Preface

CICLing 2009 was the anniversary 10th 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.

<table>
<thead>
<tr>
<th>Country or region</th>
<th>Authors Subm</th>
<th>Papers(^1) Subm</th>
<th>Authors Subm</th>
<th>Papers(^1) Subm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>4</td>
<td>1.33</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Argentina</td>
<td>3</td>
<td>1.33 1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Brazil</td>
<td>14</td>
<td>4.8</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Canada</td>
<td>6</td>
<td>2.58 1.25</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>China</td>
<td>17</td>
<td>7.25</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Colombia</td>
<td>3</td>
<td>0.75 0.75</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Croatia</td>
<td>5</td>
<td>1 1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Czech Rep.</td>
<td>9</td>
<td>5 3</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Egypt</td>
<td>2</td>
<td>1 1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Estonia</td>
<td>1</td>
<td>1 1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Finland</td>
<td>1</td>
<td>1 1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>France</td>
<td>11</td>
<td>4.37</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Germany</td>
<td>9</td>
<td>6 2</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Greece</td>
<td>3</td>
<td>1.25 1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>7</td>
<td>2.5</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Hungary</td>
<td>2</td>
<td>1 1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>India</td>
<td>32</td>
<td>13 2</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Iran</td>
<td>5</td>
<td>2 2</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Ireland</td>
<td>3</td>
<td>1 1</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Israel</td>
<td>2</td>
<td>1 1</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Total: 392 167 44

\(^1\) 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\(^{st}\) 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\(^{nd}\) 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\(^{rd}\) 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 Márquez (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.\(^2\)

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

\(^2\) 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.

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.
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); 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) 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).

Program Chair

Alexander Gelbukh

Program Committee

John Atkinson
Christian Boitet
Nicoletta Calzolari
John Carroll
Kenneth Church
Dan Cristea
Walter Daelemans
Mona Tahat Diab
Oren Etzioni
Alexander Gelbukh
Gregory Grefenstette
Yasunari Harada
Eva Hajicová
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 Wilks
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               
Yulia Ledeneva               

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. It contains information on the past CICLing conferences and satellite events, abstracts of all published papers, photos from past CICLing conferences, video recordings of most keynote talks, as well as the information on the forthcoming CICLing conferences. Contact options can be found on the website.