

Introduction to Arabic Natural Language Processing

Synthesis Lectures on Human Language Technologies

2010, 187 pages, (doi:10.2200/S00277ED1V01Y201008HLT010)

Nizar Y. Habash

Columbia University

Abstract

This book provides system developers and researchers in natural language processing and computational linguistics with the necessary background information for working with the Arabic language. The goal is to introduce Arabic linguistic phenomena and review the state-of-the-art in Arabic processing. The book discusses Arabic script, phonology, orthography, morphology, syntax and semantics, with a final chapter on machine translation issues. The chapter sizes correspond more or less to what is linguistically distinctive about Arabic, with morphology getting the lion's share, followed by Arabic script. No previous knowledge of Arabic is needed. This book is designed for computer scientists and linguists alike. The focus of the book is on Modern Standard Arabic; however, notes on practical issues related to Arabic dialects and languages written in the Arabic script are presented in different chapters.

Table of Contents: What is "Arabic"? / Arabic Script / Arabic Phonology and Orthography / Arabic Morphology / Computational Morphology Tasks / Arabic Syntax / A Note on Arabic Semantics / A Note on Arabic and Machine Translation