The Development and Assessment of Pattern Matching Algorithms Used by ZEE: A Filipino Sign Language (FSL) Dictionary and English-Learning App

Maria Dominique Daculan
University of San Carlos
16100499@usc.edu.ph

Asherrie Jaye O. Tan
University of San Carlos
18102538@usc.edu.ph

Angie M. Ceniza-Canillo
University of San Carlos
amceniza@usc.edu.ph

Abstract

"The Filipino Sign Language Act" of RA 11106 states the declaration of Filipino Sign Language (FSL) as the National Sign Language of Filipino deaf citizens. However, the lack of awareness prevails as with the cyberbullying case of Mininio Buhat. This paper takes an initial step into raising awareness with Zee, with the analysis of three different pattern-searching algorithms - Knuth-Morris Pratt, Boyer-Moore, and Robin Karp to determine which suits the application best.By implementing three different algorithms, an efficient algorithm that is consistent in all test cases was determined to be used in pattern matching. The search functionality of the application was tested with one algorithm at a time where users input an English word and the application searches the database returning the corresponding FSL gesture of that particular English word in gif format. From there, the execution time of each algorithm was recorded and the overall criteria for the analysis of the algorithms: time complexity, space complexity and overall performance was compared among the three algorithms. It was determined that the Boyer Moore algorithm gave consistent results and was generally fast and efficient in any test case, be it patterns of shorter length and patterns of longer length.

1 Introduction

According to a survey, the term disability refers to "any restriction or lack of ability (resulting from impairment) to perform an activity in the manner or within the range considered normal for a human being" (NSO, 2014). Focusing on those belonging to the deaf community, Center for Hearing and Communication statistics show that around 90% of deaf children are born to hearing parents.

In the Republic Act 7277 [1], otherwise known as Magna Carta for Disabled Persons, it is stated that people who are deaf, mute, or hearing-imapired belong to 1.23% of the entire Philippine population. Furthermore, the 2009 projected population of the deaf consists of 241,624 totally deaf citizens as well as 275, 912 partially deaf. In the aforementioned Act, Persons with Disabilities (PWD's) are given special privileges and self-development programs for them to be a part of the mainstream of society. Despite having these laws made to alleviate the burden PWD's carry, it is eminent in society that people lack awareness towards PWDs and would often regard them as "weird", "abnormal" and in most cases "special" in a negative nuance. One of the major factors contributing to these preconceived ideas is the lack of communication and understanding between the hearing and the non-hearing community.

In addition, a 2014 Philippine News article presented Mininio Buhat - a Filipina born profoundly deaf - who became a victim of cyberbullying shortly after posting on the social platform Facebook with faulty written English. It was later revealed by experts that deaf people are taught proper English grammar in the latter part of their college education, as well as the fact that her sentence construction was based upon their sign language. This further proves the lack of awareness and knowledge the hearing majority has regarding the deaf and how their language works for them to communicate. According to the list by the National Council on Disability Affairs (NCDA), there exists a number of organizations and schools offering Special Education (SpEd) that share the advocacy of inclusivity - giving education to the SpEd individuals as well as providing FSL training to both the deaf and the hearing. Mandaue City Central Special Education School (MCCSpEd), FSL National Network, Philippine Accessible Disability Services Inc.(PADS) are just a few of those that are actively expanding their advocacies to society. However, people don't seem to be

aware of such efforts and still disregard the importance of their Advocacy.

With that in mind, this study proposes Zee. It is an application that promotes interactive learning of FSL for the hearing, as well as English grammar to the deaf. Through Zee, the advocacy of inclusivity is shared between the application and its users, in hopes of closing the communication gap between the deaf and the hearing community as well as raising awareness regarding the importance of FSL in our society. In lieu of this, an assessment of the aforementioned pattern-searching algorithms is proposed to determine which suits best for the type of dataset Zee has.

This research aims to assess the efficiency of the aforementioned pattern-searching algorithms to be used in the development of Zee: A Filipino Sign Language (FSL) Dictionary and English Learning App.

The objectives of the study are as follows:
1. Gather information on the current situation of the deaf community, their education, and their relationship with the mainstream society
as well as evaluating the current applications available for learning FSL.
2. Design and develop an FSL and english learning application using pattern matching algorithms.
3. Assess the pattern matching algorithms used in developing the application.
4. Test and evaluate the performance of the application.

The following will benefit from the results of this study:
Deaf Community. This study would be of help to the said community as it will teach them to construct basic English sentences in the proper grammar as early as their high school years, as proper grammar is said to be taught only in the later part of their college education. This study would also raise awareness of the community's advocacy for inclusivity.

Families with deaf-mute relatives and friends.Similarly, families will benefit from this study as they will now be able to learn the language used by their deaf-mute relatives and acquaintances and would be able to start communicating with them and gain a bit more knowledge and understanding towards the deaf-mute community.

Teachers of SpEd Schools. These teachers have been the pioneers of the advocacy of inclusivity by giving them the education they need and deserve, as they prepare the students for living in society. Through this study, it would support and act as gratitude for the efforts of these teachers by providing the students a platform to learn, especially in this time of pandemic and online learning.

