

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, Lancaster, 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, Zurich, Switzerland

John C. Mitchell

Stanford University, Stanford, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

C. Pandu Rangan

Indian Institute of Technology Madras, Chennai, India

Bernhard Steffen

TU Dortmund University, Dortmund, Germany

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbrücken, Germany

More information about this series at <http://www.springer.com/series/7407>

Alexander Gelbukh (Ed.)

Computational Linguistics and Intelligent Text Processing

17th International Conference, CICLing 2016
Konya, Turkey, April 3–9, 2016
Revised Selected Papers, Part II

Editor
Alexander Gelbukh
CIC, Instituto Politécnico Nacional
Mexico City
Mexico

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-75486-4 ISBN 978-3-319-75487-1 (eBook)
<https://doi.org/10.1007/978-3-319-75487-1>

Library of Congress Control Number: 2018934347

LNCS Sublibrary: SL1 – Theoretical Computer Science and General Issues

© Springer International Publishing AG, part of Springer Nature 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG
part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

In Memoriam of Adam Kilgarriff

Preface

CICLing 2016 was the 17th International 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.

In 2015, Adam Kilgarriff, an influential scientist and a wonderful and brave person, passed away prematurely, at the age of only 55. Adam was a great friend of CICLing, a member of its small informal Steering Committee, one who helped to shape CICLing from its inception. Until his last days, already terminally ill, he volunteered to help us with the reviewing process for CICLing 2015. This CICLing event was dedicated to his bright memory, and its proceedings begin with a paper that attempts to summarize his scientific legacy.

This set of two books contains five 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, 8404, 9041, and 9042.

The set has been structured into 14 sections: an In Memoriam section and 13 sections representative of the current trends in research and applications of natural language processing:

- General formalisms
- Embeddings, language modeling, and sequence labeling
- Lexical resources and terminology extraction
- Morphology and part-of-speech tagging
- Syntax and chunking
- Named entity recognition
- Word sense disambiguation and anaphora resolution
- Semantics, discourse, and dialog
- Machine translation and multilingualism
- Sentiment analysis, opinion mining, subjectivity, and social media
- Text classification and categorization
- Information extraction
- Applications

The 2016 event received submissions from 54 countries. A total of 298 papers by 671 authors were submitted for evaluation by the international Program Committee (see Fig. 1 and Tables 1 and 2). This two-volume set contains revised versions of 89 regular papers selected for presentation, with the acceptance rate of 29.8%.

Table 1. Number of submissions and accepted papers by topic^a

Submissions	Topic
60	Text mining
58	Information extraction
53	Lexical resources
44	Emotions, sentiment analysis, opinion mining
43	Information retrieval
42	Semantics, pragmatics, discourse
39	Clustering and categorization
36	Under-resourced languages
35	Machine translation and multilingualism
33	Morphology
33	Practical applications
28	Social networks and microblogging
25	Named entity recognition
25	POS tagging
23	Syntax and chunking
21	Noisy text processing and cleaning
20	Formalisms and knowledge representation
16	Plagiarism detection and authorship attribution
16	Question answering
16	Summarization
16	Word sense disambiguation
13	Computational terminology
13	Natural language interfaces
12	Other
11	Natural language generation
10	Speech processing
7	Spelling and grammar check
7	Textual entailment
6	Coreference resolution

^aAs indicated by the authors. A paper may belong to more than one topic.

In addition to regular papers and an In Memoriam paper, the books features papers by the keynote speakers:

- Pascale Fung, Hong Kong University of Science and Technology, Hong Kong
- Tomas Mikolov, Facebook AI Research, USA
- Simone Teufel, University of Cambridge, UK
- Piek Vossen, Vrije Universiteit Amsterdam, The Netherlands

Publication of full-text invited papers in the proceedings is a distinctive feature of the CICLing conferences. In addition to a presentation of their invited papers, the keynote speakers organized separate lively informal events; this is also a special feature of this conference series.

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

Country or region	Authors	Submissions ^b	Country or region	Authors	Submissions ^b
Algeria	15	6.5	Libya	2	0.5
Argentina	2	1	Mexico	8	5.17
Australia	5	1.33	Morocco	13	4.81
Austria	1	0.33	Nigeria	3	1
Brazil	8	3.75	Norway	10	3.96
Canada	24	10.78	Pakistan	1	1
China	23	11.25	Peru	3	2
Colombia	7	3	Poland	2	1
Czechia	19	8	Portugal	6	2.05
Egypt	17	6	Qatar	11	2.83
Finland	1	0.25	Romania	12	5.75
France	56	19.52	Russia	12	7.25
Germany	15	5.93	Saudi Arabia	5	1.3
Greece	3	1	Singapore	8	2.5
Hong Kong	2	1	South Africa	3	1
Hungary	6	5	South Korea	7	4
India	73	37.3	Spain	9	3.78
Indonesia	3	1	Sri Lanka	12	5
Iran	5	2.33	Switzerland	4	2
Ireland	3	1.25	The Netherlands	1	0.2
Israel	10	3.24	Tunisia	86	40.35
Italy	5	1	Turkey	53	29.22
Japan	13	3.5	Turkmenistan	1	0.33
Jordan	4	3	UAE	4	1.5
Kazakhstan	19	6.75	UK	18	9.17
Latvia	2	1	USA	30	13.4
Lebanon	2	0.67	Vietnam	4	1.25
			<i>Total:</i>	671	298

^b 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.

