Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

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

Halimah Tus Sadiah[1,*], Lia Dahlia Iryani[1], Tjut Awaliyah Zuraiyah[1], Yuli Wahyuni[1], Cantika Zaddana[1]

[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

Keywords:

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

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.

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

* *Corresponding author.*
E-mail address: sadiahht@unpak.ac.id

https://doi.org/10.37934/araset.42.2.188196

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.

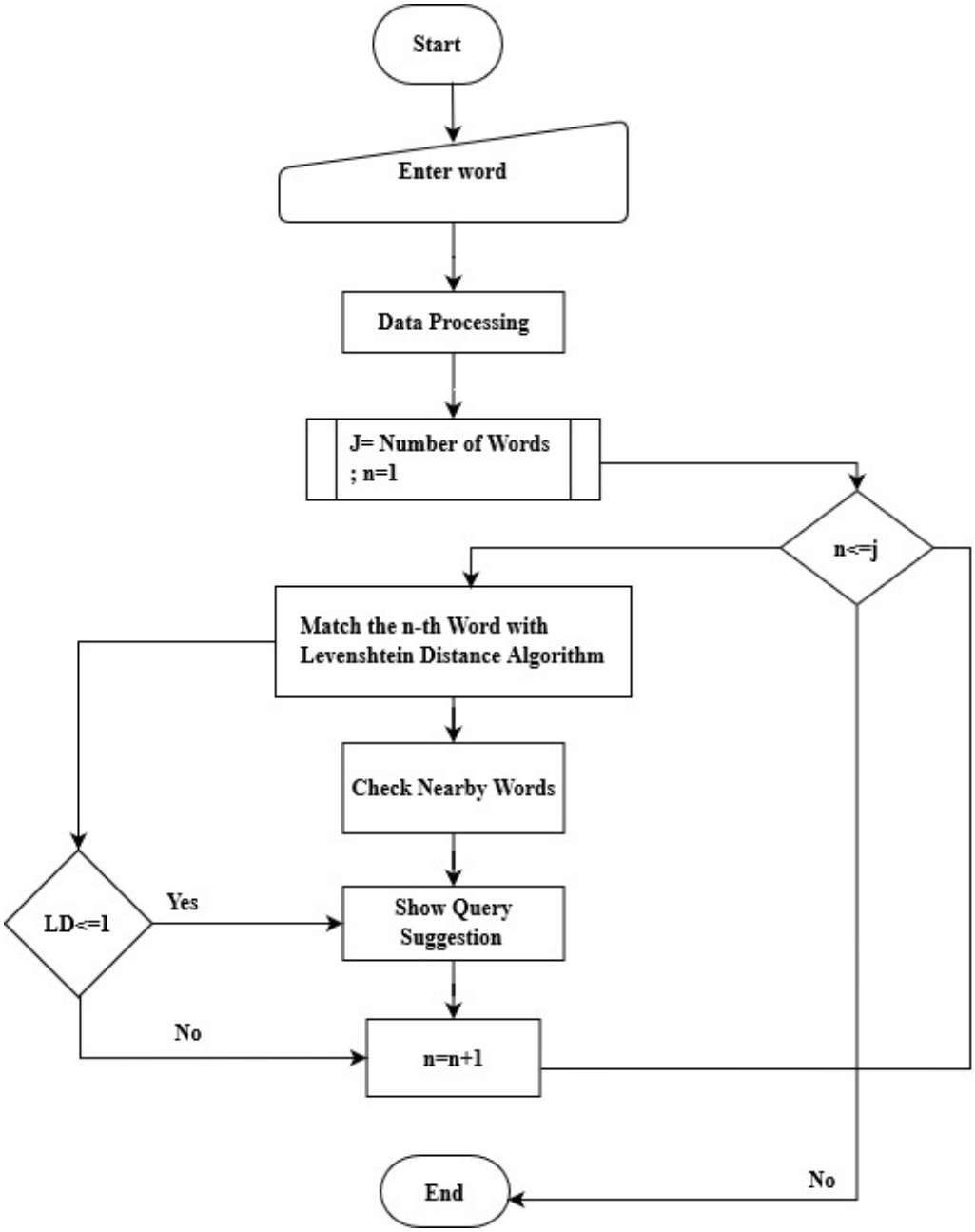

Fig. 1. Search Module Design implemented by Levenshtein Distance Algorithm [23,24]

