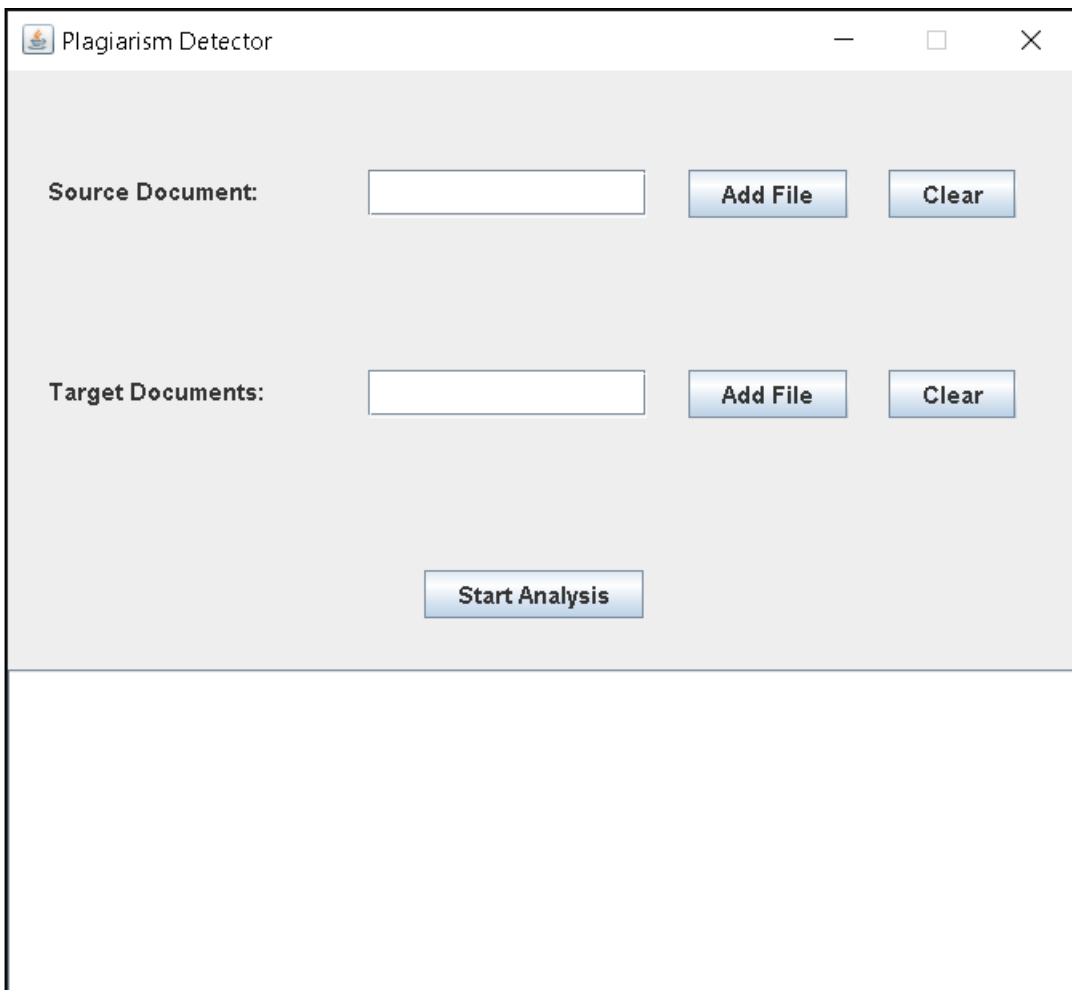


Purpose of the Application

This application is designed to detect plagiarism and calculate the similarity rate between source text document and target text documents.

