20BCE069
Practical 1
NLP

Libraries

- ☐ **NLTK**

  It is used for preprocessing the unstructured data which contains human-readable text.

  Features:

  - Used for tokenization: giving tokens to words which identifies the word that appear frequently.
  - Filtering stop words
  - Stemming : Text processing task in which you reduce words to their root, which is the core part of a word.
  - Tagging parts of speech
  - Lemmatizing
  - Chunking: identifies the phrases

  Cons:

  - NLTK is a complicated solution with a harsh learning curve and a maze of internal limitations.
  - For sentence tokenization, NLTK doesn't apply semantic analysis. Unlike *Gensim*, NLTK lacks neural network models or word embeddings.
  - NLTK is slow, whereas spaCy is said to be the fastest alternative. However, it's possible to speed up execution using Python's multiprocessing module.

- ☐ **Spacy**

  Features:

  - Parts of speech tagging
  - Making word predictions

- Morphology : morphology is the process by which a root form of a word is modified by adding prefixes or suffixes that specify its grammatical function but do not change its part-of-speech
- Lemmatization
- Tokenization
- Merging and splitting

Cons:

- less flexibility compared to NLTK

☐ Gensim

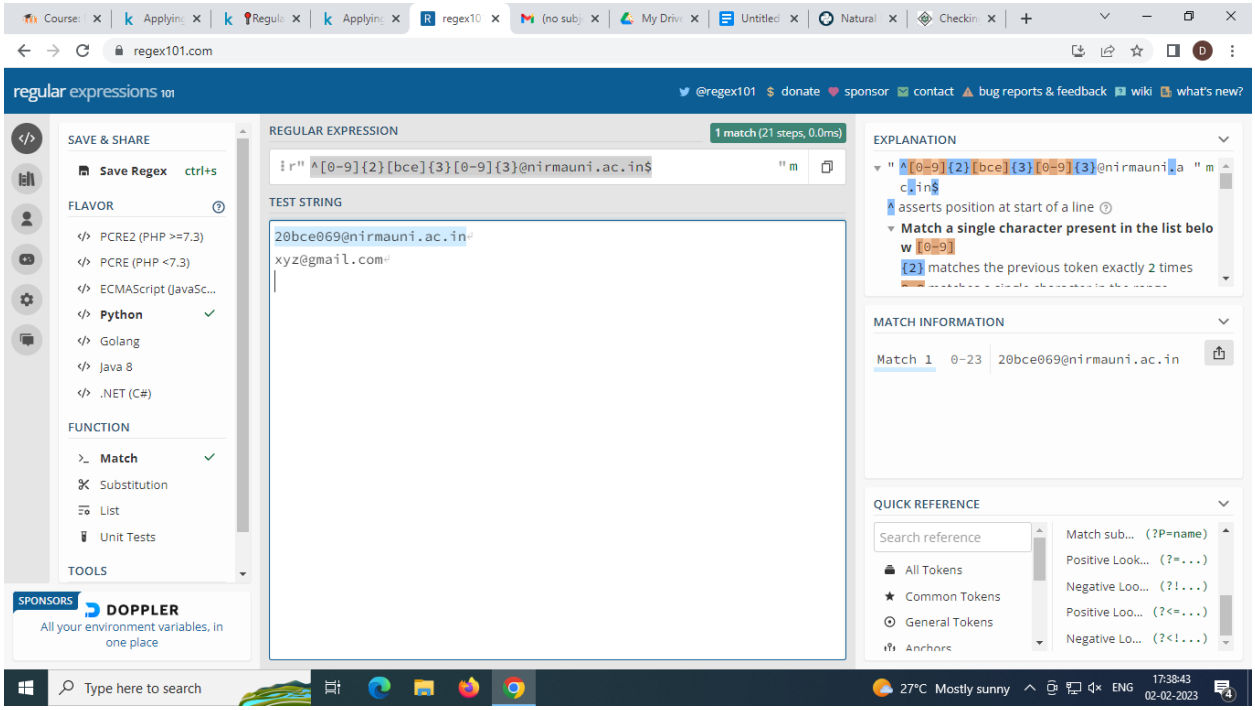
Features:

