Implementation of the String Matching Method on Anggah-Ungguhing Balinese Language Dictionary

https://doi.org/10.3991/ijim.v14i01.11109

I Made Agus Wirawan (✉) Ida Bagus Made Ludy Paryatna
Universitas Pendidikan Ganesha, Bali, Indonesia
imade.aguswirawan@undiksha.ac.id

Abstract—Although the learning media of the Anggah Ungguhing Balinese language has been done anytime and anywhere through a mobile application, this application has several obstacles. In detail, the contributions of this study are: 1) to develop a dictionary application that can be used in offline mode, so that it can be used anytime and anywhere. 2) to apply the string matching method to word search correction in the translator process so that the user gets the correct word from the type of keyword. 3) to develop a new vocabulary sending feature used for users to take part in managing the Anggah Ungguhing Balinese language. 4) to describe students’ learning responses in Anggah Ungguhing Balinese language using mobile application. Positive findings/innovations targeted at this study that support the development of science, technology, and social culture are as follows.
1) the availability of database update features for Anggah Ungguhing Balinese language that can be used for offline mode. 2) the availability of correction features using the string matching method for word search correction in the translator process in the dictionary application. 3) the availability of a new vocabulary delivery feature is used for users to take part in managing the vocabulary of Anggah Ungguhing Balinese language. 4) the students’ responses to the Anggah Ungguhing Balinese language have been described based on mobile application in the learning process with an excellent range.

Keywords—Mobile dictionary, anggah-ungguhing balinese language, string matching method, user response, offline mode, firebase.

1 Introduction

Balinese is a language that has a multilevel language system. The level of the Balinese language is referred to as the Anggah-Ungguhing Balinese language [1, 2]. This term was inaugurated in the Pasamuhani Agung in 1974 at Singaraja during the program to discuss the standardization of the Balinese language. In association with the Balinese tribe, people recognize the term word term called 'masor singgih' or 'sor singgih basa' [3]. But the present condition of young people rarely uses the level of language when communicating appropriately. The importance of protecting and preserving mother-tongue through classroom learning can enhance students' confidence, achievement [4,
5], and character-building [6]. To provide an understanding of the Anggah-Ungguhing Balinese language, it is done through learning in school.

But based on observations made in several schools, the learning process carried out in the classroom has not been carried out optimally. For this reason, various efforts need to be made through language learning theory, learning methods, sustainable learning strategies. But, that students' ability to use Balinese basics in speech texts was classified as lacking [7, 8]. This is because students often experience errors in the use of the Balinese Anggah-Ungguhing language in speech texts both in diction and in context. This is due to the mastery of vocabulary that is not well understood. The more vocabulary mastered, the more skilled the students speak and the smoother the communication process is carried out [9, 10]. The mobile learning is a media can be used anywhere and anytime for vocabulary learning [11, 12, 13, 14]. For this reason, to understand Anggah-Ungguhing Balinese Language can be addressed by mobile learning. Based on research [15], researchers have developed mobile-based media for learning vocabulary from a Balinese. But, from this study there are several disadvantages:

- A previously developed dictionary application can only be used in online mode. Where users must be connected to the internet when using a dictionary. This will be constrained if there is no internet connection. We need to develop a database that can handle both online and offline.
- The keywords used in word search in the dictionary application must be correct and complete. If the keyword is incorrect, the translation of the Anggah-Ungguhing Balinese language is not found. Word checking problems can be handled using the String Matching method. String matching [16, 17, 18] is a method of character matching from the input process like a keyboard, and the matching process is done by checking every incoming character. The brute force algorithm is a simple algorithm that can be used for word matching [19, 20].
- In the previous research, there has been no user role for adding the Anggah-Ungguhing Balinese language vocabulary. So management of vocabulary is only done by the admin.

Based on this explanation, the author was motivated to implementation of the String Matching method on a mobile dictionary of Anggah-Ungguhing Balinese language. This a mobile dictionary is also handled online and offline mode and has a vocabulary addition facility.

2 Literature Review

2.1 String matching method

String Matching is the process of searching all query occurrences which are then called patterns into longer strings of text. String matching is divided into two, namely exact matching and heuristic or statistical matching. Exact matching is used to find patterns originating from one text [21]. Exact matching algorithms are classified into three parts according to their search direction, namely:
• The direction of reading from left to right. Algorithms belonging to this category are Brute Force, Morris and Pratt (which was later developed by Knuth, Morris, and Pratt).
• The direction of reading from right to left. The algorithms included in this category are Boyer and Moore, which were later developed into the Boyer-Moore, Tuned Boyer-Moore, and Zhu-Takaoka Turbo algorithms.
• Direction specified by the programmer. The algorithms included in this category are the Colussi algorithm, Crochemore-Perrin.