Application UI

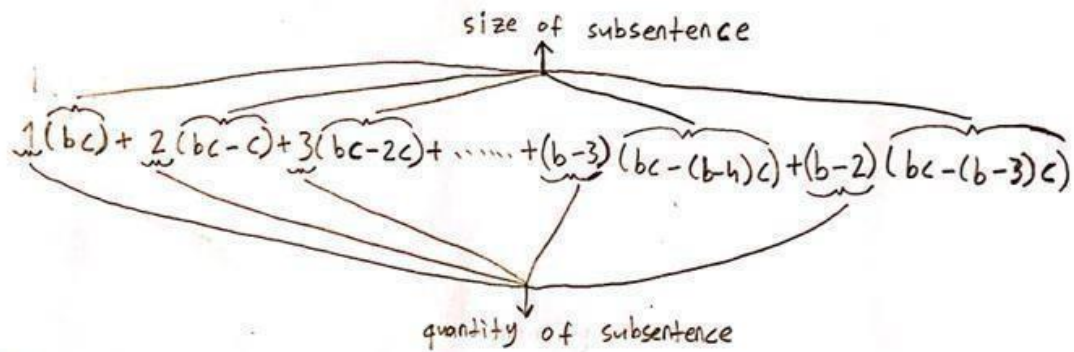


The screenshot shows a window titled "Plagiarism Detector" with a standard Windows-style title bar (minimize, maximize, close buttons). The main content area is light gray and contains two rows of input fields and buttons. The first row is labeled "Source Document:" and includes a text input field, an "Add File" button, and a "Clear" button. The second row is labeled "Target Documents:" and includes a text input field, an "Add File" button, and a "Clear" button. Below these rows, centered, is a "Start Analysis" button. The bottom half of the window is a large, empty white rectangular area, likely intended for displaying the results of the analysis.

Worst-Case Time Complexity of an Analysis

Source document: Has k sentence. Each sentence has b word, and each word has c letter.

Subsentences of source document sentence:



Target document: Has d sentence. Each sentence has e letter.

Worst-case time complexity:

$$kde(bc + 2(bc-c) + 3(bc-2c) + \dots + (b-3)(bc-(b-4)c) + (b-2)(bc-(b-3)c))$$

$$(bc + 2bc + 3bc + \dots + (b-3)bc + (b-2)bc) - (1*2c + 2*3c + 3*4c + \dots + (b-4)(b-3)c + (b-3)(b-2)c)$$

$$\frac{bc(b-2)(b-1)}{2} - \frac{(b-3)(b-2)(b-1)c}{3} = \frac{b^3c + 3b^2c - 16bc + 12c}{6}$$

worst-case time complexity is $\rightarrow O(kdeb^3c)$

* If multiple target documents are compared with source document, the largest target document in terms of number of letters among all target documents is considered

Exemplary Outputs

1-) Example 1

Source text document:

The European Values Study is large-scale, time-intensive survey on basic human values. It provides insights into the values, beliefs and preferences of citizens all over Europe. It is a unique research project on how Europeans think about life, family, work, religion, politics and society. The European Values Study was launched in 1981, when a couple of hundred citizens in the European Member States were interviewed using standardized questionnaires. Every nine years, the survey is repeated in an increasing number of countries.

Not all the respondents of the original data sample are included in the analysis. People who did not answer one or more of the questions included, are filtered out of dataset. The final number of respondent has been brought down to a sample analysis of 60077 respondents.

Target text document:

The European Values Study is a large-scale, cross-national, and longitudinal survey research program on basic human values. It provides insights into the ideas, beliefs, preferences, attitudes, values and opinions of citizens all over Europe. It is a unique research project on how Europeans think about life, family, work, religion, politics and society.

The European Values Study started in 1981, when a thousand citizens in the European Member States of that time were interviewed using standardized questionnaires. Every nine years, the survey is repeated in an increasing number of countries. The fourth wave in 2008 covers no less than 47 European countries/regions, from Iceland to Azerbaijan and from Portugal to Norway. In total, about 70,000 people in Europe are interviewed.

Output:

Similarity Rate between source.txt and target.txt is 53.125%

Most similar sentence is "It is a unique research project on how Europeans think about life, family, work, religion, politics and society." in document named source.txt, in paragraph 1, and sentence 3, and it has 100.0% similarity rate with the sentence "It is a unique research project on how Europeans think about life, family, work, religion, politics and society." in document named target.txt, in paragraph 1, and sentence 3. Plagiarised parts: "It is a unique research project on how Europeans think about life, family, work, religion, politics and society.".

Second most similar sentence is "Every nine years, the survey is repeated in an increasing number of countries." in document named source.txt, in paragraph 1, and sentence 5, and it has 100.0% similarity rate with the sentence "Every nine years, the survey is repeated in an increasing number of countries." in document named target.txt, in paragraph 2, and sentence 2. Plagiarised parts: "Every nine years, the survey is repeated in an increasing number of countries.".

Third most similar sentence is "The European Values Study is large-scale, time-intensive survey on basic human values." in document named source.txt, in paragraph 1, and sentence 1, and it has 75.0% similarity rate with the sentence "The European Values Study is a large-scale, cross-national, and longitudinal survey research program on basic human values." in document named target.txt, in paragraph 1, and sentence 1. Plagiarised parts: "The European Values Study is", "on basic human values.".

