

# WORKSHOP PROGRAM

## OPENING AND OVERVIEW

- 8:30 – 9:00 *Computer Processing of Arabic Script-based Languages: Current State and Future Directions*  
Ali Farghaly

## SESSION 1: LEXICON AND CORPORA

- 9:00 – 9:30 *Developing an Arabic Treebank: Methods, Guidelines, Procedures, and Tools*  
Mohamed Maamouri and Ann Bies
- 9:30 – 10:00 *Preliminary Lexical Framework for English-Arabic Semantic Resource Construction*  
Anne R. Diekema
- 10:00 – 10:30 *The Architecture of a Standard Arabic Lexical Database: Some Figures, Ratios and Categories from the DIINAR.1 Source Program*  
Ramzi Abbès, Joseph Dichy and Mohamed Hassoun

10:30 – 10:45 BREAK

## SESSION 2: MORPHOLOGY

- 10:45 – 11:15 *Systematic Verb Stem Generation for Arabic*  
Jim Yaghi and Sane Yagi
- 11:15 – 11:45 *Issues in Arabic Orthography and Morphology Analysis*  
Tim Buckwalter
- 11:45 – 12:15 *Finite-State Morphological Analysis of Persian*  
Karine Megerdooian

12:15 – 2:00 LUNCH & DEMO SESSIONS

## DEMONSTRATIONS

*Urdu Localization Project*  
Sarmad Hussain

*FarsiSum – A Persian Text Summarizer*  
Martin Hassel and Nima Mazdak

*Stemming the Qur'an*  
Naglaa Thabet

*Language Weaver Arabic->English MT*  
Daniel Marcu, Alex Fraser, William Wong and Kevin Knight

---

**INVITED SPEAKER**

2:00 – 2:45 *Arabic Script-Based Languages Deserve to be Studied Linguistically*  
Martin Kay

**SESSION 3: STATISTICAL APPROACHES**

2:45 – 3:15 *An Unsupervised Approach for Bootstrapping Arabic Sense Tagging*  
Mona T. Diab

3:15 – 3:45 *Automatic Arabic Document Categorization Based on the Naïve Bayes Algorithm*  
Mohamed El Kourdi, Amine Bensaïd and Tajje-eddine Rachidi

---

3:45 – 4:00      **BREAK**

**SESSION 4: SPEECH PROCESSING**

4:00 – 4:30 *A Transcription Scheme for Languages Employing the Arabic Script Motivated  
by Speech Processing Applications*  
Shadi Ganjavi, Panayiotis G. Georgiou and Shrikanth Narayanan

4:30 – 5:00 *Automatic Diacritization of Arabic for Acoustic Modeling in Speech Recognition*  
Dimitra Vergyri and Katrin Kirchhoff

5:00 – 5:30 *Letter-to-Sound Conversion for Urdu Text-to-Speech System*  
Sarmad Hussain

5:30 – 6:00 Discussion and Closing  
Ali Farghaly and Karine Megerdoomian