WatsaQ: Repository of Al hadith in Bahasa (Case study: Hadith Bukhari)

Conference Paper · September 2017

DOI: 10.1109/EECSI.2017.8239077

4 authors, including:

Dewi Khairani
Syarif Hidayatullah State Islamic University Jakarta
20 PUBLICATIONS **8** CITATIONS
SEE PROFILE

Rizal Broer Bahaweres
Syarif Hidayatullah State Islamic University Jakarta
48 PUBLICATIONS **68** CITATIONS
SEE PROFILE

Nashrul Hakiem
Syarif Hidayatullah State Islamic University Jakarta
42 PUBLICATIONS **68** CITATIONS
SEE PROFILE

WatsaQ: Repository of Al Hadith in Bahasa (Case Study: Hadith Bukhari)

Atqia Aulia, Dewi Khairani, Rizal Broer Bahaweres, and Nashrul Hakiem

Department of Informatics
UIN Syarif Hidayatullah
Jakarta, Indonesia

atqia.aulia13@mhs.uinjkt.ac.id, {dewi.khairani, rizalbroer, hakiem}@uinjkt.ac.id

***Abstract* - The Hadith is one of the two sources of Islamic law after the Qur'an. It is a fact that there are a number of false hadith, recognised by Muslim scholars since the end of the first century of Hijra, and even earlier. In addition to the breadth of false hadith circulating among the public at this time, it is difficult to determine the source of authenticity and distinguish false from genuine. This is due to the configuration of the genuine documents which are revealed in Arabic. There are applications and websites that provide online functionalities related to hadith includes hadith search. However, all these hadith websites are displayed information as plain text, thus this situation can result in difficulties to analyse and extract the information or linkage of one hadith to another. The main contribution of this research is that we develop the ontology can support the repositories of Hadith, and provide a base for an online hadith corpus. The authors implemented a repository of translation in Bahasa of Bukhari Hadith using XML schema. To study the repository performance, we use a web presentation using PHP employing brute-force string match algorithms to display the search results based on keywords entered by the user. We analyse the results of our proposed repository implementation average searching time is faster by 0.85 milliseconds compared with the repository based on the unstructured one.**

Keywords - corpus, repository, XML, hadith Bukhari, brute force.

I. INTRODUCTION

Most Muslims believe that the hadith is the vehicle of the Sunnah (teachings, deeds, and saying of the Islamic Prophet Muhammad) and that the hadith is the guidance that cannot be ignored when attempting to understand Allah's revelations. As one of the sources of Islamic authority second only to the Qur'an (the Islamic sacred book), a number of hadith literature has a decisive influence are considered a source of legal and religious inspiration. Various scholars have striven to collect, clarify and sort the authentic hadith from the false [1].

The fact that there are a number of false hadith, in addition to the authentic has been recognized by Muslim scholars since the end of the first century of Hijra (Islamic Calendar), or even earlier. It has been assumed that the collection of hadith, which flourished in the first century of Hijra, is a collection of both fake and authentic hadith [1].

As for the causes of counterfeiting hadith it is (1) an act of heresy as the people who change the meaning of the Qur'an and Sunnah have literally broken a trust and are contrary to the basic principles of the Qur'an, (2) fanaticism or wish a to win over a certain group, (3) those who falsify according to their fancies, and (4) to promote the existence of political strife.

Unfortunately, due to the large amount of false hadith circulating today, this can cause great difficulties in society especially as the hadith also act as a source of law after the Qur'an. It can be very difficult to determine and distinguish the source of authenticity of hadith due to the configuration of such lengthy documents which are revealed in Arabic. This causes ordinary people who do not understand Arabic to find it increasingly difficult to find a particular topic or learn the content of the hadith.

Currently, we have a lot of mobile applications and websites that provide online information and functionalities related to hadith and other Islamic law in Indonesian language which includes searching. However, all these sources are displayed information as plain text, thus this situation can result in difficulties to learn, analyse and extracted the Hadith.

The purpose of undertaking this research is to build a repository source of Islamic law in the Indonesian language with a case study of Hadith Bukhari that provide the ontology that can support the functionality of the repositories.

The contents of this paper are as follows: Literature Review is in Section II, Related work which also contains previous result in Section III; Section III shows the Repository Development; next Section is Result and discussion; and finally, conclusion and future works.

II. LITERATURE REVIEW

A. Knowledge Management

The role of knowledge management can be understood through the activities that seek to develop and maintain the dynamism and competitiveness of companies relying on such the resources (knowledge assets).

According to [2], the cycle of Knowledge Management is described in which there are five stages: creating, sharing, structuring, using, and auditing. Ontology symbolizes the hierarchical structuring of knowledge by subcategorizing them per their essential qualities [3]. Ontologies capable of represent a well-defined, standardized and basic form of knowledge or repositories.

