Lecture Notes in Computer Science 3237

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

Carol Peters Julio Gonzalo
Martin Braschler Michael Kluck (Eds.)

Comparative Evaluation of Multilingual Information Access Systems

4th Workshop of the
Cross-Language Evaluation Forum, CLEF 2003
Trondheim, Norway, August 21-22, 2003
Revised Papers

Springer

Volume Editors

Carol Peters
Istituto di Scienza e Tecnologie dell'Informazione
Consiglio Nazionale delle Ricerche (ISTI-CNR)
Via G. Moruzzi 1, 56124 Pisa, Italy
E-mail: carol.peters@isti.cnr.it

Julio Gonzalo
Universidad Nacional de Educación a Distancia (UNED)
Departamento de Lenguajes y Sistemas Informáticos
c/ Juan del Rosal, 16, 28040 Madrid, Spain
E-mail: julio@lsi.uned.es

Martin Braschler
Eurospider Information Technology AG
Schaffhauser Str. 18, 8006 Zürich, Switzerland
E-mail: martin.braschler@eurospider.com

Michael Kluck
Informationszentrum Sozialwissenschaften der
Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e.V. (IZ)
Lennéstr. 30, 53113 Bonn, Germany
E-mail: kluck@bonn.iz-soz.de

Library of Congress Control Number: 2004115726

CR Subject Classification (1998): H.3, I.2, H.4

ISSN 0302-9743
ISBN 3-540-24017-9 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11342052 06/3142 5 4 3 2 1 0

Preface

The fourth campaign of the Cross-language Evaluation Forum (CLEF) for European languages was held from January to August 2003. Participation in this campaign showed a slight rise in the number of participants from the previous year, with 42 groups submitting results for one or more of the different tracks (compared with 37 in 2002), but a steep rise in the number of experiments attempted. A distinctive feature of CLEF 2003 was the number of new tracks and tasks that were offered as pilot experiments. The aim was to try out new ideas and to encourage the development of new evaluation methodologies, suited to the emerging requirements of both system developers and users with respect to today's digital collections and to encourage work on many European languages rather than just those most widely used. CLEF is thus gradually pushing its participants towards the ultimate goal: the development of truly multilingual systems capable of processing collections in diverse media.

The campaign culminated in a two-day workshop held in Trondheim, Norway, 21–22 August, immediately following the 7th European Conference on Digital Libraries (ECDL 2003), and attended by more than 70 researchers and system developers. The objective of the workshop was to bring together the groups that had participated in the CLEF 2003 campaign so that they could report on the results of their experiments. Attendance at the workshop was thus limited to participants in the campaign plus several invited guests with recognized expertise in the multilingual information access field. This volume contains thoroughly revised and expanded versions of the preliminary papers presented at the workshop accompanied by detailed analyses of the results in the various track overview papers and thus provides an exhaustive record of the CLEF 2003 campaign.

CLEF 2003 was conducted within the framework of a project of the Information Society Technologies programme of the European Commission (IST-2000-31002). The campaign was organized in collaboration with the US National Institute of Standards and Technology (NIST) and with the support of the DELOS Network of Excellence for Digital Libraries. The support of NIST and DELOS in the running of the evaluation campaign is gratefully acknowledged. We would also like to thank the other members of the Workshop Steering Committee for their assistance in the coordination of this event.

June 2004

Carol Peters
Julio Gonzalo
Martin Braschler
Michael Kluck

CLEF 2003 Workshop Steering Committee

Martin Braschler, Eurospider Information Technology, Switzerland
Khalid Choukri, Evaluations and Language Resources Distribution Agency, Paris, France
Marcello Federico, Centro per la Ricerca Scientifica e Tecnologica, Istituto Trentino di Cultura, Italy
Julio Gonzalo Arroyo, Departamento de Lenguajes y Sistemas Informáticos, Universidad Nacional de Educación a Distancia, Madrid, Spain
Donna Harman, National Institute of Standards and Technology, USA
Gareth Jones, University of Exeter, UK
Noriko Kando, National Institute of Informatics, Japan
Michael Kluck, IZ Sozialwissenschaften, Bonn, Germany
Bernardo Magnini, Centro per la Ricerca Scientifica e Tecnologica, Istituto Trentino di Cultura, Italy
Douglas W. Oard, University of Maryland, USA
Carol Peters, ISTI-CNR, Pisa, Italy
Mark Sanderson, University of Sheffield, UK
Peter Schäuble, Eurospider Information Technology, Switzerland
Ellen Voorhees, National Institute of Standards and Technology, USA

Table of Contents

Part IV. Cross-Language Question Answering

Part V. Cross-Language Image Retrieval

Part VI. Cross-Language Spoken Document Retrieval