

Lecture Notes in Computer Science

Edited by G. Goos and J. Hartmanis

146

Research and Development in Information Retrieval

Proceedings, Berlin, May 18–20, 1982

Edited by Gerard Salton and Hans-Jochen Schneider



Springer-Verlag
Berlin Heidelberg New York 1983

Editorial Board

D.R. Barstow, W. Brauer P. Brinch Hansen D. Gries D.C. Luckham
C. Moler A. Pnueli G. Seegmüller J. Stoer N. Wirth

Editors

Gerard Salton

Dept. of Computer Science, Cornell University,
Ithaca, N.Y. 14853, USA

Hans-Jochen Schneider

Technische Universität Berlin, Fachbereich Informatik
Franklinstr. 28–29, 1000 Berlin 10, FRG

CR Subject Classifications (1982): H.3

ISBN 3-540-11978-7 Springer-Verlag Berlin Heidelberg New York

ISBN 0-387-11978-7 Springer-Verlag New York Heidelberg Berlin

Library of Congress Cataloging in Publication Data. Main entry under title: Research and development in information retrieval. (Lecture notes in computer science; 146)

1. Information storage and retrieval systems—Research—Congresses. I. Salton, Gerard.

II. Schneider, Hans-Jochem. III. Series. Z699.A1R47 1983 025'.04'072 82-19621

ISBN 0-387-11978-7 (U.S.)

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically those of translation, reprinting, re-use of illustrations, broadcasting, reproduction by photocopying machine or similar means, and storage in data banks. Under § 54 of the German Copyright Law where copies are made for other than private use, a fee is payable to "Verwertungsgesellschaft Wort", Munich.

© by Springer-Verlag Berlin Heidelberg 1983

Printed in Germany

Printing and binding: Beltz Offsetdruck, Hemsbach/Bergstr.

2145/3140-543210

PREFACE

The Special Interest Group on Information Retrieval of the Association for Computing Machinery in the US (SIGIR/ACM) is organizing every year an international conference on research and development in information retrieval. Normally every second conference takes place outside the USA.

The fifth conference was held at the Technische Universität Berlin, May 18-20, 1982.

This international conference was organized by the Gesellschaft für Informatik (GI) in cooperation with ACM and the British Computer Society (BCS).

The aim of the conference was to present basic research ideas in information storage and retrieval as well as interesting new applications. Papers covering Methodology, Theory and Applications in the fields of

- System Modelling
- Evaluation Problems
- Natural Language Processing
- Automatic Indexing
- Query Languages
- Complexity Problems
- File Organization
- Data Models
- Integration of IRS into DBMS
- New Technology

were invited.

39 submitted papers were reviewed by the following Program Committee members:

PROGRAMME COMMITTEE

W.S. COOPER

School of Library and Information Studies, University of California, Berkeley, Cal. 94720, U.S.A.

V.S. CHERNIAVSKY, TU Braunschweig, Informatik C, Gaußstr. 11,
D-3300 Braunschweig

M. KOLL, National Bureau of Standards, 255 Technology Bldg., Washington, D.C. 20234, U.S.A.

IV

D. KRAFT, Computer Science Dept., Louisiana State University, Baton Rouge, Louisiana 70803, U.S.A.

G. LUŠTIG, TH Darmstadt, Fachbereich Informatik, Alexanderstr. 22a, D-6100 Darmstadt

C.J. VAN RIJSBERGEN, Dept. of Computer Science, University College Dublin, Belfield, Dublin, 4, Ireland

S. ROBERTSON, Centre for Information Science, The City University, Northampton Square, London EC1V OHB/England

G. SALTON, Dept. of Computer Science, Cornell University, Ithaca, New York 14853, U.S.A.

H.-J. SCHEK, IBM Wissenschaftszentrum, Tiergartenstr. 15, D-6900 Heidelberg

H.-J. SCHNEIDER, Technische Universität Berlin, Institut für Angewandte Informatik, Franklinstr. 28/29, D-1000 Berlin 10

who together with the

ORGANIZING COMMITTEE

F. BENGS, TU Berlin

P. BOLLMANN (Chairman), TU Berlin

I. EHMKE, TU Berlin

prepared the conference under the general chairmanship of G. Salton and H.-J. Schneider.

