

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

New York University, NY, 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

Andrés Montoyo Rafael Muñoz
Elisabeth Métais (Eds.)

Natural Language Processing and Information Systems

10th International Conference on Applications
of Natural Language to Information Systems, NLDB 2005
Alicante, Spain, June 15-17, 2005
Proceedings

Volume Editors

Andrés Montoyo
Universidad de Alicante
Departamento de Lenguajes y Sistemas Informáticos
Apartado de correos, 99, 03080 Alicante, Spain
E-mail: montoyo@dlsi.ua.es

Rafael Muñoz
Universidad de Alicante
Departamento de Lenguajes y Sistemas Informáticos
Apartado de correos, 99, 03080 Alicante, Spain
E-mail: rafael@dlsi.ua.es

Elisabeth Métails
Cedric Laboratory, CNAM
292 rue Saint-Martin, 75003 Paris, France
E-mail: elsa@cnam.fr

Library of Congress Control Number: Applied for

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

ISSN 0302-9743
ISBN-10 3-540-26031-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-26031-8 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

springeronline.com

© Springer-Verlag Berlin Heidelberg 2005
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik
Printed on acid-free paper SPIN: 11428817 06/3142 5 4 3 2 1 0

Preface

NLDB 2005, the 10th International Conference on Applications of Natural Language to Information Systems, was held on June 15–17, 2005 at the University of Alicante, Spain. Since the first NLDB conference in 1995 the main goal has been to provide a forum to discuss and disseminate research on the integration of natural language resources in information system engineering.

The development and convergence of computing, telecommunications and information systems has already led to a revolution in the way that we work, communicate with each other, buy goods and use services, and even in the way that we entertain and educate ourselves. The revolution continues, and one of its results is that large volumes of information will increasingly be held in a form which is more natural for users than the data presentation formats typical of computer systems of the past. Natural language processing (NLP) is crucial in solving these problems, and language technologies will make an indispensable contribution to the success of information systems. We hope that NLDB 2005 was a modest contribution to this goal.

NLDB 2005 contributed to advancing the goals and the high international standing of these conferences, largely due to its Program Committee, composed of renowned researchers in the field of natural language processing and information system engineering. Papers were reviewed by three reviewers from the Program Committee. This clearly contributed to the significant number of papers submitted (95). Twenty-nine were accepted as regular papers, while 18 were accepted as short papers.

We would like to express here our thanks to all the reviewers for their quick and excellent work. We extend these thanks to our invited speakers, Ruslan Mitkov and Branimir Boguraev, for their valuable contribution, which undoubtedly increased the interest in the conference. We are also indebted to a number of individuals for taking care of specific parts of the conference program, specially to Miguel Angel Varó who built and maintained all Web services for the conference.

March 2005

Andres Montoyo
Rafael Muñoz
Elisabeth Métais

Organization

Conference Chairs

Rafael Muñoz (University of Alicante, Spain)

Elisabeth Métais (CEDRIC/CNAM, France)

Program Chair

Andres Montoyo (University of Alicante, Spain)

Organization Committee

Patricio Martínez-Barco (University of Alicante, Spain)

Andres Montoyo (University of Alicante, Spain)

Paloma Moreda (University of Alicante, Spain)

Rafael Muñoz (University of Alicante, Spain)

Maximiliano Saiz (University of Alicante, Spain)

Armando Suárez (University of Alicante, Spain)

Elisa Noguera (University of Alicante, Spain)

Program Committee

Kenji Araki (Hokkaido University, Japan)

Mokrane Bouzeghoub (PRiSM, Université de Versailles, France)

Gary A. Coen (Boeing, USA)

Isabelle Comyn-Wattiau (CEDRIC/CNAM, France)

Günther Fliedl (Universität Klagenfurt, Austria)

Alexander Gelbukh (Instituto Politécnico Nacional, Mexico)

Jon Atle Gulla (Norwegian University of Science and Technology, Norway)

Harmain Harmain (United Arab Emirates University, United Arab Emirates)

Helmut Horacek (Universität des Saarlandes, Germany)

Paul Johannesson (Stockholm University, Sweden)

Zoubida Kedad (PRiSM, Université de Versailles, France)

Nadira Lammari (CEDRIC/CNAM, France)

Jana Lewerenz (sd&m Düsseldorf, Germany)

Robert Luk (Hong Kong Polytechnic University, Hong Kong, China)

Bernardo Magnini (IRST, Italy)

Paul McFetridge (Simon Fraser University, Canada)

Elisabeth Métais (CEDRIC/CNAM, France)

Farid Meziane (Salford University, UK)

