COMP 7022 NATURAL LANGUAGE PROCESSING

Credit Points 10

Legacy Code 301313

Coordinator Vernon Asuncion (https://directory.westernsydney.edu.au/search/name/Vernon Asuncion/)

Description Natural Language Processing involves machine reading comprehension and the technologies using it are becoming increasingly widespread. This subject provides a foundation in using the Natural Language Toolkit, which is a leading platform for building Python programs working with 'human language' data, as well as an introduction to Python for Natural Language Processing. Students will use algorithms and explore accessing text corpora and processing raw text; categorising words and classifying text; understanding information from text and analysing sentence structures; and understanding semantic meanings of sentences. Students also gain real-world hands-on experience with Natural Language Processing through the practical tasks and assignments.

School Computer, Data & Math Sciences

Discipline Artificial Intelligence

Student Contribution Band HECS Band 2 10cp

Check your fees via the Fees (https://www.westernsydney.edu.au/currentstudents/current_students/fees/) page.

Level Postgraduate Coursework Level 7 subject

Assumed Knowledge

There are no assumed knowledge for this subject although an undergraduate degree with some probability and statistics is advantageous.

Learning Outcomes

On successful completion of this subject, students should be able to:

1. Apply the Natural Language Toolkit to real world Natural Language Processing tasks.
2. Use text corpora for the engineering and evaluation of Natural Language Processing systems.
3. Determine the role of classification of words and text in Natural Language Processing algorithms.
4. Clarify the function of Natural Language Processing algorithms in understanding information from text and the semantic meanings of sentences.
5. Analyse artificial intelligence techniques employed in Natural Language Processing algorithms.
6. Manage linguistic data.

Subject Content

1.Installing and introducing the Natural Language Toolkit (NLKT):
-Introducing the Python programming language with emphasise for Natural Language Processing (NLP);
-Using Python string datatypes for the processing of words and texts.
2.Accessing text corpora (i.e., ?etext repositories?f) and the processing of raw text:
-Introduces the Gutenberg, Reuters, Inaugural Address and Annotated text corpus as well as Web and Chat text;
-Looking at text corpus structure.
3.Categorizing words and classifying text:
-Tagging corpora;
-Mapping words to properties through Python Dictionaries;
-Transformation based tagging;
-Determining the category of a word;
-Supervised, decision trees, naive bayes and ent

Assessment

The following table summarises the standard assessment tasks for this subject. Please note this is a guide only. Assessment tasks are regularly updated, where there is a difference your Learning Guide takes precedence.

Type	Length	Percent	Threshold	Individual/ Group Task	Mandatory
Quiz	1 hour (per Quiz)	30	Y	Individual	N
Practical	2 hours	30	N	Individual	Y
Report	1,500 words	40	Y	Individual	Y

Prescribed Texts

- Jurafsky, D., & Martin, J. H. (2020). Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition (3rd Draft ed.). Stanford. https://web.stanford.edu/~jurafsky/slp3/

Teaching Periods

Autumn (2025)

Parramatta - Victoria Rd
On-site
Subject Contact Vernon Asuncion (https://directory.westernsydney.edu.au/search/name/Vernon Asuncion/)

View timetable (https://classregistration.westernsydney.edu.au/odd/timetable/?subject_code=COMP7022_25-AUT_PS_1#subjects)