KLIK: Kajian Ilmiah Informatika dan Komputer
ISSN 2723-3898 (Media Online)
Vol 4, No 1, Agustus 2023, Hal 164-172
DOI 10.30865/klik.v4i1.1090
https://djournals.com/klik

Searching and Comparing Isim Ma'rifat with Diacritic Removal in the Quran and Sahih Muslim Hadiths

Ryan Fahreza Maliki, Eko Darwiyanto*, Moch. Arif Bijaksana

[1]School of Computing, Software Engineering, Telkom University, Bandung, Indonesia
Email: [1]ryanfahrez@student.telkomuniversity.ac.id, [2,*]ekodarwiyanto@telkomuniversity.ac.id, [3]arifbijaksana@telkomuniversity.ac.id
Email Penulis Korespondensi: ekodarwiyanto@telkomuniversity.ac.id

Abstract−This research aims to address the scarcity of comprehensive websites providing detailed lists of Isim Ma'rifat in the Quran and Sahih Muslim Hadith. The absence of a comprehensive resource hinders the ability to study and compare Isim Ma'rifat between these significant Islamic texts. To overcome this issue, the study develops a natural language processing approach utilizing an integrated Java tokenizer program with a MySQL database containing the Sahih Muslim Hadith and Quranic texts. The program identifies the occurrence of the alif lam prefix, followed by diacritic removal to facilitate accurate verse comparison between the two texts. The research focuses on identifying alif lam prefixed Isim Ma'rifat exclusively present in the Quran, exclusive to Sahih Muslim Hadith, and similarities between them. The analysis yields a comprehensive understanding of the distinctions and similarities of alif lam prefixed Isim Ma'rifat between the Quran and Sahih Muslim. These findings provide valuable input for the Al-Quran project, contributing to the development of comprehensive and accessible resources for Islamic studies. It is expected that this research will enhance the understanding of Isim Ma'rifat in the religious and linguistic context, offering a significant contribution to the field of natural language processing especially in the Arabic language.

Keywords: Diacritics; Prefix; Sahih Muslim; Tokenizer; Quran

1. INTRODUCTION

Wikipedia is a website that provides discussion like an encyclopedia. Each object of discussion presented the results of the latest research on it. Ability to explain Wikipedia related to a topic discussion, reference for beginners to understand it. Search on the Quranpedia web, haven't found it yet Wikipedia-like appearance, see Appendix A. From here, the idea of Quranpedia emerged, where each the object (noun) which is the focus of working on Quranpedia in the Quran is sufficiently explained, explanation of relevant verses of Quran, explanation of relevant Kutubus Sittah hadiths, and explanation from Wikipedia itself. Project Quranpedia is an adaptation of Corpus based Quran benchmarks that have flaws, especially in the hadith. Sahih Muslim hadith is a collection book that contains the hadiths of the Prophet Muhammad. whose constituents are well known as trustworthy people because of their personal integrity and intellectual certainty. He is known as Imam Muslim. His full name is Abu Husain Muslim bin Hajjaj bin Muslim bin Kausyaz Al-Qusyairi an Naisaburi. He was born in Naisabur, Iran in 204 H. His book is very important to know, study, understand and make as a reference, especially for Muslims [21]. Arabic, as a Semitic language, possesses a unique morphology distinct from English and other languages. It holds a significant position as the language of the Holy Quran [1], which has been extensively translated into various languages over the past century. However, the interpretation of the Quran requires utmost sensitivity, considering the profound significance of each letter and its potential indications [2]. Automatic summarization has been the subject of extensive research, especially in the context of the Arabic language. Arabic is widely spoken worldwide, serving as the native language for nearly 300 million people and utilized by 1.2 billion Muslims in religious ceremonies. Its unique characteristics, such as its right-to-left script and rich vocabulary, make it an intriguing language for summarization research [3]. To facilitate the comprehensive listing of word classes, a software tool was employed to separate each sentence component based on its respective word class.

The database incorporates the display of word classes aligned with each inserted phrase or sentence, alongside their corresponding translations [4]. In particular, handling large volumes of Arabic text data demands robust database management systems capable of effectively managing and processing such data. MySQL, a widely used relational database management system, offers promising capabilities in facilitating the processing of extensive Arabic text datasets [5]. Arabic text preprocessing plays a crucial role in various natural language processing tasks, including information retrieval, sentiment analysis, and machine translation [6]. Among the key preprocessing steps, diacritic restoration and tokenization are vital for accurate linguistic analysis and semantic understanding of Arabic texts, will make it possible to move Arabic Natural Language Processing (NLP) forward and to facilitate the reuse of already existing preprocessing algorithmic resources [7]. a Java-based toolkit for the processing of Arabic text. It supports the most important preprocessing steps, such as diacritic and punctuation removal and tokenization [8].

