

## Preface

CICLing 2009 was the anniversary 10<sup>th</sup> Annual Conference on Intelligent Text Processing and Computational Linguistics. The CICLing conferences provide a wide-scope forum for discussion of the art and craft of natural language processing research as well as the best practices in its applications.

This volume contains five invited papers and the regular papers accepted for oral presentation at the conference. The papers accepted for poster presentation were published in a special issue of another journal; see information on the website. Since 2001, the proceedings of CICLing conferences have been published in Springer's Lecture Notes in Computer Science series, as volumes 2004, 2276, 2588, 2945, 3406, 3878, 4394, and 4919.

The volume is structured into 12 sections:

- Trends and Opportunities
- Linguistic Knowledge Representation Formalisms
- Corpus Analysis and Lexical Resources
- Extraction of Lexical Knowledge
- Morphology and Parsing
- Semantics
- Word Sense Disambiguation
- Machine Translation and Multilinguism
- Information Extraction and Text Mining
- Information Retrieval and Text Comparison
- Text Summarization
- Applications to the Humanities

A total of 167 papers by 392 authors from 40 countries were submitted for evaluation by the International Program Committee, see Tables 1 and 2. This volume contains revised versions of 44 papers, by 120 authors, selected for oral presentation; the acceptance rate was 26.3%. It also features invited papers by

- Jill Burstein, Educational Testing Service, USA
- Ken Church, Microsoft, USA
- Dekang Lin, Google, USA
- Bernardo Magnini, Fondazione Bruno Kessler, Italy

who presented excellent keynote lectures at the conference. Publication of extended full-text invited papers in the proceedings is a distinctive feature of the CICLing conferences. What is more, in addition to presentation of their invited papers, the keynote speakers organized separate vivid informal events; this is also a distinctive feature of this conference series.

The 2009 event was accompanied by a five-day pre-conference Lexicom Americas Workshop 2009, organized by Lexicography MasterClass and led by Adam Kilgarriff, Lexicography MasterClass, UK, and Jan Pomikálek, Masaryk University, Czech Republic. The main conference program also included a discussion panel on the future of corpus design, organized by Adam Kilgarriff.

**Table 1.** Statistics of submissions and accepted papers by country or region.

Country or region	Authors		Papers <sup>1</sup>		Country or region	Authors		Papers <sup>1</sup>	
	Subm	Subm	Subm	Accp		Subm	Subm	Accp	
Algeria	4	1.33	–		Italy	14	5.93	2.43	
Argentina	3	1.33	1		Japan	12	4.5	1	
Brazil	14	4.8	–		Jordan	1	1	–	
Canada	6	2.58	1.25		Korea	12	5	–	
China	17	7.25	–		Lithuania	2	2	–	
Colombia	3	0.75	0.75		Macao	4	1	–	
Croatia	5	1	1		Mexico	44	15.38	3.88	
Czech Rep.	9	5	3		Myanmar	2	1	–	
Egypt	2	1	1		Norway	2	1	1	
Estonia	1	1	1		Portugal	4	2	–	
Finland	1	1	1		Romania	8	4	–	
France	11	4.37	–		Russia	2	2	–	
Germany	9	6	2		Spain	60	19.95	6.87	
Greece	3	1.25	1		Sweden	3	3	1	
Hong Kong	7	2.5	–		Switzerland	13	5	1	
Hungary	2	1	–		Tunisia	9	4	–	
India	32	13	2		Turkey	10	4	1	
Iran	5	2	–		UK	9	3.32	1.32	
Ireland	3	1	1		USA	40	21.75	7.5	
Israel	2	1	1		Vietnam	2	2	–	
					<i>Total:</i>	<i>392</i>	<i>167</i>	<i>44</i>	

<sup>1</sup> Counted by authors. E.g., for a paper by 3 authors, 2 from Mexico and 1 from USA, we added  $\frac{2}{3}$  to Mexico and  $\frac{1}{3}$  to USA.

The following papers received the Best Paper Awards and the Best Student Paper Award, correspondingly:

- 1<sup>st</sup> Place: *Cross-Language Frame Semantics Transfer in Bilingual Corpora*, by Roberto Basili, Diego De Cao, Danilo Croce, Bonaventura Coppola, and Alessandro Moschitti (page 332);
- 2<sup>nd</sup> Place: *Detecting Protein-Protein Interactions in Biomedical Texts Using a Parser and Linguistic Resources*, by Gerold Schneider, Kaarel Kaljurand, and Fabio Rinaldi (page 406);
- 3<sup>rd</sup> Place: *Learning to Learn Biological Relations from a Small Training Set*, by Laura Alonso i Alemany and Santiago Bruno (page 418);
- Student: *Enriching Statistical Translation Models Using a Domain-Independent Multilingual Lexical Knowledge Base*, by Miguel García, Jesús Giménez, and Lluís Màrquez (page 306).

The best student paper was selected from papers the first author of which was a full-time student, excluding the papers that received a Best Paper Award. The authors of the awarded papers were given extended time for their presentations.

In addition, the Best Presentation Award and the Best Poster Award winners were selected by a ballot among the attendees of the conference.

**Table 2.** Statistics of submissions and accepted papers by topic.<sup>2</sup>

Accepted	Submitted	Topic
13	39	Information extraction
12	31	Text mining
11	31	Clustering and categorization
8	19	Syntax and chunking (linguistics)
8	20	Statistical methods (mathematics)
8	29	Lexical resources
7	22	Information retrieval
7	24	Semantics and discourse
7	29	Formalisms and knowledge representation
7	31	Other
6	18	Word Sense Disambiguation
5	29	Symbolic and linguistic methods
3	11	Natural language interfaces
2	3	Emotions and humor
2	7	Text generation
2	8	Machine translation
2	9	Morphology
2	10	Summarization
1	1	Textual entailment
1	2	Parsing algorithms (mathematics)
–	2	Spell checking
–	2	POS tagging
–	3	Speech processing
–	3	Anaphora resolution

<sup>2</sup> According to the topics indicated by the authors. A paper may be assigned to more than one topic.

