CHAPTER 1

INTRODUCTION

1.1 Introduction

The Holy Quran, in Arabic, is the most important source of knowledge in Islam. It is considered a primary reference with guidance, valuable stories, prophetic traditions, and wisdom (Khafajeh et al., 2010).

Arabic is a language that holds the miracle of the Holy Quran and accomplishes all the requirements of Arabic and Islamic civilisation at its flourishing peak (Kanaan & Wedyan, 2006). There are many kinds of Arabic fonts in various ways of writing. The shape of the letters will change according to their position in a word. The current research builds on recent findings within the extant literature and aims to provide a better understanding while contributing to the subject.

This chapter discusses the history of Quran printing and publications until the modern world of the digital Quran.

1.2 History of Quran Printing and Publications

Quran is the sacred, the most authentic, and tamper-proof book of Allah SWT since its revelation over 14 centuries ago. Usually, people read the Quran using the traditional printed version called Mushaf. The purpose of this section is to show evidence of how

careful the Prophet's companions (sahaba) were in collecting, validating, and finally producing the Mushaf Uthmani for the Muslim world. This is the model of Quran preservation from the *sahaba* to ensure the authenticity of the Quran. Even towards Rasulullah wafat, the angel *Gabriel* (r.a) came to revise the Quran with Rasulullah to ensure preservation.

According to Sheikh/ Khalifa *Ezzat*, the compilation of the Holy Quran took place after Prophet Muhammad SAW demise in three stages. The initial collection of the Quran took place during the Caliphate of Abu Bakr. During the Caliphate of Uthman, the Quran was compiled without diacritics. Dots were added to certain letters, diacritical marks, and finally, punctuation in the form of pause marks and circled verse numbering (Ezzat, 2009). The Caliph Abu Bakr kept this first compilation. After his death, the copy was passed to his successor, Caliph Umar, who, on his deathbed, handed it to his daughter, Hafsa, one of the *Ummahatul Mu'minin* (Mohammad Taqi et al., 2000; Hadith, 2012).

According to Leaman (2006), Prophet Muhammad SAW assigned Zayd ibn Thabit as the primary scribe whose duty was to gather all of the Quranic text. The task required Zayd ibn Thabit to collect written copies of the Quran and validate each verse with the oral testimony of at least two companions. The Quran was organised under the auspices committee of four senior ranking Companions headed by Zayd ibn Thabit. This canonical corpus is closed and fixed because nothing can be changed or modified in the Quran.

According to Foda et al. (2013), the Digital Qur'an consists of 68 different characters, where 44 characters can be stored in an ANSI text file, whilst the other 24 characters can be in a UTF-8 text file. They use UTF-8 character encoding for the digital Quran to represent chapter name, verse number and page number.

The recent advent of innovative technologies like smartphones, digital devices and tablets has transformed daily life routines under a single touch. Muslims are also adopting these new tools with exponential growth.

The presence of Islamic applications on the Internet is very impressive where. Muslims have been using online technology for everything imaginable and in all different aspects of life, from spreading their religion to educating themselves about their faith and communicating. Muslims have been using the Internet for online education and distance education, Quran memorisation, Quran teaching, online businesses, banking, socialising, politics and communication, sharing news, research works, and spreading their religion (Mohammed Zakariah, 2017).

According to Sameera et al. (2017), there are 451 online Islamic applications, and 209 are digitalization of Quran applications. Converting such Quranic data into digital format is a challenging task. However, online Quran and Islamic books are lagging in employing structured digital content (Larsson & Hoffman, 2012).

Hakak et al. (2017) defined content integrity protection as the approach where all possible techniques employed to protect certain content or can be used for protection are put together. Hakak divided content integrity authenticity into searching and verification.

The first is to search specific verses and match the same with a verifiable source. In the searching and matching phase, one of the fundamental requirements is the availability of verified content, which means the authentic database. If there is no verified content, the data integrity check fails. Next, the verification is in the final phase with several verification mechanisms such as Hashing, Brute Force and Watermarking.

Therefore, there is a need to develop a mechanism for an authentic digital Quran with anti-tampering and content integrity. This becomes an issue because previously, the Quran was protected through the memorization of hafiz before technology adoption. Today most paper-based documents are digitized with the support of new technologies. There are many multilingual translated copies of the Holy Quran available online, initially written in the Arabic language. All verses in the Quran are identical for all Arabic versions.

1.3 Problem Statement

Al Quran is the foundation of various disciplines in Quranic studies, including Tafseer, recitation, teaching and translation. Quoting Quranic verses has been used by scholars in supporting the messages they are passing across, including their basis of information evaluation (Noordin & Othman, 2006; Alshareef & El Saddik, 2012). Traditionally, Quran publishing has a strict and tight discipline with zero error tolerance and 100% accuracy.

