Design Analysis and Algorithm

Report

Description

This Text Search Application enables users to efficiently search for specific words or phrases across multiple text files. It offers two search algorithms—Brute Force and Knuth-Morris-Pratt (KMP)—to ensure efficient and precise string matching. Users can customize their search by enabling options like Match Case and Whole Word Match. The application displays each match in context, including the file name, line number, column position, and a snippet of the surrounding text.

Functionalities

1. Search Algorithms

- **Brute Force**: A straightforward algorithm that checks each character in the text sequentially to find matches.
- **KMP (Knuth-Morris-Pratt)**: An optimized search algorithm that skips parts of the text where matches are impossible, improving search speed.

2. Search Options

- **Match Case**: Makes the search case-sensitive. For instance, searching for "Dioxide" will not match "dioxide."
- **Whole Word Match**: Matches only standalone words rather than substrings within larger words. For example, searching "Washington" will not match "Washingtonian."

Display Results

For each match found, the results display:

- **File Name**: The name of the file where the match was found.
- **Line Number**: The specific line in the file where the match is located.
- **Column Number**: The starting position of the matched word in the line.
- **Context**: A snippet around the matched word for contextual clarity.

4. Clear Function

Clears the input search term and any previous search results.

5. Scrollable Results
Includes a scroll bar in the results area, allowing users to view extensive search results comfortably.

3. Language Used

This application is built with Python, using the tkinter library for the graphical user interface (GUI). Python was selected for the following reasons:

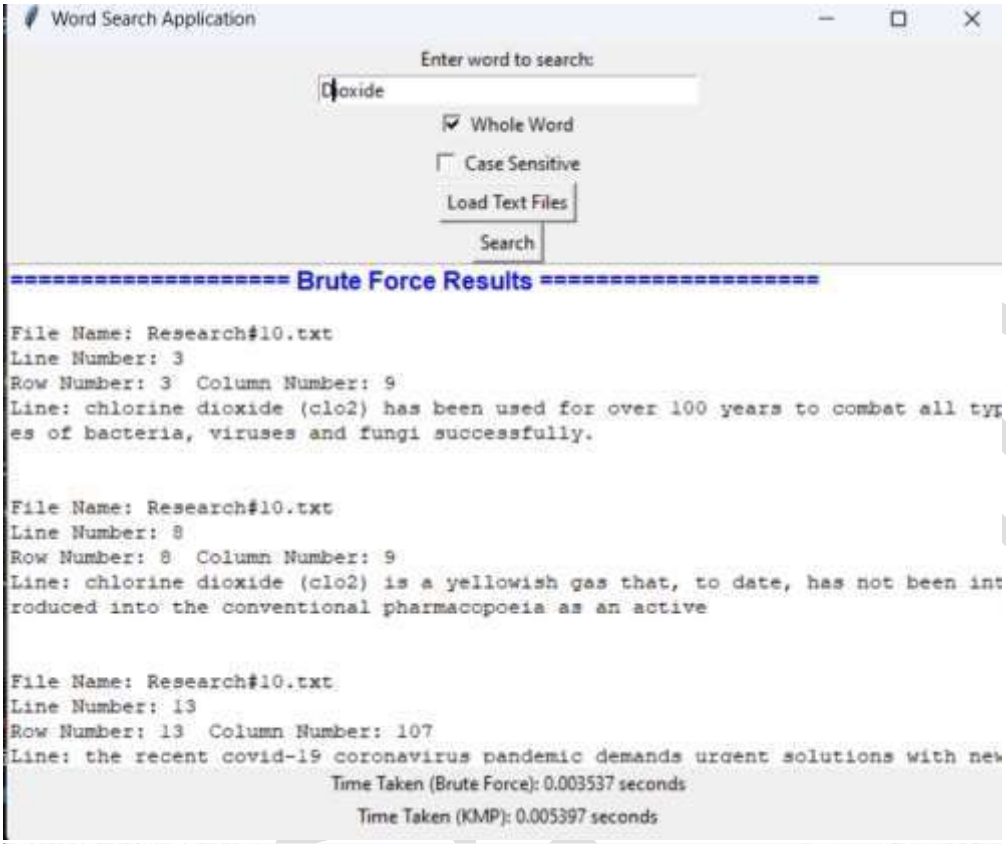
- **Rich Libraries**: Python offers extensive libraries for GUI development, string manipulation, and data handling.
- **Readable and Simple Syntax**: This makes it straightforward to implement and maintain complex algorithms like KMP.
- **Cross-Platform Compatibility**: Python's cross-platform nature allows the application to run on different operating systems without any code changes.

