INT344:NATURAL LANGUAGE PROCESSING

L:2 T:0 P:1 Credits:3

Course Outcomes: Through this course students should be able to

CO1 :: Define the perceptions of Logistic Regression, Classification and Vector Spaces, Machine Translation, Probabilistic Models, Sequence Models, Attention Models in Natural Language Processing.

CO2 :: Understand the concepts of Sentiment Analysis, Vector Space Models, Hidden Markov Models, Language Models, Recurrent Neural Networks, Siamese Networks uses for Natural Language understanding and generation.

CO3 :: Apply Machine Learning algorithms, Semantic analysis, Syntactic analysis to Natural Language Processing leads to design Real-time NLP applications, NLP tools and systems.

CO4 :: Analyze the notions of Autocorrect, Autocomplete, Word Embeddings with Neural Networks and Syntax, Semantics, and Pragmatics of a Statement written in a Natural Language.

CO5 :: Evaluate the systems using appropriate Descriptions, Visualizations, and Statistics to communicate the problems of English language for Natural Language Processing through Semantic and Syntactic analysis.

CO6 :: Develop NLP tools to Translate Words, Translate Languages, Text Generation, Summarize Text, Word Embeddings, Build Chatbots and Question-Answering.

Unit I

Natural Language Processing with Classification and Vector Spaces : Sentiment Analysis with Logistic Regression: Extract Features from Text into Numerical Vectors, Binary Classifier using a Logistic Regression, Sentiment Analysis with Naïve Bayes: Bayes' rule for Conditional Probabilities, Naive Bayes Classifier

Unit II

Vector Space Models and Machine Translation : Vector Space Models: Vector Space Models Capture Semantic Meaning, Relationships between Words, Create Word Vectors, Capture Dependencies between Words, Visualize the Relationships in Two Dimensions Using PCA, Machine Translation and Document Search: Transform Word Vectors, Assign to Subsets using Locality Sensitive Hashing, Machine Translation and Document Search

Unit III

Natural Language Processing with Probabilistic Models : Autocorrect: Minimum Edit Distance, Dynamic Programming, Spellchecker to Correct Misspelled Words, Part of Speech Tagging and Hidden Markov Models: About Markov Chains and Hidden Markov Models, Part-Of-Speech Tags using a Text Corpus, Autocomplete and Language Models: N-gram Language Models work by Calculating Sequence Probabilities, Autocomplete Language Model using a Text Corpus, Word Embeddings with Neural Networks: Word Embeddings, Semantic Meaning of Words, NLP Tasks, Continuous Bag-Of-Words

Unit IV

Natural Language Processing with Sequence Models : Neural Networks for Sentiment Analysis: Neural Networks for Deep Learning, Positive or Negative Sentiment Categories, Recurrent Neural Networks for Language Modelling: Traditional Language Models, RNNs and GRUs, Sequential Data for Text Prediction, Next-Word Generator using a Simple RNN, LSTMs and Named Entity Recognition: Long Short-Term Memory units (LSTMs), Vanishing Gradient Problem, Named Entity Recognition Systems, Named Entity Recognition System using an LSTM, Siamese Networks: Neural Network made of Two Identical Networks and Merged Together, Identifies Duplicates in a Dataset

Unit V

Natural Language Processing with Attention Models : Neural Machine Translation: Shortcomings of a Traditional seq2seq Model, Attention Mechanism, Neural Machine Translation Model with Attention, Text Summarization: Compare RNNs and other Sequential Models, Modern Transformer Architecture, Text Summaries

Unit VI

Building Models/ Case Studies : Question Answering: Transfer Learning with State-Of-The-Art Models, T5 and Bert, Model for Answering Questions, Chatbot: Examine Unique Challenges, Transformer Models Face and their Solutions, Chatbot using a Reformer Model

List of Practicals / Experiments:

List of Practical/ Experiments

- Build a Binary Classifier for Tweets using a Logistic Regression.
- Build a Naive Bayes Tweet Classifier.
- Create Word Vectors that Capture Dependencies between Words, then Visualize their relationships in Two Dimensions using PCA.
- Transform Word Vectors and Assign them to Subsets using Locality Sensitive Hashing.
- Build your own Spellchecker to Correct Misspelled Words.
- Create Part-of-Speech Tags for a Wall Street Journal Text corpus using Markov models.
- Build your own Autocomplete Language model using a Text corpus from Twitter using N-gram Language models.
- Build your own Continuous Bag-of-Words model to Create Word Embeddings from Shakespeare text.
- Build a Sophisticated Tweet Classifier that places Tweets into Positive or Negative Sentiment categories, using a Deep Neural Network.
- Build your own Next-Word Generator using a simple RNN on Shakespeare Text data.
- Build your own Named Entity Recognition system using an LSTM and Data from Kaggle.
- Build your own Siamese Network that Identifies Question Duplicates in a Dataset from Quora.
- Build a Neural Machine Translation model with Attention that Translates English Sentences into German.
- Create a Tool that Generates Text Summaries.

Text Books:

1. NATURAL LANGUAGE PROCESSING by ELA KUMAR, DREAMTECH PRESS

References:

1. SPEECH AND LANGUAGE PROCESSING: AN INTRODUCTION TO NATURAL LANGUAGE PROCESSING, COMPUTATIONAL LINGUISTICS AND SPEECH RECOGNITION by DANIEL JURAFSKY, JAMES H. MARTIN, PEARSON