

Lecture Notes in Computer Science

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

1392

Springer-Verlag Berlin Heidelberg GmbH

Andreas Barth Michael Breu Albert Endres
Arnoud de Kemp (Eds.)

Digital Libraries in Computer Science: The MeDoc Approach



Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany
Juris Hartmanis, Cornell University, NY, USA
Jan van Leeuwen, Utrecht University, The Netherlands

Volume Editors

Andreas Barth
FIZ Karlsruhe
Postfach 2465, D-76012 Karlsruhe, Germany
E-mail: AB@fiz-karlsruhe.de

Michael Breu
FAST e.V.
Arabellastr. 17, D-81925 München, Germany
E-mail: breu@fast.de

Albert Endres
Universität Stuttgart
Breitwiesenstr. 20/22, D-70565 Stuttgart, Germany
E-mail: endres@informatik.uni-stuttgart.de

Arnoud de Kemp
Springer-Verlag
Tiergartenstr. 17, D-69121 Heidelberg, Germany
E-mail: deKemp@springer.de

Cataloging-in-Publication data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Digital libraries in computer science : the MeDoc approach /
Andreas Barth ... (ed). - Berlin ; Heidelberg ; New York ; Barcelona
; Budapest ; Hong Kong ; London ; Milan ; Paris ; Santa Clara ;
Singapore ; Tokyo : Springer, 1998
(Lecture notes in computer science ; Vol. 1392)
ISBN 3-540-64493-8

CR Subject Classification (1991): H.2-6, C.2, E.4-5, I.4.1, I.7, J.1

ISSN 0302-9743

ISBN 3-540-64493-8 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 1998
Printed in Germany

Typesetting Camera-ready by author
SPIN 10636926 06/3142 - 5 4 3 2 1 0 Printed on acid-free paper

Preface

Electronic publishing and digital libraries have recently become a major new area of development and research in computer science. The interest is caused on one side by the tremendous growth of all types of publications and the need to provide immediate and selective access, on the other side by advances in computer technology that have brought unprecedented digital storage capacity and network transmission power to the desktop of every scientist.

Electronic fulltext and multimedia libraries have the potential of changing the entire publication process, particularly in science and related fields. This will affect many scientists in their roles of authors, reviewers, and readers of scientific literature. It will also allow for new types of information to be disseminated at low cost and high speed that cannot be represented on paper. We refer to visualizations, animations, and simulations that in many fields are crucial for explaining and understanding scientific phenomena. The roles of publishers, distributors, and librarians will also undergo changes because different ways will be found to bring information from the producer to the consumer.

Any project addressing this area will be confronted with technical and non-technical problems at the same time. Authors expect advice on how to prepare information that is to be consumed in an electronic environment. Publishers are concerned about user acceptance, asset protection, and accounting. An academic librarian wants to learn how to administer users and the users want to know where to look for relevant information and how to get at it swiftly and quickly. These are just examples of the questions that come to mind.

For computer scientists, the topic of digital libraries has twofold implications. As in all other fields there will be new ways and means for publication and communication. In addition, many questions raised are genuine research topics in computer science. Examples of the latter are the storage and structuring of information, the search and query operations to be applied, and the intelligence behind user interfaces; but also security aspects, accounting methods, and transaction models have to satisfy new requirements.

This volume documents results of a major digital library pilot project, initiated by the German Informatics Society (GI), and involving authors, publishers, librarians, and computer science institutes in Germany. The project MeDoc (abbreviation for Multimedia Electronic Documents) brought together some 30 academic institutions and several industrial partners in Germany and Switzerland in a two-year effort. These institutions participated in the project either as developers or as pilot users. In the first case, the researchers involved generated novel ideas and concepts, then designed and developed prototype software functions to be used and evaluated by themselves and the pilot users. The representatives of the pilot users defined requirements, evaluated existing tools, converted source texts of books to displayable format, and assessed the use of electronic contents at their sites.

The library contents used in the project came from very different sources. Of highest importance were the books and journals provided by some 15 scientific publishers. Other material came from ACM, the IEEE Computer Society, IBM Germany, and several academic institutions contributing hundreds of research reports

and other online material. Access was also provided to several existing or newly created reference databases.

The papers in this volume present the technical and non-technical answers arrived at during the MeDoc project. After an overview of the entire project, a detailed discussion follows of the design considerations that went into the MeDoc system. This is a prototype of a distributed digital library facility. Several complementary activities provided important results which are not part of the MeDoc system proper. Valuable experiences were obtained during the conversion of books and journals for the electronic library, experiences that authors and publishers will benefit from. Since the MeDoc digital library is operated in a countrywide trial for several months, first feedback on user behavior was obtained.

