Journal of Theoretical and Applied Information Technology
15th September 2024. Vol.102. No. 17
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# USING LONGEST COMMON SUBSTRING ALGORITHM TO FIND SIMILAR QURANIC VERSES

**MAJED ABUSAFIYA**
Associate Professor, Al-Ahliyya Amman University, Software Engineering Department, Jordan
E-mail: mabusafeyeh@ammanu.edu.jo, majedabusafiya@gmail.com

**ABSTRACT**

In this paper, the longest common substring algorithm is used to find similar Quranic verses (ayahs). To show the potential of using computation in this context, pairs of similar ayahs that were found computationally are compared with those that are explicitly stated in a specialized scholarly book in this field. It was found that only about one third of these pairs exactly matched those that are computationally found. However, most of the other pairs of similar ayahs that were mentioned in the book showed less similarity than those that were found computationally. The main contribution of this work is showing the value of the proposed computational solution through finding similar Quranic ayahs beyond those that are documented in the specialized books.

**Keywords:** *String Similarity Algorithm, Quran, Quranic Verse Similarity*

## 1. INTRODUCTION

Quran is the book of God and the holy book of Muslims. It is composed of 114 chapters (*surahs*), each composed of a varying number of verses (*ayahs*). One observation about Quranic ayahs is the obvious similarities among them. Because of the perfect eloquence of Quran, this similarity has attracted the attention of many Quranic interpretation scholars. They tried to explore why a given *ayah* comes in one form in one place and comes in another very similar form in another place. Moreover, extracting similar ayahs, comparing them and reasoning about their differences will be useful to better-understand their meanings. It also uncovers more and more glimpses of miraculous eloquence of Quran. That is why numerous books were authored to identify and to study these similarities under a specialized Quranic science that is called "*Mutashabeh An-nathem*" (متشابه النظم) or *syntactically similar ayahs*.

The main questions that will be answered in this research are: (1) is there a computational approach to discover similarity between Quranic ayahs? (2) how to prove that this approach is useful in discovering similarity between Quranic ayahs. To answer the first question, an algorithm that is based on longest common substring algorithm is proposed and implemented to extract similar Quranic ayahs. To answer the second question, the proposed algorithm is implemented algorithm and used to extract similar ayahs. Then, a comparison between the similarity results in a selected specialized book and the results that were found computationally are compared.

The main motivation of this work is that the automation of extracting similar ayahs may uncover ayahs' similarities that could have been overlooked by the scholars in this field. As a side product of this research, a new algorithm that measures the similarity between two strings is proposed. In addition, this work offers a new application of string algorithms in the context of natural language processing in general and in serving this holy book in specific.

The algorithm (named MAIN) that is proposed to computationally find similar ayahs is based on a model for defining the similarity between a pair of ayahs. The similarity between a pair of ayahs is modeled using *similarityRecord* structure. A similarity record encapsulates a pair of similar ayahs: the ayah for which similar ayahs are found and is referred to as *firstAyah,* and the ayah that shows obvious similarity with *firstAyah* which is referred to as *secondAyah.* To show the validity of the proposed computational approach, this algorithm was *not* used to extract similar ayahs of Quran. Instead, the MAIN algorithm takes the list of *firstAyahs* for which similar ayahs are sited in one of the well-known books in this field. This is

done by creating a similarity record for every pair of similar ayahs that are mentioned in the selected book. These similarity records are then taken one by one where two sub-algorithms are applied on each of them. The first sub-algorithm, called MEASURE-SIMILARITY, computes the similarity between *firstAyah* (of the current similarity record) and every ayah in Quran. To maintain this information, it creates a similarity record for the *firstAyah* (of current similarity record) and each Quranic ayah along with a similarity measure in a sorted vector. This sorted vector will be the basis of comparing the ayah similarity records that are found computationally with those similarity records that are stated in the selected book. The second sub-algorithm, PROJECT-TO-LEVELS, projects the similarity records of the book into one of eleven similarity levels. This projection is based on the rank of the similar second ayah as stated in the selected book in the sorted list of similar ayahs that is found computationally by MEASURE-SIMILARITY.

A literature review for related work was conducted by the author. In the context of string similarity measuring algorithms, a number of them were proposed in literature. The following algorithms may be found: *Dice* [1], *Levenshtein*[2], *Smith Waterman* [3], *Jaro* [4], *Jaro Winkler* [5], *soundex* [6] and *longest common substring* (*LCS)* [7]. Among these algorithms, the most appealing algorithm in finding Quranic ayahs similarity was the LCS algorithm. This algorithm will be used in MAIN algorithm for measuring the similarity between two Quranic ayahs as will be shown below.

In the context of finding similar Quranic ayahs computationally, a number of research papers were found. The work in [8] proposed an MQVC approach for retrieving most similar ayahs in comparison with a user input ayah as query. For comparison, a reference to specialists who identified all relevant surahs and ayahs to the targeted ayah was used. However, it was not clear from this paper the details of how the similarity was measured. In [9], an NLP method was found to detect the semantic-based similarity between the verses of Quran. They exploit the distributed representation of text to learn an informative representation of the Quranic passage. Then a mapping is done between Quranic ayahs to numerical vectors that encode the semantic properties of the text, then similarity is measured between these vectors. In [10], semantically similar or related ayahs are linked together. In [11], they used clustering techniques to classify ayahs from Chapter 2 in Quran (Surat Al-Baqarah). They used English translations and not Arabic text. In [12], they used *Word2Vec* and *Sent2Vec* models to find similarity. Work in [9],[10],[11] and [12] differ from our work in the sense that it focuses on semantic similarity, while our work is focused in measuring syntactic similarity. They also used English translations of Quran instead of Arabic text.

This paper is structured as follows: section 2 presents the MAIN algorithm; in section 3, a discussion to prove the validity of the MAIN algorithm is presented; the paper ends up with a conclusion and a number of references.

