

# Search and Comparison of Isim Ma'rifat with Remove Diacritic in the Qur'an and Hadith of Abu Daud

T M Ikhsan\*, Eko Darwiyanto, Moch Arif Bijaksana

School of Computing, Software Engineering, Telkom University, Bandung, Indonesia

Email: <sup>1,\*</sup>tmuhammadikhsan@students.telkomuniversity.ac.id, <sup>2</sup>ekodarwiyanto@telkomuniversity.ac.id, <sup>3</sup>arifbijaksana@telkomuniversity.ac.id

Correspondence Author Email: tmuhammadikhsan@students.telkomuniversity.ac.id

Submitted: 23/07/2023; Accepted: 21/08/2023; Published: 25/08/2023

**Abstract**—This research aims to address the issues of searching for and comparing Isim Ma'rifat (definite nouns) with the prefix "AL" (ل) in the Quran and Hadith Abu Daud. Accurate sources of information regarding nouns in Islam are crucial, and the sayings of the Prophet (Hadith) serve as the primary reference for explaining Isim Ma'rifat. The Remove Diacritics method allows the removal of punctuation marks in the Arabic language, facilitating the search for Isim Ma'rifat with the prefix "AL" (ل) in both sources. The primary objective of this study is to evaluate the ability of Hadith Abu Daud to explain the Isim Ma'rifat found in the Quran. The comparison results reveal 376 instances of identical Isim Ma'rifat between the Quran and Hadith Abu Daud. Additionally, there are 611 instances of Isim Ma'rifat found solely in the Quran and 1388 instances found exclusively in Hadith Abu Daud. These findings provide insights into the capacity of Hadith Abu Daud to explain Isim Ma'rifat in the Quran. The search and comparison of Isim Ma'rifat in the Quran and Hadith Abu Daud hold significant contributions to the development of Quranpedia. Valid information about Isim Ma'rifat and the comparisons between the Quran and Hadith Abu Daud can enrich the content on Quranpedia. In creating the Quranpedia website, the search results for Isim Ma'rifat offer accurate information about nouns mentioned in the Quran, aiding users in better understanding the meanings and uses of Isim Ma'rifat in the religious context. The comparison between the Quran and Hadith Abu Daud provides insights into Hadith Abu Daud's ability to explain Isim Ma'rifat in the Quran. This information can be presented as articles or comparisons to enrich users' knowledge about the relationship between the Quran and hadith concerning the use of nouns in Islam.

**Keywords:** Isim Ma'rifat; Remove Diacritics; Al-Quran; Hadith Abu Daud

## 1. INTRODUCTION

Wikipedia is a very popular online encyclopedia around the world, being a source of information that is frequently visited by millions of people every day [1]. The concept of Wikipedia as a collaborative platform for sharing knowledge has inspired the idea of Quranpedia. Quranpedia is a website that aims to provide information about noun data or Isim Ma'rifat contained in the Al-Quran and the Kutubus Sittah Hadith. Isim Ma'rifat, or which is also known as a noun in Indonesian, refers to anything that can be categorized as an object that has a clear meaning and a form that can be felt by everyone. One of the characteristics of Isim Ma'rifat is the prefix AL(ل) [2].

Al-Qur'an linguistically comes from the word qara'a - yaqra'u - qira'atan - qur'an, which is something that is read or recited. Meanwhile, in terminology, it is the Kalamullah that was revealed to the Prophet Muhammad. and reaches us mutawatir, and reading it functions as worship[3]. The Al-Quran holy book consists of 30 chapters, which include 114 letters, 6,236 verses, and 77,845 words [4]. Hadith is one of the main sources of Islamic teachings after the Qur'an. The term "hadith" is often used to refer to everything that is associated with the Prophet Muhammad [5], such as words, actions, agreements, and their nature both physically and psychologically, both that occurred before and after he became a Prophet. Sometimes the term "hadith" is also used interchangeably with the term "sunnah" [6]. Some hadith scholars consider the two terms synonymous, while others make a distinction between the two [7]. Abu Dawud is a hadith scholar who is famous for his Sunan book. He was born in Sijistan in 202 H and died in 275 H. The book of Sunan Abu Dawud is named after his name. He is one of the scholars who is recognized for his expertise in the field of hadith, as evidenced by the many works he has written. Abu Dawud did not explain the criteria for the validity of the hadith in direct detail. However, in determining its validity, he is almost the same as the approach used by the majority of hadith scholars. The narrators who narrated the hadith have a teacher-student relationship. Abu Dawud in assessing the hadith in his Sunan book divides it into five categories, namely *Ṣaḥīḥ*, *yusybihuhū* (which resembles *ṣaḥīḥ*) which according to hadith scholars is called *ṣaḥīḥ liḡairihi*, *yuqāribuhu* (which is close to *ṣaḥīḥ*) which according to hadith scholars is called *ḡasan liḡāṭihi*, *wahn syadīḡun* (hadith which very weak), and *Ṣāliḡ* which according to hadith scholars is called *ḡasan liḡairihi*. Abu Dawud in compiling his book only includes 2-4 hadiths in one chapter [8]. Arabic, being the language of the Qur'an, has a long history [9]. Several theories try to explain the origins of Arabic. The first theory suggests that the Prophet Adam as was the first human to use Arabic. In this view, the Prophet Adam used Arabic when he was still in heaven before coming down to earth. In several narrations it is stated that the language used by the inhabitants of heaven is Arabic, so it can be concluded that the Prophet Adam and his descendants used Arabic. This is one of the reasons why Arabic became the language of the Qur'an [10].

