الإسترجاع الذكي للمعلومات من القرآن الكريم: تطبيق للأجهزة الذكية

هدى عمر الجلعود

بحث مقدم لنيل درجة الماجستير في علوم الحاسب

إشراف
د. محمد ذهب

مشرف مساعد

د. محمود كامل

كلية الحاسبات وتقنية المعلومات
جامعة الملك عبد العزيز
جدة – المملكة العربية السعودية

المستخلص

القرآن الكريم يشجع المسلمين في جميع آياته على التعلم والبحث في مختلف مجالات المعرفة في لغة غنية بالمصطلحات. ووفقا لدراسة حديثة فإن الأجهزة النقالة ستتجاوز أجهزة الكمبيوتر كأداة رئيسية بحلول عام ٢٠٢٠. هذا التطور ألهم الباحثين لاستغلال هذا الهاتف الذكي في مختلف المجالات وخاصة في مجال استرجاع المعلومات. ونتيجة لزيادة استخدام الهواتف النقالة في العالم جذب استرجاع المعلومات من خلال الهاتف الذكي اهتماما متزايدا.

الرد على استفسارات مستخدمي الهواتف النقالة بذكاء هي واحدة من التحديات الكبيرة في استرجاع المعلومات في الأنظمة الذكية. في هذه الدراسة تمكنا من حل مشكلة استرجاع أيات القران الكريم باستخدام تحويل المويجات المتقطعة (DWT) واستخراج النتائج على أساس قرب كلمات البحث في نصوص الآيات . مع العلم ان جميع الدراسات التي أجريت على استرجاع المعلومات القرآنية لم تتطرق لدراسة قرب كلمات البحث وإظهار النتائج معتمدة على ذلك.

في هذا البحث نقوم بتصميم وتنفيذ نظام ذكي لاسترجاع أيات القرآن والذي يستغل إشارة الكلمات وتحويل المويجات المتقطعة ثم إظهار نتائج البحث حسب قرب الكلمات في نصوص الآيات في تطبيق للأجهزة الذكية. سنركز على مشكلة القرب بين الكلمات القرآنية كمثال للاسترجاع الذكي للمعلومات من الأجهزة في محاولة لتطبيق التقنيات الذكية لاسترجاع المعلومات وتحقيق أداء عالي وتجاوز القيود في الأجهزة الذكية . باستخدام خوارزمية (SBIRM) طورنا خوارزمية جديدة باسم (DDMDWT) تستغل اختلاف اطوال المستندات والتحويلات الرياضية لتمثيل المستندات. وقد تم تطوير نموذج (DDMDWT) الذي طور إشارة الكلمة من خلال تناسب عدد المقاطع مع عدد الكلمات بالمستند باستخدام إشارة الكلمة الديناميكية. نموذج (DDMDWT) ساهم في تقليل تعقيد نموذج (SBIRM) وتصغير حجم الإنديكس ٢٠,٩٨ %. كما حقق زيادة ملحوظه في دقة النتائج عند المعيار (P@1) و (P@3).

في نظامنا بالبحث الذكي في أيات القران ومن أجل المعالجة المسبقة لأيات القرآن الكريم فقد تم تطوير قائمة (Stopwords) ومراجعتها يدويا، كما تم تطبيق اثنين من تقنيات التجذير في اللغة العربية. واحدة استنادا لمعجم الجذور المجردة للقران الكريم [١] والأخرى استنادا لجذع الكلمة. وأظهرت النتائج أن الأولى تفوقت على الأخيرة في (RECALL) و (F-measure) . وفي المقابل تقنية استخدام الجذع تفوقت في معيار (precision) و (MAP).

ولإنتاج برنامج ذو كفائه عالية تم عمل ثلاث تجارب عمليه على الأجهزة الذكية. الاولى كانت باستخدام طريقتين لتخزين بيانات البرنامج. التجربة الثانية تمت معالجة البيانات بطريقتين (Offline)و(Online). التجربة الأخيرة كانت لدراسة تأثير استخدام تقنيات التجذير على كفائه البرنامج في الهاتف الذكي.

[١] من تأليف محمد الدباغ

Intelligent Information Retrieval in Holy Quran: Smartphone Application

By Huda Omar Aljalaoud

A thesis submitted for the requirements of the degree of Master of Computer Science

Supervised By
Dr. Mohamed Dahab
Dr. Mahmod Kamel (Co-Advisor)

FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY

KING ABDULAZIZ UNIVERSITY

JEDDAH – SAUDI ARABIA

Abstract

The Holy Quran encourages Muslims throughout all its verses to learn and search in various fields of knowledge, with a language rich in its terminology. According to a recent study mobile devices will surpass computers as the primary tool by 2020. This development has inspired researchers to exploit this smartphone mobile in various areas, especially in the field of mobile information retrieval (IR). Mobile IR is considered a subset of traditional IR. Answering mobile users' queries intelligently is one of the significant challenges in Information Retrieval in intelligent systems. In this research, an attempt to solve the verse retrieval problem by using Discrete Wavelet Transform (DWT) with IR. While most studies on Quranic information retrieval have been conducted for Quranic online software or web-portals, this research focuses on the on-device Quranic IR.

An intelligent Quranic retrieval (IQR) system was design and implemented which exploits term signal and the DWT to answer the user query in smartphone application. Based on Spectral-Based Information Retrieval Model (SBIRM) we propose a novel document model, termed Dynamic Document Model with Discrete Wavelet Transforms DDMDWT. The DDMDWT exploits the variation of documents length and mathematical transforms for document representation. The proposed model will enhance the existing term signal concept by additionally taking into consideration differing lengths document. In IQR, the DDMDWT model contributes to reducing the time complexity of SBIRM and decreasing the index size by 20.98%, all while achieving improvement in precision, recall, and F-measure with compared to SBIRM. Also demonstrates how the DDMDWT model delivers a notable increase in the precision of the P@1 and P@3.

To preprocess the Holy Quran in IQR system, the Stopwords file was developed then applied two different techniques of stemming. By used Dictionary-Lookup and light 10 stemmers to study the impact of different approach of stemmer's in an accuracy of Quranic IR and mobile performance. The results show that the former outperformed the latter in terms of recall and F-measure. On the other hand, the Light 10 precision and MAP results exceeded the other.

In IQR system three mobile experimental studies applied to optimizing smartphone performance by a tradeoff between CPU efficiency and memory capacity. In first experiment used two kinds of storage techniques for a dataset. Second investigated the ability of offline and online stemmers to raise mobile efficiency. The last, study the influence of different stemmer technique in smartphone performance. Based on the results, IQR make offline preprocessing by use Light10 stem based and store dataset in text format.