

Requirements Specification

Introduction

The Plagiarism Checker is a web application where a user will be able to submit some text through a text box on their screen. After pressing the submit button, the program will execute and output the results to their screen. The output will include the website, if any, containing text matching the text from the input, the percentage of the input text that was found to be matching, and a list of the matching text. The intended customers will be teachers whose students submit their work electronically. Many, if not all, students will refer to texts found on the internet and teachers will need to verify that the work the students submit will not be plagiarized from these texts. The Plagiarism Checker will provide an easy to use way for teachers to check whether or not the student's work contains plagiarism.

User Requirements Specification

1. The user shall be able to input text and submit it to determine if there is any plagiarism.
2. The application shall find websites online with some text matching the text given by the user. *(The user will need to know from which website a text was plagiarized from to justify an accusation of plagiarism.)*
3. The application shall find all text from each website matching the text that was submitted by the user. *(A user will need to know which specific text was matching and determine themselves if it is actually plagiarism or just a phrase that is difficult to word in any other way.)*
4. The application shall determine the percentage of the user's submitted text that is plagiarized from each website. *(A percentage statistic may be useful to see how much of a text is plagiarized from one site.)*
5. At the top of the results page shall be a button that will take the user back to the webpage where they can input text and submit it. *(A user will usually have to check more than one piece of text every time the Plagiarism Checker is used.)*

6. The results shall be shown to the user on a new webpage after pressing the button to submit their text. *(A user should be shown the results to justify an accusation of plagiarism.)*
7. The Plagiarism Checker should not take longer than one minute per 200 words to show the results after the user submits the text. *(A user may become impatient and give up using the application if it takes too long to get results.)*

Figure 1: Sequence diagram for the Plagiarism Checker application.

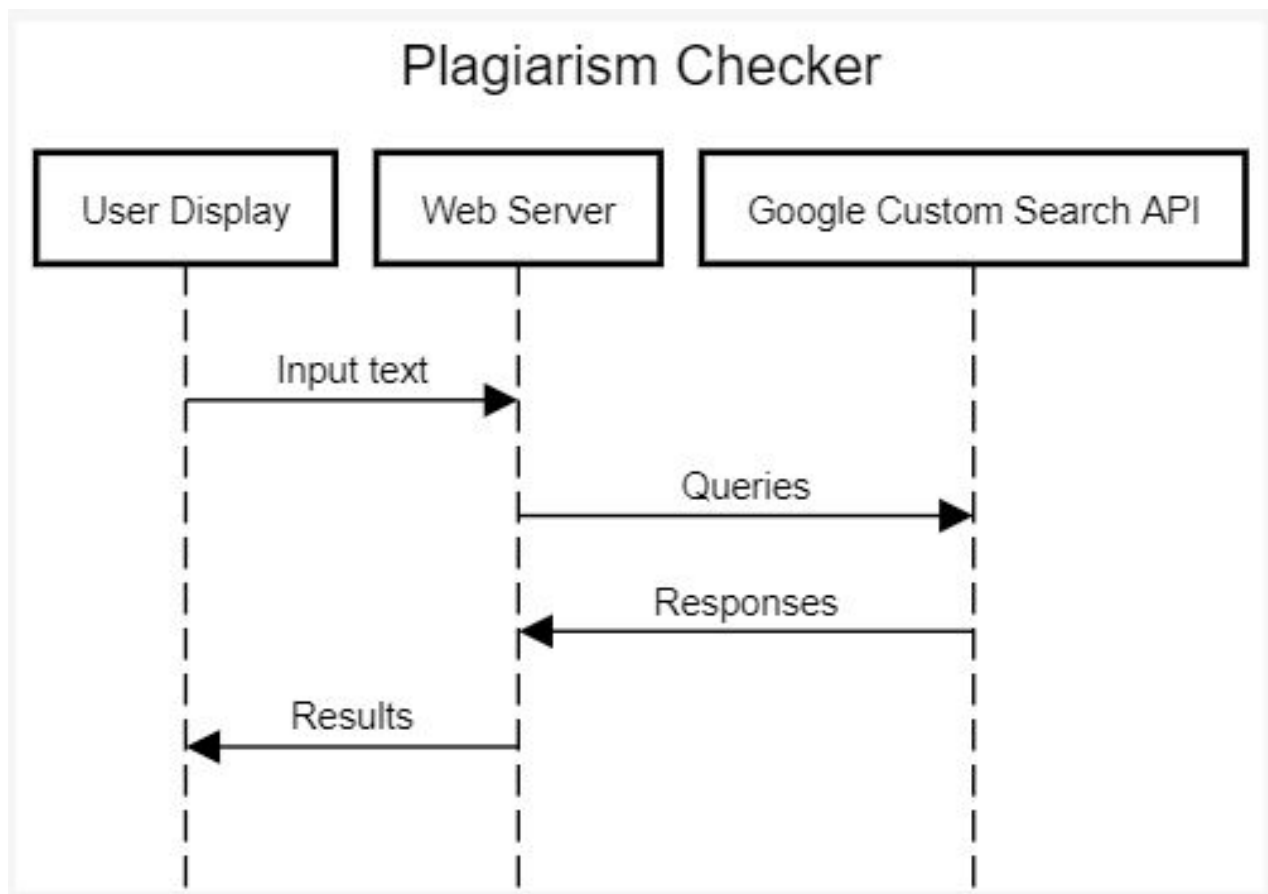


Figure 2: Structured specification for requirement 2.

