

Proceedings of the 12th ASIS&T SIG/CR
Classification Research Workshop

November 4, 2001

Held at the 64th ASIS&T Annual Meeting
November 2-8, 2001
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 and Technology

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 and Technology



Copyright © 1997 by the American Society for Information Science and
Technology

1320 Fenwick Lane, Suite 510, Silver Spring, Maryland 20910

Phone: (301) 495-0900

Fax: (301) 495-0810

E-mail: asis@asis.org

Homepage: <http://www.asis.org>

PREFACE

General Comments

These proceedings constitute a working copy of the contributions to the 12th ASIST SIG/CR Workshop on Classification Research, held in Washington, D.C., at the ASIST Annual Meeting on Sunday, November 4, 2001. 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 Therefore, please do not cite these papers.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. In addition, the discussion and commentary for each paper during the Workshop will be audio taped, transcribed and added to the final proceedings. 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 ASIST monograph by Information Today, Inc., under the title *Advances in Classification Research: Volume 12*. 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 multidisciplinary 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.

The 12th SIG/CR workshop will open with a keynote address on "the many uses of classification: enriched thesauri as knowledge sources" that will be given by professor **Dagobert Soergel**, College of Information Studies, University of Maryland. Five papers on a variety of themes will follow in this year's Workshop.

Carlyle and Ranger investigate methods for the systematic retrieval and meaningful display of records that represent works that appear in many editions. So far, retrieval systems have not addressed this issue. The authors propose that one means of achieving retrieval and meaningful display of such works is to discover attributes that could be used to classify records as belonging to a particular work set. Examples of such attributes include author name, title, and Library of Congress classification number. These attributes may be discovered automatically using software that would employ existing information in both authority and bibliographic records. A simulation of automatic attribute identification and record classification records shows that automatic methods of work record classification may be largely successful in achieving their purpose.

Endres-Niggemeyer, Hertenstein, Villiger, and Zieger describe an ontology for WWW summarization in Bone Marrow Transplantation. It is a user-centered text-based ontology. Medical knowledge is represented using first-order logic extended with contexts. The ontology

will serve query scenario formulation, text passage retrieval, and summarization proper in a summarization system. It will be stored and managed by an XML database server.

Detailed information about patients' medical diagnoses is recorded for virtually every encounter with hospital emergency departments. Despite the wide availability of such information and the standardized reporting of diagnosis data with the International Classification for Diseases, Ninth Edition, Clinical Modification (ICD-9-CM), the fine-grained raw data have not proven useful in describing the primary reasons why patients visit the emergency departments. **Haas, Travers, Waller, Hilligoss, Cahill, and Pearce** suggest that a solution to this problem is to use a scheme of diagnosis clusters to group similar codes and they study diagnosis clusters in the context of emergency medicine. Their questions relate to: Where do diagnosis clusters fit among the myriad representations used for medical information? What are the principles by which clusters should be defined? Does a diagnosis cluster scheme designed for one use work for other uses?

Priss argues that the two approaches to knowledge representation, i.e., formal, classical or symbolic approaches on one hand, and biologically inspired, fuzzy, or category-based approaches on the other, should be viewed as complementary forms. Both forms of representation serve a purpose and can be combined. This is especially important with respect to formal ontologies, which currently only use formal, classical representations but which should utilize both approaches.

In order to facilitate subject access interoperability a mechanism must be built that allows the different controlled vocabularies to communicate meaning, relationships, and levels of extension and intension so that different user groups using different controlled vocabularies could access collections across the network. Switching languages, the tools of controlled vocabulary compatibility, consist of a single layer that does not allow for a flexible control of the semantic levels of meaning, relationships, and extension or intension. **Tennis** proposes a multilayered conceptual framework wherein the levels of meaning, relationships and extension and intension are each controlled as individual parameters, rather than in a single switching language.

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 the list of pre-registered participants to the Workshop.

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 eleven 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 ASIST administration and office staff who have supported and subsidized this Workshop and its Proceedings over the past eleven years deserve thanks also. Special thanks go to the Program Committee who reviewed this year's contributions:

Helen Barsky Atkins, ISI, USA
Clare Beghtol, University of Toronto, Canada
Allyson Carlyle, University of Washington, USA
Efthimis N. Efthimiadis, University of Washington, USA
Susanne M Humphrey, National Library of Medicine, NIH, USA
Traugott Koch, Lund University Libraries, Sweden
Barbara Kwasnik, Syracuse University, USA
Ia C. McIlwaine, University College London, UK
Jens-Erik Mai, University of Washington, USA
Philip Smith, Ohio State University, USA
Paul Solomon, University of North Carolina, USA
Paul Thompson, St. Thomas University, USA
Diane Vizine-Goetz, OCLC, USA
Marcia Lei Zeng, Kent State University, USA

I would also like to gratefully acknowledge the generous support of The Information School of the University of Washington, and the specific assistance of graduate assistant Lorie Vik.

-ENE

Comments or questions may be sent to:

Efthimis N. Efthimiadis

The Information School, University of Washington

Box 352840, Seattle, Washington 98195-2840

Phone: (206) 616-6077, Fax: (206) 616-3152, E-mail: efthimis@u.washington.edu

TABLE OF CONTENTS

PREFACE	III
FACILITATING RETRIEVAL OF FICTION WORKS IN ONLINE CATALOGS	1
by Allyson Carlyle and Sara Ranger	
CONSTRUCTING AN ONTOLOGY FOR WWW SUMMARIZATION IN BONE MARROW TRANSPLANTATION (BMT).....	13
by Brigitte Endres-Niggemeyer, Bernd Hertenstein, Claudia Villiger, and Carsten Ziegert	
DEFINING CLINICAL SIMILARITY AMONG ICD-9-CM DIAGNOSIS CODES: DIAGNOSIS CLUSTER SCHEMES.....	31
by Stephanie W. Haas, Debbie A. Travers, Anna E. Waller, Brian Hilligoss, Molly Cahill, and Patricia F. Pearce	
MULTILEVEL APPROACHES TO CONCEPTS AND FORMAL ONTOLOGIES.....	53
by Uta Priss	
LAYERS OF MEANING: DISENTANGLING SUBJECT ACCESS INTEROPERABILITY	67
by Joseph T. Tennis	
LIST OF PRE-REGISTERED PARTICIPANTS	79