Luisa Mich (University of Trento, Italy)
Ruslan Mitkov (University of Wolverhampton, UK)
Ana Maria Moreno (Universidad Politécnica de Madrid, Spain)
Diego Mollá Aliod (Macquarie University, Australia)
Andrés Montoyo (Universidad de Alicante, Spain)
Rafael Muñoz (Universidad de Alicante, Spain)
Jian-Yun Nie (Université de Montréal, Canada)
Manuel Palomar (Universidad de Alicante, Spain)
Odile Piton (Université Paris I Panthéon-Sorbonne, France)
Reind van de Riet (Vrije Universiteit Amsterdam, The Netherlands)
Hae-Chang Rim (Korea University, Korea)
Vijay Sugumaran (Oakland University Rochester, USA)
Veda Storey (Georgia State University, USA)
Bernhard Thalheim (Kiel University, Germany)
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)
Hans Weigand (Tilburg University, The Netherlands)
Werner Winiwarter (University of Vienna, Austria)
Christian Winkler (Universität Klagenfurt, Austria)
Stanislaw Wrycza (University of Gdansk, Poland)

Additional Reviewers

Andrea, Mulloni	McFetridge, Paul
Bergholtz, Maria	Montejo Ráez, Arturo
Isabelle, Comyn-Wattiau	Moreda, Paloma
Nadira, Lammari	Navarro, Borja
Del Jesus Diaz, Maria Jose	Pekar, Viktor
Echizen-ya, Hiroshi	Peral, Jesús
Evans, Richard	Prost, Jean-Philippe
Ferrández, Antonio	Puscasu, Georgiana
García Cumbresas, Miguel Ángel	Roger, Sandra
Ha, Le An	Rzepka, Rafal
Haddad, Hatem	Sasaoka, Hisayuki
Kabilan, Vandana	Selima, Besbes
Kozareva, Zortnisa	Storey, Veda
Llopis, Fernando	Vadera, Sunil
Maria, Bergholtz	Vazquez, Sonia
Martínez-Barco, Patricio	Xue, Xiaohui
Martin-Valdivia, M.Teresa	

Table of Contents

Regular Papers

Extracting Semantic Taxonomies of Nouns from a Korean MRD Using a Small Bootstrapping Thesaurus and a Machine Learning Approach	1
<i>SeonHwa Choi and HyukRo Park</i>	
On the Transformation of Sentences with Genitive Relations to SQL Queries	10
<i>Zsolt T. Kardkovács</i>	
Binary Lexical Relations for Text Representation in Information Retrieval	21
<i>Marco Gonzalez, Vera Lúcia Strube de Lima, and José Valdeni de Lima</i>	
Application of Text Categorization to Astronomy Field	32
<i>Huaizhong Kou, Amedeo Napoli, and Yannick Toussaint</i>	
Towards an XML Representation of Proper Names and Their Relationships	44
<i>Béatrice Bouchou, Mickael Tran, and Denis Maurel</i>	
Empirical Textual Mining to Protein Entities Recognition from PubMed Corpus	56
<i>Tyne Liang and Ping-Ke Shih</i>	
Automatic Extraction of Semantic Relationships for WordNet by Means of Pattern Learning from Wikipedia	67
<i>Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells</i>	
Combining Data-Driven Systems for Improving Named Entity Recognition	80
<i>Zornitsa Kozareva, Oscar Ferrández, Andres Montoyo, Rafael Muñoz, and Armando Suárez</i>	
Natural Language Processing: Mature Enough for Requirements Documents Analysis?	91
<i>Leonid Kof</i>	
Improving Text Categorization Using Domain Knowledge	103
<i>Jingbo Zhu and Wenliang Chen</i>	
A Process and Tool Support for Managing Activity and Resource Conflicts Based on Requirements Classification	114
<i>Hwasil Yang, Minseong Kim, Sooyong Park, and Vijayan Sugumaran</i>	

Web-Assisted Detection and Correction of Joint and Disjoint Malapropos Word Combinations 126
Igor A. Bolshakov and Sofia N. Galicia-Haro

Web Directory Construction Using Lexical Chains 138
Sofia Stamou, Vlassis Krikos, Pavlos Kokosis, Alexandros Ntoulas, and Dimitris Christodoulakis

Email Categorization with Tournament Methods 150
Yunqing Xia, Wei Liu, and Louise Guthrie

Knowledge-Based Information Extraction:
 A Case Study of Recognizing Emails of Nigerian Frauds 161
Yanbin Gao and Gang Zhao

Extended Tagging and Interpretation Tools
 for Mapping Requirements Texts to Conceptual (Predesign) Models 173
Günther Fliedl, Christian Kop, Heinrich C. Mayr, Martin Hölbling, Thomas Horn, Georg Weber, and Christian Winkler

Improving Question Answering Using Named Entity Recognition 181
Antonio Toral, Elisa Noguera, Fernando Llopis, and Rafael Muñoz

