

# Design Requirements for Multilingual Hadith Corpus

Samah Mohamed Osman Hassan<sup>1</sup>, Dr. Eric Atwell<sup>2</sup>

<sup>1</sup>Sudan University of Science and Technology, Khartoum, Sudan

<sup>2</sup> University of Leeds, Leeds, UK

**Abstract:** *This paper discusses the requirements of users who are willing to be familiarized with Hadith. The first set of questions aimed to personal information. The remaining eight questions about Hadith started by identifying why users reading Hadith also requested them to declare their opinion regarding finding Hadith explanation and word meaning in Arabic, moreover Finding moral for each Hadith, And Hadith explanation, word meaning, and moral in one website. Muslim finding Hadith in different languages, along with the source and classification of each Hadith respectively. Finally, the users had been requested to indicate their search methods preferences whether like to read Hadith from websites or from books.*

**Keywords:** Hadith, Corpus, Multilingual, Correct Hadith, Weak Hadith, Good Hadith

## 1. Introduction

Corpus is an increasingly important area in applied linguistics and some of it focus on QURAN. Hadith for Muslims comes in the second level after Quran. Hadith is considered the second source of religion practices and legislations for Muslims. Hadith explain the full life of Muslim mention in Prophet Mohammed (Peace be upon him) words which are called (Sunnah). So all Muslims irrespective of their age, occupation, nationality, their mother tongue language and place of residence must learn about Hadith, in addition knowing the meaning of each word as well as the moral of these Hadiths.

In this paper, the intention is to investigate the need for building new Comprehensive Multilingual Hadith Corpus (MHC). Hadith is Prophet Mohammed (Peace be upon him) speeches. The finding of searching for corps specialized in Hadith was one corpus containing some of the Hadiths the KSUCCA. It is important to state that the mentioned corpus has a limited number of Hadiths, moreover, these Hadith were written in Arabic language only. The main objective is to have all Hadith in one location along with an explanation, words meaning, morals (lessons learned) and translation in different languages

## 2. Related Work

There had been a lot of work done designing of corpora ,But there is only one corpus mention clearly that contain Hadith and that is the project is done by Alrabiah for her Ph.D.[3]. project at king Saud University design of a Classical Arabic corpus (KSUCCA). She had collected around 50 Million words in her corpus for classical Arabic only. The data is generally classified into six categories, the Religion, Literature, Linguistics, Science, Biography, and Sociology. The corpus file is using UTF-8 for character encoding [3]. she said her campus contains 44 document of Hadith.

## 3. Why Collect Multilingual Hadith

People have a drawer to deal with the Hadith directly from books and this applies to its own language Arabic or uses for translations in other languages such as English, French, and Russian. On the other hand, some work had been done by groups or by individuals in the work of the electronic scanning of many modern books have been incorporated computerized and has been available on the internet in the form of (PDF) and can be downloaded and viewed, but these books are found in several different locations, and to my knowledge until today no one worked to collect this Ahadith in different languages in one place. Besides, the purpose of the collection of a large number of Hadith, including this word meanings for each word and a full explanation of the Hadith translation of talk in several languages that will allow Muslims easily accessible, used to identify the Hadith and the meanings of the words that come from the Hadith, as well as being able to find a translation in English, French and Russian. In addition, there are many words and meanings, interpretations and different associated semantics to each other making it easier for linguistics and researchers in the definition of vocabulary, the coherence of data, data analysis, concepts and phrases contained in the Hadith .The target users to use this source is the Muslims who wish to learn or studying Hadith as education. Also, the Hadith Corpus can be used by non-Muslims who would like to get to know about the great of Islam and its values, researchers and linguistics.

