

Information and Communication Technologies International Symposium

DIGEST

CO-EDITORS

Mohammad Essaaidi

*Abdelmalek Essaadi University
Tetuan, Morocco*

Mohammed El Mohajir

*Sidi Mohamed Ben Abdelah University
Fez, Morocco*

Paolo Rosso

*Polytechnic University of Valencia
Valencia, Spain*

Information and Communication Technologies International Symposium

Digest

Jnan Palace Hotel, Fez, Morocco
April 3-5, 2007

Co-editors

Mohammad Essaïdi
*Abdelmalek Essaadi University
Tetuan, Morocco*

Mohammed El Mohajir
*Sidi Mohamed Ben Abdelah University
Fez, Morocco*

Paolo Rosso
*Polytechnic University of Valencia
Valencia, Spain*

Copyright © 2007 ICTIS'07

All rights reserved. This book or parts thereof may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the authors.

ISBN: 9954-8577-0-2
Fez, Morocco

Foreword.....	5
Welcome to ICTIS'07.....	6
Workshop on Arabic Natural Language Processing	7
Chairmen Biographies	8
ICTIS'07 Committees	9
Workshop on Arabic NLP Committees	10
Partners.....	11
ICTIS'07 Technical Sessions.	12
Keynote talks.....	13
Tutorial on Arabic NLP.....	17
List of Papers	18
ICTIS'07 Program.....	35

Owing to the success of its first edition ICTIS'05, which was organized in Tetuan (Morocco) in June 3-6, 2005, and the important added value it has for Information and Communication Technologies (ICT) Moroccan engineers, scientists and students, the Moroccan Association for Electrical, Electronics and Computer Engineers (MAECEE) jointly with the Institute of Electrical and Electronics Engineers (IEEE) Morocco Section have adopted this symposium as one of their major events with the main objective to contribute to the development of R&D activities and investments in all the fields related with ICT in academia, industry and government in Morocco, and by so, to contribute to promoting the input of their added value to the national economy and jobs generation. Other very important objectives set for this symposium were to offer a forum for scientists and engineers from around the world where they can meet, present and discuss their research findings. It intended also to offer them the opportunity to interact with their peers and to stimulate fruitful cooperation ties among them.

The current edition ICTIS'07 is really very rich in terms of the number of papers and the diversity of topics they cover which contributed to the very interesting and rich technical program they have allowed us to have.

We set in ICTIS'05 a tradition of organizing invited workshops in the framework of the symposium's activities by organizing the NATO Advanced Research Workshop on Information Security Assurance and Security, which I had the honor to co-chair with Prof. Johnson Thomas from Oklahoma State University in USA. In ICTIS'07, my friend Prof. Paolo Rosso, from Polytechnic University of Valencia in Spain, is thankfully organizing and chairing the Workshop on Arabic Natural Language Processing which is giving a special flavor to this edition.

One of the incentives of ICTIS is the publication of special issues of international journals and books with peer reviewed papers from those presented in this symposium. As it was done for the first edition, we are planning to publish two special issues of "International Journal of Computer Science and Applications" and "International Journal on Computer Science and Information Systems" with some papers presented in this edition.

I would like to extend my most sincere thanks and gratitude to all the keynote speakers of ICTIS'07 for their important added value to this edition and to the Scientific Committee Members who helped us in the review process. I would like also to express my thanks to the IEEE Communication Society and to IEEE Antennas and Propagation Society for their support through their Distinguished Lecturers Programs. I am also very glad to express my most sincere gratitude for the local organizing committee members, especially ICTIS'07 Local Chair, Prof. Mohammed El Mohajir and to ICTIS'07 Electronic Communications Coordinator Prof. Badr Eddine El Mohajir for their full dedication and professional organization of this edition. I will not forget the IEEE Morocco Section Secretary, Mr. Jon Raines, for his very active involvement in the different organizational tasks of this edition.

