Implementation of Levenshtein Distance Algorithm for Product Search Query Suggestions on Koro Pedang Edutourism E-Commerce

Halimah Tus Sadiah¹,*, Lia Dahlia Iryani¹, Tjut Awalyyah Zuraiyah¹, Yuli Wahyuni¹, Cantika Zaddana¹

¹ Universitas Pakuan, Kecamatan Bogor Tengah, Kota Bogor, Jawa Barat 16129, Indonesia

ARTICLE INFO

Article history:
Received 18 August 2023
Received in revised form 14 October 2023
Accepted 6 March 2024
Available online 3 April 2024

ABSTRACT

Users sometimes write queries that are inaccurate or typos in the product search contained in the Koro Pedang Educational Tourism e-commerce, so the system is not find product search results because the query entered in the system is incorrect. This can frustrate users because they cannot find the product they are looking for, so the users leave the website. According to these problems, it is necessary to suggest a query on the product search function. This is expected to assist users in finding the product they are looking for if there is an error in typing the query. This research purposes were to implement the Levenshtein Distance Algorithm for product search query suggestions on Koro Pedang Educational Tourism e-commerce. The stages of this research, namely the development of the search module, implementation of the Levenshtein Distance Algorithm and testing. The implementation of the Levenshtein Distance Algorithm in the search function for Koro Pedang Educational Tourism e-commerce products, a Suggestion Query is generated for Query typos in the search function with an accuracy value of 90%, Precision 95% and Recall 90.9%. This shows that the performance of the algorithm that has been applied to the search function for query suggestion is very good. The application of the Levenshtein Distance Algorithm gives a positive value to the usability of searching for e-commerce products for Koro Pedang Educational Tourism.

Keywords:
Algorithm; Levenshtein distance; Query suggestion; Searching; Koro Pedang edutourism

1. Introduction

E-commerce is buying or selling transaction, that is conducted online or via the internet [1-5]. E-commerce has become popular since the Covid-19 pandemic among Small and Medium Enterprises (SMEs) and Small, Micro and Medium Enterprises (SMMEs) [5-10]. In addition to SMEs, SMMEs, Cooperatives in Indonesia have also implemented e-commerce to sell their products [11,12]. One of the cooperatives that have implemented e-commerce is the Paramasera Cooperative. The Paramasera Cooperative sells its processed products through the Koro Pedang Educational Tourism e-commerce. In the Koro Pedang Educational Tourism e-commerce, it sells Koro Pedang seed
products and processed products such as cakes, chips, snacks, sago, peanuts, soy sauce, and other processed products of Koro Pedang.

On the Koro Pedang Educational Tourism e-commerce page there is a search form which is the main function is to make it easier for users to find products. But the problem is that sometimes users enter incorrect or typo queries in the search for products contained in the Koro Pedang Educational Tourism e-commerce so that the system cannot find product search results because the query entered in the system is incorrect. This can frustrate users because they cannot find the product they are looking for, so users leave the website. According to these problems, it is necessary to implement an algorithm on the search form to provide query suggestions on the product search function. Query Suggestions help users find the product they are looking for if there is a query typo [13-15]. The addition of the query suggestion feature is done by implementing the algorithm on the search form. The algorithm that will be implemented is the Levenshtein Distance Algorithm.

The Levenshtein Distance Algorithm is an algorithm made by Vladimir Levenshtein in 1965 [16,17]. This algorithm is represented in the form of a matrix to calculate the number of differences in the string, entered by the user with the string contained in the database [18-22]. The final result of the algorithm value is in the lower right corner of the matrix [23-25]. The final result of the matrix calculation shows the number of operations that must be executed by the algorithm. These types of operations are character addition operations, character substitution operations and character deletion operations. Research on the implementation of the Levenshtein Distance Algorithm has been implemented for query suggestions by Khalidah and Sadiah et al., where the algorithm is implemented in a dictionary search form [23-25]. This research purposes to apply the Levenshtein Distance Algorithm for product search query suggestions on Koro Pedang Educational Tourism e-commerce.

