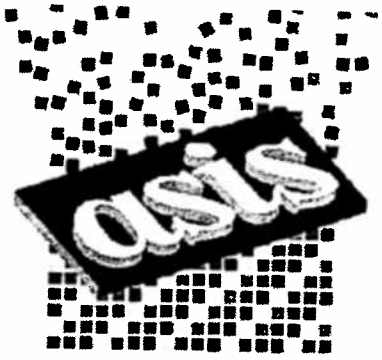
Proceedings of the 8th ASIS SIG/CR Classification Research Workshop

November 2, 1997

Held at the 60th ASIS Annual Meeting
November 1-6, 1997
Washington, D.C.

Efthimis N. Efthimiadis
Editor and Workshop Chair



Printed and Distributed by the
Special Interest Group / Classification Research
American Society for Information Science

ISSN: 2324-9773

The opinions expressed by the contributors to this publication do not necessarily reflect the position or official policy of the American Society for Information Science

Copyright ©1997 by the American Society for Information Science
8720 Georgia Avenue, Suite 501, Silver Sping, MD 20910-3602
Phone: (301) 495-0900
Fax: (301) 495-0810
E-mail: asis@cni.org
Homepage: http://www.asis.org

ISSN: 2324-9773

Preface

General Comments

These proceedings constitute a working copy of the contributions to the 8th ASIS SIG/CR Workshop on Classification Research, held in Washington, D.C., at the ASIS Annual Meeting on Sunday, November 2, 1997. The purpose of publishing working papers is to provide participants and presenters with an opportunity to review the contributions before, during, and after the Workshop. As working papers, the contributions printed here are not in their final form. Therefore, please do not cite these papers. After the Workshop, authors will have an opportunity to revise and edit their papers. Workshop participants are invited to provide the authors or the editor with any comments that they feel will improve the papers or help the authors with their research. Subsequently the Proceedings will be published as an ASIS monograph by Information Today, Inc., under the title *Advances in Classification Research: Volume 8.* This title will be available for purchase as soon as possible.

Again this year, we are pleased with the diversity of viewpoints represented in these papers. The rationale for the Workshop from the beginning has been to showcase the multi-disciplinary bases of classification research. The authors of the papers in this volume have accordingly been asked to write for an audience that may be unfamiliar with the terminology and issues current in the author's own field, and they have tried to follow this suggestion with good humor. The communication across disciplines that the Workshop supports increases the value of these papers for all who take advantage of them as, in seeking to explain assumptions and perspectives that are not shared, participants in the Workshop can deepen and broaden their understanding of classificatory issues in general.

We received papers on a variety of themes for this year's Workshop. Classification systems provide both an overall structure and a set of concepts to guide the information searcher into the domains or the semantics of databases in information systems. However, even in cases where multidimensional classificatory approaches have been implemented such structures seldom come directly forward as guiding instruments supporting the user's search activity. This very important information about the semantics of the databases is often hidden away. Albrechtsen and Jacob suggest that classifications systems should be designed and implemented as visible and ready-to-hand systematic domains in information systems so that they can function as boundary objects in the mediation of database semantics.

The process of thesaurus construction involves the acquisition, analysis and presentation of concepts and terms. The analysis of the domain of the thesaurus is an important part of the process. Nielsen investigates how word association tests can be used as part of the necessary domain analysis.

Automatic classification offers the possibility of improved production efficiencies to large document publishers in print and online environments. Curran and Thompson explore the

ISSN: 2324-9773

possibilities of automating the classification of statutory legal materials and consider strategies currently available for doing so.

In indexing terms may be selected from a redefined (controlled) vocabulary or from the text itself (uncontrolled). Text categorization involves the assignment of terms from a predefined (controlled) vocabulary. Ruiz and Srinivasan present the results of an automatic text categorization study using neural networks.

Finding information on the Web can be a daunting and time-consuming task for most users. Wang and Pouchard present the results of a study of end-user searching of Web resources. Lin and Chan discuss *"Knowledge Class,"* a new method for organizing Web resources for individual users. It is a conceptual building block for Web access and it is similar to a classified mini thesaurus, consisting of hierarchically structured terms on a specific topic or discipline. The diversity of these interests and approaches suggests another exciting Workshop.

Bringing together classification researchers on an annual basis is one of the major accomplishments of the Workshop because we might not otherwise have the opportunity to meet, exchange ideas, and debate issues. Registrants for the Workshop will receive the Proceedings as soon as possible so that the papers can be read before the presentations are given. Once again, we are asking presenters to address the high points of their papers and not to present their papers verbatim. Participation from the audience is one of the most important aspects of the Workshop, so we provide time for questions and discussion as well as breaks for follow-up discussion. All present are invited to participate in advancing the state of classification research.

Organization of these Proceedings

The Proceedings contain two parts. The first section consists of papers submitted by the Workshop presenters arranged in alphabetic order by author. The second section contains brief personal statements of Workshop presenters again arranged in alphabetic order by author.

Acknowledgments

The continuing success of the ASIS SIG/CR Classification Research Workshop is the result of the work of many people. The present Editor would like especially to thank all of those who have been instrumental in organizing and supporting the seven previous Workshops, in reviewing papers, and in editing the Proceedings. We would also like to thank all of the contributors/authors for their interest in the Workshop and their willingness to follow our suggestions for their papers. The ASIS administration and office staff who have supported and subsidized this Workshop and its Proceedings over the past eight years deserve thanks also.

ISSN: 2324-9773

Special thanks go to the Program Committee who reviewed this year's contributions:

Marcia Bates, UCLA
Allyson Carlyle, University of Washington
Clare Begthol, University of Toronto
Raya Fidel, University of Washington
Elin Jacobs, Indiana University
Barbara Kwasnik, Syracuse University
Ray Schwartz, Rutgers University
Paul Solomon, University of North Carolina

I would also like to gratefully acknowledge the general support of the Graduate School of Library and Information Science, University of Washington, and the specific assistance of graduate assistant Peggy Burns. -ENE

Comments or questions may be sent to:

Efthimis N. Efthimiadis
Graduate School of Library and Information Science
University of Washington
233A Suzzallo Library
Box 352930
Seattle, Washington 98195-2930
Phone: (206) 616-6077
Fax: (206) 616-3152
E-mail: efthimis@u.washington.edu

ISSN: 2324-9773

ISSN: 2324-9773

Table of Contents

Preface iii

Presented Papers

Classification Systems as Boundary Objects in Diverse Information Ecologies 1
by Hanne Albrechtsen & Elin Jacob

Automatic Categorization of Statute Documents 19
by Thomas Curran & Paul Thomson

Knowledge Class -- A dynamic structure for subject access on the web 31
by Xia Lin & Lois Mai Chan

The word association test in the methodology of thesaurus construction 41
by Marianne Lykke Nielsen

Automatic Text Categorization Using Neural Networks 58
by Miguel Ruiz & Padmini Srinivasan

End-user Searching of Web Resources: Subject Needs and Zero-hits 69
by Peiling Wang & Line Pouchard

Personal Statements Of Workshop Participants 81

List of Preregistered Participants 91

ISSN: 2324-9773

ISSN: 2324-9773