Concept Search Tool for Multilingual Hadith Corpus

Samah Mohamed Osman Hassan[1], Eric Atwell[2]

[1]Sudan University of Science and Technology, Khartoum, Sudan
[2]University of Leeds, Leeds, UK

Abstract: ***The search for the Hadith and understanding of the concepts and meaning is the important science. Moreover searching using concepts is a technique used with large data and purpose to obtain relevant results and more accurate. Accordingly the idea of building a search tool for Hadith by concepts to facilitate searches for users on the Web and make the Hadith is available in several ways. In this paper, we will design a concept search tool for Hadith. Our plan to work in three phases the first phase is select the dataset from Ahadith in Sahih Bukhari and define the concepts which will be extracted from the same book. Sahih Bukhari book is contains 100 concepts (book). Second phase generation of (HTML) files, each page will contain one Hadith text with four languages: Arabic, English, French, Russian, plus its concepts. The third phase builds a website to contain this tool.***

Keywords: Concept, Hadith, Multilingual, Parallel corpora, Tree Concept

1. Introduction

The Hadith contains many of the concepts. On the way to help Muslims to find and understand these concepts we decided to design Concept Search Tool for Hadith(CSTH), which is a tool for multilingual concepts (Arabic, English, French, Russian).The purpose of Hadith Searches to enhance precision and accuracy when searching for specific concepts as well as abstract concepts in Hadith. Besides, it is the only tool that allows users to search by concept overall hierarchy of topics or abstract concept in the Hadith, using the imported knowledge from the book of Sahih Bukhari as one of the most important specialized books in Hadith and the correct one based Muslim scholars [1] which divides the Hadith into to 100 different concept (book).

Hadith is Arabic single word (plural is Ahadith) are collections of the reports claiming to word what the prophet Mohammed(peace be upon him)said[2], the original Hadith language is Arabic and we decided to select three different languages**.** As a result plus Arabic, we take Hadith in English, French and Russian

2. Related Work

Concept tool for Quran by Noura Abbass [5] had allowed the user to search in holy Quran by two-way first one search by keyword and second by the concept, she generated a tree-view concept for Quran from 'Mushaf Al Tajweed' ontology of the Quran. She claims that her tool is more accuracy than another search tool for the Quran [4].Qurany concept search tool is a search for a semantic and syntactic function to recover syntactic information extends with semantics, thus controlling the benefits of retrieving information both syntactic and semantic is used. They show that the combined approach yields better results than the search for syntactic information [5].

3. Concept Search

Concept search is an automated search concepts information used for full-text search electronic information relevant provisions of concepts on how research organized and stored information. In other words, ideas retrieving information in response to a search query terms to relevant ideas in the query text [3].

4. Understanding the Concepts of the Hadith

The importance of the Hadith for Muslims comes in the second level after Quran. Hadith is considered the second source of religion practices and legislations for Muslims. Hadith show Muslims ways of doing everything in life [7].So Muslims must understand the concepts and meanings of the talk is a matter of great importance [7] which require us to permanent work and continuing to seek to clarify and facilitate dealing with everything related to Hadith and make it available on the electronic network in the best pictures, easy to access and May God accept our work.

5. Methodology

For the purpose of building our tool we create a website to contain the concept search tool see Figuer-1.for the dataset we are using the Multilingual Hadith Corpus [8] in this corpus we depend on Sahih Bukhari book for Hadith in this corpus the data was collected as four languages Arabic, English, French and Russian along with the related concept in four languages. For generated the concept we extracted from Sahih Bukhari which is categorized as 100 books(concept).

6. Dataset of Hadith

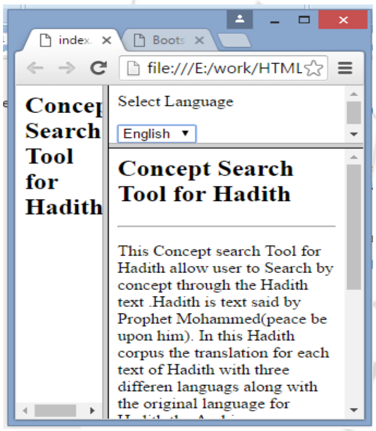
The data was in excel file each record represents one Hadith each column have each language like Hadith_Arabic, Hadith_English, Hadith_French, Hadith_Russain, concept name, concept ID and sub Concept Arabic. This data was collected to build the Multilingual Hadith Corpus [2].we are just using the data here the specification of how the data was been collected and organize by the researcher [2].

- Take the excel file imported to Structured query language (SQL) database to generate one table called Hadith Concept.

- Researcher builds a small script to take each row from the table generated the HTML page for each Hadith with the four languages along with the concept see (Figure 7).

7. The Tree of Concept Search

To build the concept tree we take all the concepts depends on the language as you see in (Table 1).First, we start with the Arabic language with the 100 concept we build the HTML tree view as in (Figure 2).The Arabic tree concept contains 100 nodes each node represent one concept. Accordingly we did the same process for the English , French and Russian so by the end we have four concept tree one tree for each language on a separate page in (Figure 3, 4, 5) show the concept tree for English , French and Russian respectively. If we want to see the Ahadith under the concept "بدءالوحي as example so will see the result as it show in (Figure 6), and if we want to see each Hadith click the link it will open a new page have the Hadith text in Arabic, English, French, Russian along with the specific concept also by the same four languages as you see in (Figure 7).



