

*Commenced Publication in 1973*

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

## Editorial Board

David Hutchison

*Lancaster University, UK*

Takeo Kanade

*Carnegie Mellon University, Pittsburgh, PA, USA*

Josef Kittler

*University of Surrey, Guildford, UK*

Jon M. Kleinberg

*Cornell University, Ithaca, NY, USA*

Friedemann Mattern

*ETH Zurich, Switzerland*

John C. Mitchell

*Stanford University, CA, USA*

Moni Naor

*Weizmann Institute of Science, Rehovot, Israel*

Oscar Nierstrasz

*University of Bern, Switzerland*

C. Pandu Rangan

*Indian Institute of Technology, Madras, India*

Bernhard Steffen

*University of Dortmund, Germany*

Madhu Sudan

*Massachusetts Institute of Technology, MA, USA*

Demetri Terzopoulos

*University of California, Los Angeles, CA, USA*

Doug Tygar

*University of California, Berkeley, CA, USA*

Moshe Y. Vardi

*Rice University, Houston, TX, USA*

Gerhard Weikum

*Max-Planck Institute of Computer Science, Saarbruecken, Germany*

Christian Kop Günther Fliedl  
Heinrich C. Mayr Elisabeth Métais (Eds.)

# Natural Language Processing and Information Systems

11th International Conference on Applications  
of Natural Language to Information Systems, NLDB 2006  
Klagenfurt, Austria, May 31 – June 2, 2006  
Proceedings

## Volume Editors

Christian Kop  
Günther Fliedl  
Heinrich C. Mayr  
Alpen-Adria Universität Klagenfurt  
Institute of Business Informatics and Application Systems  
Klagenfurt, Austria  
E-mail: {chris, fliedl, mayr}@ifit.uni-klu.ac.at

Elisabeth Métais  
CNAM, Chaire d'Informatique d'Entreprise  
292 rue Saint-Martin, 75141 Paris, France  
E-mail: metais@cnam.fr

Library of Congress Control Number: 2006926265

CR Subject Classification (1998): H.2, H.3, I.2, F.3-4, H.4, C.2

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN 0302-9743  
ISBN-10 3-540-34616-3 Springer Berlin Heidelberg New York  
ISBN-13 978-3-540-34616-6 Springer 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. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media  
springer.com

© Springer-Verlag Berlin Heidelberg 2006  
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India  
Printed on acid-free paper SPIN: 11765448 06/3142 5 4 3 2 1 0

# Preface

Information systems and natural language processing are fundamental fields of research and development in informatics. The combination of both is an exciting and future-oriented field which has been addressed by the NLDB conference series since 1995. There are still many open research questions but also an increasing number of interesting solutions and approaches.

NLDB 2006 with its high-quality contributions tersely reflected the current discussion and research: natural language and/or ontology-based information retrieval, question-answering methods, dialog processing, query processing as well as ontology- and concept creation from natural language. Some papers presented the newest methods for parsing, entity recognition and language identification which are important for many of the topics mentioned before. In particular, 53 papers were submitted by authors from 14 nations. From these contributions, the Program Committee, based on 3 peer reviews for each paper, selected 17 full and 5 short papers, thus coming up with an overall acceptance rate of 32% (41% including short papers).

Many persons contributed to making NLDB 2006 a success. First we thank all authors for their valuable contributions. Secondly, we thank all members of the Program Committee for their detailed reviews and discussion. Furthermore we thank the following people for their substantial organizational collaboration: Kerstin Jörgl, who did a lot of work to compose these proceedings, our Conference Secretary, Christine Seger, Stefan Ellersdorfer for his technical support, and Jürgen Vöhringer and Christian Winkler, who provided additional last-minute reviews.

This year, NLDB was a part of a multi-conference event on Information Systems: UNISCON — United Information Systems Conference. Thus, participants could get into scientific contact with experts from more technical (ISTA 2006) or more business-oriented (BIS 2006) fields. In either case, they profited from UNISCON's organizational environment. We, therefore, express our thanks also to the UNISCON organization team: Markus Adam, Jörg Kerschbaumer and all the students who supported the participants during the NLDB 2006 conference.

March 2006

Christian Kop  
Günther Fliedl  
Heinrich C. Mayr  
Eliabeth Métais

# Organization

## Conference Co-chairs

|                  |                                    |
|------------------|------------------------------------|
| Christian Kop    | Alpen-Adria Universität Klagenfurt |
| Günther Fliedl   | Alpen-Adria Universität Klagenfurt |
| Heinrich C. Mayr | Alpen-Adria Universität Klagenfurt |
| Elisabeth Métais | Cedric Laboratory CNAM, Paris      |