The main sources for explaining Isim Ma'rifat are the hadiths of the Prophet, one of which is the Hadith of Abu Daud. Isim Ma'rifat, is often found in hadiths and the Al-Quran, including the Hadith of Abu Daud. In the research conducted by [11], it is stated that all words starting with the prefix "اَل" in Arabic are considered Isim or nouns, except for the letters of Muqata'at, which are fourteen specific letters (e.g., Alif Lam Mim, Alif Lam Raa, Ta Ha, Ya, Sin, etc.). However, it is essential to validate this claim by examining whether there are any exceptions among words with the "اَل" prefix that are not Isim. Additionally, another study carried out by [2] utilized a light stemming algorithm to determine the types of words in the Arabic language, including Isim (nouns) and Fi'il (verbs). This research achieved an impressive average accuracy rate of 97.54%. Regrettably, the study did not provide specific findings on the identified Isim or Fi'il, which would have been valuable as a reference. Therefore, it is important to have a system that can search and find Isim Ma'rifat in the Al-Quran and Hadith of Abu Daud. In this case, an efficient and accurate search mechanism is needed. In addition, the data found by Isim Ma'rifat are similar to the Al-Quran, it is necessary to look for an explanation in the Hadith of Abu Daud, to find out the extent to which the Hadith of Abu Daud can explain Isim Ma'rifat contained in the Al-Quran. To overcome this problem, the author wants to offer a solution in the form of designing a search system for Isim Ma'rifat in the Al-Quran and Abu Daud Hadith that uses the prefix AL (اَل). The method used in this system is Remove Diacritics, which is a method that allows the removal of vowels or punctuation in Arabic. This algorithm will look for words that have the prefix AL (اَل) in the verses of the Al-Quran and Hadith of Abu Daud. The search results will be displayed as data or information on the Quranpedia website.

In this context, the author intends to conduct research entitled "SEARCH AND COMPARISON OF ISIM MA'RIFAT WITH REMOVE DIACRITIC IN THE QURAN AND ABU DAUD HADITH" which aims to develop a search system for Isim Ma'rifat with the prefix AL (اَل) in the Quran and Abu Daud Hadith using the Remove Diacritics method.

## 2. RESEARCH METHODOLOGY

### 2.1 Research Steps

At the system design stage, a design was carried out using a flowchart to be able to produce a system that can search for and compare nouns (Isim) in the Al Quran and Abu Daud Hadith which can be used on the Quranpedia website. The output from Abu Daud's Hadith will be used to explain verses that contain the same noun.



Figure 1. Flowchart of Ma'rifat (اَل) Search System

### 2.2 Remove Diacritics

Diacritics in the Indonesian dictionary are defined as additional marks on letters that slightly change the phonetic value of the letters [12]. In the context of the current study of Al-Quran sciences, it is stated that the science which discusses diacritical signs in the Qur'an is called al-dabt science or the science of naqt wa syakal. This

knowledge cannot be separated from the science of rasm mushaf because the relationship between letters and punctuation (diacritics) functions to complement each other so that it makes it easier for someone to read the Al-Quran [13].

Diacritical punctuation or vowels are written above or below the letters. It should be noted that the speed of typing in Arabic script can be halved if we include diacritics in the written text [14]. The diacritic removal algorithm works to remove diacritical punctuation marks or vowels. The aim is to simplify the search process for Isim Ma'rifat (ال) and reduce duplication. For example, in the word "الله" (Allah) if you don't remove the diacritics, the program will find variations such as "الله" (Allaha), "الله" (Allahi), and "الله" (Allahu) which all mean the same thing, i.e. Allah. This can cause duplication in search results for Isim Ma'rifat (ال) which is not needed.

### 2.3 Unicode

In computing, character encoding techniques are the most appropriate working standards used to represent character sets for both storage and transmission. These techniques were developed alongside the Universal Character Code Standard (UCS) and published as the Unicode Standard. The latest version of Unicode contains more than 128,000 characters including 135 modern and historical scripts, as well as several symbol sets, with many encoding types such as ANSI, UTF-8, UTF-16, UTF-32, and so on [15]. Unicode is designed to represent almost all characters in every language in the world. All Arabic characters are now encoded according to the Universal Unicode Principles [16]. Unicode is large enough to cover all the characters likely to be used in a common text exchange, including characters in the major international, national, and industry character sets. The unicode technique requires more space in memory when storing [17].

### 2.4 Quranpedia

Quranpedia is a website that displays a noun and root words using Arabic. The noun (isim) is found in the Al-Quran and hadith based on explanations in verses of the Al-Quran and explanations of Kutubus Sittah hadiths. This Quranpedia is inspired by one of the websites, namely Wikipedia, which provides very complete sources of knowledge and includes a very wide discussion. It's just that this Quranpedia topic is for knowing nouns and root words in the Al-Quran and Hadith.

