CENG464 Text Mining

Dr. Buket ERŞAHİN

Teaching

- Instructor: Buket Erşahin
- Res. Assistant:
  - Güliz Akkaya
  - Cansu Özkan

Reference Books

- Speech and Language Processing, An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Daniel Jurafsky & James H. Martin
- Natural Language Processing Fundamentals, Sohom Ghosh & Dwight Gunning.

Summary of Contents

Grading

- %20 Individual Project
- %40 Team Project
- %40 Quizzes (biweekly)

- Students are expected to work on two programming projects: (i) an individual project involving implementing an NLP algorithm on a standard data set with evaluation, and (ii) a semester-long group project involving a state-of-the-art NLP problem: students will choose between topics. We will be using common data sets to facilitate evaluation wherever possible. The project requirements will be discussed in detail in the first two weeks. You will receive guidance regarding data collection, algorithms, evaluation methodology during the semester. Students will be required to present their final group project during the last two weeks of class. Students are also required to write a technical paper describing their project and experiments. You will work in groups for the class Project.

What is this course about ?

- A comprehensive set of topics in natural language processing (NLP)
- The required background is a combination of information retrieval, machine learning, and programming expertise in Python.
- Knowledge-based, traditional feature-based approaches and deep learning approaches
- In the first part
  - Fundamental algorithms , language models, POS analysis, entity recognition…
- In the second part
  - Deep learning -> neural embeddings, encoder-decoder models, transformers, transfer learning using pre-trained models
- These topics are presented in the context of NLP tasks such as machine translation and sentiment analysis.
- Each session will have a lecture component followed by a recitation involving interactive and code demonstration session for hands-on learning.