Search Optimization of PIP Scholarship Recipients In Web-Based Student Data Application Using The Levenshtein Distance Algorithm

Yoga Handoko Agustin1,*, Yosep Septiana2, Arbi Yuan Aspahany3

1,2,3) Institut Teknologi Garut
1)yoga.handoko@itg.ac.id, 2)yoseptiana@itg.ac.id, 3)1906113@itg.ac.id

Submitted : Aug 17, 2023 | Accepted : Aug 28, 2023 | Published : Oct 1, 2023

Abstract: Realizing that education is very important, the government supports every citizen to get education. One of the government programs is the Smart Indonesia Program. PIP is a scholarship designed to help school-age children from poor/vulnerable families to continue to receive education services, both through formal elementary to high school/vocational schools and non-formal pathways from package a to package c and special education. SDN II Babakanloa has not been touched by technology for processing student data. So that the student section has difficulties in recording and updating student data. Student names have unique identities and errors often occur in typing the keywords to be searched. This results in an information that is desired or sought can not be found. Therefore we need a web-based data application that can provide keyword corrections in searching for student names. This study aims to create a web-based student data application by optimizing corrections to typing keywords searched by implementing the Levenshtein Distance Algorithm and also making it easier to process and search student data. The development method used is the Rational Unified Process (RUP) with the stages of Inception, Elaboration, Construction, and Transition. Designed using the CodeIgniter Framework with the PHP and JavaScript programming languages. The application of the Levenshtein Distance Algorithm can optimize the search for student data and reduce the occurrence of search errors by School Operators. The application of the Levenshtein Distance Algorithm produces a very good accuracy rate of 94% of the results of student data correction. accordance with the expectations of the School Operator. So it shows that the application of the Levenshtein Distance Algorithm is appropriate to use in optimizing the search.

Keywords: keyword correction, levenshtein distance, RUP, scholarship, PIP

INTRODUCTION

Realizing that education is very important, the government supports every citizen to get an education. One of the government programs is the Smart Indonesia Program. PIP is a scholarship designed to help school-age children from poor, vulnerable, or priority families continue to receive education services until the end of secondary education, both through formal elementary to high school or vocational schools and non-formal pathways from package A to package C and special education. SDN II Babakanloa is an educational institution located at KP. Sirnarasa Desa Babakanloa Kec. Pangatikan, Garut Regency. Therefore, technology has not been touched for processing student data, which is carried out at SD Negeri II Babakanloa, which still uses ledgers. Data processing like this results in the student department often experiencing difficulties in recording and updating student data.

*Yoga Handoko Agustin

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.
Which can result in a lot of student data being the same and not being updated. Under these conditions, it is difficult to distribute PIP scholarships (Smart Indonesia Program) because the latest data from students is not known. Data on students who were previously entitled to get scholarships are not entitled to get scholarships, and vice versa is very dependent on the latest data from students and the number of students who excel at SD Negeri II Babakanloa but are unable to pay school fees because parents' income is not balanced with the number of dependents. as well as the difficulty of the student section getting the latest student data. Therefore, a web-based application is needed to make it easier for the student section to record, update, and search for the latest data from students. Web-based applications have ease in terms of installation and access. Website-based applications are easier for users to use because accessing a browser is easy to do. In contrast to other mobile and desktop-based systems that must be installed first.

In searching for data, especially student data that has unique identities, sometimes errors occur in typing the words to be searched, resulting in the desired or sought information not being found. Therefore, we need an approach that can later provide corrections for input errors. The Levenshtein Distance Algorithm is an algorithm commonly used to overcome typing errors in the word being searched for. The Levenshtein Distance Algorithm works by finding the distance between the words typed by the user and the words in the database so that they can display the corrections. The first reference research was conducted by Apriliani with the theme Implementation of the Levenshtein Distance Algorithm for Searching Thesis Titles and Adding Autocomplete Features to the Thesis Repository Website for the Informatics Engineering Department, University of Palangka Raya. This research examines the accuracy of the system in detecting spelling errors in input, and the results show that the use of the Levenshtein distance algorithm can improve accuracy in detecting spelling errors by measuring the distance between the words produced by students and the correct word, as well as providing more accurate correction suggestions. (Aprilian et al., 2021). Furthermore, the second reference research is by Isbahiakana Larasati Marisa with the theme of Application of the Levenshtein Distance Algorithm in Mail Processing Applications. Research to measure the value of similarity between two words (strings) So that it can make it easier to search for archives as well as operate certain operations. (Larasati & Marisa, 2019).