Regarding the explanation, the process of searching for Isim Ma'rifat within the Quran and Sahih Muslim requires a tool to store data from each text in the books. MySQL is used as the tool to store the data, while Java is employed during the search process as a toolkit to facilitate preprocessing steps. These steps involve removing diacritics to enable comparison of each verse without considering vowels, as well as tokenizing each verse when encountering specific prefixes [9]. The objective of this study is to identify Isim Ma'rifat [10], [11], in the Quran and Sahih Muslim hadith, compare each verse, and assess the similarities between two verses using aggregate similarity [12]. Furthermore, the study aims to determine the extent to which the identified isim ma'rifat aligns with the nouns [13] identified using Java tools. This is achieved by leveraging a tokenizer and removing diacritics to enhance the depth of Arabic text processing. Incorporating the implemented word tokenizer and character tokenizer into the analysis of the Quran and Sahih Muslim

hadiths has enabled the successful detection of every alif lam prefix present in each verse. This has facilitated a thorough examination of the alif lam prefixes to determine whether they represent Isim Ma'rifat or not. To ensure the accuracy and reliability of the findings, a stringent validation process was employed, particularly when verifying the data found in Sahih Muslim, by conducting a comprehensive comparison, the research calculated the number of prefix similarities between the Quran and the hadiths of Sahih Muslim. This comparison has proven to be invaluable, as it simplifies the process of identifying nouns that begin with the letter alif lam, appearing in both the Quran and Sahih Muslim hadiths. The integration of advanced preprocessing techniques, such as diacritic removal and tokenization, has further enhanced the efficiency and accuracy of the program. These preprocessing steps have streamlined the analysis of the Arabic texts, allowing for a more focused examination of the sentence structure and word relationships within the texts. the utilization of tools and coupled with strict validation procedures, has enriched the research findings. By exploring the similarities and variations between the Quran and Sahih Muslim, this study has offered a comprehensive understanding of the alif lam prefixes present in both sources. The utilization of tools and coupled with strict validation procedures, has enriched the research findings. By exploring the similarities and variations between the Quran and Sahih Muslim, this study has offered a comprehensive understanding of the alif lam prefixes present in both sources.

2. RESEARCH METHODOLOGY

The overall system design based on their respective tasks is shown in figure 1.