Schools offering SpEd.Similar to the teachers, it is through the schools that deaf students are given a comfortable learning environment wherein SpEd and non-SpEd students can interact with each other and can lessen the 3 language barrier that exists between them. This study would benefit these institutions as it reflects their advocacy as it allows the deaf and the hearing community an opportunity to minimize the language barrier that exists.

Hearing Community. As sign language resources are a challenge to find, this study would provide the hearing community an opportunity to learn sign language - the official language of the deaf-mute community. Through this, they would be able to understand those that are deaf-mute and would be able to communicate with them even through basic conversations.

Researchers. The researchers will benefit from this study because this allows them to gain more knowledge and experience in creating an application, as well as knowing the best suited algorithm for this type of application.

Future Researchers. The study will benefit future researchers on their research on similar topics.

This study is for the assessment of different pattern-matching algorithms (Boyer-Moore, Rabin-Karp, and Knuth-Morris Pratt algorithm) for the development of a web, mobile-responsive application for learning FSL and English. This teaches the deaf community in their high school years to construct basic English sentences according to the correct grammar rules. In addition, this application brings forth awareness of the deaf community and serves as an opportunity for us - the hearing community - to learn the different signs and to be able to conjure a conversation with a deaf or hard of hearing individual. Sign Languages consist of manual (hand movements) and non manual signals (head nod, eye blink, etc.). In this study, only the manual signals are considered, with

slight inclusion of non manual signals concerning signs about questions.

The different signs will be in a form of Graphics Interchange Format (GIF). Each sign will be demonstrated by Elizabeth Ann Daculan, a high school deaf student. The demonstration clips will then be sent to the contacted advisers of Mandaue City Central Special Education School to validate the GIF clips. The 4 proposed study is primarily a web application, accessible through either Google Chrome, Microsoft Edge, and Mozilla Firefox. Furthermore, Zee will also be mobile responsive; hence, can be accessed through a mobile browser.

The Dictionary module consists of the different FSL signs in different categories, along with their corresponding meaning and guidelines on how to do the sign. In addition, users would also be able to search certain words (if available). The content is taken from a scanned reference book owned by the researchers in 2019; furthermore, the ones included in this study will be limited to the Preparatory level English signs which consists of the basic signs (ordinal and cardinal numbers, color, family, common phrases and vocabulary), the Filipino alphabet.

The application tracks the progress of its users - both the deaf and hearing - through different categories wherein each category has several levels of difficulty and with assessment tests after every level in the grammar module. On one hand, the categories for the hearing users run from the Filipino alphabet to basic vocabulary pertaining to color, numbers, family and relatives, days of the week and months, common phrases, and ultimately the Philippine National Anthem. Assessments are done through matching-type and/or fill in the blank type of questions.

On the other hand, categories for the deaf users cover the four basic sentences - Declarative, Interrogative, Imperative, and Exclamatory sentences. All sentences will be in the active voice for simplicity and ease of understanding. When relating the signs to the sentence, each pair of signs and its corresponding word in the sentence will be color-coded for identification. Similar to how the hearing users are assessed, matching type and fill in the blanks are the questions that will be used for assessment.

A review quiz will be given to the users after every level. Should the user not be able to get everything correct, he/she is not allowed to move to the next level and must restart the level. A successful attempt on the review quiz rewards the user with a star, and the total number of stars to collect per category is equal to the number of levels it has. The assessment tests will be given after every category - containing jumbled questions from the different levels the category contains. Similar to the review quiz, a user is not able to continue to the next category until he/she has passed the assessment test.

2 Related Works

2.1 The Deaf Community

In a research by Calimpusan and Silva-dela Cruz (2018), those belonging to the Deaf community are defined as people with a “hidden disability”, as they lack physical marks and attendants (wheelchair, walking cane, black eyeglasses, service dogs, etc.) as opposed to other disabilities. The only recognition they get is when they begin communicating using sign language.

Cabreros (2020) studied on the English proficiency level of 23 college Special Education (SPED) students taking the Associate in Computer Technology program under the Bachelor of Science (BS) in the Information Technology department. Results showed that a large percentage of the respondents scored below the mean score, given an interpretation of Needs Improvement. The average result determined that most of the respondents portrayed below average skills on their English proficiency in terms of their vocabulary. Respondents gained a significant improvement on their English grammar, but portrayed low comprehension skills.

2.2 Filipino Sign Language (FSL)

Several studies, including one by Rinaldi, Caselli, Lucioli, Lamano, and Volterra (2018), have emphasized the difference between Signed Languages (SLs) and spoken languages. One of which being the difference between their basic linguistic units. Spoken languages sequentially produce phonemes - the smallest unit of sound in a word that makes a difference in its pronunciation, as well as its meaning (LiteraryDevices) - and morphemes - the smallest syntactical 7 and meaningful linguistic unit that contains a word (LiteraryDevices) - whereas signed languages produce them simultaneously. Signed languages are also divided into two parameters, the manual parameters (movement, hand location, orientation) as well as the non-manual parameters (facial expression, mouthing, body movements) which gives a

different meaning to every sign used when communicating.

