Course title: Natural Language Processing (1902736) معالجة اللغات الطبيعية

Credit Hours: 3 Credit hours
Prerequisite: Artificial Intelligence Course and Good Programming Skills (Java, Python or Perl)
Course Lecturer: Dr. Majdi Sawalha

Office Location	Office Phone	Office hours	E-mail
KASIT, CIS Dept	Ext. 22619	TBA	sawalha.majdi@ju.edu.jo

Course Description:

Origins of Natural Language Processing (NLP); Language structure representation; The role of knowledge; Knowledge representation; Parsing techniques; Finite-state techniques; Recursive and augmented transition networks; Language ambiguity; Well-Formed constructs; Features and the lexicon; Language semantics; Applications.

Intended Learning Outcomes:

Successful completion of this course enables a student to:

A- Knowledge and Understanding:

A1) Know and Understand the basic concepts of Natural Language Processing
A2) Understand some problems and solutions of NLP, and their relation to linguistics and statistics.
A3) Understand Natural Language Processing (Morphology, Syntax and Semantics).
A4) Understand the language grammar and context-free parsing techniques.
A5) Understand the principles of Information Retrieval and Speech Recognition Systems.
A6) Understand the principles of Question Answering Systems.

B- Cognitive and Intellectual skills: with the ability to …

B1) Recognize NL processing techniques and how they are applied to NL applications.
B2) Design set of tools to support a NL system.
B3) Recognize and employ analytical skills to solve problems.

C- Subject specific skills – with ability to …

C1) Analyze a NL problem and develop a NL tools to solve it.
C2) Evaluate different NLP systems and propose solutions for NL problems.
C3) Discuss different NL applications.

D- Transferable skills – with ability to

D1) Work individually and within a group to design a solution for a NL problem (IR system, IE system, QA system, Text summarization system).

D2) Work effectively to implement a Java/Python program that implement the NL application under study.

D3) Interact with the other study groups to make use of different implementations and NL tools.

D4) Present the final project and make a demo.

<u>**Weekly Course Contents, and T/L & A with ILOs**</u>

Week	Part	Topic Details	Teaching/Learning and Assessment Methods	ILOs
1	**0- Introduction**	Knowledge in speech and language processing; Ambiguity; Models and Algorithms; Language, Thoughts, and Understanding; The State of the Art; Some brief history	T: Lecture L: Reading lecture notes and Chap 1 A: in Class questions	A1, A2
2	**I - Words**	**Regular expressions and Automata:** Regular expressions, Finite-state Automata, Regular languages and FSAs	T: Lecture L: Reading lecture notes and Chap 2 A: Programming Assignment 1 on regular expressions	A2, B1, D1
3+4		**Words and Transducers:** Finite-State Transducers; FSTs for Morphological Parsing; Transducers and Orthographic Rules; The Combination of FST Lexicon and Rules; Lexicon-Free FSTs: The Potter Stemmer; Survey of English Morphology; Finite-State Morphological Parsing; Construction of a Finite-State Lexicon; Word and Sentence Tokenization	T: Lecture L: Reading lecture notes and Chap 3 A: Programming Assignment 2 implementation of Potter Stemmer/ word and sentence tokenization for Arabic	A3, B2, C1, D1, D2
5		**N-Grams:** Counting Words in Corpora; Simple (Unsmoothed)	T: Lecture L: Reading lecture notes and Chap 3	A3, B2, C1, D1, D2