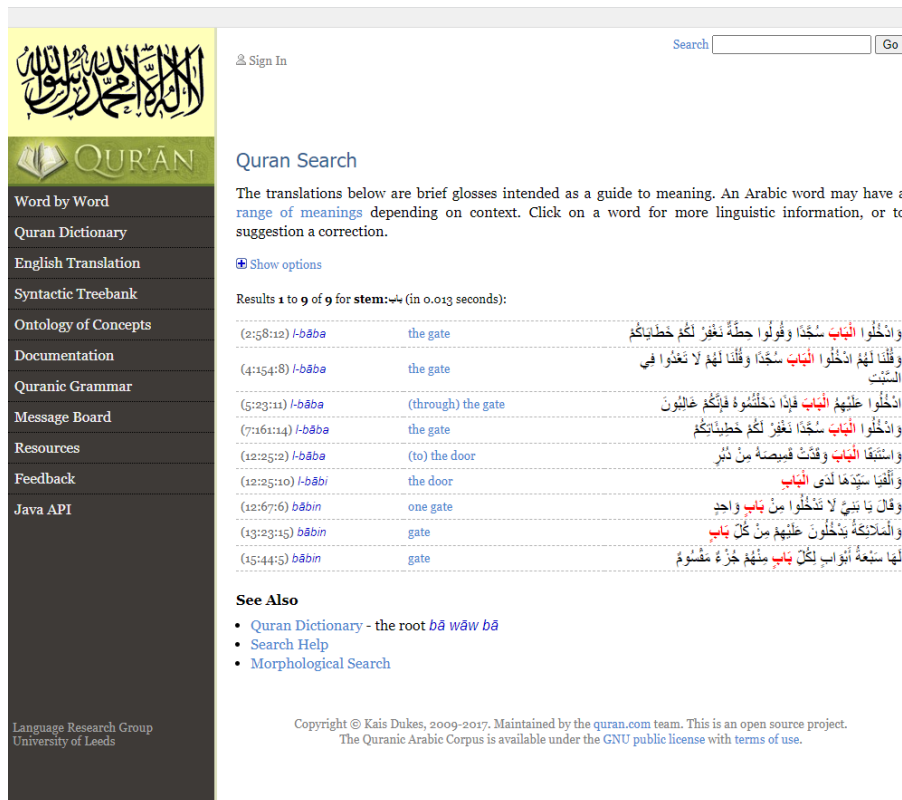


Figure 2. Word search results on the corpus.quran.com site

It can be seen from the noun search in Figure 2 above. The word to search for is a noun door or in Arabic chapter. There are also sentences in the Al-Quran which contain the noun door in them. For Quranpedia it's easier to use because you don't need to search anymore, just click on it and it immediately displays a page like the one on corpus.quran.com

### 2.5 Website Benchmark

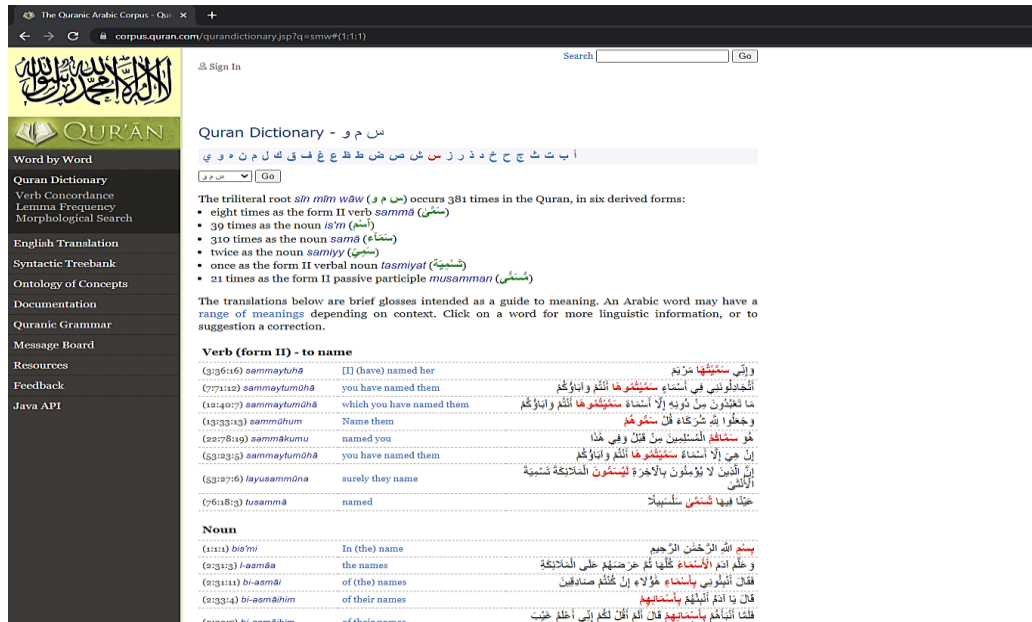


Figure 3. View corpus.quran.com/qurandictionary

The Corpus Quran website (Figure 3 Views corpus.quran.com/qurandictionary) is the main source of reference in creating the Quranpedia project. This site provides a linguistic explanation of the Al-Quran as well as features that allow users to display the words of each verse, such as word-by-word dictionaries, and so on. Quranpedia will collect Isim Ma'rifat (ال) data from the Al-Quran and the Kutubus Sittah Hadith, as well as display verses that have the same root word.

### 3. RESULT AND DISCUSSION

#### 3.1 Solution Design

To overcome this problem, there are several proposed solutions, including:

1. Build a database containing the texts of the Al-Quran and the Hadith of Abu Daud.
2. Develop a program for removing diacritics to make text analysis easier.
3. Validate any Isim Ma'rifat (ال) findings found.
4. Measuring the level of accuracy of the algorithm used.
5. Process findings data to look for similarities and differences between the data.
6. Assessing the ability of Abu Daud's Hadith in explaining the Isim Ma'rifat (ال) contained in the Al-Quran.