The aforementioned study also did an experiment with three different deaf age groups (younger children, older children, adults) and tasked them to reproduce the sign language shown on the computer. Results showed that younger children tend to omit a larger number of signs then the other groups, while the remaining groups did not show any significant differences. Younger children presented a pattern of reproducing the signs different from the older children and adult groups.

Signed by President Duterte on the 30th of October 2020, the Republic Act 11106 - otherwise known as The Filipino Sign Language Act - officially recognizes FSL as the National Sign Language and be officially used in all types of transactions with the Deaf community. Section 11 of said Act states the promotion of FSL by the appropriate agencies towards propagating the hearing people's competency in learning FSL by offering it as an elective in the mainstream curriculum, particularly of the State Universities and Colleges (SUCs) or schools that are government-funded. However, a section that talks about family members learning FSL in conjunction with their deaf child/children's learning is not present.

Uploaded on Google Playstore in February of 2018, Filipino Sign Language is a simple informative program that teaches sign-language gestures or "fingerspell" of the basics (letters, numbers, special characters) and offers an FSL translation of the Philippine National Anthem. Considered as the first of its kind due to similar apps only up to its first release and has never been updated.

Another application named FSL Buddy was uploaded last July 2018. The goal of the app - as well as other apps similar to them - have a goal of teaching people the national sign language. The FSL vocabulary featured in this app is 8 part of a course taught at De La Salle – College of Saint Benilde. This app was only up to its first release. Recent applications lack the feature of teaching users how to use the signs in a sentence.

2.3 Sign Language Translation

In this study by Orbay and Akarun (year), a pre-processing method called tokenization was used as well as utilizing token learning from sign language videos wherein supervised data is at hand. As annotated data is costly and scarce altogether, Transfer-Learning from semi-supervised tokenization approach is utilized. Three experimental setups were determined; right hand keypoints, both hand keypoints, as well as full body pose while ignoring the non-manual signs (face keypoints). The said study utilizes RWTHPHOENIX-Weather 2014T which is a Continuous Sign Language Benchmark dataset. This study only focuses on the manual portion of the signs and did not put into consideration the non-manual portion as well as the context of the sign that would help produce better results. This study does not in any way discuss the linguistic properties of sign language. Another study by Camgoz, Hadfield, et. al., utilizes the Neural Machine Translation (NMT) framework inorder to determine the spatial representations, its language model, as well as the mapping in between the sign and spoken language.

In this study by Orbay and Akarun (year), a pre-processing method called tokenization was used as well as utilizing token learning from sign language videos wherein supervised data is at hand. As annotated data is costly and scarce altogether, Transfer-Learning from semi-supervised tokenization approach is utilized. Three experimental setups were determined; right hand keypoints, both hand keypoints, as well as full body pose while ignoring the non-manual signs (face keypoints). The said study utilizes RWTHPHOENIX-Weather 2014T which is a Continuous Sign Language Benchmark dataset. This study only focuses on the manual portion of the signs and did not put into consideration the non-manual portion as well as the context of the sign that would help produce better results. This study does not in any way discuss the linguistic properties of sign language. Another study by Camgoz, Hadfield, et. al., utilizes the Neural Machine Translation (NMT) framework inorder to determine the spatial representations, its language model, as well as the mapping in between the sign and spoken language.

Lugman and Mahmoud (2018), developed a semantic rule-based machine translation system from Arabic into Arabic Sign Language (ArSL). This system involves three main translation stages (morphological analysis, syntactic analysis, and ArSL generation. The user inputs an Arabic sentence. The system then outputs an ArSL sentence represented by the gloss notation which is displayed as a sign sequence of GIF images. The sentence is first

morphologically analyzed using the MADAMIRA toolkit to extract the structure of each of its words. This output is then syntactically analyzed using a dependency parser to obtain the relation between the words in the sentence. An Arabic parse tree is generated and then transformed into its equivalent tree in ArSL by applying some transformation rules. This phase also involves lexically translating the Arabic phrases and words into their equivalent signs in ArSL. In the event of an out-of vocabulary (OOV) problem, the synonym of the word is used. A statistical language model in the synonym selection is used to ensure that the synonym preserves the meaning of the sentence.

3 Methodology

This study consists of two user categories - deaf and hearing. In terms of the deaf respondents, the researchers will be collaborating with Ms. Dioscora Sollano, an adviser of the Hearing Impaired (HI) department of Mandaue City Central Special Education School. Considering the current online setting, the exact number of respondents for the deaf category cannot be determined, as there exists the challenge of contacting the parents of the students as well as the availability of internet connection in each household.