- Parallelized implementations of fastText, word2vec and doc2vec algorithm
- Latent semantic analysis (LSA, LSI, SVD)
- Non-negative matrix factorization (NMF)
- Latent Dirichlet allocation (LDA)
- tf-idf

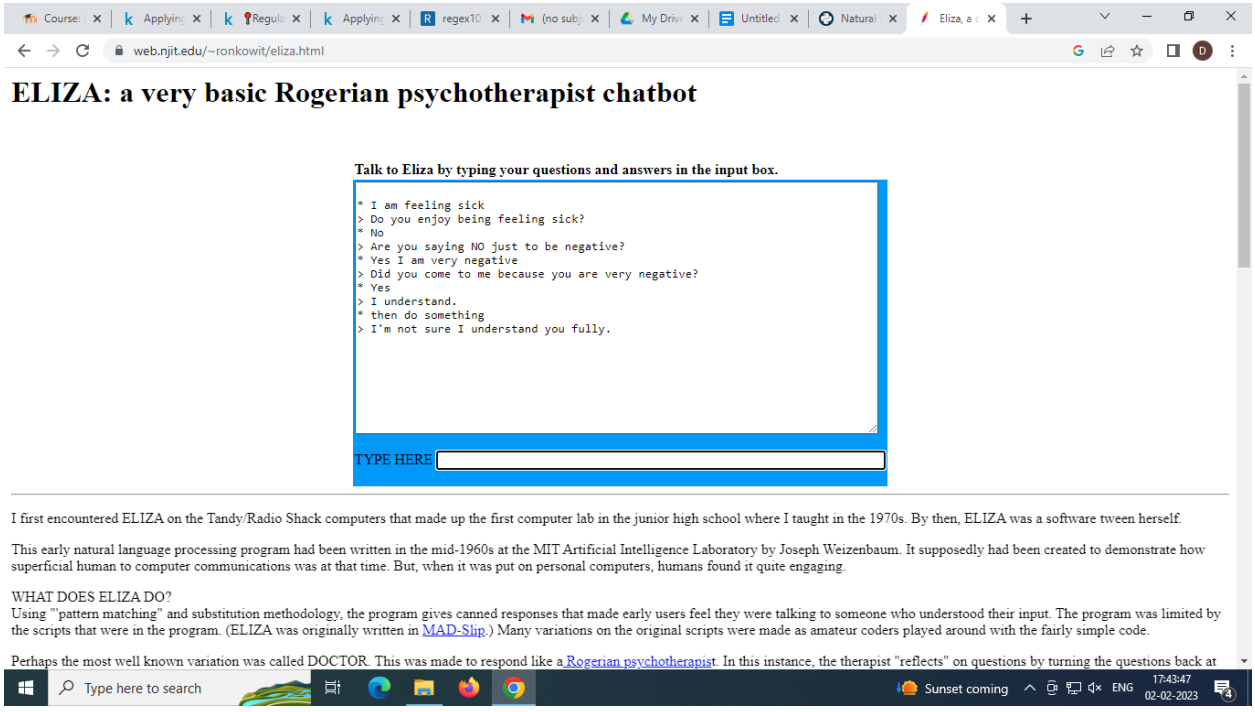
Cons:

- designed primarily for unsupervised text modeling
- don't implement full NLP pipeline, should be used with other libarary like Spacy or NLTK

REGULAR EXPRESSION FOR EMAIL



TALK WITH ELIZA



Features:

- Eliza is a very simple chatbot to implement as it relies on a limited set of rules.
- It can simulate a conversation by asking questions, rephrasing, and reflecting the user's input back to them.
- Eliza can appear to be empathetic, supportive, and understanding by following the Rogerian therapy approach.
- It can handle a limited range of topics, such as emotions, relationships, and personal experiences.

Limitations:

- Eliza's responses are limited to a pre-programmed set of rules and patterns, which can make it repetitive and predictable.
- It cannot understand the meaning behind the user's input, which can lead to irrelevant or nonsensical responses.
- Eliza does not learn from user interactions or improve over time, making it less effective at understanding and adapting to individual users.
- It lacks the ability to provide helpful advice, guidance, or solutions to users' problems as it only reflects what the user says.
- It is not capable of handling complex conversations, such as those involving sarcasm, humor, or multiple topics.