Lecture Notes in Computer Science 2785

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

Springer-Verlag
Berlin Heidelberg GmbH

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

Advances in Cross-Language Information Retrieval

Third Workshop of the
Cross-Language Evaluation Forum, CLEF 2002
Rome, Italy, September 19-20, 2002
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

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

Julio Gonzalo
Universidad Nacional de Educación a Distancia
Lenguajes y Sístemas Informáticos
Ciudad Universitaria, 28040 Madrid, Spain
E-mail: julio@lsi.uned.es

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

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, I.2

ISSN 0302-9743
ISBN 978-3-540-40830-7 ISBN 978-3-540-45237-9 (eBook)
DOI 10.1007/978-3-540-45237-9

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 2003
Originally published by Springer-Verlag Berlin Heidelberg New York in 2003

Typesetting: Camera-ready by author, data conversion by DA-TeX Gerd Blumenstein
Printed on acid-free paper SPIN: 10931776 06/3142 5 4 3 2 1 0

Preface

The third campaign of the Cross-Language Evaluation Forum (CLEF) for European languages was held from January to September 2002. Participation in this campaign showed a slight rise in the number of participants with, 37 groups from both academia and industry and a steep rise in the number of experiments they submitted for one or more of the five official tracks. The campaign culminated in a two-day workshop held in Rome, Italy, 19–20 September, immediately following the Sixth European Conference on Digital Libraries (ECDL 2002), attended by nearly 70 researchers and system developers. The objective of the workshop was to bring together the groups that had participated in CLEF 2002 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 a complete run-down and detailed analysis of the results, and it thus provides an exhaustive record of the CLEF 2002 campaign.

CLEF 2002 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 2003

Carol Peters
Martin Braschler
Julio Gonzalo
Michael Kluck

CLEF 2002 Workshop Steering Committee

Martin Braschler	Eurospider Information Technology AG, Switzerland
Khalid Choukri	Evaluations and Language Resources Distribution Agency, France
Julio Gonzalo Arroyo	Universidad Nacional de Educación a Distancia, Spain
Donna Harman	National Institute of Standards and Technology, USA
Noriko Kando	National Institute of Informatics, Japan
Michael Kluck	IZ Sozialwissenschaften, Bonn, Germany
Patrick Kremer	Institute de Information Scientifique et Technique, Vandoevre, France
Carol Peters	Italian National Research Council, Pisa, Italy
Peter Schäuble	Eurospider Information Technology AG, Switzerland
Laurent Schmitt	Institute de Information Scientifique et Technique, Vandoevre, France
Ellen Voorhees	National Institute of Standards and Technology, USA

Table of Contents

Monolingual Experiments

Mainly Domain-Specific Information Retrieval

Interactive Track

Cross-Language Spoken Document Retrieval

II Cross-Language Systems Evaluation Initiatives, Issues and Results

III Appendix