For the hearing respondents, the 155 BS Computer Science (BSCS) students from the Department of Computer, Information Sciences and Mathematics of the University of San Carlos - Talamban Campus is our chosen population.

$$n = \frac{N}{1 + N(e)^2}$$

Figure 1. Yamane Formula

Figure 1 shows how the sample size from the 155 BSCS students are calculated. n represents the sample size, N represents the total population - in this case, 155 -, while e is the margin of error which is set to 0.05 based on the research conditions. These values would then generate a sample size of 111 BSCS students.

For the initial information gathering, two sets of researcher-made questionnaires; one for the deaf respondents and the other for the hearing respondents. The questions are structured in a way to acquire qualitative data on their opinion on the awareness of the deaf community in society. Aforementioned questions are inspired from the personal experiences of one of the researchers, news articles, past interviews, and conversations with people directly related to the deaf community.

In the Knuth-Morris Pratt algorithm, let w be the substring and s be the given string and is the basis of comparison. A one-dimensional array (variable arr) is utilized and initialized according to the number of characters that can be skipped after a mismatch is met in every iteration. For the initialization of arr, a failure function f(i) is used. Such function is basing its procedure on the fact that all previous characters are correctly matched once a mismatch occurs. Hence, when a w prefix occurs in the set of matched characters, it is also a w suffix. In the function proper, a string with a length of 1 is given the value of 0 to arr.

For the Robin-Karp algorithm; let pattern be the string to be searched, m be the length of the pattern, text as which the pattern will be searched from, as well as N to be the number of characters. To start, the hash value of pattern is calculated and utilized later for comparison in pattern_hash. M characters from text are then calculated of their own hash value in text_hash. Both values of pattern_hash as well as text_hash are then compared; if they are equal, a brute force comparison is then executed wherein individual characters are compared from pattern and text, otherwise the next M characters of text are then calculated for their hash value and comparison begins. The process ends once the end of text is reached or the brute force comparison results in a TRUE.

In the case of the Boyer-Moore algorithm; let text be the origin string, pattern be the string to be searched, and N being the length of the pattern . Beginning 16 from the rightmost character of pattern (place variable i), compare that character with its equivalent placement in text (place variable j). If those initial characters mismatch, and if the character from text is not detected anywhere in pattern, pattern is then shifted N characters to the right. But if that character is detected, the pattern is shifted until the occurrence of that character in the pattern is aligned with the character in text. This process continues until all characters have a match or if it reaches the end of text.

Other sources such as websites, videos are used for data gathering. For the in-application assessment, a pre-test and post-test is done after each category in the application. The same questions are used in both tests to properly compare the change of scores. Evaluation is then done through the use of a T-Test.

By having respondents in their high school years involved, the researchers will apply for a research ethics review. Any information gathered from the respondents will be treated confidential between the researchers and the respondents themselves.

Both qualitative and quantitative approaches are used in this study. A qualitative approach is used in the initial data collection of the study to demonstrate the current awareness of the respondents regarding the deaf community in society. Letters were sent to the respective school heads to obtain permission on having a few of their students as respondents to the study. Once approval is received, contact information of the respondents are then acquired, and questionnaires will be transformed into Google Forms wherein the links will be sent to each of the respondents.

For the quantitative approach, a pre-test is administered before every category of the application. Individual scores are then recorded to be used in analysis. After which, respondents will then undergo a series of lessons in the form of levels and taking quizzes in between. Their progress per test is also recorded. After each category, a post-test is administered having the same questions as the pre-test. Results of the post-test are also recorded and will be analyzed together with the pre-test results.

A t-test is a statistical test used to compare the means of two groups, in this case, between the deaf and hearing respondents. It is often used to determine whether or not a process or treatment is effective on the population of interest. (Bevans, 2021)

Paired t-test is specifically used in this study to determine the mean difference of the English proficiency test scores taken by the respondents before and after learning each lesson category in the application. The results will then be analyzed and presented through graphs. SPSS software will be used for this purpose. ("LibGuides: SPSS Tutorials: Paired Samples t Test", 2021)

$$s_{\bar{x}} = \frac{s_{\text{diff}}}{\sqrt{n}}$$

Figure 2. Paired t-test Formula

Figure 2 shows the formula for the Paired t-test. Sdiff represents the sample standard deviation of the differences, $s_{\bar{x}}$ represents the estimated standard error of the mean (s/sqrt(n)) and n represents the sample size (i.e., number of observations).

Three different pattern-searching algorithms are utilized in this study, namely the Knuth-Morris Pratt, Boyer-Moore, as well as the Robin-Karp algorithm. Each of these algorithms will be implemented in the Dictionary module and analyzed based on a number of factors. To analyze each of the three algorithms, a variable O is defined to represent the running time and another variable n is defined to represent the space complexity. Two other variables Sn, representing the number of characters skipped from the input string in the event of a mismatch and Cn, are defined to represent the number of comparisons done overall.

