DIGITAL QURAN WITH STORAGE OPTIMIZATION THROUGH DUPLICATION HANDLING AND COMPRESSED SPARSE MATRIX METHOD

Ashraf Saleh Mohammad Alomoush

Thesis submitted in partial fulfilment for the degree of
DOCTOR OF PHILOSOPHY
SCIENCE AND TECHNOLOGY

UNIVERSITI SAINS ISLAM MALAYSIA

November 2022

AUTHOR DECLARATION

I hereby declare that the work in this thesis is my own except for quotation andsummaries, which have been duly acknowledged.

Date: 15th November 2022 Signature

Name: Ashraf Saleh Mohammad Alomoush

Matric No: 4120083

Address: Amman-Jordan

ACKNOWLEDGEMENTS

Praise is to ALLAH, the Almighty, Most Gracious, and Most Merciful, who enabled me and gave me everything to complete this thesis. I would also like to extend my thanks and appreciation to my supervisor, Prof. Ts. Dr Norita Md Norwawi, who was the source of inspiration at all stages of this research. Much of this work would have been impossible without her guidance, comments, and encouragement.

In the life of Saleh, the father

In the life of Aziza, the mother

To my wife

To Darling Dawood, the son

To Darling Hor, the daughter

To Darling Kenz, the daughter

To my friends

May ALLAH bless you all .....

ABSTRAK

Penyelidikan ini mengkaji topik Al-Quran Digital sebagai aplikasi yang digunakan oleh sebilangan besar orang di seluruh dunia dan melalui pelbagai peranti (cth. telefon pintar, tablet dan komputer). Ketika ini Quran Digital berformatkan teks, memanafaatkan perwakilan Hexadesimal. Namun , ruang storan tidak dioptimumkan oleh penggunaan penyatuan rentetatan aksara dalam Unicode bagi suatu kalimah Quran. Maka. saiz storan akan berkadar terus dengan panjang kalimah apabila ditukar bentuk Unicode. Oleh kerana itu kajian ini mencadangkan model Al-Quran Digital (DQM) yang baharu bertujuan untuk mengurangkan saiz ruang penyimpanan melalui perwakilan kalimah dalam Hexadecimal menggantikan penyatuan rentetan aksara. Selanjutnya saiz storan dioptimumkan lagi dengan pengendalian kalimah Quran yang berulang. Ini selari dengan struktur matriks jarang yang dimampatkan untuk perwakilan ayat-ayat dan surah-surah Al-Quran. Fokus penyelidikan ini adalah pengoptimuman storan untuk Quran Digital menggunakan format teks. DQM telah dibangunkan menggunakan Visual Studio dan pelayan Java dimana kualiti penyelesaian diukur dengan perbandingan saiz fail sebelum dan selepas menggunakan model DQM. Untuk surah Al-Baqarah, pengurangan saiz storan ialah 50.76%, manakala surah Al-Fatihah ialah 69.81%. Hasil penyelidikan ini konsisten dengan kajian lain dalam konteks yang sama dan memberikan implikasi dalam kedua-dua domain teori dan praktikal. Selain dari analisis pengurangan saiz ruang, pakar bidang al-Quran juga ditemuduga berkaitan Al-Quran Digital yang dibangunkan. Hasil kajian ini membantu pembangun aplikasi membina Quran Digital yang boleh dimanfaatkan oleh pengguna kerana kebolehpercayaan, kesahihan dan pengoptimuman storan supaya ia boleh digunakan sebagai aplikasi standard.

ABSTRACT

This research looks into the topic of the Digital Quran as an application used by many people worldwide and via various devices (e.g. smartphones, tablets, and computers). The current digital Quran text-based format benefitted the Hexadecimal representation. However, space was not optimized due to the string concatenation approach of the Unicode of each letter that occurs in a word. Thus, the storage size for the Unicode representation is directly proportional with the length of a word. In this regard, the current study proposes a new Digital Quran Model (DQM) that aims to reduce storage requirements through word conversion into Hexadecimal instead of combining letter sequences. This approach is further optimized through Quranic word duplication handling. This is paralleled with a compressed sparse matrix representing Quranic verses and surah. The focus of this research is storage optimization for text-based Quran content formats. DQM was implemented using Visual Studio and Java servers, where the solution quality was measured by comparing the file size before and after applying the DQM model. For surah Al-Baqarah, the reduction in the storage size was 50.76%, and Al-Fatihah was 69.81%. The results of this research are consistent with other studies in the same context and provide implications in both theoretical and practical domains. In addition to space reduction analysis, the current research interviews domain experts on the proposed digital Quran applications. The outcome helps the researchers to develop a Digital Quran that users can benefit from due to its reliability, validity, and optimized storage so that it can be used as a standard application.

الملخص

يبحث هذا البحث في موضوع القرآن الرقمي كتطبيق يستخدمه العديد من الأشخاص في جميع أنحاء العالم وعبر الأجهزة المختلفة (مثل الهواتف الذكية والأجهزة اللوحية وأجهزة الكمبيوتر). أستفاد التنسيق الرقمي الحالي المستند إلى النص القرآني من التمثيل السداسي العشري. ومع ذلك، لم يتم تحسين المساحة بسبب نهج تجميع السلسلة لتشفير كل حرف يحدث في الكلمة الواحدة. وبالتالي ، فإن حجم التخزين لتمثيل اليونيكود يتناسب طرديًا مع طول الكلمة. في هذا الصدد، تقترح الدراسة الحالية نموذجًا جديدًا للقرآن الرقمي يهدف إلى تقليل متطلبات التخزين من خلال تحويل الكلمات إلى نظام سداسي عشري بدلاً من الجمع بين تسلسل الحروف. تم تحسين هذا النهج بشكل أكبر من خلال تقليل التكرار للكلمات و يتوازى هذا مع مصفوفة متفرقة مضغوطة تمثل الآيات والسور القرآنية. يركز هذا البحث على تحسين التخزين لتنسيقات محتوى القرآن النصية. تم تنفيذ نموذج القرآن الرقمي باستخدام البرمجة المرئية وخوادم الجافا حيث تم قياس جودة الحل من خلال مقارنة حجم الملف قبل وبعد تطبيق نموذج القرآن الرقمي. بالنسبة لسورة البقرة ، كان الانخفاض في حجم التخزين ٥٠.٧٦٪ ، والفاتحة ٦٩.٨١٪. تتوافق نتائج هذا البحث مع دراسات أخرى في نفس السياق وتوفر آثارًا في كل من المجالات النظرية والعملية. بالإضافة إلى تحليل تقليل المساحة، يقوم البحث الحالي بإجراء المقابلات مع خبراء المجال حول تطبيقات القرآن الرقمي المقترحة. تساعد النتيجة الباحثين على تطوير مصحف رقمي يمكن للمستخدمين الاستفادة منه نظرًا لموثوقيته وصلاحيته وتخزينه الأمثل بحيث يمكن استخدامه كتطبيق قياسي.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AMR	Adaptive Multi-Rate
API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
ASR	Automatic Speech Recognizer
DCT	Discrete Cosine Transform
HEX	Hexadecimal
DQM	Digital Quran Model
DCT	Discrete-Cosine Transform
DWT	Discrete Wavelet Transform
FFT	Fast-Fourier Transform
LSB	Least Significant Bit
PoS	Part of Speech
QoC	Quran-on-Chip
QR	Quick Response Code
Qu Hex	Quran using Hexadecimal
SME	Subject Matter Expert
SVD	Singular-Value Decomposition
UTF	Unicode Transformation Format
XOR	Exclusive OR