2. Methodology

This research step consists of building a search module product were implementing in the Levenshtein Distance Algorithm and test [23].

2.1 Search Module Development

The stage to build the query suggestion feature is to build a search module. The search module contained in e-commerce products is designed to be able to implement the Levenshtein Distance Algorithm. Figure 1 is the design of the search module.
In Figure 1, the user inputs a product query in the E-commerce search form Koro Pedang Educational Tourism. The query inputted by the user becomes pre-processing data where the number of words is symbolized by \( j \) and the initialization value is symbolized by \( n \). Initialize values starting from \( n=1 \). Then the next stage the system will check the value of \( n \) and \( j \). If the value of \( n \) is less than the value of \( j \), the system will match the \( n \)-th word using the Levenshtein Distance Algorithm where adjacent words will be checked. If the value of the Levenshtein Distance Algorithm is less than 1, the system will display a query suggestion. However, if the value of the Levenshtein Distance Algorithm is more than 1, the system will immediately return to the word matching process.
2.2 Implementation of the Levenshtein Distance Algorithm

The algorithm implementation stage is conducted, after there is a search module. Levenshtein Distance Algorithm in pseudocode language is converted into PHP-MYSQLi programming language and then implemented in PHP script search form. Figure 2 is the pseudocode for the Levenshtein Distance Algorithm.

![Pseudocode Levenshtein Distance Algorithm](image)

**Fig. 2.** Pseudocode Levenshtein Distance Algorithm [26,27]

2.3 Testing

The testing phase is conducted to find out, whether the functions that have been built are working properly [28,29]. At the testing stage there is a test scenario where the user must perform tests according to the scenario. Test scenarios for the Query Suggestion features is seen in Table 1.

<table>
<thead>
<tr>
<th>No</th>
<th>Test Scenarios</th>
<th>Query Suggestion</th>
<th>Test Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>User input query by writing the correct e-commerce product query on the search form</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>User input query by writing the wrong e-commerce product query, namely lack of letters in the search form input</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>User input Query with the wrong e-commerce product query writing, namely excess letters in the search form input</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>User input Query with the wrong e-commerce product query writing, namely the letters that are swapped in the search form input</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>User input Query by writing the wrong e-commerce product query, namely inputting other categories of products in the search form input that is not in the database</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Results

The product search form in the e-commerce Educational Tourism is seen in Figure 3. The user enters a query in the product search form. If the product you are looking for is found, the system will display the product you are looking for.

![Figure 3. Display of product search form in e-commerce edutourism](image)

For example, if the user inputs the query “koro” it will display all products containing the word “koro” Figure 4.

![Figure 4. Display of the search results for the “koro” product in the e-commerce edutourism](image)

The search module is implemented with the Levenshtein Distance Algorithm where the pseudocode algorithm is converted into the PHP-MYSQLi programming language. The results of the implementation of the Levenshtein Distance Algorithm on the search form generate query suggestion. For example, the user entered the query incorrectly, that is, entered the query incorrectly. The user inputs the query “Aban” then the system will display a query suggestion in the form of "What do you mean by Abon? (Figure 5).
The computation of the Levenshtein Distance Algorithm for the Query Suggestion query “Aban” to “Abon” is seen in Figure 6.

In Figure 6 it is known that the result of the Levenshtein Distance Algorithm is the value 1 which is in the lower right corner of the matrix. This means that one operation is performed. The type of operation performed depends on where the results are obtained whether from the side or from above or from the diagonal. In Figure 6 the value of 1 is generated from the diagonal where the cost value is 0 so the minimum value is 1. This means that 1 type of operation is performed, namely substitution of A into O so that the query suggestion for the “Aban” query is “Abon”.