We are greatly indebted to the referees for their and competent work. The help of a number of well-known specialists provided the basis for the success of this conference. We would like to thank all of them for their contributions.

We hope that this book provides a challenge to the information retrieval community and a stimulus to researchers, implementers and users.

October 1982

Gerard Salton
Cornell University. USA

Hans-Jochen Schneider
Technische Universität Berlin

LIST OF AUTHORS

- BAERTSCHI, Martin, Herzogstr. 2, CH-8044 Zürich
- BILLER, H., SOTEC, Software-Technik GmbH & Co., Nonnendammallee 101,
1000 Berlin 13
- BOLLMANN, P., Inst.f.Angew.Informatik, Technische Universität Berlin,
Franklinstr. 28/29, 1000 Berlin 10
- BOOKSTEIN, A., Graduate Library School, University of Chicago,
1100 East 57th Street, Chicago, Il 60637, U.S.A.
- BUCKLEY, C., Dept. of Computer Science, Cornell University, Ithaca,
N.Y. 14853, U.S.A.
- BUELL, D.A., Dept. of Computer Science, Louisiana State University,
Baton Rouge, Louisiana 70803, U.S.A.
- COOPER, W.S., School of Library and Information Studies, University
of California, Berkeley, Cal. 94720, U.S.A.
- CROFT, W.B., Dept. of Computer and Information Science, University of
Massachusetts, Amherst, Mass. 01003, U.S.A.
- DOSZKOCS, T.E., National Library of Medicine, 9627 Lawndale Drive,
Silver Spring MD, U.S.A.
- DREWES, B., Boston University, Metropolitan College, Computer Science,
755 Commonwealth Ave., Boston, Mass. 02215, U.S.A.
- FAURE, J.C., STERIA, 26, Av.de l'Europe, 78140 Velizy, France
- FRAENKEL, A.S., Dept. of Appl. Mathematics, The Weizmann Inst. of
Science, Rehovot, Israel 76 100
- FREI, H.P., Inst.f.Informatik, ETH-Zentrum, CH-8092, Zürich
- HENRICHS, H., GID, Lyoner Str. 44, 6000 Frankfurt-Main 71
- IP, M.Y.L., Dept. of Computer Science, University of Regina, Regina
Sask. S4S 0A 2
- KNORZ, G., FB Informatik, FG Datenverwaltungssysteme II, TH Darmstadt
Karolinenpl. 5, 6100 Darmstadt
- KONRAD, Erhard, Inst.f.Angew.Informatik, Technische Universität
Berlin, Franklinstr. 28/29, 1000 Berlin 10
- KRAFT, D.H., Dept. of Computer Science, Louisiana State University,
Baton Rouge, Louisiana 70703, U.S.A.
- LINGGARD, R.J., Dept. of Computer Science, The Queen's University of
Belfast, Belfast BT7 1NN, Northern Ireland
- LITWIN, W., INRIA, BP 105, Domaine de Voluceau, F-78150 Le Chesnay
France
- MARON, M.E., School of Library and Information Studies, University of
California, Berkeley, Cal. 94720, U.S.A.