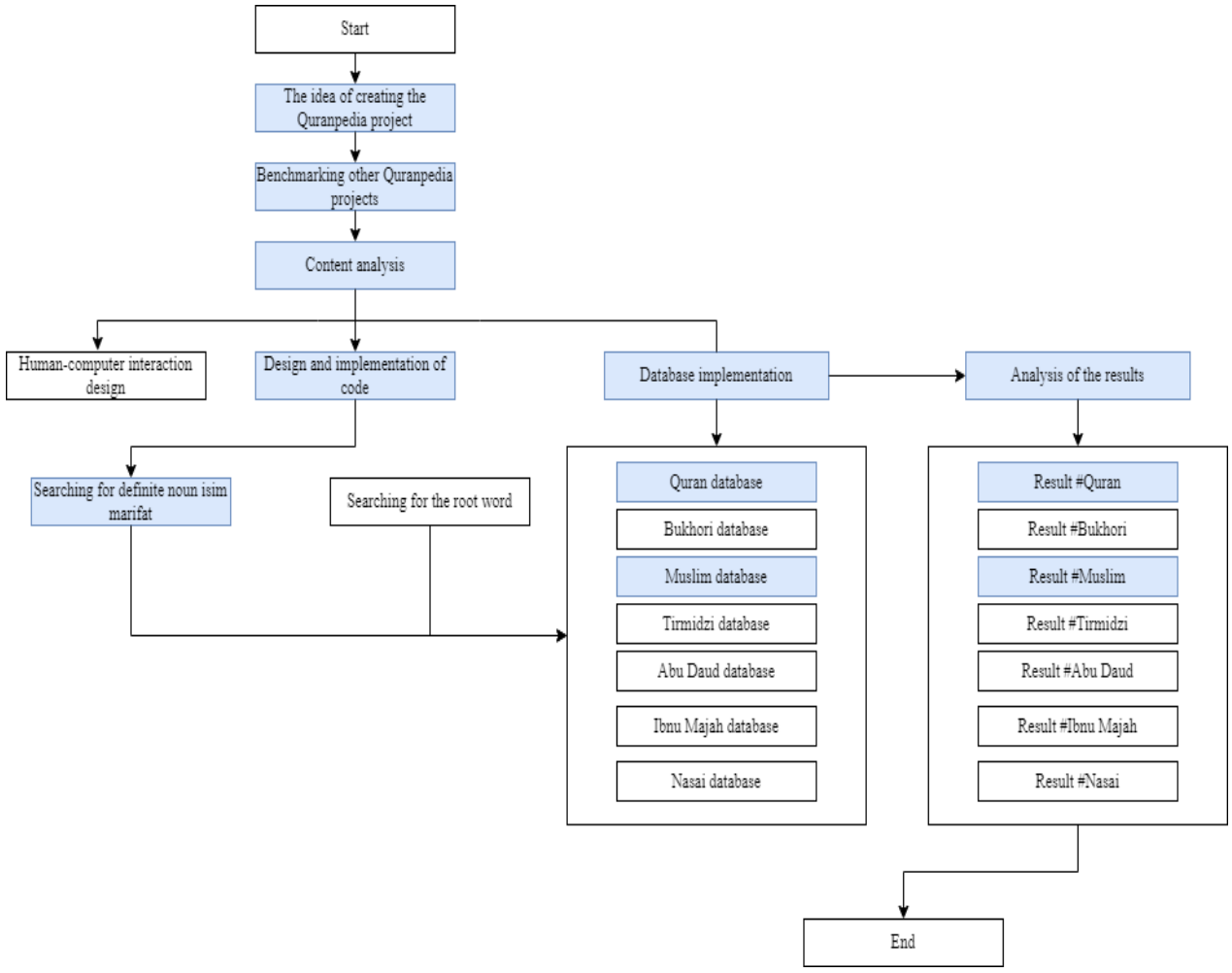

Figure 1. System Design

Figure 1 illustrates the process flow of the Quran project, which aims to develop a website capable of performing natural language processing on religious texts, specifically the Quran and Hadith. The primary focus of this project is to facilitate the retrieval of word roots and nouns from the texts. Each step in the process plays a distinct role in achieving this goal. The core objective of this work lies in the domain of natural language processing, wherein the system utilizes specific prefixes to identify and extract noun entities from the Quran and Hadith. These extracted data are subsequently compared, analyzed, and leveraged in the development of the website.

Flowchart of the General Process Illustrated in Figure 2.

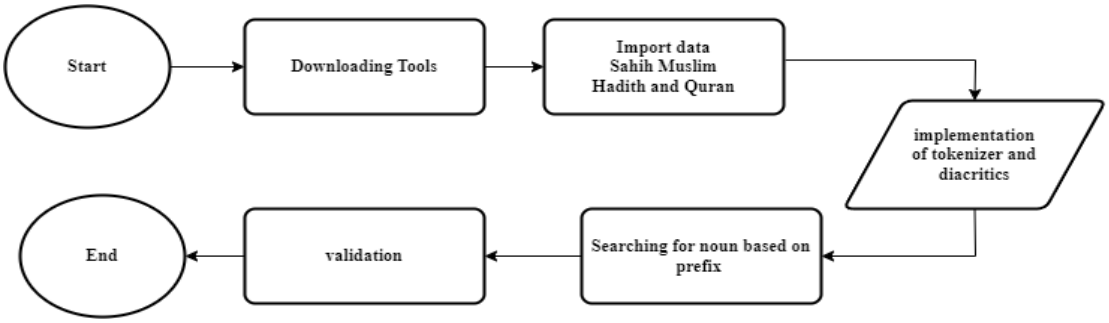

Figure 2. Flowchart General Process

2.1 Download Tools for Processing Text on Desktop

Table 1. Tools

Tools	Version	Download Source
NetBeans	17	https://netbeans.apache.org/download/index.html
XAMPP	7.4.27.2	https://www.apachefriends.org/download.html
MySQL Connector	8.0.33	https://dev.mysql.com/downloads/connector/j/?os=26

When installing the mentioned tools in Table 1, there are several considerations to bear in mind to ensure smooth program execution without any errors. These considerations will be explained below.

a. NetBeans version 17 for Java programming
   Ensure that the system meets the minimum requirements for running NetBeans. Download the NetBeans installer from Table 1 and follow the installation wizard. Select the appropriate JDK (Java Development Kit) version during the installation process. Configure the necessary settings, such as the installation directory and user preferences. Verify that NetBeans is successfully installed by launching the IDE and checking for any error messages.
b. XAMPP for Local Database Management
   Follow the installation wizard and select the desired components, such as Apache, MySQL, and PHP. Choose an installation directory and ensure that no conflicting services are already using the required ports. Once the installation is complete, start the XAMPP Control Panel and verify that the necessary modules (Apache and MySQL) are running. Configure the security settings for MySQL to ensure proper access and protection of database.
c. MySQL Connector for connectivity between Java and XAMPP
   Download the MySQL Connector/J from the official MySQL website on the Table 1, make sure that the downloaded JAR file is of version 8, as using any other version might result in errors. Extract the downloaded archive and locate the JAR file for the connector. In NetBeans, create a new Java project or open an existing one. Right-click on the project in the Projects view and select "Properties." In the project properties window, navigate to the "Libraries" tab and click on the "Add JAR/Folder" button. Browse to the location where extracted the MySQL Connector JAR file and select it. Click "OK" to confirm and apply the changes.

2.2 Importing Data from Sahih Muslim Hadith and the Quran

Table 2. Dataset

Public API database	Download Source
Quran	https://tanzil.net/download/
Sahih Muslim Hadith	https://github.com/irsyadulibad/hadits-database/blob/main/shahih-muslim.sql

In the process of importing data from Sahih Muslim Hadith and the Quran, the following steps were taken. The databases for Sahih Muslim Hadith and the Quran were downloaded from their respective public APIs. These databases contain the required data for further analysis.

a. XAMPP
   The next step involved importing the downloaded databases into XAMPP. First, the XAMPP Control Panel was opened, and both Apache and MySQL services were activated to ensure the local server is running.
b. Accessing phpMyAdmin
   To import the databases, phpMyAdmin was accessed through a web browser. PhpMyAdmin provides a user-friendly interface to manage MySQL databases. Within phpMyAdmin, a new database was created for storing the imported data. The name of the database was determined based on the project requirements.
c. Importing the Database
   Once the new database was created, the import functionality of phpMyAdmin was used to upload the downloaded databases. The "Import" option was selected, and the database was chosen for uploading. After selecting the database file, the import process was initiated by clicking the "Go" button. PhpMyAdmin read the file and executed the necessary SQL queries to create the tables and import the data into the newly created database.

