Software Engineering Assignment 2 Keyword Extraction

---

1. Overview

This project is a software engineering text analyzer that takes a transcript file as input and outputs a list of keywords related to software engineering. The list of keywords is displayed in a GUI window where users can select a word to see its definition. The definitions are stored in a separate text file called "definitions.txt".

2. Requirements

- Python 3.7 or higher
- tkinter
- re
- spacy

3. How to run

1. Install required libraries

```
pip install tk re spacy
```

2. Save the transcript file as "transcript.txt" in the same directory as the python script.
3. Run the script in the terminal or IDE, the output keyword list will be stored in "input.txt" for future use.
4. The GUI window will pop up, select a word to see its definition.
5. Definitions are stored in definitions.txt.

4. How it works

4.1. Keyword Extraction

The keyword extraction process uses spaCy's NLP library to extract noun chunks from the transcript text. A list of keywords related to software engineering is created by filtering the noun chunks based on a predefined list of keywords. The filtered list of keywords is then saved to the "input.txt" file.

4.2 GUI Window

The GUI window displays the filtered list of keywords in a scrollable window. Users can select a word to see its definition. The definitions are stored in the "definitions.txt" file and are displayed in a separate window when a word is selected.

5. Customizing the Project

You can customize the project by adding, removing, or modifying the definitions in the "definitions.txt" file. Additionally, you can modify the list of keywords used for filtering the noun chunks to suit your needs.