Allan Hanbury	Vienna University of Technology, Austria
Donna Harman	NIST, USA
Djoerd Hiemstra	Twente University, The Netherlands
Jen-Shin Hong	Department of Computer Science, National ChiNan University, Taiwan
Leonid Kalinichenko	Russian Academy of Sciences, Russia
Sarantos Kapidakis	Ionian University, Greece
Claus-Peter Klas	University of Duisburg, Germany
Traugott Koch	Max Planck Digital Library, Germany
Harald Krottmaier	Graz University of Technology, Austria
Carl Lagoze	Cornell University, USA
Mounia Lalmas	Queen Mary University of London, UK
Ronald Larsen	University of Pittsburgh, USA

Ray Larson	University of California, Berkeley, USA
Clifford Lynch	Coalition for Networked Information, USA
Antonio Polo Márquez	University of Extremadura, Spain
Catherine C. Marshall	Microsoft Corporation, Redmond,WA, USA
András Micsik	MTA SZTAKI, Hungary
Reagan Moore	SDSC, USA
Marc Nanard	University of Montpellier, France
Liddy Nevile	La Trobe University, Australia
Fernando López Ostenero	UNED, Spain
Anselmo Penas̃	UNED, Spain
Dimitris Plexousakis	FORTH, Greece
Andy Powell	Eduserv Foundation, UK
Hansen Preben	SICS, Sweden
Andreas Rauber	University of Technology, Vienna, Austria
Thomas Risse	Fraunhofer IPSI, Germany
Laurent Romary	Laboratoire Loria CNRS, France
Lloyd Rutledge	CWI, The Netherlands
J. Alfredo Sánchez	Universidad de las Americas Puebla, Mexico
Heiko Schuldt	University of Basel, Switzerland
Timos Sellis	National Technical University of Athens, Greece
Dagobert Soergel	University of Maryland, USA
Ingeborg Solvberg	Norwegian University of Technology and Science, Norway
Jela Steinerova	Comenius University in Bratislava, Slovakia
Shigeo Sugimoto	Graduate School of Library, Information and Media Studies, University of Tsukuba, Japan
Tamara Sumner	University of Colorado, Boulder, USA
Jesús Tramullas	University of Zaragoza, Spain
Omar Valdiviezo	Universidad de las Americas Puebla, Mexico
Herbert Van de Sompel	Los Alamos National Laboratory, USA
Ian Witten	University of Waikato, New Zealand

Table of Contents

Ontologies

Digital Libraries and the Web

Models

Multimedia and Multilingual DLs

Grid and Peer-to-Peer

Preservation

User Interfaces

Document Linking

Information Retrieval

Personal Information Management

New DL Applications

User Studies

Panels

Posters and Demos

The Semantic GrowBag Algorithm: Automatically Deriving Categorization Systems

Jörg Diederich and Wolf-Tilo Balke

L3S Research Center and Leibniz Universität Hannover, Hanover, Germany
{diederich,balke}@l3s.de

Abstract. Using keyword search to find relevant objects in digital libraries often results in way too large result sets. Based on the metadata associated with such objects, the faceted search paradigm allows users to structure and filter the result set, for example, using a publication type facet to show only books or videos. These facets usually focus on clear-cut characteristics of digital items, however it is very difficult to also organize the actual semantic content information into such a facet. The *Semantic GrowBag* approach, presented in this paper, uses the keywords provided by many authors of digital objects to automatically create light-weight topic categorization systems as a basis for a meaningful and dynamically adaptable *topic facet*. Using such emergent semantics enables an alternative way to filter large result sets according to the objects' content without the need to manually classify all objects with respect to a pre-specified vocabulary. We present the details of our algorithm using the DBLP collection of computer science documents and show some experimental evidence about the quality of the achieved results.

Keywords: faceted search, category generation, higher-order co-occurrence.

1 Introduction

Due to today's sophisticated ranking techniques, the simple keyword search paradigm has been remarkably successful in finding relevant resources in huge data collections, such as digital libraries or even the world wide web. One remaining problem, however, is that users are often unsure which actual keywords to choose so that finding a particular resource often involves several search queries, which then have to be manually refined step-by-step according to the result set of the previous keyword search.

The *faceted search* [1,2,3,4] paradigm makes this process of refining queries explicit and presents the results along with several orthogonal facets, which characterize the result set (e.g., a 'publication type' facet might reveal that there are only two videos among possibly 10,000 relevant results) and thus allow the user to restrict the result set in an easy and intuitive way by exploiting metadata.

This paper is focused on facets based on the actual content of the objects, which allow to restrict the result set to a specific topic. To limit the size of such a *topic facet*, a hierarchical system is typically used to structure the facet, for example, using the Dewey Decimal Classification System, the ACM curriculum or any other cataloguing system (cf. Fig. 1). Such a topic facet can be used in several different ways:

L. Kovács, N. Fuhr, and C. Meghini (Eds.): ECDL 2007, LNCS 4675, pp. 1–13, 2007.