

# **Advances in Natural Language Processing**

---

# Research in Computing Science

---

## Series Editorial Board

Comité Editorial de la Serie

### Editors-in-Chief:

Editores en Jefe

*Juan Humberto Sossa Azuela (Mexico)*  
*Gerhard Ritter (USA)*  
*Jean Serra (France)*  
*Ulises Cortés (Spain)*

### Associate Editors:

Editores Asociados

*Jesús Angulo (France)*  
*Jihad El-Sana (Israel)*  
*Jesús Figueroa (Mexico)*  
*Alexander Gelbukh (Russia)*  
*Ioannis Kakadiaris (USA)*  
*Serguei Levachkine (Russia)*  
*Petros Maragos (Greece)*  
*Julian Padget (UK)*  
*Mateo Valero (Spain)*

### Editorial Coordination:

Coordinación Editorial

*Blanca Miranda Valencia*

### Formatting:

Formación

*Sulema Torres Ramos*

**Research in Computing Science** es una publicación trimestral, de circulación internacional, editada por el Centro de Investigación en Computación del IPN, para dar a conocer los avances de investigación científica y desarrollo tecnológico de la comunidad científica internacional. **Volumen 18**, Febrero, 2006. Tiraje: 500 ejemplares. *Certificado de Reserva de Derechos al Uso Exclusivo del Título* No. 04-2004-062613250000-102, expedido por el Instituto Nacional de Derecho de Autor. *Certificado de Licitud de Título* No. 12897, *Certificado de licitud de Contenido* No. 10470, expedidos por la Comisión Calificadora de Publicaciones y Revistas Ilustradas. El contenido de los artículos es responsabilidad exclusiva de sus respectivos autores. Queda prohibida la reproducción total o parcial, por cualquier medio, sin el permiso expreso del editor, excepto para uso personal o de estudio haciendo cita explícita en la primera página de cada documento. Imagen de la portada: [www.absolutewallpapers.com/wallpapers/3dwallpapers/fractal/fractal\\_8.jpg](http://www.absolutewallpapers.com/wallpapers/3dwallpapers/fractal/fractal_8.jpg). Impreso en la Ciudad de México, en los Talleres Gráficos del IPN – Dirección de Publicaciones, Tres Guerras 27, Centro Histórico, México, D.F. Distribuida por el Centro de Investigación en Computación, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738, México, D.F. Tel. 57 29 60 00, ext. 56571.

**Editor Responsible:** *Juan Humberto Sossa Azuela*, RFC SOAJ560723

*Research in Computing Science* is published by the Center for Computing Research of IPN. **Volume 18**, February, 2006. Printing 500. The authors are responsible for the contents of their articles. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of Centre for Computing Research. Printed in Mexico City, February, 2006, in the IPN Graphic Workshop – Publication Office.

---

## Volume 18

Volumen 18

---

# Advances in Natural Language Processing

**Volume Editor:**

Editor del Volumen

*Alexander Gelbukh*

Instituto Politécnico Nacional  
Centro de Investigación en Computación  
México 2006



**ISSN: 1665-9899**

---

Copyright © Instituto Politécnico Nacional 2006  
*Copyright © by Instituto Politécnico Nacional*

Instituto Politécnico Nacional (IPN)  
Centro de Investigación en Computación (CIC)  
Av. Juan de Dios Bátiz s/n esq. M. Othón de Mendizábal  
Unidad Profesional “Adolfo López Mateos”, Zacatenco  
07738, México D.F., México

<http://www.ipn.mx>  
<http://www.cic.ipn.mx>

**Printing: 500**  
Impresiones: 500

**Printed in Mexico**  
Impreso en México

## Preface

Natural Language Processing is a branch of Artificial Intelligence aimed at enabling the computers to perform meaningful human-like activities related to written or spoken human language, such as English or Spanish. For traditional computer programs, a text is just a long string of characters—which the program can copy, print, or erase. In contrast, intelligent natural language processing programs are designed for operations involving the meaning of the text.

The most important applications of natural language processing include information retrieval and information organization, machine translation, and natural language interfaces, among others. However, as in any science, the activities of the researchers are mostly concentrated on its internal art and craft, that is, on the solution of the problems arising in analysis or synthesis of natural language text or speech, such as syntactic and semantic analysis, disambiguation, or compilation of dictionaries and grammars necessary for such analysis.

This volume presents 19 original research papers written by 63 authors representing 16 different countries: Argentina, Austria, Canada, China, Finland, Hong Kong, India, Japan, Korea, Lebanon, Mexico, Norway, Spain, UK, United Arab Emirates, and USA. The volume is structured in 7 thematic areas of both theory and applications of Natural Language Processing:

- Lexical Resources
- Morphology and syntax
- Word Sense Disambiguation and Semantics
- Speech Processing
- Information Retrieval
- Question Answering and Text Summarization
- Computer-Assisted Language Learning

The papers included in this volume were selected on the base of rigorous international reviewing process out of 43 submissions received for evaluation; thus the acceptance rate of this volume was 44%.