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.

```
Program Pseudocode Algoritma Levenshtein
Distance
int LevenshteinDistance (char s[1..m], char
t[1..n])
{
declare int d[0..m, 0..n]
declare int cost
  for i from 0 to m d[i,0] :=i
  for j from 0 to n d[0,j] := j
    for j from 1 to n{
      for i from 1 to m {
        if s[i]!=t[j] then cost := 1
        else cost := 0
        d[i,j] := minimum(
          d[i-1,j] + 1,
          d[i, j-1]+ 1,
          d[i-1, j-1] + cost
        )
      }
    }
return d[m,n]
}
```

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 1
Test Scenario of Query Suggestion Feature Search

No	Test Scenarios	Query Suggestion	Test Results
1	User input query by writing the correct e-commerce product query on the search form		
2	User input query by writing the wrong e-commerce product query, namely lack of letters in the search form input		
3	User input Query with the wrong e-commerce product query writing, namely excess letters in the search form input		
4	User input Query with the wrong e-commerce product query writing, namely the letters that are swapped in the search form input		
5	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		

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

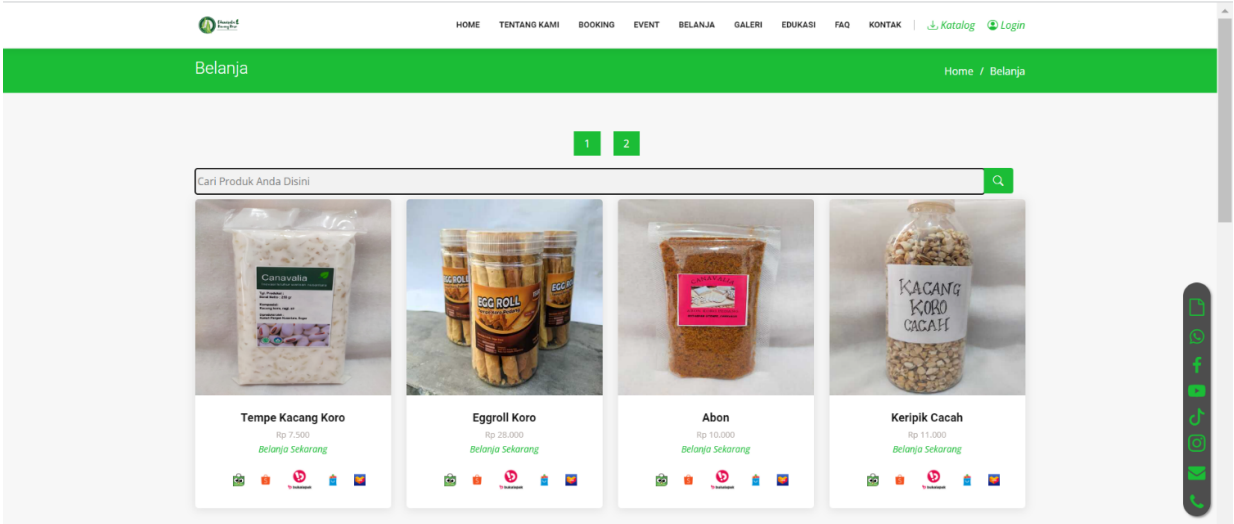

Fig. 3. Display of product search form in e-commerce edutourism

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

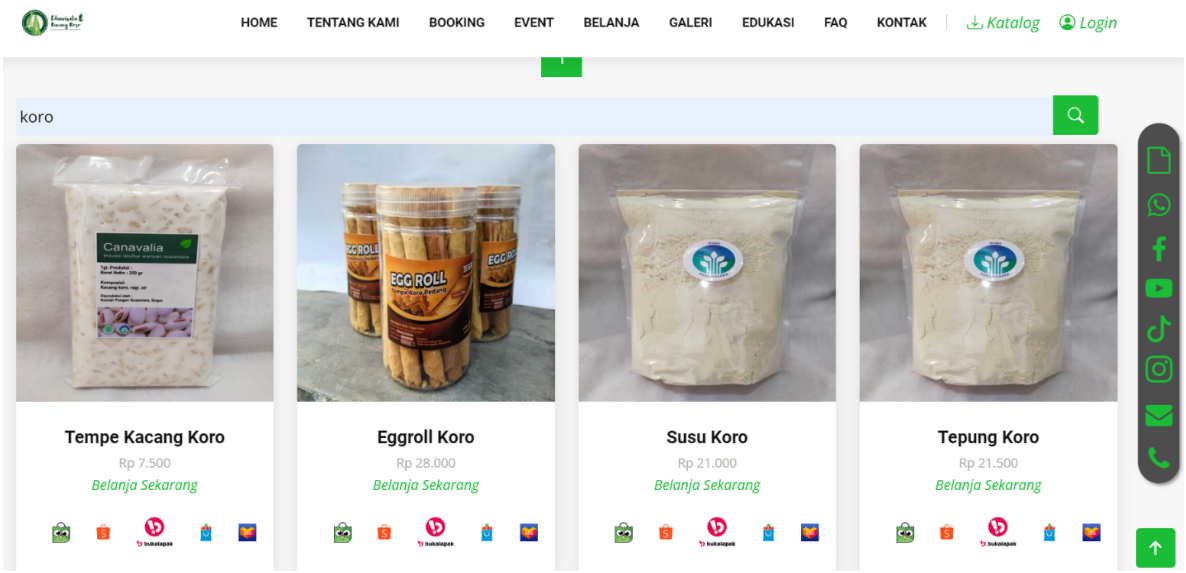

Fig. 4. Display of the search results for the “koro” product in the e-commerce edutourism

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).

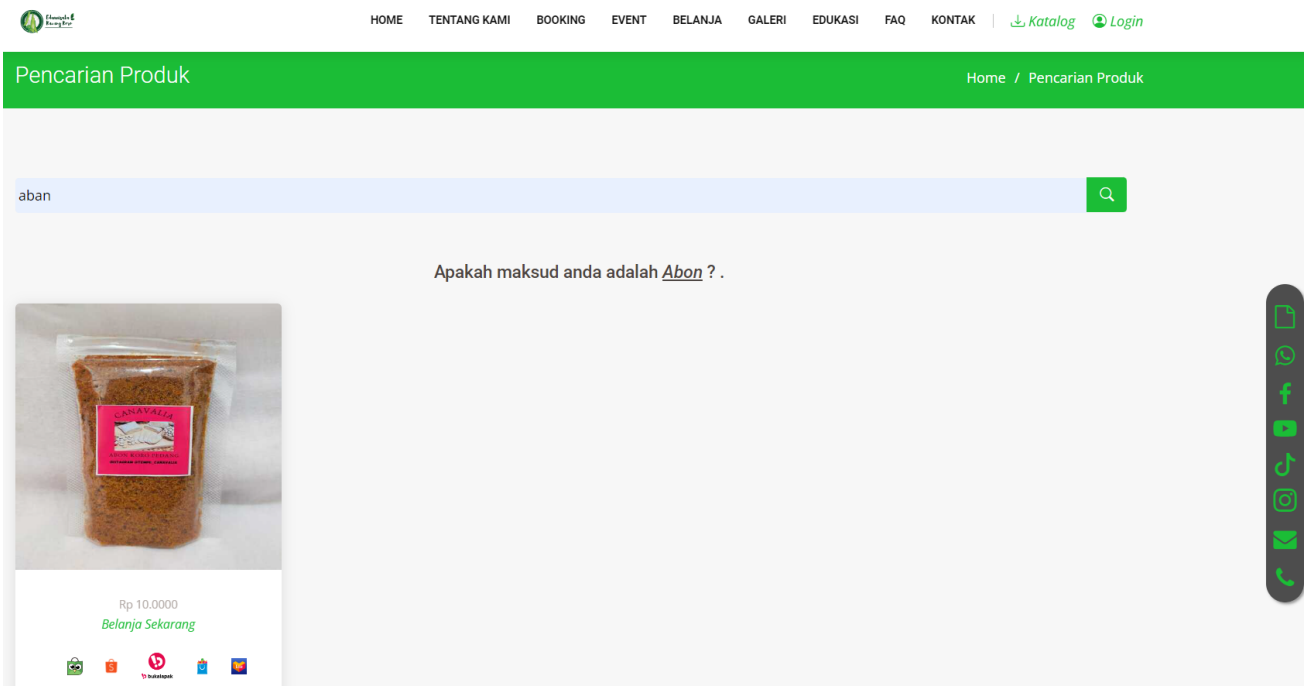

Fig. 5. Display of product search form in e-commerce edutourism

The computation of the Levenshtein Distance Algorithm for the Query Suggestion query "Aban" to "Abon" is seen in Figure 6.