Last but not least, it is my enormous pleasure to extend all the participants of ICTIS'07 my warmest welcome to this edition wishing them a pleasant and a fruitful stay in this wonderful Moroccan imperial city of Fez.

Mohammad Essaaidi
ICTIS'07 General Chair
IEEE Morocco Section Chair



Workshop on Arabic Natural Language Processing



Salam Aleikum!

Natural Language Processing (NLP) is a subfield of artificial intelligence and linguistics: it studies the problems of automated generation and understanding of natural human languages (http://en.wikipedia.org/wiki/Natural_language_processing). Arabic is one of the six official languages of the United Nations, one of the five working languages of the African Union and spoken by more than 300 millions of people all around the world. Its characteristics - highly inflectional, complex morphology, word position-dependent form of a character, long vowels and short vowels (often not used), no capital letters etc.- makes the Arabic NLP very challenging.

The aim of this workshop is to bring together researchers interested in the area of NLP with a special emphasis on Arabic in order to provide a wide-scope forum for discussion. A total of 16 papers were finally accepted from both the Arab world – *Maghreb, Egypt, Syria, Jordan and United Arab Emirates* – as well as those countries where the Arab authors have carried out their research. Each paper was reviewed by at least two Program Committee members.

I would like to thank all those involved in this workshop. In the first place Mona Diab from the Center for Computational Learning Systems, Columbia University, who accepted with enthusiasm to give the 4-hour tutorial on *Introduction to statistical Natural Language Processing with a special emphasis on Arabic Processing* and also the keynote speech on *Advances and Challenges in Arabic Natural Language Processing*: Maghrebian M.Sc. and Ph.D. students will find out what the challenges in this research field are!

I thank the Program Committee members for accepting to review the submissions in a short time and feeding the authors with their knowledge on Arabic NLP. Very special thanks go to the authors of the papers who made this workshop become a reality. I would like to express my most cordial thanks to the ICTIS'07 General Chair, Mohammad Essaadi, and to the local Organising Committee for supporting and hosting this workshop in the imperial city of Fez, a UNESCO World Heritage Site. I express also my gratitude to Prof. Badr Eddine El Mohajir for the effective communication we had during this event.

Last but not least, I deeply appreciate that Spanish Agency for International Cooperation AECI (Agencia Española de Cooperación Internacional) believed in NLP research projects with the Arab world in general and with Morocco (PCI A/7067/06) and the Maghreb in particular, because it is through the collaboration and the (natural) *language* (processing) that we will strengthen the much we share.

Shucran, thanks, grazie, gracias to everybody for attending the Arabic NLP workshop: a gathering in the Arab NLP word, *inshallah!*

Paolo Rosso
Workshop on Arabic NLP Chair

Chair: P. Rosso, Polytechnic Univ. of Valencia, Spain

Technical Program Committee

T. Buckwalter, LDC University of Pennsylvania, USA
V. Cavalli-Sforza, San Francisco State University, USA
K. Choukri, ELDA
M. Diab, Columbia University, USA
G. Grefenstette, LIC2M/CEA-LIST
N. Habash, Columbia University, USA
A. Lyhyaou, Abdelmalek Essaadi University, Morocco
M. Maamouri, LDC University of Pennsylvania, USA
O. Rambow, Columbia University, USA
H. Rodriguez, Universitat Politècnica de Catalunya, Spain
D. Oard, University of Maryland, USA

Biography: P. Rosso,



Paolo Rosso received his Ph.D. degree in Computer Science (1999) from the Trinity College Dublin, University of Ireland. He is currently a member of the Engineering Language Processing Laboratory and a permanent lecturer at the Informatic Systems and Computation Department of the Polytechnic University of Valencia, Spain. His research interests are mainly focused on Natural Language Engineering tasks and include: knowledge-based word sense disambiguation, question answering, clustering of narrow-domain short texts, the Web as a lexical resource, and recently Arabic Natural Language Processing.

