1	**Course title**	Natural Language Processing
2	**Course number**	1905380
3	**Credit hours**	3
	Contact hours (theory, practical)	3
4	**Prerequisites/corequisites**	1905320
5	**Program title**	Artificial Intelligence
6	**Program code**	1905
7	**Awarding institution**	The University of Jordan
8	**School**	King Abdullah II School of Information Technology
9	**Department**	Department of Artificial Intelligence
10	**Level of course**	Bsc
11	**Year of study and semester (s)**	2022-2023, Second Semester
12	**Final Qualification**	Bsc
13	**Other department (s) involved in teaching the course**	-
14	**Language of Instruction**	English
15	**Teaching methodology**	☐Blended　　☐Online　☒ In classroom (Lab)
16	**Electronic platform(s)**	☒Moodle　☒Microsoft Teams ☐Skype　☐Zoom ☐Others…………
17	**Date of production/revision**	09/02/2023

18 Course Coordinator:

Office numbers, office hours, phone numbers, and email addresses should be listed.
Office number: 221
Office hours: 12:30 pm - 1:30 pm: Sun., Tues., Thur.
phone number: 22619
email addresses: sawalha.majdi@ju.edu.jo

19 Other instructors:

Name:
Office number:
Phone number:
Email:

20 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.

21 Course aims and outcomes:

A- Aims:

The aim of the course is to introduce students to the concepts of Natural Language processing and its applications. It discusses the linguistic theories and computational techniques for machine translation, information retrieval, text summarization, reference resolution, question answering, parsing, sense disambiguation, morphological analysis, speech analysis and synthesis.

B- Intended Learning Outcomes (ILOs):
Upon successful completion of this course, students will be able to:

OA-Knowledge and Understanding (students should)
   (OA1) have some understanding of the basic concepts of Natural Language Processing.
   (OA2) have some understanding of the problems and solutions of NLP, and their relation to linguistics and statistics.
   (OA3) have some understanding of the Natural Language Processing (Morphology, Syntax and Semantics).
   (OA4) have some understanding of the language grammar and context-free parsing techniques.
   (OA5) have some understanding of the principles of Information Retrieval and Speech Recognition Systems.
   (OA6) have some understanding of the principles of Question Answering Systems.

OB-Intellectual skills-with ability to
   (OB1) Recognize NL processing techniques and how they are applied to NL applications.
   (OB2) Design set of tools to support a NL system.
   (OB3) Decide the suitability of innovative methods for developing business plans.
   (OB4) Recognize and employ analytical skills to solve problems.

OC- Practical Skills-With ability to
   (OC1) Analyze a NL problem and develop a NL tools to solve it.
   (OC2) Evaluate different NLP systems and propose solutions for NL problems.
   (OC3) Discuss different NL applications.

OD-Transferable Skills-With ability to
   (OD1) Work individually and within a group to design a solution for a NL problem (IR system, IE system, QA system, Text summarization system).
   (OD2) Work effectively to implement a Java/Python program that implement the NL application under study.
   (OD3) Interact with the other study groups to make use of different implementations and NL tools.
   (OD4) effectively present the final work in a demo.

22. Topic Outline and Schedule:

Week	Topic	ILOs	Program SOs[1]	TLA (teaching, learning and Assessment)
1 & 2	**Induction and Welcome Introduction** - Knowledge in speech and language processing; - Ambiguity; - Models and Algorithms; - Language, Thoughts, and Understanding; - The State of the Art; - Some brief history	OA1, OA2	1, 2	T: Lecture L: Reading lecture notes and Chap 1 A: in Class questions
3	**Regular expressions and Automata:** - Regular expressions, - Finite-state Automata, - Regular languages and FSAs	OA2, OB1, OD1	1, 2	T: Lecture L: Reading lecture notes and Chap 2 A: Programming Assignment 1 on regular expressions
4 & 5	**Words and Transducers:** - 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	OA3, OB2, OC1, OD1, OD2	1, 2	T: Lecture L: Reading lecture notes and Chap 3 A: Programming Assignment 2 implementation of Potter Stemmer/ word and sentence tokenization for Arabic
6	**N-Grams:** - Counting Words in Corpora; - Simple (Unsmoothed) N-grams; - More on N-grams and their sensitivity to the training corpus;		1, 2	T: Lecture L: Reading lecture notes and Chap 3 A: Programming Assignment 3 implementation of spelling checker and corpora analysis

---

[1] The ABET outcomes

	– Smoothing; – N-grams for Spelling and Pronunciation; – Context-Sensitive Spelling Error Correction; – N-grams for Pronunciation Modeling			
7 & 8	– **Programming** Introduction to Language Processing using **Python and NLTK (Natural Language Processing)** – **Midterm exam**	OB2, OB3, OD1, OD2	1, 2	T: In Lab exercises L: Chapters 1-4 book 2 A: Post Lab Quizzes
9 & 10	**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	OA3, OB2, OC1, OC2, OD1, OD2	1, 2	T: Lecture L: Reading lecture notes and Chap 6 A: Programming Assignment 4 implementation of HMM tagging algorithm