B. Repository

According to [4] a repository is a physical space (building, room, area) which is used for permanent storage backup or temporary storage (manuscripts, rare books, government documents, papers, photos, etc.).

C. Corpus Analysis

Corpus analysis is a form of text analysis that allows comparisons to be made between textual objects on a large scale. It allows the determination of things that are not

necessary when text is read as a human being. If there are several collections of documents, it may be possible to find a pattern in the use of grammar, or a phrase that is often repeated in a corpus. It may be required to search for phrases that are liked or not liked statistically by an author or a particular type of text, a certain type of grammatical structure or many instances of a particular concept in a large number of documents in context. Corpus analysis is very useful for testing the intuition of text and/or triangulation results from other digital methods [5].

D. Extensible Mark Up Language (XML)

XML (Extensible Markup Language) is a standard to define data in a format that is simple and flexible where the web service supports communication among applications and application integration using XML and the Web [6]. Typical uses of XML are as follows:

a. By using an XML database system, make it easier to store and query data that has format or semi-structured documents [7].
b. XML is able to display the data in a structured format that can be easily understood by an application or a human (human application-usable) [8].
c. An XML document has to be transparent, i.e: easily accessed, read, and modified, so any computer that is connected can easily exchange data information [8].

E. Hadith

According to the experts in hadith [9], hadith understanding involves all the words of the Prophet Muhammad teachings, deeds and things. The meaning of all the happenings are narrated from the Prophet SAW associated with characteristics, history, and habits.

F. Imam Bukhari

Imam Bukhari's full name was Abu Abdullah Muhammad bin Ismail bin Ibrahim bin al-Mughiroh Bardizbahal bin al-Bukhari. He was born in Bukhara Uzbekistan, into a wealthy and respectable family. His father, who died when Bukhari was a child, was a pious and erudite man.

Since childhood Bukhari revealed a noble personality, and the ability to memorise hadith with a powerful intelligence and outstanding capabilities. He studied the hadith before he was even 10 years old. At age 11, he was able to correct errors in hadith quotations made by his teachers.

To expand and deepen his knowledge of hadith, Bukhari visited various countries such as Iraq, Syria, Egypt, Kuffah, Basroh, Khurasa, and others. In these countries, he studied with many leading scholars of hadiths. In his wanderings, he always collected and wrote the hadiths he heard. He spent sixteen years compiling it, and ended up with 2,602 hadith (9,082 with repetition). His criteria for acceptance into the collection were amongst the most stringent of all the scholars of Hadith [10].

III. RELATED WORKS

The current authors used several sources as a literature source from previous studies. Paper [11] described about the construction of UML model repository named Online Img2UML. Corpus repository UML models in particular are UML class diagram using XMI file as a storage place corpus. Display this repository by using a Web-based user interface.

In [12] the researchers used Corpus Models that were annotated semantically to improve the performance of the QA system. However, in their work the QA system only used three types of question, such as Who, When and Where.

Paper [13] described the collection repository Indonesian words with OAIS model of storage formats and file data storage using Fedora Commons. Display this repository using a RESTful web service interface.

This paper [14] described the making Repository music using music data sets that can be analyzed, modified, and updated. The format of the corpus storage is by using MIDI, kern, or also MUSITECH format to save it as an XML file. The display of this repository is a web-based platform.

The paper [15] presents a work in generating Weighted Vector for each Concept in Indonesian Translation of Quran (ITQ). This task is done in aiming to provide a resource needed in implementing a semantic-based question answering system (QAS) for Indonesian ITQ. It used stemming so as to increase the ability of the system to find the required documents, it showed better results than a previous study using Lucene ITQ and the study also demonstrated improved indexing. However the paper was limited to three QAS, such as person, location, and time. Further, this paper only supported translations from English.

IV. REPOSITORY DEVELOPMENT

When performing research, it is necessary to determine the restrictions of a problem such that the research can be more focused and facilitate the discussion in order that the research objectives can be achieved. This study was limited to using only one book [16]. This study was limited to the Indonesian language text. Our methods for repositories development explain in the Figure 1.

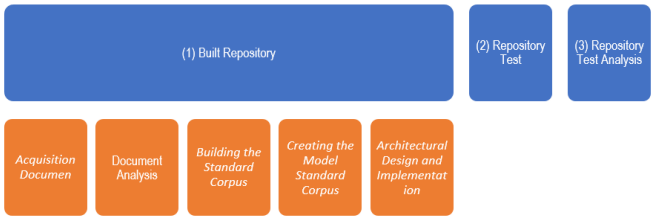

Fig. 1. Overview of Hadith Repository Development

A. Build Repository

1) Acquisition Document

The current authors built a repository based on the reference book Encyclopedia Hadith 1: Sahih Al-Bukhari 1 Almahira published in 2013.

