

Communications in Computer and Information Science

782

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

Editors

Abdelmonaime Lachkar
Ex ENSA-USMBA
Fez
Morocco

and

ENSA-UAE
Tangier
Morocco

Karim Bouzoubaa
EMI, UM5
Rabat
Morocco

Azzedine Mazroui
FS, UMP
Oujda
Morocco

Abdelfettah Hamdani
IERA, UM5
Rabat
Morocco

Abdelhak Lekhouaja
FS, UMP
Oujda
Morocco

ISSN 1865-0929

ISSN 1865-0937 (electronic)

Communications in Computer and Information Science

ISBN 978-3-319-73499-6

ISBN 978-3-319-73500-9 (eBook)

<https://doi.org/10.1007/978-3-319-73500-9>

Library of Congress Control Number: 2017962905

© Springer International Publishing AG 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 Springer Nature

The registered company is Springer International Publishing AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

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 Sharia
- 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 Elkhatabi	ENSA, USMBA, Fez, Morocco
Abdelhakim Essbaai	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 Lakhrissi (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

Wasfi Al-Khatib	King Fahd University of Petroleum and Minerals, Saudi Arabia
Aziza Alouazizi	USMBA, Fez, Morocco
Abdul Malik Al Salman	King Saud University, Saudi Arabia
Mohammed Tabili Alaoui	FS UMP, Oujda, Morocco
Almoataz Bellah Al-Said	Cairo University, Egypt
Yassine Benajiba	Thomson Reuters, IP and Science
Karim Bouzoubaa	EMI, UM5, Rabat, Morocco
Violetta Cavalli-Sforza	Al Akhawayn University, Morocco
Khalid Choukri	ELDA, European Language Resource Association, France
Oumayma Dakkak	Higher Institute for Applied Science and Technology, Syrian Arab Republic
Kareem Darwish	Qatar Computing Research Institute, Qatar
Said Desouki	Higher Institute for Applied Science and Technology, Syrian Arab Republic
Mahmoud El-Haj	Lancaster University, UK
Abdelhamid El Jihad	IERA, UM5, Rabat, Morocco
Abderrahim Elqadi	EST, UMI, Meknès, Morocco
Ramy Eskander	Columbia University, USA
Ali Farghaly	DataFlux A SAS Company, USA
Bilel Gargouri	Université de Sfax, Tunisia
Nada Ghneim	Higher Institute for Applied Science and Technology, Damascus, Syria
Ahmed Guessoum	University of Science and Technology Houari Boumediene, Algeria
Hatem Haddad	Université Libre de Bruxelles, Belgium
Kais Haddar	University of Sfax, Tunisia
Salwa Hamada	Taibah University, Madinah, Saudi Arabia
Abdelfettah Hamdani	IERA, UM5, Rabat, Morocco
Yannis Haralambous	Institut Mines-Télécom and UMR CNRS 6285 Lab-STICC, France
Hamid Jaafar	UCA, Marrakech, Morocco
Mustafa Jarrar	Birzeit University, Sina Institute, Palestine
Ali Mohamed Jaoua	Qatar University, Qatar
Mohammed Kissi	FS, UCD, El Jadida, Morocco
Abdelmonaime Lachkar	ENSA, USMBA, Fez, Morocco
Abouenour Lahsen	EMI UM5, Rabat, Morocco
Azzedine Lazrek	UCA, Marrakech, Morocco
Abdelhak Lekhouaja	UMP, Oujda, Morocco
Mourad Loukam	Hassiba Benbouali University, Algeria
Sabri A. Mahmoud	King Fahd University of Petroleum and Minerals, Saudi Arabia
Azzedine Mazroui	UMP, Oujda, Morocco
Slim Mesfar	Université de Carthage, Tunisia
Farid Meziane	University of Salford, UK

Vito Pirrelli
 Noureddine Rais
 Allan Ramsay
 Salim Rami
 Horacio Rodríguez
 Paolo Rosso
 Nasredine Semmar
 Khaled Shaalan
 Kamel Smaili
 Abderrahim Tragha
 Adnan Yahya
 Abdellah Yousfi
 Wajdi Zaghouani
 Mohamed Zakaria Kurdi
 Arssalan Zarghili
 Petr Zemánek
 Taha Zerrouki
 Ahmed ZineEddine
 Farah Benamara Zitoune
 Imed Zitouni
 Mounir Zrigui

Institute for Computational Linguistics CNR, Pisa, Italy
 FSDM, USMBA, Fez, Morocco
 University of Manchester, UK
 UCA Marrakech, Morocco
 Universitat Politècnica de Catalunya, Spain
 Universidad Politécnica de Valencia, Spain
 CEA Saclay Nano-INNOV, France
 The British University, United Arab Emirates
 University of Lorraine Nancy, France
 UH2, Casablanca, Morocco
 Birzeit University, Palestine
 Mohammed V University, Morocco
 Carnegie Mellon University, Qatar
 University of Delaware, USA
 USMBA, Fez, Morocco
 Charles University in Prague, Czech Republic
 University of Bouira, Algeria
 USMBA, Fez, Morocco
 IRT, Université Paul Sabatier, France
 Microsoft Research, USA
 University of Monastir, Tunisia

Sponsors



USMBA University

Université Sidi Mohamed Ben Abdellah
 (USMBA)



Arabic Language Engineering Society in Morocco
 ALESMA

Arabic Language Engineering Society in
 Morocco (ALESMA)



Centre National pour
la Recherche Scientifique
et Technique (CNRST-Morocco)



Académie
Hassan II
des Sciences
& Techniques

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 Scharia

Contents

Machine Translation Systems

An Automatic Approach for WordNet Enrichment Applied to Arabic WordNet	3
<i>Mohamed Seghir Hadj Ameur, Ahlem Chérifa Khadir, and Ahmed Guessoum</i>	
Word Embedding-Based Approaches for Measuring Semantic Similarity of Arabic-English Sentences	19
<i>El Moatez Billah Nagoudi, Jérémy Ferrero, Didier Schwab, and Hadda Cherroun</i>	
A POS-Based Preordering Approach for English-to-Arabic Statistical Machine Translation	34
<i>Mohamed Seghir Hadj Ameur, Ahmed Guessoum, and Farid Meziane</i>	

Speech Recognition and Synthesis

Towards a High-Quality Lemma-Based Text to Speech System for the Arabic Language	53
<i>Oumaima Zine, Abdelouafi Meziane, and Mohamed Boudchiche</i>	
Towards a Speech Recognizer for Multiple Languages Using Arabic Acoustic Model Application to Amazigh Language	67
<i>Ali Sadiqui and Ahmed Zinedine</i>	
A Game with a Purpose for Automatic Detection of Children's Speech Disabilities Using Limited Speech Resources	79
<i>Reem Elhady, Mohamed Elmahdy, Injy Hamed, and Slim Abdennadher</i>	
Building a Rich Arabic Speech and Language Corpus Based on the Holy Quran	90
<i>Ali Mefah, Yasser Seddiq, Yousef Alotaibi, and Sid-Ahmed Selouani</i>	

Text Categorization, Clustering and Summarization

Combining Words and Concepts for Automatic Arabic Text Classification . . .	105
<i>Alaa Alahmadi, Arash Joorabchi, and Abdulhussain E. Mahdi</i>	
Modern Standard Arabic Readability Prediction	120
<i>Naoual Nassiri, Abdelhak Lakhouaja, and Violetta Cavalli-Sforza</i>	

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