In evaluating algorithm performance, the following factors need to be considered: correctness, finiteness, and efficiency. An algorithm is said to be correct if it produces the desired output for any particular set of inputs. It is vital that an algorithm should terminate after a finite number of steps. Otherwise, should loops be used, an infinite loop will occur and the program cannot progress until the error is resolved by the programmer. Lastly, an algorithm should be efficient meaning it should take up as less time and use less space, memory and resources as possible. ("Time and Space Complexity Analysis of Algorithm", 2019)

To further compare the three algorithms, certain formulae will be used to determine their respective efficiency and correctness. In measuring algorithm efficiency, the Big-O notation is used to determine its running time. The general formula for getting the running time is O(n)=cn, where c is some variable that represents the number of statements executed by the program. Using this formula, time complexity is calculated depending on the number of operations and the size of the input. Below shows the running time for different types of operations performed:

O(1) - when the algorithm performs a constant number of operations regardless of the size of the input.

O(log n) - when the algorithm takes as many steps as it can by performing repeated operations, for example division.

O(n) - for every loop executed
("Running Time and Big-O - Learneroo", n.d.)

An algorithm is considered to be correct whenever the initial state fulfills the precondition and the program terminates. Thus the formula for total correctness is:

Total Correctness = Partial correctness + Termination (Tran, 2017)

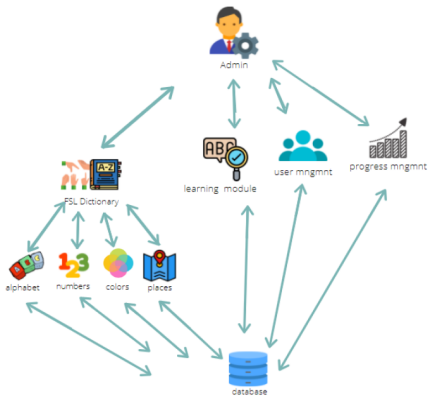

Figure 3 Administrator Conceptual Framework

Zee is an English language learning web and mobile-responsive application that aims to improve the communication between the deaf community and the hearing community. The learning application consists of several modules: an FSL Dictionary and a grammar module. The grammar module is subdivided into two - an FSL to sentence and a sentence to FSL submodule. The users of this application are the administrators, the deaf-mute community, and the hearing community. The administrators handle the database and updates whenever needed as well as accessing other modules. Both communities can access all modules but are unable to access and update the database itself.

As shown in Figure 3, the administrator can access all the modules, input and edit data into the database. The dictionary module is used in the learning module as a resource for lectures and tests, as well as allowing users to view the different signs. A Graphic Interchange Format (GIF) is used to creatively display how each sign is done as well as a textual description of how to act the sign. Each category of signs is assigned a color for ease of identification, which will be useful on the second module.

On the other hand, the learning module - separated into two submodules - aids deaf-mute users in constructing sentences in the basic types (declarative, interrogative, imperative, exclamatory) in the English grammar system through means of a game that assesses their knowledge in between. The FSL equivalent word in the sentence will be tagged according to the sign's category in the dictionary. This same means is applied to the hearing users to understand how the different signs are done and used in a sentence.

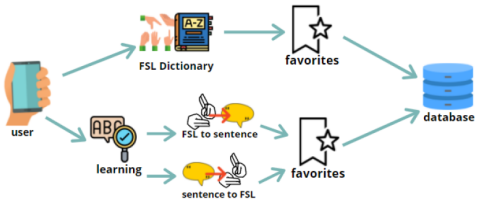

Figure 4 User Conceptual Framework

Users can access all modules in the application as shown above in Figure 4. When accessing each module and submodule, users can opt to add certain items - in the dictionary and learning modules - into Favorites to save for future access.

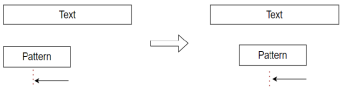

Figure 5 Boyer Moore algorithm

The Boyer-Moore Algorithm, specifically the Bad-Character Rule shown above, works this way. A given pattern is checked with the beginning of a certain string/text. Checking is done from right to left to see if any mismatched 22 characters exist between the two strings. If a mismatched character is found, the pattern moves one character past an occurrence of the mismatched character to the left. If no occurrence of the mismatched character is found to the left, the pattern moves one character to the right of the mismatched character. From there, checking continues until the end of the string is reached. (Langmead, 2015)

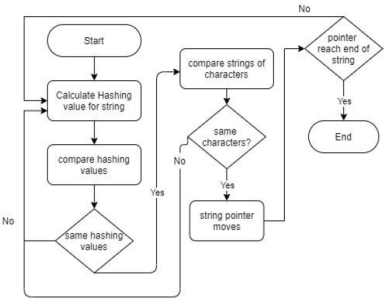

Figure 6 Rabin Karp Algorithm

The Rabin-Karp algorithm in the figure above makes use of hash functions and associates each character with a value from these functions. Pattern matching is done by adding the hash values of each character from a given pattern and

doing the same for the entire string to be checked. If the sum of the hash values of both patterns are not equal, a mismatch occurs and the pattern moves one space to the right. The process repeats until the end of the given string is reached. (Bari, 2018)

The figure below shows how the Knuth-Morris Pratt algorithm works. It starts by initially setting traversal variables to the beginning of both the pattern and given string. When a mismatch is met, the pattern traversal variable goes back to the beginning of the pattern to restart comparison. The algorithm 23 continues until the end of the given string is reached, whether or not a match is found.

