

Antje Düsterhöft, Bernhard Thalheim (Eds.)

**Natural Language Processing
and Information Systems**

**8th International Conference on Applications
of Natural Language to Information Systems**

**June 2003
in Burg (Spreewald), Germany**

Gesellschaft für Informatik 2003

Lecture Notes in Informatics (LNI) - Proceedings

Series of the Gesellschaft für Informatik (GI)

Volume P-29

ISSN 1617-5468

ISBN 3-88579-358-X

Volume Editors

Prof. Dr. Antje Düsterhöft

HS-Wismar

FB Elektrotechnik und Informatik

Postfach 1210

D-23952 Wismar

E-Mail: duest@et.hs-wismar.de

Prof. Dr. Bernhard Thalheim

Brandenburgische Technische Universität Cottbus

Institut für Informatik

Postfach 101344

D-03013 Cottbus

E-Mail: thalheim@informatik.tu-cottbus.de

Series Editorial Board

Heinrich C. Mayr, Universität Klagenfurt, Austria (Chairman, mayr@ifit.uni-klu.ac.at)

Jörg Becker, Universität Münster, Germany

Ulrich Furbach, Universität Koblenz, Germany

Axel Lehmann, Universität der Bundeswehr München, Germany

Peter Liggesmeyer, Universität Potsdam, Germany

Ernst W. Mayr, Technische Universität München, Germany

Heinrich Müller, Universität Dortmund, Germany

Heinrich Reinermann, Hochschule für Verwaltungswissenschaften Speyer, Germany

Karl-Heinz Rödiger, Universität Bremen, Germany

Sigrid Schubert, Universität Dortmund, Germany

Dissertations

Dorothea Wagner, Universität Konstanz, Germany

Seminars

Reinhard Wilhelm, Universität des Saarlandes, Germany

© Gesellschaft für Informatik, Bonn 2003

printed by Köllen Druck+Verlag GmbH, Bonn

Preface

Since 1995 the NLDB conference has aimed at bringing together researchers, industrial and potential users interested in various applications of NATURAL LANGUAGE in the DATABASE and INFORMATION SYSTEMS field.

The integration of databases and natural language has been an utopia for a long time. Nowadays, this is an accessible convergent point on which a lot of researchers are focusing, mainly due to the large progress of research in natural language and to the development of new technologies which allow the storage of real semantic electronic dictionaries. Each aspect of an information system life cycle may be improved by natural language techniques: database design (specification, validation, conflicts resolution), database query languages and application programming that use new software engineering research (e.g. natural language program specifications). Furthermore, natural language based query languages and user interfaces facilitate the access to software systems for anyone and allow for new paradigms in the usage of computerized services. As information systems are now evolving into the communication area, the term databases should be considered in the broader sense of information and communication systems. The NLDB'2003 contributions are a balanced mix of full paper reports and extended abstracts from research and application giving a broad insight into the state of the art concerning problems and solutions within the context of natural language processing and information systems. The selected papers are assigned to the major topics: natural language for conceptual modelling, information retrieval and information extraction, linguistic resources for dialogue modelling, natural language for database querying, referencing and categorization, as well as building ontologies for web applications.

We thank all authors for their interesting papers and we also take pleasure in thanking those who have contributed to the realization of the conference and of these proceedings, especially Karla Kersten, Aleksander Binemann-Zdanowicz and Thomas Kobienia.

Wismar and Cottbus, in June 2003

Antje Düsterhöft
Bernhard Thalheim

Programme Committee:

Diego Mollá Aliod, Macquarie University, Australia
Kenji Araki, Hokkaido University, Japan
Alfs T. Bertziss, University of Pittsburgh, USA
Mokrane Bouzeghoub, PRiSM, Université de Versailles, France
Hans Burg, Ordina Alignment Consulting, The Netherlands
Key-Sun Choi, NHK Science and Technology Research Lab., Japan
Gary A Coen, Boeing, USA
Isabelle Comyn-Wattiau, CEDRIC/CNAM, France
Walter Daelemans, University of Antwerp, Belgium
Antje Düsterhöft, University of Wismar, Germany
Günther Fliedl, Universität Klagenfurt, Austria
Alexander Gelbukh, Instituto Politecnico Nacional, Mexico
Rafael Muñoz Guillena, Universidad de Alicante, Spain
Jon Atle Gulla, Borgegian Institute of Technology, Norway
Helmut Horacek, Universität Saarbrücken, Germany
Paul Johannesson, Stockholm University, Sweden
Zoubida Kedad, PRiSM, Université de Versailles, France
Christian Kop, Universität Klagenfurt, Austria
Winfried Lenders, Universität Bonn, Germany
Jana Lewerenz, sd&m Düsseldorf, Germany
Robert Luk, Hong Kong Polytechnic University, Hong Kong

