Verification of Qur'anic Quotations Embedded in Online Arabic and Islamic Websites

Yasser M. Alginahi[1,2,a], Omar Tayan[1,2,b] and Muhammed N. Kabir[1,2,c]

IT Research Center for the Holy Qur'an (NOOR) and Department of Computer Science
College of Computer Science and Engineering
Taibah University
P.O Box 344
Madinah, Saudi Arabia
{yginahi[a], otayan[b], mkabir[c]}@taibahu.edu.sa

ABSTRACT

With the Internet penetration in the Middle East exceeding the world's average penetration and the growth of Internet users exceeding the 2000% in the last decade, the number of Arabic documents online has seen an exponential growth. Therefore, with the increase of Arabic online publications, the presence of Qur'anic quotations in online articles is becoming more notable. However, very few online documents use complete diacritics and Tajweed symbols for all words. Therefore, the verification of the quoted Quran text, the diacritics and the Tajweed symbols is essential for verse authenticity checking. In this paper, a verification algorithm, is proposed in order to verify the authenticity of Qur'anic quotes/verses (including full and partial verses). The system developed was tested, using Qur'an quotes found in online Arabic documents and proved to be effective in detecting any tampering or unintentional mistakes during the dissemination of such documents.

Keywords: Qur'an, verification, Arabic, authentication

1. Introduction

The recent advancements in technology have changed society and the way we live. This is more noticeable in developing countries, and no doubt that with the increasing dependence on technology, more emphasis is laid on the digitization of documents transmitted and published online. Therefore, with the wide use of the Internet in the Middle East and the increase in Internet penetration, the Arabic content on the Internet is growing exponentially. According to www.internetworldstats.com, this has pushed Arabic to be ranked the 7th top language in the Internet in 2010, with the increase in Internet penetration reported as above the world's average in 2012.

Qur'anic verses are quoted in many online Islamic and Arabic articles. Hence, it is important that quoted verses are written with no mistakes and that the Qur'anic verses must be authenticated to guarantee that the word of Allah is not tampered-with; even a single misplacement or omitting of a letter or diacritic considers the text as invalid. The objective of this work is to provide a framework for verification and authentication of the fundamental text of the Qur'anic verses embedded within other texts. Typically, the Qur'anic verses to be verified, are completely or partially decorated with diacritics (Tashkeel) and Tajweed symbols (recitation rules symbols/signs).

The Qur'an consists of 114 chapters of varying lengths (called Surahs). Each chapter consists of several verses (called Ayat). The number of verses and length of each verse differs from chapter to chapter. The total number of verses in the Qur'an exceeds 6000 verses. The Qur'an text is in Arabic and it uses diacritics. In addition to use of diacritics, the Qur'an text contains recitation symbols, which guide readers to recite the Qur'an correctly. These Tajweed symbols are not used in Arabic writing and they are unique to the Holy Qur'an. These diacritics and/or Tajweed symbols are used with every letter in the Qur'an. Figure 1 shows an example of a verse from the Qur'an with complete diacritics and Tajweed symbols. Table 1 shows the different diacritics used in Arabic and Table 2 shows some of the Tajweed symbols used in the Qur'an.
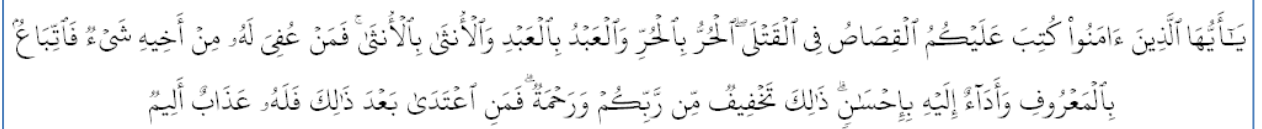

Fig. 1: A Sample Ayah from the Qur'an [Surah No. 2, Ayah 177]

Tajweed is considered as a discipline by itself; through learning the Tajweed rules a person can learn how to read/recite the Qur'an; how each syllable is pronounced, location of pauses, pronunciation of letters …. etc.

Table 1: Arabic Diacritics

Unicode	Name	Symbol
064B	TanweenFattha	ً
064C	TanweenDummah	ٌ
064D	TanweenKesra	ٍ
064E	Fattha	َ
064F	Dummah	ُ
0650	Kesra	ِ
0651	Shaddah	ّ
0652	Sukkun	ْ

Table 2: Examples of some Tajweed Symbols

Symbol	Unicode	Symbol	Unicode
؞	061E	ۜ	06DC
ۖ	06D6	ۢ	06E2
ۗ	06D7	ۣ	06E3
ۘ	06D8	ۥ	06E5
ۙ	06D9	ۨ	06E8

The use of diacritics in regular/daily Arabic writing, is not too much noticed. Diacritics are more common in classical Arabic books and poetry writing. In many documents, the random use of diacritics may be noticed on the last letters of some words when it is ambiguous to read. The use of diacritics in online Arabic documents is rarely seen; however, it is always noticed in Qur'anic quotes. These diacritics may be fully, partially or even randomly used in such documents. In addition, it is very common to use diacritics (whether fully or partially) with classical poetry and literature. For Arabic speakers, the only way to differentiate the diacritic-less words is to locate them within the context. In (Debili, et al., 2002), an analysis of 23,000 Arabic scripts showed that there is an average of 11.6 possible ways to assign diacritics for every diacritic-less word.