2.3 Implementation of Tokenizer and Diacritics

The first step in any NLP (natural language processing) pipeline is to split the text into individual tokens [14] . Word Tokenizer splits words based on whitespaces [8]. In this research, the Word Tokenizer are implemented on the Quran and Sahih Muslim Hadith to search for the prefix "ال" (alif lam) [10]. These tokenizers are used to segment the text into smaller units [6], which allows for efficient identification and extraction of words with the alif lam prefix. Based in Figure 3 When applied to the Quran and Sahih Muslim Hadith, it helps to identify words that start with the prefix "ال" alif lam. For example, shown in Table 3, consider the following verse from the Quran.

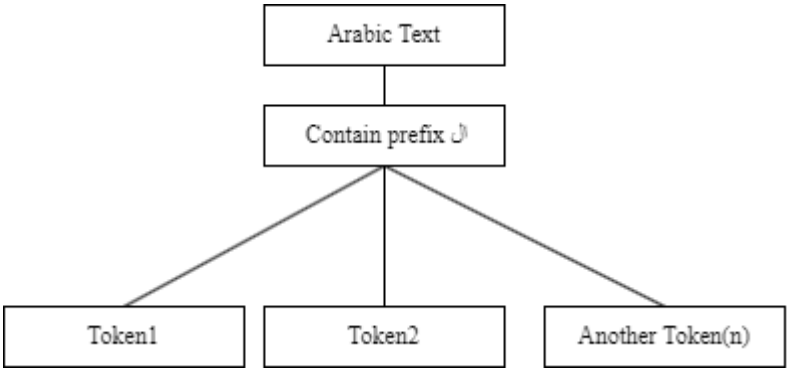

Figure 3. Process Tokenizer

Table 3. Example Word Tokenizer

Original Arabic Text	Token1	Token2	Token3	Another Token(n)
بِسْمِ اللَّهِ الرَّحْمَـٰنِ الرَّحِيمِ	اللَّهِ	لرَّحْمَـٰنِ	الرَّحِيمِ	…
مٰلِكِ يَوْمِ الدِّيْنِ۟	الدِّيْنِ۟	…	…	…
۞اِهْدِنَا الصِّرَاطَ الْمُسْتَقِيْمَ	الصِّرَاطَ	الْمُسْتَقِيْمَ	…	…

Arabic script consists of two classes of symbols: letters and diacritics. Letters comprise long vowels such as A, y, and w as well as consonants. Diacritics on the other hand comprise short vowels, gemination markers, nunation markers [15]. In this study, short vowel diacritics refer to the eight short vowels in Modern Standard Arabic (MSA) [15]. Table 4 consists of various diacritic marks commonly found in Arabic text. The diacritics, known as harakat in Arabic, serve the purpose of indicating short vowels in the verses found in the Quran and Sahih Muslim hadith. By removing the harakat, the analysis can focus on the sentence structure and word relationships within the text, which is relevant in searching for Isim Ma'rifat with the prefix alif lam. This process enables a more efficient and flexible implementation on large datasets, allowing for effective analysis and identification of relevant linguistic patterns.

Table 4. Diacritics

Unicode	Name	Pronunciation	symbol
\U064b	Fathatan	An	ً
\U064c	Dammatan	Un	ٌ
\U064d	Kasratan	In	ٍ
\U064e	Fatha	A	َ
\U064f	Damma	U	ُ
\U0650	Kasra	I	ِ
\U0651	Shadda	Doubling	ّ
\U0652	Sukun	None	ْ

Table 5. Example using Non-Diacritics

Token Using Diacritics	Token Non diacritics
الصِّرَاطَ	الصراط
الْمُسْتَقِيْمَ	المستقيم
اللَّهِ	الله
الرَّحْمَـٰنِ	الرحمٰن