## 4. Method

Since the main idea is to build Hadith corpus to offer fully supported and more trusted Hadith source for the benefit of the users, it is planned to identify the requirements of Hadith users. An online survey has been prepared and conducted, which consists of two parts. The first part is relating to personal information of the users; it is about their religions, genders, age, mother tongue languages and occupations with avoidance of their private information such as name, telephone etc. .The second part is composed of nine

questions, which summarizes the main features of the intended Hadith corpus. Two versions have been created, one in the Arabic language while the other is in The English language, to allow easy collecting potential users' requirements. In the 22<sup>nd</sup> of May 2015, the survey was launched and distributed over the electronic emails, Facebook, and other social media. Within two days the number of the replies was a Figure up to 154 from different countries, nationalities, and occupations, as expected. The received information has been analyzed, discussed and thereafter many decisions has been taken as illustrated in section 5 hereunder.

## 5. Result and Discussion

Here are the results and discussions depending on the gathered data.

### 5.1. Religion

One of the most important pointers of the questionnaire findings is the Religion of participants, which Shows that (99%) are Muslim ( Figure.1). It is Obvious that potential users will most probably be Muslims. But other users with different religions will be targeted to use the intended MHC for their purposes i.e. educational and / or religious. Therefore, the MHC will be built to serve both Muslims and non-Muslims.

### 5.2. Gender

There are 154 Persons has participated in the survey, with 71(46%) females 83(54%) are males. This indicates that the MHC is appreciated by both genders. Therefore, Muslims and non-Muslims, regardless their gender, can benefit from the intended MHC as illustrated in (Figure.2).

### 5.3. Age

Age is usually used to compare different groups of people used to meet the different interests and way to get different views of different ages .In our case, this will lead to discover what the user needs to read and find about Hadith and this will give us about what to consider when we design these requirements. The survey results show that (19%) of the contributors are 15-25 years old, and (19%) are 41-60 years old while (0%) is above 60 years old. The remaining (62%) is between 26-40 years old. This indicates that the people above 60 are not included in our survey as some report separated them into two part first one those interested in this technology, but they have hesitation about the use of the fear of extreme complexity or high costs. This category needs to step on the feet of the use of the network to some support and practical help and encouragement from the social environment, second type who do not care about the Internet and find internet not useful to them[6].so that people in middle age and youth users who are already familiarized with using the new technologies will constitute the main target group for the anticipated MHC as they prefer to exploit the latest internet application. Figure.3 is showing the contribution of target users from different ages.