With this event, we continued 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 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.

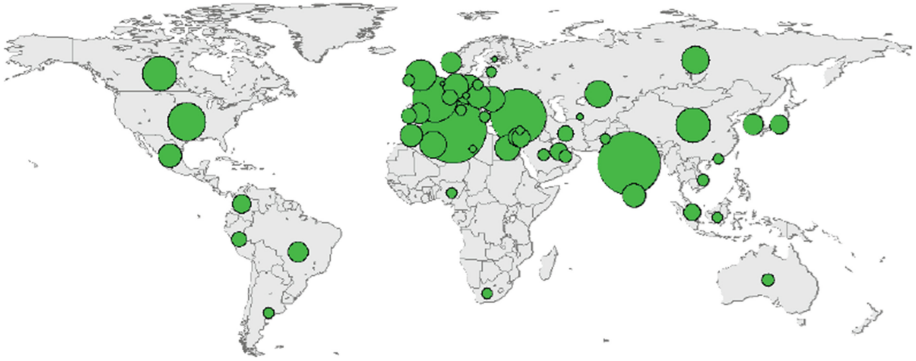


Fig. 1. Submissions by country or region. The area of a circle represents the number of submitted papers

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—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.

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

- | | |
|-----------------------|--|
| Best Paper 1st Place: | <p>“Mining the Web for Collocations: IR Models of Term Associations,” by Rakesh Verma, Vasanthi Vuppuluri, An Nguyen, Arjun Mukherjee, Ghita Mammar, Shahryar Baki, and Reed Armstrong, USA</p> <p>“Extracting Aspect Specific Sentiment Expressions Implied Negative Opinions,” by Arjun Mukherjee, USA</p> |
| Best Paper 2nd Place: | <p>“Word Sense Disambiguation Using Swarm Intelligence: A Bee Colony Optimization Approach,” by Saket Kumar and Omar El Ariss, USA</p> |

- Best Paper 3rd Place: “Corpus Frequency and Affix Ordering in Turkish,” by Mustafa Aksan, Umut Ufuk Demirhan, and Yeşim Aksan, Turkey
- Best Student Paper¹: “Generating Bags of Words from the Sums of Their Word Embeddings,” by Lyndon White, Roberto Togneri, Wei Liu, and Mohammed Bennamoun, Australia
- Verifiability 1st Place: “Pluralizing Nouns in isiZulu and Related Languages,” by Joan Byamugisha, C. Maria Keet, and Langa Khumalo, South Africa
- Verifiability 2nd Place: “A Free/Open-Source Hybrid Morphological Disambiguation Tool for Kazakh,”² by Zhenisbek Assylbekov, Jonathan North Washington, Francis Tyers, Assulan Nurkas, Aida Sundetova, Aidana Karibayeva, Balzhan Abduali, and Dina Amirova, Kazakhstan, USA, and Norway
- Verifiability 3rd Place: “An Informativeness Approach to Open IE Evaluation,” by William Léchelle and Philippe Langlais, Canada

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 the 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 Tinaztepe Cavern and nearby attractions; the Underground City and the amazing Cappadocia landscape, probably the most astonishing natural landscape I have seen so far; as well as the attractions and historical monuments of the city of Konya, the ancient capital of the Seljuk Sultanate of Rum and the Karamanids.

The conference was accompanied by two satellite events: the Second International Conference on Arabic Computational Linguistics, ACLing 2016, and the First International Conference on Turkic Computational Linguistics, TurCLing 2016, both founded by CICLing. 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. The Program Committees of ACLing 2016 and TurCLing 2016 have helped the CICLing 2016 committee in evaluation of a number of submissions on the corresponding narrow topics.

¹ The best student paper was selected 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 the proceedings of a satellite event of the conference and not in this book set.

I would like to thank all those involved in the organization of this conference. In the first place, the authors of the papers that constitute this book: It is the excellence of their research work that gives value to the book and meaning to the work of all other people. I thank all those who served on the Program Committee of CICLing 2016, the Second International Conference on Arabic Computational Linguistics, ACLing 2016, and the First International Conference on Turkic Computational Linguistics, TurCLing 2016; the Software Reviewing Committee, the Award Selection Committee, as well as additional reviewers, for their hard and very professional work. Special thanks go to Ted Pedersen, Manuel Vilares Ferro, and Soujanya Poria for their invaluable support in the reviewing process.

I also want to cordially thank the conference staff, volunteers, and the members of the local Organizing Committee headed by Hatem Haddad, as well as the organizers of the satellite events, Samhaa R. El-Beltagy of Nile University, Egypt, Khaled Shaalan of the British University in Dubai, UAE, and Bahar Karaoglan of Ege University, Turkey. I am deeply grateful to the administration of the Mevlana University for their helpful support, warm hospitality, and in general for providing this wonderful opportunity of holding CICLing in Turkey. I acknowledge support from the project CON-ACYT Mexico–DST India 122030 “Answer Validation Through Textual Entailment” and SIP-IPN grant 20161958.