The research from Noor Kamala Sari and colleagues with the theme Application of the Levenshtein Distance Algorithm for Searching in the Library Information System, Faculty of Medicine, University of Palangka Raya. applied to calculate the difference between each book title or author name in library data with user query input. This algorithm calculates the minimum number of delete, addition, or substitution operations needed to change one string into another. (Noor Kamala Sari et al., 2019). Research from Cynthia Natalie with the theme Optical Character Recognition Using Ulpath and Certificate Data Matching. With the Levenshtein Distance Algorithm, this researcher tests the accuracy of the input as well as for matching strings, where file comparisons are made between the input data and the OCR result data by using certain operations such as the character insertion operation (Natalie, 2023).

Research from M.Fahrur and his friends with the theme of Application of the Levenshtein Distance Algorithm there is an English Teaching Support Application for Teachers. This study tested the accuracy in checking essays on students' assignments in the preprocessing stages of the text. This process includes the steps of tokenization, cleaning, deleting common words (stopwords), stemming, and sorting. To obtain text similarity scores between students, calculations are run using the Levenshtein Distance algorithm with a text similarity threshold of 70%. If the text similarity between students is less than 70%, the document will be accepted by the system. The text similarity exceeds 70%, then the document will be rejected. (Azhri et al., 2019).

Comparison of all references and research conducted by the author, namely the goal of implementing the Levenshtein Distance Algorithm which is different with the same algorithm function. The algorithm commonly used to overcome typing errors in the word you are looking for, the levenshtein distance algorithm works by finding the distance between the words typed by the user and the words in the database so that they can display the corrections. but maybe each study has different weaknesses when testing this algorithm for the purpose of the research. As well as the author's research that was conducted only focused on focusing on optimizing the search for PIP scholarship recipients on web-based student data applications using the Levenshtein Distance Algorithm without applying other applications such as previous studies.

*Yoga Handoko Agustin
Based on the problems and research references mentioned above, the title "Optimizing the Search for PIP Scholarship Recipients in Web-Based Student Data Applications with the Levenshtein Distance Algorithm" is obtained.

**LITERATURE REVIEW**

**Smart Indonesia Program**

The Smart Indonesia Program (PIP) is an educational program launched by the Ministry of Education and Culture of the Republic of Indonesia. The aim of this program is to improve access to and quality of education in Indonesia, especially for students from underprivileged families. PIP provides various types of educational assistance to students who meet certain criteria. Some of the forms of assistance provided include:

1. Assistance Program for Underprivileged Students (BSKM): This program is provided to students who come from families with economic limitations to support their educational needs, including school fees, study materials, uniforms, and other learning equipment.
2. School Operational Assistance (BOS): This assistance is provided to schools as support in providing educational facilities and services to students. BOS funds can be used to finance school activities, such as purchasing books, stationery, infrastructure improvements, and other activities.
3. Smart Indonesia Program (PIPC): This program provides additional educational assistance to students who excel and have high academic potential. PIPC includes scholarships, support for gifted students, and academic competency development programs.

Through the Smart Indonesia Program, the government seeks to increase educational opportunities for all Indonesian children, especially those from underprivileged families. This program aims to create wider access, equalize social inequality, and improve the quality of education in this country.