Concep Search Tool for Hadith

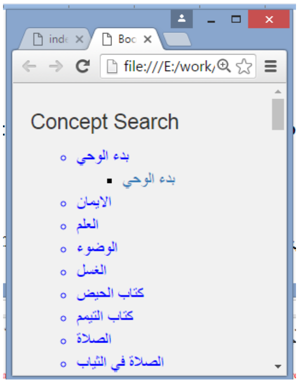
Select Language

English ▾

Concept Search Tool for Hadith

This Concept search Tool for Hadith allow user to Search by concept through the Hadith text .Hadith is text said by Prophet Mohammed(peace be upon him). In this Hadith corpus the translation for each text of Hadith with three differen languags along with the original language for

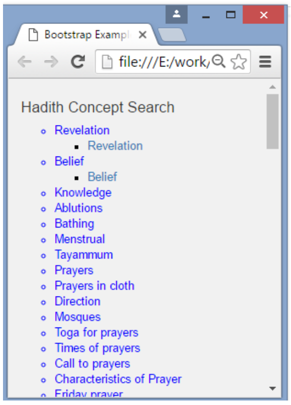
Figure 1: The search Tool interface



Concept Search

- بدء الوحي
  - بدء الوحي
- الايمان
- العلم
- الوضوء
- الغسل
- كتاب الحيض
- كتاب التيمم
- الصلاة
- الصلاة في الثياب

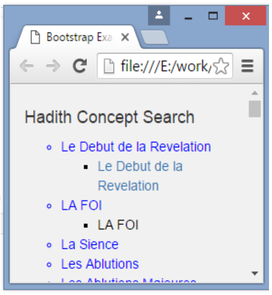
Figure 2: Snapshot of the Concept Tree for Arabic



Hadith Concept Search

- Revelation
  - Revelation
- Belief
  - Belief
- Knowledge
- Ablutions
- Bathing
- Menstrual
- Tayammum
- Prayers
- Prayers in cloth
- Direction
- Mosques
- Toga for prayers
- Times of prayers
- Call to prayers
- Characteristics of Prayer
- Friday prayer

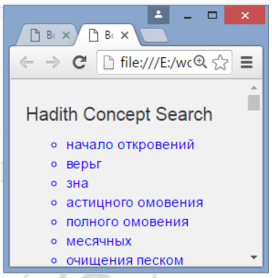
Figure 3: Snapshot of the Concept Tree for English



Hadith Concept Search

- Le Debut de la Revelation
  - Le Debut de la Revelation
- LA FOI
  - LA FOI
- La Sience
- Les Ablutions
- Les Ablutions Majeures

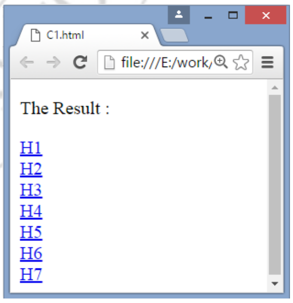
Figure 4: Snapshot of the Concept Tree for French



Hadith Concept Search

- начало откровений
- верьг
- зна
- астицного омовения
- полного омовения
- месячных
- очищения песком

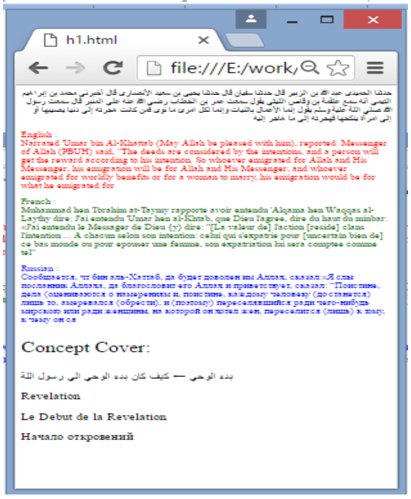
Figure 5: Snapshot of the Concept Tree for Russian



The Result :

H1
H2
H3
H4
H5
H6
H7

Figure 6: The result of the selected concept "بدءالوحي



h1.html
file:///E:/work/

حدثنا الحميدي عبد الله بن الزبير قال حدثنا سفيان قال حدثنا يحيى بن سعيد الأنصاري قال أخبرني محمد بن إبراهيم التيمي أنه سمع علقمة بن وقاص الليثي يقول سمعت عمر بن الخطاب رضي الله عنه على المنبر قال سمعت رسول الله صلى الله علية وسلم يقول إنما الأعمال بالنيات وإنما لكل امرئ ما نوى فمن كانت هجرته إلى دنيا يصيبها أو إلى امرأة ينكحها فهجرته إلى ما هاجر إليه