Heinrich C. Mayr, Universität Klagenfurt, Austria
Paul McFetridge, Simon Fraser University, Canada
Elisabeth Metais, CEDRIC/CNAM, France
Farid Meziane, Salford University, UK
Ana Maria Moreno, Universidad Politecnica de Madrid, Spain
Kazunori Muraki, NEC Custum Tecnica Ltd., Japan
Jian-Yun Nie, Université de Montréal, Canada
Odile Piton, Université Paris I Panthéon-Sorbonne, France
Reind van de Riet, Vrije Universiteit Amsterdam, The Netherlands
Hae-Chang Rim, Korea University, Korea
Hongchi Shi, University of Missouri-Columbia, USA
Ishizaki Shun, Keio University, Japan
Vijay Sugumaran, Oakland University Rochester, USA
Veda Storey, Georgia State University, USA
Lua Km Teng, National University of Singapore, Singapore
Bernhard Thalheim, University of Cottbus, Germany
Babis Theodoulidis, University of Surrey, UK
Benkt Wangler, University of Skövde, Sweden
Hans Weigand, Tilburg University, The Netherlands
Werner Winiwarter, University of Vienna, Austria
Christian Winkler, Universität Klagenfurt, Austria
Mustafa Yaseen, Amman University of Applied Sciences, Jordan

Additional Reviewers:

Per Backlund, University of Skövde, Sweden
Eva Söderström, University of Skövde, Sweden

Organizing Committee:

chair: Vojtech Vestenický, BTU Cottbus
Aleksander Binemann-Zdanowicz, BTU Cottbus
Antje Düsterhöft, University of Applied Sciences Wismar
Carola Kadow, University of Applied Sciences Wismar
Karla Kersten, BTU Cottbus
Thomas Kobienia, BTU Cottbus

Contents

Invited Paper

K.-R. Fellbaum Speech Input and Output Technology – State of the Art and Selected Applications	7
--	---

Conference Papers

S. Armstrong, A. Clark, G. Coray, M. Georgescu, V. Pallotta, A. Popescu-Belis, D. Portabella, M. Rajman, M. Starlander Natural Language Queries on Natural Language Data: a Database of Meeting Dialogues	14
I. A. Bolshakov, A. Gelbukh On Detection of Malapropisms by Multistage Collocation Testing	28
V. Boonjing, C. Hsu Natural Language Interaction Using a Scalable Reference Dictionary	42
A. Burton-Jones, V. C. Storey, V. Sugumaran, P. Ahluwalia Assessing the Effectiveness of the DAML Ontologies for the Semantic Web	56
R. Camps, J. Daudé Improving the Efficacy of Approximate Searching by Personal-Name	70
P. Cimiano Ontology-Driven Discourse Analysis in GenIE	77
G. Fliedl, C. Kop, H. C. Mayr From Scenarios to KCPM Dynamic Schemas: Aspects of Automatic Mapping	91
G. Gardarin, H. Kou, K. Zetourni, X. Meng, H. Wang SEWISE : An Ontology-based Web Information Search Engine	106
A. Gelbukh, M. Alexandrov, A. Bourek , P. Makagonov Selection of Representative Documents for Clusters in a Document Collection	120
E. Kapetanios, D. Baer, P. Groenewoud Simplifying Syntactic and Semantic Parsing of NL Based Queries in Advanced Application Domains	127
H. Kou, G. Gardarin, K. Zeitouni Approaches to Feature Selection for Document Categorization	141

K. C. Lan, K. S. Ho Accessing Financial News Using Dialogues	155
M. Martinovic, G. Sampath, R. Wagner, S. Briening A Model of USENET Newsgroups Dynamics: Implementation and Results	168
F. Meziane, M. Khairudin Kasiran Extracting Unstructured Information from the WWW to Support Merchant Existence in eCommerce	175
B. Navarro, M. Palomar, P. Mart3nez-Barco A General Proposal to Multilingual Information Access Based on Syntactic-Semantic Patterns	186
O. Piton, T. Grass, D. Maurel Linguistic Resource for NLP: Ask for “Die Drei Musketiere” and meet “Les Trois Mousquetaires”	200
G. Ramakrishnanan, P. Bhattacharyya Text Representation with WordNet Synsets using Soft Sense Disambiguation	214
I. Renz, A. Ficzy, H. Hitzler Keyword Extraction for Text Characterization	228
N. Stratica, L. Kosseim, B. C. Desai NLIDB Templates for Semantic Parsing	235
K. Thirunarayan, A. Berkovich, S. Grace, D. Sokol Information Extraction for Reorganizing Specifications	242