#### 3.2 Implementation

1. Successfully create a database of Hadith Abu Daud.

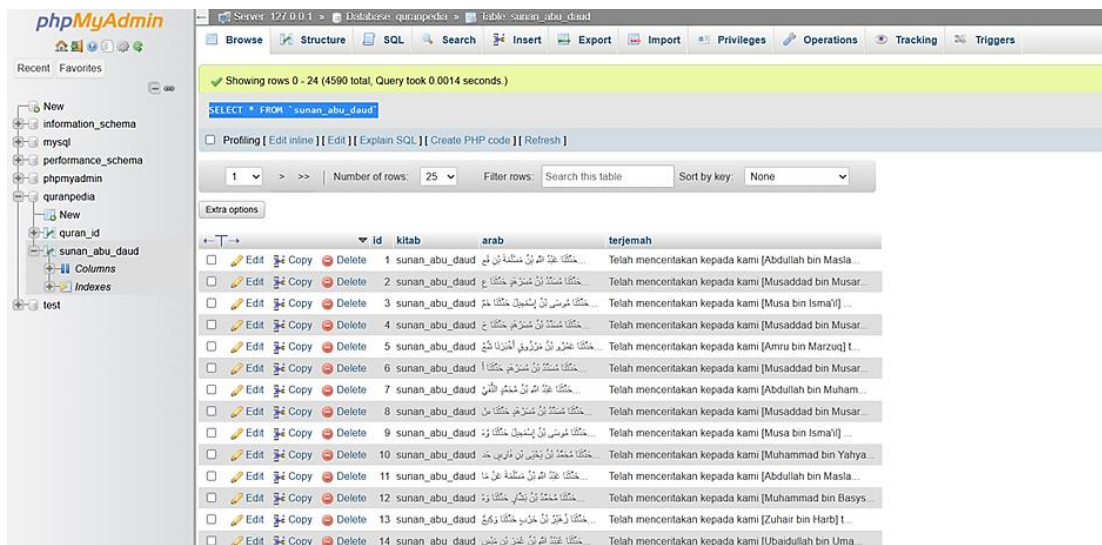


Figure 4. Database screenshots from H.R. Abu Dawud. Data source: carihadis.com

## 2. Remove Diacritics Algorithm

```
import daud.DiacriticsRemover;
import java.io.*;
import java.sql.*;
import java.util.StringTokenizer;

public class daud {

    public static void main(String args[]) {
        String DB_URL = "jdbc:mysql://localhost/quranpedia?characterEncoding=UTF-8";
        String DB_USER = "root";
        String DB_PASS = "";
        Connection conn;
        Statement stmt;
        ResultSet rs;
        DiacriticsRemover diacriticRemover = new DiacriticsRemover();
        try {
            // Set the console's encoding to UTF-8
            System.setOut(
                new PrintStream(new FileOutputStream(FileDescriptor.out), true, "UTF-8")
            );

            //Untuk menyimpan token, dan kata benda
            String[] token = new String[600_000];
            String[] kataBenda = new String[600_000];
            String[] kataBendaWa = new String[600_000];
            String[] kataBendaBa = new String[600_000];
            StringTokenizer tokenizer;

            int i, j, k;
            //inisialisasi token
            for (i = 0; i < 600_000; i++) {
                token[i] = "";
                kataBenda[i] = "";
            }
        }
    }
}
```

Figure 5. Main Function

To get Isim Ma'rifat with the prefix "AL" (اَل), the startsWith() method is used in line 64 which functions to check whether the string value starts with the specified prefix, in this case, the prefix (اَل). When the program encounters a prefix (اَل), it tokenizes each verse. tokenizing is a process that converts a sentence into several words (strings) [18]. The search was carried out on the database of Al-Quran and Hadith Abu Daud which had been prepared.

```
package daud;
import java.util.ArrayList;

public class DiacriticsRemover {

    // remove all diacritics from text
    public String removeDiacritics(String currentText) {
        ArrayList<String> tokens = new ArrayList<String>();
        StringBuffer modifiedWord = new StringBuffer("");
        ArrayList<String> modifiedText = removeExtraSpaces(currentText);

        // for each token in the text
        for (int i = 0; i < modifiedText.size(); i++) {
            modifiedWord.setLength(0);
            String ctoken = modifiedText.get(i);
            for (int j = 0; j < ctoken.length(); j++) {
                if (!Constants.diacritics.contains(ctoken.substring(j, j + 1))) {
                    modifiedWord.append(ctoken.substring(j, j + 1));
                }
            }
            tokens.add(modifiedWord.toString());
        }

        StringBuilder result = new StringBuilder();
        for (String t : tokens) {
            result.append(t).append(" ");
        }

        if (result.length() > 0) {
            result.setLength(result.length() - 1);
        }

        return result.toString();
    }

    private ArrayList<String> removeExtraSpaces(String currentText) {
        ArrayList<String> tt = new ArrayList<String>();
        StringBuilder word = new StringBuilder();
        currentText = currentText + " ";

        for (int i = 0; i < currentText.length(); i++) {
            // if the character is not a space, add it to a word
            if (!Character.isWhitespace(currentText.charAt(i))) {
                word.append(currentText.charAt(i));
            } else {
                if (word.length() != 0) {
                    tt.add(word.toString());
                    word.setLength(0);
                }
            }
        }

        return tt;
    }
}
```

