This is a PHP script that takes a Word document and checks it for plagiarism by searching each sentence on Google and comparing it with the results obtained from the search. The script uses the PHPWord library to extract the sentences from the Word document and the cURL library to retrieve the HTML content of the Google search results. It also uses the DOMDocument and DOMXPath classes to parse the HTML content and extract the search results. The similarity between the original document and the searched content is calculated as a percentage and returned as output.

The script starts by setting the maximum execution time to 300 seconds and loading the necessary libraries and starting the session. It then checks if the ID parameter is set in the URL and retrieves the name of the Word document from the database using the ID. It then loads the Word document using the PHPWord library and initializes an empty array to store the sentences. It then loops through each section in the document and each element in the section and extracts the text from the elements. It splits the text into sentences using a regular expression and adds them to the sentences array.

The script then initializes another empty array to store the search results obtained from Google. It loops through each sentence in the sentences array and constructs a Google search URL by encoding the sentence. It then sends a cURL request to the URL and retrieves the HTML content of the search results. It uses the DOMDocument and DOMXPath classes to extract the search results from the HTML content and adds them to the search results array. It then sleeps for 2 seconds to avoid getting blocked by Google.

The script then initializes two more empty arrays to store the plagiarized sentences and their similarity weights. It loops through each sentence in the sentences array and splits it into words. It then loops through each search result in the search results array and compares each word in the sentence with each word in the search result. It calculates a similarity weight for each search result based on the number of words that match between the sentence and the search result. It then checks if the similarity weight is greater than or equal to a threshold weight for plagiarism. If it is, it tags the sentence as plagiarized and adds it to the plagiarized sentences array. It also adds the similarity weight to the similarity weights array. It then breaks out of the loop to avoid checking the remaining search results for the same sentence.

After looping through all the sentences, the script calculates the similarity between the original document and the searched content as the sum of all the similarity weights divided by the total number of sentences in the document, multiplied by 100. It then returns the similarity as output.