The entire submission and reviewing process was supported for free by the EasyChair system (www.EasyChair.org). Last but not least, I deeply appreciate the Springer team’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.

December 2016

Alexander Gelbukh

Organization

CICLing 2016 was hosted by the Mevlana University, Turkey, and was organized by the CICLing 2016 Organizing Committee in conjunction with the Mevlana University, the Natural Language and Text Processing Laboratory of the Centro de Investigación en Computación (CIC) of the Instituto Politécnico Nacional (IPN), Mexico, and the Mexican Society of Artificial Intelligence (SMIA).

Organizing Chair

Hatem Haddad

Université Libre de Bruxelles, Belgium

Organizing Committee

Hatem Haddad (Chair)

Université Libre de Bruxelles, Belgium

Armağan Ozkaya

Mevlana University, Turkey

Niyazi Serdar Tunaboğlu

Mevlana University, Turkey

Alaa Eleyan

Mevlana University, Turkey

Mustafa Kaiiali

Mevlana University, Turkey

Mohammad Shukri Salman

Mevlana University, Turkey

Gülden Eleyan

Mevlana University, Turkey

Hasan Ucar

Mevlana University, Turkey

Program Chair

Alexander Gelbukh

Instituto Politécnico Nacional, Mexico

Program Committee

Ajith Abraham

Machine Intelligence Research Labs (MIR Labs), USA

Rania Al-Sabbagh

University of Illinois at Urbana-Champaign, USA

Galia Angelova

Bulgarian Academy of Sciences, Bulgaria

Marianna Apidianaki

LIMSI-CNRS, France

Alexandra Balahur

European Commission Joint Research Centre, Italy

Sivaji Bandyopadhyay

Jadavpur University, India

Leslie Barrett

Bloomberg, LP, USA

Roberto Basili

University of Rome Tor Vergata, Italy

Anja Belz

University of Brighton, UK

Christian Boitet

Université Joseph Fourier and LIG, GETALP, France

Igor Bolshakov

Independent Researcher, Russia

Nicoletta Calzolari

Istituto di Linguistica Computazionale – CNR, Italy

Erik Cambria

Nanyang Technological University, Singapore

Nick Campbell	TCD, Ireland
Michael Carl	Copenhagen Business School, Denmark
Niladri Chatterjee	IIT Delhi, India
Dan Cristea	A. I. Cuza University of Iași, Romania
Samhaa R. El-Beltagy	Nile University, Egypt
Michael Elhadad	Ben-Gurion University of the Negev, Israel
Anna Feldman	Montclair State University, USA
Alexander Gelbukh	Instituto Politécnico Nacional, Mexico
Dafydd Gibbon	Universität Bielefeld, Germany
Roxana Girju	University of Illinois at Urbana-Champaign, USA
Gregory Grefenstette	IHMC, USA
Hatem Haddad	Université Libre de Bruxelles, Belgium
Eva Hajičová	Charles University, Czechia
Sanda Harabagiu	The University of Texas at Dallas, USA
Yasunari Harada	Waseda University, Japan
Ales Horak	Masaryk University, Czechia
Nancy Ide	Vassar College, USA
Diana Inkpen	University of Ottawa, Canada
Aminul Islam	University of Louisiana at Lafayette, USA
Guillaume Jacquet	JRC, Italy
Miloš Jakubiček	Lexical Computing, Czech Republic
Doug Jones	MIT, USA
Sylvain Kahane	Modyco, Université Paris Ouest Nanterre La Défense and CNRS/Alpage, Inria, France
Dimitar Kazakov	University of York, UK
Alma Kharrat	Microsoft, USA
Philipp Koehn	Johns Hopkins University, USA
Leila Kosseim	Concordia University, Canada
Mathieu Lafourcade	LIRMM, France
Bing Liu	University of Illinois at Chicago, USA
Cerstin Mahlow	University of Stuttgart, Germany
Suresh Manandhar	University of York, UK
Diana McCarthy	University of Cambridge, UK
Alexander Mehler	Goethe University Frankfurt am Main, Germany
Farid Meziane	University of Salford, UK
Rada Mihalcea	University of North Texas/Oxford University, USA
Evangelos Milios	Dalhousie University, Canada
Ruslan Mitkov	University of Wolverhampton, UK
Dunja Mladenic	Jozef Stefan Institute, Slovenia
Hermann Moisl	Newcastle University, UK
Masaki Murata	Tottori University, Japan
Preslav Nakov	Qatar Computing Research Institute, Qatar Foundation, Qatar
Costanza Navarretta	University of Copenhagen, Denmark
Nicolas Nicolov	J.D. Power and Associates/McGraw-Hill, USA
Joakim Nivre	Uppsala University, Sweden