The reason the current authors chose Hadith Al-Bukhari as a reference compared to other hadiths is because according to [16], some scholars have argued that the value of Sahih Bukhari is higher than Sahih Muslim. The reason is because of the conditions set by Imam Bukhari before recording any hadith were more stringent than the requirements established by Imam Muslim.

These requirements include, Bukhari set liqo (convergence between narrators convey the narrators

receiving). While Imam Muslim, the conditions are quite mu'assaroh (narrators convey and narrators receive enough to live in the future).

2) Document Analysis

The authors used as many as 134 hadith based on the first three books in the collection. Each of the three books contained the following number of hadith: (1) *Kitab Permulaan Wahyu* contained seven starter hadith, (2) *Kitab Iman* contained 51 hadith, and (3) *Kitab Ilmu* contained 76 hadith.

3) Building the Standard Corpus

At this stage, the authors built a standardisation model corpus based on the documents selected in the previous stage, using 134 hadith. For each hadith, the authors took five sections to be used as a standardisation model for each hadith, the name of the Book (first line), Chapter Name (second line), Number of the Hadith (third line), Hadith Content (fourth line) and Reference Number(fifth line) as shown in Figure 2.

The authors took five sections from the reference books. However, for the Reference Number there were some traditions that were not found in that part, so the authors allocated the sign (-) when modelling the standardisation corpus.

4) Creating the Model Standard Corpus

The authors converted the model standardisation corpus into a text file format, which in this case was achieved by using the Notepad application to implement the model standardisation corpus.

Implementation of the model standardisation corpus was based on the model created in the previous stage, including the five parts to each hadith. The implementation of the standardisation model can be seen in the Figure 2.

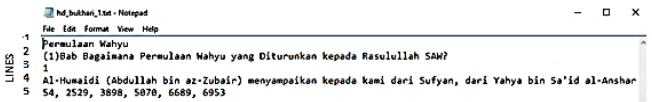

Fig. 2. Text File example: hd_bukhari_1.txt

5) Architectural Design and Implementation

Figure 3 shows the use case diagram of a proposed work which has 2 cases namely Upload File Case and Search Case. Upload File activity represent the user upload the .txt file that contain of digital form of hadith in standardize model. The file uploaded then consumed by the system to be structured and stored into .xml based file.

Fig. 3. Use Case Diagram

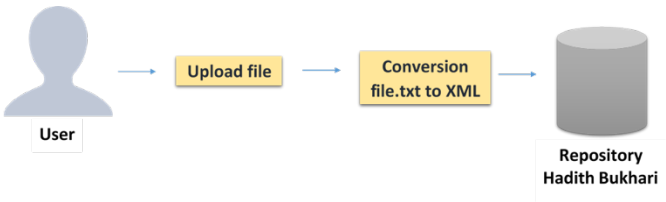

Fig. 4. Architecture Repository

The authors built a repository using the programming language PHP and apply brute force algorithm for searching. The repository architectural drawing can be seen in the process as depicted in Figure 4.

B. Repository Test

At this stage, the authors tested the repository that had been built in order to test the search results of the keywords entered by the user. The authors used two test scenarios by manually entering 20 keywords into the system, and observing the search results according to the categories of answers. Furthermore, the results of these tests were analysed as described in the next chapter.

C. Repository Analysis Test Results

At this stage, the authors observed the results of the test repository then analysed the results of the keyword search.

V. Results And Discussion

The authors examined the repository by calculating the timing of 20 manually entered keyword searches. The authors used the convenience sampling method, as the sampling technique. This was chosen with consideration of the ease of obtaining the keywords. The list of 20 keywords that were tested was as shown in Table 1.

TABLE 1
LIST OF KEYWORDS

Number	Keyword	Number	Keyword
1.	Mencari	11.	Mati
2.	Hidayah	12.	Kitab
3.	Kurma	13.	Menghukum
4.	Mewahyukan	14.	Mengumpulkan
5.	Menurunkan	15.	Mengajari
6.	Kiblat	16.	Shalat
7.	Subuh	17.	Diam
8.	Saudara	18.	Jalan
9.	Beriman	19.	Mengeluarkan
10.	Masuk	20.	Malu

A. Scenario Testing

Testing is done by entering 20 keywords into the repository and viewing the search time. Tests compare

between structured (tagging) and unstructured XML files (no tagging) using command time to compare search results.

