COURSE DESCRIPTION FORM

INSTITUTION Department of Computer Science & IT, University Of Sargodha

PROGRAM (S) TO BE EVALUATED For MS and PhD Level

A. Course Description

Course Code	CS-5840
Course Title	Natural Language Processing
Credit Hours	3 CR
Prerequisites by Course(s) and Topics	Math and Programming Skills
Assessment Instruments with Weights (homework, quizzes, midterms, final, programming assignments, lab work, etc.)	Quizzes and Homework: 10% Midterm: 30%, Final Term: 50 % Term Paper: 10%
Course Coordinator	Dr. Qaiser Abbas
URL (if any)	http://www.clsp.org/qabbas/nlp.html
Current Catalog Description	Not Available as per curriculum
Textbook (or **Laboratory Manual** for Laboratory Courses)	Daniel Jurafsky and James H. Martin. 2008. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Second/Third Edition. Prentice Hall.
Reference Material	• Christopher D. Manning and Hinrich Schütze. 1999. Foundations of Statistical Natural Language Processing. MIT Press. • Steven Bird. 2009. Natural Language Processing with Python. O'Reilly. (Free on SearchWorks) • Philipp Koehn. 2010. Statistical Machine Translation. Cambridge. • Yoshua Bengio. 2009. Learning Deep Architectures for AI. Technical Report. (Free from Stanford network) • Frederick Jelinek. 1998. Statistical Methods for Speech Recognition. MIT Press. • James Allen. 1995. Natural Language Understanding.

	Benjamin/Cummings, 2ed. • Psycholinguistics: Tanenhaus & Trueswell (2006), Human Sentence Processing website
Course Goals	This course covers a broad range of topics in computational linguistics/natural language processing, including word and sentence tokenization, text classification and sentiment analysis, spelling correction, information extraction, parsing, meaning extraction, and question answering, We will also introduce the underlying theory from probability, statistics, and machine learning that are crucial for the field, and cover fundamental algorithms like n-gram language modeling, naive bayes and maxent classifiers, sequence models like Hidden Markov Models, probabilistic dependency and constituent parsing, and vector-space models of meaning.
Topics Covered in the Course, with Number of Lectures on Each Topic (assume 15-week instruction and one-hour lectures)	1. Introduction [TB1, Ch. 1: Week 1] 2. Regular Expressions, Text Normalization, and Edit Distance: Regular Expressions, Words, Corpora, Text Normalization, Minimum Edit Distance, etc. [TB1, Ch. 2: Week 2-3] 3. N-gram Language Models: N Grams, Evaluating Language Models, Generalization and Zeros, Smoothing, etc. [TB1, Ch.3: Week 4-5] 4. Naive Bayes and Sentiment Classification: Naïve Bayes Classifiers and their training with examples, Its optimization for sentiment analysis and for other text classification, etc. [TB1, Ch.4: Week 6] 5. Logistic Regression: Classification: The sigmoid; Learning in Logistics Regression; The Cross Entropy loss function; Gradient Descent, Regularization, etc. [TB1, Ch. 5: Week 7] 6. Vector Semantics: Lexical Semantics, Vector Semantics, Words and Vectors, Cosine for measuring similarity, TF-IDF Weighting terms in the vector, etc. [TB1, Ch. 6: Week 8-9] 7. Neural Networks and Neural Language Models: Units, The XOR problem, Feed Forward Neural Network, Training Neural Nets, etc. [TB1, Ch. 7: Week 10-11] 8. Part-of-Speech Tagging: Word Classes, Penn POS Tagset, POS Tagging, HMM POS Tagging, etc. [TB1, Ch. 8: Week 12] 9. Sequence Processing with Recurrent Networks: Simple Recurrent Networks, Applications of RNNs, Deep Networks, Managing Context in RNNs, etc. [TB1, Ch. 9: Week 13-14] 10. Formal Grammars of English/Urdu: Constituency, Context

	Free Grammars, Rules of English/Urdu, Treebanks, etc. [TB1, Ch. 10: Week 15] 11. Syntactic Parsing: Ambiguity, CKY Parsing, Partial Parsing, etc. [TB1, Ch. 11: Week 16]			
Laboratory Projects/Experiments Done in the Course	Through assignments and final project			
Programming Assignments Done in the Course	Through assignments and final project			
Class Time Spent on (in credit hours)	**Theory**	**Problem Analysis**	**Solution Design**	**Social and Ethical Issues**
	24 Hrs	12 Hrs	12 hrs	Not Applicable
Oral and Written Communications	Students are required to submit at least 3 written reports in the form of assignments. Presentations in the form of groups would be arranged in the last week if course would be covered before the 16th week.			

Instructor Name **Dr. Qaiser Abbas**

Instructor Signature _____

Date _____