Kjetil Nørvåg	Norwegian University of Science and Technology, Norway
Attila Novák	Pázmány Péter Catholic University, Hungary
Nir Ofek	Ben-Gurion University, Israel
Partha Pakray	National Institute of Technology Mizoram, India
Ivandre Paraboni	University of São Paulo, Brazil
Patrick Saint-Dizier	IRIT-CNRS, France
Maria Teresa Pazienza	University of Rome Tor Vergata, Italy
Ted Pedersen	University of Minnesota, USA
Viktor Pekar	University of Birmingham, UK
Anselmo Peñas	UNED, Spain
Soujanya Poria	Nanyang Technological University, Singapore
Marta R. Costa-Jussà	Institute for Infocomm Research, Singapore
Fuji Ren	University of Tokushima, Japan
German Rigau	IXA Group, UPV/EHU, Spain
Fabio Rinaldi	University of Zurich, Switzerland
Horacio Rodriguez	Universitat Politècnica de Catalunya, Spain
Paolo Rosso	Universitat Politècnica de València, Spain
Vasile Rus	The University of Memphis, USA
Franco Salvetti	University of Colorado at Boulder/Microsoft, USA
Rajeev Sangal	Language Technologies Research Centre, India
Kepa Sarasola	Euskal Herriko Unibertsitatea, Spain
Fabrizio Sebastiani	Qatar Computing Research Institute, Qatar
Serge Sharoff	UoLeeds, UK
Bernadette Sharp	Staffordshire University, UK
Grigori Sidorov	Instituto Politécnico Nacional, Mexico
Kiril Simov	Linguistic Modelling Laboratory, IICT-BAS, Bulgaria
John Sowa	VivoMind Intelligence, USA
Efstathios Stamatatos	University of the Aegean, Greece
Maosong Sun	Tsinghua University, China
Jun Suzuki	NTT, Japan
Stan Szpakowicz	University of Ottawa, Canada
Hristo Tanev	Independent Researcher, Italy
Juan-Manuel Torres-Moreno	Laboratoire Informatique d'Avignon/UAPV, France
George Tsatsaronis	Technical University of Dresden, Germany
Olga Uryupina	University of Trento, Italy
Manuel Vilares Ferro	University of Vigo, Spain
Aline Villavicencio	Federal University of Rio Grande do Sul, Brazil
Piotr W. Fuglewicz	TiP Sp. z o. o., Poland
Marilyn Walker	UCSC, USA
Andy Way	ADAPT Centre, Dublin City University, Ireland
Bonnie Webber	The University of Edinburgh, UK
Alisa Zhila	IBM, USA

Software Reviewing Committee

Ted Pedersen	University of Minnesota, USA
Florian Holz	University of Leipzig, Germany
Miloš Jakubiček	Masaryk University, Czech Republic
Sergio Jiménez Vargas	Instituto Caro y Cuervo, Colombia
Miikka Silfverberg	University of Colorado, USA
Ronald Winnemöller	University of Hamburg, Germany

Award Committee

Alexander Gelbukh	Instituto Politécnico Nacional, Mexico
Eduard Hovy	Carnegie Mellon University, USA
Rada Mihalcea	University of Michigan, USA
Ted Pedersen	University of Minnesota, USA
Yorick Wilks	University of Sheffield, UK

Additional Reviewers

Abhijit Mishra	Kyoko Kanzaki
Abidrahman Moh'D	Leonardo Zilio
Aitor García	Magdalena Jankowska
Alisa Zhila	Mahsa Forati
Aljaz Kosmerlj	Maite Giménez
Amir Hossein Razavi	Majid Laali
Avi Hayoun	Marc Franco Salvador
Benjamin Marie	Marcelo Sardelich
Bernardo Cabaleiro	Michael Mohler
Bryan Rink	Miguel Angel Rios Gaona
Carla Parra Escartín	Olga Kolesnikova
Diana Trandabăț	Petya Osenova
Egoitz Laparra	Pidong Wang
Elnaz Davoodi	Pistol Ionut Cristian
Enrique Flores	Richard Evans
Francisco J. Ribadas-Pena	Rodrigo Wilkens
Hanna Bechara	Silvio Ricardo Cordeiro
Helena Gómez Adorno	Sonia Ordoñez Salinas
Hiram Calvo	Svetla Boytcheva
Iliia Markov	Tal Baumel
Irazú Hernández Farias	Travis Goodwin
Jan R. Benetka	Victor Darriba
Jan Rupnik	Victoria Yaneva
Janez Starc	Vinita Nahar
Jared Bernstein	Vojtech Kovar
Jie Mei	Xin Chen
Jumana Nassour-Kassis	Zuzana Neverilova
Kazuhiro Takeuchi	

Second Arabic Computational Linguistics Conference

Program Chairs

Samhaa R. El-Beltagy
Alexander Gelbukh
Khaled Shaalan

Nile University, Egypt
Instituto Politécnico Nacional, Mexico
The British University in Dubai, UAE

Program Committee

Bayan Abushawar
Hanady Ahmed
Hend Alkhalifa
Mohammed Attia
Aladdin Ayeshe
Karim Bouzoubaa
Violetta Cavalli-Sforza
Khalid Choukri
Samhaa R. El-Beltagy
Ossama Emam
Aly Fahmy
Ahmed Guessoum

Nizar Habash
Hatem Haddad
Kais Haddar

Lamia Hadrich Belguith
Imtiaz Khan
Alma Kharrat
Sattar Izwaini
Mark Lee
Sherif Mahdy Abdou
Farid Meziane
Farhad Oroumchian
Hermann Moisl
Ahmed Rafea
Allan Ramsay
Mohsen Rashwan
Horacio Rodriguez
Paolo Rosso
Nasredine Semmar
Khaled Shaalan
William Teahan
Imed Zitouni

