Communications
in Computer and Information Science

Commenced Publication in 2007
Founding and Former Series Editors:
Alfredo Cuzzocrea, Xiaoyong Du, Orhun Kara, Ting Liu, Dominik Ślęzak, and Xiaokang Yang

Editorial Board
Simone Diniz Junqueira Barbosa
Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil
Phoebe Chen
La Trobe University, Melbourne, Australia
Joaquim Filipe
Polytechnic Institute of Setúbal, Setúbal, Portugal
Igor Kotenko
St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, St. Petersburg, Russia
Krishna M. Sivalingam
Indian Institute of Technology Madras, Chennai, India
Takashi Washio
Osaka University, Osaka, Japan
Junsong Yuan
Nanyang Technological University, Singapore, Singapore
Lizhu Zhou
Tsinghua University, Beijing, China
More information about this series at http://www.springer.com/series/7899
Abdelmonaime Lachkar · Karim Bouzoubaa
Azzedine Mazroui · Abdelfettah Hamdani
Abdelhak Lekhouaja (Eds.)

Arabic Language Processing
From Theory to Practice
6th International Conference, ICALP 2017
Fez, Morocco, October 11–12, 2017
Proceedings

Springer
Welcome to the CCIS proceedings of the 6th International Conference on Arabic Language Processing – ICALP 2017. The conference was hosted in the imperial city Fez, Morocco, and organized by the Sidi Mohammed Ben Abdellah University, National School of Applied Sciences (ENSA-USMBA) in collaboration with the Arabic Language Engineering Society in Morocco (ALESM).

ICALP (ex-CITALA) started in 2006 and is held every two years. It is the flagship conference of the Arabic Language Engineering Society in Morocco (ALESM), which is the largest Arabic NLP institute in Morocco.

ICALP (ex-CITALA) is one of the premier international forums for disseminating new scholarly and technological work in Arabic computational linguistics; it has become one of the leading international conferences in the field of Arabic language processing and related applications. Each ICALP (ex-CITALA) conference attracts more authors from different countries in different continents thanks to the success of the preceding editions.

ICALP 2017 (ex-CITALA) provided excellent opportunities for scientists, researchers, industrial engineers, and university students to present their research achievements and to develop new collaborations and partnerships with experts in the field.

In this notable edition, we were indeed privileged to have had not one but two esteemed speakers. Our distinguished speakers were Dr. Khalid Choukri (ELRA Secretary General, ELDA founder, and CEO Evaluations and Language Resources Distribution Agency; ELDA), representing industry, and Prof. Allan Ramsay from Manschether University, UK, representing academia.

The keynote speakers shared their expertise with the aim of presenting to the participants a wide spectrum of their recent research developments in the field of Arabic NLP and related applications.

Also in this sixth edition, we were indeed privileged to have Springer publish the ICALP 2017 conference proceedings in their CCIS series.

ICALP 2017 focused on the following topics:
- Information retrieval systems (IRS)
- Question/answering systems (QAS)
- Information extraction (IE)
- Opining mining and sentiment analysis
- Text categorization and clustering
- Named entity recognition (NER)
- Information extraction
- Part-of-speech tagging (POS)
- Word sense disambiguation (WSD)
- Multi-word term extraction (MWTE)
- Paraphrasing and textual entailment
Building ontologies
Text summarization
Machine translation systems (MTS)
Cross-language applications: Arabic–English, Arabic–Spanish, Arabic–French
Study of linguistic phenomena
Language resources: linguistic dictionary-based resources (phonology, morphology, semantics)
Development and normalization of linguistic resources and their exploitation in Arabic NLP applications
Spell-check and automatic corrections
Generation and analysis (morphology, syntax, semantics)
Linguistic resources and NLP applications
Comparative linguistic studies
Speech recognition and synthesis
Optical character recognition (OCR)

It is gratifying to note that the program of the ICALP conference covered a wide range of very interesting items (papers and posters) relating to Arabic NLP, from theory to practice, resources, and ontologies.

For this sixth edition, we received 64 submissions from 15 countries (Morocco, Algeria, Belgium, Canada, Egypt, France, Germany, Ireland, Palestine Qatar, Saudi Arabia, Tunisia, Turkey, USA, and UK).

Over 65 technical experts from all over the world participated in the peer-review process. Based on the review results, the Program Chairs accepted only 18 papers for oral presentation and to be included in the CCIS proceedings published by Springer.

In addition, 15 papers were selected for poster presentation. We received many good submissions, but could only accept a small number of them for oral presentation. The overall quality of submissions was extremely high, and we have had to decline a number of very good papers like those selected as posters.