Using Semantic Roles in Information Retrieval Systems 192
Paloma Moreda, Borja Navarro, and Manuel Palomar

Text Categorization Based on Subtopic Clusters 203
Francis C.Y. Chik, Robert W.P. Luk, and Korris F.L. Chung

Interpretation of Implicit Parallel Structures.
 A Case Study with “vice-versa” 215
Helmut Horacek and Magdalena Wolska

Text2Onto – A Framework for Ontology Learning
 and Data-Driven Change Discovery 227
Philipp Cimiano and Johanna Völker

Interaction Transformation Patterns Based on Semantic Roles 239
Isabel Díaz, Lidia Moreno, Oscar Pastor, and Alfredo Matteo

Query Refinement Through Lexical Clustering
 of Scientific Textual Databases 251
Eric SanJuan

Automatic Filtering of Bilingual Corpora
 for Statistical Machine Translation 263
Shahram Khadivi and Hermann Ney

An Approach to Clustering Abstracts 275
Mikhail Alexandrov, Alexander Gelbukh, and Paolo Rosso

Named Entity Recognition for Web Content Filtering	286
<i>José María Gómez Hidalgo, Francisco Carrero García, and Enrique Puertas Sanz</i>	
The Role of Word Sense Disambiguation in Automated Text Categorization	298
<i>José María Gómez Hidalgo, Manuel de Buenaga Rodríguez, and José Carlos Cortizo Pérez</i>	
Combining Biological Databases and Text Mining to Support New Bioinformatics Applications	310
<i>René Witte and Christopher J.O. Baker</i>	
A Semi-automatic Approach to Extracting Common Sense Knowledge from Knowledge Sources	322
<i>Veda C. Storey, Vijayan Sugumaran, and Yi Ding</i>	
A Phrasal Approach to Natural Language Interfaces over Databases	333
<i>Michael Minock</i>	
Information Extraction for User's Utterance Processing on Ubiquitous Robot Companion	337
<i>Hanmin Jung, Choong-Nyong Seon, Jae Hong Kim, Joo Chan Sohn, Won-Kyung Sung, and Dong-In Park</i>	
Investigating the Best Configuration of HMM Spanish PoS Tagger when Minimum Amount of Training Data Is Available	341
<i>Sergio Ferrández and Jesús Peral</i>	
An Approach to Automatic Construction of Lexical Relations Between Chinese Nouns from Machine Readable Dictionary	345
<i>Yi Hu, Ruzhan Lu, and Xuening Li</i>	
Automatic Acquisition of Adjacent Information and Its Effectiveness in Extraction of Bilingual Word Pairs from Parallel Corpora	349
<i>Hiroshi Echizen-ya, Kenji Araki, and Yoshio Momouchi</i>	
Text Mining from Categorized Stem Cell Documents to Infer Developmental Stage-Specific Expression and Regulation Patterns of Stem Cells	353
<i>Hyun Seok Park, Min Kyung Kim, Eun Jeong Choi, and Young Joo Seol</i>	
Simple But Useful Algorithms for Identifying Noun Phrase Complements of Embedded Clauses in a Partial Parse	357
<i>Sebastian van Delden</i>	
An Add-On to Rule-Based Sifters for Multi-recipient Spam Emails	361
<i>Vipul Sharma, Puneet Sarda, and Swasti Sharma</i>	

Semantic Annotation of a Natural Language Corpus
for Knowledge Extraction 365
Borja Navarro, Patricio Martínez-Barco, and Manuel Palomar

mySENSEVAL: Explaining WSD System Performance
Using Target Word Features 369
Harri M.T. Saarikoski

Information Extraction from Email Announcements 372
Viktor Pekar

An Application of NLP Rules to Spoken Document Segmentation Task . . . 376
*Rafael M. Terol, Patricio Martínez-Barco, Fernando Llopis,
and Trinitario Martínez*

A Generalised Similarity Measure for Question Answering 380
Gerhard Fliedner

Multi-lingual Database Querying and the Atoms of Language 384
Epaminondas Kapetanios and Panagiotis Chountas

Extracting Information from Short Messages 388
Richard Cooper, Sajjad Ali, and Chenlan Bi

Automatic Transition of Natural Language Software Requirements
Specification into Formal Presentation 392
M.G. Ilieva and Olga Ormandjieva

Automatic Description of Static Images in Natural Language 398
*Azucena Montes Rendón, Pablo Sánchez Luna, Gerardo Reyes Salgado,
Juan G. González Serna, and Ricardo Fuentes Covarrubias*

On Some Optimization Heuristics for Lesk-Like WSD Algorithms 402
Alexander Gelbukh, Grigori Sidorov, and Sang-Yong Han

Author Index 407