

Communications in Computer and Information Science

1108

Commenced Publication in 2007

Founding and Former Series Editors:

Phoebe Chen, Alfredo Cuzzocrea, Xiaoyong Du, Orhun Kara, Ting Liu,
Krishna M. Sivalingam, Dominik Ślęzak, Takashi Washio, Xiaokang Yang,
and Junsong Yuan

Editorial Board Members

Simone Diniz Junqueira Barbosa 

*Pontifical Catholic University of Rio de Janeiro (PUC-Rio),
Rio de Janeiro, Brazil*

Joaquim Filipe 

Polytechnic Institute of Setúbal, Setúbal, Portugal

Ashish Ghosh

Indian Statistical Institute, Kolkata, India

Igor Kotenko 

*St. Petersburg Institute for Informatics and Automation of the Russian
Academy of Sciences, St. Petersburg, Russia*

Lizhu Zhou

Tsinghua University, Beijing, China

More information about this series at <http://www.springer.com/series/7899>

Kamel Smaïli (Ed.)

Arabic Language Processing

From Theory to Practice

7th International Conference, ICALP 2019
Nancy, France, October 16–17, 2019
Proceedings

Editor
Kamel Smaili
University of Lorraine
Nancy, France

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-030-32958-7 ISBN 978-3-030-32959-4 (eBook)
<https://doi.org/10.1007/978-3-030-32959-4>

© Springer Nature Switzerland AG 2019

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, expressed 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.

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

Preface

This book constitutes the refereed proceedings of the 7th International Conference on Arabic Language Processing ICALP 2019 (ex-CITALA), held in Nancy, France, in October 2019. The 21 full papers presented were carefully reviewed and selected from about 40 submissions. 60% of the articles have been reviewed by three reviewers and the others by four reviewers using the double-blind review process.

The conference highlighted new approaches related to the Arabic language from basic theories to applications. All the branches of natural language processing (NLP) related to Arabic spoken or text language processing constituted the main kernel of ICALP 2019. The papers covered the following topics: modern standard Arabic, sentiment analysis and opinions mining, code-switching, deep learning, and other aspects of NLP. The volume is organized in four parts: the first is devoted to Arabic dialects and sentiment analysis, an area for which there are several challenges; the second part contains papers focusing on neural techniques for text and speech; the third part comprises papers describing different aspects of modeling modern standard Arabic; and the last one is dedicated to the resources that play an important role in NLP.

October 2019

Kamel Smaïli

Organization

General Chair

Kamel Smaïli University of Lorraine, France

Program Committee Chair

Kamel Smaïli University of Lorraine, France

Steering Committee

Karima Abidi	University of Lorraine, France
Karim Bouzoubaa	University of Mohammed V, Morocco
Olivia Brenner	University of Lorraine, France
Joseph Di Martino	University of Lorraine, France
Abdelhamid EL Jihad	University of Mohammed V, Morocco
Abdelfettah Hamdani	University of Mohammed V, Morocco
Abdelmonaime Lachkar	Université Sidi Mohamed Ben Abdellah, Morocco
David Langlois	University of Lorraine, France
Azzedine Lazrek	Université Cadi Ayyad, Morocco
Abdelhak Lekhouaja	Université Mohammed I ^{er} , Morocco
Azzedine Mazroui	Université Mohammed I ^{er} , Morocco
Mohamed-Amine Menacer	University of Lorraine, France

Program Committee

Ahmed Ali	Qatar Computing Research Institute (QCRI), Qatar
Hassina Aliane	Centre de Recherche sur l'Information Scientifique et Technique, Algeria
Frédéric Béchet	University of Aix Marseille Université, France
Almoataz Bellahal-Said	Cairo University, Egypt
Laurent Besacier	University of Grenoble, France
Karim Bouzoubaa	University of Mohammed V, Morocco
Violetta Cavalli-Sforza	Al Akhawayn University, Morocco
Khalid Choukri	European Language Resource Association (ELRA), France
Kareem Darwish	Qatar Computing Research Institute, Qatar
Joseph Di Martino	Loria, University of Lorraine, France
Mona Diab	George Washington University, USA
Abdelhamid El Jihad	University of Mohammed V, Morocco
Mahmoud El-Haj	School of Computing and Communications Lancaster University, UK

Yannick Estève	LIUM University of Maine, France
Albert Gatt	Institute of Linguistics and Language Technology, University of Malta, Malta
Nada Ghneim	Higher Institute for Applied Science and Technology, Syria
Ahmed Guessoum	University of Science and Technology Houari Boumediene, Algeria
Hatem Haddad	Université Libre de Bruxelles, Belgium
Abdelfetah Hamdani	University of Mohammed V, Morocco
Yannis Haralambous	Institut Mines-Télécom, UMR CNRS 6285 Lab-STICC, France
Salima Harrat	École Normale Supérieure Bouzaréah, Algeria
Mark Hasegawa-Johnson	University of Illinois, USA
Mustafa Jarrar	Birzeit University, Sina Institute, Palestine
Denis Jouvét	Loria, Inria, France
Abdelmonaime Lachkar	Université Sidi Mohamed Ben Abdellah, Morocco
David Langlois	Loria, University of Lorraine, France
Abdelhak Lakhouaja	Université Mohammed I ^{er} , Morocco
Azzedine Lazrek	Université Cadi Ayyad, Morocco
Yves Lepage	University of Waseda, Japan
Azzedine Mazroui	Université Mohammed I ^{er} , Morocco
Karima Meftouh	University of Annaba, Algeria
Farid Meziane	University of Salford, UK
Vito Pirreli	Institute for Computational Linguistics CNR, Italy
Violaine Prince	LIRMM, France
Allan Ramsay	School of Computer Science, University of Manchester, UK
Horacio Rodríguez	Universitat Politècnica de Catalunya, Spain
Paolo Rosso	Universidad Politècnica de Valencia, Spain
Motaz Saad	University Islamic of Gaza, Palestine
Fatiha Sadat	University of Québec, Montréal, Canada
Khaled Shaalan	The British University, UAE
Olivier Siohan	Google, USA
Kamel Smaili	Loria, University of Lorraine, France
Adnan Yahya	Birzeit University, Palestine
Abdella Yousfi	Mohammed V University, Morocco
Wajdi Zaghouani	Carnegie Mellon University, Qatar
Imed Zitouni	Microsoft Research, USA