Fourth most similar sentence is "The European Values Study was launched in 1981, when a couple of hundred citizens in the European Member States were interviewed using standardized questionnaires." in document named source.txt, in paragraph 1, and sentence 4, and it has 75.0% similarity rate with the sentence "The European Values Study started in 1981, when a thousand citizens in the European Member States of that time were interviewed using standardized questionnaires." in document named target.txt, in paragraph 2, and sentence 1. Plagiarised parts: "citizens in the European Member States", "The European Values Study", "in 1981, when a", "were interviewed using standardized".

Fifth most similar sentence is "It provides insights into the values, beliefs and preferences of citizens all over Europe." in document named source.txt, in paragraph 1, and sentence 2, and it has 71.42857% similarity rate with the sentence "It provides insights into the ideas, beliefs, preferences, attitudes, values and opinions of citizens all over Europe." in document named target.txt, in paragraph 1, and sentence 2. Plagiarised parts: "It provides insights into the", "of citizens all over Europe.".

Elapsed time \approx 8ms – 12ms

2-) Example 2

Source text document:

ABC College for Women is one of the most prestigious institutions of London with a full time enrollment of about 8000 students. The glorious academic values of this oldest premier post-graduate female institution have been shaped by its institutional history, which is spread over a span of 64 years. In January 2002, the University made all strong decisions for the improvement in Higher Education. Established in May 1962 as an Intermediate residential college and affiliated with the University of Oxford, it was housed in a building on XYZ Road, with a strength of 90 students and then the progress flourished with full shot. And College started programs like Electronics, Environmental Science, Fine Arts, Economics and Mass Communication. Various national industries and linkages with foreign Colleges helped a lot...

Target text document:

Since the establishment of ABC College for Women and in early January 2002, the University has tried its level best for improvement in Higher Education. Government did various national industries and linkages with foreign universities. MoU with various national industries and linkages with foreign universities have been established in the field of Pharmacy, Electronics, Environmental Science, Fine Arts, Economics and Mass Communication. This is how they made the glorious academic values of this oldest premier post-graduate female institution very nicely.

Output:

Similarity Rate between source.txt and target.txt is 30.46875%

Most similar sentence is "And College started programs like Electronics, Environmental Science, Fine Arts, Economics and Mass Communication." in document named source.txt, in paragraph 1, and sentence 5, and it has 64.28571% similarity rate with the sentence "Government did various national industries and linkages with foreign universities MoU with various national industries and linkages with foreign universities have been established in the field of Pharmacy, Electronics, Environmental Science, Fine Arts, Economics and Mass Communication." in document named target.txt, in paragraph 1, and sentence 2. Plagiarised parts: "Electronics, Environmental Science, Fine Arts, Economics and Mass Communication.".

Second most similar sentence is "Various national industries and linkages with foreign Colleges helped a lot..." in document named source.txt, in paragraph 1, and sentence 6, and it has 63.636364% similarity rate with the sentence "Government did various national industries and linkages with foreign universities MoU with various national industries and linkages with foreign universities have been established in the field of Pharmacy, Electronics, Environmental Science, Fine Arts, Economics and Mass Communication." in document named target.txt, in paragraph 1, and sentence 2. Plagiarised parts: "Various national industries and linkages with foreign".

Third most similar sentence is "In January 2002, the University made all strong decisions for the improvement in Higher Education." in document named source.txt, in paragraph 1, and sentence 3, and it has 53.333336% similarity rate with the sentence "Since the establishment of ABC College for Women and in early January 2002, the University has tried its level best for improvement in Higher Education." in document named target.txt, in paragraph 1, and sentence 1. Plagiarised parts: "January 2002, the University", "improvement in Higher Education.".

Fourth most similar sentence is "The glorious academic values of this oldest premier post-graduate female institution have been shaped by its institutional history, which is spread over a span of 64 years." in document named source.txt, in paragraph 1, and sentence 2, and it has 40.74074% similarity rate with the sentence "This is how they made the glorious academic values of this oldest premier post-graduate female institution very nicely." in document named target.txt, in paragraph 1, and sentence 3. Plagiarised parts: "The glorious academic values of this oldest premier post-graduate female institution".

Fifth most similar sentence is "ABC College for Women is one of the most prestigious institutions of London with a full time enrollment of about 8000 students." in document named source.txt, in paragraph 1, and sentence 1, and it has 18.181818% similarity rate with the sentence "Since the establishment of ABC College for Women and in early January 2002, the University has tried its level best for improvement in Higher Education." in document named target.txt, in paragraph 1, and sentence 1. Plagiarised parts: "ABC College for Women".

Elapsed time \approx 16ms – 22ms