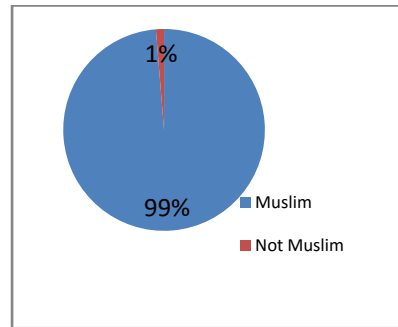


Figure 1: MHC Muslims & non-Muslims Participants

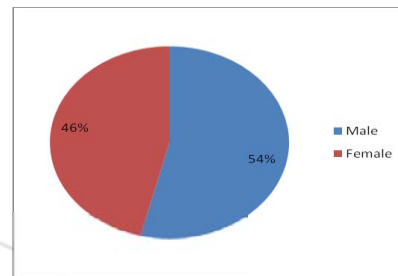


Figure 2: MHC participation of both Genders

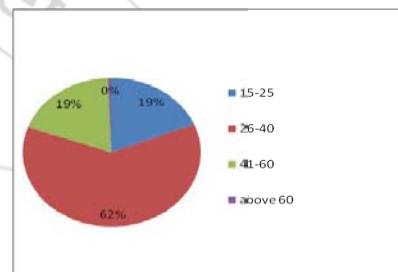


Figure 3: MHC participants of different age

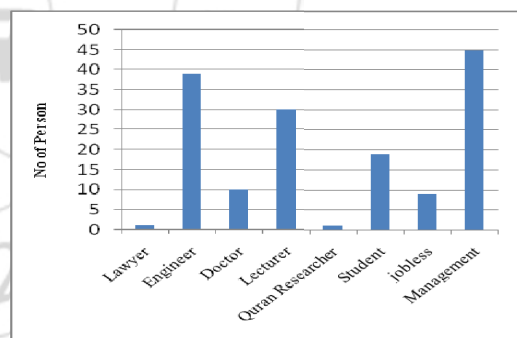


Figure 4: MHC participants of different occupations

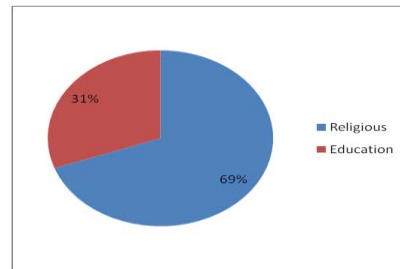
### 5.4. Occupations

Occupations are classification use in the research are to specify if their special job professional makes people have more or less interested in Hadith. The survey findings show that users from different carriers and professionals are interested in the comprehensive MHC to gain knowledge about Hadith, words meanings and the morals from each Hadith as shown in (Figure 4).

### 5.5. Religious and Educations purposes users

The MHC is projected to serve many purposes, the most important are the religious one. There is a need for an all-inclusive website that contains Prophet Mohammed (peace be upon him) Hadiths, meaning and moral of each Hadith.

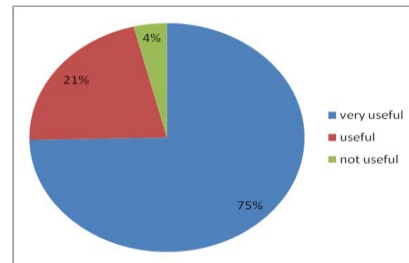
The survey's question are about why the users wish to study about Hadith in MHC shows that 69% assuring that they do it for religious reasons while 31% required to use this MHC for both motives religious as well as educational motives (Figure 5). Hadith is a very important for Muslims because the MHC will enable them to read more about prophet Mohammed (peace be upon him), about his doctrine his private life and his companions. Also Muslims effort their best to learn from Prophet Mohammed (peace be upon him) and from his Hadith to turn into good Muslims and to amplify their faith to obey Allah instructions.



**Figure 5:** MHC religious and educational motives

**5.6. Hadith Explanation in Arabic**

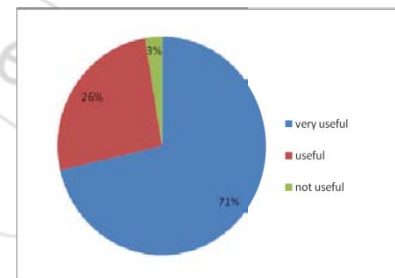
This category helps us to see if the Arabic speaker needs the explanation for each Hadith .The questionnaire's participants respond to the question regarding preferring to have in each Hadith explanation in Arabic was Figure up to (75%) believe that it is very useful and (21%) assume it is useful while only (4%) consider it is not useful (Figure.6).Therefore, from this result realized that explanation in Arabic is very important to be there in the Hadith corpus.



**Figure 6:** MHC Hadith explanation in Arabic

**5.7. Word Meaning**

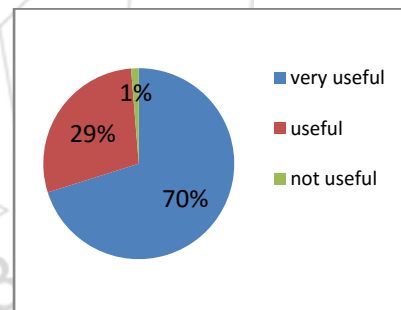
Similar percentages had been observed regarding the response of the participants to survey's question concerning their desire to have for each Hadith meaning of each word in Arabic. (71%) of them seems it is very useful and 26% believes it is useful while 3% only consider it as not useful (Figure 7).Therefore knowing the word meaning for each Hadith is useful as per the decision of (97%) participants, the thing that confirms their interest in having the word meaning of each Hadith and this is the very high percentage in which the purposed MHC expectations can depend on.