Figure 6. Diacritics Remover

The purpose of this class is to remove diacritics or vowels in program output. The purpose of removing these diacritics is to simplify the search process for Isim Ma'rifat (ال) and reduce the possibility of duplication. For example, when deleting the vowel from lafadz الله (Allah), the program will avoid findings like الله (Allaha), الله (Allahi), and الله (Allahu) which have the same meaning, namely Allah. Thus, removing these diacritics helps reduce unnecessary duplication in search results for Isim Ma'rifat (ال).

```

package daud;
import java.util.ArrayList;

public class DiacriticsRemover {

    // remove all diacritics from text
    public String removeDiacritics(String currentText) {
        ArrayList<String> tokens = new ArrayList<String>();
        StringBuffer modifiedWord = new StringBuffer("");
        ArrayList<String> modifiedText = removeExtraSpaces(currentText);

        // for each token in the text
        for (int i = 0; i < modifiedText.size(); i++) {
            modifiedWord.setLength(0);
            String ctoken = modifiedText.get(i);
            for (int j = 0; j < ctoken.length(); j++) {
                if (!(Constants.diacritics.contains(ctoken.substring(j, j + 1)))) {
                    modifiedWord.append(ctoken.substring(j, j + 1));
                }
            }
            tokens.add(modifiedWord.toString());
        }

        StringBuilder result = new StringBuilder();

        for (String t : tokens) {
            result.append(t).append(" ");
        }

        if (result.length() > 0) {
            result.setLength(result.length() - 1);
        }

        return result.toString();
    }

    private ArrayList<String> removeExtraSpaces(String currentText) {
        ArrayList<String> tt = new ArrayList<String>();
        StringBuilder word = new StringBuilder();
        currentText = currentText + " ";

        for (int i = 0; i < currentText.length(); i++) {
            // if the character is not a space, add it to a word
            if (!Character.isWhitespace(currentText.charAt(i))) {
                word.append(currentText.charAt(i));
            } else {
                if (word.length() != 0) {
                    tt.add(word.toString());
                    word.setLength(0);
                }
            }
        }

        return tt;
    }
}

```

Figure 7. Constant

In this class, an encoding process is carried out for each verse found so that the verse no longer uses diacritics or punctuation. This process is carried out by removing excess spaces in the Arabic text and removing all vowels or diacritics that generally appear, such as FATHA, DAMMA, KASRA, SHADDA, SUKUN, and so on. To run the remove diacritics process, the code given above needs to be converted into a library named "Constant".

3. Search Result of Alif Lam (ال)

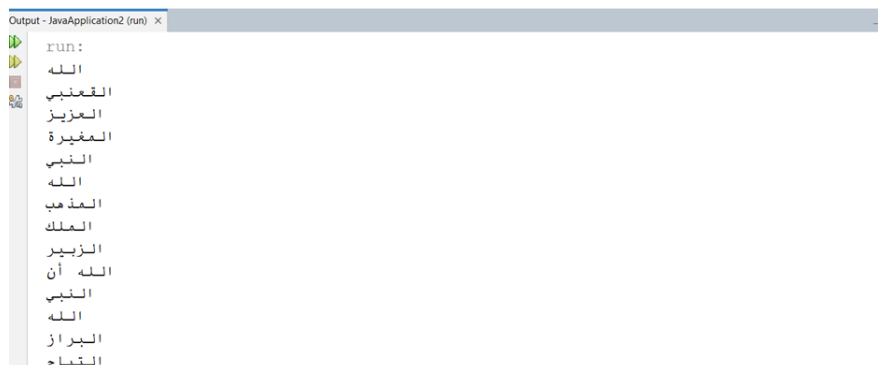


Figure 8. Search Result





| Isim Ma'rifat (ال) |                      |          |                 |
|--------------------|----------------------|----------|-----------------|
| No (Al-Quran)      | No (Hadith Abu Daud) | Al-Quran | Hadith Abu Daud |
| 4                  | 4                    | المصن    | الاثنين         |
| 5                  | 5                    | المن     | الاحتباء        |
| 6                  | 6                    | التي     | الاحتلام        |
| 7                  | 7                    | التي     | الاختصار        |
| 8                  | 8                    | الأثمين  | الاستبضاع       |
| 9                  | 9                    | الأخر    | الاستتابة       |
| 10                 | 10                   | الأخرين  | الاستحداد       |
| n - 1635           | n - 3685             | ...      | ...             |

The table above is the result of removing duplication in the Al-Quran and Abu Daud Hadith data. After deletion, found 1635 unique data on the Al-Quran out of a total of 9851 data that was originally found. Whereas in the Hadith of Abu Daud, there are 3685 unique data out of a total of 45184 initial data.

This finding indicates the existence of significant duplicate data in the Al-Quran and Hadith of Abu Daud. The duplicate removal process is mostly done by using the duplicate removal feature provided by Microsoft Excel. Nevertheless, these data still need to be further validated to ensure their accuracy and validity as Isim Ma'rifat (ال).