The objective of this work is to propose a verification algorithm to check the authentication of Qur'anic quotes with full, partial or random diacritics and Tajweed symbols found in online Arabic documents.

This paper is organized as follows: section 2 presents the background on related work, followed by the methodology for the proposed Qur'an quote verification mechanism in section 3. The results and discussions are presented in section 4, and finally, concluding remarks are provided in section 5.

2. Literature Survey

The research area of Qur'an-IT is an emerging area of research in which researchers have until recently started to give more attention. One particularly important challenge in the Qur'an-IT domain is in digital Qur'an verification and authentication, an essential requirement in this era of technology where digital text documents are disseminated over the Internet as the dominant media type. Therefore, a mechanism to verify and authenticate Qur'anic quotes/verses is necessary to make sure that no tampering has been made to even a single letter or symbol of the Qur'anic text. The literature in this area is very scarce, and the only work that provides a close resemblance to our work is found in (Alshareef & El Saddik, 2012). Here, the Qur'an e-citation is verified as compared to the original text of the Qur'an.

Other related work includes the Qur'anic meta-codification (Shamsudin & Farooq, 2000). Here the authors proposed the use of an "atomization structure" and Unicode codification on the Qur'an in terms of the number of verses in each chapter, the number of characters in each verse and so on, in order to protect the digital Qur'an from any distortion or tampering.

Other search methods in the literature are based on semantic search, information retrieval and grammar dependency; however, to the best knowledge of the authors, no work has been based on the Qur'anic text. Therefore, in this paper, a three-stage model is proposed to illustrate the verification and authentication of Qur'anic quotes which use full, partial or random diacritics and Tajweed symbols as found in electronic Arabic text documents.

3. Methodology

The use of Qur'anic quotes in online Arabic and Islamic documents emphasizes the need for a Qur'an quotation verification and authentication algorithm. Normally, Qur'anic quotes in online documents appear between brackets and the reference, Surah name and Ayah-number are provided after the quotation. Figure 2 shows two examples of Qur'anic quotes

from online documents. Figure 2 (a) is a part of an online document obtained from www.islamhouse.com and it contains full diacritics and Tajweed symbols. On the other hand, Figure 2(b) is part of an online document obtained from www.articles.islamweb.net/ which contains no diacritics or Tajweed symbols.

اعلم أن الأمة اعتنت بنص القرآن الكريم وحفظه، وعلى رأس الأمة	بعد أن ذكر سبحانه صفات المؤمنين الشارين أنفسهم طلباً لمرضاة الله، وبعد أن ذكر صفات
نبينا محمد ﷺ فقد بلغ من عنايته بالقرآن الكريم وحرصه على حفظه	الكافرين المعجبين بأقوالهم، المخاصمين لشرعه، والصادين عن أمره ونهيه، دعا سبحانه
أن كان يعاجل جبريل حين يقرئه القرآن حتى أنزل الله عليه : ﴿ لَا تُحَرِّكۡ بِهِۦ	عباده المؤمنين إلى التزام شرعه، والتمسك بدينه، فخاطبهم بقوله: {يَا أَيُّهَا الَّذِينَ آمَنُوا
لِسَانَكَ لِتَعۡجَلَ بِهِۦٓ ۞ إِنَّ عَلَيۡنَا جَمۡعَهُۥ وَقُرۡءَانَهُۥ ۞ فَإِذَا قَرَأۡنَٰهُ فَٱتَّبِعۡ قُرۡءَانَهُۥ ۞ ثُمَّ إِنَّ عَلَيۡنَا بَيَانَهُۥ ۞ ﴾	ادخلوا في السلم كافة ولا تتبعوا خطوات الشيطان انه لكم عدو مبين} (البقرة: 208).
(القيامة : ١٦–١٩) وأنزل عليه : ﴿ إِنَّا نَحۡنُ نَزَّلۡنَا ٱلذِّكۡرَ وَإِنَّا لَهُۥ لَحَٰفِظُونَ ۞ ﴾	تتجه هذه السطور إلى الوقوف على سبب نزول هذه الآية، ومن ثم تنعطف لإلقاء بعض
(الحجر: ٩) فعندئذ اطمأن الرسول ﷺ على القرآن وكان جبريل	الضوء على المراد من شقها الأول. وهو قوله سبحانه: {يا أيها الذين آمنوا ادخلوا في السلم
يعارض الرسول ﷺ القرآن في كلِّ سنة، وعارضه القرآن في سنته التي	كافة}.
(a)	(b)