3. RESULT AND DISCUSSION

3.1 Database

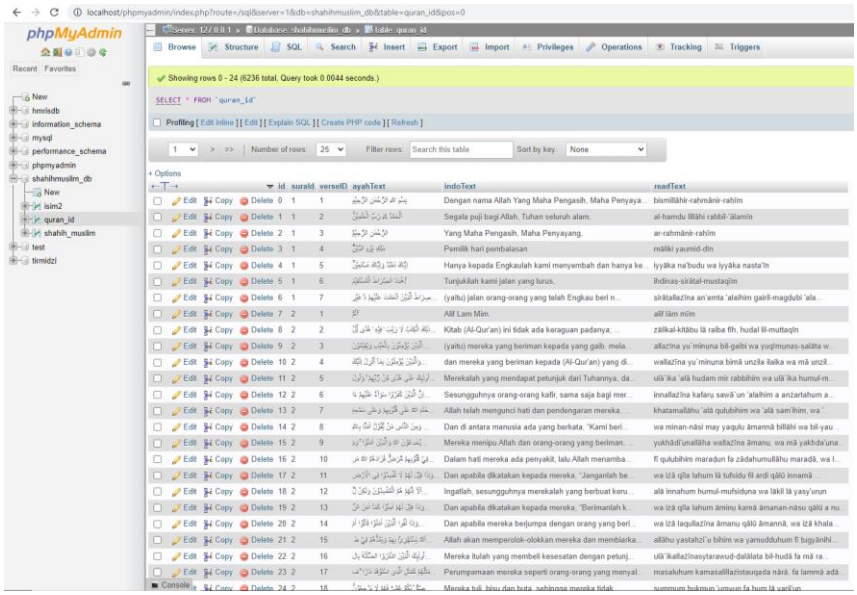

Figure 3. Database MySQL

Figure 3 depicts the process of dataset storage for Sahih Muslim Hadith and the Quran utilizing MySQL. This database serves as a testing platform for the implemented toolkit program. Each data entry from the Quran and Sahih Muslim Hadith undergoes tokenization based on verses containing the "alif lam" prefix. The Quran dataset comprises 6235 unique identifiers (IDs), encompassing translations and verses. Meanwhile, the Sahih Muslim Hadith dataset consists of 5362 unique identifiers (IDs) comprising both hadiths and translations.

3.2 Program Result

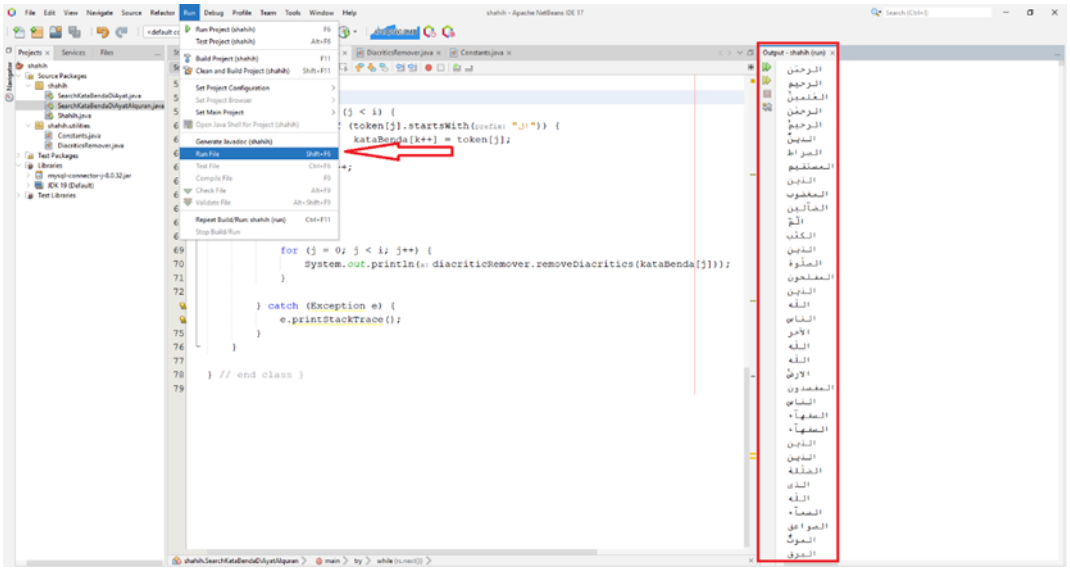

Figure 4. Program Tokenizer and Removal Diacritics

In Figure 4, the implementation of tokenization is evident, which involves searching for each dataset within the Quran and Sahih Muslim Hadiths containing the "alif lam" prefix. Each verse in the Quran or Sahih Muslim Hadith is transformed into a token once the program identifies verses with the "alif lam" prefix. These tokens are displayed as outputs based on predetermined IDs. Subsequently, the Arabic text is tokenized, and the diacritics are removed from each token. This critical step aims to facilitate the validation process, enabling the identification of similarities between each text present in the Quran and Sahih Muslim Hadiths. By eliminating diacritics, the validation becomes more efficient, especially when matching each token against a specific source, such as Corpus, which offers diverse types of post-taggers, streamlining the process of searching for source verses. This systematic approach ensures precise and comprehensive

analysis of the "alif lam" prefixed verses, enhancing the capability to discern Isim Ma'rifat, thus contributing to the advancement of Natural Language Processing (NLP) techniques in the context of Arabic texts. Moreover, by leveraging tokenization and diacritics removal, the research significantly enhances the program's accuracy in identifying patterns and associations between Quranic and Hadith verses, elevating the overall efficacy of the toolkit in uncovering linguistic and religious nuances within the sacred texts.

