# Preface

CICLing 2015 was the 16th 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 set of two books contains four invited papers and a selection of regular papers accepted for presentation at the conference. Since 2001, the proceedings of the CICLing conferences have been published in Springer's *Lecture Notes in Computer Science* series as volumes 2004, 2276, 2588, 2945, 3406, 3878, 4394, 4919, 5449, 6008, 6608, 6609, 7181, 7182, 7816, 7817, 8403, and 8404.

The set has been structured into 13 sections representative of the current trends in research and applications of Natural Language Processing:

- Lexical Resources
- Morphology and Chunking
- Syntax and Parsing
- Anaphora Resolution and Word Sense Disambiguation
- Semantics and Dialogue
- Machine Translation and Multilingualism
- Sentiment Analysis and Emotions
- Opinion Mining and Social Network Analysis
- Natural Language Generation and Summarization
- Information Retrieval, Question Answering, and Information Extraction
- Text Classification
- Speech Processing
- Applications

The 2015 event received submissions from 62 countries, a record high number in the 16-year history of the CICLing series. A total of 329 papers (second highest number in the history of CICLing) by 705 authors were submitted for evaluation by the International Program Committee; see Figure 1 and Tables 1 and 2. This two-volume set contains revised versions of 95 regular papers selected for presentation; thus, the acceptance rate for this set was 28.9%.

In addition to regular papers, the books feature invited papers by:

- Erik Cambria, Nanyang Technical University, Singapore
- Mona Diab, George Washington University, USA
- Lauri Karttunen, Stanford University, USA
- Joakim Nivre, Uppsala University, Sweden

who presented excellent keynote lectures at the conference. Publication of full-text invited papers in the proceedings is a distinctive feature of the CICLing conferences.

**Table 1.** Number of submissions and accepted papers by topic[1]

| Accepted | Submitted | % accepted | Topic |
|---|---|---|---|
| 19 | 51 | 37 | Emotions, sentiment analysis, opinion mining |
| 19 | 56 | 34 | Text mining |
| 17 | 65 | 26 | Arabic |
| 17 | 58 | 29 | Information extraction |
| 17 | 49 | 35 | Lexical resources |
| 15 | 53 | 28 | Information retrieval |
| 14 | 35 | 40 | Under-resourced languages |
| 12 | 45 | 27 | Semantics, pragmatics, discourse |
| 11 | 40 | 28 | Clustering and categorization |
| 11 | 33 | 33 | Machine translation and multilingualism |
| 10 | 29 | 34 | Practical applications |
| 8 | 37 | 22 | Social networks and microblogging |
| 8 | 21 | 38 | Syntax and chunking |
| 7 | 17 | 41 | Formalisms and knowledge representation |
| 7 | 23 | 30 | Noisy text processing and cleaning |
| 5 | 21 | 24 | Morphology |
| 4 | 12 | 33 | Question answering |
| 4 | 10 | 40 | Textual entailment |
| 3 | 9 | 33 | Natural language generation |
| 3 | 8 | 38 | Plagiarism detection and authorship attribution |
| 3 | 13 | 23 | Speech processing |
| 3 | 21 | 14 | Summarization |
| 3 | 12 | 25 | Word sense disambiguation |
| 2 | 10 | 20 | Computational terminology |
| 2 | 8 | 25 | Co-reference resolution |
| 2 | 16 | 12 | Named entity recognition |
| 2 | 9 | 22 | Natural language interfaces |
| 1 | 1 | 100 | Computational humor |
| 1 | 15 | 7 | Other |
| 1 | 11 | 9 | POS tagging |
| 0 | 7 | 0 | Spelling and grammar checking |

[1] As indicated by the authors. A paper may belong to more than one topic.

Furthermore, 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.

With this event, we continued with our policy of giving preference to papers with verifiable and reproducible results: in addition to the verbal description of their findings given in the paper, we encouraged the authors to provide a proof of their claims in electronic form. If the paper claimed experimental results, we asked the authors to make available to the community all the input data necessary to verify and reproduce these results; if it claimed to introduce an algorithm, we encourage the authors to make the algorithm itself, in a programming language, available to the public. This additional