Variables related to the Smart Indonesia Program for schoolchildren include:

1. Economic Condition: This variable is used to determine the eligibility of students to receive PIP assistance. This program is aimed at students from underprivileged families who meet certain income criteria.
2. School Attendance: PIP also considers student attendance at school as an assessment variable. Students who have good attendance tend to be more likely to benefit from this program.
3. Academic Achievement: Academic achievement can also be a factor to be considered in the provision of PIP assistance. While not the sole determining factor, good academic performance can increase a student's chances of benefiting from the program.
4. Education Level: The Smart Indonesia Program applies to students from elementary (SD) to secondary (SMP/SM/A/SMK) education levels. This variable determines who is eligible to receive PIP assistance based on their level of education.
5. Students who hold KIP
6. Students from poor or vulnerable families and/or with special considerations such as:
   a. Students from poor or vulnerable families and/or with special considerations such as:
   b. Students from families participating in the Family Hope Program
   c. Students from families who hold the Prosperous Family Card
   d. Students who are orphans from schools, social institutions, or orphanages
   e. Students affected by natural disasters
   f. Students who do not attend school (drop out) are expected to return to school.
   g. Students who experience physical abnormalities, victims of natural disasters, from parents who have had their employment terminated, in conflict areas, from families of convicts, are in correctional institutions, or have more than three siblings living in the same house
   h. Participants in course institutions or other non-formal education units. (Kemendikbud, n.d.)

**Levenshtein Distance Algorithm**

Levenshtein Distance was made by Vladimir Levenshtein in 1965. The edit distance calculation is obtained from the matrix used to calculate the number of differences between the two strings. The calculation of the distance between these two strings is determined by the minimum number of change

*Yoga Handoko Agustin*
operations needed to make string A into string B. There are three main types of operations that can be performed by this algorithm, namely:

1. Character Change Operation
   The character change operation is an operation to exchange a character with another character; for example, the writer writes the string 'yamg' to 'yang'. In this case the character 'm' is replaced by the letter 'n'.

2. Character Addition Operation
   The operation of adding characters means adding characters to a string. For example, the string "topad" becomes the string "to", and the character "a" is added at the end of the string. The addition of characters is not only done at the end of the word but can be added at the beginning or inserted in the middle of the string.

3. Character Erase Operation
   The character deletion operation is performed to remove characters from a string. For example, the string "new" removes the last character so that it becomes a "new" string. In this operation, the character "r" is deleted.

This algorithm starts from the top left corner of a two-dimensional array that has been filled with a number of initial and target string characters and is given a cost value. The cost value at the bottom right becomes the edit distance value, which describes the difference between the two strings.

![Figure 1. Edit Distance Calculation Matrix Table](image)

An example of levenshtein calculations using two different strings and then calculating the edit distance is in Figure 1. It can be seen that the results of calculating the edit distance between the two strings "sya" and "me" are 1. The check starts with the initial iteration of the two strings, then an addition operation is performed, followed by insertion and deletion. The edit distance value is at the bottom right of the matrix.

There is only one insertion process that is carried out, namely the insertion of the character "a" in the string "sya" so that it becomes "me". In the case of spell checking, this calculation process is carried out by a number of words in the database. Of course, for the best advice, a complete list of Indonesian words is needed. So that the suggested word can be close to what is expected by the user. The Levenshtein distance formula is a method used to measure the degree to which two strings or sequences of characters differ from one another.

The Levenshtein distance is the minimum number of insertion, deletion, and substitution operations required to convert one string into another.

**METHOD**

This research is an activity stage in the process of designing a system which is presented in the form of diagrams, so that it can be easily read and understand the direction in outline of the logic flow of a study. These stages can be seen in the picture 2.
Figure 2. Conceptual
The description of the stages in Figure 2 is as follows:

1. Inception
   In the first stage, it begins with interviews with school operators and then generates a problem statement. After that, determine the identification of business processes in the system. After identifying business processes, determine system specifications, which ensure that the system built meets user needs and can run properly. After that, enter the identification of the actor, which identifies the actor, in the application.

2. Elaboration
   In the second stage, namely the process of making documentation for programmers, such as use case diagrams, activity diagrams, and sequence diagrams, after that, it goes into the design of solving the Levenshtein Distance algorithm, in correcting user input, and after that, it is compared by the Levenshtein Distance algorithm so that you get the output/data you are looking for.

3. Construction
   Construction is the stage where the programmer will implement the elaboration results into the code structure. In this RUP method, the programmer will carry out software testing simultaneously in the construction stage, which is tested not only for the functional system but also for the application of the algorithm, whether it is successful or not, and how much accuracy the algorithm has in optimizing search.