Arab Open University, Jordan
Qatar University, Qatar
King Saud University, Saudi Arabia
Al-Azhar University, Egypt
De Montfort University, UK
Mohamed V University, Morocco
Al Akhawayn University, Morocco
ELDA, France
Nile University, Egypt
ALTEC, Egypt
Cairo University, Egypt
University of Science and Technology Houari
Boumediene, Algeria
New York University Abu Dhabi, UAE
Université Libre de Bruxelles, Belgium
MIRACL Laboratory, Faculté des sciences de Sfax,
Tunisia
MIRACL Laboratory, Tunisia
King Abdulaziz University, Jeddah, Saudi Arabia
Microsoft, USA
American University of Sharjah, UAE
University of Birmingham, UK
RDI, Egypt
University of Salford, UK
University of Wollongong in Dubai, UAE
Newcastle University, UK
American University in Cairo, Egypt
The University of Manchester, UK
Cairo University, Egypt
Universitat Politècnica de Catalunya, Spain
Universitat Politècnica de València, Spain
CEA, France
The British University in Dubai, UAE
Bangor University, UK
IBM Research, USA

Additional Reviewers

Asma Aouichat
Hind Saddiki
Mohamed Hadj Ameur
Muhammad Shuaib Qureshi

Muhammad Umair
Nora Al-Twairesh
Riadh Belkebir

First International Conference on Turkic Computational Linguistics

Program Chairs

Bahar Karaoglan (Chair)	Ege University, Turkey
Tarık Kışla (Co-chair)	Ege University, Turkey
Senem Kumova (Co-chair)	İzmir Ekonomi University, Turkey
Hatem Haddad (Co-chair)	Université Libre de Bruxelles, Belgium

Program Committee

Yeşim Aksan	Mersin University, Turkey
Adil Alpkocak	Dokuz Eylül University, Turkey
Ildar Batyrshin	Instituto Politécnico Nacional, Mexico
Cem Bozsahin	Middle East Technical University, Turkey
Fazli Can	Bilkent University, Turkey
Ilyas Cicekli	Hacettepe University, Turkey
Gülşen Eryiğit	Istanbul Technical University, Turkey
Alexander Gelbukh	Instituto Politécnico Nacional, Mexico
Tunga Gungor	Boğaziçi University, Turkey
Hatem Haddad	Université Libre de Bruxelles, Belgium
Bahar Karaoglan	Ege University, Turkey
Tarık Kışla	Ege University, Turkey
Senem Kumova Metin	İzmir University of Economics, Turkey
Altynbek Sharipbayev	L. N. Gumilyov Eurasian National University, Kazakhstan
Dzhavdet Suleymanov	Academy of Sciences of Tatarstan, Russia

Additional Reviewers

A. Sumru Özsoy
Ali Erkan
Cem Rıfkı Aydın
Dilyara Yakubova

Ezgi Yıldırım
Nihal Yağmur Aydın
Razieh Ehsani

Website and Contact

The website of the CICLing conference series is <http://www.CICLing.org>. It provides information about past CICLing conferences and their satellite events, including links to published papers (many of them in open access) or their abstracts, as well as photos and video recordings of keynote talks. In addition, it provides data, algorithms, and open-source software accompanying accepted papers, in accordance with the CICLing verifiability, reproducibility, and working description policy. It also provides information about the forthcoming CICLing events, as well as contact options.

Contents – Part II

Machine Translation and Multilingualism

Enabling Medical Translation for Low-Resource Languages	3
<i>Ahmad Musleh, Nadir Durrani, Irina Temnikova, Preslav Nakov, Stephan Vogel, and Osama Alsaad</i>	
Combining Phrase and Neural-Based Machine Translation: What Worked and Did Not	17
<i>Marta R. Costa-jussà and José A. R. Fonollosa</i>	
Combining Machine Translated Sentence Chunks from Multiple MT Systems	27
<i>Matīss Rikters and Inguna Skadiņa</i>	
Forest to String Based Statistical Machine Translation with Hybrid Word Alignments	38
<i>Santanu Pal, Sudip Kumar Naskar, and Josef van Genabith</i>	
Instant Translation Model Adaptation by Translating Unseen Words in Continuous Vector Space	51
<i>Shonosuke Ishiwatari, Naoki Yoshinaga, Masashi Toyoda, and Masaru Kitsuregawa</i>	
Fast-Syntax-Matching-Based Japanese-Chinese Limited Machine Translation	63
<i>Wuying Liu, Lin Wang, and Xing Zhang</i>	
A Classifier-Based Preordering Approach for English-Vietnamese Statistical Machine Translation	74
<i>Việt Hong Tran, Huyen Thuong Vu, Vinh Van Nguyen, and Minh Le Nguyen</i>	
Quality Estimation for English-Hungarian Machine Translation Systems with Optimized Semantic Features.	88
<i>Zijian Győző Yang, László János Laki, and Borbála Siklósi</i>	
Genetic-Based Decoder for Statistical Machine Translation.	101
<i>Douib Ameur, Langlois David, and Smaïli Kamel</i>	
Bilingual Contexts from Comparable Corpora to Mine for Translations of Collocations	115
<i>Shiva Taslimipoor, Ruslan Mitkov, Gloria Corpas Pastor, and Afsaneh Fazly</i>	

