



Certification in Natural Language Processing

Syllabus

Duration: 120 Hours

1. N-gram Model

- Next Word Prediction
- Learning n-gram models
- Text Generation using n-gram models
- Handling low frequency n-grams
- Google n-grams
- Evaluation of n-gram models
- Information Retrieval using language models
- Query Likelihood Model
- Smoothed Query Likelihood Model
- Laplace Smoothing
- Jelinek-Mercer Smoothing
- Dirichlet Smoothing and Two-Stage Smoothing
- Overall IR Language Model
- Building N-Gram models
- Discussion and Doubt Clear
- Assignment and Test

2. Named Entity Recognition

- What is NER?
- Why is NER challenging?
- Applications of NER
- Annotation and Evaluation for NER
- Broad Approaches for NER
- Rule based Approaches for NER: List lookup approach
- Rule based Approaches for NER: Shallow parsing approach
- Rule based Approaches for NER: Shallow parsing approach with context
- Learning based Approaches for NER
- Read text file, extract sentences and words
- Part of Speech Tagging and NER
- Chunking/NER visualization
- Get complete Person Names and Location Names from any text

3. NLP

- **What is NLP?**
- **List of NLP Tasks**
- **Why is NLP challenging?**
- **Tokenization**
- **Lemmatization and Stemming**
- **Sentence Segmentation**
- **Phrase Identification**
- **Word Sense Disambiguation**
- **Parsing**
- **Word Tokenization with nltk**
- **Stemming and Lemmatization with nltk**
- **Tokenization, Word Counts, Stop Word removal, and Text Normalization**
- **Text Processing with Conference Abstracts Dataset**
- **Text Classification for Reuters Dataset**
- **Discussion and Doubt Clear**
- **Assignment and Test**

4. Sentiment Analysis

- **Applications of Sentiment Analysis**
- **Word Classification based Approach for Sentiment Analysis**
- **Naïve Bayes for Sentiment Analysis**
- **Challenges in Sentiment Analysis**
- **Sentiment Lexicons**
- **Learning Sentiment Lexicons: “and” and “but”**
- **Learning Phrasal Sentiment Lexicons**
- **Turney’s Algorithm , WordNet approach , Domain specific**
- **Basic Sentiment Analysis using Naive Bayes and sentiment dictionaries**
- **Sentiment Analysis on Movie Reviews Dataset**
- **Sentiment analysis on Twitter Data obtained via Tweepy**
- **Discussion and Doubt Clear**
- **Assignment and Test**

5. Summarization

- **What is Summarization? What are its applications?**
- **Genres and Types of Summaries**

- Position-based, cue phrase-based and word frequency-based approaches for extractive summarization
- Lex Rank
- Problems with Extractive Summarization Methods
- Cohesion-based Methods
- Lexical Chains Method for Extractive Summarization
- Information Extraction based Method for Extractive Summarization
- Interpretation Methods for Summarization
- Multi-document Summarization
- Evaluating Summaries – Extrinsic vs Intrinsic and ROUGE and BLEU
- Write a Simple Summarizer in Python from Scratch
- Text Summarization using Gensim (uses TextRank based summarization)
- Text Summarization using sumy (LSA, Word freq method, cue phrase method)
- LexRank using sumy
- Summarization using PyTeaser
- Discussion and Doubt Clear
- Assignment and Test

6. Topic Modeling

- What are topic models? Why do you need them?
- Plate diagrams, unigram models, mixture of unigrams
- Application of topic modeling to matrices with high dimensionality
- Singular Value Decomposition
- Latent Semantic Indexing/Analysis (LSI/LSA) as an application of SVD
- Latent Semantic Indexing/Analysis (LSI/LSA): Examples, Advantages and Drawbacks
- Probabilistic Latent Semantic Analysis (PLSA)
- Comparison between LSI and PLSA/PLSI
- Motivation for LDA
- Dirichlet Distributions
- LDA Model Details
- Comparison between various topic models: unigrams, mixture of unigrams, PLSI, LDA
- LDA Hyper-parameters
- Other Topic Models



- LDA using gensim and scikit learn
- Topic Modeling with Gensim
- Discussion and Doubt clear with Assignment and Test

7. Word Representation Learning

- What are word representations? Where can you use word vectors?
- Neural Network Language Model (NNLM)
- Word2Vec
- CBOW and Skip-gram
- GloVe (Global vectors for word representation)
- Using gensim to train your first Word2Vec model
- Finding similar words using gensim Word2Vec model
- More stuff with word2vec models: Find odd one out, compute accuracy, get the actual vector, and save model.
- Discussion and Doubt Clear
- Assignment and Test

8. Project Work

9. Placement Assistance Sessions

- Mock Interviews
- GDs
- Preplacement Talks
- Industry Exposer Sessions

10. Internship

***In Assignment there is Mini Projects.**