Internship Project Report

Name: Sanika Sunil Ulane

Company: Elevate Labs

Project Title: Text Similarity Checker (Plagiarism Detector)

1. Introduction

In today's digital world, a lot of content is created and shared. So, checking whether content is copied or original is very important. This project is about creating a tool that compares two text files and tells how much similar they are. The idea is to check for copied content easily.

I made this project as part of my internship at Elevate Labs. The application is simple to use and helps anyone who wants to check if content is copied from somewhere else. It shows results in percentage and also lists words that are common in both files. It even lets you save the result as a text or PDF file.

2. Abstract

The Text Similarity Checker is a desktop tool made using Java. It allows the user to select two input files and an optional stopword file. The tool processes both files by removing punctuation and common words (stopwords), and then checks how much they are similar.

It uses a method called Cosine Similarity, which looks at how often each word appears and then compares the files. After checking, the result is shown in the application and saved to an output file. You can also click a button to save the result as a PDF, and it opens automatically. The tool is useful for students, teachers, and writers.

3. Tools Used

- Java 8+: Programming language

- Java Swing: Used for creating the GUI

- Java I/O: Reading and writing files

- HashMap, Set, List: To store and compare words

- Cosine Similarity: To calculate how similar the files are

- iText PDF Library: To create and open the result as a PDF

- Eclipse IDE: Used to write and run the code

4. Steps Involved in Building the Project

1. Created a Java Swing interface to select files.

2. Read the contents of both files.

3. Removed punctuation and common stopwords.

4. Counted word frequency in both files.

5. Used Cosine Similarity to calculate similarity.

6. Showed result in percentage and common words.

7. Saved the result to a text file.

8. Added a feature to export and open the result as a PDF.

5. Conclusion

This project helped me learn how to work with files, create a GUI, and use basic natural language processing ideas. It also taught me how to use external libraries like iText. I now understand how text similarity works using word comparison.

This tool can be helpful in checking assignments, blog content, or any kind of writing. It is easy to use and gives quick results. In the future, I can improve it by adding language support, better result display, or making it available online.

Overall, the project gave me good hands-on experience and improved my Java skills.