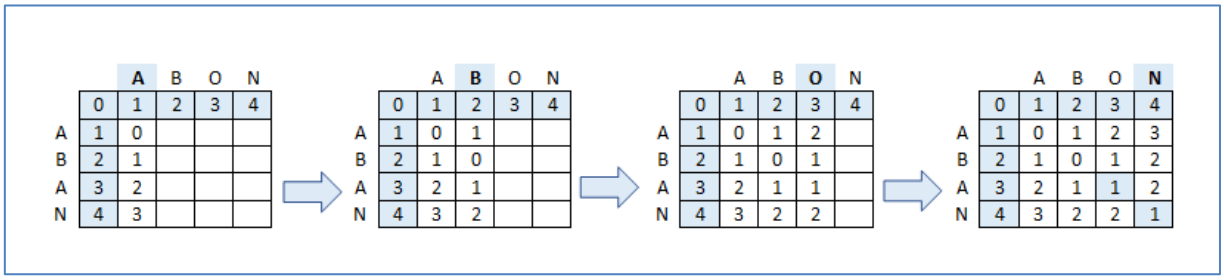

Fig. 6. Levenshtein Distance Algorithm Computing the query "Aban" becomes "Abon"

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

No	Test Scenarios	Query Suggestion	Test Results
1	User input query by writing the correct e-commerce product query on the search form Query input: koro	-	Output: Showing all products containing the word koro Valid
2	User input query by writing the wrong e-commerce product query, namely lack of letters in the search form input Input: kripik	There is an operation to add letters to the Query Suggestion E → add Do you mean "Keripik" ?	Output : Show products according to Query Suggestion, namely Produk Keripik Valid
3	User input Query with the wrong e-commerce product query writing, namely excess letters in the search form input Input : Bibiit	There is a letter deletion operation in Query Suggestion I → delete Do you mean "Bibit"?	Output : Show products according to Query Suggestion, namely Bibit Products Valid
4	User input Query with the wrong e-commerce product query writing, namely the letters that are swapped in the search form input Input: Aban	There is a character substitution operation in Query Suggestion A → O Do you mean "Abon"?	Output: Show products according to Query Suggestion, namely Abon Products Valid
5	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 Input : Matah	The algorithm process is conducted, namely Query suggestion obtained from the closest word and the minimum number of operations Do you mean "Kacang Mentah"?	Output : Show products according to Query Suggestion, namely Kacang Mentah Products Valid

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%.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} * \ 100\% = \frac{20+7}{20+1+2+7} * 100\% = 90\% \quad (1)$$

$$\text{Precission} \ = \frac{TP}{TP+FP} * \ 100\% = \frac{20}{20+1} * 100\% = 95\% \quad (2)$$

$$\text{Recall} \ = \frac{TP}{TP+FN} * \ 100\% = \frac{20}{20+2} * 100\% = 90{,}9\% \quad (3)$$

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

[1] Bawack, Ransome Epie, Samuel Fosso Wamba, Kevin Daniel André Carillo, and Shahriar Akter. "Artificial intelligence in E-Commerce: a bibliometric study and literature review." *Electronic markets* 32, no. 1 (2022): 297-338. https://doi.org/10.1007/s12525-022-00537-z

[2] Haryanti, Tining, and Apol Pribadi Subriadi. "E-commerce acceptance in the dimension of sustainability." *Journal of Modelling in Management* 17, no. 2 (2022): 715-745. https://doi.org/10.1108/JM2-05-2020-0141

[3] Sanny, Lim, Tita Dwi Julianto, Serafim Savionus, and Beni Widarman bin Yus Kelena. "Purchase intention in the fashion industry on local and international E-commerce in Indonesia." *International Journal of Asian Business and Information Management (IJABIM)* 13, no. 2 (2022): 1-12. https://doi.org/10.4018/IJABIM.20220701.oa4