4. Transitions
   At the transition stage of testing by the user, if there are problems during testing by the user, the system will be repaired, but if during testing by the user there are no problems, then the application will be deployed or released.

RESULT

Inception
A requirement on the design of search optimization for PIP scholarship recipients on student data applications with the Levenshtein Distance Algorithm is to be made. Details of the requirements that must be included in the application are display requirements and system requirements.

   a. Display requirements, this application is expected to provide an admin view that is used to manage data and is also expected to provide this web-based archiving feature. This application must have a menu for login and logout.
   b. System Requirements to meet the needs of users Search Optimization for PIP Scholarship Recipients in Student Data Applications with the Levenshtein Distance Algorithm has the following specifications:

<table>
<thead>
<tr>
<th>No</th>
<th>System requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The system can be accessed by admin</td>
</tr>
<tr>
<td>2</td>
<td>The system can display search correction results using the Levenshtein Distance algorithm</td>
</tr>
<tr>
<td>3</td>
<td>The system can login</td>
</tr>
<tr>
<td>4</td>
<td>The system can logout</td>
</tr>
<tr>
<td>5</td>
<td>The system can display student data information</td>
</tr>
<tr>
<td>6</td>
<td>The system can display teacher data information</td>
</tr>
<tr>
<td>7</td>
<td>The system can add Teacher and Student data</td>
</tr>
<tr>
<td>8</td>
<td>The system can delete teacher and student data</td>
</tr>
<tr>
<td>9</td>
<td>The system can change teacher and student data</td>
</tr>
</tbody>
</table>

*Yoga Handoko Agustin
Actor identification
Actor identification was carried out during the identification process. The result was that one actor was identified, namely the admin, who functions as a manager and receives information about student data in the system.

<table>
<thead>
<tr>
<th>No</th>
<th>Actor</th>
<th>Type of activity</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>School operator</td>
<td>Who manages the System</td>
<td>1. Login to enter the system.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2. Add student data</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>3. Deleting student data</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4. Edit student data</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5. Do a student data search</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>6. logout</td>
</tr>
</tbody>
</table>

Elaboration
At this stage, a system design activity is carried out, namely use case diagrams, activity diagrams, sequence diagrams, and class diagrams, as well as designing the interface (interface) of the system to be created.

Figure 3. Use case diagram for search optimization in Web-Based Student Data Applications With Levenshtein Distance Algorithm

*Yoga Handoko Agustin
The design phase defines the static structure of the system, such as subsystems, classes, and interfaces and their respective relationships within the framework of the system or software to be developed. At this stage, we used class diagrams. Class diagrams can be seen in Figure 4:

**Figure 4** Search optimization class diagram on Web-Based Student Data Application With Levenshtein Distance Algorithm

**Constructions**

Implementation stage to the program stage. At this stage implementing a component to perfect the system as a whole. The constructions stage produces a prototype program according to the previous design. This stage uses the main menu display of the resulting application. The main menu can be seen in Figure 5.
At this stage, black box testing is carried out, which is tested based on the activities in the application, this stage functions to test whether the application runs as it should without any errors. The following is testing with alpha testing:

<table>
<thead>
<tr>
<th>No</th>
<th>Test Case</th>
<th>Test Step</th>
<th>Expected Result</th>
<th>Result</th>
<th>Status</th>
</tr>
</thead>
</table>
| T001| login and go to Dashboard | 1. Login as school operator
2. Input username
3. Input password
4. Klik button login | 1. Showing Page Dashboards
2. Users can see information on the amount of student and teacher data | 1. Show dashboard page
2. Users can see information on the amount of student and teacher data | OK     |

Transitions
At this stage, it is more about the installation so that it is understood by the user and testing the application involving the user. The users who tested were two people, namely school operators and school principals. From the test results, it can be concluded that applications and menus can function properly. To clarify the results of testing, researchers use beta testing in order to get convincing results. In carrying out beta testing, the operator and principal open the Application and fill out the questionnaire that has been made by the author.

The following is the alpha testing carried out by the researcher. The calculation of the levenshtein distance continues from the sixth calculation until all matrix values are filled. The levenshtein distance is the value on the right-hand side of the matrix, and in the case of the input string RANDY and the database target string RANDI are in D(5,5). After all matrix calculations have been performed, it is known that the result of calculating the distance between the RANDY input string and the RANDI database target string is 1 (one) as shown in matrix Table 2.