3.3 Result of Total Data based on Program

Table 6. Total Prefix in Two Sources

Number of Data	Result of total prefix (ال) in Quran	Number of Data	Result of Total Prefix (ال) in Sahih Muslim Hadith
1	ٱللَّهِ	1	الأثر
2	ٱلرَّحۡمَٰنِ	2	المشهور عن
3	ٱلرَّحِيمِ	3	الله
4	ٱلۡعَٰلَمِينَ	4	الله
5	ٱلرَّحۡمَٰنِ	5	الكاذبين
6	ٱلرَّحِيمِ	6	الحكم
7	ٱلدِّينِ	7	الرحمن
…	…	…	…
9851	الجنة	61290	الرحمن

In this section, Table 6. presents the results obtained from the Java program, which implemented tokenizer and diacritics removal. However, it should be noted that the total number of findings displayed in the "Number of Data" column 9851 data includes duplicated entries. In addition, another table, referred to as Table 6, is presented to display the results of the program's analysis on the Sahih Muslim Hadith collection, this column represents the total count of data found in the Sahih Muslim Hadith collection. The value displayed is 61290 data, indicating the number of entries processed during the program's execution.

3.4 Sorting and Removing Duplicate data in Quran and Sahih Muslim

For the Quran and Sahih Muslim, the table includes a column labeled "Unique Prefix (ال) in Quran." And "Unique Prefix (ال) in Sahih Muslim." Before removing duplicates, the total number of data entries in the Quran was 9851 data. After applying the sorting and duplicate removal process, the number reduced to 1635 data unique entries. The sorting and removal of duplicates were performed using the feature in Microsoft Excel. Similarly, for the Sahih Muslim Hadith collection, this table presents the results after sorting and removing duplicate entries using the same approach. Prior to removing duplicates, the total number of data entries in Sahih Muslim was 61290. After the sorting and duplicate removal process, the number reduced to 3506 unique entries shown in Table 7. The sorting and removal of duplicates ensures that the final dataset contains only distinct entries, eliminating any repeated data. This step is crucial in obtaining accurate and reliable results for further analysis and interpretation.

Table 7. Prefix Unique Value

Sorting and Remove Duplication			
Number of Data Quran	Unique prefix (ال) in Quran	Number of Data Sahih Muslim	Unique prefix (ال) in Sahih Muslim
1	الٓرٰ	1	الابتداء
2	الٓمٓ	2	الاثنين
3	الٓمٓرٰ	3	الاحتلام
4	الٓمٓصٓ	6	الاختتان
5	الٓنٓ	7	الاستجمار
6	الٓيٓ	8	الاستسقاء
7	الٓيٓ	9	الاستعجال
…	…	…	…
1635	ٱلۡيَوۡمٌ	3506	اليومين

3.5 Comparing (ال) prefix in Quran and Sahih Muslim Hadith

Table 8. Comparing Prefix Two Source

Comparing (ال) prefix					
Data Similar	Similarities Found in the Quran and Sahih Muslim	Data Quran	Found Only in the Quran	Data Sahih Muslim	Found Only in Sahih Muslim
1	الاسم	1	الٓرٰ	1	الابتداء
2	الإبل	2	الٓمٓ	2	الاثنين
3	الإثم	3	الٓمٓرٰ	3	الاحتلام

4	الإحسان	4	الْمَصَ	4	الاختتان
5	الإربة	5	الٓمٓ	5	الاستجمار
6	الإسلام	6	الٓمٓيٓ	6	الاستسقاء
7	الإنجيل	7	الٓمٓيٓ	7	الاستعجال
…	…	…	…	…	…
447	اليوم	1188	اليومٌ	3059	اليومين

In this section, compare the occurrences of the prefix (ال) in the Quran and the Sahih Muslim Hadith collection. Table 8 presents the results of this comparison, consisting of three main points.

a. Similar Prefix (ال) in Quran and Sahih Muslim
   The analysis reveals that there are 447 instances where the prefix (ال) appears in both the Quran and the Sahih Muslim Hadith collection. These instances indicate similarities between the two sources.
b. Prefix (ال) Unique to the Quran
   After separating and comparing the data, it was found that there are 1188 occurrences of the prefix (ال) that are exclusively present in the Quran. These instances are not found in the Sahih Muslim Hadith collection.
c. Prefix (ال) Unique to Sahih Muslim
   Similarly, the analysis identified 3059 occurrences of the prefix (ال) that are exclusively present in the Sahih Muslim Hadith collection. These instances are not found in the Quran.

To separate and compare each data point, the VLOOKUP function in Excel was utilized. For example, the first data point from the Quran was compared with all the data points in the Sahih Muslim Hadith collection. If a match was found, a new table titled "Similarities Found in the Quran and Sahih Muslim" was created to store the corresponding data. On the other hand, if the data point did not match, it was separated into the "Found Only in the Quran" table. The same process was repeated for each data point in the Sahih Muslim Hadith collection, comparing it with the Quran. By applying this approach, we can accurately identify the occurrences of the prefix (ال) in both the Quran and the Sahih Muslim Hadith collection and categorize them based on their similarities or uniqueness to each source. This comparison provides valuable insights into the overlap and distinctions between the two texts.