String Matching is formulated as follows:

\[ x = x[0...m-1] \]
\[ y = y[0...n-1] \]

Where:
- \( x \) is a pattern
- \( m \) is the pattern length
- \( y \) is text
- \( n \) is the length of the text

Both strings consist of a set of characters called alphabets which are denoted by \( \Sigma \) (Sigma) and have a size of \( \sigma \) (Tao). String matching is divided into two, namely exact matching and heuristic or statistical matching. Exact matching is used to find patterns originating from one text [22]. An example of searching exact matching is the search for the word "student" in the sentence "I am a student" or "I am a learner". The system will give results that the first sentence contains the word "student" while the second sentence does not, even though in reality students and learners are synonymous words. The Brute Force algorithm is the simplest and most widely used exact matching algorithm. The brute force algorithm is an algorithm used to match patterns with all text between 0 and \( n-m \) to find the existence of a text pattern. The brute force algorithm solves problems very simply, directly, and clearly. The brute-force algorithm is a technique commonly used when the compiler of the algorithm is more concerned with getting a solution to the problem directly as it is [23]. In detail the steps used by the brute force algorithm to match strings are as follows:

1. The brute force algorithm starts matching patterns from the beginning of the text.
2. From left to right, the brute force algorithm will match characters per character pattern with the characters in the corresponding text, until one of the following conditions is met:
   a) Characters in the pattern that are compared match, then the search is complete.
   b) If there is a discrepancy between the pattern and the text, then the search is not suitable and has not been successful.

• Then the brute force algorithm continues to shift the pattern one by one character to the right, and repeat step 2 until the pattern is at the end of the text. As with other algorithms, brute force algorithms also have advantages and disadvantages. Brute force advantages are:
   a) The brute force algorithm can be used to solve almost any problem.
   b) The brute force algorithm is simple and easy to understand.
While the weakness of the brute force algorithm is not as constructive as other problem-solving techniques. The pseudocode of the search for brute force algorithms, namely:

Procedure Algorithm
Input:
\( m, n \) as the length of the pattern and text
\( x \) as a pattern
\( y \) as text
process
looping \( i = 0 \) to \( m-n \) then
\( j = 0 \)
if \( j \) is smaller than \( n \) and \( x[i+1] = y[j] \) then
\( j = j + 1 \)
end if
if \( j \) is greater than \( n \) then
pattern = true (found)
end if
end of the loop
output:
pattern found.

The string matching method can be divided into two [23], namely:

- Exact string matching, is matching strings correctly with the arrangement of characters in matching strings having the number or sequence of characters in the same string. For example, the word drug will show a match only with the word drug
- Inexact string matching or fuzzy string matching is a cryptic match of strings that are matching strings where matching strings have similarities but both have different character arrangements (maybe the number or sequence) but the strings have similarities in textual/writing or similarity greeting.

Matching strings based on textual/writing similarities includes the number of characters, the arrangement of characters in a document is called an approximate string matching while matching strings based on speech similarity in terms of pronunciation is called phonetic string matching. The following is an example of an approximate string matching and phonetic string matching:

a) The word 'panitia' (committee) with the word 'panitea', has the same number of characters but there are different characters. If this character difference can be tolerated as a writing error, the two strings are said to be suitable
b) The word 'obat' with 'obad' of different writing but in similar pronunciation so that the two strings are considered suitable. Another example is the word 'obat' (drug) with 'obbat', 'obaat', 'obatt', 'obate'.
2.2 Firebase

Firebase is a Cloud Service Provider and Backend as a Service owned by Google. Firebase is a solution offered by Google to simplify the development of mobile and web applications. So, there is no need to build the features created on the backend and infrastructure from the beginning so that it can focus on developing high-quality applications without the need to spend a lot of effort [24]. Firebase has many SDK (Software Development Kits) that allow it to integrate this service with Android, iOS, JavaScript, C++ to Unity. One feature of Firebase is Firebase Storage. Firebase Storage allows mobile apps to store or download files on Google's cloud storage. The SDK provided even has been equipped with an auto-resume download/upload feature when the connection is lost. Firebase Storage also includes a rules engine to determine access rights of files stored on the server and can be modified according to the needs of each application [25].

