IT 335 Introduction to Natural Language Processing

Introduction

Assist. Prof. Dr. Dželila Mehanović

IBU

IBU

International Burch University

Lecturers

- Dželila Mehanović - dzelila.mehanovic@ibu.edu.ba
- Naida Fatić - naida.fatic@ibu.edu.ba

Grading

- Midterm exam: 30%
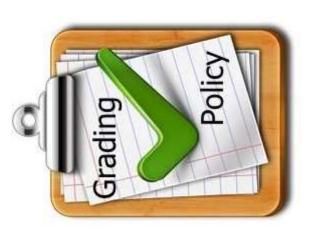- Final exam: 30%

- Assignments 10% (every week)
- Projects 30%

- 55% of attendance on lectures and labs is mandatory

We count on academic honesty !

Textbooks

1. Steven Bird, Ewan Klein, and Edward Loper. 2009. Natural Language Processing with Python. Second Edition. O'Reilly

Topics

1. Introduction to NLP, what is NLP, applications and challenges
2. Recap of Python and pandas fundamentals
3. Accessing Text Corpora and Lexical Resources
4. Processing Raw Text
5. Categorizing and Tagging Words
6. Introduction to Machine Learning
7. Learning to Classify Text, Naive-Bayes & Logistic Regression Classifiers
8. MIDTERM EXAM
9. Extracting Information from Text
10. Intro to Scikit-learn for NLP
11. Case Study: Tweet Classification
12. Intro to word embeddings
13. Chatbots
14. Named Entity Recognition using SpaCy
15. Project presentations

Project Requirements

PART 1:

- Practical experiments and implementation of the selected topic

PART 2:

- Research paper about the selected topic
- Paper should include following topics

INTRODUCTION

THEORETICAL BACKGROUND / LITERATURE REVIEW

DATASET

METHODOLOGY

RESULTS

DISCUSSION AND CONCLUSION

Sentiment Analysis research paper example available [here](#)

Project Ideas

- Developing a Chatbot in Python Using Natural Language Processing and Neural Networks
- AI and Language Models - Twitter Data Analysis
- Exploratory Data Analysis for Amazon reviews dataset
- Text Summarization of News Articles
- Amazon Book Review Sentiment Analysis
- Analysis and Processing of Spam Messages
- Fake news Classifier using Long Short Term Memory

Project Defense / Presentation

- There will not be submission on the LMS
- You should create folder with your name and surname [here](#) and add all your materials (dataset, code and paper) there
- Add your topic [here](#)
- Presentation and defense of the projects will be in the last week of semester

Course name on the LMS

Technical Elective 7 - Introduction to Natural Language Processing

What is Natural Language Processing

By "natural language" we mean a language that is used for everyday communication by humans; languages such as English. In contrast to artificial languages such as programming languages and mathematical notations, natural languages have evolved as they pass from generation to generation, and are hard to pin down with explicit rules.

We will take Natural Language Processing - or NLP for short- in a wide sense to cover any kind of computer manipulation of natural language.

Technologies based on NLP are becoming increasingly widespread.

For example, phones and handheld computers support predictive text and handwriting recognition; web search engines give access to information locked up in unstructured text; machine translation allows us to retrieve texts written in Chinese and read them in Spanish.

By providing more natural human-machine interfaces, and more sophisticated access to stored information, language processing has come to play a central role in the multilingual information society.

Key applications of NLP

NLP has seen significant advancements and widespread adoption in various industries. Let's explore some key applications of NLP in different domains:

Sentiment Analysis and Opinion Mining:
Sentiment analysis involves determining the sentiment or emotional tone of a given text. It finds application in analyzing customer feedback, social media monitoring, brand reputation management, and market research. By analyzing sentiment, businesses can make data-driven decisions, identify customer needs, and tailor their products or services accordingly.

Text Classification and Document Categorization:
NLP techniques are used to classify documents into predefined categories or topics. This application finds applications in news categorization, spam filtering, content moderation, and legal document analysis. By automatically categorizing large volumes of text, organizations can efficiently organize and retrieve information, automate workflows, and improve productivity.

Key applications of NLP

Named Entity Recognition (NER) and Entity Linking:

NER is the task of identifying and classifying named entities such as names, organizations, locations, and dates in text. Entity linking aims to associate these entities with knowledge bases or ontologies. These techniques are crucial for information extraction, question answering systems, recommendation engines, and content generation. They enable machines to understand the context and meaning of textual data.

Machine Translation:

NLP plays a vital role in machine translation systems, enabling communication across different languages. Advanced models like neural machine translation (NMT) have significantly improved translation accuracy and fluency. These systems find applications in cross-border communication, localization of software and websites, and multilingual content generation.

Key applications of NLP

Chatbots and Virtual Assistants:

Chatbots and virtual assistants powered by NLP technologies are becoming increasingly common in customer support, helpdesk services, and online interactions. These systems utilize techniques such as intent recognition, dialogue management, and sentiment analysis to provide automated responses, answer queries, and assist users in various domains.

Question Answering and Information Retrieval:

NLP techniques are used to build question answering systems that can comprehend and answer questions posed in natural language. These systems find applications in search engines, customer support, and knowledge base systems. By extracting relevant information and providing precise answers, they enhance user experience and enable efficient information retrieval.

Key applications of NLP

Text Summarization and Automatic Document Generation:
NLP algorithms can generate concise summaries of long texts, enabling users to quickly grasp the main points without reading the entire document. This application finds utility in news summarization, document summarization for research purposes, and generating executive summaries for business reports.

Fake News Detection:
With the proliferation of misinformation, NLP techniques are employed to detect and combat fake news. By analyzing the linguistic patterns, credibility of sources, and fact-checking information, NLP models can identify suspicious content and aid in maintaining the integrity of information.

Medical Text Mining and Healthcare Applications:
NLP is increasingly used in the healthcare industry for analyzing medical records, clinical notes, and scientific literature. It aids in medical coding, adverse event monitoring, clinical decision support, and drug discovery. NLP techniques help healthcare professionals extract meaningful insights from unstructured textual data, leading to improved patient care and research outcomes.

Key applications of NLP

Social Media Analysis and Influencer Marketing:
NLP techniques are utilized to analyze social media conversations, detect trends, and identify influencers. This information helps businesses understand customer preferences, sentiment towards their brand, and market trends. It enables targeted advertising, influencer marketing strategies, and social media management.

These are just a few examples of the wide range of applications of NLP in 2023.

As NLP continues to advance, it is expected to have an even greater impact on various industries, transforming the way we interact with and process textual data.

Questions?