3.6 Isim Ma'rifat Noun Prefix Validation (ال)

In this section, validation for the Isim Ma'rifat (noun with prefix ال) was found in the Quran. To ensure the accuracy of our findings,. This paper provided valuable insights into the classification of Isim Ma'rifat and identified seven types, including using the prefix alif lam [11]. POS Tagging is a major step in most NLP applications to determine the proper grammatical tag or syntactic category of a word depending on its context [16]. For example, the words in the sentence: "Ali has the can" are tagged as noun, verb, determiner, and noun respectively[17]. However, some words have more than one tag with respect to the context. In this study Tables 9 and 10 each prefix found was given 3 labels including muqatta'at, noun and not found. not found because it does not have a definite entity in the source being tested.

Table 9. Validation Quran

Validation Of Quran Only			
Number of Data	Prefix	Part of speech	The Source of Verses
1	الٓرٰ	muqatta'at	-
2	الٓمٓ	muqatta'at	-
178	البارئ	Noun	Al-Hasyr verse 24
189	البسط	Noun	Al-Isra verse 29
194	البصير	Noun	Al-An'aam verse 50
…	…	…	…
1188	اليومٌ	Not Found	…

Table 10. Validation Similarities

Validation Of Similarities The Quran And Sahih Muslim			
Number of Data	Prefix	Part of speech	The Source of Verses
1	الاسم	Noun	Al-Hujuraat verse 11
2	الإبل	Noun	Al-An'aam verse 144
3	الإثم	Noun	Al-Baqarah verse 85
4	الإحسان	Noun	Ar-Rahman verse 60
5	الإربة	Not Found	-
6	الإسلام	Noun	Ali-Imran verse 19
7	الإنجيل	Noun	Al-Maidah verse 46
…	…	…	…
447	اليوم	Noun	Al-Baqarah verse 249

3.7 Summary Count

All information collected from the previous step is important to calculate the similarity for each sentence [7] Refer to the data provided in Table 11, based on these calculations, it can be concluded that approximately 27.35% of the "alif lam" prefixes found in the Quran also have similarities in Sahih Muslim Hadith. Similarly, about 12.75% of the "alif lam" prefixes found in Sahih Muslim Hadith also have similarities in the Quran. The analysis of these percentages indicates a significant correlation between the use of "alif lam" prefixes in the Quran and Sahih Muslim Hadith.

In Table 12, present the calculation of valid Isim Ma'rifat, it is noted that among these Isim Ma'rifat some of them contain the (ال) prefix [11] [14] . and refers to the golden standard corpus in the lemma frequency section [18]. The analysis of the Quran data reveals a total of 1635 instances of the (ال) prefix. After conducting the validation process, 724 instances were confirmed as valid Isim Ma'rifat with the (ال) prefix. This indicates that approximately 45.44% of the nouns with the (ال) prefix in the Quran are valid and based on similarity data found in Sahih Muslim hadiths of 12.20%, some nouns that are not found within the context have separate linguistic entities with meanings [19], [20] . In the comparison between the Quran and Sahih Muslim Hadith, a total of 447 instances of the (ال) prefix were identified. Following the validation process, 428 instances were determined to be valid Isim Ma'rifat. This implies that approximately 95.74% of the nouns with the (ال) prefix found in both the Quran and Sahih Muslim Hadith are valid, .

Table 11. Score Percentage Similarities

Summary Count Similarities				
Prefix (ال)	Total Prefix (ال)	Only Found Quran / Sahih Muslim	Similarities between Quran and Sahih Muslim	Percentage Similarities
Quran	1635	1188	447	27,35%
Sahih Muslim Hadith	3506	3059		12.75%

Table 12. Score Percentage System

Summary Count Validation System			
Prefix (ال)	Total Prefix (ال)	Total Prefix (ال) Valid Noun	Percentage Valid
Quran	1635	724	45,44%
Sahih Muslim (validation Quran)	3506	428	12,20%
Similarities Quran and Sahih Muslim Hadith	447	428	95,74%

4. CONCLUSION

In conclusion, this study demonstrates that utilizing Java tokenizer and diacritic removal tools enables the identification of Isim Ma'rifat (definite nouns) with the alif lam prefix (ال). The obtained results reveal that the program successfully identifies 45.44% of the Isim Ma'rifat present in the Quran. Moreover, the comparison of the alif lam prefix (ال) between Sahih Muslim Hadith and the Quran demonstrates a high level of validity, reaching 95.74%. Therefore, it can be inferred that the shared prefix between the two books represents valid Isim Ma'rifat. However, it should be noted that the absence of certain prefixes in the validation results is attributed to the fact that these prefixes have separate linguistic entities with different meanings or belong to the category of Muqattaʿat (disjointed letters). It is important to consider that not all words with the alif lam prefix (ال) are nouns, as some may represent Muqattaʿat or belong to other word categories. This study provides valuable insights into the identification and validation of Isim Ma'rifat with the alif lam prefix (ال) in the Quran and Sahih Muslim Hadith. Further research is encouraged to explore the linguistic nuances and variations in the usage of these prefixes to enhance our understanding of the Arabic language and its textual analysis, This study uses the Isim Ma'rifat with the prefix alif lam, there are several other prefixes with different prefixes in the isim ma'rifat, this research can be further developed to look for other prefixes, and validation can be improved in certain contexts.