- MOR, M., Dept. of Appl. Mathematics, The Weizmann Inst. of Science,
Rehovot, Israel 76100
- MORRISSEY, J., Dept. of Computer Science, University College Dublin,
Belfield, Dublin, Ireland
- MOULINOX, C., INRIA-SIRIUS, BP 105, Domaine de Voluceau,
F-78153 Le Chesnay Cedex
- NELSON, M., The University of Western Ontario, School of Library and
Information Science, London N6A 5B9, Canada
- PERL, Y., Dept. of Mathematics and Computer Science, Bar-Ilan Univer-
sity, Ramat Gan, Israel
- RADECKI, T., Main Library and Scientific, Information Centre, Techni-
cal University of Wroclaw, Wybrzeze Wyspianskiego 27,
Wroclaw/Poland
- RAGHAVAN, V.V., Dept. of Computer Science, University of Regina,
Regina Sask. S4S 0A2, Canada
- ROBERTSON, S.E., Centre f. Information Science, The City University,
Northampton Square, London EC1V OHB
- RUGGLES, L., Dept. of Computer and Information Science, University of
Massachusetts, Amherst, Mass. 01003, U.S.A.
- SALTON, G., Dept. of Computer Science, Cornell University, Ithaca,
N.Y. 14853, U.S.A.
- SMITH, F.J., Dept. of Computer Science, The Queen's University of
Belfast, Belfast BT7 1NN, Northern Ireland
- SMITH, L.C., Grad.School of Libr. & Inf. Science, University of
Illinois, 410 David Kinley Hall, 1407 W. Gregory St., Urbana,
Il. 61801, U.S.A.
- TAGUE, J., The University of Western Ontario, School of Library and
Information Science, London N6A 5B9, Canada
- YU, C.T., Dept. of Information Engineering, University of Illinois,
Chicago Circle, Chicago, Il. 60680
- ZUSE, H., FB Informatik, Rechnerbetrieb, Technische Universität Berlin,
Franklinstr. 28/ 9, 1000 Berlin 10

CONTENTS	Page
Norbert Henrichs	1
The Growing Crisis of Traditional Information Retrieval Systems - what is to follow?	
P.Bollmann, E. Konrad, H. Zuse	13
FAKYR - a Method Base System for Education and Research in In- formation Retrieval	
D. A. Buell, D. H. Kraft	20
LIARS: A Software Environment for Testing Query Processing Strategies	
W. B. Croft	28
The Implementation of a Document Retrieval System	
Joan Morrissey	38
An Intelligent Terminal for Implementing Relevance Feedback on Large Operational Retrieval Systems	
C. Moulinoux, J.C. Faure, W. Litwin	51
MESSIDOR: A Distributed Information Retrieval System	
M. Bärtschi and H.P. Frei	62
Adapting a Data Organization to the Structure of Stored Information	
Horst Biller	80
On the Architecture of a System Integrating Data Base Management and Information Retrieval	
M.E. Maron	98
Probabilistic Approaches to the Document Retrieval Problem	

S.E. Robertson, London, U.K., M.E. Maron, W.S. Cooper The Unified Probabilistic Model for IR	108
Abraham Bookstein Explanation and Generalization of Vector Models in Information Retrieval	118
Tadeusz Radecki Incorporation of Relevance Feedback into Boolean Retrieval Systems	133
G. Salton, C. Buckley, C.T. Yu An Evaluation of Term Dependence Models in Information Retrieval	151
Gerhard Knorz A Decision Theory Approach to Optimal Automatic Indexing	174
Jean Tague and Michael Nelson Simulation of Bibliographic Retrieval Databases Using Hyperterms	194
Vijay V. Raghavan, M.Y.L. Ip Techniques for Measuring the Stability of Clustering: A Comparative Study	209
Bernd Drewes Retrieval of Abstracts by Analogy	238
Tamas E. Doszkocs From Research to Application: The Cite Natural Language Information Retrieval System	251

	Page
Linda C. Smith	263
Machine Intelligence VS. Machine-Aided Intelligence in Information Retrieval: A Historical Perspective	
F.J. Smith and R.J. Linggard	275
Information Retrieval by Voice Input and Output	
A.S. Fraenkel, M.Mor and Y. Perl	289
Is Text Compression by Prefixes and Suffixes Practical?	