## 2. MAIN ALGORITHM

To show the value of this work in finding similar Quranic ayahs, the similar ayahs that are mentioned in one of the specialized books in this context is regenerated by computation. This will prove that by computation, more similar pairs of ayahs may be found. One of the most known books in this context – titled *Malak At-taweel (*ملاك التأويل) authored by *Ghirnati* (الغرناطي) in the thirteen's century– was chosen for this aim. This book presents pairs of Quranic ayahs that show obvious similarity. It also presents a justification to show that this similarity is not a useless repetition by showing that each ayah perfectly serves its context in Quran. To model similarity between a pair of ayahs in this paper, the *SimilarityRecord* structure is defined (Fig-1). This structure encapsulates a pair of ayahs and a measure of their similarity. The *computedSimilarityRecords* field is a sorted vector of similarity records for the *firstAyah* with every ayah in Quran. This field will be the mean to compare between similarity pairs that are documented in *Ghirnati's* book with those that are found by computation.

```
class SimilarityRecord {
    Ayah firstAyah;
    Ayah secondAyah;
    double measure;
    Vector computedSimilarityRecords;
}
```

*Figure 1: SimilarityRecord structure*

The MAIN algorithm is shown in Fig-2. The goal of this algorithm is to define a similarity record structure for every pair of similar ayahs that is mentioned in the *Ghirnati's* book. By comparing the similarity pair of *Ghirnati's* book with the similarity found by computation (using

the *computedSimilarityRecords* field), it may be proven that computation uncovers more similarity pairs than those that are mentioned in *Ghirnati's* book.

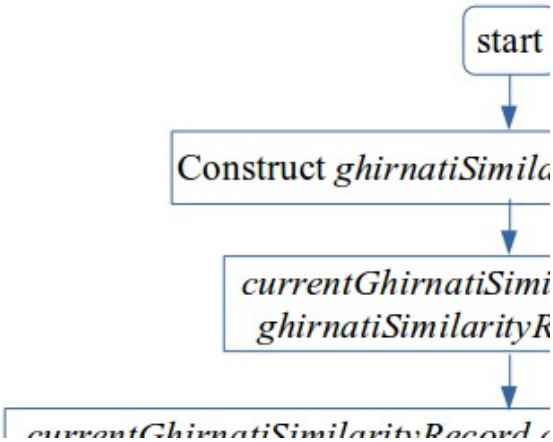


*Figure 2: MAIN algorithm*

### 2.1 Construction of Ghirnati Similarity Records Vector

The first step of MAIN algorithm is to create a similarity record for every pair of similar ayahs that is mentioned in *Ghirnati*'s book. These records are maintained in *ghirnatiSimilarityRecords* vector. The construction of this vector is essential because the rest of the MAIN algorithm processes the similarity records of this vector. Table-1 shows an example of a similarity record for a similarity pair that is mentioned in *Ghirnati's* book. Only the first two fields are set with values at this point. The other two fields will be computed latter by the MEASURE-SIMILARITY sub-algorithm. *Ghirnati's* book contains 456 pairs of highly similar ayahs.

*Table 1:* A similarity pair record for a selected ayah from *Ghirnati's* Book.

| Field | Value |
|---|---|
| *firstAyah* | (واتقوا يوما لا تجزي نفس عن نفس شيئا ولا يقبل منها شفاعة ولا يؤخذ منها عدل ولا هم ينصرون) |
| *secondAyah* | (واتقوا يوما لا تجزي نفس عن نفس شيئا ولا يقبل منها عدل ولا تنفعها شفاعة ولا هم ينصرون) |
| *measure* | To be calculated by MEASURE-SIMILARITY |
| *computedSimilarityRecords* | To be calculated by MEASURE-SIMILARITY |

The *ghirnatiSimilarityRecords* vector is constructed manually. In other words, a similarity record is explicitly constructed and added to the *ghirnatiSimilarityRecords* vector for every pair of similar ayahs that is mentioned in *Ghirnati*'s book. The addition of the similarity record to *ghirnatiSimilarityRecords* vector is shown in Fig-3. This record is added because it was mentioned in *Ghirnati's* book that the ayah 48 of the second surah showed high similarity with ayah 123 of the same surah. By the end of this step, all pairs of similar ayahs that are mentioned in the *Ghirnati's* book are recorded in this vector.

```
    :
ghirnatiSimilarityRecords.add(new SimilarityRecord(
        Quran.getAyah(2, 48), Quran.getAyah(2, 123));
    :
```

*Figure 3: ghirnatiSimilarityRecords* vector construction

### 2.2 Building *computedSimlarityRecords* vector field

Once constructed, the similarity records in the *ghirnatiSimilarityRecords* vector are taken one by one through a loop (Fig-2). For each Ghirnati similarity record from this vector - name it *currentGhirnatiSimilarityRecord* - the MEASURE-SIMILARITY algorithm is applied (Fig-4). This algorithm computationally measure the similarity of the ayah stated by *firstAyah* field of the *currentGhirnatiSimilarityRecord* with every ayah in Quran. It returns a sorted vector that is assigned to the *computedSimilarityRecords* field of the *currentGhirnatiSimilarityRecord.*

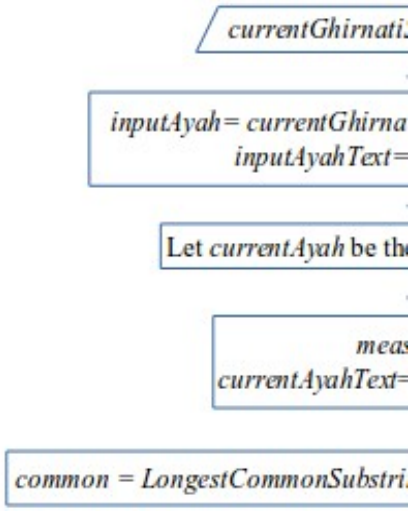


*Figure 4: MEASURE-SIMILARITY algorithm*

MEASURE-SIMILARITY algorithm is composed of two nested loops. The outer loop takes the ayahs of Quran one by one. Name the

ayah that is being compared against the *inputAyah* in the current iteration as *currentAyah*. The variable *measure* is an integer that represents a quantitative measure of similarity between *inputAyah* and *currentAyah*. It is initialized to zero in the beginning of each iteration of the outer loop. In other words, it is initialized to zero for every *currentAyah.* The variables *inputAyahText* and *currentAyahText* are initialized to be the texts of the *inputAyah* and the *currentAyah* respectively. These variables are used as a side copy of the text of the tested pair of ayahs so that the original text of the ayahs' is not changed. This is needed because the similarity measure calculation requires repeated deletion of *common* substrings between *inputAyah* and *currentAyah* texts.