Fig 2: Examples of online documents containing Quran Quote with (a) full diacritics (b) no diacritics

The methodology used in this paper consists of three stages. The first stage begins with the extraction of the input, i.e., the Qur'anic quotation (usually a verse (Ayah)). Depending on the way the article is written, the text is extracted either automatically or manually, then the surah name and ayah number are verified in order to eliminate any mistakes during the entry/typing process of the article. The second stage involves the removal of all the diacritics and Tashkeel in order to filter the fundamental text (e.g. characters comprising the words) within the verse. Finally, the third stage involves the verification process: The Unicode representation of the characters is used to verify the fundamental text against the original Unicode text of the Qur'an in order to authenticate the fundamental text of the Qur'anic quotation. Thereafter, the diacritics and Tashkeel are verified, and in the case of any tampering or misallocating of any character or symbol, the Qur'anic quotation is considered invalid, regardless if the fundamental text is authentic or not. The proposed Quran quotation verification and authentication system is shown in Fig. 3.
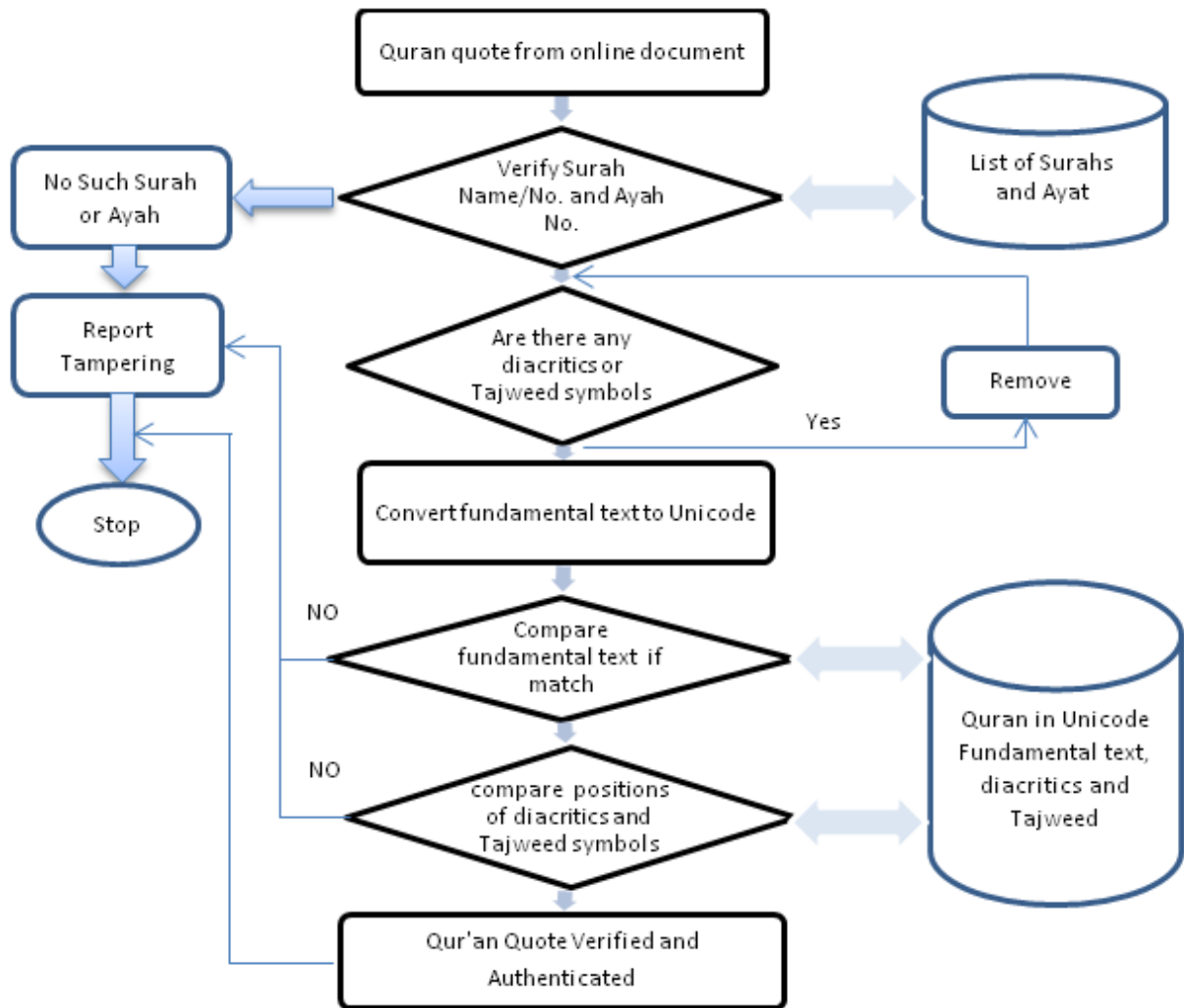
Fig. 3: Proposed Quran Quote Verification and Authentication System