**Table 2.** Number of submitted and accepted papers by country or region

| Country or region | Authors Subm. | Papers[2] Subm. | Accp. | Country or region | Authors Subm. | Papers[2] Subm. | Accp. |
|---|---|---|---|---|---|---|---|
| Algeria | 6 | 2.5 | 0.5 | Lebanon | 2 | 1 | – |
| Argentina | 2 | 0.5 | 0.5 | Malaysia | 3 | 3 | – |
| Australia | 4 | 1.33 | 1.33 | Mexico | 13 | 5.95 | 1.08 |
| Belgium | 7 | 2 | 1 | Morocco | 16 | 6 | – |
| Brazil | 19 | 8.5 | 2.5 | Myanmar | 3 | 1.5 | – |
| Canada | 20 | 9.5 | 6 | The Netherlands | 2 | 1 | – |
| Chile | 5 | 1 | – | New Zealand | 4 | 2 | 1 |
| China | 37 | 15.05 | 4.3 | Nigeria | 5 | 2 | 1 |
| Colombia | 6 | 2 | 1 | Oman | 1 | 1 | – |
| Czech Rep. | 11 | 3.33 | 2 | Peru | 6 | 3.5 | 0.5 |
| Egypt | 89 | 42.8 | 13.33 | Philippines | 2 | 1 | – |
| Estonia | 1 | 1 | – | Poland | 3 | 2 | – |
| Ethiopia | 2 | 1 | – | Portugal | 8 | 3 | 2 |
| Finland | 6 | 1.75 | 1.75 | Qatar | 1 | 1 | – |
| France | 30 | 13.05 | 2.58 | Romania | 7 | 5 | 1 |
| Georgia | 1 | 1 | – | Russia | 11 | 7.33 | 1.67 |
| Germany | 15 | 6.83 | 1.33 | Saudi Arabia | 10 | 4.42 | 0.5 |
| Greece | 3 | 1 | – | Serbia | 3 | 1 | – |
| Hong Kong | 10 | 3.7 | 2.7 | Singapore | 1 | 1 | 1 |
| Hungary | 2 | 2 | 2 | South Africa | 1 | 0.83 | 0.5 |
| India | 98 | 57 | 15 | Spain | 16 | 6.92 | 3.92 |
| Iran | 1 | 1 | – | Sri Lanka | 7 | 3 | 0.67 |
| Iraq | 2 | 1.33 | 0.5 | Sweden | 2 | 0.67 | – |
| Ireland | 4 | 1.33 | 1.33 | Switzerland | 4 | 2.25 | 1.25 |
| Israel | 6 | 2 | – | Tunisia | 57 | 23.58 | 4 |
| Italy | 9 | 3.33 | 0.33 | Turkey | 29 | 14.42 | 2.83 |
| Japan | 17 | 7.25 | 1.25 | Ukraine | 2 | 0.33 | – |
| Jordan | 1 | 0.5 | – | UAE | 4 | 2.83 | 0.5 |
| Kazakhstan | 4 | 1 | 1 | UK | 23 | 10.75 | 3.58 |
| Korea, South | 1 | 0.5 | – | USA | 36 | 13.78 | 6.58 |
| Kuwait | 1 | 0.17 | 0.17 | Vietnam | 3 | 1.67 | – |
| | | | | *Total:* | 705 | 329 | 95 |

[2] By the number of authors: e.g., a paper by two authors from the USA and one from UK is counted as 0.67 for the USA and 0.33 for UK.

electronic material will be permanently stored on the CICLing's server, www.CICLing.org, and will be available to the readers of the corresponding paper for download under a license that permits its free use for research purposes.

In the long run, we expect that computational linguistics will have verifiability and clarity standards similar to those of mathematics: in mathematics, each claim is accompanied by a complete and verifiable proof, usually much longer than the claim itself; each theorem's complete and precise proof—and not just a description of its general idea—is made available to the reader. Electronic media allow computational linguists to provide material analogous to the proofs and formulas in mathematic in full length—
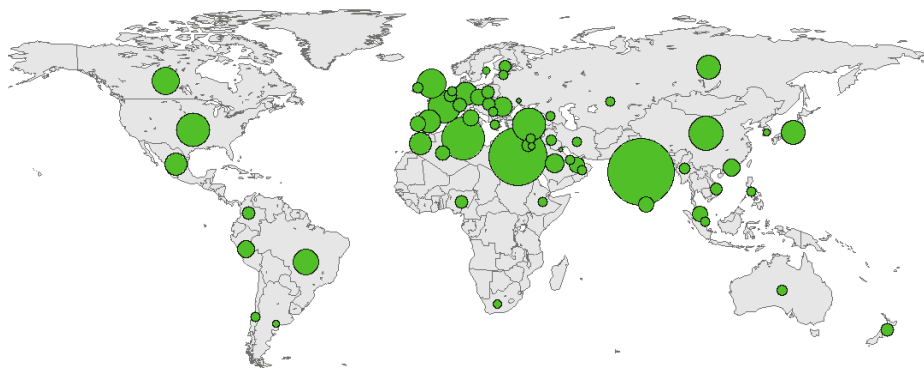
**Fig. 1.** Submissions by country or region. The area of a circle represents the number of submitted papers.

which can amount to megabytes or gigabytes of data—separately from a 12-page description published in the book. More information can be found on http://www.CICLing. org/why_verify.htm.

To encourage providing algorithms and data along with the published papers, we selected three winners of our Verifiability, Reproducibility, and Working Description Award. The main factors in choosing the awarded submission were technical correctness and completeness, readability of the code and documentation, simplicity of installation and use, and exact correspondence to the claims of the paper. Unnecessary sophistication of the user interface was discouraged; novelty and usefulness of the results were not evaluated—instead, they were evaluated for the paper itself and not for the data.