B. Analysis of Testing Results

Based on the testing from the two test scenarios, the authors analysed the test results as follows in Figure 5:

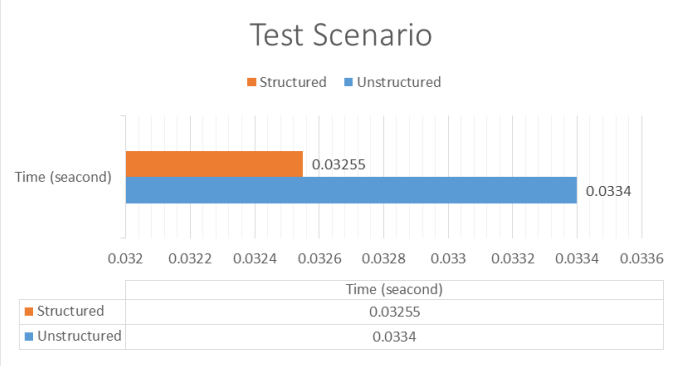

Fig. 5. Testing Result

From the testing that was completed by entering 20 keywords into the repository, it was found that the test results that use of structured XML files faster 0.00085 seconds compared with the use of unstructured XML file.

Fig. 6. Search result

Next, we test the reference ontology we develop that enables user to get the referenced hadith related to the search result as shown in Figure 6. Currently we develop the reference strictly based on the book.

VI. CONCLUSION

A repository of hadith has been built which is able to display the search results of the keywords entered by the user. From the scenario test, it shows the result that storing the hadith document by using a structured XML file displays search results faster than the use of unstructured XML files.

This is because in a structured XML file has been tagging-tagging that can help in the search process. Thus, the repository directly searches for the intended tagging. While in the unstructured XML file using only one large tagging, so the repository will perform more searches than the use of structured XML files.

For further research, the authors suggest developing a case study not only for the particular hadith of Bukhari Hadith but also to cover the entire hadith and the Qur'an in the Indonesian language and expand the linkage of reference not only based on the related hadith, but also for the scholar name that collect the Hadith.

REFERENCES

[1] P. K. Amin, *Menguji Kembali Keakuratan Metode Kritik Hadis*. Jakarta: PT Mizan Publika, 2009.

[2] M. Sagsan, "A new life cycle model for processing of knowledge management," *2nd Int. Conf. Business, Manag. Econ.*, pp. 1–9, 2006.

[3] M. Laclavik, "Ontology and agent based approach for knowledge management," Slovak Academy of Sciences, 2005.

[4] J. . Reitz, "ODLIS : Online Dictionary of Library and Information Science," 2014. .

[5] H. Froehlich, "Corpus Analysis with Antconc," *Programming Historian*, 2015. .

[6] A. Muzakir, "Pemanfaatan dan Implementasi Library XMLSEC Untuk Keamanan Data Pada XML Encryption," *Semin. Nas. Inform. 2013*, pp. 1–6, 2013.

[7] S. Hartati and E. Zuliarso, "Aplikasi Pengolah Bahasa Alami untuk Query Basisdata XML," *J. Teknol. Inf. Din.*, vol. XIII, no. 2, pp. 168–175, 2008.

[8] B. R. Suteja, "Implementasi XML Signature pada Dokumen XML untuk Transkrip Nilai Online," 2003.

[9] S. Abdullah and S. A. Manas, *Introduction to the sciences of Hadīth*, 1st ed. Kuala Lumpur: Research Centre, International Islamic University Malaysia, 2008.

[10] M. M. Khan, *Sahih Bukhari*. 1985.

[11] B. Karasneh and M. R. V. Chaudron, "Online Img2UML Repository : An Online Repository for UML Models," no. October, 2013.

[12] H. T. Sukmana, R. H. Gusminta, Y. Durachman, and A. F. Firmansyah, "Semantically annotated corpus model of Indonesian Translation of Quran: An effort in increasing question answering system performance," in *Proceedings of 2016 4th International Conference on Cyber and IT Service Management, CITSM 2016*, 2016.

[13] R. Manurung, B. Distiawan, and D. D. Putra, "Developing an Online Indonesian Corpora Repository," *PACLIC 24 Proc.*, pp. 243–249, 2010.

[14] T. Noll, J. Garbers, K. Hothker, C. Spevak, and T. Weyde, "Opuscope - Towards a Corpus-Based Music Repository," no. January, 2002.

[15] S. J. Putra, K. Hulliyah, N. Hakiem, R. P. Iswara, and A. F. Firmansyah, "Generating weighted vector for concepts in indonesian translation of Quran," in *Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services - iiWAS '16*, 2016, pp. 293–297.

[16] A. A. M. bin I. Al-Bukhari, *Buku Ensiklopedia Hadits 1; Shahih al-Bukhari 1*. Jakarta: Penerbit Almahira, 2013.