User Guide

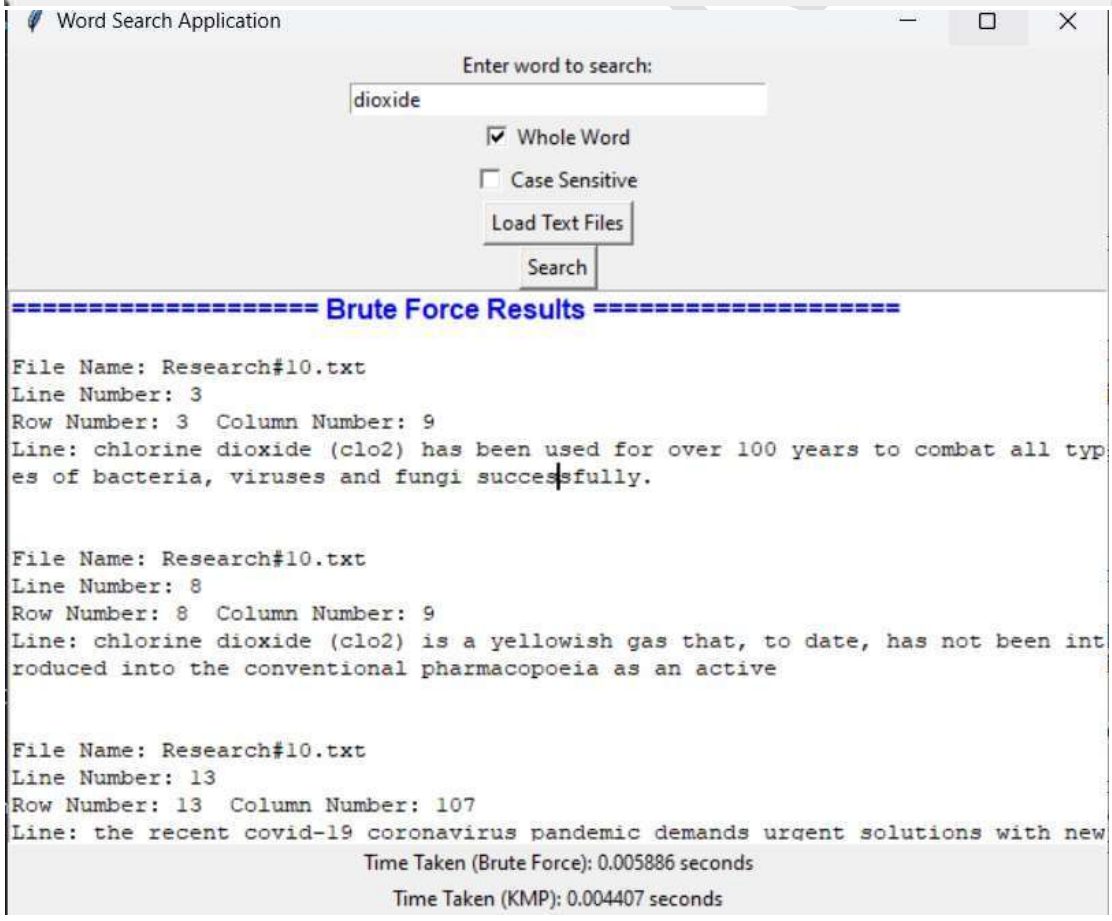
How to Use the Application

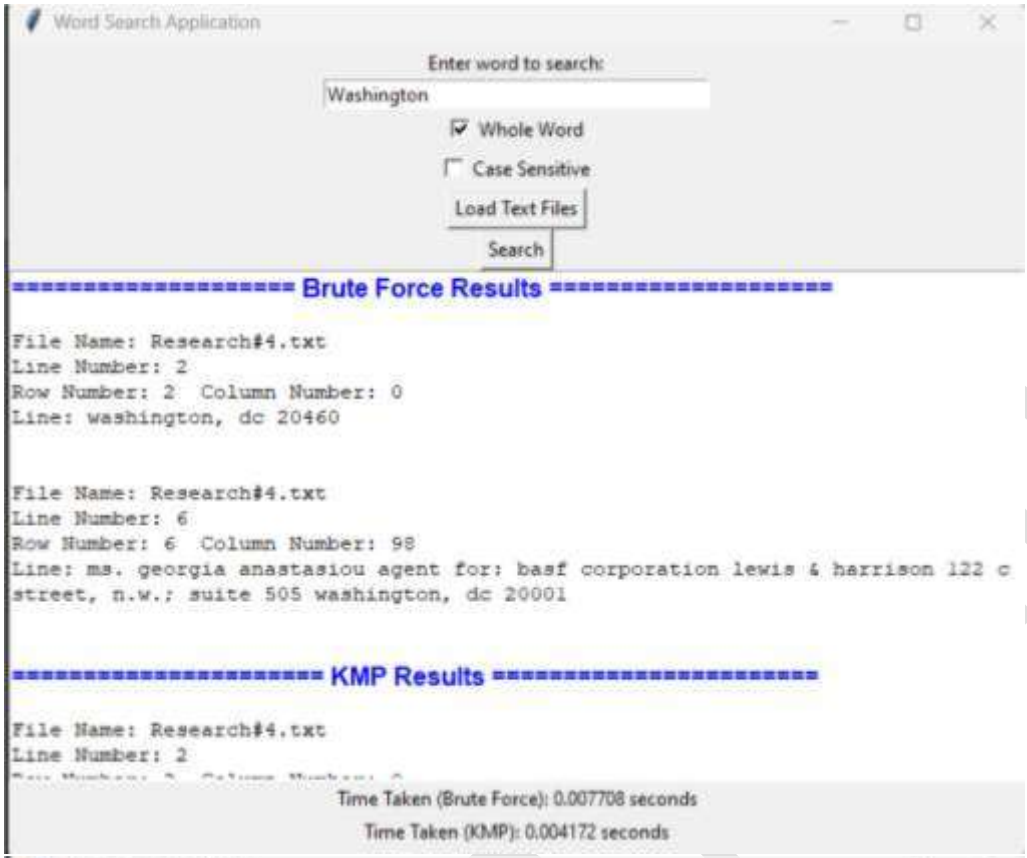
1. **Enter Search Term**: Type the word or phrase you want to search for, such as "Dioxide" or "Washington."
2. **Select Options**:
    - **Case Sensitive**: Check this box if you want the search to be case-sensitive. For example, checking "Match Case" will distinguish "Dioxide" from "dioxide."
    - **Whole Word**: Check this box if you want to match only whole words. For example, searching for "Washington" will ignore instances like "Washingtonian."
3. **Search**:
    - Click **Search** to start the search process.
    - Select the text files you want to search through.
4. **View Results**:
    - The results will appear in the text area, showing details like the file name, line number, Row number, column position, and a snippet around the matched text.
    - Use the scroll bar to browse through all the search results. Matches are highlighted for easier visibility.

