ISSN 2349-7831

International Journal of Recent Research in Social Sciences and Humanities (IJRRSSH)
Vol. 3, Issue 2, pp: (100-104), Month: April 2016 - June 2016, Available at: www.paperpublications.org

Design and Implementing Of Multilingual Hadith Corpus

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

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

Abstract: **In this paper, we want to establish the first design of Multilingual Hadith Corpus. The Hadith original language is Arabic and we decide to select English, French and Russian as extra languages for Hadith translation. Design the Hadith corpus will be in four steps, the first step is data collection, which will be from the internet because it is considered as the biggest corpora, second step cleaning the data, step three file generation and the last step is file annotation using XML.**

Keywords: **Multilingual, parallel corpora, Ahadith, corpus, Hadith.**

1. INTRODUCTION

There is persuasive evidence that the Hadith plays a crucial role in regulating Muslims life .A lot of work has been done in the creation of Arabic corpora like the big one created by Alrabia [2], but we notice that there is no special corpus for Hadith, on the other hand, we have more than one corpus special for Quran[4] ,include Quran text [2] and one created for specific purpose like what Alfife did in his corpus[1].Hadith is the single word the plural is (Ahadith)[6], are words of prophet Mohammed(Peace be upon him). The Hadith is considered as the second important book after the Quran for Muslims. In the Islamic Rules (Shree Al-Islamia), the Hadith is considered the second source of religious knowledge for Muslims besides that Hadith contain teachings on all areas of Muslims life mentioned by prophet Mohammed(Peace be upon him) words.The Hadith guides Muslims how to be good Muslims through the explanation of prophet Mohammed(Peace be upon him) to every necessary thing in Muslims life: how to eat, how to drink ,how to sleep, how to deal with other people, how to pray, how to obey Allah and how to live a simple and clear life , in other words prophet Mohammad explained every matter in Muslims life. Therefore, a multilingual Hadith corpus would be useful for Muslims all around the world as it will allow them to know what each word is, what it means, and what it teach us about our religion.

2. IMPORTANT OF HADITH

Hadith are mainly used and read in printed form, from printed books; this applies to original Arabic Hadith, and mores to translations into English, French and Russian. Some Hadith books (or parts of books) have been scanned and put online; but these samples are on different websites, and no-one has collected together a set of Hadith in several languages in parallel corpora.

3. PARALLEL CORPORA

Parallel corpora are the corpus contains text from more than one language. In our MHC we had four version of text for each Hadith, started by Arabic (the original Hadith language), English, French and Russian.

ISSN 2349-7831

International Journal of Recent Research in Social Sciences and Humanities (IJRRSSH)
Vol. 3, Issue 2, pp: (100-104), Month: April 2016 - June 2016, Available at: www.paperpublications.org

4. MULTILINGUAL HADITH CORPUS

We want to develop MHC that will help Muslims reading the Hadith understand the meaning and moral that comes from each Hadith .Our MHC contain Hadith texts including their explanations, the semantics or meanings for each word and the full semantics or explanation of each Hadith verse. Building this corpus will allow Muslims who access it to use it to learn about the Hadith and the meanings of the words which come from the Hadith, as well as being able to find translations for it in English, French, and Russian. This MHC could be used by Muslims who want to learn or teach the Hadith and other people who want to know about Islam.

5. METHODOLOGY

Our plan is to create the MHC in four steps, which we will describe further now:

A. Data Collection:

1. Searching the internet consider as the biggest corpora, which we can derive our texts.
2. Selecting and organizing the texts, which will include written texts in multilingual Arabic, English, French and Russian. Text collection was done manually because the text was coming in different format like (.doc,.pdf).
3. The process for the (.doc) file was done smoothly copy and paste, on the other hand, the (.pdf) file it had to be converted first to editable text then we can copy it for that job we used Nitro Pro 10 converter. This software works perfectly with the English text but for Arabic and French, we have to run manually through the text for correction and make the text clean.

B. Data Cleaning:

The cleaning mean remove the irrelevant data and disproportionate, as well as the data that we do not want Cleaning the data is considered as time-consuming job because it takes too much time and effort. There are two ways to do that ;

1. Manually to remove the unwanted word or character, or to correct some wrong words ,or to replace one word by another word.

2. Automatically: by using find and replace tool available by Excel to do.

- Replacement of the wrong word by the right one
- Replacement of character by another character

C. File generation:

Copying a text from the Internet and pasting it into Microsoft Word, every text was encoded choosing Unicode UTF-8.Finally, we had to clean data in the plain text saved into four separated files:Arabic.txt see in (Figure.1), English.txt see in (Figure.2),French.txt see in (Figure.3)and Russian.txt see in (Figure.4), and the words in each file will be counted.

حدثنا الحميدي عبد اللة بن الزبير قال حدثنا سفيان قال حدثنا يحي بن سعيد الانصاري قال اخبرني محمد بن ابراهيم التيمي انة سمع علقمة بن وقاص الليثي يقول سمعت عمر بن الخطاب رضي اللة عنة قال سمعت رسول اللة صلي اللة علية وسلم يقول انما الاعمال بالنيات وانما لكل امري ما نوي فمن كانت هجرتة الي دنيا يصيبها او الي امراة ينكحها فهجرتة الي ما هاجر الية

Figure.1: Plain text for Arabic

On the authority of Omar bin Al-Khattab, who said I heared the messenger of Allah salla Allah u alihi wa sallam say : (Actions are but by intention and every man shall have but that which he intended. Thus he whose migration was for Allah and His messenger, his migration was for Allah and His messenger, and he whose migration was to achieve some worldly benefit or to take some woman in marriage, his migration was for that for which he migrated.) related by Bukhari and Muslim

Figure.2: Plain text for English

ISSN 2349-7831

International Journal of Recent Research in Social Sciences and Humanities (IJRRSSH)
Vol. 3, Issue 2, pp: (100-104), Month: April 2016 - June 2016, Available at: www.paperpublications.org

Le Commandeurdes Croyants, Aboû Hafç Omar ben El-Kattâb (que Dieu soit satisfait de lui) a dit: J`ai entendu l` Envoyé de Dieu, salla Allah u alihi wa sallam, (à lui, bénédiction et salut) dire: «Les actions ne valent que par leurs intentions "Leurs Niyates: « Chacun ne recevra la récompense qu`il mérite que selon ce qu`il a entendu faire.

