Computational Linguistics (CS-493)

Instructor: Qaiser Abbas
Credit Hours: 3(3-0)
Spring 2010, BSCS, 8th Semester, Regular & Self Support
Department of Computer Science, University of Sargodha
Sargodha, 40100, Pakistan.
Email: qaiser.abbas@uos.edu.pk,

Objectives

The course is aimed to develop an understanding of linguistics and how it may be modeled and processed. The course will focus on modeling words and phrases, with some reference to higher structures, including meaning and discourse. Challenges associated with multilingual text processing and their solutions will also be addressed.

Pre-requisites

Theory of Automata/Computation

Textbook

Speech and Language Processing, Second Edition, by Daniel Jurafsky and James Martin *Additional reading material will be provided for sections in italics.*

Course Outline

Corpora	*Scripts*
	Unicode Encoding and Processing
	Normalization and Collation
	Tokenization
Words	Spell Checking
	Morphology and Finite State Transducers
	N-Grams
	Word Segmentation
Phrases	Word classes and POS tagging
	Chunking
	CFG and Language Grammars
	Rule-Based and Probabilistic Parsing of CFGs
	Features and Unifications
	Annotated Grammars and Lexical Functional Grammar
Semantics	Lexical Semantics
	Compositional Semantics
	Word Sense Disambiguation
Discourse	

Grading Policy:

Mid Term:	30%
Final Term:	20%+30% (Final exam + Term paper)
Quiz/Assignments:	20%

*In case of plagiarism and cheating, F grade will be awarded.