Bi-text Alignment of Movie Subtitles for Spoken English-Arabic Statistical Machine Translation	127
<i>Fahad Al-Obaidli, Stephen Cox, and Preslav Nakov</i>	

A Parallel Corpus of Translationese.	140
<i>Ella Rabinovich, Shuly Wintner, and Ofek Luis Lewinsohn</i>	

A Low Dimensionality Representation for Language Variety Identification. . .	156
<i>Francisco Rangel, Marc Franco-Salvador, and Paolo Rosso</i>	

Sentiment Analysis, Opinion Mining, Subjectivity, and Social Media

Invited Paper:

Towards Empathetic Human-Robot Interactions	173
<i>Pascale Fung, Dario Bertero, Yan Wan, Anik Dey, Ricky Ho Yin Chan, Farhad Bin Siddique, Yang Yang, Chien-Sheng Wu, and Ruixi Lin</i>	

Best Paper Award, First Place:

Extracting Aspect Specific Sentiment Expressions Implying Negative Opinions.	194
<i>Arjun Mukherjee</i>	

Aspect Terms Extraction of Arabic Dialects for Opinion Mining Using Conditional Random Fields.	211
<i>Alawya Alawami</i>	

Large Scale Authorship Attribution of Online Reviews	221
<i>Prasha Shrestha, Arjun Mukherjee, and Thamar Solorio</i>	

Discovering Correspondence of Sentiment Words and Aspects	233
<i>Geli Fei, Zhiyuan (Brett) Chen, Arjun Mukherjee, and Bing Liu</i>	

Aspect Based Sentiment Analysis: Category Detection and Sentiment Classification for Hindi	246
<i>Md Shad Akhtar, Asif Ekbal, and Pushpak Bhattacharyya</i>	

A New Emotional Vector Representation for Sentiment Analysis	258
<i>Hanen Ameer, Salma Jamoussi, and Abdelmajid Ben Hamadou</i>	

Cascading Classifiers for Twitter Sentiment Analysis with Emotion Lexicons	270
<i>Hiram Calvo and Omar Juárez Gambino</i>	

A Multilevel Approach to Sentiment Analysis of Figurative Language in Twitter.	281
<i>Braja Gopal Patra, Soumadeep Mazumdar, Dipankar Das, Paolo Rosso, and Sivaji Bandyopadhyay</i>	
Determining Sentiment in Citation Text and Analyzing Its Impact on the Proposed Ranking Index	292
<i>Souvik Ghosh, Dipankar Das, and Tanmoy Chakraborty</i>	
Combining Lexical Features and a Supervised Learning Approach for Arabic Sentiment Analysis	307
<i>Samhaa R. El-Beltagy, Talaat Khalil, Amal Halaby, and Muhammad Hammad</i>	
Sentiment Analysis in Arabic Twitter Posts Using Supervised Methods with Combined Features	320
<i>Rihab Bouchlaghem, Aymen Elkhelifi, and Rim Faiz</i>	
Interactions Between Term Weighting and Feature Selection Methods on the Sentiment Analysis of Turkish Reviews	335
<i>Tuba Parlar, Selma Ayşe Özel, and Fei Song</i>	
Developing a Concept-Level Knowledge Base for Sentiment Analysis in Singlish	347
<i>Rajiv Bajpai, Danyuan Ho, and Erik Cambria</i>	
Using Syntactic and Semantic Features for Classifying Modal Values in the Portuguese Language	362
<i>João Sequeira, Teresa Gonçalves, Paulo Quaresma, Amália Mendes, and Iris Hendrickx</i>	
Detecting the Likely Causes Behind the Emotion Spikes of Influential Twitter Users	374
<i>Calkin Suero Montero, Hatem Haddad, Maxim Mozgovoy, and Chedi Bechikh Ali</i>	
Age Identification of Twitter Users: Classification Methods and Sociolinguistic Analysis.	385
<i>Vasiliki Simaki, Iosif Mporas, and Vasileios Megalooikonomou</i>	
Mining of Social Networks from Literary Texts of Resource Poor Languages	396
<i>Pattabhi R. K. Rao and Sobha Lalitha Devi</i>	
Collecting and Annotating Indian Social Media Code-Mixed Corpora	406
<i>Anupam Jamatia, Björn Gambäck, and Amitava Das</i>	

Turkish Normalization Lexicon for Social Media	418
<i>Seniz Demir, Murat Tan, and Berkay Topcu</i>	

Text Classification and Categorization

Introducing Semantics in Short Text Classification	433
<i>Ameni Bouaziz, Célia da Costa Pereira, Christel Dartigues-Paliez, and Frédéric Precioso</i>	
Topics and Label Propagation: Best of Both Worlds for Weakly Supervised Text Classification.	446
<i>Sachin Pawar, Nitin Ramrakhiyani, Swapnil Hingmire, and Girish K. Palshikar</i>	
Deep Neural Networks for Czech Multi-label Document Classification	460
<i>Ladislav Lenc and Pavel Král</i>	
Turkish Document Classification with Coarse-Grained Semantic Matrix.	472
<i>İlknur Dönmez and Eşref Adalı</i>	
Supervised Topic Models for Diagnosis Code Assignment to Discharge Summaries.	485
<i>Mohamed Dermouche, Julien Velcin, Rémi Flicoteaux, Sylvie Chevret, and Namik Taright</i>	