[4] Scutariu, Adrian-Liviu, Ștefăniță Șușu, Cătălin-Emilian Huidumac-Petrescu, and Rodica-Manuela Gogonea. "A cluster analysis concerning the behavior of enterprises with e-commerce activity in the context of the COVID-19 pandemic." *Journal of Theoretical and Applied Electronic Commerce Research* 17, no. 1 (2021): 47-68. https://doi.org/10.3390/jtaer17010003

[5] Fernández-Bonilla, Fernando, Covadonga Gijón, and Bárbara De la Vega. "E-commerce in Spain: Determining factors and the importance of the e-trust." *Telecommunications Policy* 46, no. 1 (2022): 102280. https://doi.org/10.1016/j.telpol.2021.102280

[6] Cassia, Fabio, and Francesca Magno. "Cross-border e-commerce as a foreign market entry mode among SMEs: the relationship between export capabilities and performance." *Review of International Business and Strategy* 32, no. 2 (2022): 267-283. https://doi.org/10.1108/RIBS-02-2021-0027

[7] Tolstoy, Daniel, Emilia Rovira Nordman, and Uyen Vu. "The indirect effect of online marketing capabilities on the international performance of e-commerce SMEs." *International Business Review* 31, no. 3 (2022): 101946. https://doi.org/10.1016/j.ibusrev.2021.101946

[8] Kawasaki, Tomoya, Hisayuki Wakashima, and Ryuichi Shibasaki. "The use of e-commerce and the COVID-19 outbreak: A panel data analysis in Japan." *Transport Policy* 115 (2022): 88-100. https://doi.org/10.1016/j.tranpol.2021.10.023

[9] Witek-Hajduk, Marzanna Katarzyna, Anna Małgorzata Grudecka, and Anna Napiórkowska. "E-commerce in the internet-enabled foreign expansion of Polish fashion brands owned by SMEs." *Journal of Fashion Marketing and Management: An International Journal* 26, no. 1 (2022): 51-66. https://doi.org/10.1108/JFMM-10-2020-0225

[10] Hussain, Arsalan, Minhas Akbar, Arfan Shahzad, Petra Poulova, Ahsan Akbar, and Rohail Hassan. "E-commerce and SME performance: The moderating influence of entrepreneurial competencies." *Administrative Sciences* 12, no. 1 (2022): 13. https://doi.org/10.3390/admsci12010013

[11] Puspasari, Novita, Kiky Sri Rejeki, Intan Shaferi, and Ratno Purnomo. "Adopsi E-Marketing oleh Koperasi dan Usaha Kecil dan Menengah (UKM) di Kabupaten Banyumas." *Universitas Jenderal Soedirman* (2016).

[12] Azmi, Meri, Yance Sonatha, Rasyidah Rasyidah, Indri Rahmayuni, and Fazrol Rozi. "Pelatihan Pemanfaatan Teknologi E-Commerce pada Koperasi." *Jurnal Abdimas: Pengabdian dan Pengembangan Masyarakat* 2, no. 2 (2020): 19-23.