REFERENCES

[1] F. Beirade, H. Azzoune, and D. E. Zegour, "Semantic query for Quranic ontology," Journal of King Saud University - Computer and Information Sciences, vol. 33, no. 6, pp. 753–760, Jul. 2021, doi: 10.1016/j.jksuci.2019.04.005.

[2] M. N. Al Salem, S. Alghazo, I. Alrashdan, N. Abusalim, and M. Rayyan, "On English translation variation of similar plural nouns in the Holy Quran," Cogent Arts Humanit, vol. 10, no. 1, 2023, doi: 10.1080/23311983.2023.2196136.

[3] A. M. Abu Nada, E. Alajrami, A. A. Al-Saqqa, and S. S. Abu-Naser, "Arabic Text Summarization Using AraBERT Model Using Extractive Text Summarization Approach", [Online]. Available: www.ijeais.org/ijaisr

[4] T. R. M. Romli, M. Z. Othman, M. H. Abdullah, and M. Z. A. Hamat, "Word Classification in the Online Database of Malay-Arabic Comparable Phrases," International Journal of Academic Research in Progressive Education and Development, vol. 7, no. 4, Nov. (2018), doi: 10.6007/ijarped/v7-i4/4853.

[5] S. S. Saloum, "DAD: A Detailed Arabic Dataset for Online Text Recognition and Writer Identification, a New Type," Journal of Computer Science, vol. 17, no. 1, pp. 19–32, (2021), doi: 10.3844/jcssp.2021.19.32.

[6] M. Alkaoud and M. Syed, "On the Importance of Tokenization in Arabic Embedding Models," 2020. [Online]. Available: https://github.com/attardi/wikiextractor

[7] A. Hallberg, "Principles of variation in the use of diacritics (taškīl) in Arabic books," Language Sciences, vol. 93, Sep. 2022, doi: 10.1016/j.langsci.2022.101482.
[8] Z. Alyafeai Dhahran, S. Arabia, and M. S. Al-Shaibani Dhahran, "ARBML: Democratizing Arabic Natural Language Processing Tools," (2020).
[9] R. G. Disclaimer, "The Mysterious Disjointed Letters in The Qur'an: Evidence of Divine Authorship", doi: 10.13140/RG.2.2.12311.60327.
[10] B. Justice and M. Ahmad Mughal, "Kinds of Alif, Lām (: [Online]. Available)." **یک الم الف ا ق سام** http://ssrn.com/author=1697634Electroniccopyavailableat:https://ssrn.com/abstract=3655831
[11] M. Zakki Mubarok, M. Irham, and S. Darul Fattah Bandar Lampung, "Analisis Isim Ma'rifat dalam Al-Qur'an Surat Ash-Shaff," (2021).
[12] A. A. Amer, M. H. Mohamed, and K. Al_Asri, "ASGOP: An aggregated similarity-based greedy-oriented approach for relational DDBSs design," Heliyon, vol. 6, no. 1, Jan. (2020), doi: 10.1016/j.heliyon.2020.e03172.
[13] A. Al Qifari, "Shaut Al-'Arabiyah Nakirah dan Ma' Rifah Fii Al-Qur'an," vol. 10, no. 1, 2022, doi: 10.24252/saa.v10i1.29432.
[14] Z. Alyafeai, M. S. Al-shaibani, M. Ghaleb, and I. Ahmad, "Evaluating Various Tokenizers for Arabic Text Classification," Jun. (2021), [Online]. Available: http://arxiv.org/abs/2106.07540
[15] O. Hamed and T. Zesch, "The Role of Diacritics in Adapting the Difficulty of Arabic Lexical Recognition Tests." [Online]. Available: http://www.rdi-eg.com/RDI/TrainingData/
[16] A. Alhasan and A. T. Al-Taani, "POS Tagging for Arabic Text Using Bee Colony Algorithm," in Procedia Computer Science, Elsevier B.V., 2018, pp. 158–165. doi: 10.1016/j.procs.2018.10.471.
[17] Q. Bsoul, R. A. Salam, J. Atwan, and M. Jawarneh, "Arabic Text Clustering Methods and Suggested Solutions for Theme-Based Quran Clustering: Analysis of Literature," Journal of Information Science Theory and Practice, vol. 9, no. 4, pp. 15–34, 2021, doi: 10.1633/JISTaP.2021.9.4.2.
[18] N. Hizbullah and A. Mutaali, "Quranic Corpus Models for Corpus-Based Studies," (2019).
[19] "Exploring the Meaning of Huroof-e-Muqatta'at (Abbreviated / Disjointed Letters) in the Quran Exploring the Meaning of Huroof-e-Muqatta'at (Abbreviated / Disjointed Letters) in the Quran Exploring the Meaning of Huroof-e-Muqatta'at (Abbreviated / Disjointed Letters) in the Quran," (2021).
[20] L. Nahda Sahib Hashim, "The Muqatta'at Disjointed Letters in the Holy Qur'an Analytical Study."(2019)
[21] Muhammad Misbah,etc "STUDI KITAB HADIS: Dari Muwaththa' Imam Malik hingga Mustadrak Al Hakim"(2020)