Natural Language Processing Course Syllabus – Artificial Intelligence College – AASTMT (Al-Alamin)

Reference books:

- Practical Natural Language Processing
- Natural Language Processing with Transformers

Weeks	Topics

1	• Introduction to NLP:
	  • What is NLP in Real world
	  • NLP tasks
	
	• What is language?
	  • Building blocks of language (Phonemes, Morphemes & Lexemes, Syntax, and context)
	
	• Introduction to Approaches to NLP
	  • Heuristics-Based NLP
	  • Machine Learning for NLP
	  • Deep Learning for NLP

2	• NLP pipeline in detail:
	  • Data acquisition
	  • Text cleaning
	  • Pre-processing
	  • Feature engineering
	  • Modeling
	  • Evaluation
	  • Deployment
	  • Monitoring and model updating

3	

4	• Text Representation:
	  • Vector Space Models
	
	  • Basic Vectorization Approaches
	  • One-Hot Encoding
	  • Bag of Words
	  • Bag of N-Grams
	  • TF-IDF
	
	  • Distributed Representations

	• Word Embeddings

5
6
7
	• Text classification:
	  • Using Machine learning approaches:
	  • Naïve Bayes Classifier
	  • Logistic Regression
	  • SVM
	  • Using Neural embeddings
	  • Word embedding
	  • Subword embedding and fastText
	  • Document Embedding
	  • Using Deep learning:
	  • CNN
	  • Vanilla RNN
	  • GRU
	  • LSTM
	  • Text Classification with Large, Pre-Trained Language Models

8
9
	• Information Extraction:
	  • Pipeline for IE
	  • Keyphrase extraction
	  • Named Entity Recognition
	  • Building an NER System
	  • NER Using an Existing Library
	  • NER Using Active Learning
	  • Relationship extraction
	  • Temporal Information Extraction
	  • Event Extraction
	  • Template Filling

10
11
	• Transformer Anatomy
	  • The Transformer Architecture
	  • The Encoder
	  • Self-Attention
	  • The Feed-Forward Layer
	  • Adding Layer Normalization
	  • Positional Embeddings
	  • Adding a Classification Head
	  • The Decoder

12	• NLP application 1: Text Generation
	  • The Challenge with Generating Coherent Text
	  • Greedy Search Decoding
	  • Beam Search Decoding
	  • Sampling Methods
	  • Top-k and Nucleus Sampling
	  • Which Decoding Method Is Best?

13	• NLP application 2: Summarization:
	  • The CNN/DailyMail Dataset
	
	  • Text Summarization Pipelines
	  ▪ Summarization Baseline
	  ▪ GPT-3
	  ▪ T5
	  ▪ BART
	  ▪ PEGASUS
	
	  • Comparing Different Summaries
	
	  • Measuring the Quality of Generated Text
	  • BLEU 148
	  • ROUGE 152
	
	  • Chatbots (ChatGPT) and Prompt Engineering

14	Project Discussions
15	Revision
16	Final Exam