At several points in the project, close contacts were made to related efforts in Germany, other European countries and the USA. Two of these projects have submitted material for this volume, outlining the state of the art by showing work being done elsewhere. In a final workshop at Schloß Dagstuhl, the contents of the selected papers for this volume were presented and the writeups subsequently critically reviewed by international experts.

As members of the steering team that led the project, we want to give credit to all participants in the project, be they developers, pilot users, librarians, contributing publishers, or sponsors. The remarkable results achieved are due to a strong technical interest in the matter and to dedicated work carried out in a rather short period of time. We appreciate the excellent cooperation and the explorative spirit of all partners involved. We would also like to acknowledge the indispensable funding received for this effort from the German Federal Ministry of Education, Science, Research, and Technology (BMBF) in Bonn.

Karlsruhe, Munich, Stuttgart, Heidelberg in December 1997

Andreas Barth, Michael Breu, Albert Endres, Arnoud de Kemp

Table of Contents

Part I: Project Overview

- The MeDoc Digital Library Project - Its Goals and Major Achievements . . . 1
A. Barth (FIZ Karlsruhe), M. Breu (FAST e.V. München), A. Brüggemann-Klein (TU München), A. Endres (Universität Stuttgart), A. de Kemp (Springer Verlag Heidelberg)

Part II: Design Considerations of the MeDoc System

- The MeDoc System - A Digital Publication and Reference Service for Computer Science 13
D. Boles, J. Meyer, G. Möller (OFFIS Oldenburg), M. Dreger, S. Lohrum (FU Berlin), K. Großjohann (Universität Dortmund), C. Haber, R. Weber (TU München), A. Kusserow (Universität Bonn), D. Menke (FU Hagen)
- Design and Implementation of the MeDoc Fulltext System 21
J. Meyer, H.-J. Appelrath (OFFIS Oldenburg)
- Conception of the User Agent - The User Interface of the MeDoc System . 35
D. Boles, G. Möller, H.-J. Appelrath (OFFIS Oldenburg), A. Kusserow, D. W. Fellner (Universität Bonn)
- Access to Bibliographic Databases - The Provider Agents in MeDoc 51
D. Menke, C. Gnörlich, H. Helbig (FU Hagen)
- Provider Selection - Design and Implementation of the MeDoc Broker 67
M. Dreger, S. Lohrum (FU Berlin), N. Fuhr, K. Großjohann (Universität Dortmund)
- Query Transformation for Heterogeneous Provider Systems in MeDoc 79
K. Großjohann (Universität Dortmund), D. Menke (FU Hagen)
- The MeDoc Communication Protocol 89
M. Dreger, S. Lohrum (FU Berlin), P. Müller (ZIB Berlin)
- The MeDoc Library: Security Aspects 103
A. Brüggemann-Klein, C. Haber (TU München)
- Cost Models and Accounting in MeDoc 115
R. Weber (TU München), M. Breu (FAST e. V. München), A. Endres (Universität Stuttgart)

Part III: Complimentary Activities to the MeDoc System

- Ariadne, an Interactive Navigation and Search System for Computer -
Science Information on the World-Wide Web 127
*M. Dreger, S. Lohrum, H Schweppe (FU Berlin), C D Ziegler (FIZ
Karlsruhe)*
- Grey Literature and Multiple Collections in NCSTRL 145
*S. Adler, W. Lamersdorf, M. Münke, S. Rücker, H. Spahn (Universität
Hamburg), U. Berger (Universität Stuttgart), A. Brüggemann-Klein,
C. Haber (TU München)*
- Digital Publishing for Conference Proceedings on CD-ROM 171
D W. Fellner (Universität Bonn)

Part IV: Application of MeDoc Results and User Experiences

- Books and Journals in MeDoc - Experiences on Bringing Publications Online 179
U. Schwab (Springer Verlag Heidelberg)
- Experiences in Document Conversions During the MeDoc Project 187
H. Feyrer (FH Regensburg)
- Digital Books in MeDoc - A First Analysis of Usage Patterns 193
M. Breu (FAST e.V München)
- Use of MeDoc Offerings in Technical Education 207
A. Spillner (Hochschule Bremen)

Part V: Reports from Related Projects

- The Stanford InfoBus and Its Service Layers: Augmenting the Internet
with Higher-Level Information Management Protocols 213
*M. Röscheisen, M. Baldonado, K. Chang, L. Gravano, S. Ketchpel,
A. Paepcke (Stanford University)*
- Objectives, Results and Conclusions of the European DECOMATE Project:
Delivery of Copyright Material to End Users 231
J. Dijkstra (Tilburg University)
- Author Index** 239