**Figure 7:** MHC responds to having the Meaning of words in Arabic

**5.8. Moral and Lesson Learn**

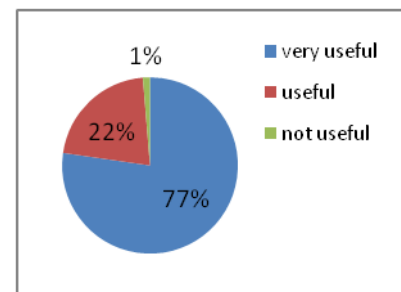
One of the most significant constructive usages of the intentional MHC is the inclusion of the moral (lesson learn) from each Hadith. (70%) of the participants consider the inclusion of the moral from each Hadith in the MHC is very useful and 29% suppose it is useful while 1% had selected the not useful answer ( Figure 8). As a conclusion, (99%) of all participants voted for very useful and useful for the inclusion of the moral of each Hadith in the MHC, which will facilitate Muslim to Figure out the right directions for their religious life and personal life as well.



**Figure 8:** MHC moral gained from Hadith

**5.9. One website**

Searching on the internet is somehow time-consuming, so in the survey one question had been dedicated to discovering the users opinion of having Hadith explanation, word meaning and moral(lesson learn) on one website. The questionnaire outcome proved the participant willing to have Hadith explanation, words meaning and moral from each Hadith in one location is very useful as (77%) of them decided that it will be very useful and 22% of them thought it well is useful i.e. (99 %)are willing to get their information from one location on the internet ( Figure.9). This is, of course, will save the user time and will be worth significance.



**Figure 9:** MHC one website feature

**5.9. One website**

Searching on the internet is somehow time-consuming, so in the survey one question had been dedicated to discovering the users opinion of having Hadith explanation, word meaning and moral(lesson learn) in one website. The

questionnaire outcome proved the participant willing to have Hadith explanation, words meaning and moral from each Hadith in one location is very useful as (77%) of them decided that it will be very useful and (22%) of them thought it well is useful i.e. (99 %) are willing to get their information from one location on the internet (Figure 9). This is, of course, will save the user time and will be worth significance.

### 5.10. Hadith in different languages

The original language for Hadith is Arabic, However for Muslims learning of Arabic language will be added value. To measure the importance of Hadith translations in different languages in the survey participants had been requested to answer the question if they find Hadith in different translation is very useful, useful or not useful ?. The answer to the question had been showed that (65%) of the participant consider it as very useful, (25%) of them think it is useful and 10% of the participants believe that the matter is not useful( Figure 10). Therefore, users think that if they find Hadith in their original language that will allow them to understand Hadith very well, moreover this will make the Hadith statements so clear for them, normally when if someone wants to study something it better to study it in his mother language for better understanding and implementation of gained knowledge.

### 5.11. Hadith source

The most recognized Hadith sources are Sahih Al-Bukhari, Sahih Muslim, Riyad Al-Saliheen and many other books. The questionnaire outcomes confirmed that 64% of the participants prefer and consider the information about from which source the Hadiths are copied in the intended MHC is very useful, 33% believe it is useful and 3% declare that is not useful ( Figure.11).

### 5.12. Hadith Classification

Reading Hadith is very important for Muslim life so when he/she find any Hadith they prefer to check if this Hadith is strong(correct Hadith(Sahih) or weak Hadith(Daeef) or good Hadith (Hassan).In the survey, the participants respond to the question regarding their willingness to have the classification of each Hadith include in the intended MHC e.g. correct(Sahih) or weak(Daeef) or good (Hassan).The final responds to this question were (86%) of participants agreed that is worthy and very useful and (13%) of them decided that is useful and the remaining (1%) are not agreed and declared that is not useful ( Figure 12).

### 5.13. Hadith Search

Regarding the searching facilities in the intended MHC, the survey inquired the prospective users whether they prefer to search about Hadith from books, from one website or from different websites. The perspective users had responded to this inquiry as follow, (33%) of the participants prefer to have searching facilities for all information within one MHC,(22%) of them like better to obtain the required information from different websites ,(23% ) would like to depend on books only for collecting the desired information

,(22%) of them think it is worthy to look for the information on the websites as well as in the books ( Figure 13). From the above it is obviously that users divided into two major groups, the first group prefer to utilize the new technologies, such as websites over the internet and the second is traditional users who used the hard copies books as have lag of trusting in website information, whoever by developing authenticated and authorized MHC to serve the user's needs in searching for the required information and data, then the users from the second group can be converted to a backup target group for the intended the MHC .

