**Computational Linguistics**

**Credits:** 3

**Objective:**

Students will understand problems and methodscomputational linguistics and natural language processing. Students will learn the fundamentals of computational linguistics and its applications in text mining. Students will be able to apply the pre-processing and parsing methods for natural languages. Students will employ techniques and models for NLP problem scenarios, design, and implement solutions of NLP applications

**Learning Outcomes**: Students, on successful completion of the course, will be able to

1. Explain the fundamentals of computational linguistics and natural language processing (NLP)
2. Apply state-of-the-art pre-processing and parsing methods for natural languages,
3. Describe and employ suitable machine learning techniques and models for NLP problem scenarios,
4. Design and implement NLP applications.

**Prerequisites:** AI, ML, Computer Programming for Data Science

**Course Outline**:

1. Introduction to computational linguistics and natural language processing
2. Pre-processing
   1. Tokenization and sentence splitting, morphology
   2. Regular Expressions, Edit Distance
3. Language processing
   1. POS tagging
   2. Sequence labelling and recurrent neural networks
   3. Sequence to sequence transformation
4. Word Representations (vector semantics and embeddings)
   1. Vector space model and classification
   2. Static word embeddings
5. Syntactic processing
   1. Constituent parsing
   2. Dependency parsing
6. Semantic Analysis
   1. Word sense disambiguation
   2. Semantic role labeling
7. Classification models
   1. Naïve Bayes
   2. Feedforward neural networks
8. Neural NLP and transfer learning
9. Applications of Computational Linguistics
   1. Information extraction (NER and Relation Extraction)
   2. Sentiment analysis (Document level sentiment analysis and aspect based sentiment analysis)
   3. Dialogue systems /Conversational agents
   4. Machine translation

**Learning Resources:**

**Textbooks / Reference Books:**

* Textbook: Jurafsky, D., Martin, J. H., “Speech and Language Processing”, 3rd edition (online, 2019) <https://web.stanford.edu/~jurafsky/slp3/>
* One chapter from: Eisenstein, J., “Natural Language Processing” (online, 2018) <https://github.com/jacobeisenstein/gt-nlp-class/tree/master/notes>

**Practical work:**

Tutorials:

* Language processing with Spacy: <https://realpython.com/natural-language-processing-spacy-python/>
* Scikit-learn: Working with text data. <https://scikit-learn.org/stable/tutorial/text_analytics/working_with_text_data.html>
* CRFsuite: sequence labelling for Named entity recognition <https://sklearn-crfsuite.readthedocs.io/en/latest/tutorial.html>
* Word embeddings in Python with gensim: <https://machinelearningmastery.com/develop-word-embeddings-python-gensim/>
* BERT word embeddings in Huggingface: <https://mccormickml.com/2019/05/14/BERT-word-embeddings-tutorial/>
* GPT2-2 text generation tutorial: <https://minimaxir.com/2019/09/howto-gpt2/>

Assignments:

* Assignment\_1\_text\_categorization.pdf
* Assignment\_2\_sequence\_labelling.pdf
* Assignment\_3\_sentiment\_analysis.pdf

**Teaching and Learning Methods**:

1. *Direct instruction*-*based teaching* for the lectures using visual aid via slides
2. *Self-learning* method for NLP models using tutorials
3. *Inquiry based* *teaching* for the labs through one medium scale project and assignments

**Evaluation Components**

1. Examinations
2. Assignments
3. Project on Computational Linguistics