He has published over 70 papers in different conferences, workshops and journals. Paolo Rosso organised several international workshops and conferences and he was co-editor of the Proceedings of four of these events. He has also been involved in many Spanish and international research projects with Italy, Morocco, Mexico and India.

Paolo was invited to present his research work in numerous seminars globe-trotting from Latin America and Australia to the Arab World.

At the moment, he is in charge of a research project funded by the Spanish Association of International Cooperation AECI between Spain and Morocco on Arabic question answering.

The symposium accommodates a large number of technical plenary sessions addressing a wide variety of Information and Communication (I&C) topics of general interest. In order to provide an open forum and encourage participation, original papers describing significant results and developments have been solicited in a wide range of fields addressing the topics listed in the following three sessions.

Session I. Information and Communication SYSTEMS Engineering

1. Algorithms & Applications
2. Information systems and GIS
3. Expert Systems, Databases & Software Systems, Parallel & Distributed Systems
4. Virtual Reality Systems
5. Computer: Networks, Architecture, Simulation, Ethics, Security
6. Machine Learning
7. Data integration & data mining
8. Neural Networks
9. Software Engineering
10. Human Computer Interface
11. Multimedia & Visual Programming
12. Internet & Web engineering
13. Pattern Recognition
14. Reliability & Fault-Tolerance
15. Telecoms, wireless networks
16. Multimedia data processing
17. Standardization, regulatory and policy issues

Session II. Information and Communication Technologies

1. Signal and image processing
2. Digital Telecommunications
3. Optical and optoelectronic communications
4. Nanotechnology
5. RF and microwaves technologies 6. 2D-3D engineering
7. Navigation and Positioning
8. Emerging satcom systems
9. Network and Access issues
10. Communications Payload, Techniques and Equipment

Session III. Information and Communication Applications and Services

1. Telecommunication and Remote Sensing
2. Dissemination of Remote Sensing Data
3. Multimedia Communications and web services
4. Tele-Education; Tele-Medicine
5. Satellite Navigation and Positioning Services
6. Video, Sound and Data Broadcasting

7. VSAT systems
8. Personal Communications via Satellite
9. Integrated Applications and Services
10. Emergency and Disaster Monitoring, Management and Mitigation
11. The Role of Satellite Communications in the Developing World

Workshop on Arabic Natural Language Processing

Nowadays, Natural Language Processing (NLP) techniques are being widely investigated because they represent the key to develop tools able to process the huge amount of digital information available both on the Internet and the local repositories held by institutions and companies.

In comparison with other languages, the Arabic language is in a need of great investigations in the NLP techniques to provide the necessary resources to the interested researchers and to discern the relevant and appropriate techniques for this language.

Many of the computational linguistics conferences and competitions give the chance to the researchers to discuss and compare language-independent techniques. However, the Arabic language is quite different from other languages and a special focus is needed from the researchers to tackle the challenges it presents.

The final aim of this event is to bring together researchers interested in the area of the NLP in the Arabic language to foster the discussion of the future of this research area in a language which is one of the six official languages of the United Nations, one of the five working languages of the African Union and spoken by more than 300 millions of people all around the world.

Papers concerned all the areas of Natural Language Processing in the Arabic language as:

- Information Retrieval
- Information Extraction
- Question Answering
- Text Clustering
- Text Summarization
- Text Mining
- Named Entity Recognition
- Word Sense Discrimination
- Ontologies
- Lexical resources

Langley Research Center working on calibration techniques for dielectric measurements and an ionization (plasma) sensor for an experimental reentry spacecraft. In 1992 he joined the Applied Technology organization of Motorola's Paging Product Group and in 1995 he moved to Corporate Research to start an advanced modeling effort. While at Motorola he has worked on several projects from product design to measurement systems and the development of proprietary software tools for electromagnetic design. He currently manages the Microwave Technologies Research Lab within Motorola Labs in Schaumburg, IL. Recent activities of the group include high frequency communications systems design, modeling and measurements of complex electromagnetic problems, RFID systems as well as TIA standards work on RF propagation and RF exposure.