**Table 3.** Isim Ma'rifat (ال) validation results on the Al-Quran and Hadith of Abu Daud (Ascending)

| Isim Ma'rifat (ال) |                      |          |                 |
|--------------------|----------------------|----------|-----------------|
| No (Al-Quran)      | No (Hadith Abu Daud) | Al-Quran | Hadith Abu Daud |
| 1                  | 1                    | الز      | الابن           |
| 2                  | 2                    | الم      | الاثنان         |
| 3                  | 3                    | المر     | الاثنين         |
| 4                  | 4                    | المصن    | الاثنين         |
| 5                  | 5                    | المن     | الاحتباء        |
| 6                  | 6                    | التي     | الاحتلام        |
| 7                  | 7                    | التي     | الاختصار        |
| 8                  | 8                    | الأثمين  | الاستبضاع       |
| 9                  | 9                    | الأخر    | الاستتابة       |
| 10                 | 10                   | الأخرين  | الاستحداد       |
| n - 1635           | n - 3685             | ...      | ...             |

The table above is the result of Isim Ma'rifat (ال) validation using the sites quran.bblm.go.id and context.reverso.net. After the validation process, 987 Isim Ma'rifat (ال) were found in the Al-Quran and 1951 Isim Ma'rifat (ال) in the Hadith of Abu Daud. These data will be used to make comparisons and find similarities between the Isim Ma'rifat (ال) contained in the Al-Quran and the Hadith of Abu Daud.

**Table 4.** Comparison of the Isim Ma'rifat (ال) Al-Quran with the Hadith of Abu Daud

| Comparison of Isim Ma'rifat (ال) |                    | Comparison of Isim Ma'rifat (ال) |          |
|----------------------------------|--------------------|----------------------------------|----------|
| Al-Quran                         | Hadith of Abu Daud | Hadith of Abu Daud               | Al-Quran |
| الابتر                           | الابن              | الابن                            | الابتر   |
| الابرار                          | الابنة             | الابنة                           | الابرار  |
| الابصار                          | الاثنين            | الاثنين                          | الابصار  |
| الابل                            | الاجتناب           | الاجتناب                         | الابل    |
| الابواب                          | الاحتلام           | الاحتلام                         | الابواب  |

The table above is valid and unique Isim Ma'rifat (ال) data. The isim data in the Al-Quran (orange color) will be compared with all isim data in the Hadith of Abu Daud (yellow color). If no similarities are found, then the data is only found in the Al-Quran. However, if similarities are found, then the data will be found in both the Al-Quran and the Hadith of Abu Daud. This comparison process uses the "Match" function in Microsoft Excel for each data in the Al-Quran and Hadith of Abu Daud. The purpose of this comparison is to identify similarities and differences in data and to measure the extent to which Abu Daud's Hadith is able to explain the Isim Ma'rifat (ال) contained in the Al-Quran.

**Table 5.** Similarities and differences between the Isim Ma'rifat (ال) Al-Quran and the Hadith of Abu Daud (Ascending)

| Isim Ma'rifat (ال) |  |     |                            |      |                                  |
|--------------------|--|-----|----------------------------|------|----------------------------------|
| No                 | The Similarities Between the Al-Quran and the Hadith of Abu Daud | No  | Only Exist in the Al-Quran | No   | Only Exist in Hadith of Abu Daud |
| 1                  | الاسم  | 1   | الابتر                     | 1    | الابن                            |
| 2                  | الإبل  | 2   | الابرار                    | 2    | الاثنان                          |
| 3                  | الإثم  | 3   | الابصار                    | 3    | الاثنين                          |
| 4                  | الإحسان  | 4   | الابيض                     | 4    | الاحتلام                         |
| 5                  | الإسلام  | 5   | الاثيم                     | 5    | الاختصار                         |
| 6                  | الإصلاح  | 6   | الاجداث                    | 6    | الاستبضاع                        |
| 7                  | الإنجيل  | 7   | الاجل                      | 7    | الاستنابة                        |
| 8                  | الإنسان  | 8   | الاحمال                    | 8    | الاستخارة                        |
| 9                  | الإيمان  | 9   | الاخ                       | 9    | الاستسقاء                        |
| 10                 | الأبواب  | 10  | الاخت                      | 10   | الاستطالة                        |
| n -                | ...  | n - | ...                        | n -  | ...                              |
| 310                |  | 641 |                            | 1641 |                                  |

The table shows the same and different data between the Al-Quran and the Hadith of Abu Daud in terms of Isim Ma'rifat (ال). After doing a comparison, it was found that 310 Isim Ma'rifat (ال) were the same between the Al-Quran and the Hadith of Abu Daud. In addition, there are 641 Isim Ma'rifat (ال) which are only found in the Al-Quran and are not found in the Hadith of Abu Daud, and 1641 Isim Ma'rifat (ال) which are only found in the Hadith of Abu Daud and are not found in the Al-Quran. These findings provide an indication of the ability of Abu Daud's Hadith in explaining the Isim Ma'rifat (ال) contained in the Al-Quran.

### 3.4 Analysis of Test Results

**Table 6.** Analysis of validation results

| Validation Result of Al-Quran |      | Validation Result of Hadith of Abu Daud |        |
|-------------------------------|------|---|--------|
| Data                          | 1183 | Data                                    | 3659   |
| Isim                          | 987  | ISIM                                    | 1951   |
| Not Isim                      | 196  | Not Isim                                | 1708   |
| Accuracy                      | 83%  | Accuracy                                | 53,32% |

The table is the result of an analysis of the validation of Isim Ma'rifat (ال) on the Al-Quran and Hadith of Abu Daud. Of the 1183 data on Isim Ma'rifat (ال) in the Al-Quran, 987 valid data were found, with an accuracy rate of 83%. This means that as much as 83% of the data that is considered as Isim Ma'rifat (ال) in the Al-Quran proves to be correct according to the validation carried out. Whereas in the Hadith of Abu Daud, from 3,306 Isim Ma'rifat (ال) data, 1,764 valid data were found, with an accuracy rate of 53%. This means that as much as 53% of the data that is considered as Isim Ma'rifat (ال) in Abu Daud's Hadith is proven correct based on the validation carried out.