Function	Find websites where input text is plagiarized from.
Description	Search the entire web for potential texts where the user inputted text has plagiarized from.
Inputs	Text inputted by user.
Source	User submitted text via text box.
Output	List of potential websites.
Destination	Store in an array for more processing.
Action	Search through the entire web using the Google Custom Search API with the user input as a query. For each website returned, store the URL into an array.
Requires	The user input, URLs array, and use of the Google Custom Search API.
Precondition	The user to have submitted some text to be processed.
Postcondition	The array of URLs is populated.
Side effects	None.

Figure 3: Structured specification for requirement 3.

Function	Determine the matches from each website.
Description	Use a matching algorithm to record any substrings of the input text matching any substrings of each website stored in the URLs array.
Inputs	Text inputted by user and a URL.
Source	Input text from user and URL from array of URLs found using Google Custom Search API.
Output	Array of substrings of the input text matching substrings of the website's text.
Destination	Array whose index is the same as the index of the URL in the URLs array.
Action	For each URL in the URLs array, get the text from the HTML. Using that text and the input text, use a matching algorithm to find all the common substrings. Store the array of common substrings in an array whose index is the same as the index of the URL in the URLs array.
Requires	User input, URLs array, common substrings array.
Precondition	Text to have been inputted. URLs array to have been populated.
Postcondition	Common substrings array is populated.
Side effects	None.

Figure 4: Structured specification for requirement 4.

Function	Determine percent match
Description	Computes the percentage of matching text between each website and the input text.
Inputs	Input text, all matching substrings of a website.
Source	User input, array of matching substrings.
Outputs	Percent match.
Destination	Array in same index as the array of matching substrings.
Action	Count words in input text. Count words in all matching substrings. Divide the count of words in matching substrings over count of words in the input text and multiply by 100 to get the percentage.
Requires	Array of matching substrings, input text, array of percent match.
Precondition	User has inputted text, array of URLs has been populated, array of matching substrings for each website has been populated.
Postcondition	Array of percent match has been populated.
Side effects	None.

Figure 5: Structured specification for requirement 5.

Function	Back button.
Description	User will be taken back to the screen where they can input text to be submitted.
Inputs	Back button is clicked.
Source	User clicked back button on user's display screen.
Output	Page where user may input text to be checked for plagiarism.
Destination	User's display screen.
Action	User will press on the back button to take them back to the page where they can input text to check for plagiarism.
Requires	User to have already submitted input and gotten results.
Precondition	Display is of results screen.
Postcondition	Display is of the text input screen.
Side effects	Display of previous results is lost.

Figure 6: Structured specification for requirement 6.

Function	Output results
Description	Output the URL, percent match, and list of matching substrings for each website the application has determined the input text to have been plagiarized from.
Inputs	Website URL, list of matching substrings, percent match
Source	Array of URLs, array of matching substrings, array of percent match.
Output	Each website there is a match, the list of substrings determined to be plagiarism, the percentage of the input text that matches the website.
Destination	User's display.
Action	For each website, output the URL, the percent match, and the list of substrings determined to be plagiarism.
Requires	URLs array, matching substrings array, percent match array.
Precondition	URLs array to have been populated, matching algorithm to have found list of all matching substrings for that website and stored in an array, percent match array to have been populated.
Postcondition	User's display show the results.
Side effects	None.

System Requirements Specification

- 1.1 The user shall be provided an option to input text using a text box.
- 1.2 The user shall be provided an option to input text using a file upload button.
- 1.3 If a file was uploaded, the contents of it shall be displayed in the text box.
- 1.4 The user shall submit the text to be checked for plagiarism with a submit button.

- 2.1 The application shall take the input text and search through the entire web for texts that the input text has potentially plagiarized from.
- 2.2 The application shall take the top websites and place them into an array.

- 3.1 The application shall, for each website, use a matching algorithm and find all matching substrings between the input text and the text from the website.
- 3.2 If the matching substrings contains at least 6 contiguous matching words, place them into an array.
- 3.3 At the end of the matching algorithm, place the array of matching substrings into an array at the index equal to the index where the URL for that website is placed.

- 4.1 Make a count of how many words there are in the input text.
- 4.2 For each website, make a count of all words in every matching substring. Divide this count by the count of words in the input text and multiply by 100 to find the percentage.
- 4.3 For each website, place the percentage into an array whose index is equal to the index of the website in the URLs array.

- 5.1 For each website in the URLs array, display to the user's screen the URL.
- 5.2 Under each URL, display the percent match between the input text and that website.
- 5.3 Under the percent match, display the list of all matching substrings containing at least 6 contiguous matching words.

Platform Requirements Specification

- A browser which supports HTML5 should be used on the user's machine. Older versions may work but is not recommended.
- Web server to host PHP version 7.4. Older versions may work but is not recommended.
- Network connection between the user's machine and the web server hosting the PHP.
- Network connection between the Google Custom Search API and web server hosting the PHP.