The inner loop repeatedly finds the longest common substring – called *common* - between *inputAyahText* and *currentAyahText*. In every iteration, the *measure* variable is updated by adding the value of 2 raised to a power that is equal to the length of *common*. Then the texts of *inputAyahText* and *currentAyahText* are updated by deleting *common* from both. The inner loop terminates when the length of *common* is less than two letters. By the end of the inner loop, the computation of the similarity *measure* between *inputAyah* and *currentAyah* is done. So a new *SimilarityRecord* is created and is then inserted into the *SimilarityRecordsVector* which is a sorted vector. The *similarityRecordsVector* will grow in size until its length becomes equal to the number of ayahs of Quran (at the end of the execution of MEASURE-SIMILARITY algorithm). In other words, the outer loop terminates when the similarity measure between *firstAyah* and every quranic ayah is calculated. The algorithm returns the sorted *similarityRecordsVector*. Referring to MAIN algorithm (Fig-2), this sorted *similarityRecordsVector* will be assigned to the *computedSimilarityRecords* field of *currentGhirnatiSimilarityRecord.* At this point, the similarity data between the *firstAyah* in *currentGhirnatiSimilarityRecord* and every ayah in Quran is computed and recorded.

Table-2 shows a *part* of the *similarityRecordsVector* that is constructed by the MEASURE-SIMILARITY algorithm for the *Ghirnati* similarity record shown in Table-1. It is described to be '*part*' because it is actually composed of a very long list of records that is equal to the number of Quranic ayahs. The first three fields are set with values, while the *computedSimilarityRecords* field is left with no value because it is not needed. The *measure* column is decreasing because the *similarityRecordsVector* is sorted according to this column. The first similarity record is the record that has *secondAyah* to be the same as *firstAyah*. This is expected because the most similar ayah to the *firstAyah* is the *firstAyah* itself. This record is excluded from the *similarityRecordsVector*. The *secondAyah* field in the first similarity record is supposed to be the *computationally-found* most similar ayah to the *firstAyah* (excluding the *firstAyah* itself). Based on the approach that is used to measure similarity, the most similar ayah is the ayah that has the longest common substring with the ayah under consideration and hence the highest similarity measure value. In this specific example (Table-1), it happened that the computationally-found most similar ayah is the same similar ayah that was stated in the *Ghirnati's* book for *firstAyah*.

*Table 2: SimilarityRecordsVector* that is computed by MEASURE- SIMILARITY algorithm for the *Ghirnati similarity record* shown in Table 1.

| | *firstAyah* | *secondAyah* | *Measure* |
|---|---|---|---|
| 1 | واتقوا يوما لا تجزي نفس عن نفس شيئا ولا يقبل منها شفاعة ولا يؤخذ منها عدل ولا هم ينصرون | واتقوا يوما لا تجزي نفس عن نفس شيئا ولا يقبل منها عدل ولا تنفعها شفاعة ولا هم ينصرون | 1.1259E+15 |
| 2 | | يوم لا يغني مولى عن مولى شيئا ولا هم ينصرون | 16470.0 |
| 3 | | لو يعلم الذين كفروا حين لا يكفون عن وجوههم النار ولا عن ظهورهم ولا هم ينصرون | 16462.0 |
| 4 | | فاليوم لا تظلم نفس شيئا ولا تجزون إلا ما كنتم تعملون | 16454.0 |
| : | | : | : |

## 2.3 Projecting *Ghirnati* records into Levels

By reaching this point in MAIN, a *currentGhirnatiSimilarityReccord* will have the *computedSimilarityRecords* vector field calculated. As mentioned earlier, the *similarityRecordsVector* that was calculated by MEASURE-SIMILARITY is assigned to this field (i.e. *computedSimilarityRecords,* refer to Fig-2). Running the PROJECT-TO-LEVELS algorithm (Fig-5) is the last processing step for the *currentGhirnatiSimilarityRecord* in MAIN algorithm loop. The purpose of this algorithm is to add the *currentGhirnatiSimilarityRecord* to one of eleven similarity levels that together compose

the *SimilarityLevelsArray*. For each of the eleven similarity levels, there is a corresponding vector of Ghirnati similarity records. The *currentGhirnatiSimilarityRecord* will be added to the *SimilarityLevelsArray[l]* vector if the *secondAyah* of *currentGhirnatiSimilarityReccord* is the same as the *secondAyah* of the *l*-th similarity record in the *computedSimilarityRecords* vector of this *currentGhirnatiSimilarityReccord.* In other words, a *similarity level* is the rank of the ayah that is stated to be similar – for a given *firstAyah* - in *Ghirnati's* book within the list of computationally-found similar ayahs for this given *firstAyah*.

In the following, how PROJECT-TO-LEVELS algorithm works is illustrated. Its input is *currentGhirnatiSimilarityRecord* (Fig-2). Initially, the variable *level* is set to 1. This algorithm is composed of a loop where the first ten records from *computedSimilarityRecords* vector of *currentGhirnatiSimilarityRecord* are taken one after another in each iteration. Name this taken record as *currentComputedSimilarityRecord.* Next, the *secondAyah* field of *currentGhirnatiSimilarityRecord* is compared with *secondAyah* field in *currentComputedSimilarityRecord*. If they are the same, then the level for *currentGhirnatiSimilarityRecord* is the current value of *level*. In this case, *currentGhirnatiSimilarityRecord* is added to *similarityLevelsArray[level]* vector and the algorithm terminates. Otherwise, the *level* is incremented and a new iteration is needed. The continuation condition of the loop is checked. If the *level* is less than 11, then a new iteration starts and the next *computedSimilarityRecord* is taken. Otherwise (i.e. when the value of *level* is 11), *currentComputedSimilarityRecord* should be – anyway - added to *similarityLevelsArray[11]* and the loop terminates.