Each paper was assigned to five reviewers, and generally each one was reviewed by no less than three reviewers.

We take this opportunity to extend a heartfelt thanks to the reviewers for providing high-quality and insightful reviews under a tight schedule, and also to thank the authors for their interest and contributions, which resulted in an outstanding technical program.

We are extremely grateful to the sponsors of the conference.

We extend a special word of thanks to all our colleagues of the Organizing Committee and secretariat for their hard work in organizing the conference, and to Springer for their assistance in publishing the ICALP 2017 proceedings in the CCIS series.

Finally, we would like to take this opportunity to express our sincere thanks and deep appreciation to the organizers and in particular:

Prof. Omar Assobhei, President of USMBA University in Fez
Prof. Mustapha Ijaali, Director of ENSA, Dean of FST, USMBA University, Fez
Prof. Hassan Zaher, Dean of Faculty of Shari’a Sciences, USMBA University, Fez

All of them worked with us from the very beginning of the planning stage. We truly appreciate their dedication. The conference program would not have been possible without the following:
Arabic Language Engineering Society in Morocco (ALESM)
Université Sidi Mohamed Ben Abdellah (USMBA)
Ecole Nationale des Sciences Appliquées de Fès (ENSA-Fès)
Faculté des Sciences et Techniques (FST-Fès)
Faculté des Sciences de la Scharia
Centre National pour la Recherche Scientifique et Technique (CNRST-Morocco)
Académie Hassan II des sciences et techniques

On behalf of the Steering, Program, and Organizing Committees, we hope you enjoy the contributions in this volume.

October 2017

Abdelmonaime Lachkar
Karim Bouzoubaa
Azzedine Mazroui
Abdelfettah Hamdani
Abdelhak Lekhouaja
Organization

General Chair
Abdelmonaime Lachkar (Ex ENSA-USMBA, Fez), ENSA-UAE, Tangier, Morocco

Honorary Committee
Omar Assobhei President of USMBA University, Fez, Morocco
Moulhime El Bekkali Vice President of Sidi Mohamed Ben Abdellah University in Fez, Morocco
Brahim Akdim Vice President of Sidi Mohamed Ben Abdellah University in Fez, Morocco
Mustapha Ijaali Director of ENSA and Dean of FST, USMBA University, Fez, Morocco
Hassan Zaher Dean of Faculty of Shari’a Sciences, USMBA University, Fez, Morocco

Steering Committee
Karim Bouzoubaa EMI, UM5, Rabat, Morocco
Abdelhamid El Jihad IERA, UM5, Rabat, Morocco
Abdelfettah Hamdani IERA, UM5, Rabat, Morocco
Azzedine Lazrek UCA, Marrakech, Morocco
Abdelhak Lekhouaja UMP, Oujda, Morocco
Azzedine Mazroui UMP, Oujda, Morocco

Communication and Publicity Committee
Jaoua Ali Mohamed Qatar University, Qatar
Hassina Aliane Centre de Recherche sur l’Information Scientifique et Technique, Algeria
Khalid Choukri ELDA, European Language Resource Association, France
Haikal El Abed Institute for Communications Technology, Germany
Ali Farghaly DataFlux A SAS Company, USA
Salwa Hamada Taibah University, Madinah, Saudi Arabia
Abdul Malik Al-Salman King Saud University, Riyadh, Saudi Arabia
Elmohajir Mohammed Chair of IEEE Morocco Section, FSDM, USMBA, Fez, Morocco
Ouafae Nahli Institute for Computational Linguistics CNR, Pisa, Italy
Nasredine Semmar CEA Saclay Nano-INNOV, France
Khaled Shaalan  British University in Dubai, UAE  
Nadi Tomeh  Paris 13 University, France  
Imed Zitouni  Microsoft Research, USA 

**Local Organizing Chair**

Abdelmajid Saka  ENSA, USMBA, Fez, Morocco

**Local Organizing Committee**

Mohammed Berrada  ENSA, USMBA, Fez, Morocco  
Zakariya Chalh  ENSA, USMBA, Fez, Morocco  
Saad Bennani Dosse (IEEE Member)  ENSA, USMBA, Fez, Morocco  
Hakim El Fadili  ENSA, USMBA, Fez, Morocco  
Abdelmajid El Katani  USMBA, Fez, Morocco  
Youness Elkhamlchi  ENSA, USMBA, Fez, Morocco  
Souad Elkhattabi  ENSA, USMBA, Fez, Morocco  
Abdelhakim Essbaai (IEEE Member)  USMBA, Fez, Morocco  
Abdellatif Ezzouhairi (IEEE Member)  ENSA, USMBA, Fez, Morocco  
Khalid Fardousse  USMBA, Fez, Morocco  
Adil Kenzi (IEEE Member)  ENSA, USMBA, Fez, Morocco  
Youness Lakhirssi (IEEE Member)  ENSA, USMBA, Fez, Morocco  
Anas Mansouri (IEEE Member)  ENSA, USMBA, Fez, Morocco  
Elhaj Ben Ali Safae  ENSA, USMBA, Fez