I would like to cordially thank all people involved in the preparation of this volume. In the first place I want to thank the authors of the published paper for their excellent research work giving sense to the work of all other people involved, as well the authors of rejected papers for their interest and effort. I also thank the members of the Editorial Board of the volume and additional reviewers for their hard work on reviewing and selecting the papers. I thank Sulema Torres, Ignacio Garcia Araoz, Oralia del Carmen Pérez Orozco, and Raquel López Alamilla for their valuable collaboration in preparation of this volume. The submission, reviewing, and selection process was supported for free by the EasyChair system, [www.EasyChair.org](http://www.EasyChair.org).

# Table of Contents

## Índice

Page/Pág.

### Lexical Resources

Generating Multilingual Grammars from OWL Ontologies.....	3
<i>Guillermo Pérez, Gabriel Amores, Pilar Manchón and David González</i>	
Clustering of English-Korean Translation Word Pairs Using Bi-grams .....	15
<i>Hanmin Jung, Hee-Kwan Koo, Won-Kyung Sung and Dong-In Park</i>	

### Morphology and syntax

Modified Makagonov's Method for Testing Word Similarity and its Application to Constructing Word Frequency List.....	27
<i>Xavier Blanco, Mikhail Alexandrov, Alexander Gelbukh</i>	
HMM based POS Tagger for a Relatively Free Word Order Language .....	37
<i>Arulmozhi Palanisamy and Sobha Lalitha Devi</i>	
Extended Tagging in Requirements Engineering.....	49
<i>Günther Fliedl, Christian Kop, Heinrich C. Mayr, Jürgen Vöhringer, Georg Weber and Christian Winkler</i>	
Analogical Modeling with Bias allowing Feedback and Centering .....	57
<i>Christer Johansson and Lars G. Johnsen</i>	

### Word Sense Disambiguation and Semantics

Word Sense Disambiguation with Basic-Level Categories.....	71
<i>Steve Legrand</i>	
A New Proposal of Word Sense Disambiguation for Nouns on a Question Answering System.....	83
<i>S. Ferrández, S. Roger, A. Ferrández, A. Aguilar and P. López-Moreno</i>	
On Types, Intension and Compositionality.....	93
<i>Walid S. Saba</i>	

### Speech Processing

Acoustic Model Adaptation for Codec Speech based on Learning-by-Doing Concept .....	105
<i>Shingo Kuroiwa, Satoru Tsuge, Koji Tanaka, Kazuma Hara and Fuji Ren</i>	
Specific Speaker's Japanese Speech Corpus over Longand Short Time Periods.....	115
<i>Satoru Tsuge, Masami Shishibori, Fuji Ren, Kenji Kita and Shingo Kuroiwa</i>	
A Task Analysis of Nursing Activites Using Spoken Corpora .....	125
<i>Hiromi itoh Ozaku, Akinori Abe, Kaoru Sagara, Noriaki Kuwahara and Kiyoshi Kogure</i>	

## **Information Retrieval**

Improved Focused Web Crawling Based on Incremental Q-Learning .....	139
<i>Yunming Ye, Yan Li, Joshua Huang and Xiaofei Xu</i>	
Automatic Identification of Chinese Stop Words .....	151
<i>Feng Zou, Fu Lee Wang, Xiaotie Deng and Song Han</i>	
Fast Search Algorithm for Sequences of Hierarchically Structured Data .....	163
<i>Masaki Murata, Masao Utiyama, Toshiyuki Kanamaru and Hitoshi Isahara</i>	

## **Question Answering and Text Summarization**

Cross-Lingual Question Answering Using InterLingual Index Module of EuroWordNet .....	177
<i>Sergio Ferrández and Antonio Ferrández</i>	
Using Terminology and a Concept Hierarchy for Restricted-Domain Question-Answering .....	183
<i>Hai Doan-Nguyen and Leila Kosseim</i>	
Efficient Randomized Algorithms for Text Summarization .....	195
<i>Ahmed Mohamed and Sanguthevar Rajasekaran</i>	

## **Computer-Assisted Language Learning**

Arabic Error Feedback in an Online Arabic Learning System.....	203
<i>Khaled F. Shaalan and Habib E. Talhami</i>	