By studying this loop, it can be seen that it iterates at least one iteration and at most eleven iterations. It iterates only *once* in case the *secondAyah* of *currentGhirnatiSimilarityRecord* is the same as *secondAyah* for the first similarity record in *computedSimilarityRecords* of this *currentGhirnatiSimilarityRecord.* On the other hand, it iterates eleven iterations in case the *secondAyah* of *currentGhirnatiSimilarityRecord* is *not* the same as the *secondAyah* of any of the first ten records of *computedSimilarityRecords* of *currentGhirnatiSimilarityRecord.* In between these two extremes, the number of iterations will be equal to the similarity level of the *currentGhirnatiSimilarityRecord.* This similarity level is the rank of the similarity record within *computedSimilarityRecords* whose *secondAyah* matches the *secondAyah* for *currentGhirnatiSimilarityRecord.*

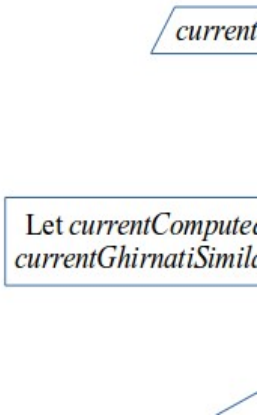


*Figure 5: PROJECT-TO-LEVELS algorithm*

To show how this algorithm works with an example, consider the Ghirnati Similarity record that is shown in Table-1. If PROJECT-TO-LEVELS algorithm is applied on this similarity record, then it will be added to *SimilarityLevelsArray[1]*. This is because the *secondAyah* field in this similarity record is the same as the *secondAyah* field of the *first* similarity record in *computedSimilarityRecords* of this Ghirnati Similarity record (refer to Table-2). Assume for example, that the *secondAyah* of the Ghirnati similarity record matched the *secondAyah* field of the *fourth* record of *computedSimilarityRecords*, then this Ghirnati Similarity record should be added to the *similarityLevelsArray[4]* vector. Moreover, assume for example, that the *secondAyah* field of the Ghirnati similarity record did not match the *secondAyah* field of any of the first ten records of *computedSimilarityRecords* vector of this Ghirnati similarity record, then it should be added to the *similarityLevelsArray[11]*.

## 3. DISCUSSION

The MAIN algorithm was implemented and applied to the 456 similarity records of the Ghirnati's book. Table-3 shows these similarity records projected into similarity levels. Note that only 38% of the Ghirnati's similarity records

matched the computationally-found most similar ayah. By studying the patterns of similarity, a number of observations can be easily seen:

First, many of the similarity pairs that were found in the Ghirnati's book are not based on the length of matching substrings. Many of these similarities consider semantic similarity in addition to the syntactic similarity. Second, many of Ghirnati's similarity records are based on certain portions of the ayahs while the similarity record that is found by computation may detect similarity in other portions of the ayahs. In other words, it tries to find similarity based on the largest substrings that are common between the two ayahs, regardless of the location of these common substring within these two ayahs. Third, it can be noticed that the same ayah has many similar ayahs because similarity may exists between different portions of the same ayah with the other similar ayahs. In other words, you will find two ayahs that are similar to a given ayah where similarity occurs in different portions of this ayah.

*Table 3: SimilarityLevelsArray* vectors sizes

| *level* | *SimilarityLevelsArray[level]* vector size | percentage |
|---|---|---|
| 1 | 175 | 38% |
| 2 | 60 | 13% |
| 3 | 31 | 7% |
| 4 | 22 | 5% |
| 5 | 18 | 4% |
| 6 | 10 | 2% |
| 7 | 10 | 2% |
| 8 | 13 | 3% |
| 9 | 8 | 2% |
| 10 | 6 | 1% |
| 11 | 103 | 23% |
| **sum** | **456** | **99%** |

In the following, samples of some similarity ayahs from various levels, is shown. The shaded ayah, is the ayah that was cited in *Ghirnati's* book as the similar ayah for the ayah in *firstAyah* column.

**Level-1 Sample**

In this level, the computationally-found most similar ayah matches the one that is stated in Ghirnati book.

| | *firstAyah* | *secondAyah* | Measure |
|---|---|---|---|
| 1 | واتقوا يوما لا تجزي نفس عن نفس شيئا ولا يقبل منها عدل ولا تنفعها شفاعة ولا هم ينصرون | | 1.1259E+15 |
| 2 | | يوم لا يغني مولى عن مولى شيئا ولا هم ينصرون | 16470.0 |
| 3 | | لو يعلم الذين كفروا حين لا يكفون عن وجوههم النار ولا عن ظهورهم ولا هم ينصرون | 16462.0 |
| 4 | | فاليوم لا تظلم نفس شيئا ولا تجزون إلا ما كنتم تعملون | 16454.0 |
| 5 | | يوم لا يغني عنهم كيدهم شيئا ولا هم ينصرون | 16450.0 |
| 6 | | أولئك الذين اشتروا الحياة الدنيا بالآخرة فلا يخفف عنهم العذاب ولا هم ينصرون | 16430.0 |
| 7 | | وذر الذين اتخذوا دينهم لعبا ولهوا وغرتهم الحياة الدنيا وذكر به أن تبسل نفس بما كسبت ليس لها من دون الله ولي ولا شفيع وإن تعدل كل عدل لا يؤخذ منها أولئك الذين أبسلوا بما كسبوا لهم شراب من حميم وعذاب أليم بما كانوا يكفرون | 8386.0 |
| 8 | | واتقوا يوما ترجعون فيه إلى الله ثم توفى كل نفس ما كسبت وهم لا يظلمون | 4170.0 |
| 9 | | يا أيها الناس اتقوا ربكم واخشوا يوما لا يجزي والد عن ولده ولا مولود هو جاز عن والده شيئا إن وعد الله حق فلا تغرنكم الحياة الدنيا ولا يغرنكم بالله الغرور | 2318.0 |
| 10 | | يا أيها النبي إذا جاءك المؤمنات يبايعنك على أن لا يشركن بالله شيئا ولا يسرقن ولا يزنين ولا يقتلن أولادهن ولا يأتين ببهتان يفترينه بين أيديهن وأرجلهن ولا يعصينك في معروف فبايعهن واستغفر لهن الله إن الله غفور رحيم | 2234.0 |

**Level-2 Sample**

This sample shows that a computationally found similarity is not necessarily based on substrings that are composed of complete words. For example, the computationally found most similar ayah found the common sub-string (وما هو بقول شـ) with a part of a word (شـ). Moreover, it may be noticed that some of the similar ayahs have the same measure value.