Besides its high scientific level, one of the success factors of CICLing conferences is their excellent cultural program. CICLing 2009 was held in Mexico, a wonderful country rich in culture, history, and nature. The participants of the conference had a chance to see the legendary 2000-years-old Teotihuacan pyramids, a monarch butterfly wintering site where the old pines are covered with millions of butterflies as if they were leaves, a great cave with 85-meter halls and a river flowing out of it, Aztec warriors dancing in the street in their colorful plumages, and the largest anthropological museum in the world; see photos at [www.CICLing.org](http://www.CICLing.org).

I would like to thank all those involved in the organization of this conference. In the first place these are the authors of the papers constituting this book: it is the excellence of their research work that gives value to the book and sense to the work of all other people involved. I thank the Program Committee members for their hard and very professional work. Very special thanks go to Manuel Vilares and his group, Rada Mihalcea, and Ted Pedersen for their invaluable support in the reviewing process.

## VIII

I express my most cordial thanks to the members of the local Organizing Committee for their considerable contribution to making this conference become a reality. I thank Mexican Government for providing financial support, Mexican Society of Artificial Intelligence for valuable collaboration, and the Center for Computing Research (CIC) of the National Polytechnic Institute (IPN), Mexico, for hosting the conference.

The entire submission, reviewing, and selection process, as well as putting together the proceedings, was supported for free by the EasyChair system ([www.EasyChair.org](http://www.EasyChair.org)); I express my gratitude to its author Andrei Voronkov for his constant support and help. Last but not least, I deeply appreciate the Springer staff's patience and help in editing this volume—it is always a great pleasure to work with them.

January 2009

Alexander Gelbukh

## Organization

CICLing 2009 was organized by the Natural Language and Text Processing Laboratory ([nlp.cic.ipn.mx](http://nlp.cic.ipn.mx)) of the Center for Computing Research (CIC) of the National Polytechnic Institute (IPN), Mexico, in collaboration with the Mexican Society of Artificial Intelligence (SMIA, [www.smia.org.mx](http://www.smia.org.mx)).

### Program Chair

Alexander Gelbukh

### Program Committee

John Atkinson  
 Christian Boitet  
 Nicoletta Calzolari  
 John Carroll  
 Kenneth Church  
 Dan Cristea  
 Walter Daelemans  
 Mona Talat Diab  
 Oren Etzioni  
 Alexander Gelbukh  
 Gregory Grefenstette  
 Yasunari Harada  
 Eva Hajičová  
 Graeme Hirst  
 Eduard Hovy  
 Nancy Ide  
 Diana Inkpen  
 Frederick Jelinek  
 Alma Kharrat  
 Adam Kilgarriff  
 Alexander Koller  
 Henry Lieberman

Dekang Lin  
 Aurelio López  
 Igor Mel'čuk  
 Rada Mihalcea  
 Masaki Murata  
 Nicolas Nicolov  
 Kemal Oflazer  
 Constantin Orasan  
 Ted Pedersen  
 Viktor Pekar  
 Stelios Piperidis  
 Fuji Ren  
 Fabio Rinaldi  
 Horacio Rodríguez  
 Vasile Rus  
 Franco Salvetti  
 Serge Sharoff  
 Grigori Sidorov  
 Thamar Solorio  
 Maosong Sun  
 Karin Verspoor  
 Manuel Vilares Ferro

### Award Committee

Alexander Gelbukh  
 Eduard Hovy  
 Rada Mihalcea

Ted Pedersen  
 Yorick Wiks

## Additional Referees

Mohamed Abdel Fattah	Joji Maeno
Mikhail Alexandrov	Nobuaki Minematsu
Muath Alzghool	Tomoko Ohkuma
Chris Biemann	Juan Otero Pombo
Maya Carrillo Ruiz	Ryo Otoguro
Victor Darriba	Feng Pan
Georgiana Dinu	Michael Piotrowski
Iustin Dornescu	Natalia Ponomareva
Milagros Fernández Gavilanes	Prokopis Prokopidis
Oana Frunza	Jonathon Read
René Arnulfo García Hernández	Francisco Jose Ribadas Pena
Byron Georgantopoulos	Gerold Schneider
Chikara Hashimoto	Petr Sgall
Laura Hasler	Kiyoaki Shirai
Laritza Hernández Rojas	Takeo Tatsumi
Kaarel Kaljurand	Lorenzo Thione
Fazel Keshtkar	Martin Volk
Rob Koeling	Christopher Walker
Marco Kuhlmann	Zdenek Zabokrtsky
Yulia Ledeneva	Carlos Mario Zapata Jaramillo

## Organizing Committee

Ignacio García Araoz, CIC	Raquel López Alamilla, CIC
Rosalía García, SMIA	Oralia del Carmen Pérez Orozco, CIC
Alexander Gelbukh, CIC & SMIA	Carlos Alberto Reyes García, SMIA
Olga Kolesnikova, CIC	Grigori Sidorov, CIC & SMIA
Yulia Ledeneva, UAEM	Sulema Torres Ramos, CIC & SMIA

## Website and Contact

The website of the CICLing conferences is [www.CICLing.org](http://www.CICLing.org). It contains information on the past CICALing conferences and satellite events, abstracts of all published papers, photos from past CICALing conferences, video recordings of most keynote talks, as well as the information on the forthcoming CICALing conferences. Contact options can be found on the website.