Arabic characters are represented by UTF-8 character encoding having compatibility with the ASCII code in a backward manner. UTF-8 is a variable-sized coding method to encode the text; each character needs two bytes to code (Almazrooie et al., 2020; Hilal & Hilal, 2019). Within the scope of current research, it has been noted in the relevant literature that design systems should be introduced or implemented to enhance the current features (e.g. Almazrooie et al., 2020; Alanazi, Khan & Gutub, 2020). In a recent study by Almazrooie et al. (2020), it was noted that the Unicode aspect could be a moulder to be considered for future studies. While this study emphasizes a sparse compression matrix regarding storage optimization for the digital Quran, this model is based on recent studies that direct the current research direction. Additionally, integrity verification, compression methods and other relevant elements have been recommended as scholarly paths toward a better understanding of the subject. This is because matrixes and assessment methods can significantly vary based on context and usage (Almazrooie et al., 2020).

Digital Holy Quran comes in many forms, such as portable data files (pdf) containing image-based Holy Quran, text files, applications books, and raw data (Unicode format). Multimedia forms such as images, audio, and video consume memory space, which is not lightweight. Almazrooie et al. (2020) proposed a cryptographic hash function for digital Holy Quran verses and converted the Arabic alphabet to hexadecimal for compression. Saada and Zhang (2015) proposed a similar approach, which converts the

English alphabet to hexadecimal for compression using DNA sequences compression algorithm based on hexadecimal representation.

The compression method proposed by Almazrooie et al. (2020) includes the sample representing the verse **الله الصمد** with the size of 33 Bytes, and after the compression, it becomes 17 Bytes. The same algorithm was applied to the verse **اياك نعبد وإياك نستعين** then the sample size was reduced from 73 Bytes to 38 Bytes.

Almazrooie and Hakak adopt strings concatenation, where the length correlates with the number of letters in a word, thus requiring space. However, the algorithm ignores duplication of word اياك where space can be optimized with handling words duplication mechanism.

On the other hand, Larsson and Hoffman (2012) and Alsmadi and Zarzour (2015) also mentioned that the digital Quran is lagging in employing structured digital content. Almazrooie et al. (2020) used 6236 rows of elements to represent each verse in the Quran. However, Hakak et al. (2017) used 6234 rows. The vulnerabilities within these features that convey the Quran's contents are unclear and need more research efforts (Hakak et al., 2017). Thus, a mechanism for optimizing the space of the digital Quran is proposed by handling duplications and string manipulation with compressed structured content of the table representation for the Quran, which relies on the length of each verse.

Over the last several years, Internet users accessing content connected to the digital Quran and Hadith have increased. This has boosted the demand for Quran authentication systems to distinguish between authentic and forged verses (Hakak et al., 2017).

Therefore, the vulnerabilities of the Digital Quran are blurry and thus lack robustness and are prone to threats (Alsmadi & Zarzour, 2017). Considering the importance of the integrity of the Holy Quran and its sciences among all Muslims, all those Quranic verses must be authentic and free from tampering and distortion.

Additionally, Hakak et al. (2017) said that cryptography, stenography, watermarking, and digital signature were less authentic and not suitable for preserving the integrity of sensitive documents for the Digital Quran that adopt a text-based format as compared to image-based, thus requiring a new digital representation method to ensure its content integrity.

1.4 Research Questions

In light of the problems mentioned earlier, the research questions are;

a. What are the challenges in digital Quran computing, issues related to digital content structure and optimized storage space?

b. How to improve the current representation of the Digital Quran that can optimize space, thus enabling a lightweight version?

c. To what extent can the elements be enhanced for the Digital Quran model by optimizing memory space yet maintaining content integrity?

d. How does the Digital Quran Model (DQM) perform in terms of space optimization and still maintain content integrity?

1.5 Research Objectives

The main objective of this research is to propose a digital Quranic model with memory optimizations using hexadecimal representation and compressed sparse matrix yet preserving the content integrity. This is embedded within the scope of this research as the final desired outcome is to enhance the Digital Quran. Based on its optimized size, it will be more convenient for users to install it on their devices as a lightweight standard application. This is the core objective of this research that directs its conduct. To achieve this goal, the current research follows a strategic approach that is as follows:

a. To review the extant literature on the digital Quran and issues related to digital content structure and authentication.
b. To propose a new Digital Quran model using hexadecimal representation and compressed sparse matrix techniques for content structure with Quranic word duplication handling.
c. To evaluate the performance of the proposed Digital Quran model in terms of space optimization in parallel to preserving the content integrity.

1.6 Research Scope and Limitation

By what was noted above, the scope of this research is to enhance the storage usage of the Digital Quran to enhance its usability as an end-product that users can install as a lightweight standard application on their devices. Therefore, the algorithm for words and verses conversion in hexadecimal representation based on UTF-8 Arabic Presentation

Forms A and B Unicode Standard 7.0 for Arabic characters is examined (UTF8, 2015; Unicode Consortium,2015). As for DQM performance evaluation purposes, Surah Al-Baqarah was selected as a test case, the longest chapter in Al-Quran. Besides that, Surah Al-Fatihah was included since it is the first chapter of the Al-Quran, which the Muslims also recite in the compulsory five-time daily prayers. This study uses ***tanzeel.net*** as the primary source, built upon the recent findings and conceptual frameworks in the relevant literature of the subject (e.g. Almazrooie et al., 2020), and extended their work by using non-diacritic Quranic text.

