



Course Description

CAI2300C | Introduction to Natural Language Processing | 3.00 credits

Students will learn the fundamental concepts of Natural Language Processing (NLP) and text processing. In addition, focus will be on knowledge and skills necessary to create a language recognition application. Prerequisite: CAI2100C and recommended preparation: COP1047C or equivalent knowledge of Python programming language.

Course Competencies:

Competency 1: Students will describe standard techniques in Natural Language Processing and associated applications by:

1. Exploring AI For NLP, Applications of NLP, NLP data processing, BOW, and Algorithms in NLP
2. Processing textual data by sentence segmentation, tokenization, lemmatization, stop word removal, etc
3. Applying data preprocessing techniques like document similarity, Word Vectors, Cosine similarity, etc
4. Distinguishing between NLP models and algorithms

Competency 2: The students will describe the data acquisition process in NLP by:

1. Comparing different types of NLP datasets
2. Identifying and examining various data storage methods
3. Examining curated data sources and interpreting their usage in NLP domain
4. Downloading and processing data using the NLTK library. e) Applying data visualization techniques specific to NLP

Competency 3: The students will explore NLP Data Preprocessing by:

1. Utilizing proprietary and open-source libraries and data visualization techniques
2. Exploring and applying various vectorization techniques
3. Exploring and applying the methods of document similarity and vector visualization
4. Distinguishing between various distance measurement techniques
5. Defining and understanding the various processes associated with the NLP data pipeline

Competency 4: The students will describe, compare, and train different machine learning models by:

1. Describing and applying NLP classifiers to train machine learning models
2. Describing neural networks and their working principles. Understanding various language models
3. Defining and summarizing various Neural Language Models, N-gram Models, and Sequential Models
4. Defining and demonstrating Recurrent Neural Networks and Named Entity Recognition (NER) models through various activities and use cases

Competency 5: The students will explore NLP Model deployment by:

1. Identifying and exploring various machine learning model deployment platforms
2. Describing and classifying various types of chatbots by their applications
3. Implementing Language Detection, Transliteration, Translation, and Sentiment Analysis for different language scenarios
4. Using different tools to create and deploy chatbots, pre-existing chatbot frameworks, and Chatter On
5. Using cosine similarity in neural networks to train chatbots

Competency 6: The students will discuss and describe advanced models in NLP by:

1. Exploring the most recent developments in the NLP space
2. Explaining the workings of LSTM, Transformers, and BERT
3. Examining several NLP pre-trained models
4. Comparing the workings, performance, and architectures of Distil BERT, RoBERT, GPT, GPT-2, BERT, and BART models of NLP
5. Summarizing various types of learning fast learning techniques, Zero-shot, One-shot, and Few-shot
6. Evaluating various ethical issues in language models

Learning Outcomes:

- Use quantitative analytical skills to evaluate and process numerical data
- Solve problems using critical and creative thinking and scientific reasoning
- Formulate strategies to locate, evaluate, and apply information
- Demonstrate knowledge of ethical thinking and its application to issues in society
- Use computer and emerging technologies effectively