## Organization and Local Arrangements

|                     |                                    |
|---------------------|------------------------------------|
| Markus Adam         | Alpen-Adria Universität Klagenfurt |
| Stefan Ellersdorfer | Alpen-Adria Universität Klagenfurt |
| Kerstin Jörgl       | Alpen-Adria Universität Klagenfurt |
| Christine Seger     | Alpen-Adria Universität Klagenfurt |
| Christian Winkler   | Alpen-Adria Universität Klagenfurt |

## Program Committee

|                        |                                                        |
|------------------------|--------------------------------------------------------|
| Kenji Araki            | Hokkaido University, Japan                             |
| Akhilesh Bajaj         | University of Tulsa, USA                               |
| Mokrane Bouzeghoub     | PRiSM, Université de Versailles, France                |
| Andrew Burton-Jones    | University of British Columbia, Canada                 |
| Roger Chiang           | University of Cincinnati, USA                          |
| Gary A. Coen           | Boeing, USA                                            |
| Isabelle Comyn-Wattiau | CEDRIC/CNAM, France                                    |
| Antje Düsterhöft       | University of Wismar, Germany                          |
| Günther Fliedl         | Universität Klagenfurt, Austria                        |
| Alexander Gelbukh      | Instituto Politecnico Nacional, Mexico                 |
| Nicola Guarino         | CNR, Italy                                             |
| Jon Atle Gulla         | Norwegian University of Science and Technology, Norway |
| Karin Harbusch         | Universität Koblenz-Landau, Germany                    |
| Helmut Horacek         | Universität des Saarlandes, Germany                    |
| Cecil Chua Eng Huang   | Nanyang Technological University, Singapore            |
| Paul Johannesson       | Stockholm University, Sweden                           |
| Zoubida Kedad          | PRiSM, Université de Versailles, France                |
| Christian Kop          | University of Klagenfurt, Austria                      |
| Leila Kosseim          | Concordia University, Canada                           |

|                            |                                               |
|----------------------------|-----------------------------------------------|
| Nadira Lammari             | CEDRIC/CNAM, France                           |
| Winfried Lenders           | Universität Bonn, Germany                     |
| Jana Lewerenz              | sd&m Düsseldorf, Germany                      |
| Stephen Liddle             | Brigham Young University, USA                 |
| Deryle Lonsdale            | Brigham Young University, USA                 |
| Robert Luk                 | Hong Kong Polytechnic University, Hong Kong   |
| Heinrich C. Mayr           | University of Klagenfurt, Austria             |
| Elisabeth Métais           | CEDRIC/CNAM, France                           |
| Farid Meziane              | Salford University, UK                        |
| Luisa Mich                 | University of Trento, Italy                   |
| Diego Mollá Aliod          | Macquarie University, Australia               |
| Andrés Montoyo             | Universidad de Alicante, Spain                |
| Ana Maria Moreno           | Universidad Politecnica de Madrid, Spain      |
| Rafael Muñoz               | Universidad de Alicante, Spain                |
| Günter Neumann             | DFKI, Germany                                 |
| Jian-Yun Nie               | Université de Montréal, Canada                |
| Manual Palomar             | Universidad de Alicante, Spain                |
| Sandeep Purao              | Pennsylvania State University, USA            |
| Odile Piton                | Université Paris I Panthéon-Sorbonne, France  |
| Yacine Rezgui              | University of Salford, UK                     |
| Reind van de Riet          | Vrije Universiteit Amsterdam, The Netherlands |
| Hae-Chang Rim              | Korea University, Korea                       |
| Veda Storey                | Georgia State University, USA                 |
| Vijay Sugumaran            | Oakland University Rochester, USA             |
| Bernhard Thalheim          | Kiel University, Germany                      |
| Krishnaprasad Thirunarayan | Wright State University, USA                  |
| Juan Carlos Trujillo       | Universidad de Alicante, Spain                |
| Luis Alfonso Ureña         | Universidad de Jaén, Spain                    |
| Sunil Vadera               | University of Salford, UK                     |
| Panos Vassiliadis          | University of Ioannina, Greece                |
| Jürgen Vöhringer           | University of Klagenfurt, Austria             |
| Hans Weigand               | Tilburg University, The Netherlands           |
| Werner Winiwarter          | University of Vienna, Austria                 |
| Christian Winkler          | University of Klagenfurt, Austria             |

## External Referees

Birger Andersson  
 Maria Bergholtz  
 Miguel Ángel García Cumbreiras  
 Theodore Dalamagas  
 Hiroshi Echizen-ya  
 Yasutomo Kimura  
 Nadia Kiyavitskaya  
 Francisco Javier Ariza López

Borja Navarro  
Lluís Padró  
Hideyuki Shibuki  
Dariusz Strasunskas  
Stein L. Tomassen  
Sonia Vazquez  
Chih-Sheng Yang  
Nicola Zeni

### **Organized by:**

NLDB was organized by the Institute of Business Informatics and Applications Systems, Alpen-Adria University of Klagenfurt, Austria.

