

Contents

Papers

<i>Computer Processing of Arabic Script-based Languages: Current State and Future Directions</i> Ali Farghaly	1
<i>Developing an Arabic Treebank: Methods, Guidelines, Procedures, and Tools</i> Mohamed Maamouri and Ann Bies	2
<i>Preliminary Lexical Framework for English-Arabic Semantic Resource Construction</i> Anne R. Diekema	10
<i>The Architecture of a Standard Arabic Lexical Database: Some Figures, Ratios and Categories from the DIINAR.1 Source Program</i> Ramzi Abbès, Joseph Dichy and Mohamed Hassoun	15
<i>Systematic Verb Stem Generation for Arabic</i> Jim Yaghi and Sane Yagi	23
<i>Issues in Arabic Orthography and Morphology Analysis</i> Tim Buckwalter	31
<i>Finite-State Morphological Analysis of Persian</i> Karine Megerdooian	35
<i>Arabic Script-Based Languages Deserve to be Studied Linguistically</i> Martin Kay	42
<i>An Unsupervised Approach for Bootstrapping Arabic Sense Tagging</i> Mona T. Diab	43
<i>Automatic Arabic Document Categorization Based on the Naïve Bayes Algorithm</i> Mohamed El Kourdi, Amine Bensaïd and Tadjeddine Rachidi	51
<i>A Transcription Scheme for Languages Employing the Arabic Script Motivated by Speech Processing Applications</i> Shadi Ganjavi, Panayiotis G. Georgiou and Shrikanth Narayanan	59
<i>Automatic Diacritization of Arabic for Acoustic Modeling in Speech Recognition</i> Dimitra Vergyri and Katrin Kirchhoff	66
<i>Letter-to-Sound Conversion for Urdu Text-to-Speech System</i> Sarmad Hussain	74

Demonstrations

<i>Urdu Localization Project</i> Sarmad Hussain	80
<i>FarsiSum – A Persian Text Summarizer</i> Martin Hassel and Nima Mazdak	82
<i>Stemming the Qur'an</i> Naglaa Thabet	85
<i>Language Weaver Arabic->English MT</i> Daniel Marcu, Alex Fraser, William Wong and Kevin Knight	89