| | *firstAyah* | *secondAyah* | Measure |
|---|---|---|---|
| 1 | وما هو بقول شاعر قليلا ما تؤمنون | وما هو بقول شيطان رجيم | 8194.0 |
| 2 | | ولا بقول كاهن قليلا ما تذكرون | 2118.0 |
| 3 | | وما يستوي الأعمى والبصير والذين آمنوا وعملوا الصالحات ولا المسيء قليلا ما تتذكرن | 2078.0 |

| | | | |
|---|---|---|---|
| 4 | | قل هو الذي أنشأكم وجعل لكم السمع والأبصار والأفئدة قليلا ما تشكرون | 2074.0 |
| 5 | | اتبعوا ما أنزل إليكم من ربكم ولا تتبعوا من دونه أولياء قليلا ما تذكرون | 2070.0 |
| 6 | | ثم سواه ونفخ فيه من روحه وجعل لكم السمع والأبصار والأفئدة قليلا ما تشكرون | 2062.0 |
| 7 | | أمن يجيب المضطر إذا دعاه ويكشف السوء ويجعلكم خلفاء الأرض أإله مع الله قليلا ما تذكرون | 2062.0 |
| 8 | | وهو الذي أنشأ لكم السمع والأبصار والأفئدة قليلا ما تشكرون | 2062.0 |
| 9 | | ولقد مكناكم في الأرض وجعلنا لكم فيها معايش قليلا ما تشكرون | 2062.0 |
| 10 | | وقالوا قلوبنا غلف بل لعنهم الله بكفرهم فقليلا ما يؤمنون | 562.0 |

**Level-3 Sample**

In this sample, it can be seen that Ghirnati sometimes considered semantic similarity. Syntactically, the *firstAyah* is more similar to the first two ayahs in the *secondAyah* column than the *secondAyah* in the third line.

| | *firstAyah* | *secondAyah* | Measure |
|---|---|---|---|
| 1 | وهو الذي يرسل الرياح بشرا بين يدي رحمته حتى إذا أقلت سحابا ثقالا سقناه لبلد ميت فأنزلنا به الماء فأخرجنا به من كل الثمرات كذلك نخرج الموتى لعلكم تذكرون | أمن يهديكم في ظلمات البر والبحر ومن يرسل الرياح بشرا بين يدي رحمته أإله مع الله تعالى الله عما يشركون | 4.29497E9 |
| 2 | | وهو الذي أرسل الرياح بشرا بين يدي رحمته وأنزلنا من السماء ماء طهورا | 1.07374E9 |
| 3 | | الله الذي يرسل الرياح فتثير سحابا فيبسطه في السماء كيف يشاء ويجعله كسفا فترى الودق يخرج من خلاله فإذا أصاب به من يشاء من عباده إذا هم يستبشرون | 262554 |
| 4 | | وهو الذي أنزل من السماء ماء فأخرجنا به نبات كل شيء فأخرجنا منه خضرا نخرج منه حبا متراكبا ومن النخل من طلعها قنوان دانية وجنات من أعناب والزيتون والرمان مشتبها وغير متشابه انظروا إلى ثمره إذا أثمر وينعه إن في ذلكم لآيات لقوم يؤمنون | 33614.0 |
| 5 | | أيود أحدكم أن تكون له جنة من نخيل وأعناب تجري من تحتها الأنهار له فيها من كل الثمرات وأصابه الكبر وله ذرية ضعفاء فأصابها إعصار فيه نار فاحترقت كذلك يبين الله لكم الآيات لعلكم تتفكرون | 33274.0 |
| 6 | | مثل الجنة التي وعد المتقون فيها أنهار من ماء غير آسن وأنهار من لبن لم يتغير طعمه وأنهار من خمر لذة للشاربين وأنهار من عسل مصفى ولهم فيها من كل الثمرات ومغفرة من ربهم كمن هو خالد في النار وسقوا ماء حميما فقطع أمعاءهم | 32966.0 |
| 7 | | ألم تر أن الله أنزل من السماء ماء فأخرجنا به ثمرات مختلفا ألوانها ومن الجبال جدد بيض وحمر مختلف ألوانها وغرابيب سود | 32946.0 |
| 8 | | الذي جعل لكم الأرض مهدا وسلك لكم فيها سبلا وأنزل من السماء ماء فأخرجنا به أزواجا من نبات شتى | 32912.0 |
| 9 | | ثم كلي من كل الثمرات فاسلكي سبل ربك ذللا يخرج من بطونها شراب مختلف ألوانه فيه شفاء للناس إن في ذلك لآية لقوم يتفكرون | 32914.0 |
| 10 | | وهو الذي مد الأرض وجعل فيها رواسي وأنهارا ومن كل الثمرات جعل فيها زوجين اثنين يغشي الليل النهار إن في ذلك لآيات لقوم يتفكرون | 170338.0 |

**Level-4 Sample**

In this sample, three more similar ayahs to the *firstAyah* are found computationally. It also shows that with the proposed approach, the similarity is independent from the sizes of the similar ayahs. Computational similarity may happen between ayahs with large difference in size.

| | *firstAyah* | *secondAyah* | Measure |
|---|---|---|---|
| 1 | كذبت قبلهم قوم نوح وعاد وفرعون ذو الأوتاد | وإن يكذبوك فقد كذبت قبلهم قوم نوح وعاد وثمود | 3.35544E7 |
| 2 | | ألم يأتهم نبأ الذين من قبلهم قوم نوح وعاد وثمود وقوم إبراهيم وأصحاب مدين والمؤتفكات أتتهم رسلهم بالبينات فما كان الله ليظلمهم ولكن كانوا أنفسهم يظلمون | 2097178 |
| 3 | | كذبت قبلهم قوم نوح والأحزاب من بعدهم وهمت كل أمة برسولهم ليأخذوه وجادلوا بالباطل ليدحضوا به الحق فأخذتهم فكيف كان عقاب | 1048602 |
| 4 | | كذبت قبلهم قوم نوح وأصحاب الرس وثمود | 1048590 |
| 5 | | كذبت قبلهم قوم نوح فكذبوا عبدنا وقالوا مجنون وازدجر | 524306.0 |
| 6 | | ألم يأتكم نبأ الذين من قبلكم قوم نوح وعاد وثمود والذين من بعدهم لا يعلمهم إلا الله جاءتهم رسلهم بالبينات فردوا أيديهم في أفواههم وقالوا إنا كفرنا بما أرسلتم به وإنا لفي شك مما تدعوننا إليه مريب | 65578.0 |
| 7 | | مثل دأب قوم نوح وعاد وثمود والذين من بعدهم وما الله يريد ظلما للعباد | 32790.0 |
| 8 | | وعاد وفرعون وإخوان لوط | 4098.0 |
| 9 | | ولقد فتنا قبلهم قوم فرعون وجاءهم رسول كريم | 2126.0 |
| 10 | | أوعجبتم أن جاءكم ذكر من ربكم على رجل منكم لينذركم واذكروا إذ جعلكم خلفاء من بعد قوم نوح وزادكم في الخلق بسطة فاذكروا آلاء الله لعلكم تفلحون | 1066.0 |