Firebase Storage is used to store data such as images, audio, and video. Some applications must use image, audio or video files. Firebase is an online database that can be used as a data storage medium from applications. Data is stored in JSON form and can be synchronized in real-time to each connected client. Firebase has 3 core capabilities, namely [26]:

- **a) Real-time:** If there is a change in the database data, then all clients that are connected automatically will get the changes quickly.
- **b) Offline mode:** Applications that use this feature will remain responsive even when offline. This is because the Firebase SDK (Software Development Kit) can maintain data and changes to client storage media. When the client connects to the internet network, the Firebase SDK will make automatic adjustments to the records of changes in data stored on the client storage media with the current conditions of the Firebase Server.
- **c) Accessible from client devices:** This service offers the convenience to access the firebase real-time database directly from a mobile device or a web browser without the need for an application server.

The process of uploading to storage can be done through the user application using AngularJS. This is easier to implement because it does not need to use server-based programings such as PHP or ASP. When the file has been uploaded successfully to the storage, it will return a download link (download link) that can be saved to Firebase Database.

2.3 Anggah-ungguhing of Balinese language

Balinese has a multi-level language system. The first person to discuss the level of Balinese language in a written description was the Dutchman J. de Vroom in 1872 with the term ‘Masor Singgih’. B. Van Eck in 1874 with the terms Speech Levels. The level of the Balinese language is now referred to as ‘Anggah-Ungguhing Balinese’. The term was inaugurated at the 1974 Agung Basa Bali gathering at Singaraja in terms of dis-
cussing Balinese Language Standardization Problems. Previously in association in Balinese society, it was known as ‘alus kasar’, ‘masor singgih’, and ‘sor singgih basa’. [27, 28]

According to [29], the existence of the level of Balinese language in accordance with the times. The level of the Balinese language (especially the Balinese language) is the communication system of the Balinese community which is influenced by the stratification of the community towards mutual respect for its use in daily life in accordance with the speech situation that exists in the community. The division of the level of Balinese language as follows: [30, 31]

Table 1. Balinese language level

<table>
<thead>
<tr>
<th>No.</th>
<th>Word Levels (Rasa Basa Kruna Basa Bali)</th>
<th>Sentence Level (Rasa Basa Lengkara Basa Bali)</th>
<th>Paragraph levels (Rasa Basa, Basa Bali)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Courteous soft words (kruna alus singgih)</td>
<td>Courteous soft sentence (lengkara alus singgih)</td>
<td>Courteous soft paragraph (basa alus)</td>
</tr>
<tr>
<td>2</td>
<td>Moderate soft words (kruna alus madia)</td>
<td>Moderate soft sentence (lengkara alus madia)</td>
<td>Moderate soft paragraph (basa madia)</td>
</tr>
<tr>
<td>3</td>
<td>Low soft words (kruna alus sor)</td>
<td>Low soft words (lengkara alus sor)</td>
<td>Low soft words (basa andap)</td>
</tr>
<tr>
<td>4</td>
<td>Soft neutral word (kruna alus mider)</td>
<td>Soft sentences (lengkara andap)</td>
<td>Rough paragraph (basa kasar)</td>
</tr>
<tr>
<td>5</td>
<td>Neutral word (kruna mider)</td>
<td>Rough sentences (lengkara kasar)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Soft word (kruna andap)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Rough word (kruna kasar)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to [29] the ability to use the words of the Balinese language will be able to increase language skills in accordance with (graciously, Balinese language). The types of words used in (including the language of Bali) are as follows:
The Kruna Basa Bali flavor (based on the level of the Balinese word) is divided into 7 parts:

a) Kruna alus singgih (K. ASI): the word alus singgih is used to honor people who deserve respect. Example. ‘Seda’ (died), ‘mobot’ (pregnant), ‘kayun’ (want), ‘wik’ (smart), and etc. Kruna alus madia (K. AMA): the word alus madia is used between the words alus singgih and alus sor. Example. ‘Niki’ (this), ‘niki’ (that), ‘ampun’ (already), ‘ngajeng’ (eating), ‘sirep’ (sleeping), and etc.Kruna alus sor (K. ASO): the word alus sor is used to humble oneself and those who deserve to be humbled. Example. ‘Mawasta’ (named), ‘maurip’ (animate), ‘mapajar’ (said), ‘manah’ (mind), and etc.Kruna alus mider (K. AMI): the word alus mider is used for people who deserve respect and are not respected. Example: ‘rauh’ (coming), ‘lali’ (forgetting), ‘malih’ (again), ‘olih’ (by), ‘pacang’ (going), ‘sampun’ (already), and etc. Kruna Mider (A word that has a neutral value). Kruna mider is neutral word that does not have a rough and smooth variety. Example. ‘Tembok’ (wall), ‘kambing’ (goat), ‘lidi’ (sticks), and etc.
f) Kruna Andap (Lower than soft word). Krina andap is a word of respect, a word that has a lower language than subtle words. Example. ‘apa’ (what), ‘suba’ (already), ‘pipis’ (money), ‘nasi’ (rice), ‘panak’ (children), ‘gedeg’ (angry), and etc.

g) Kruna Kasar (Word which has a rough value). Is a word that has a rough language usually used when angry and quarreling. Example. ‘Bangka’ (dead), ‘ngamah’ (eating), and etc.

