Introduction to the Special Issue on Arabic Natural Language Processing

K. SHAALAN
The British University in Dubai
and
A. FARGHALY
Monterey Institute of International Studies

OVERVIEW OF THE ARTICLES

This special issue is dedicated to the reporting of the recent Arabic natural language processing advances. A special issue of a journal allows, of course, only a partial representation of the current development in a field. However, we believe that the articles in this special issue are representative of at least some Arabic natural language processing fields. The following is a brief summary of each of the main articles in this issue.

"Arabic Natural Language Processing: Challenges and Solutions," by Ali Farghaly and Khaled Shaalan, introduces an account of challenges and solutions to significant issues in Arabic natural language processing. It covers many aspects of Arabic that are important to know by different researchers of Arabic language processing, such as Arabic diglossia, the levels (morpheme, word, syntax) of studying the Arabic language, non-concatenative morphology, and the agglutinative nature of the word structure, etc.

"Discriminative Phrase-Based Models for Arabic Machine Translation," by Cristina Espana-Bonet, Jesus Gimenez, and Lluis Marquez, presents an Arabic-to-English machine translation system that follows the phrased-based statistical translation architecture. The accuracy of the translation is increased by training the classifier on phrase selection using linguistic and

Author's address: K. Shaalan, The British University in Dubai, Alsufuh St., Knowledge Village, Block 17, Dubai United Arab Emirates (UAE), P.O. Box 502216, Dubai; email: khaled.shaalan@buid.ac.ae; A. Farghaly, Monterey Institute of International Studies, Monterey, CA 93940; email: afarghal@miis.edu.

context information. The evaluation shows an improvement at the lexical, syntactic, and semantic levels. It is also concluded that the classifier has resolved some semantic ambiguities of Arabic.

"Morphology-Based Segmentation Combination for Arabic Mention Detection," by Yassine Benajiba and Imed Zitouni, describes the importance of word segmentation as an initial step for Arabic language processing. An evaluation of Arabic mention detection models using different segmentation schemes shows a better performance, especially when Arabic Treebank and morphological segmentations are combined.

"Cross-Language Information Propagation for Arabic Mention Detection" by Imed Zitouni and Radu Florian, presents an approach that tries to overcome the unavailability of Arabic linguistic resources for some applications, such as mention detection, by propagating information from a resource-rich language. The approach is applied using a mention detection system and statistical machine translation system that translates text from English to Arabic. The result of experiments shows improvement in the Arabic mention detection system performance.

"Automatic Speech-to-Text Transcription in Arabic," by Lori Lamel, Abdelkhalek Messaoudi, and Jean-Luc Gauvain Limsi-Cnrs, reports on research carried out over the last few years on the incremental improvements to a system for the automatic speech-to-text transcription of broadcast data in Arabic. Arabic texts are written without diacritics, yet the diacritics provide useful information for pronunciation modeling and higher level processing. So rules to generate pronunciations with a generic vowel have been proposed, and this method has been used to significantly facilitate training on nonvocalized data. The results show that the explicit modeling of gemination and the introduction of pronunciation variants led to significant improvements in speech-to-text transcription performance.

"Sura Length and Lexical Probability Estimation in Cluster Analysis of the Qur'an," by Hermann Moisl, addresses the problem of clustering of the shorter suras in order to generate their accurate classifications. It proposes a solution of the problems found in a previous work to the reanalysis of the Qur'an.

ACKNOWLEDGMENTS

We are very pleased with the response we received to our call for contributions from the research community in the field. We received 29 submissions to this special issue and accepted five. We would like to thank all those who submitted articles for this special issue. Space and time limitations caused the exclusion of some good articles.

We would like to thank the following reviewers for their valuable help in producing this special issue. Without their help and contribution, the present issue would not have been possible:

Sherif Abdou (Cairo University, Egypt); Azza Abd El Moem (Ain Shams University, Egypt); Mohammed Abdel-Aal Attia (Al-Azhar University, Egypt); Galia Angelova (Bulgarian Academy of Sciences, Bulgaria); Mohamed Attia (The Engineering Company for the Development of Computer Systems,