**Table 7.** Analysis of data processing results

| Isim Ma'rifat (ال)                             | Total |
|--|-------|
| Isim that found in the Al-Quran                | 987   |
| Isim that found in the Hadith of Abu Daud      | 1951  |
| Isim that found in the both of them            | 641   |
| Isim that only exist in the Al-Quran           | 1641  |
| Isim that only exist in the Hadith of Abu Daud | 310   |

The table contains the data values generated in this study after going through the stages of eliminating duplication, validation, and comparison. This data will be used to calculate the percentage of similarities and differences between the Al-Quran and the Hadith of Abu Daud.

**Table 8.** Percentage of Al-Quran data

| Percentage of Similarities and Differences in data on the Al-Quran | Total |
|--|-------|
| The same amount of data on Al-Quran                                | 310   |
| The amount of data that is not the same on Al-Quran                | 677   |
| The same percentage of data in the Al-Quran                        | 38%   |

| Percentage of Similarities and Differences in data on the Al-Quran | Total |
|--|-------|
| The percentage of data that is not the same in the Al-Quran        | 62%   |

In the Al-Quran, there are 677 data on Isim Ma'rifat (ال) which are different from the data in the Hadith of Abu Daud. The percentage of data that is the same between the Al-Quran and the Hadith of Abu Daud is 38%, while the percentage of different data reaches 62%. This means that out of the 677 different words of Isim Ma'rifat (ال), there is no explanation or finding in Abu Daud's Hadith.

**Table 9.** Abu Daud Hadith data percentage

| Persentase Persamaan dan Perbedaan data pada Hadis Abu Daud           | Total |
|---|-------|
| The same amount of data on Hadith of Abu Daud                         | 310   |
| The amount of data that is not the same on Hadith of Abu Daud         | 1641  |
| The same percentage of data in the Hadith of Abu Daud                 | 16%   |
| The percentage of data that is not the same in the Hadith of Abu Daud | 84%   |

In the Hadith of Abu Daud, found 1641 data on Isim Ma'rifat (ال) which are different from the data in the Al-Quran. The percentage of data that is the same between the Hadith of Abu Daud and the Al-Quran is 16%, while the percentage of different data reaches 84%. This shows that Abu Daud's Hadith can only explain 16% of the total nouns or Isim Ma'rifat (ال) contained in the Al-Quran. In addition, there are also 1641 Isim Ma'rifat (ال) in Abu Daud's Hadith which are not found in the Al-Quran.

#### 4. CONCLUSION

In this study, researchers thoroughly investigated Isim Ma'rifat, the Arabic construct with the prefix "AL" (ال), by utilizing the Remove Diacritics method in both the Al-Quran and the Hadith of Abu Daud. By removing diacritical marks like vowels and tasykil from the Arabic text, they searched for valid Isim Ma'rifat instances. The validation results revealed that the Al-Quran exhibited a higher accuracy (83%) compared to the Hadith of Abu Daud (53.32%). Moreover, the Al-Quran contained more valid Isim Ma'rifat occurrences (987 data) than the Hadith of Abu Daud (1951 data). However, both sources also contained a significant amount of non-Isim Ma'rifat data, with 196 instances in the Al-Quran and 1708 instances in the Hadith of Abu Daud. The primary objective of comparing the findings of Isim Ma'rifat between the Al-Quran and the Hadith of Abu Daud was to assess the latter's ability to explain Isim Ma'rifat (ال) present in the former. By using validated data, the researchers sought to identify similarities and differences between the two sources. The comparison unveiled 376 shared instances of Isim Ma'rifat between the Al-Quran and the Hadith of Abu Daud. Additionally, there were 611 instances unique to the Al-Quran and 1388 instances exclusive to the Hadith of Abu Daud. This analysis shed light on the extent to which Abu Daud's Hadith can elucidate the Isim Ma'rifat found in the Al-Quran. The implications of the search and comparison of Isim Ma'rifat on the Al-Quran and Hadith of Abu Daud are significant for developing the Quranpedia website. The validated information on Isim Ma'rifat and the comparative analysis between the two sources can enrich the content presented on Quranpedia. By incorporating these findings, the website can offer accurate and comprehensive insights into the nouns mentioned in the Quran, helping users better understand their meanings and usage within a religious context. Furthermore, the comparison between the Al-Quran and the Hadith of Abu Daud can provide valuable insights into how the Hadith of Abu Daud can serve as an explanatory source for the Isim Ma'rifat present in the Al-Quran. This knowledge can be presented through articles or comparisons, enhancing users' understanding of the relationship between the Al-Quran and hadith concerning the usage of nouns.