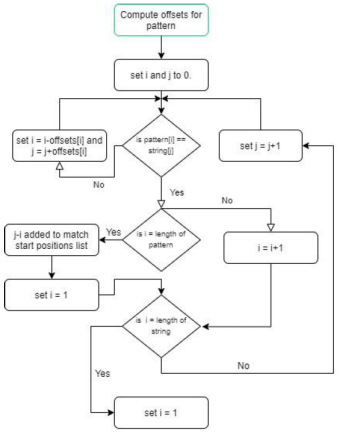

Figure 7 Knuth Morris Pratt Algorithm

According to Figure 8, the lexical table entitled ‘entries’ is taken from the “MySQL English Dictionary” which was parsed from the OPTED database and does not have any relation with the other tables. Every FSL sign corresponds to only one word in the entries table. In every favorite, there could be one or more FSL signs, and can be marked by one or more users. In terms of the progress, each entry of progress is to only one user, while every user can have one or more progress entries.

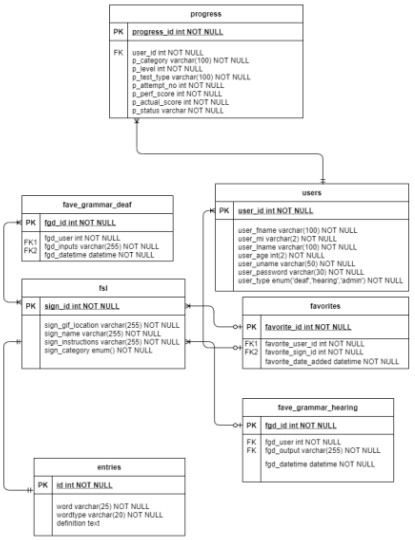

Figure 8. Entity Relationship Diagram

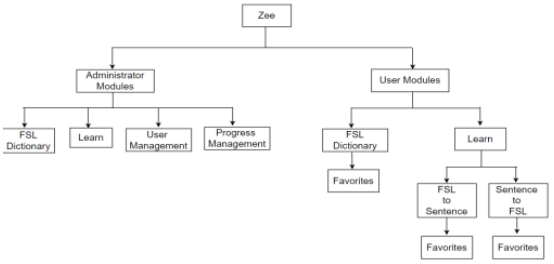

Figure 10. Top-Down Approach

This project utilizes the top-down approach for the development of the application. The approach allows a more defined state of development in every step as undefined submodules are defined after each step.

4 Results and Findings

The three pattern matching algorithms were implemented in the Zee FSL to English dictionary. Each algorithm was implemented one at a time as the search feature was tested. Users are to input a certain English word or pattern where the algorithm searches for a match in the database consisting of 201 entries and returns the corresponding Filipino sign language gesture with the corresponding gif format.

Each algorithm was analyzed based on the following metrics: Time Complexity, Space Complexity and Performance (Execution time).

pattern	Knuth Morris	Rabinn Karp	Boyer Moore
teen	6527	5017	5008
ther	6601	5009	5012
ch	6584	5020	5013
o	6594	5076	5013
day	6518	5017	5010

Table 1 Execution of all three Algorithms in searching small partial patterns

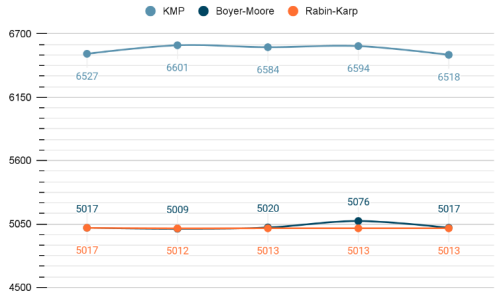

Figure 7 Screenshot to compare the execution time of the three Algorithms with searching small partial patterns

pattern	Knuth Morris	Rabinn Karp	Boyer Moore
small	6655	5006	5012
classmate	6778	5012	5005
read	6570	5007	5005
different	6717	5011	5015
why	6752	5014	5007

Table 2 Execution of all three Algorithms in searching full patterns

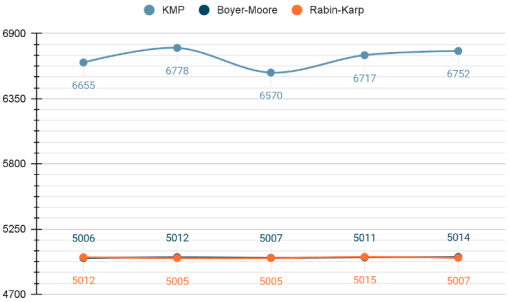

Figure 7 Screenshot to compare the execution time of the three Algorithms with searching full patterns

pattern	Knuth Morris	Rabinn Karp	Boyer Moore
invalid	6734	5005	5013
mountain	6654	5009	5009
xxx	6654	5005	5014
presents	6614	5006	5008
up	6619	5012	5016