Sponsors

Google

University of Lorraine

Loria: Laboratoire lorrain de recherche en informatique et ses applications

ELRA: European Language Resources Association

SIGUL: Special Interest Group: Under-resourced Languages

IDMC: Institut des Sciences Digitales

OLKI: Open Language and Knowledge for Citizens

ALESM: The Arabic Language Engineering Society in Morocco



Contents

Arabic Dialects and Sentiment Analysis

Generating a Lexicon for the Hijazi Dialect in Arabic	3
<i>Fatimah Abdullah Alqahtani and Mark Sanderson</i>	

Natural Arabic Language Resources for Emotion Recognition in Algerian Dialect	18
<i>Habiba Dahmani, Hussein Hussein, Burkhard Meyer-Sickendiek, and Oliver Jokisch</i>	

An Empirical Evaluation of Arabic-Specific Embeddings for Sentiment Analysis.	34
<i>Amira Barhoumi, Nathalie Camelin, Chafik Aloulou, Yannick Estève, and Lamia Hadrach Belguith</i>	

A Fine-Grained Multilingual Analysis Based on the Appraisal Theory: Application to Arabic and English Videos	49
<i>Karima Abidi, Dominique Fohr, Denis Juvet, David Langlois, Odile Mella, and Kamel Smaïli</i>	

Neural Techniques for Text and Speech

Extractive Text-Based Summarization of Arabic Videos: Issues, Approaches and Evaluations.	65
<i>Mohamed Amine Menacer, Carlos-Emiliano González-Gallardo, Karima Abidi, Dominique Fohr, Denis Juvet, David Langlois, Odile Mella, Fatiha Sadat, Juan-Manuel Torres-Moreno, and Kamel Smaïli</i>	

Automatic Identification Methods on a Corpus of Twenty Five Fine-Grained Arabic Dialects	79
<i>Salima Harrat, Karima Meftouh, Karima Abidi, and Kamel Smaïli</i>	

Aggregation of Word Embedding and Q-learning for Arabic Anaphora Resolution.	93
<i>Saoussen Mathlouthi Bouzid and Chiraz Ben Othmane Zribi</i>	

LSTM-CNN Deep Learning Model for Sentiment Analysis of Dialectal Arabic	108
<i>Kathrein Abu Kwaik, Motaz Saad, Stergios Chatzikyriakidis, and Simon Dobnik</i>	

Sentiment Analysis of Code-Switched Tunisian Dialect: Exploring RNN-Based Techniques	122
<i>Mohamed Amine Jerbi, Hadhemi Achour, and Emna Souissi</i>	

Modeling Modern Standard Arabic

Building an Extractive Arabic Text Summarization Using a Hybrid Approach	135
<i>Said Moulay Lakhdar and Mohamed Amine Chérargui</i>	

On Arabic Stop-Words: A Comprehensive List and a Dedicated Morphological Analyzer.	149
<i>Driss Namly, Karim Bouzoubaa, Rachida Tajmout, and Ali Laadimi</i>	

Negation in Standard Arabic Revisited: A Corpus-Based Metaoperational Approach	164
<i>Mohamed-Habib Kahlaoui</i>	

Skew Correction and Text Line Extraction of Arabic Historical Documents . . .	181
<i>Abdelhay Zoizou, Arsalane Zarghili, and Ilham Chaker</i>	

Authorship Attribution of Arabic Articles.	194
<i>Maha Hajja, Ahmad Yahya, and Adnan Yahya</i>	

Keyphrase Extraction from Modern Standard Arabic Texts Based on Association Rules.	209
<i>Mourad Loukam, Djamila Hammouche, Freha Mezzoudj, and Fatma Zohra Belkredim</i>	

HPSG Grammar Supporting Arabic Preference Nouns and Its TDL Specification	221
<i>Samia Ben Ismail, Sirine Boukédi, and Kais Haddar</i>	

Resources: Analysis, Disambiguation and Evaluation

“AlkhalilDWS”: An Arabic Dictionary Writing System Rich in Lexical Resources	237
<i>Mohammed Reqqass, Abdelhak Lakhouaja, and Mohamad Bebah</i>	

T-HSAB: A Tunisian Hate Speech and Abusive Dataset	251
<i>Hatem Haddad, Hala Mulki, and Asma Oueslati</i>	

Towards Automatic Normalization of the Moroccan Dialectal Arabic User Generated Text	264
<i>Ridouane Tachicart and Karim Bouzoubaa</i>	

Arabic Search Results Disambiguation: A Set of Benchmarks.	276
<i>Haytham Salhi, Radi Jarrar, and Adnan Yahya</i>	
An Arabic Corpus of Fake News: Collection, Analysis and Classification . . .	292
<i>Maysoon Alkhair, Karima Meftouh, Kamel Smaili, and Nouha Othman</i>	
Author Index	303