Nick is a senior member of the IEEE, and serves on a MTT Technical Program Committee. He has also recently served as chair of a TIA committee on RF exposure.

Advances and Challenges in Arabic Natural Language Processing

M. Diab, Center for Computational Learning Systems, Columbia University

Diglossia is a term rarely heard outside the field of linguistics. But Arabic computationalists should be very familiar with the term. Native speakers of Arabic are living a diglossic experience. Diglossia is in reference to a linguistic environment where people speak one form of a language but write and read a different form of that same language. In the Arabic language case, natives speak a "colloquial" vernacular, Egyptian, Palestinian, Moroccan, etc., but read and write in Modern Standard Arabic (MSA). As people working in the field of computational linguistics and are concerned with modeling the spoken and written forms of languages, we should be aware of this distinction and take it seriously. Especially with the advent of the web, we see an emergence of a special place for our so called vernaculars, we see a pervasive presence of colloquialisms in blogs, newsgroups, chats and emails. We see a ubiquity of code switching --- where languages are mixed on the lexical/syntactic levels --- between different forms of the language (intra language dialect-MSA, if we concede that the different forms of the language are all Arabics) and interlanguage where we see an interspersed use of English/French.

In this talk, the differences between the languages will be highlighted from a computational perspective that makes them relevant to our different approaches to processing Arabic, both within rule based techniques and machine learning techniques. In the process, the similarities and divergences will be illustrated with stress on real applications such as segmentation, pos tagging, parsing, speech recognition and machine translation.

Finally, the talk will stress on the importance of understanding our data and getting a good enough handle on it, in order to allow for better feature engineering techniques. This is the privilege and yet the challenge we are faced with as professionals interested in modeling Arabics.

Biography



Mona Diab received her PhD in 2003 in the Linguistics department and the University of Maryland Institute of Advanced Computer Studies (UMIACS), University of Maryland College Park. Her PhD work focused on lexical semantic issues and was titled Word Sense Disambiguation

within a Multilingual Framework.

Mona is currently an associate research scientist at the Center for Computational Learning Systems, Columbia University. Her research includes work on word sense disambiguation, automatic acquisition of natural language resources such as dictionaries and taxonomies, unsupervised learning methods, lexical semantics, cross language knowledge induction from both parallel and comparable corpora, Arabic NLP in general, tools for processing Arabic(s), computational modeling of Arabic dialects, Arabic syntactic and semantic parsing.

Dr. Diab served as co-chair - together with Kareem Darwish and Nizar Habash - of the Workshop on Computational Approaches to Semitic Languages (ACL 2005). She was also a senior member in the 2005 JHU summer workshop on Parsing Arabic Dialects. In 2005, she co-founded the Columbia Arabic Dialect Modeling (CADIM) group together with Nizar Habash and Owen Rambow.

She has published over 20 articles in different conferences, journals and workshops. Mona has presented her work in numerous lectures and tutorials both for academic and industrial audiences.

Advances in Wireless Sensor Networks

H. Mouftah, University of Ottawa, Canada

In recent years, advances in miniaturization; low-power circuit design; simple, low power, yet reasonably efficient wireless communication equipment; and improved small-scale energy supplies have combined with reduced manufacturing costs to make a new technological vision possible: Wireless sensor networks. A sensor network is composed of a large number of sensor nodes, which are densely deployed either inside the phenomenon or very close to it. The position of nodes need not be engineered or pre-determined. This allows random deployment in inaccessible terrains or disaster relief operation. We will present an overview of advances in wireless sensor networks technology and its future trends and its applications.