# Table of Contents

## Concepts Extraction and Ontology

|                                                                                                                                                                          |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| An Automated Multi-component Approach to Extracting Entity Relationships from Database Requirement Specification Documents<br><i>Siqing Du, Douglas P. Metzler</i> ..... | 1  |
| Function Point Extraction Method from Goal and Scenario Based Requirements Text<br><i>Soonhwang Choi, Sooyong Park, Vijayan Sugumaran</i> .....                          | 12 |
| Unsupervised Keyphrase Extraction for Search Ontologies<br><i>Jon Atle Gulla, Hans Olaf Borch, Jon Espen Ingvaldsen</i> .....                                            | 25 |
| Studying Evolution of a Branch of Knowledge by Constructing and Analyzing Its Ontology<br><i>Pavel Makagonov, Alejandro Ruiz Figueroa, Alexander Gelbukh</i> .....       | 37 |

## Ontologies and Task Repository Utilization

|                                                                                                                                         |    |
|-----------------------------------------------------------------------------------------------------------------------------------------|----|
| Document Space Adapted Ontology: Application in Query Enrichment<br><i>Stein L. Tomassen, Jon Atle Gulla, Darijus Strasunskas</i> ..... | 46 |
| The Language of Folksonomies: What Tags Reveal About User Classification<br><i>Csaba Veres</i> .....                                    | 58 |
| A Task Repository for Ambient Intelligence<br><i>Porf rio Filipe, Nuno Mamede</i> .....                                                 | 70 |

## Query Processing

|                                                                                                                                             |     |
|---------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Formulating Queries for Assessing Clinical Trial Eligibility<br><i>Deryle Lonsdale, Clint Tustison, Craig Parker, David W. Embley</i> ..... | 82  |
| Multi-lingual Web Querying: A Parametric Linguistics Based Approach<br><i>Epaminondas Kapetanios, Vijayan Sugumaran, Diana Tanase</i> ..... | 94  |
| Using Semantic Knowledge to Improve Web Query Processing<br><i>Jordi Conesa, Veda C. Storey, Vijayan Sugumaran</i> .....                    | 106 |



## Information Retrieval and Dialog Processing

|                                                                                                                            |     |
|----------------------------------------------------------------------------------------------------------------------------|-----|
| Using Semantic Constraints to Improve Question Answering<br><i>Jamileh Yousefi, Leila Kosseim</i> .....                    | 118 |
| An Abstract Model of Man-Machine Interaction Based on Concepts<br>from NL Dialog Processing<br><i>Helmut Horacek</i> ..... | 129 |
| The PHASAR Search Engine<br><i>Cornelis H.A. Koster, Olaf Seibert, Marc Seutter</i> .....                                  | 141 |

## NLP Techniques

|                                                                                                                                                            |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Language Identification in Multi-lingual Web-Documents<br><i>Thomas Mandl, Margaryta Shramko, Olga Tartakovski,<br/>Christa Womser-Hacker</i> .....        | 153 |
| DILUCT: An Open-Source Spanish Dependency Parser Based on<br>Rules, Heuristics, and Selectional Preferences<br><i>Hiram Calvo, Alexander Gelbukh</i> ..... | 164 |
| Fine Tuning Features and Post-processing Rules to Improve Named<br>Entity Recognition<br><i>Óscar Ferrández, Antonio Toral, Rafael Muñoz</i> .....         | 176 |
| The Role and Resolution of Textual Entailment in Natural Language<br>Processing Applications<br><i>Zornitsa Kozareva, Andrés Montoyo</i> .....             | 186 |

## Short Paper Session I

|                                                                                                                       |     |
|-----------------------------------------------------------------------------------------------------------------------|-----|
| An Information Retrieval Approach Based on Discourse Type<br><i>D.Y. Wang, R.W.P. Luk, K.F. Wong, K.L. Kwok</i> ..... | 197 |
| Natural Language Updates to Databases Through Dialogue<br><i>Michael Minock</i> .....                                 | 203 |

## Short Paper Session II

|                                                                                                                                                          |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Automatic Construction of a Japanese Onomatopoeic Dictionary Using<br>Text Data on the WWW<br><i>Manabu Okumura, Atsushi Okumura, Suguru Saito</i> ..... | 209 |
|----------------------------------------------------------------------------------------------------------------------------------------------------------|-----|

|                                                                                                                                                                                                                     |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Category-Based Audience Metrics for Web Site Content Improvement<br>Using Ontologies and Page Classification<br><i>Jean-Pierre Norquet, Benjamin Tshibasus-Kabeya,<br/>Gianluca Bontempi, Esteban Zimányi</i> ..... | 216 |
| Automatic Turkish Text Categorization in Terms of Author, Genre and<br>Gender<br><i>M. Fatih Amasyalı, Banu Diri</i> .....                                                                                          | 221 |
| <b>Author Index</b> .....                                                                                                                                                                                           | 227 |