This research is performed under three phases: reviewing the state of the art of Digital Quran-related work, new Digital Quran model development, and performance evaluation phase. The undertaken phases are explained in detail in chapter 3 of this study. The model was implemented as a proof of concept using a development method (Construction of techniques) with compact, lightweight, authentic and content integrity features. Figure 1.1 illustrates the undertaken phases for this study.

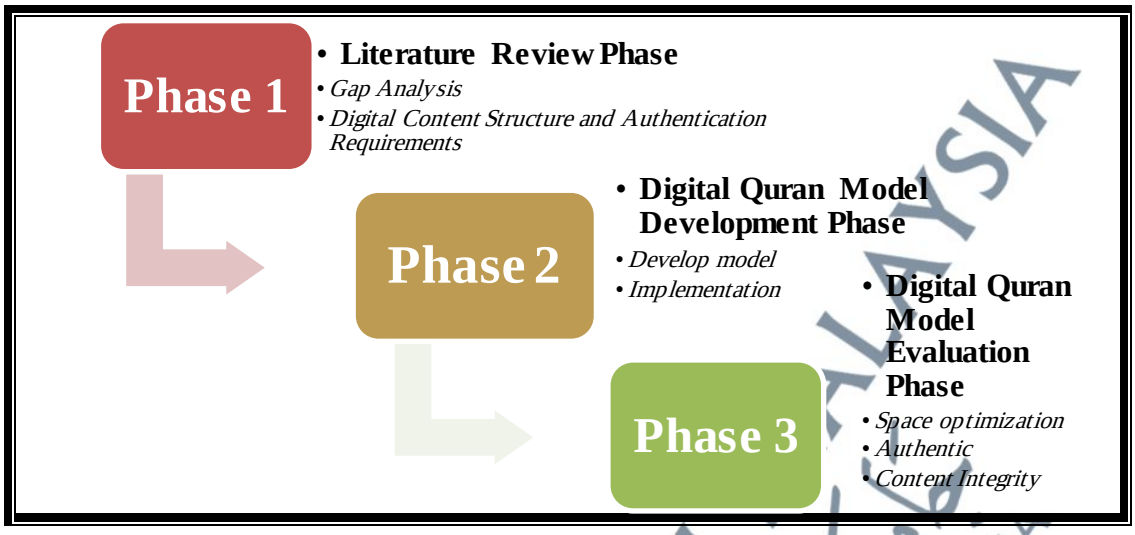

Figure 1.1: Overview of Research Activities in Phases

1.7 Research Contributions

This research aims to contribute to the knowledge of space optimization for the digital Quran model and its academic literature. This research uses a new representation for Digital Quran Model through Hexadecimal UTF-8 Arabic Presentation Forms A and B Unicode Standard 7.0. Furthermore, DQM is assessed regarding the optimization of memory space and tamperproof elements. Prior studies have established a pathway for the current study, and thus, the setting for this research is built upon previous findings and recommendations (Almazrooie et al., 2020; Mouratidis et al., 2013).

1.8 Organization of the Thesis

This research consists of six chapters; the structure of the thesis is as follows:

Chapter 1 Introduction In this introductory chapter, the rationale for this study is explained, and an overview of the thesis is provided. The Chapter presents the context in which this study was conducted and the research background. Furthermore, it consists of the introduction and scope, research problem and overview of the research.

Chapter 2 Literature Review This chapter begins with a comprehensive history of the digital Quran to understand the digital Quran's detailed functionality in the real world from an Islamic perspective. This is followed by explaining the role of the Digital Quran in the real world in addition to Vulnerability Issues for the Digital Quran and Content Integrity. Chapter 2 provides a conceptual framework for the current research based on the works of scholars to answer the research questions.

Chapter 3 Research Methodology In this chapter, the research methodology identifies the current research approach. Requirements of the model, block diagrams, step-by-step functions and an overview of the design are among the topics of this chapter.

Chapter 4 New Digital Quran Model to Optimize Storage This chapter presents the development of the DQM model by word conversion with a compressed table for the content structure and handling Quran words' duplications that optimize space.

Chapter 5, Digital Quran Model Implementation and Evaluation, focus on developing the DQM prototype as proof of concept and implementation, the performance comparison evaluation result and implementation for surah Al-Fatihah and surah Al-Baqarah.

Chapter 6 Conclusion and Future Work This chapter includes the conclusion of this thesis. Suggestions for future work are presented, which direct other scholars to conduct more comprehensive studies in this context to enhance further the model presented by this research.

1.9 Summary

This chapter presents a summary of the thesis, including the research background, problem statement, research questions, research objectives, research contributions, research scope and limitation, and organization of the thesis. The research background discusses the literature on the research domain and the study's motivation. The problem statement reviews the main related work closer to the study area and highlights the gap and problems under study. The next chapter discusses the literature review for the study, which mainly summarizes the literature on Digital Quran.