Biography



Hussein Mouftah joined the School of Information Technology and Engineering (SITE) of the University of Ottawa in September 2002 as a Canada Research Chair (Tier 1) Professor in Optical Networks. He has been with the Department of Electrical and Computer Engineering at Queen's University (1979-

Introduction to statistical Natural Language Processing with a special emphasis on Arabic Processing

M. Diab, Center for Computational Learning Systems, Columbia University, USA

In this tutorial, basic linguistic and probabilistic principles will be reviewed and linked together to get at the underlying notions of Statistical Natural Language Processing (NLP). Different NLP approaches will be reviewed such as rule based approaches and machine learning techniques including supervised and unsupervised methods. The approaches will be highlighted within the context of discussing a couple of NLP applications. Finally, we will discuss the different approaches to Arabic Processing.

Machine translation oriented syntactic normalization of noun phrases in Arabic <i>Khaled Elghamry</i> <i>Faculty of Al-Alsun, Ain Shams University, Cairo, Egypt</i>	C4.1
Arabic Anaphora Resolution: A Distributional, Monolingual and Bilingual Approach <i>Rania Al-Sabbagh, Khaled Elghamry</i> <i>Faculty of Al-Alsun, Ain Shams University, Cairo, Egypt</i>	C4.2
Using Text Transcriptions For Summarizing Arabic News Video <i>H. Karray, M. Ellouze, A. M. Alimi</i> <i>Research Group on Intelligent Machines, University of Sfax, Tunisia</i>	C4.3
An Arabic to English Example-Based Translation System <i>K. Bar, Y. Choueka, N. Dershowitz.</i> <i>University Mideale East, Mideale East</i>	C5.1
Implementation of the ArabiQA Question Answering System's Components <i>Y. Benajiba, P. Rosso, Abdelouahid LYHYAOUI</i> <i>Departamento de Sistemas Informaticos y Computacio,</i> <i>Universidad Politecnica de Valencia, Spain</i>	C5.2
Arabic-English Cross-Language Information Retrieval using Latent Semantic Indexing with Single-Value Matrix Decomposition <i>Ghassan Kanaan, Riyad Al-Shalabi, Abed al-Raouf K. Bsoul, Tarek Kanaan</i> <i>Amman al-Ahlyiah University, Jordan</i>	C5.3
Introduction to Arabic Speech Recognition Using CMU Sphinx System <i>H. Satori, M. Harti, N. Chenfour</i> <i>Faculty of sciences, Fez, Morocco</i>	C5.4
A Morphological Analyzer for Unvoweled Arabic Words <i>A. Boudlal, R. Belahbib, A. Lakhouaja, A. Mazroui, A. Meziane, M. Ould Abdallahi Ould Bebah</i> <i>University Mohamed I, Faculty of sciences, Oujda, Morocco</i>	C5.5
Applying Partition Around Medoids like Clustering Algorithm for Enhancing Retrieval Processes in Arabic Text using different Similarity Measures <i>Mustafa Al-Yaseen, Ghassan Kana'n, Riyad Al-Shalabi, Tarek Kanaan</i> <i>Amman al-Ahlyiah University, Amman, Jordan</i>	C5.6
Communicating in Arabic in the Cyberspace <i>Sameh Alansary, Magdy Nagi, Noha Adly</i> <i>Bibliotheca Alexandrina, Alexandria, Egypt</i>	C6.1
A Hidden-Markov-Model-based Named Entity Extractor for Arabic <i>Fatma Al Shamsi, Ahmed Guessoum</i> <i>Computer Science Department, University of Sharjah, United Arab Emirates</i>	C6.2
Approximate String Matching for Fast Arabic Speech Recognition <i>Ahmed Khorsi</i> <i>Department of Computer Science, Djillali Liabes University, Bel Abbes, Algeria</i>	C7.1
LoLo: A Tool for Extracting Statistical Information and Lexical Resources from Arabic Corpora <i>Yousif Almas</i> <i>Department of Computing, University of Surrey, United Kingdom</i>	C7.2

Arabic lexical database
S. Attar, M. Bawab, O. Al Dakkak
HIAST, Damascus, Syria..... C7.3

Morpho-syntactic tagging system for Arabic texts
A. El jihad, and A. Yousofi
Institut d'Etudes et de Recherches pour l'Arabisation, Rabat, Morocco.....C6.3