In this year, we introduced a policy of allowing—and encouraging—papers longer than the 12-page limit included in the fee. The reason was an observation that longer papers tend to be more complete and useful for the reader. In contrast, when a restrictive page limit is enforced, very often the authors have to omit important details; to the great frustration of the readers, this usually renders the whole paper largely useless because the presented results cannot be reproduced. This our observation was strongly confirmed by the fact that all four papers selected for the Best Paper Awards, especially the winners of the first two places, were much over the usual 12-page limit imposed by other conferences.

The following papers received the Best Paper Awards, the Best Student Paper Award, as well as the Verifiability, Reproducibility, and Working Description Awards, correspondingly:

Best Paper *Automated Linguistic Personalization of Targeted Marketing Messages*
1st Place: *Mining User-generated Text on Social Media*, by Rishiraj Saha Roy, Aishwarya Padmakumar, Guna Prasaad Jeganathan, and Ponnurangam Kumaraguru, India;

| | |
|---|---|
| Best Paper 2nd Place: | *Term Network Approach for Transductive Classification*, by Rafael Geraldeli Rossi, Solange Oliveira Rezende, and Alneu de Andrade Lopes, Brazil; |
| Best Paper 3rd Place: | *Building Large Arabic Multi-domain Resources for Sentiment Analysis*, by Hady ElSahar and Samhaa R. El-Beltagy, Egypt; |
| Best Student Paper:[1] | *Translation Induction on Indian Language Corpora using Translingual Themes from Other Languages*, by Goutham Tholpadi, Chiranjib Bhattacharyya, and Shirish Shevade, India; |
| Verifiability 1st Place: | *Domain-specific Semantic Relatedness from Wikipedia Structure: A Case Study in Biomedical Text*, by Armin Sajadi, Evangelos E. Milios, and Vlado Keselj, Canada; |
| Verifiability 2nd Place: | *Translation Induction on Indian Language Corpora using Translingual Themes from Other Languages*, by Goutham Tholpadi, Chiranjib Bhattacharyya, and Shirish Shevade, India; |
| Verifiability 3rd Place: | *Opinion Summarization using Submodular Functions: Subjectivity vs Relevance Trade-off*, by Jayanth Jayanth, Jayaprakash S., and Pushpak Bhattacharyya, India.[2] |

The authors of the awarded papers (except for the Verifiability award) 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.

Besides its high scientific level, one of the success factors of CICLing conferences is their excellent cultural program in which all attendees participate. The cultural program is a very important part of the conference, serving its main purpose: personal interaction and making friends and contacts. The attendees of the conference had a chance to visit the Giza Plateau with the Great Pyramid of Cheops and the Sphinx—probably the most important touristic place on Earth; The Egyptian Museum, the home to the largest collection of Pharaonic or ancient Egyptian relics and pieces; and the Old Cairo, to mention only a few most important attractions.

In this year we founded, and held in conjunction with CICLing, the First Arabic Computational Linguistics conference, which we expect to become the primary yearly event for dissemination of research results on Arabic language processing. This is in accordance with CICLing's mission to promote consolidation of emerging NLP communities in countries and regions underrepresented in the mainstream of NLP research and, in particular, in the mainstream publication venues. With founding this new conference, and with the very fact of holding CICLing in Egypt in a difficult moment of its history, we expect to contribute to mutual understanding, tolerance, and confidence between the Arabic world and the Western world: the better we know each other the more lasting peace between peoples.

---

[1] The best student paper was selected from among papers of which the first author was a full-time student, excluding the papers that received Best Paper Awards.

[2] This paper is published in a special issue of a journal and not in this book set.

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 that constitute 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. I thank all those who served on the Program Committee of CICLing 2015 and of the First Arabic Computational Linguistics conference, the Software Reviewing Committee, Award Selection Committee, as well as additional reviewers, for their hard and very professional work. Special thanks go to Ted Pedersen, Savas Yıldırım, Roberto Navigli, Manuel Vilares Ferro, Kenneth Church, Dafydd Gibbon, Kais Haddar, and Adam Kilgarriff for their invaluable support in the reviewing process.

I would like to thank Prof. Tarek Khalil, president of Nile University (NU), for welcoming CICLing at NU. I also want to cordially thank the conference staff, volunteers, and members of the Local Organization Committee headed by Prof. Samhaa R. El-Beltagy and advised by Prof. Hussein Anis. In particular, I am very grateful to Ms. Aleya Serag El-Din for her great effort in coordinating all the aspects of the conference. I wish to thank the Center for Informatics Science for all the support they have provided. I am deeply grateful to the administration of the Nile University for their helpful support, warm hospitality, and in general for providing this wonderful opportunity of holding CICLing in Egypt. I am also grateful to members of the Human Foundation for their great effort in planning the cultural program. I acknowledge support from Microsoft Research, the project CONACYT Mexico–DST India 122030 "Answer Validation through Textual Entailment," and SIP-IPN grant 20150028.

The entire submission and reviewing process was supported for free by the Easy-Chair system (www.EasyChair.org). Last but not least, I deeply appreciate the Springer staff's patience and help in editing these volumes and getting them printed in very short time—it is always a great pleasure to work with Springer.

March 2015                                                                 Alexander Gelbukh