**Finance Committee**

Bouklata Ahmed  EST, USMBA, Fez, Morocco  
Abdelhamid El Jihad  IERA, Rabat, Morocco  
El Faik Hanane  ENSA, USMBA, Fez, Morocco

**Webmaster and Social Media**

El Bekkali Mohammed  ENSA, USMBA, Fez, Morocco

**Program Committee**

M'Hamed AitKbir  FST, UAE Tanger, Morocco  
Hassina Aliane  Centre de Recherche sur l’Information Scientifique et Technique, Algeria  
Hend Alkhalifa  King Saud University, Saudi Arabia
<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wasfi Al-Khatib</td>
<td>King Fahd University of Petroleum and Minerals, Saudi Arabia</td>
</tr>
<tr>
<td>Aziza Alouaazizi</td>
<td>USMBA, Fez, Morocco</td>
</tr>
<tr>
<td>Abdul Malik Al Salman</td>
<td>King Saud University, Saudi Arabia</td>
</tr>
<tr>
<td>Mohammed Tabit Alouai</td>
<td>FS UMP, Oujda, Morocco</td>
</tr>
<tr>
<td>Almoataz Bellah Al-Said</td>
<td>Cairo University, Egypt</td>
</tr>
<tr>
<td>Yassine Benajiba</td>
<td>Thomson Reuters, IP and Science</td>
</tr>
<tr>
<td>Karim Bouzoubaa</td>
<td>EMI, UM5, Rabat, Morocco</td>
</tr>
<tr>
<td>Violetta Cavalli-Sforza</td>
<td>Al Akhawayn University, Morocco</td>
</tr>
<tr>
<td>Khalid Choukri</td>
<td>ELDA, European Language Resource Association, France</td>
</tr>
<tr>
<td>Oumayma Dakkak</td>
<td>Higher Institute for Applied Science and Technology, Syrian Arab Republic</td>
</tr>
<tr>
<td>Kareem Darwish</td>
<td>Qatar Computing Research Institute, Qatar</td>
</tr>
<tr>
<td>Said Desouki</td>
<td>Higher Institute for Applied Science and Technology, Syrian Arab Republic</td>
</tr>
<tr>
<td>Mahmoud El-Haj</td>
<td>Lancaster University, UK</td>
</tr>
<tr>
<td>Abdelhamid El Jihad</td>
<td>IERA, UM5, Rabat, Morocco</td>
</tr>
<tr>
<td>Abderrahim Elqadi</td>
<td>EST, UMI, Meknès, Morocco</td>
</tr>
<tr>
<td>Ramy Eskander</td>
<td>Columbia University, USA</td>
</tr>
<tr>
<td>Ali Farghaly</td>
<td>DataFlux A SAS Company, USA</td>
</tr>
<tr>
<td>Bilel Gargouri</td>
<td>Université de Sfax, Tunisia</td>
</tr>
<tr>
<td>Nada Ghneim</td>
<td>Higher Institute for Applied Science and Technology, Damascus, Syria</td>
</tr>
<tr>
<td>Ahmed Guessoum</td>
<td>University of Science and Technology Houari, Boumediene, Algeria</td>
</tr>
<tr>
<td>Hatem Haddad</td>
<td>Université Libre de Bruxelles, Belgium</td>
</tr>
<tr>
<td>Kais Haddar</td>
<td>University of Sfax, Tunisia</td>
</tr>
<tr>
<td>Salwa Hamada</td>
<td>Taibah University, Madinah, Saudi Arabia</td>
</tr>
<tr>
<td>Abdelfettah Hamdani</td>
<td>IERA, UM5, Rabat, Morocco</td>
</tr>
<tr>
<td>Yannis Haralambous</td>
<td>Institut Mines-Télécom and UMR CNRS 6285 Lab-STICC, France</td>
</tr>
<tr>
<td>Hamid Jaafar</td>
<td>UCA, Marrakech, Morocco</td>
</tr>
<tr>
<td>Mustafa Jarrar</td>
<td>Birzeit University, Sina Institute, Palestine</td>
</tr>
<tr>
<td>Ali Mohamed Jaoua</td>
<td>Qatar University, Qatar</td>
</tr>
<tr>
<td>Mohammed Kissi</td>
<td>FS, UCD, El Jadida, Morocco</td>
</tr>
<tr>
<td>Abdelmonaime Lachkar</td>
<td>ENSA, USMBA, Fez, Morocco</td>
</tr>