#### REFERENCES

- [1] S. Haque, Z. Eberhart, A. Bansal, and C. McMillan, "Semantic Similarity Metrics for Evaluating Source Code Summarization," in IEEE International Conference on Program Comprehension, IEEE Computer Society, 2022, pp. 36–47. doi: 10.1145/nnnnnnn.nnnnnnn.
- [2] Nurul Maliya and Mudafiq Riyan Pratama, "SISTEM DETEKSI JENIS KATA DALAM BAHASA ARAB DENGAN MENGGUNAKAN ALGORITMA LIGHT STEMMING," Jember, 2018.
- [3] F. Faizin, "Kisah Al-Qur'an dalam Tinjauan Sains (Studi atas Serial Tafsir Ilmi Kementerian Agama RI)," AL QUDS: Jurnal Studi Alquran dan Hadis, vol. 4, no. 1, p. 77, May 2020, doi: 10.29240/alquds.v4i1.1106.
- [4] M. Satria Perwira Negara et al., "APLIKASI TEMU KEMBALI CITRA AYAT AL-QURAN DENGAN METODE GRAY LEVEL CO-OCCURRENCE MATRIX DAN HISTOGRAM OF ORIENTED GRADIENT STUDI KASUS AL-QURAN JUZ 30," 2018. [Online]. Available: <http://ejournal.unib.ac.id/index.php/rekursif/>

- [5] N. A. P. Rostam and N. H. A. H. Malim, "Text categorisation in Quran and Hadith: Overcoming the interrelation challenges using machine learning and term weighting," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 6, pp. 658–667, Jul. 2021, doi: 10.1016/j.jksuci.2019.03.007.
- [6] H. Yusupov, "THE DEVELOPMENT OF THE SCIENCE OF HADITH IN MOVAROUNNAHR IN THE IX CENTURY," 2021, doi: 10.37547/pedagogics-crjp-02-09-35.
- [7] Leni Andariati, "HADIS DAN SEJARAH PERKEMBANGANNYA," *Jurnal Ilmu Hadis*, vol. 4, no. 2, pp. 153–166, Mar. 2020.
- [8] A. C. Nadhiroh, "ISTILAH-ISTILAH HADITS MENURUT ABU DAWUD (Studi atas kitab sunan abu dawud)," *Al-Tadabbur: Jurnal Ilmu Al-Quran dan Tafsir*, pp. 259–274, 2023, doi: 10.30868/at.v7i02.2713.
- [9] I. Sa'dudin, J. Z. Hafizd, and E. Safitri, "The Arabic's Significant Role in the Understanding of Islamic Law," *Jurnal Pendidikan Bahasa Asing dan Sastra*, vol. 6, no. 2, 2022, [Online]. Available: <https://ojs.unm.ac.id/eralingua>
- [10] M. H. Bashir et al., "Arabic natural language processing for Qur'anic research: a systematic review," *Artif Intell Rev*, vol. 56, no. 7, pp. 6801–6854, Jul. 2023, doi: 10.1007/s10462-022-10313-2.
- [11] Muhammad Saeed and Amjad Hayat, "Difference of Opinions among Scholars in the Explanation and Interpretation of Al- Huruful Muqataat: A Comparative Study," *UOCHJRS*, vol. 3, no. 2, pp. 165–180, 2020, doi: 10.33195/uochjrs-v3i21952020.
- [12] Ehta Setiawan, "<https://www.kbbi.web.id/>," versi 2.9, 2023.
- [13] Ibnu Rawandhy N. Hula and Berti Arsyad, "DIAKRITIK AL-QUR'AN MENURUT PREFERENSI ABU DAWUD", doi: 10.31314/ajamiy.9.2.265-284.2020.
- [14] A. Mijlad and Y. El Younossi, "A Comparative Study of Some Automatic Arabic Text Diacritization Systems," *Advances in Human-Computer Interaction*, vol. 2022. Hindawi Limited, 2022. doi: 10.1155/2022/3613710.
- [15] B. Vijayalakshmi and N. Sasirekha, "LOSSLESS TEXT COMPRESSION FOR UNICODE TAMIL DOCUMENTS," *ICTACT JOURNAL ON SOFT COMPUTING*, p. 2, 2018, doi: 10.21917/ijsc.2017.0227.
- [16] E. A. Jrai, S. Alsharari, L. Almazaydeh, K. Elleithy, and O. Abu-Hamdan, "Improving LZW Compression of Unicode Arabic Text Using Multi-Level Encoding and a Variable-Length Phrase Code," *IEEE Access*, vol. 11, pp. 51915–51929, 2023, doi: 10.1109/ACCESS.2023.3279021.
- [17] T. A. Hilal and H. A. Hilal, "Arabic text lossless compression by characters encoding," in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 618–623. doi: 10.1016/j.procs.2019.08.087.
- [18] A. Hassani, A. Iranmanesh, and N. Mansouri, "Text Mining using Nonnegative Matrix Factorization and Latent Semantic Analysis," Nov. 2019, [Online]. Available: <http://arxiv.org/abs/1911.04705>
- [19] M. A. Permana, E. Darwiyanto, and M. Arif Bijaksana, "Pembobotan dan Pemeringkatan Ayat Al-Quran Berdasarkan Compound, Term Frequency dan Prinsip Pareto Untuk Membantu Hafalan," 2021.
- [20] B. Zemni and Chaouki Bounaas, "Audiovisual Translation and contextual dictionaries: An exploratory comparative study of Reverso Context and Almaany uses," *Asian EFL Journal Research*, vol. 27, no. 5, pp. 274–309, 2020, [Online]. Available: <https://www.researchgate.net/publication/348150615>