*Yoga Handoko Agustin
Based on the sample above, the Levenshtein Distance Algorithm can be used to identify writing errors or student data discrepancies. As well as accuracy in determining database corrections from user input. From the 100 algorithms testing data in searching using the Levenstein distance, it can display the corrections, and from the tests carried out to get the level of accuracy in searching for student data, it is 90%. So this algorithm is suitable for use in student data applications.

1. Comparing Between Algorithm And SQL
a. Search with SQL

Based on figure 6, the school operator looks for names with the word wash while those in the grandchild database. By using SQL, the data cannot be found because the strings do not match those in the database.

*Yoga Handoko Agustin

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.
b. Search with algorithm

![Figure 7. using the Levenshtein Distance algorithm](image)

Based on Figure 7, using the Levenshtein Distance algorithm, when the school operator is wrong in inputting a search, the Levenshtein Distance algorithm will optimize the search by displaying the results of the correction.

With the results of the data from these trials, it can be concluded that the Levenshtein Distance algorithm can optimize the search for search input errors and can display the corrected results, compared to the SQL structure, which must be the same between the input string being searched for and the string in the database. However, from a comparison of time, SQL is faster than the Levenshtein distance algorithm. Because SQL only detects that the input string from the search must be the same as the string in the database, if it is different, then the data will not be displayed. Unlike the Levenshtein Distance, which is quite long compared to SQL because the Levenshtein Distance processes between the user search input string and the string in the database. Even if the input string is different from the database string, this algorithm can process it and display the result of the correction.

The following is testing the application with the beta testing stage by showing the function of the application to the Babakanloa 2 SDN school operator, and after that, an assessment of the results of the answers is carried out using a Likert scale.

### Table 4. testing and calculating using the Likert formula

<table>
<thead>
<tr>
<th>No</th>
<th>Statement</th>
<th>Response</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The appearance of this application is interesting</td>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Does the entire menu in the application work</td>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Is this application easy to use</td>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Does search in the student data menu minimize errors in inputting</td>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Is the menu in the application easy to understand</td>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>Does this application answer the problems that exist at SDN 2 Babakanloa</td>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>This application makes it easy</td>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>Overall this app is useful</td>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Amount**

| 11 |

By testing and calculating using the Likert formula, the final result is that the percentage of answers is 94% with the following formula calculation.

**Final result** = Total Score / Highest Score x 100% so that the following values are obtained:

- **Total value**
  = (Total Voters x Likert Score)
  = (11x5) + (5x4) + (0x3) + (0x2) + (0x1)
  = 55 + 20 + 0 + 0 + 0
  = 70

- **The highest score**
  =Highest Likert Score x Number of Questions x Number of Examiners
  = 5 x 8 x 2
  = 20

- **The final result**
  = (Total Score / Highest Score x 100%)
  = 75/80 x 100%
  = 94%

*Yoga Handoko Agustin
DISCUSSIONS

With the results of the data from the trials above, it can be interpreted that the Levenshtein Distance algorithm is more optimal for conducting searches because it can display the results of the corrections, whereas the sql structure between the input words being searched for and the words in the database must be 100% the same. Therefore, the use of the Levenshtein Distance algorithm can optimize word search compared to SQL. The following optimization is obtained using the Levenshtein Distance algorithm.

CONCLUSION

Based on the results of the discussion regarding Optimizing the Search for PIP Scholarship Recipients in Student Data Applications with the Levenshtein Distance Algorithm, the following conclusions are obtained:

1. The application of the Levenshtein Distance method can optimize student data searches and minimize the occurrence of search input errors by the school operator. The application of the Levenshtein distance method produces a very good accuracy rate of 94% from the results of correcting student data at SDN II Babakanloa. So this shows that the application of the Levenshtein Distance method is appropriate to use in optimizing search and minimizing the occurrence of input errors in SDN II Babakanloa student data.
2. This application can make it easier to update parents’ income, as well as optimize searches and minimize errors in searches.

REFERENCES


*Yoga Handoko Agustin