[13] Chen, Wanyu, Fei Cai, Honghui Chen, and Maarten De Rijke. "Personalized query suggestion diversification in information retrieval." *Frontiers of Computer Science* 14 (2020): 1-14. https://doi.org/10.1007/s11704-018-7283-x
[14] Zhong, Jianling, Weiwei Guo, Huiji Gao, and Bo Long. "Personalized query suggestions." In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1645-1648. 2020. https://doi.org/10.1145/3397271.3401331
[15] Mustar, Agnès, Sylvain Lamprier, and Benjamin Piwowarski. "On the study of transformers for query suggestion." *ACM Transactions on Information Systems (TOIS)* 40, no. 1 (2021): 1-27. https://doi.org/10.1145/3470562
[16] Saputra, Reza, Sapri Sapri, and Ricky Zulfiandry. "Book Catalog Application In Man 2 Bengkulu City Using The Levenstein Distance Algorithm." *Jurnal Media Computer Science* 1, no. 2 (2022): 105-116.
[17] Sander, Vobey Alex, Liza Yulianti, and Yupianti Yupianti. "The Implementation of Data Mining to Get The Pattern for Selecting Students' Thesis Title." *Jurnal Komputer, Informasi dan Teknologi (JKOMITEK)* 1, no. 2 (2021): 165-173. https://doi.org/10.53697/jkomitek.v1i2.200
[18] Diyasa, I. Gede Susrama Mas, Gideon Setya Budiwitjaksono, Alfiatun Masrifah, and Muhammad Rif'an Dzulqornain. "Word Search on the" SITUK" Application Using the Levenshtein Distance Algorithm." *Internasional Journal of Data Science, Engineering, and Anaylitics* 1, no. 2 (2021): 55-62. https://doi.org/10.33005/ijdasea.v1i2.13
[19] Setiabudi, Reza, Ni Made Satvika Iswari, and Andre Rusli. "Enhancing text classification performance by preprocessing misspelled words in Indonesian language." *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 19, no. 4 (2021): 1234-1241. https://doi.org/10.12928/telkomnika.v19i4.20369
[20] Berger, Bonnie, Michael S. Waterman, and Yun William Yu. "Levenshtein distance, sequence comparison and biological database search." *IEEE transactions on information theory* 67, no. 6 (2020): 3287-3294. https://doi.org/10.1109/TIT.2020.2996543
[21] Soyusiawaty, Dewi, and Denny Hilmawan Rahmatullah Wolley. "Hybrid spelling correction and query expansion for relevance document searching." *International Journal of Advanced Computer Science and Applications* 12, no. 8 (2021). https://doi.org/10.14569/IJACSA.2021.0120838
[22] Logan, Robert, Zoe Fleischmann, Sofia Annis, Amy Wangsness Wehe, Jonathan L. Tilly, Dori C. Woods, and Konstantin Khrapko. "3GOLD: optimized Levenshtein distance for clustering third-generation sequencing data." *BMC bioinformatics* 23, no. 1 (2022): 1-18. https://doi.org/10.1186/s12859-022-04637-7
[23] Khalidah, Khalidah. "Implementation of Levenshtein Distance Algorithm in the Digital Biology Dictionary Search Function." *Indonesian Journal of Artificial Intelligence and Data Mining* 4, no. 2: 89-96.
[24] Sadiah, H. T., M. S. N. Ishlah, and N. N. Rokhmah. "Query Suggestion on Drugs e-Dictionary Using the Levenshtein Distance Algorithm." *Lontar Komputer* 10, no. 3 (2019): 193-202. https://doi.org/10.24843/LKJITI.2019.v10.i03.p07
[25] Sadiah, Halimah Tus, Muhamad Saad Nurul Ishlah, and Nisa Najwa Rokhmah. "Autocorrect on Drugs e-Dictionary Search Module Using Levenshtein Distance Algorithm." *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)* 4, no. 1 (2020): 64-69. https://doi.org/10.29207/resti.v4i1.1401
[26] Syarafina, Nadhia Nurin, Jozua Ferjanus Palandi, and Nira Radita. "Designing a word recommendation application using the Levenshtein Distance algorithm." *Matrix: Jurnal Manajemen Teknologi dan Informatika* 11, no. 2 (2021): 63-70. https://doi.org/10.31940/matrix.v11i2.2419
[27] Oktaviyani, Enny Dwi, Sherly Christina, and Deddy Ronaldo. "Keywords search correction using Damerau Levenshtein distance algorithm." In *Conference SENATIK STT Adisutjipto Yogyakarta*, vol. 5, pp. 167-176. 2019. https://doi.org/10.28989/senatik.v5i0.344
[28] Sadiah, Halimah Tus. "Usability Testing on Android-based KMS for Pregnant Women using the USE Questionnaire." *International Journal of Quantitative Research and Modeling* 1, no. 3 (2020): 164-173. https://doi.org/10.46336/ijqrm.v1i3.61
[29] Sadiah, Halimah Tus. "Website management training and ICT literacy evaluation of Kelurahan Ciwaringin Bogor authorities." *International Journal of Quantitative Research and Modeling* 1, no. 2 (2020): 67-75. https://doi.org/10.46336/ijqrm.v1i2.26
[30] Ramli, Nurul Ehsan, Zainor Ridzuan Yahya, and Nor Azinee Said. "Confusion Matrix as Performance Measure for Corner Detectors." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 29, no. 1 (2022): 256-265. https://doi.org/10.37934/araset.29.1.256265