

CSIT 261

Foundations of Natural Language Processing and Information Retrieval

3 Credits

Community College of Baltimore County
Common Course Outline

Description

CSIT 261 – Foundations of Natural Language Processing and Information Retrieval: provides an overview of natural language processing (NLP) and information retrieval (IR) techniques. Students apply problem-solving skills using NLP and IR algorithms, methodologies, and concepts. Additional topics include text, link, and sentiment analysis, classification algorithms, language theory, search, recommender systems, and examination of emerging trends in the field.

Pre-requisites: CSIT 210 and MATH 153 or permission of the program director

Overall Course Objectives

Upon completion of this course, students will be able to:

1. design programs using algorithms for NLP;
2. describe techniques used in IR;
3. compare search engine indexing techniques;
4. perform sentiment analysis on text or other types of data;
5. discuss the syntax of grammars and languages in NLP and IR;
6. recognize ways to represent text and information;
7. compare techniques for data classification and clustering;
8. examine various feature extraction techniques;
9. create programs that utilize text analysis;
10. create programs that utilize classification methods;
11. perform link analysis on structured or unstructured data;
12. design programs that perform web scraping and information extraction;
13. evaluate methods used for search and IR;
14. explain algorithms used for recommender systems;
15. identify ethical issues, bias, and problems of fairness in NLP and IR systems; and
16. explain emerging trends and technologies related to NLP and IR.

Major Topics

- I. Introduction of natural language processing
- II. Overview of information retrieval
- III. Text representation
- IV. Text analysis and classification
- V. Grammars and language
- VI. Link analysis
- VII. Feature extraction techniques
- VIII. Web scraping and information extraction

The Common Course Outline (CCO) determines the essential nature of each course.
For more information, see your professor's syllabus.

- IX. Sentiment analysis
- X. Foundations of information retrieval and search algorithms
- XI. Data classification and clustering
- XII. Search engine indexing
- XIII. Recommender systems
- XIV. Ethics, bias, and fairness in natural language processing and information retrieval
- XV. Emerging trends

Course Requirements

Grading will be determined by the individual faculty member, but shall include the following, at minimum:

- Six quizzes or assignments
- Six projects
- Two exams
- One comprehensive final exam and/or final project

Written assignments and research projects: Students are required to use appropriate academic resources in their research and cite sources according to the style selected by their professor.

Date Revised: 10/19/2021