4. Results and Discussion

The proposed system was implemented and tested on hundreds of Ayahs which were taken from the text version of the Holy Quran produced by the King Fahd Quran Complex for Printing the Holy Quran. Thereafter, several Ayahs were modified by switching the positions of some letters and/or diacritics or Tajweed symbols. The system proved to be very effective for detecting any tampering or distortion to any letter, diacritic or Tajweed symbols. Table 3 shows several examples where the system detected some tampering in the Qur'anic quotation and reported that the quote is tampered with.

Table 3: Results for the Proposed Qur'an Quotation Verification System

No	Verse text	Surah Name	Verse no.	Fundamental Text Verified	Diacritics &Tajweed symbols verified	Report
1	وَرَفَعْنَا لَكَ ذِكْرَكَ (a) وَرَفَعْنَا لَكَ ذِكْرَكَ (b)	الشرح	4	✔	✔	Verified and Authenticated
2	فَأَمَّا مَن ثَقُلَتْ مَوَازِينُهُ (a) فَأَمَّا مَن ثَقُلت مَوَازِيُهُ (b)	القارعة	6	✘	✘	Tampered with
3	لَيْلَةُ ٱلْقَدْرِ خَيْرٌ مِّنْ أَلْفِ شَهْرٍ (a) لَيْلَةُ ٱلقدر خَيْرٌ مِّنْ أَلْف شهر (b)	القدر	3	✔	✔	Verified and Authenticated
4	إِنَّ إِلَى رَبِّكَ ٱلرُّجْعَىٰ (a) إِنَّ إِلَى رُبِّكَ ٱلرُّجْعَىٰ (b)	العلق	8	✔	✘	Tampered with

Table 3 shows four examples with each verse written twice. The original, which has not been tampered with is shown as (a), and the verse shown as (b) represents the verse being scrutinized. Example 1 shows that no tampering or removing of any diacritics or Tajweed symbols performed on the ayah. Example 2 shows that the fundamental text was tampered with, since one letter was removed, which at the same time shifted the diacritic to the letter before reporting an unauthentic tampered verse. For the third verse in example 3, some of the diacritics and Tajweed symbols have been removed without modifying the text or adding any other symbols or letters, which means that the ayah is considered authentic even though some diacritics have been removed. In example 4, a diacritic was replaced with another diacritic; therefore, the ayah is considered as tampered-with, even though the fundamental text was untouched.

Finally, the proposed system shows that it is capable of detecting any small tampering in the text, diacritics and/or Tajweed symbols, which will invalidate the ayahs and consider them unauthentic. The goal here is to report such tampering to the organizations publishing such articles containing Qur'anic quotations so that they correct those mistakes, or otherwise be blacklisted and reporting of such websites as propagating false information (e.g. incorrect Qur'anic verses) due to their continued propagation of distorted ayah/s of the Qur'an.

5. Conclusion

The Qur'an quote verification algorithm developed in this paper helps the user to verify the Qur'anic e-citations over the Internet, and to provide confidence in the accuracy of the Qur'anic e-citations. The proposed system executes three stages, where each stage detects any intentional or unintentional tampering or distortion in Qur'anic quotations used in online text documents. Finally, such a system could help in warning an organization/digital-publisher of any tampered/distorted Qur'anic quotes posted on their websites and at the same time provide a list of verified/authentic Arabic and Islamic websites that publish Qur'anic verses.

Acknowledgement

References

Abdulrhman Alshareef and Abdulmotaleb El Saddik, "A Qur'anic Quote Verification Algorithm for Verses Authentication", 2012 International Conference on Innovations in Information Technology (IIT), pp. 339 – 343.

Debili, F., Achour, H., & Souissi, E. (2002).Del'etiquetage grammatical a' la voyellationautomatique de l'arabe.Correspondances (Vol. 71, pp. 10-28). Tunis: Institut de Recherchesur le Maghreb Contemporain.

Shamsudin A. F. and Farooq, A., "AI natural language in metasynthetics of Al-Qur'an," 2000 TENCON Proceedings. Intelligent Systems and Technologies for the New Millennium (Cat.No.00CH37119), pp. 464-467.