<tr>
<td>Abouenour Lahsen</td>
<td>EMI UM5, Rabat, Morocco</td>
</tr>
<tr>
<td>Azzedine Lazrek</td>
<td>UCA, Marrakech, Morocco</td>
</tr>
<tr>
<td>Abdelhak Lekhouaja</td>
<td>UMP, Oujda, Morocco</td>
</tr>
<tr>
<td>Mourad Loukam</td>
<td>Hassiba Benouali University, Algeria</td>
</tr>
<tr>
<td>Sabri A. Mahmoud</td>
<td>King Fahd University of Petroleum and Minerals, Saudi Arabia</td>
</tr>
<tr>
<td>Azzedine Mazroui</td>
<td>UMP, Oujda, Morocco</td>
</tr>
<tr>
<td>Slim Mesfar</td>
<td>Université de Carthage, Tunisie</td>
</tr>
<tr>
<td>Farid Meziane</td>
<td>University of Salford, UK</td>
</tr>
<tr>
<td>Name</td>
<td>Institution</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Vito Pirreli</td>
<td>Institute for Computational Linguistics CNR, Pisa, Italy</td>
</tr>
<tr>
<td>Noureddine Rais</td>
<td>FSDM, USMBA, Fez, Morocco</td>
</tr>
<tr>
<td>Allan Ramsay</td>
<td>University of Manchester, UK</td>
</tr>
<tr>
<td>Salim Rami</td>
<td>UCA Marrakech, Morocco</td>
</tr>
<tr>
<td>Horacio Rodríguez</td>
<td>Universitat Politècnica de Catalunya, Spain</td>
</tr>
<tr>
<td>Paolo Rosso</td>
<td>Universidad Politècnica de Valencia, Spain</td>
</tr>
<tr>
<td>Nasredine Semmar</td>
<td>CEA Saclay Nano-INNOV, France</td>
</tr>
<tr>
<td>Khaled Shaalan</td>
<td>The British University, United Arab Emirates</td>
</tr>
<tr>
<td>Kamel Smaili</td>
<td>University of Lorraine Nancy, France</td>
</tr>
<tr>
<td>Abderrahim Tragha</td>
<td>UH2, Casablanca, Morocco</td>
</tr>
<tr>
<td>Adnan Yahya</td>
<td>Birzeit University, Palestine</td>
</tr>
<tr>
<td>Abdellah Yousfi</td>
<td>Mohammed V University, Morocco</td>
</tr>
<tr>
<td>Wajdi Zaghouani</td>
<td>Carnegie Mellon University, Qatar</td>
</tr>
<tr>
<td>Mohamed Zakaria Kurdi</td>
<td>University of Delaware, USA</td>
</tr>
<tr>
<td>Arssalan Zarghili</td>
<td>USMBA, Fez, Morocco</td>
</tr>
<tr>
<td>Petr Zemánek</td>
<td>Charles University in Prague, Czech Republic</td>
</tr>
<tr>
<td>Taha Zerrouki</td>
<td>University of Bouira, Algeria</td>
</tr>
<tr>
<td>Ahmed ZineEddine</td>
<td>USMBA, Fez, Morocco</td>
</tr>
<tr>
<td>Farah Benamara Zitoune</td>
<td>IRIT, Université Paul Sabatier, France</td>
</tr>
<tr>
<td>Imed Zitouni</td>
<td>Microsoft Research, USA</td>
</tr>
<tr>
<td>Mounir Zrigui</td>
<td>University of Monastir, Tunisia</td>
</tr>
</tbody>
</table>

### Sponsors

**USMBA University**

Université Sidi Mohamed Ben Abdellah (USMBA)

**Arabic Language Engineering Society in Morocco (ALESM)**
Centre National pour la Recherche Scientifique et Technique (CNRST-Morocco)

Académie Hassan II des sciences et techniques

Ecole Nationale des Sciences Appliquées de Fès (ENSA-Fès)

LISA

Faculté des Sciences et Techniques, (FST-Fès)

Faculté des Sciences de la Sharia
Contents

Machine Translation Systems

An Automatic Approach for WordNet Enrichment Applied to Arabic WordNet .......................................................... 3
Mohamed Seghir Hadj Ameur, Ahlem Chérifa Khadir, and Ahmed Guessoum