<b>Author Index</b> .....	213
---------------------------	-----

<b>Editorial Board of the Volume</b> .....	215
--	-----

<b>Additional Reviewers</b> .....	215
-----------------------------------	-----

## Author Index

### Índice de autores

Abe, Akinori	125	Li, Yan	139
Aguilar, A.	83	López-Moreno, P.	83
Alexandrov, Mikhail	27	Manchón , Pilar	3
Amores, Gabriel	3	Mayr, Heinrich C.	49
Blanco, Xavier	27	Mohamed, Ahmed	195
Deng, Xiaotie	151	Murata, Masaki	163
Ferrández, Antonio	83, 177	Ozaku, Hiromi itoh	125
Ferrández, Sergio	83, 177	Palanisamy , Arulmozhi	37
Fliedl, Günther	49	Park, Dong-In	15
Gelbukh, Alexander	27	Pérez, Gillermo	3
González, David	3	Rajasekaran, Sanguthevar	195
Hai, Doan-Nguyen	183	Ren, Fuji	105, 115
Han, Song	151	Roger, S.	83
Hara, Kazuma	105	Saba, Walid S.	93
Huang, Joshua	139	Sagara, Kaoru	125
Isahara, Hitoshi	163	Shaalán, Khaled F.	203
Johansson, Christer	57	Shishibori, Masami	115
Johnsen, Lars G.	57	Sung , Won-Kyung	15
Jung, Hanmin	15	Talhami, Habib E.	203
Kanamaru, Toshiyuki	163	Tanaka, Koji	105
Kita, Kenji	115	Tsuge, Satoru	105, 115
Kogure, Kiyoshi	125	Utiyama, Masao	163
Koo, Hee-Kwan	15	Vöhringer, Jürgen	49
Kop, Christian	49	Wang, Fu Lee	151
Kosseim, Leila	183	Weber , Georg	49
Kuroiwa, Shingo	105, 115	Winkler, Christian	49
Kuwahara, Noriaki	125	Xu, Xiaofei	139
Lalitha Devi, Sobha	37	Ye, Yunming	139
Legrand, Steve	71	Zou, Feng	151



## **Editorial Board of the Volume**

### **Comité editorial de volumen**

Eneko Agirre	Rada Mihalcea
Christian Boitet	Ruslan Mitkov
Igor Bolshakov	Masaki Murata
Nicoletta Calzolari	Vivi Nastase
John Carroll	Olga Nevzorova
Dan Cristea	Nicolas Nicolov
Barbara Di Eugenio	Sergei Nirenburg
Gregory Grefenstette	Constantin Orasan
Linda van Gulder	Manuel Palomar
Cătălina Hallett	Ted Pedersen
Yasunari Harada	Viktor Pekar
Eduard Hovy	Stelios Piperidis
Nancy Ide	James Pustejovsky
Diana Inkpen	Fuji Ren
Frederick Jelinek	Fabio Rinaldi
Aravind Krishna Joshi	Horacio Rodriguez
Martin Kay	Vasile Rus
Alma Kharrat	Ivan Sag
Adam Kilgarriff	Franco Salvetti
Richard Kittredge	Serge Sharoff
Kevin Knight	Grigori Sidorov
Alexander Koller	Thamar Solorio
Grzegorz Kondrak	Carlo Strapparava
Sandra Kuebler	Maosong Sun
Ken Litkowski	John Tait
Hugo Liu	Benjamin Ka-yin T'sou
Aurelio López López	Felisa Verdejo
Bernardo Magnini	Karin Verspoor
Daniel Marcu	Manuel Vilares Ferro
Carlos Martín-Vide	Yorick Wilks
Igor Mel'čuk	

## **Additional Reviewers**

### **Árbitros adicionales**

Mohamed Abdel Fattah	Bogdan Babych
Mustafa Abusalah	Verginica Barbu Mititelu
Farooq Ahmad	Fco. Mario Barcala Rodríguez
Iñaki Alegria	Francesca Bertagna
Muath Alzghool	Dimitar Blagoev

Hiram Calvo Castro  
Anna Clark  
Daoud Clarke  
Andras Csomai  
Victor Manuel Darriba Bilbao  
Jeremy Ellman  
Davide Fossati  
Oana Frunza  
Irbis Gallegos  
Jorge Graña  
Samer Hassan  
David Hope  
Scott K. Imig  
Aminul Islam  
Shih-Wen Ke  
Stephan Kepser  
Rob Koeling  
Alberto Lavelli  
Fennie Liang  
Christian Loza  
Xin Lu

Fernando Magán Muñoz  
Raquel Martínez  
Jaime Mendez  
Dragos Stefan Munteanu  
Crystal Nakatsu  
Apostol Natsev  
Matteo Negri  
Michael Oakes  
Octavian Popescu  
Oana Postolache  
Christoph Reichenbach  
Francisco Ribadas Peña  
German Rigau  
Tarek Sherif  
Radu Soricut  
Chris Stokoe  
Rajen Subba  
Hristo Tanev  
Martin Thomas  
Jesus Vilares Ferro  
Zhuli Xie

Impreso en los Talleres Gráficos  
de la Dirección de Publicaciones  
del Instituto Politécnico Nacional  
Tresguerras 27, Centro Histórico, México, D.F.  
Febrero de 2006.  
Printing 500 / Edición 500 ejemplares.