The search form that has been implemented with the Levenshtein Distance Algorithm is tested according to the test scenario in Table 1. Table 2 is an example of some of the results of testing the Query Suggestion for Product Search in E-commerce Koro Pedang Educational Tourism E-commerce.
Table 2
Test Results for Product Search Query Suggestions in E-commerce Koro Pedang Edutourism E-commerce

<table>
<thead>
<tr>
<th>No</th>
<th>Test Scenarios</th>
<th>Query Suggestion</th>
<th>Test Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>User input query by writing the correct e-commerce product query on the search form</td>
<td>-</td>
<td>Output: Showing all products containing the word koro</td>
</tr>
<tr>
<td></td>
<td>Query input: koro</td>
<td></td>
<td>Valid</td>
</tr>
<tr>
<td>2</td>
<td>User input query by writing the wrong e-commerce product query, namely lack of letters in the search form input</td>
<td>There is an operation to add letters to the Query Suggestion</td>
<td>Output : Show products according to Query Suggestion, namely Produk Keripik Valid</td>
</tr>
<tr>
<td></td>
<td>Input: kripik</td>
<td>E \rightarrow add</td>
<td>Do you mean “Keripik”?</td>
</tr>
<tr>
<td>3</td>
<td>User input Query with the wrong e-commerce product query writing, namely excess letters in the search form input</td>
<td>There is a letter deletion operation in Query Suggestion</td>
<td>Output : Show products according to Query Suggestion, namely Bibit Products Valid</td>
</tr>
<tr>
<td></td>
<td>Input : Bibiit</td>
<td>I \rightarrow delete</td>
<td>Do you mean “Bibit”?</td>
</tr>
<tr>
<td>4</td>
<td>User input Query with the wrong e-commerce product query writing, namely the letters that are swapped in the search form input</td>
<td>There is a character substitution operation in Query Suggestion</td>
<td>Output: Show products according to Query Suggestion, namely Abon Products Valid</td>
</tr>
<tr>
<td></td>
<td>Input: Aban</td>
<td>A \rightarrow O</td>
<td>Do you mean “Abon”?</td>
</tr>
<tr>
<td>5</td>
<td>User input Query by writing the wrong e-commerce product query, namely inputting other categories of products in the search form input that is not in the database</td>
<td>The algorithm process is conducted, namely Query suggestion obtained from the closest word and the minimum number of operations</td>
<td>Output : Show products according to Query Suggestion, namely Kacang Mentah Products Valid</td>
</tr>
<tr>
<td></td>
<td>Input : Matah</td>
<td></td>
<td>Do you mean “Kacang Mentah”?</td>
</tr>
</tbody>
</table>

In the testing phase, there are 30 queries tested. According to the results of the confusion matrix, there are 20 True Positive (TP) data, 1 False Positive (FP), 2 False Negative (FN) data and 7 True Negative (TN) data. According to the confusion matrix and Eq. (1) to Eq. (3) [30], the resulting accuracy values are 90%, precision is 95% and recall is 90.9%.

\[
\text{Accuracy} = \frac{TP}{TP+FN+TN} \times 100\% = \frac{20+7}{20+1+2+7} \times 100\% = 90\% \tag{1}
\]

\[
\text{Precision} = \frac{TP}{TP+FP} \times 100\% = \frac{20}{20+1} \times 100\% = 95\% \tag{2}
\]

\[
\text{Recall} = \frac{TP}{TP+FN} \times 100\% = \frac{20}{20+2} \times 100\% = 90.9\% \tag{3}
\]
4. Conclusions

According to the implementation of the Levenshtein Distance Algorithm in the search function of Koro Pedang Educational Tourism e-commerce products, a Suggestion Query is generated for Query typos in the search function, namely the accuracy value is 90%, Precision is 95% and Recall is 90.9%. This shows that the performance of the algorithm that has been applied to the search function for query suggestion is very good. The application of the Levenshtein Distance Algorithm gives a positive value to the usability of searching for e-commerce products for Koro Pedang Educational Tourism. This research has only implemented a per-word query search, not yet in a sentence search. It is hoped that there will be further research that can provide Query Suggestions for inputting sentences on the search form.

Acknowledgement

This research was funded by the 2022 Kedaireka Matching Fund Grant. The authors would like to thank Pakuan University and LPPM Pakuan University for supporting this research.

References