Word Embedding-Based Approaches for Measuring Semantic Similarity of Arabic-English Sentences ..................... 19
El Moatez Billah Nagoudi, Jérémy Ferrero, Didier Schwab, and Hadda Cherroun

A POS-Based Preordering Approach for English-to-Arabic Statistical Machine Translation ...................................... 34
Mohamed Seghir Hadj Ameur, Ahmed Guessoum, and Farid Meziane

Speech Recognition and Synthesis

Towards a High-Quality Lemma-Based Text to Speech System for the Arabic Language .................................................. 53
Oumaima Zine, Abdelouafi Meziane, and Mohamed Boudchiche

Towards a Speech Recognizer for Multiple Languages Using Arabic Acoustic Model Application to Amazigh Language ...... 67
Ali Sadiqui and Ahmed Zinedine

A Game with a Purpose for Automatic Detection of Children’s Speech Disabilities Using Limited Speech Resources ........ 79
Reem Elhady, Mohamed Elmahdy, Injy Hamed, and Slim Abdennadher

Building a Rich Arabic Speech and Language Corpus Based on the Holy Quran ......................................................... 90
Ali Meftah, Yasser Seddiqu, Yousef Alotaibi, and Sid-Ahmed Selouani

Text Categorization, Clustering and Summarization

Combining Words and Concepts for Automatic Arabic Text Classification ........................................................... 105
Alaa Alahmadi, Arash Joorabchi, and Abdulhussein E. Mahdi

Modern Standard Arabic Readability Prediction ................................................................. 120
Naoual Nassiri, Abdelhak Lakhouaja, and Violetta Cavalli-Sforza
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Document Similarity for Arabic and Cross-Lingual Web Content</td>
<td>134</td>
</tr>
<tr>
<td>Ali Salhi and Adnan H. Yahya</td>
<td></td>
</tr>
<tr>
<td>Empirical Evaluation of Word Representations on Arabic Sentiment</td>
<td>147</td>
</tr>
<tr>
<td>Analysis</td>
<td></td>
</tr>
<tr>
<td>Mourad Gridach, Hatem Haddad, and Hala Mulki</td>
<td></td>
</tr>
<tr>
<td>A Survey of Extractive Arabic Text Summarization Approaches</td>
<td>159</td>
</tr>
<tr>
<td>Samira Lagrini, Mohammed Redjimi, and Nabiha Azizi</td>
<td></td>
</tr>
<tr>
<td><strong>Information Retrieval Systems</strong></td>
<td></td>
</tr>
<tr>
<td>Toward a Context-Aware Multilingual Personalized Search</td>
<td>175</td>
</tr>
<tr>
<td>Mohamed Seghir Hadj Ameur, Youcef Moulahoum, Lamia Berkani, and Ahmed Guessoum</td>
<td></td>
</tr>
<tr>
<td>Fusion Based Authorship Attribution-Application of Comparison Between</td>
<td></td>
</tr>
<tr>
<td>the Quran and Hadith</td>
<td>191</td>
</tr>
<tr>
<td>Halim Sayoud and Hassina Hadjadj</td>
<td></td>
</tr>
<tr>
<td>Automatic Identification of Moroccan Colloquial Arabic</td>
<td>201</td>
</tr>
<tr>
<td>Ridouane Tachicart, Karim Bouzoubaa, Si Lhoussaine Aouragh, and Hamid Jaafa</td>
<td></td>
</tr>
<tr>
<td><strong>Arabic NLP Tools and Applications</strong></td>
<td></td>
</tr>
<tr>
<td>Encoding Prototype of Al-Hadith Al-Shareef in TEI</td>
<td>217</td>
</tr>
<tr>
<td>Hajer Maraoui, Kais Haddar, and Laurent Romary</td>
<td></td>
</tr>
<tr>
<td>A New Tool for Benchmarking and Assessing Arabic Syntactic Parsers</td>
<td>230</td>
</tr>
<tr>
<td>Younes Jaafar and Karim Bouzoubaa</td>
<td></td>
</tr>
<tr>
<td>Syntactic Parsing of Simple Arabic Nominal Sentence Using the NooJ</td>
<td>244</td>
</tr>
<tr>
<td>Linguistic Platform</td>
<td></td>
</tr>
<tr>
<td>Said Bourahma, Samir Mbarki, Mohammed Mourchid, and Abdelaziz Mouloudi</td>
<td></td>
</tr>
<tr>
<td><strong>Author Index</strong></td>
<td>259</td>
</tr>
</tbody>
</table>