

Name of the Programme: MCA

Course Code: CSA 524

Title of Course: Natural Language Processing

Number of Credits: 4 (2L-2T-0P)

Effective from AY: 2022-23

<u>Prerequisites for the course</u>	Fundamentals of Artificial Intelligence; Mathematical Foundations for Artificial Intelligence. Machine Learning and Programming background. Introduction to NLP (Theory), Mathematical foundations for AI.	
<u>Objectives</u>	This course will focus on understanding the essentials of Natural Language Processing (NLP), areas in NLP, algorithms, and NLP tasks. Students who complete this course will gain a foundational understanding in natural language processing methods and strategies. They will also learn how to evaluate the strengths and weaknesses of various NLP technologies and frameworks as they gain practical experience in the NLP toolkits available.	
<u>Content</u>	Part I: Foundations of Natural Language Processing Introduction ● Natural Language Processing - Problems and perspectives ● Introduction/Recall to/of probability calculus ○ N-grams and Language Models ○ Markov Models ● Introduction to Machine Learning and Deep Learning ● Recurrent Neural Network Language Models ● The evaluation of NLP applications Corpora ● Corpora and their construction: representativeness ● Concordances, collocations and measures of words association ● Methods for Text Retrieval ● Regular expressions	8 hours
	Part II: Natural Language Processing ● Computational Phonetics and Speech Processing ○ Speech samples: properties and acoustic measures ○ Analysis in the frequency domain, Spectrograms ○ Applications in the acoustic-phonetic field. ○ Speech recognition with HMM and Deep Neural Networks ● Tokenisation and Sentence splitting ● Computational Morphology ○ Morphological operations ○ Static lexica, Two-level morphology ● Computational Syntax ○ Part-of-speech tagging ○ Grammars for natural language ○ Natural language Parsing ○ Supplementary worksheet: formal grammars for NL ■ Formal languages and Natural languages. ■ Natural language complexity ■ Phrase structure grammars, Dependency Grammars ■ Treebanks ■ Modern formalisms for parsing natural languages ● Computational Semantics	16 hours

	○ Lexical semantics: WordNet and FrameNet ○ Word Sense Disambiguation ○ Distributional Semantics & Word-Space models ○ Logical approaches to sentence semantics	
	Part III: Applications and Case studies: ● Solving Downstream Tasks: Document classification, Sentiment Analysis, Named Entity Recognition, Semantic Textual Similarity ● Prompting Pre-Trained Language Models ● Network Embedding	6 hours
	Sample list of Assignments to be carried out during the Tutorial Slots - Assignment -1 -Import nltk and download the 'stopwords' and 'punkt' packages. Assignment-2 -Import spacy and load the language model. Assignment -3 -How to tokenize a given text? Assignment-4 -How to get the sentences of a text document? Assignment- 5-How to tokenize a text using the 'transformers' package? Assignment -6 - How to tokenize text with stopwords as delimiters? Assignment- 7- How to remove stop words in a text?	30 hours
	Assignment -8- How to add custom stop words in spaCy? Assignment- 9 -How to remove punctuations? Assignment-10 - How to perform stemming? Assignment -11 -How to lemmatize a given text? Assignment-12 -How to extract usernames from emails? Assignment -13-How to find the most common words in the text excluding stopwords Assignment -14- How to do spell correction in a given text? Assignment -15- How to tokenize tweets? Assignment -16- How to extract all the nouns in a text? Assignment -17- How to extract all the pronouns in a text? Assignment - 18 - How to find similarity between two words? Assignment -19- How to find similarity between two documents? Assignment -20 -How to find the cosine similarity of two documents?	
<u>Pedagogy</u>	Hands-on assignments/tutorials / peer-teaching / pair programming/presentations / mini-project. Lectures / Practical / tutorials / assignments / self-study / mini-project	
<u>References/ Readings</u>	1. Allen, James, Natural Language Understanding, Second Edition, Benjamin/Cumming, 1995. 2. Charniack, Eugene, Statistical Language Learning, MIT Press, 1993. 3. Jurafsky, Dan and Martin, James, Speech and Language Processing, Second Edition, Prentice Hall, 2008. 4. Manning, Christopher and Heinrich, Schutze, Foundations of Statistical 5. Natural Language Processing, MIT Press, 1999. 6. Tamburini, F.. Neural Models for the Automatic Processing of Italian, Bologna: Pàtron. 2022 7. T. McEnery and A. Wilson. Corpus Linguistics, EUP. 2001 8. https://corpora.ficlit.unibo.it/NLP/ 9. https://www.machinelearningplus.com/nlp/nlp-exercises/ 10. Deep Learning by Goodfellow, Bengio, and Courville free online 11. Machine Learning — A Probabilistic Perspective by Kevin Murphy	

	online 12. Natural Language Processing by Jacob Eisenstein free online Speech and Language Processing by Dan Jurafsky and James H. Martin (3rd ed. draft)	
<u>Course Outcomes</u>	1. Learners will learn about the concepts in natural language processing. 2. Learners will have a fair idea of different areas in NLP 3. Learners will appreciate the complexities involved in natural language processing. 4. Through lectures and practical assignments, students will learn the necessary tricks for making their models work on practical problems. 5. They will learn how to contribute towards the development of NLP Resources and Tools.	