Information Extraction

Invited Papers:

Identity and Granularity of Events in Text	501
<i>Piek Vossen and Agata Cybulska</i>	
An Informativeness Approach to Open IE Evaluation	523
<i>William Léchelle and Philippe Langlais</i>	
End-to-End Relation Extraction Using Markov Logic Networks	535
<i>Sachin Pawar, Pushpak Bhattacharya, and Girish K. Palshikar</i>	
Knowledge Extraction with NooJ Using a Syntactico-Semantic Approach for the Arabic Utterances Understanding	552
<i>Chahira Lhioui, Anis Zouaghi, and Mounir Zrigui</i>	
Adapting TimeML to Basque: Event Annotation.	565
<i>Begoña Altuna, María Jesús Aranzabe, and Arantza Díaz de Ilarraza</i>	

Applications

Invited Papers:

Deeper Summarisation: The Second Time Around: An Overview and Some Practical Suggestions	581
<i>Simone Teufel</i>	
Tracing Language Variation for Romanian	599
<i>Daniela Gîfu and Radu Simionescu</i>	
Aoidos: A System for the Automatic Scansion of Poetry Written in Portuguese	611
<i>Adiel Mittmann, Aldo von Wangenheim, and Alckmar Luiz dos Santos</i>	
Author Index	629

Contents – Part I

In Memoriam

Invited Paper:

Adam Kilgarriff’s Legacy to Computational Linguistics and Beyond	3
<i>Roger Evans, Alexander Gelbukh, Gregory Grefenstette, Patrick Hanks, Miloš Jakubiček, Diana McCarthy, Martha Palmer, Ted Pedersen, Michael Rundell, Pavel Rychlý, Serge Sharoff, and David Tugwell</i>	

General Formalisms

Invited Papers:

A Roadmap Towards Machine Intelligence.	29
<i>Tomas Mikolov, Armand Joulin, and Marco Baroni</i>	
Algebraic Specification for Interoperability Between Data Formats: Application on Arabic Lexical Data.	62
<i>Malek Lhioui, Kais Haddar, and Laurent Romary</i>	
Persianp: A Persian Text Processing Toolbox	75
<i>Mahdi Mohseni, Javad Ghofrani, and Heshaam Faili</i>	

Embeddings, Language Modeling, and Sequence Labeling

Best Student Paper Award:

Generating Bags of Words from the Sums of Their Word Embeddings	91
<i>Lyndon White, Roberto Togneri, Wei Liu, and Mohammed Bennamoun</i>	
New Word Analogy Corpus for Exploring Embeddings of Czech Words	103
<i>Lukáš Svoboda and Tomáš Brychcín</i>	
Using Embedding Models for Lexical Categorization in Morphologically Rich Languages	115
<i>Borbála Siklósi</i>	
A New Language Model Based on Possibility Theory	127
<i>Mohamed Amine Menacer, Abdelfetah Boumerdas, Chahnez Zakaria, and Kamel Smaili</i>	

Combining Discrete and Neural Features for Sequence Labeling	140
<i>Jie Yang, Zhiyang Teng, Meishan Zhang, and Yue Zhang</i>	
New Recurrent Neural Network Variants for Sequence Labeling	155
<i>Marco Dinarelli and Isabelle Tellier</i>	

Lexical Resources and Terminology Extraction

Best Paper Award, First Place:

Mining the Web for Collocations: IR Models of Term Associations	177
<i>Rakesh Verma, Vasanthi Vuppuluri, An Nguyen, Arjun Mukherjee, Ghita Mammari, Shahryar Baki, and Reed Armstrong</i>	
A Continuum-Based Model of Lexical Acquisition	195
<i>Pierre Marchal and Thierry Poibeau</i>	
Description of Turkish Paraphrase Corpus Structure and Generation Method	208
<i>Bahar Karaoglan, Tarık Kışla, and Senem Kumova Metin</i>	
Extracting Terminological Relationships from Historical Patterns of Social Media Terms	218
<i>Daoud Daoud and Mohammad Daoud</i>	
Adaptation of Cross-Lingual Transfer Methods for the Building of Medical Terminology in Ukrainian	230
<i>Thierry Hamon and Natalia Grabar</i>	
Adaptation of a Term Extractor to Arabic Specialised Texts: First Experiments and Limits	242
<i>Wafa Neifar, Thierry Hamon, Pierre Zweigenbaum, Mariem Ellouze Khemakhem, and Lamia Hadrich Belguith</i>	

Morphology and Part-of-Speech Tagging

Best Paper Award, Third Place:

Corpus Frequency and Affix Ordering in Turkish	257
<i>Mustafa Aksan, Umut Ufuk Demirhan, and Yeşim Aksan</i>	
Pluralising Nouns in isiZulu and Related Languages	271
<i>Joan Byamugisha, C. Maria Keet, and Langa Khumalo</i>	
Morphological Analysis of Urdu Verbs	284
<i>Aneeta Niazi</i>	