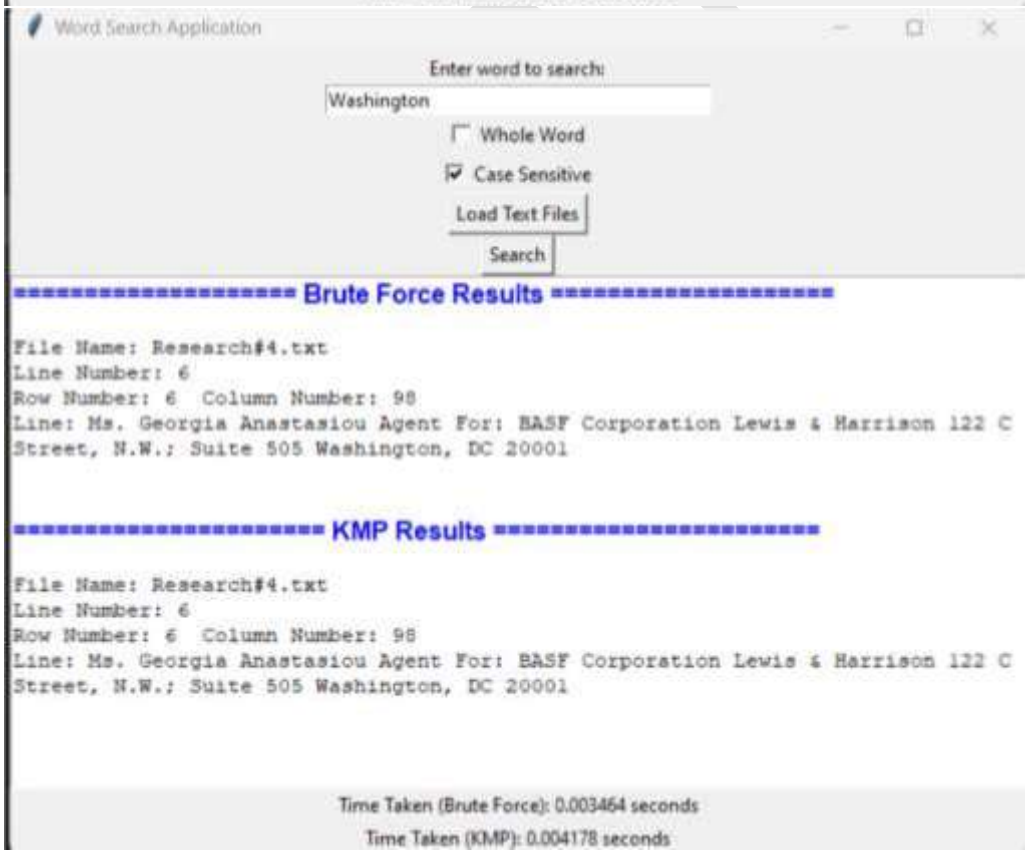
Screenshots




Word Search Application
Enter word to search:
dioxide
Whole Word
Case Sensitive
Load Text Files
Search
========== KMP Results ==========
File Name: Research#10.txt
Line Number: 3
Row Number: 3 Column Number: 9
Line: chlorine dioxide (clo2) has been used for over 100 years to combat all typ
es of bacteria, viruses and fungi successfully.
File Name: Research#10.txt
Line Number: 8
Row Number: 8 Column Number: 9
Line: chlorine dioxide (clo2) is a yellowish gas that, to date, has not been int
roduced into the conventional pharmacopoeia as an active
File Name: Research#10.txt
Line Number: 13
Row Number: 13 Column Number: 107
Line: the recent covid-19 coronavirus pandemic demands urgent solutions with new
Time Taken (Brute Force): 0.005886 seconds
Time Taken (KMP): 0.004407 seconds


Word Search Application
Enter word to search:
dioxide
Whole Word
Case Sensitive
Load Text Files
Search
========== Brute Force Results ==========
File Name: Research#10.txt
Line Number: 3
Row Number: 3 Column Number: 9
Line: chlorine dioxide (clo2) has been used for over 100 years to combat all typ
es of bacteria, viruses and fungi successfully.
File Name: Research#10.txt
Line Number: 8
Row Number: 8 Column Number: 9
Line: chlorine dioxide (clo2) is a yellowish gas that, to date, has not been int
roduced into the conventional pharmacopoeia as an active
File Name: Research#10.txt
Line Number: 13
Row Number: 13 Column Number: 107
Line: the recent covid-19 coronavirus pandemic demands urgent solutions with new
Time Taken (Brute Force): 0.005886 seconds
Time Taken (KMP): 0.004407 seconds


Word Search Application
Enter word to search:
Washington
Whole Word
Case Sensitive
Load Text Files
Search
========== Brute Force Results ==========
File Name: Research#4.txt
Line Number: 2
Row Number: 2 Column Number: 0
Line: washington, dc 20460
File Name: Research#4.txt
Line Number: 6
Row Number: 6 Column Number: 98
Line: ms. georgia anastasiou agent for: basf corporation lewis & harrison 122 c
street, n.w.; suite 505 washington, dc 20001
========== KMP Results ==========
File Name: Research#4.txt
Line Number: 2
Time Taken (Brute Force): 0.007708 seconds
Time Taken (KMP): 0.004172 seconds


Word Search Application
Enter word to search:
Washington
Whole Word
Case Sensitive
Load Text Files
Search
========== Brute Force Results ==========
File Name: Research#4.txt
Line Number: 6
Row Number: 6 Column Number: 98
Line: Ms. Georgia Anastasiou Agent For: BASF Corporation Lewis & Harrison 122 C
Street, N.W.; Suite 505 Washington, DC 20001
========== KMP Results ==========
File Name: Research#4.txt
Line Number: 6
Row Number: 6 Column Number: 98
Line: Ms. Georgia Anastasiou Agent For: BASF Corporation Lewis & Harrison 122 C
Street, N.W.; Suite 505 Washington, DC 20001
Time Taken (Brute Force): 0.003464 seconds
Time Taken (KMP): 0.004178 seconds