Table 3 Execution of all three Algorithms in searching invalid patterns

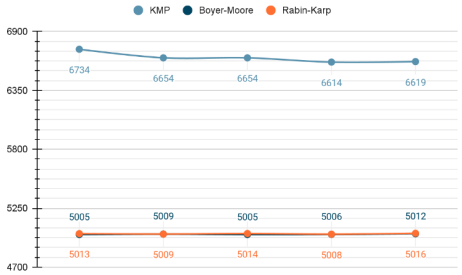

Figure 8 Screenshot to compare the execution time of the three Algorithms with searching invalid patterns

5 Conclusion and Recommendations

After running some tests, it was determined that the Knuth Morris Pratt Algorithm executes the slowest. This is considering that the words in the database are mostly of shorter length, with an average of 5 characters. Since the Knuth Morris Pratt Algorithm is designed to check for repeating characters known as prefixes and suffixes, then it behaves differently when all the characters in the string are unique..

Overall, the Boyer Moore Algorithm seems to be the most efficient. This is because the Boyer Moore algorithm is consistent in skipping a couple of characters when a mismatch is encountered.

However, when both patterns have similar subpatterns known as prefixes and suffixes, the Knuth Morris Pratt Algorithm works faster as it immediately goes to a common subpattern and begins checking from there until the last index

In terms of space complexity, the Rabin Karp Algorithm utilizes the least amount of space compared to the two algorithms. However, this low cost space is at the expense of a high time complexity as it only shifts one character to the right at a time.

While the algorithms analyzed in the paper may have an old approach, it is encouraged that future researchers who wish to study pattern matching make their own algorithms that are hopefully faster and more efficient in checking and comparing the existence of patterns than these existing algorithms. It will be even better if the new future algorithm proposed by other researchers will work smoothly as intended and hopefully easy to comprehend. New pattern matching algorithms will pave the way for faster pattern comparisons to be used by professionals not only in the field of Computer Science but also other fields that utilize pattern matching algorithms like Bioinformatics and DNA sequencing.(Fainstein, 2005)

References

Abdallah, E., & Fayyoumi, E. (2016). Assistive Technology for Deaf People Based on Android Platform. Procedia Computer Science, 94, 295-301.doi: 10.1016/j.procs.2016.08.044.

Cabreros, M. R. Q.. English Language Proficiency Profile: A Case Study of the Communication Skills of Deaf Students in the Undergraduate Program in Quezon, Philippines. International Journal of Educational Management and Development Studies, 1(1), 20-42.

Fainstein, J., 2005. The application of pattern matching algorithms in bioinformatics. [online] (UMI 1430487). Available at: <https://www.proquest.com/openview/d206c1963532e01b391e32cdf1c8b516/1?pq-origsite=gscholar&cbl=18750&diss=y> [Accessed 20 May 2022].

Luqman, H., & Mahmoud, S. (2018). Automatic translation of Arabic text-to-Arabic sign language. Universal Access In The Information Society, 18(4), 939-951. doi: 10.1007/s10209-018-0622-8.

Rinaldi, P., Caselli, M., Lucioli, T., Lamano, L., & Volterra, V. (2019). Sign Language Skills Assessed Through a Sentence Reproduction Task. Journal of Deaf Studies and Deaf Education, 408–421. doi:10.1093/deafed/eny021.

Silva-Dela Cruz, F. W. & Calimpusan, E. C.. Status and Challenges of the Deaf in One City in the Philippines: Towards the Development of Support Systems and Socio-Economic Opportunities. Asia Pacific Journal of Multidisciplinary Research, 6(4), 61-74.

Sleit, A., AlMobaideen, W., et. al., 2007. An Efficient Pattern Matching Algorithm. Journal of Applied Sciences, 7, 2691-2695.

Suszczanska, N., Szmal, P., & Kulikow, S. (2007). Continuous Text Translation Using Text Modeling in the Thetos System, 1(8). doi: doi.org/10.5281/zenodo.1331657.

Access Innovation Media. 4 Reasons People Use Sign Language Other Than Hearing Loss. Retrieved 7 March from https://blog.ai-media.tv/blog/4-reasons-people-use-sign-language-other-th an-hearing-loss.

All About Cookies. About your web browser. Retrieved 5 May from https://www.allaboutcookies.org/browsers/.

Bevans, R. (2021). An Introduction to T-Tests | Definitions, Formula and Examples. Retrieved 5 May 2021, from https://www.scribbr.com/statistics/.

Center for Hearing and Communication. Statistics and Facts About Hearing Loss. Retrieved 7 March from https://chchearing.org/facts-about-hearing-loss/.

Codecacademy. What Is a Relational Database Management System? Retrieved 5 May from https://www.codecademy.com/articles/.

Codecacademy. What Is an IDE? Retrieved 28 April from https://www.codecademy.com/articles/what-is-an-ide.

Deafblind Support Philippines. Persons with Disabilities: Status in the Philippines. Retrieved 6 March from http://web.nlp.gov.ph/nlp/sites/default/files/.