Stemming and Segmentation for Classical Tibetan.	294
<i>Orna Almogi, Lena Dankin, Nachum Dershowitz, Yair Hoffman, Dimitri Pauls, Dorji Wangchuk, and Lior Wolf</i>	

Part of Speech Tagging for Polish: State of the Art and Future Perspectives	307
<i>Łukasz Kobyliński and Witold Kieraś</i>	

Turkish PoS Tagging by Reducing Sparsity with Morpheme Tags in Small Datasets	320
<i>Burcu Can, Ahmet Üstün, and Murathan Kurfalı</i>	

Part-of-Speech Tagging for Code Mixed English-Telugu Social Media Data	332
<i>Kovida Nelakuditi, Divya Sai Jitta, and Radhika Mamidi</i>	

Syntax and Chunking

Analysis of Word Order in Multiple Treebanks.	345
<i>Vladislav Kuboň, Markéta Lopatková, and Jiří Mirovský</i>	

A Framework for Language Resource Construction and Syntactic Analysis: Case of Arabic	356
<i>Nabil Khoufi, Chafik Aloulou, and Lamia Hadrach Belguith</i>	

Enhancing Neural Network Based Dependency Parsing Using Morphological Information for Hindi	366
<i>Agnivo Saha and Sudeshna Sarkar</i>	

Construction Grammar Based Annotation Framework for Parsing Tamil	378
<i>Vigneshwaran Muralidaran and Dipti Misra Sharma</i>	

Comparative Error Analysis of Parser Outputs on Telugu Dependency Treebank	397
<i>Silpa Kanneganti, Himani Chaudhry, and Dipti Misra Sharma</i>	

Gut, Besser, Chunker – Selecting the Best Models for Text Chunking with Voting	409
<i>Balázs Indig and István Endrédy</i>	

Named Entity Recognition

A Deep Learning Solution to Named Entity Recognition	427
<i>V. Rudra Murthy and Pushpak Bhattacharyya</i>	

Deep Learning Approach for Arabic Named Entity Recognition	439
<i>Mourad Gridach</i>	
Hybrid Feature Selection Approach for Arabic Named Entity Recognition	452
<i>Miran Shahine and Mohamed Sakre</i>	
Named-Entity-Recognition (NER) for Tamil Language Using Margin-Infused Relaxed Algorithm (MIRA)	465
<i>Pranavan Theivendiram, Megala Uthayakumar, Nilusija Nadarasamoorthy, Moganarangan Thayaparan, Sanath Jayasena, Gihan Dias, and Surangika Ranathunga</i>	

Word Sense Disambiguation and Anaphora Resolution

Best Paper Award, Second Place:

Word Sense Disambiguation Using Swarm Intelligence: A Bee Colony Optimization Approach	479
<i>Saket Kumar and Omar El Ariss</i>	
Verb Sense Annotation for Turkish PropBank via Crowdsourcing	496
<i>Gözde Gül Şahin</i>	
Coreference Resolution for French Oral Data: Machine Learning Experiments with ANCOR	507
<i>Adèle Désoyer, Frédéric Landragin, Isabelle Tellier, Anaïs Lefevre, Jean-Yves Antoine, and Marco Dinarelli</i>	
Arabic Anaphora Resolution Using Markov Decision Process	520
<i>Fériel Ben Fraj Trabelsi, Chiraz Ben Othmane Zribi, and Saoussen Mathlouthi</i>	
Arabic Pronominal Anaphora Resolution Based on New Set of Features	533
<i>Souha Mezghani Hammami and Lamia Hadrich Belguith</i>	

Semantics, Discourse, and Dialog

GpSense: A GPU-Friendly Method for Commonsense Subgraph Matching in Massively Parallel Architectures	547
<i>Ha-Nguyen Tran and Erik Cambria</i>	
Parameters Driving Effectiveness of LSA on Topic Segmentation	560
<i>Marwa Naili, Anja Chaibi Habacha, and Henda Hajjami Ben Ghezala</i>	

A New Russian Paraphrase Corpus. Paraphrase Identification and Classification Based on Different Prediction Models	573
<i>Ekaterina Pronoza and Elena Yagunova</i>	
Constructing a Turkish Corpus for Paraphrase Identification and Semantic Similarity	588
<i>Asli Eyecioglu and Bill Keller</i>	
Evaluation of Semantic Relatedness Measures for Turkish Language	600
<i>Ugur Sopaoglu and Gonenc Ercan</i>	
Using Sentence Semantic Similarity to Improve LMF Standardized Arabic Dictionary Quality	612
<i>Wafa Wali, Bilel Gargouri, and Abdelmajid Ben Hamadou</i>	
Multiword Expressions (MWE) for Mizo Language: Literature Survey.	623
<i>Goutam Majumder, Partha Pakray, Zoramdinthara Khiangte, and Alexander Gelbukh</i>	
Classification of Textual Genres Using Discourse Information	636
<i>Elnaz Davoodi, Leila Kosseim, Félix-Hervé Bachand, Majid Laali, and Emmanuel Argollo</i>	
Features for Discourse-New Referent Detection in Russian.	648
<i>Svetlana Toldova and Max Ionov</i>	
A Karaka Dependency Based Dialog Act Tagging for Telugu Using Combination of LMs and HMM	663
<i>Suman Dowlagar and Radhika Mamidi</i>	
Author Index	675