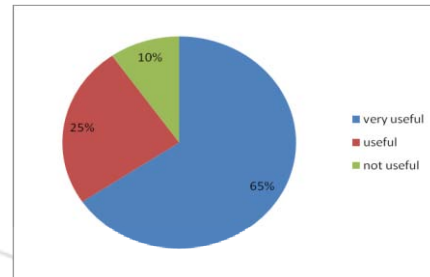


Figure10: MHC Hadith in different languages

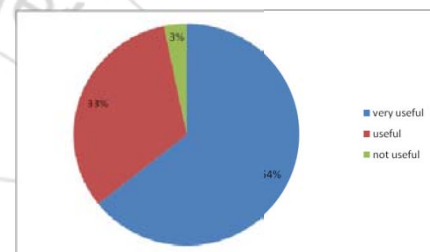


Figure11: MHC Hadith Sources

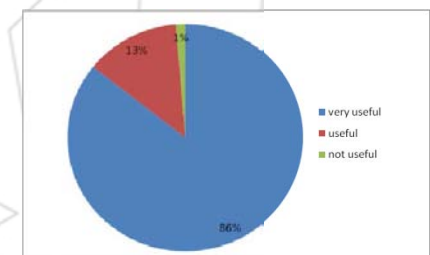


Figure12: MHC Hadith Classification

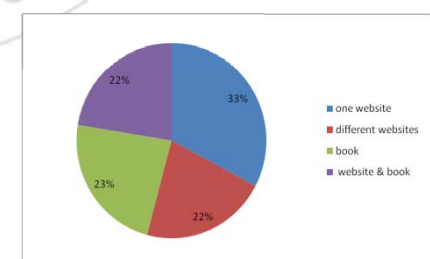


Figure13: MHC Searching Facilities

## 6. Conclusion

From the above the findings is obviously that the users are looking for trusted MHC, which will satisfy their all needs in searching for Hadith along with the each word explained in Arabic besides the moral of each Hadith, moreover if the MHC will offer the sources of the Hadith as well as the classification of each Hadith this is will added value for the users of the MHC, so building of the intended MHC will be

significance and worthy project and the expectation is to collect 6,000 Hadith in Arabic along with the explanation of around 70 words so the total of Arabic words will Figure up 4,200,00 words, and 6,000 Hadith in English which consist of 30 words of each Hadith i.e. The total of the English words will reach 1,800,00 words .The same calculation will be applicable for French and Russian, and thus, the total words will be 1,500,000 words for the four languages Arabic, English, French, and Russian. The future MHC have great expectations to become a wonderful useful source for the most Muslim all around the world as well as for other researchers from another religious background who are seeking for information regarding Hadith in different languages.

## References

- [1] L.Al-Sulaiti and E.Atwell, “ The design corpus of contemporary Arabic ”.International journal of corpus Linguistics,vol.11,pp.135-171,2006.
- [2] Kais Dukes, Eric Atwell, and Abdul-Baqueem.Shareef. “Syntactic annotation guidelines for the Quranic Arabic treebank”. in proceedings of the language Resources and Evaluation Conference (LREC 2010).Valletta, Malta.
- [3] M.Alrabiah, A.Al-Salman and E.Atwell. “The design and construction of the 50 Million Words KSUCCA King Saud University-Corpus of classical Arabic”, in the second workshop on Arabic Corpus linguistics(WACL-2), Monday 22d July, Lancaster University, UK,2013.
- [4] Imam Al-Nawawi. “40 Hadith Al-Nawawi”, published by Islamic University of Medina third edition,1984.
- [5] Alansary, Sameh, Magdy Nagi, and Noha Adly. "Building an international corpus of Arabic (ICA): Progress of compilation stage." *7th international conference on language engineering, Cairo, Egypt.* 2007.
- [6] <http://www.swissinfo.ch/>." is it not using the internet for eldest people made them in social isolation",july.09,2015.[online].Available: <http://www.swissinfo.ch/ara/>.[Accessed:April.3,2016].(General Internet Site)