NATURAL LANGUAGE PROCESSING

LESSON 1: AN INTRODUCTION TO NATURAL LANGUAGE PROCESSING

SYLLABUS

WEEK 1: An Introduction to Natural Language Processing
WEEK 2: Normalization, Lemmatization, Parsing
WEEK 3: POS, Syntax, N-grams
WEEK 4: Corpus (Features and Analysis)
WEEK 5: Part Of Speech (POS) Tagging
WEEK 6: Simple Semantic Analysis
WEEK 7: Morphological & Semantic Ambiguity
MIDTERM EXAM
WEEK 8: Lexical Similarity
WEEK 9: Semantic Similarity
WEEK 10: Dialogue Systems and Question Answering
WEEK 11: Machine Translation
WEEK 12: Keyword Extraction-Document Summarization
WEEK 13: Paraphrasing / Ontology Mapping
FINAL EXAM

LESSON 2: Normalization, Lemmatization, Parsing

- Text Normalization
  - Tokenizing words
  - Normalization of word formats
  - Segmenting sentences

- Lemmatization
  - Lemma and Stem
  - Lemmatization with similarity function
  - Lemmatization with FSA

- Parsing
  - Parsing tree
  - Ambiguity

LESSON 3: POS, Syntax, N-grams

- Part of Speech
  - Types of words and categories

- Syntax
  - Natural language rules, and context-free grammar

- N-Grams
  - Probabilistic data for NLP studies

LESSON 4: Corpus-Features and Analysis

- What Is Corpus?
- Corpus Types
- Approaches In Corpus Design
- Corpus Analysis Software
- Corpus Examples
- Corpus Evaluation Algorithms
- Examples of Corpus Tools

LESSON 5: Part Of Speech (POS) Tagging

- **POS Tagging Methods**
  - Rule-Based
  - Stochastic
  - Hybrid (Transformation-based)
- **Treebanks and POS Tags**
- **POS Taggers**
  - Stanford Log-linear Part-of-Speech Tagger
  - LingPipe POS Tagger
  - NLTK POS Taggers
  - CLAWS POS Tagger for English
  - TSCorpus POS Tagger

LESSON 6: Simple Semantic Analysis

- What is Semantic?

- Content Analysis

- Semantic Analysis in CENG

- Semantic Analysis in NLP

- Vector Space Model

- Semantic Relations

- Latent Semantic Analysis (LSA)

LESSON 7: Morphological & Semantic Ambiguity

➤ Morphological Ambiguity & Morphological Disambiguation
- ◦ Surrounding bag of words.
- ◦ POS of neighboring words
- ◦ Local collocations

➤ Semantic Ambiguity & Word Sense Disambiguation
- ◦ Knowledge-based Approaches
- ◦ Supervised Approaches
- ◦ Unsupervised Approaches

➤ WordNet Based Disambiguation

LESSON 8: Lexical Similarity

- **Lexical vs. Semantic Similarity**
- **Similarity**
  - Levenstein Distance
  - Jaccard Similarity
  - Cosine Similarity
- **Vector Space Model**
  - Binary Weighting
  - Term Frequency (TF)
  - TFIDF

LESSON 9: Semantic Similarity

- Semantic Relations
- Semantic Similarity Levels
  - Sense Level
  - Word Level
  - Text Level

- WordNet-based Similarity Methods
- Hybrid Methods Similarity
- PageRank-based Similarity

LESSON 10: Dialogue Systems & Question Answering

- **Dialogue Systems**
  - Overview of Dialogue Systems
  - Types of Dialogue Systems
  - Dialogue System Components
  - Example Systems

- **Question Answering**
  - Overview of Question Answering
  - Types of Question Answering
  - Question Answering Components
  - Question Answering in TREC

LESSON 11: Machine Translation

- What is Machine Translation (MT)

- Difficulties

- Types of MT

  - Rule-Based Machine Translation (RBMT)

  - Statistical Machine Translation (SMT)

  - Neural Machine Translation (NMT)

LESSON12:Keyword Extraction-Document Summarization

- Keyword Extraction
- Keyword Extraction Approaches
  - Rule Based Linguistic Approaches
  - Machine Learning Approaches
  - Statistical Approaches
- Document Summarization
  - Extraction-based
  - Abstraction-based

LESSON 13: Paraphrasing - Ontology Mapping

- Paraphrase
  - Methods
    - Linguistic resources
    - Corpus based)
- Ontology Mapping
  - Monolingual Ontology Mapping
  - Cross Lingual Ontology Mapping
  - CLOM Approaches