

Natural Language Processing

Course Outline

COMP9312

Spring 2024/2025

Instructors:

Mustafa Jarrar
Mohammad Khalilia

Text book:

1. Lecture Notes and slides given by the instructors

Recommended References:

1. Speech and Language Processing. Third Edition. Daniel Jurafsky and James Martin.
2. Natural Language Processing for Corpus Linguistics. Jonathan Dunn
3. Language Processing with Perl and Prolog. Pierre M. Nugues
4. Jarrar, M. (2021). The Arabic Ontology - An Arabic Wordnet with Ontologically Clean Content. Applied Ontology Journal, 16:1, 1-26. IOS Press.
5. Jarrar, M., Khalilia, M., Ghanem, S., (2022): Wojood: Nested Arabic Named Entity Corpus and Recognition using BERT. Proceedings of the 13th International Conference on Language Resources and Evaluation (LREC'22) (Submitted).
6. Al-Hajj, M., Jarrar, M., (2021). ArabGlossBERT: Fine-Tuning BERT on Context-Gloss Pairs for WSD. In Proceedings – the International Conference on Recent Advances in Natural Language Processing (RANLP 2021), PP 40--48.

Course Description:

Applied and state-of-the-art topics in natural language processing. Language encoding, orthography, morphology, syntax, text processing, probabilistic language modeling, text classification, part-of-speech tagging, and lexical semantics, autocomplete, morphological analyses, named entity recognition, chatbots, smart search, sentiment analysis, and event discovery.

معارف ومهارات متقدمة لمعالجة وفهم اللغات الطبيعية، تشمل الترميز اللغوي، والإملاء، والتصريف، والقواعد، والدلالة، ومعالجة النصوص وتصنيفها، والنمذجة اللغوية الإحصائية، وتوصيف الجمل، والتطبيقات الحديثة، مثل التصريف الآلي، واكتشاف الكينونات، وبرامج الدردشة الآلية، واكتشاف الأحداث

Objectives:

- In-depth understanding of the theoretical foundations of, and how to, parse and understand, natural language.
- Demonstrate how natural language can be processed, understood, and used in IT applications.
- Design and develop algorithms and systems that are able to understand natural language.
- Use relevant NLP tools and resources.

- فهم الخلفيات العلمية والأسس النظرية لفهم ومعالجة اللغات الطبيعية بطريقة آية.
- معرفة بكيفية معالجة وفهم اللغات الطبيعية في البرمجيات.
- تحليل وتطوير برمجيات وخوارزميات قادرة على فهم النصوص اللغوية.
- استخدام أدوات ومصادر لغوية متاحة.

Evaluation:

- Assignments (35%)
- Midterm project (15%)
- Final project (30%)
- Final paper (20%)

Schedule:

Chapter	Lectures
Introduction	0.5
Encoding and Orthography	0.5
Morphology	0.5
Syntax	0.5
Semantics	2
Corpora and Annotations	0.5
Evaluation Metrics	0.5
Information Retrieval	2
Machine Learning Basics	3
Word Vectors and Language Models (m-gram, TF-IDF, word2vec)	1
Classical Sequence Modeling - RNN and LSTM	1
Neural Sequence Modeling – Transformers (BERT, etc.)	1
NLP libraries (e.g., PyTorch)	1
NLP Tasks (NER, WSD, event extraction, text classification, etc.)	1

Office hours:

Students are encouraged to meet and discuss. If you cannot meet one of the instructors in the office, please book an appointment by email.

Student Responsibilities:

- **Class participation and independent work.** Students are expected to actively participate in all classes and perform independent work.
- **Attendance.** Attendance is mandatory. University regulations regarding this matter will be strictly enforced.
- **Academic Honesty.** Individual work must be each student's own work. Plagiarism or cheating will result in an official University disciplinary review.
- **Class Etiquette.** Please keep all cell phones and other electronic devices turned off during class. If your activities during class are deemed disruptive, you will be asked to leave.