English :
Narrated 'Umar bin Al-Khattab (May Allah be pleased with him), reported: Messenger of Allah (PBUH) said, "The deeds are considered by the intentions, and a person will get the reward according to his intention. So whoever emigrated for Allah and His Messenger, his emigration will be for Allah and His Messenger; and whoever emigrated for worldly benefits or for a woman to marry, his emigration would be for what he emigrated for

French :
Muhammad ben Ibrahim at-Taymy rapporte avoir entendu 'Alqama ben Waqqas al-Laythy dire: J'ai entendu 'Umar ben al-Khtab, que Dieu l'agree, dire du haut du minbar: «J'ai entendu le Messager de Dieu {y} dire: "[La valeur de] l'action [reside] dans l'intention ... A chacun selon son intention: celui qui s'expatrie pour [uncertain bien de] ce bas monde ou pour epouser une femme, son expatriation lui sera comptee comme tel"

Russian :
Сообщается, чт бин аль-Хаттаб, да будет доволен им Аллах, сказал:«Я слы посланник Аллаха, да благословит его Аллах и приветствует, сказал: "Поистине, дела (оцениваются о намерениям и, поистине, каждому человеку (достанется) лишь то, вамеревался (обрести), и (поэтому) переселившийся ради чего-нибудь мирского или ради женщины, на которой он хотел жен, переселится (лишь) к тому, к чему он ся

Concept Cover:

بدء الوحي — كيف كان بدء الوحي الي رسول الله

Revelation

Le Debut de la Revelation

Начало откровений

Figure 7: Snapshot of the HTML page for each Hadith along with the concept

8. Issue About Data

our plan was to working in the same number of the concept in the four languages but we found the full concept in the Arabic language because it is the original language of Hadith and regarding the other languages we did not find translated for all the concept in the Hadith book (see Table-1) and because Hadith is very sensitive and important text we cannot do the translation must be done by the expertise in Islamic for that reason we consider only the concept which we find it in the translated book by the expertise and work on it.

This study has raised important questions about the nature of Hadith concept. First, the concept must be one or two words but for the Hadith, we notice that the concept may be one sentence, second why not all the text of Hadith not translated. Finally, in the mean time we are using the sub-concept with the Arabic language only.

9. Result

The first version of the search tool will be working on the big number of Hadith see Table-2.we can search by the concept in the four different languages Arabic, English, French and Russian. Beside the Concept Search Tool for Hadith is the only tool allow the user to search Hadith by the concept in four different languages.

Table 1: Declare the different of concept number

Language	*Number of concepts*	*Number of Hadith*
Arabic	100	7009
English	90	6769
French	57	3000
Russian	68	1569
Total	315	18, 347

10. Evaluation

While we search with Google engine there is no tool found for search Hadith by the concept, but we find some tool for keyword search so to evaluate our work with the AlMuhadith project [6] is a search tool in the large database have many Hadith books. We take the first 15 concept from Sahih Bukhari book and evaluated using them by finding the precision and the recall for each (Table 2).so you will notice the CSTH will give better precision and recall.

Table 2: The precision and Recall for AlMuhadith and CSTH

	AlMuhadith	CSTH
Precision	43.9 %	100 %
Recall	47.6 %	100 %

11. Conclusion

These findings have significant implications for the understanding of how the concept of Hadith work and the important of translating the entire concept in all the language to make the Hadith text are understandable and easy to find and search by Arabic and by all other languages.

This research will serve as a base for future studies to extend the search concept tool with more language like Chinese , Urdu, Turkey and more languages.

References

[1] Bader Alden Abu Mohammed AlAenee."Omdet Alqari shrih Albukhari", Published by: Dar Alktub Alalmea. (1421H-2001M).

[2] https://en.wikipedia.org."Hadith", March.9, 2016. [Online]. Available: https://en.wikipedia.org/wiki/Hadith.(General Internet Site)

[3] https://en.wikipedia.org/wiki/.Concept search : (27\3\2016).

[4] Abbas, N.H.A.(2009).Quran search for a concept Tool and Website. University of Leeds, UK.

[5] Giunchiglia, Fausto, Uladzimir Kharkevich, and Ilya Zaihrayeu. "Concept search: Semantics enabled syntactic search." (2008).

[6] http://www.muhaddith.org/."Search Hadith", March.5, 2016. [Online].Available: http://www.muhaddith.org/cgi-bin/a_Optns.exe?.

[7] AL Imam Majd al-Din Abu Saadat Al Mubarak bin Mohammed Al-Jazari bin Al-Athir.Alnehia in a strange word for Hadith and Al-Athr. Dar Ibn Al-Jawzia. First edition 1421AH.

[8] Samah Mohammed Osman, Eric Atwell.(2016).Compilation of an Islamic Hadith Corpus. International Conference on Computing in Arabic. Sudan, Khartoum.

[9] Van Zaanen, M., Roberts, A., & Atwell, E. S. (2004). A multilingual parallel parsed corpus as gold standard for grammatical inference evaluation. In Proceedings of LREC'04 Workshop on the Amazing Utility of Parallel and Comparable Corpora (pp. 58-61). European Language Resources Association