3 Results and Discussion

The development of this application uses the Android programming language, SQLite for data storage on smartphone devices and firebase databases to store data on the server. The application performs two-way translation namely from Indonesian to Balinese (which consists of 7 variants namely Kruna Alus Singgih, Kruna Alus Madia, Kruna Alus Sor, Kruna Alus Mider, Kruna Mider, Kruna Andap, and Kruna Rough) and vice versa. The list of words used in the application comes from a book written by N. Shadeg entitled A Balinese Basic Vocabulary [32] with a number of ± 3,000 lines of vocabulary. If the inputted word is not appropriate, then the String Matching method will correct according to words that have a greater closeness value. In this application, there is a feature to add vocabulary by the user, which is then verified by the admin.

3.1 Results

Application architecture: This dictionary application can be used offline mode (using SQLite), but for the process of updating vocabulary data to firebase and submitting new words, an internet connection is required. Overall, the system architecture of the dictionary application is already a Balinese base.

Fig. 1. Mobile dictionary architecture
**Application development:** There are several features developed in this mobile dictionary application:

**Translation menu feature:** The translation feature is used to process the translation of words from Indonesian to Balinese, and vice versa. For example, the process of translating the word 'eat' (from Indonesian to Balinese). Then the results of the translation will be translated into Balinese (7 levels), the translation results can be seen as follows.

![Translation process](http://www.i-jim.org)

The Translation Results dialog displays the results according to the translation mode that has been selected, translation from Indonesian to Balinese, or from Balinese to Indonesian. For example the translation process of the word 'makan' ('eat') will result in a translation of 7 variants of the Balinese language that are displayed in a dropdown, the word 'rayung' for the variant of Kruna Alus Singgih, the word 'rayung' for the variant of Kruna Alus Madia, the word 'ngajeng' for the variant of Kruna Alus Sor, the word 'daar' for the variant of Kruna Alus Mider, the word 'daar' for the variant of Kruna Andap, and the word 'amah' for the variant of Kruna Kasar. In the translation process, there is a word correction feature inputted by the user using the string matching method. Suppose the user wants to do the translation of the word 'makan' ('eat'), but the word typed is 'makn'. The string matching method will match the words that correspond to the word 'makn' that is in the database. The results obtained show the word 'meaning' corresponds to the word 'makan' ('eat').
In the translation process, there is a word correction feature inputted by the user using the string matching method. Suppose the user wants to do the translation of the word 'makan' (eat), but the word typed is 'makn'. The string matching method will correct by matching the words that correspond to the word 'makn' that is in the database. The results obtained show the word 'makn' corresponds to the word 'makan' (eat). Users will be given 2 options, first the user is given the correct word (according to matching string matching method) and the second is given a new vocabulary added option. If the first option is chosen by the user, the word 'makn' will be changed to the word 'makan' (eat) and then translated. Whereas if the second option is chosen, the word 'makn' is used as new vocabulary and verified by the admin.

New vocabulary submission feature: The new vocabulary submission feature is used to provide access to users to take part in managing the vocabulary of Anggah Ungguhing Balinese language. The process of submitting vocabulary can be done by all users but is verified by the admin.
Admin can do vocabulary searches and take actions whether the new vocabulary is accepted or ignored. If ignored, the new vocabulary is deleted from the list, whereas if it is accepted, the new vocabulary will be entered into the main vocabulary list and will be used by the user when the next database update. The process of submitting vocabulary can also be done by the admin. Admin can search vocabulary, add new vocabulary, change and delete vocabulary data.

Database update feature: The database update feature is used for updating data/vocabulary in the database. The update process only is done by the admin.
If the admin updates the database, a notification will appear on the user application (red mark) on the database download icon. Users can download the latest version of the database online. If successfully downloaded, then it can be used offline for the translation process.

**Application testing:** The testing process of the Balinese language application is done through several stages of testing, namely white box testing, black-box testing, and User acceptance testing.

*White-box testing:* White box testing aims to test software in terms of design and program code whether it is able to produce functions, inputs, and outputs that are in accordance with the specifications of the requirements. White box testing is done by checking the logic of the program code. Testing the device in a white box which includes software unit testing, integration testing, system testing. Structural testing software on the Balinese Dictionary Application using String Matching Method is carried
out to find out the work processes of the software internally and to guarantee internal operations in accordance with predetermined specifications using the control structure of the designed procedure. White box testing is carried out by the developer during the application development process. The white box test results conclude that the work process of the software internally works well in accordance with predetermined specifications. The source code of the program works well and correctly according to its function.