Domantas, G. What is Web Hosting? Web Hosting Explained for Beginners. Retrieved 18 April 2021 from https://www.hostinger.com/tutorials/what-is-web-hosting/.

Educative. What is the Knuth-Morris-Pratt algorithm? Retrieved 5 June 2021 from https://www.educative.io/edpresso/what-is-the-knuth-morris-pratt-algorithm.

Farzand, A. Yamanes Formula - Sample Calculation. Retrieved 30 April 2021 from https://www.scribd.com/document/438058892/Yamanes-Formula-SampleCalculation.

FreeCodeCamp. The Rabin-Karp Algorithm Explained. Retrieved 07 June 2021 from https://www.freecodecamp.org/news/the-rabin-karp-algorithm-explained/.

Kent State University. Deaf Community Definition. (2021). Retrieved 29 April 2021 from Kent State University - Department of Modern and Classical Language Studies: https://www.kent.edu/mcls/deaf-community-definition.

Leonardo, B., Hansun, S. (2017). Text Documents Plagiarism Detection using Rabin-Karp and Jaro-Winkler Distance Algorithms. Retrieved 20 May from Rabin-Karp Algorithm | Download Scientific Diagram (researchgate.net).

LibGuides: SPSS Tutorials: Paired Samples t Test. (2021). Retrieved 5 May 2021, from https://libguides.library.kent.edu/SPSS/PairedSamplestTest.

Morpheme. (n.d.). In Literary Devices Definition and Examples of Literary Terms. Retrieved from https://literarydevices.net/morpheme/.

National Council on Disability Affairs. Public Elementary School. Retrieved from https://www.ncda.gov.ph/cebu/.

Opensource. What is open source? Retrieved 5 May from https://opensource.com/resources/what-open-source.

Orbay, A. & Akarun, L. (2020). Neural Sign Language Translation by Learning Tokenization. Retrieved 24 February from http://www.counseling.org.

Paired Sample T-Test - Statistics Solutions. (2021). Retrieved 5 May 2021, from https://www.statisticssolutions.com/manova-analysis-paired-sample-t-test/.

Phoneme. (n.d.). In Literary Devices Definition and Examples of Literary Terms. Retrieved from https://literarydevices.net/?s=phoneme.

Plaxton, G. (2004). String-Matching: Rabin-Karp Algorithm. Retrieved 20 May 2021 from https://www.cs.utexas.edu/~plaxton/c/337/04s/slides/StringMatching-1.pdf.

Running Time and Big-O - Learneroo. Retrived from https://www.learneroo.com/modules/106/nodes/559.

Smith, A. (2021, April). What Are Text Editors and Why Are They Important?. Retrieved 5 May from https://www.techstuffed.com/what-are-text-editors-and-why-are-they-impor tant/.

Software Testing Fundamentals (2020). Black Box Testing. Retrieved from https://softwaretestingfundamentals.com/black-box-testing/.

Time and Space Complexity Analysis of Algorithm. (2019). Retrieved from https://afteracademy.com/blog/time-and-space-complexity-analysis-of-algo rithm.

Tran, H. (2017). How to prove correctness of algorithm. Retrieved from https://medium.com/@tranduchanh.ms/partial-correctness-of-computer-pr ogram-f541490e7a21.

Umbao, E. (2014, August 7). Mininio Buhat: A Deaf Filipino Cyber Bullied Because of Faulty Written English Grammar. Philippine News. 35 https://philnews.ph/2014/08/07/mininio-buhat-deaf-filipino-cyber-bullied-because-faulty-written-english-grammar/.

What is IDE?. (n.d.). Retrieved from https://www.codecademy.com/articles/what-is-an-ide.

Videos

Bari, A. (2018). Knuth-Morris-Pratt KMP String Matching Algorithm [Video]. Retrieved from https://www.youtube.com/watch?v=V5-7GzOfADQ.

Bari, A. (2018). Rabin-Karp String Matching Algorithm [Video]. Retrieved from https://www.youtube.com/watch?v=qQ8vS2btsxI.

DeafED Philippines (2020, August 30). FILIPINO SIGN LANGUAGE EXPLAINED [Video]. YouTube. https://youtu.be/JYR8OgSViUY.

Inquirer.Net (2019, September 23). What we ought to know about Filipino Sign Language [Video]. YouTube. https://youtu.be/3yX4GN6wG0k.

Langmead, B. (2015). Boyer-Moore basics [Video]. Retrieved from https://www.youtube.com/watch?v=4Xyhb72LCX4.

Ramesh, B. (2020). Boyer Moore Pattern Matching Algorithm [Video]. Retrieved from https://www.youtube.com/watch?v=4Oj_ESzSNCk.

Government Legislation

Magna Carta for Disabled Persons 1992. (Ph) (Ph)

S.B. 2117, Sixteenth Congress of The Republic of The Philippines, First RegularSession(2011). http://legacy.senate.gov.ph/lisdata/1868815815!.pdf.

The Filipino Sign Language Act 2018. (Ph) s. No. 1455 (Ph).