11	**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 Language Syntax; – Disfluencies; – Grammar Equivalence & Normal Form; – Finite State & Context-Free Grammars; – Grammars & Human Processing	OA3, OA4, OB1, OD1	1, 2	T: Lecture L: Reading lecture notes and Chap 9 A: Assignment 5 Survey of Arabic grammar and the Dialects of Arabic
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 .	OA3, OA4, OB1, OB2, OC1, OC2, OD1, OD2	1, 2	T: Lecture L: Reading lecture notes and Chap 10 A: Programming Assignment 6 implementation of the Earley algorithm
13 & 14	**Representing Meaning:** – Computational Desiderata for Representations; – Meaning Structure of Language; – First Order Predicate Calculus; – Some Linguistically Relevant Concepts;	OA3, OA4, OB1, OB2, OC1, OC2, OC3, OD1, OD2	1, 2	T: Lecture L: Reading lecture notes and Chap 14 & 16 A: Assignment 7 Survey of Arabic WordNet and Semantic lexicons

	– 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			
15	**Information Retrieval and Question Answering systems**	OA5, OA6, OB3, OC3, OD1, OD2	1, 2	T: Lecture L: Reading lecture notes and Chap 14 & 16 A: Assignment 8 Survey of Arabic IR and QA systems
16	**Revision and Presentations**	OD1, OD4	1, 2	Presentations

- Teaching methods include: Synchronous lecturing/meeting; Asynchronous lecturing/meeting
- Evaluation methods include: Homework, Quiz, Exam, pre-lab quiz…etc

23 Evaluation Methods:

Opportunities to demonstrate achievement of the ILOs are provided through the following assessment methods and requirements:

Evaluation Activity	Mark	Topic(s)	Period (Week)	Platform
Programming Assignment 1	2	Regular expressions	1	Moodle
Programming Assignment 2	3	Potter Stemmer/ word and sentence tokenization for Arabic	1	Moodle
Programming Assignment 3	2	Spelling checker and corpora analysis	1	Moodle
Programming Assignment 4	3	HMM tagging algorithm	1	Moodle
Assignment 5	2	Syntax and grammar	1	Moodle
Programming Assignment 6	3	the Earley algorithm	1	Moodle
Assignment 7	3	WordNet and Semantic lexicons	1	Moodle
Assignment 8	2	Arabic IR and QA systems	1	Moodle
	20			

24 Course Requirements (e.g: students should have a computer, internet connection, webcam, account on a specific software/platform…etc):

Student should have:
- Hardware:
  - Computer
  - Internet connection
  - Microphone & speakers
- Software
  - Moodle account
  - MS Teams account

25 Course Policies:

A- Attendance policies:

University Regulations

B- Absences from exams and handing in assignments on time:

University Regulations

C- Health and safety procedures:

University Regulations

D- Honesty policy regarding cheating, plagiarism, misbehavior:

University Regulations

E- Grading policy + Weighting (i.e. weight assigned to exams as well as other student work)

Mid term exam 30%

Homeworks and quizzes: 20%

Final exam: 50%

F- Available university services that support achievement in the course:

Computer laboratories, data shows and JU e-learning system

G- Statement on Students with disabilities

Students with Disabilities: Students with disabilities who need special accommodations for this class are encouraged to meet with the instructor and/or their academic advisor as soon as possible. In order to receive accommodations for academic work in this course, students must inform the course instructor and/or their academic advisor, preferably in a written format, about their needs no later than the 4th week of classes.

26 References:

A- Required book (s), assigned reading and audio-visuals:

[1] Dan Jurafsky and James H. Martin (2020) "Speech and Language Processing" 3rd edition

B- Recommended books, materials, and media:

[2] دليل أبحاث" (2019) محمد زكي خضـر، محمد السعودي، مجدي صوالحة، سامي عبابنة، يوسف حمدان، مأمون حطاب
حوسبة اللغة العربية"، اللجنة الوطنية للنهوض باللغة العربية، مجمع اللغة العربية الأردني، الطبعة الأولى، عمّان – الأردن.

[3] J. Pustejovsky, A. Stubbs (2012) "Natural Language Annotation for Machine Learning", First Edition, O'Reilly.

[4] Steven Bird, Ewan Klein, Edward Loper (2009) "Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit" 1st Edition, O'Reilly.

[5] other extra materials provided by instructors on e-learning system. http://e-learning.ju.edu.jo

27 Additional information:

1. Supplementary notes are made available of the e-learning (Moodel) system.
2. Students are encouraged to make use of JU library, E-LIBRARY:
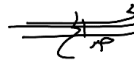   access within the university: http://e-library
   access from outside: http://ezlibrary.ju.edu.jo/login

Some important/relevant journals include:
(1) International Journal of Transitions and Innovation Systems
(2) International Journal of Innovative Information Systems and Technology Research

Name of Course Coordinator: Dr. Majdi Sawalha Signature: Date: 09.02.2023

Head of Curriculum Committee/Department: -------------------------- Signature: ------------------------

Head of Department: -- Signature: ---------------------

Head of Curriculum Committee/Faculty: ------------------------------------- Signature: ------------------

Dean: -- Signature: --

QF-AQAC-03.02.1.3