**Level-5 Sample**

In this sample, it is clear that similarity may happen between one ayah and other ayahs but at different portions. Ghirnati has considered the similarity between *firstAyah* based on a certain portion of it. On the other hand, the

computationally found most similar ayah showed similarity with the *firstAyah* but in a different portion of it.

| | *firstAyah* | *secondAyah* | Measure |
|---|---|---|---|
| 1 | قل من يرزقكم من السماء والأرض أمن يملك السمع والأبصار ومن يخرج الحي من الميت ويخرج الميت من الحي ومن يدبر الأمر فسيقولون الله فقل أفلا تتقون | يخرج الحي من الميت ويخرج الميت من الحي ويحيي الأرض بعد موتها وكذلك تخرجون | 1.0995E12 |
| 2 | | أمن يبدأ الخلق ثم يعيده ومن يرزقكم من السماء والأرض أإله مع الله قل هاتوا برهانكم إن كنتم صادقين | 2.68436E8 |
| 3 | | يا أيها الناس اذكروا نعمت الله عليكم هل من خالق غير الله يرزقكم من السماء والأرض لا إله إلا هو فأنى تؤفكون | 3.35546E7 |
| 4 | | إن الله فالق الحب والنوى يخرج الحي من الميت ومخرج الميت من الحي ذلكم الله فأنى تؤفكون | 2359606.0 |
| 5 | | قل من يرزقكم من السماوات والأرض قل الله وإنا أو إياكم لعلى هدى أو في ضلال مبين | 2097514.0 |
| 6 | | تولج الليل في النهار وتولج النهار في الليل وتخرج الحي من الميت وتخرج الميت من الحي وترزق من تشاء بغير حساب | 1048698.0 |
| 7 | | أفلم يروا إلى ما بين أيديهم وما خلفهم من السماء والأرض إن نشأ نخسف بهم الأرض أو نسقط عليهم كسفا من السماء إن في ذلك لآية لكل عبد منيب | 524474.0 |
| 8 | | ولو أن أهل القرى آمنوا واتقوا لفتحنا عليهم بركات من السماء والأرض ولكن كذبوا فأخذناهم بما كانوا يكسبون | 262230.0 |
| 9 | | والله أخرجكم من بطون أمهاتكم لا تعلمون شيئا وجعل لكم السمع والأبصار والأفئدة لعلكم تشكرون | 131242.0 |
| 10 | | ثم سواه ونفخ فيه من روحه وجعل لكم السمع والأبصار والأفئدة قليلا ما تشكرون | 131154.0 |

**Level-6 Sample**

| | *firstAyah* | *secondAyah* | Measure |
|---|---|---|---|
| 1 | ويوم نبعث في كل أمة شهيدا عليهم من أنفسهم وجئنا بك شهيدا على هؤلاء ونزلنا عليك الكتاب تبيانا لكل شيء وهدى ورحمة وبشرى للمسلمين | ثم آتينا موسى الكتاب تماما على الذي أحسن وتفصيلا لكل شيء وهدى ورحمة لعلهم بلقاء ربهم يؤمنون | 2097802.0 |
| 2 | | وما أنزلنا عليك الكتاب إلا لتبين لهم الذي اختلفوا فيه وهدى ورحمة لقوم يؤمنون | 266306.0 |
| 3 | | أولم يكفهم أنا أنزلنا عليك الكتاب يتلى عليهم إن في ذلك لرحمة وذكرى لقوم يؤمنون | 262410.0 |
| 4 | | إنا أنزلنا عليك الكتاب للناس بالحق فمن اهتدى فلنفسه ومن ضل فإنما يضل عليها وما أنت عليهم بوكيل | 262374.0 |
| 5 | | لقد كان في قصصهم عبرة لأولي الألباب ما كان حديثا يفترى ولكن تصديق الذي بين يديه وتفصيل كل شيء وهدى ورحمة لقوم يؤمنون | 262246.0 |
| 6 | | قل نزله روح القدس من ربك بالحق ليثبت الذين آمنوا وهدى وبشرى للمسلمين | 32850.0 |
| 7 | | ويوم نبعث من كل أمة شهيدا ثم لا يؤذن للذين كفروا ولا هم يستعتبون | 174606.0 |
| 8 | | ونزعنا من كل أمة شهيدا فقلنا هاتوا برهانكم فعلموا أن الحق لله وضل عنهم ما كانوا يفترون | 165006.0 |
| 9 | | ولولا فضل الله عليك ورحمته لهمت طائفة منهم أن يضلوك وما يضلون إلا أنفسهم وما يضرونك من شيء وأنزل الله عليك الكتاب والحكمة وعلمك ما لم تكن تعلم وكان فضل الله عليك عظيما | 8922.0 |
| 10 | | هو الذي أنزل عليك الكتاب منه آيات محكمات هن أم الكتاب وأخر متشابهات فأما الذين في قلوبهم زيغ فيتبعون ما تشابه منه ابتغاء الفتنة وابتغاء تأويله وما يعلم تأويله إلا الله والراسخون في العلم يقولون آمنا به كل من عند ربنا وما يذكر إلا أولو الألباب | 8382.0 |