**Black-box testing:** Black box testing is a test to measure the success of the system in providing an output based on input given by the user. Testing on the Balinese dictionary using the String Matching method is done by developers and users. Tests carried out by the developer are carried out during the application development process. While testing carried out by users (5 users) is using different smartphones for each tester. After using the application, each tester is given a process conformity questionnaire. The results of this test will also be used as the results of the application usage test capable of running on various smartphones. Functional testing that has been carried out by developers and 5 testers shows that all processes are work well. These results also show that the application can run well on a variety of different smartphones. There is no difference in function and appearance for each smartphone of each tester.

**User acceptance testing:** The purpose of using the UEQ questionnaire is to determine the rating scale based on; Attractiveness, Clarity, Efficiency, Accuracy, Stimulation, and Renewal of a product. The test involved 15 users from the general public. Tests were carried out using the UEQ assessment questionnaire (User Experience Questionnaire). After testing, the following results are obtained. Grading points use a scale from -3 to +3, where -3 is the most negative answer, 0 is a neutral answer, and +3 is the most positive answer. Based on the UEQ questionnaire calculation, the results were obtained, among others: in terms of appealing appeal is very good, in terms of excellent application clarity, in terms of application efficiency is very good, application accuracy is very good, application stimulation is very good, and in terms of novelty also very well. The average achievement of User Experience is at a very satisfying level as shown in the following figure 8.

![Fig. 8. Result of user experience](http://www.i-jim.org)
3.2 Discussion and future research

This research has produced three features in the dictionary application Anggah Ungguhing Bali language, namely:

- Word correction features
- Vocabulary management features
- Translate features with offline mode

Translation menu feature: In accordance with the background of the problem being studied, the translation feature is used to correct word typing errors in the translation process. The string matching method is successfully applied to the dictionary application Anggah Ungguhing Balinese language to correct the word translation. But the string matching method with the brute force algorithm requires a lot of time in the process of word matching because the matching process is done thoroughly per character [33]. To improve the matching performance of the Jaro Winkler Distance algorithm [34]. Besides that, further research can apply the Bastal algorithm to improve the translation sentences of the Anggah Ungguhing Balinese language [35].

New vocabulary submission feature: The new vocabulary submission feature is used for users to take part in managing the vocabulary of Anggah Ungguhing Balinese language. The process of submitting vocabulary can be done by all users but is verified by the admin. This feature can help admin manage Anggah Ungguhing Balinese language vocabulary. This feature is developed in accordance with the background of the problem, where there needs to be a user role in adding new vocabulary and verified by the admin.

Database update feature: This feature is developed in accordance the database update feature is used for updating data/vocabulary in the database. The update process only is done by the admin. This feature is developed according to the background of the problem, where the dictionary application can be run offline mode.

The three features of the Anggah Ungguhing Balinese language dictionary application: Have been successfully developed and several stages of testing have been carried out, including the white box, black box, and user acceptance testing. White box testing and black box testing are technical tests to test the success of the application being developed. While user acceptance testing is used to measure the level of user acceptance in using the Anggah Ungguhing Balinese language dictionary application. Based on the results of the user acceptance testing test, user acceptance is obtained using the dictionary application, which is considered important. Balinese is in an excellent range.

4 Conclusion

Based on the background of the problem examined in this study, it can be concluded:

- The availability of database update features for Anggah Ungguhing Balinese language that can be used for offline mode
• The availability of correction features using the string matching method for word search correction in the translator process in the dictionary application
• The availability of a new vocabulary delivery feature is used for users to take part in managing the vocabulary of Anggah Ungguhing Balinese language
• The students’ responses to the Anggah Ungguhing Balinese language have been described based on mobile applications in the learning process with an excellent range. For future research, to improve the matching performance of the Jaro Winkler Distance algorithm. Besides that, further research can apply the Bastal algorithm to improve the translation sentences of the Anggah Ungguhing Balinese language.

5 References


6 Authors

**I Made Agus Wirawan** is a lecturer at the Education of Informatics Engineering, Universitas Pendidikan Ganesha, Indonesia. His research focus is in the field of artificial intelligence, learning technology, and mobile applications.

**Ida Bagus Made Ludy Paryatna** is a lecturer in the Balinese Language Education Departement, Universitas Pendidikan Ganesha, Indonesia. His research focus is in the field of Balinese language, and anggah unguhing language.