CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Overview

This chapter summarizes the presented research work, research contributions, and future work recommendations. This thesis aimed to design and model a new representation of the digital Quran with efficient memory management using the presentation layer mechanism, compressed content structure, and handling of Quran word duplications.

A critical concern, especially for Muslims, is ensuring that information technologies be employed smoothly and safely in the service of the Holy Quran. As such, such technologies cannot be exploited or propagated (Tayan, 2014; Alshareef & El Saddik, 2012). Thus, a primary motivation for this thesis was to design a model as a digital content structure and authentication for the digital Quran to address those sensitive issues and to exert and motivate efforts to enact necessary protection measures to ensure that modern information technology is used only to propagate accurate and authentic digital Quran. Another reason is to guarantee that any deliberate/inadvertent changes to the Quran are recognized, and that required correction procedures are implemented.

The creation of digital Quran multimedia material is primarily accomplished via the use of authentication and integrity procedures, which ensure that Quran content can be delivered and stored in its whole at any time and from any location on the Internet. Integrity refers to ensuring that transmitted data is not altered or tampered with in any way and remains identical to the version delivered. Integrity may be accomplished by various approaches, including encryption, hashing, and watermarking. On the other hand, authentication relates to building trust between communication parties, such as assurance through verifying the provenance of materials from a trustworthy source/publisher. Digital signatures/certificates and digital watermarking can be used to authenticate. In comparison, confidentiality and non-repudiation considerations are less critical for Quran distribution since we are more concerned with spreading accurate/correct material than with whether the content may be accessible by other third parties during online transmission.

This study also develops a digital Quran model that not only optimizes space but has the properties of tamperproof and content integrity. In addition, this thesis demonstrated the model with Surah Al-Fatihah and Surah Al-Baqarah as the test case being the longest chapter in Al-Quran. All steps and forms are based on experts' standard criteria and consensus. In addition, this thesis provided an overview of the Digital Quran Computing research work that covers the problems in the digital Quranic model and analyses the state-of-art models.

6.2 Responses to Research Objectives

This section describes the findings, research methodology and deliverables of the defined research objectives. This thesis set out to develop a digital Quran model (DQM). The research process is divided into three main sections: literature review, development, and evaluation.

Research Objective 1: To review the extant literature on the digital Quran and issues related to digital content structure and authentication.

To achieve research objective 1, this thesis defined the following question:

Research Question 1: What are the challenges in digital Quran computing, issues related to digital content structure and optimized storage space?

To achieve the first objective, as described in Chapter 1 (Section 1.5), literature on related work and state of the art of digital Quran technologies was reviewed (Chapter 2) as the first step. The literature review is carried out using the guidelines in past work. After analyzing the studies, the research questions are formulated, zooming into reviewing the digital Quran on the related works. This is done to identify open issues, research gaps, related theories, and dominating techniques, as mentioned in Chapter 2.

Research Objective 2: To propose a new Digital Quran model using hexadecimal representation and compressed sparse matrix techniques for content structure with Quranic word duplication handling.

To achieve research objective 2, this thesis defined the following question:

Research Question 2: How to improve the current representation of the Digital Quran that can optimize space, thus enabling a lightweight version?

Research Question 3: To what extent the authenticity element can be enhanced for the Digital Quran model by optimizing memory space, tamperproof and content integrity features?

To achieve the second objective, as described in Chapter 1 (Section 1.5), This thesis proposed a new technique for representing words and verses of the Holy Quran as a new model (DQM) (Chapter 3 and Chapter 4) which a hexadecimal digital representation technique for words in Arabic using UTF-8 Arabic Presentation Forms A and B Unicode Standard 7.0 Arabic letters for character encoding which is backwards compatible with ASCII code. This representation proves more than 50% reduction in the storage memory space, thus increasing the search speed. The storage of words is optimized by using one memory space for that word rather than one memory space for each word's Arabic character. This representation provides an optimized digital representation of Arabic words.

On the other hand, since the Quran has 14,870 unique words and many repetitions, duplicated words were indexed. Using the UTF-8 character representation format backwards compatible with ASCII code, the words and verses were constructed to represent the Holy Quran in hexadecimal structure using a sparse matrix. The matrix

represents each word and verse of the digital Quran arranged according to the word ID from the index lookup table. Thus, the DQM performs well in memory management, which will help to optimize memory for the whole digital Quran.

To further achieve the second objective, a digital Quranic content structure was developed using hexadecimal representation, word ID lookup table and compressed sparse matrix (Chapter 3 and Chapter 4).

Research Objective 3: To evaluate the performance of the proposed Digital Quran model in terms of space optimization in parallel to preserving content integrity.

To achieve research objective 3, this thesis defined the following question:

Research Question 4: To what extent can the elements be enhanced for the Digital Quran model by optimizing memory space yet maintaining content integrity?

Research Question 5: How does the Digital Quran Model (DQM) perform in terms of space optimization and still maintain content integrity?

To achieve the third objective, as described in Chapter 1 (Section 1.5), the performance measures used are identified to evaluate and compare research works, as mentioned in Chapters 3 and 5. The experimental result performs the evaluation, which is analyzed and discussed in Chapter 5.

First, the results corresponding to the functional correctness of DQM were detailed. The DQM is functionally correct, and the results of measuring the performance overhead

caused by the proposed DQM and its underlying infrastructure were described. This included measuring the memory usage overhead, which depicted that the DQM caused minimal memory usage and maximised the reduction, i.e. more than 50% of memory overhead was reduced. Discussion on comparing the proposed DQM and existing techniques shows a significant reduction of Quranic text content structure representation.

Integrity tests were also conducted with error words, non-Quranic text and non-Quranic text with Arabic words in the Quran. DQM passes the integrity test and can distinguish non-Quranic text or errors in Quranic words or verses.

6.3 Contribution of this Study

This study was able to achieve the following contributions:

a. Quranic word presentation using hexadecimal based on UTF8 Unicode 7.0 Arabic characters.
b. Quranic verses representation based on unique word id cross-reference to word index lookup table due to the nature of the Quran that has many word repetitions.
c. Quranic surah representation using compressed sparse matrix for further efficient memory management and processing.
d. Digital Quran model with efficient memory management and ability to detect non-Quranic Arabic text or incorrect words or verses.

6.4 Future Research Directions

This study could be considered a corner block for future studies. It could be applied to other different documents and languages. Future work on this problem should continue to explore ways to integrate knowledge into the topic modelling DQM algorithms to create indexing for the whole Holy Quran and automate the generation of surah in hexadecimal which recognizes as an anti-tamper algorithm that will manipulate the binary bitwise operation.

This study could be considered a corner block for the Digital representation of any language. It could be applied to other languages such as Hindi, Tamil, Malay Jawi, Urdu, etc. However, the limitation of this study only focused on surah Al-Fatihah and surah Al-Baqarah.

On the other hand, this study may lead researchers to develop and modify an algorithm to build a suitable query automatically to switch between languages as an application for Arabic information retrieval. On the other hand, this study may lead researchers to develop and modify an algorithm to build suitable techniques automatically for the Arabic language, especially Quran words and verses with diacritics.

Finally, this technique can optimize the memory space & solve the problem of word duplications and. also can be extended to other non-Roman characters to serve system programmers for non-English text such as Hindi, Chinese, and Japanese categorized in Unicode standards.