**Level-7 Sample**

| | *firstAyah* | *secondAyah* | Measure |
|---|---|---|---|
| 1 | أفأمنتم أن يخسف بكم جانب البر أو يرسل عليكم حاصبا ثم لا تجدوا لكم وكيلا | أم أمنتم من في السماء أن يرسل عليكم حاصبا فستعلمون كيف نذير | 262234.0 |
| 2 | | أم أمنتم أن يعيدكم فيه تارة أخرى فيرسل عليكم قاصفا من الريح فيغرقكم بما كفرتم ثم لا تجدوا لكم علينا به تبيعا | 134174.0 |
| 3 | | أأمنتم من في السماء أن يخسف بكم الأرض فإذا هي تمور | 8242.0 |
| 4 | | وهو القاهر فوق عباده ويرسل عليكم حفظة حتى إذا جاء أحدكم الموت توفته رسلنا وهم لا يفرطون | 4170.0 |
| 5 | | يا بني إسرائيل قد أنجيناكم من عدوكم وواعدناكم جانب الطور الأيمن ونزلنا عليكم المن والسلوى | 1206.0 |
| 6 | | ولئن شئنا لنذهبن بالذي أوحينا إليك ثم لا تجد لك به علينا وكيلا | 1142.0 |
| 7 | | إذا لأذقناك ضعف الحياة وضعف الممات ثم لا تجد لك علينا نصيرا | 1074.0 |
| 8 | | يرسل عليكما شواظ من نار ونحاس فلا تنتصران | 1070.0 |
| 9 | | وما كان لبشر أن يكلمه الله إلا وحيا أو من وراء حجاب أو يرسل رسولا فيوحي بإذنه ما يشاء إنه علي حكيم | 618.0 |
| 10 | | أفأمن الذين مكروا السيئات أن يخسف الله بهم الأرض أو يأتيهم العذاب من حيث لا يشعرون | 618.0 |

**Level-8 Sample**

| | *firstAyah* | *secondAyah* | Measure |
|---|---|---|---|
| 1 | وما أرسلنا من قبلك إلا رجالا نوحي إليهم من أهل القرى أفلم يسيروا في الأرض فينظروا كيف كان | أفلم يسيروا في الأرض فينظروا كيف كان عاقبة الذين من قبلهم دمر الله عليهم وللكافرين أمثالها | 2.8823E17 |
| 2 | | أفلم يسيروا في الأرض فينظروا | 2.88 |

Journal of Theoretical and Applied Information Technology
15th September 2024. Vol.102. No. 17
© Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

| | | | |
|---|---|---|---|
| | عاقبة الذين من قبلهم ولدار الآخرة خير للذين اتقوا أفلا تعقلون | كيف كان عاقبة الذين من قبلهم كانوا أكثر منهم وأشد قوة وآثارا في الأرض فما أغنى عنهم ما كانوا يكسبون | 23E17 |
| 3 | | أولم يسيروا في الأرض فينظروا كيف كان عاقبة الذين من قبلهم وكانوا أشد منهم قوة وما كان الله ليعجزه من شيء في السماوات ولا في الأرض إنه كان عليما قديرا | 1.4411E17 |
| 4 | | أولم يسيروا في الأرض فينظروا كيف كان عاقبة الذين من قبلهم كانوا أشد منهم قوة وأثاروا الأرض وعمروها أكثر مما عمروها وجاءتهم رسلهم بالبينات فما كان الله ليظلمهم ولكن كانوا أنفسهم يظلمون | 7.206E16 |
| 5 | | أولم يسيروا في الأرض فينظروا كيف كان عاقبة الذين كانوا من قبلهم كانوا هم أشد منهم قوة وآثارا في الأرض فأخذهم الله بذنوبهم وما كان لهم من الله من واق | 1.4074E14 |
| 6 | | وما أرسلنا من قبلك إلا رجالا نوحي إليهم فاسألوا أهل الذكر إن كنتم لا تعلمون | 1.0995E12 |
| 7 | | قل سيروا في الأرض فانظروا كيف كان عاقبة الذين من قبل كان أكثرهم مشركين | 4.2950E9 |
| 8 | | وما أرسلنا قبلك إلا رجالا نوحي إليهم فاسألوا أهل الذكر إن كنتم لا تعلمون | 1.3421E8 |
| 9 | | قد خلت من قبلكم سنن فسيروا في الأرض فانظروا كيف كان عاقبة المكذبين | 4260114.0 |
| 10 | | لقد بعثنا في كل أمة رسولا أن اعبدوا الله واجتنبوا الطاغوت فمنهم من هدى الله ومنهم من حقت عليه الضلالة فسيروا في الأرض فانظروا كيف كان عاقبة المكذبين | 4260038.0 |

**Level-9 Sample**

| | *firstAyah* | *secondAyah* | Measure |
|---|---|---|---|
| 1 | وكذلك أنزلناه حكما عربيا ولئن اتبعت أهواءهم بعدما جاءك من العلم ما لك من الله من ولي ولا واق | ولن ترضى عنك اليهود ولا النصارى حتى تتبع ملتهم قل إن هدى الله هو الهدى ولئن اتبعت أهواءهم بعد الذي جاءك من العلم ما لك من الله من ولي ولا نصير | 1.0995E12 |
| 2 | | ولئن أتيت الذين أوتوا الكتاب بكل آية ما تبعوا قبلتك وما أنت بتابع قبلتهم وما بعضهم بتابع قبلة بعض ولئن اتبعت أهواءهم من بعد ما جاءك من العلم إنك إذا لمن الظالمين | 1179750 |
| 3 | | ألم تعلم أن الله له ملك السماوات والأرض وما لكم من دون الله من ولي ولا نصير | 262254 |
| 4 | | وما أنتم بمعجزين في الأرض ولا في السماء وما لكم من دون الله من ولي ولا نصير | 262246 |
| 5 | | إن الله له ملك السماوات والأرض يحيي ويميت وما لكم من دون الله من ولي ولا نصير | 262238 |
| 6 | | وما أنتم بمعجزين في الأرض وما لكم من دون الله من ولي ولا نصير | 262230 |
| 7 | | فمن حاجك فيه من بعد ما جاءك من العلم فقل تعالوا ندع أبناءنا وأبناءكم ونساءنا ونساءكم وأنفسنا وأنفسكم ثم نبتهل فنجعل لعنت الله على الكاذبين | 131242 |
| 8 | | قد كانت لكم أسوة حسنة في إبراهيم والذين معه إذ قالوا لقومهم إنا برآء منكم ومما تعبدون من دون الله كفرنا بكم وبدا بيننا وبينكم العداوة والبغضاء أبدا حتى تؤمنوا بالله وحده إلا قول إبراهيم لأبيه لأستغفرن لك وما أملك لك من الله من شيء ربنا عليك توكلنا وإليك أنبنا وإليك المصير | 32914 |
| 9 | | وكذلك أنزلناه قرآنا عربيا وصرفنا فيه من الوعيد لعلهم يتقون أو يحدث لهم ذكرا | 16998.0 |
| 10 | | وكذلك أنزلناه آيات بينات وأن الله يهدي من يريد | 16546.0 |