		N-grams; More on N-grams and their sensitivity to the training corpus; Smoothing ; N-grams for Spelling and Pronunciation; Context-Sensitive Spelling Error Correction; N-grams for Pronunciation Modeling	A: Programming Assignment 3 implementation of spelling checker and corpora analysis	
6+7	**Programming**	Introduction to Language Processing using **Python**	T: In Lab exercises L: learning programming skills A: Post Lab Quizzes	B2, B3, D1, D2
		Midterm exam	A: Written exam on materials in Chaps: 1-4 & programming skills for NLP with Python	
8+9		**Word Classes and Part-of-Speech Tagging:** (Mostly) English Word Classes; Tagsets for English; Part of Speech Tagging; Rule-based Part-of-speech Tagging; Stochastic Part-of-speech Tagging; A Motivating Example; The Actual Algorithm for HMM tagging	T: Lecture L: Reading lecture notes and Chap 5 A: Programming Assignment 4 implementation of HMM tagging algorithm	A3, B2, C1, C2, D1, D2
10	**II Syntax**	**Context-Free Grammars for English:** Constituency; Context-Free Rules and Trees; Sentence-Level Constructions; The Noun Phrase; Before the Head Noun; After the Noun; Coordination; Agreement; The Verb Phrase and Subcategorization ; Auxiliaries; Spoken	T: Lecture L: Reading lecture notes and Chap 12 A: Assignment 5 Survey of Arabic grammar and the Dialects of Arabic	A3, A4, B1, D1

		Language Syntax; Disfluencies; Grammar Equivalence & Normal Form; Finite State & Context-Free Grammars; Grammars & Human Processing		
11+12		**Parsing with Context-Free Grammars:** Parsing as Search; Top-Down Parsing; Bottom-Up Parsing; Comparing Top-down and Bottom-up Parsing; The Earley Algorithm; Finite-State Parsing Methods .	T: Lecture L: Reading lecture notes and Chap 13 A: Programming Assignment 6 implementation of the Earley algorithm	A3, A4, B1, B2, C1, C2, D1, D2
13	**III Semantics**	**Representing Meaning:** Computational Desiderata for Representations; Meaning Structure of Language; First Order Predicate Calculus; Some Linguistically Relevant Concepts; Related Representational Approaches; **Lexical Semantics:** Relations Among Lexemes and Their Senses; WordNet: A Database of Lexical Relations; The Internal Structure of Words; Creativity and the Lexicon	T: Lecture L: Reading lecture notes and Chap 17 & 16 A: Assignment 7 Survey of Arabic WordNet and Semantic lexicons	A3, A4, B1, B2, C1, C2, C3, D1, D2
14	**Applications**	**Information Extraction and Question Answering and summarization**	T: Lecture L: Reading lecture notes and Chap 22 & 23 A: Assignment 8 Survey of Arabic IR and QA systems	A5, A6, B3, C3, D1, D2
15		**Review**	T: Summary L: Demo A: Sample of exams	D4
16		**Final Exam**		

Teaching (T) Strategies: Class Contact is 3 Hours per week. The Course will be delivered using different means like lecture, presentations, seminars, discussion and case studies.

Learning (L) Methods: Students attend classes, ask questions and participate in discussions, do the home works, present the assignments and demo their works. A student will use the lab and select a programming language to implement the assignments. Students will access the e-learning platform for more instruction and supported learning materials.

Assessment (A) Methods: There will be several assessment methods of evaluation the performance of the students such as attending and class participation, grading the homework, quizzes and assignments; conducting the Midterm and the Final Exams. Every student is expected to completely adhere to the assignments and project strict deadlines, absolutely no exceptions will be given.

Assessment Weights:

Home works and quizzes	10%
Assignments and Demo	15%
Midterm	25%
Final Exam	50%

Satisfactory completion of this subject requires a 50% pass in the end-of-semester examination.

Grading Scale:

0-49	F
50-53	D-
54-57	D
58-61	D+
62-66	C-
67-71	C
72-76	C+
77-81	B-
82-86	B
87-89	B+
90-93	A-
94-100	A

Text book (TB):

Daniel Jurafsky and James H. Martin (2008) *"Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition"*, 2nd Edition, Prentice Hall.

Steven Bird, Ewan Klein, and Edward Loper (2009) *"Natural Language Processing with Python"*, 1st Edition, O'REILLY.

Ethics:

The honor code applies to all work turned in for this course including exams and assignments. It is important that you understand the solutions to all problems, and the best way to gain an understanding is to work them out and write them up by yourself. Hence the policy is that <u>you must submit your own work and clearly list your references</u>. You may not share your work with other students, unless it is allowed as group. Violating the policy will be taken as a no submission state for the assignment. University regulations will be preserved at all times.