**Level-10 Sample**

| | *firstAyah* | *secondAyah* | Measure |
|---|---|---|---|
| 1 | وإلى ثمود أخاهم صالحا قال يا قوم اعبدوا الله ما لكم من إله غيره قد جاءتكم بينة من ربكم هذه ناقة الله لكم آية فذروها تأكل في أرض الله ولا تمسوها بسوء فيأخذكم عذاب أليم | ويا قوم هذه ناقة الله لكم آية فذروها تأكل في أرض الله ولا تمسوها بسوء فيأخذكم عذاب قريب | 1.51116E23 |
| 2 | | وإلى مدين أخاهم شعيبا قال يا قوم اعبدوا الله ما لكم من إله غيره قد جاءتكم بينة من ربكم فأوفوا الكيل والميزان ولا تبخسوا الناس أشياءهم ولا تفسدوا في الأرض بعد إصلاحها ذلكم خير لكم إن كنتم مؤمنين | 1.47574E20 |
| 3 | | وإلى ثمود أخاهم صالحا قال يا قوم اعبدوا الله ما لكم من إله غيره هو أنشأكم من الأرض واستعمركم فيها فاستغفروه ثم توبوا إليه إن ربي قريب مجيب | 1.84467E19 |
| 4 | | وإلى مدين أخاهم شعيبا قال يا قوم اعبدوا الله ما لكم من إله غيره ولا تنقصوا المكيال والميزان إني أراكم بخير وإني أخاف عليكم عذاب يوم محيط | 1.75921E13 |
| 5 | | وإلى عاد أخاهم هودا قال يا قوم اعبدوا الله ما لكم من إله غيره إن أنتم إلا مفترون | 1.75921E13 |
| 6 | | وإلى عاد أخاهم هودا قال يا قوم اعبدوا الله ما لكم من إله غيره أفلا تتقون | 1.75921E13 |
| 7 | | لقد أرسلنا نوحا إلى قومه فقال يا قوم اعبدوا الله ما لكم من إله غيره إني أخاف عليكم عذاب يوم عظيم | 4.39805E12 |
| 8 | | ولقد أرسلنا نوحا إلى قومه فقال يا قوم اعبدوا الله ما لكم من إله غيره أفلا تتقون | 4.39804E12 |
| 9 | | فأرسلنا فيهم رسولا منهم أن اعبدوا الله | 4.29 |

| | | | |
|---|---|---|---|
| | | ما لكم من إله غيره أفلا تتقون | 496E9 |
| 10 | | ولا تمسوها بسوء فيأخذكم عذاب يوم عظيم | 5.36871E8 |

## 4. CONCLUSION

In this paper, an algorithm is proposed to measure the similarity between a pair of Quranic ayahs. To show the value of the proposed algorithm in finding similar *second* ayahs for a given *first* ayah, the most similar ayahs for the ayahs stated in a selected specialized book are found computationally. The most similar ayahs that are found computationally for the *first* ayahs in the book are projected into levels. A similar *second* ayah is projected into level *i*, if it is ranked as the *i*-th ayah in the list of the computationally found similar ayahs for the *first* ayah. It was found that about 38% of similar ayahs that are stated in the selected book exactly matched those that were found computationally (i.e. they were projected into the first level). However, computationally found similar ayahs show more similar ayahs than those that were stated in the selected book. The main reason behind this difference is that scholars did not only consider syntactical similarity but also semantic similarity when presenting similar pairs of ayahs. Another reason is that similarity between a given first ayah and other ayahs could be in different portions of this first ayah. Thus, the scholar may have focused on similarity in specific portions of the ayah, while there may be other portions of the same ayah showing more similarity with other ayahs. The main result of this work is showing that computation may help in discovering similar pairs of ayahs that were overlooked by the scholars in this field. As a future work, an ayah similarity dictionary may be built using the proposed algorithm. Also, the same approach may be used to measure the similarities between Quranic surahs.

## REFERENCES:

[1] Dice, and R. Lee, “Measures of the amount of ecologic association between species”, *Ecology* 26, No. 3, 1945, pp. 297-302

[2] V. Levenshtein, “Binary codes capable of correcting deletions, insertions, and reversals”, In *Soviet physics doklady*, Vol. 10, No. 8, 1966, pp. 707-710.

[3] F. Smith and F.M. Waterman, “Identification of common molecular subsequences”, *Journal of molecular biology* 147, no. 1, 1981, pp. 195-197.

[4] M. Jaro, “Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida, *Journal of the American Statistical Association,* Vol. 84, No. 406, 1989, pp. 414-420.

[5] W. Winkler, “String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage.” (1990).

[6] D. Knuth, “The Art of Computer Programming, Volume 3, Sorting and Searching.” *Addison-Wesley. Section* 5: 1998, pp. 75-80.

[7] D. Gusfield, “Algorithms on stings, trees, and sequences: Computer science and computational biology." *ACM Sigact News* 28, No. 4, 1997, pp. 41-60.

[8] M. Akour, I. Alsmadi and I. Alazzam, “MQVC: Measuring quranic verses similarity and surah classification using N-gram”. WSEAS Transactions on Computers, vol. 13, 2014.

[9] M. Alshammeri, E. Atwell, and M. Alsalka, “Detecting semantic-based similarity between verses of the Quran with Doc2vec”. Procedia Computer Science, 189, 2021, pp. 351-358.

[10] A. Sharaf and E. Atwell. “QurSim: A corpus for evaluation of relatedness in short texts.” In LREC, 2012, pp. 2295-2302.

[11] A.F. Huda, M.R. Deyana, Q.U. Safitri, W. Darmalaksana, and U. Rahmani, “Analysis Partition Clustering and Similarity Measure on Al-Quran Verses.” In 2019 IEEE 5th International Conference on Wireless and Telematics (ICWT), 2019, pp. 1-5.

[12] S. Saeed, S. Haider, and Q. Rajput, “On finding similar verses from the holy quran using word embeddings.” International Conference on Emerging Trends in Smart Technologies (ICETST), Pakistan, 2020, pp. 1-6. Technologies (ICETST), 2020.