Figure.3: Plain text for French

По свидетельству Эмира Правоверных , (Титул халифов.) Абу Хафса Умара ибн ал-Хаттаба (Второй халиф Ислама.) (да будет Аллах милостив к нему), который сказал: Я слышал, как Посланник Аллаха (да благословит его Аллах и да ниспошлет ему мир) сказал: Все действия - преднамеренны, и каждому воздается по его намерениям. Поэтому, у того, чье переселение (ЗИмеется в виду религиозное переселение, в особенности из Мекки в Медину.)было ради Аллаха и Посланника Его, переселение будет ради Аллаха и Посланника Его,а у того, чье переселение было ради мирской выгоды или ради женитьбы, переселение будет ради того, ради чего он переселялся.

Приводится двумя Имамами знатоков хадисов, Абу Абдуллахом Мухаммадом ибн Исмаилом ибн Ибрахимом ибн ал-Мугира ибн Бардизбахом ал-Бухари и Абу л-Хусаином Муслимом ибн ал-Хаджджаджем ибн Муслимом ал-Кушаири ан-Наисабури, в их двух Сахихах, которые являются самыми надежными сборниками [хадисов]

Figure.4: Plain text for Russian

D. Data Annotation:

To create the XML file for each Hadith we need to map each Arabic Hadith to the translation in the English, French and Russian. This work had done manually to recheck that each Arabic Hadith match the same translation in the other languages because Hadith is very accurate text .we have two volunteers for the French and Russian check text.

Finally, after we had one excel file have all the mapping Hadith correctly. We notice from (Table-2) that we did not have the same number of Ahadith from all the Language for that reason we will consider only the minimum number of Ahadith which is English with 652 Hadith and we mapped it with all the 3 languages.

Generate XML schema see(Table-1), for the data and apply it to the data in the excel file for each raw ,each raw will have the 4 text of Hadith from the different languages plus some information about each Hadith like Hadith number, book name, book number see (Table-2),then save the file as (H_1_E_A_R_F.xml) indicate H for Hadith,1 was the Hadith number, E for English, A for Arabic text, F for French text and R for Russian text see (Figure.5). Consequently, we include 652 XML files.

6. RESULT

The result of this work is that we have MHC with around 2 Million words of Hadith from the four languages .Besides we have each hadith along with the four languages translation in one XML file as you see in Appendix-1.All the Ahadith from SAHAI ALBUKHARI only.Table-3 shows that number of Hadith differs and that is because not all the Ahadith from ALBUKHARI are translated into the other language, people can select some texts of Ahadith and translated them. So we fail to find the translation for the entire ALBUKHARI book online.

ISSN 2349-7831

International Journal of Recent Research in Social Sciences and Humanities (IJRRSSH)
Vol. 3, Issue 2, pp: (100-104), Month: April 2016 - June 2016, Available at: www.paperpublications.org

```
<?XML version="1.0" encoding="UTF-8"?>

<!--XML database--> <Data>

<Hadith_Source> الحديث الاول في الاربعين النووية</Hadith_Source>

<Notice>الطبعة الثالثة</Notice>

<Hadith_Arabic> عَنْ أَمِيرِ الْمُؤْمِنِينَ أَبِي حَفْصٍ عُمَرَ بْنِ الْخَطَّابِ قَالَ: سَمِعْت رَسُولَ اللَّهِ يَقُولُ: " إنَّمَا الْأَعْمَالُ بِالنِّيَّاتِ، وَإِنَّمَا لِكُلِّ
امْرِئٍ مَا نَوَى، فَمَنْ كَانَتْ هِجْرَتُهُ إلَى اللَّهِ وَرَسُولِهِ فَهِجْرَتُهُ إلَى اللَّهِ وَرَسُولِهِ، وَمَنْ كَانَتْ هِجْرَتُهُ لِدُنْيَا يُصِيبُهَا أَوْ امْرَأَةٍ يَنْكِحُهَا فَهِجْرَتُهُ إلَى
مَا هَاجَرَ إلَيْهِ

</Hadith_Arabic> <Hadith_Explian> انما الاعمال بالنيات اي صحة مايقع من المكلف من قول او فعل او كمالة وترتب
الثواب علية لايكون الاحسب ماينوية والنيات هي جمع نية وهي القصد وعزم القلب علي امر من الامور وهجرتة في اللغة الخروج من
ارض الي ارض ومفارقة الوطن والاهل مشتقة من الهجر وهو ضد الوصل وشرعا هي مفارقة دار الكفر الي دار الاسلام والمراد بها هنا
الخروج من مكة الي غيرها الي مدينة رسول اللة صلي اللة علية وسلم ويصيبها اي يحصلها وينكحها اي يتزوجها وفهجرتة الي ماهاجر
الية اي جزاء عملة الغرض الدنيوي الذي قصدة ان حصلة والافلا شئ له</Hadith_Explian><Hadith_Narrated_by> رواة
مسلم والبخاري</Hadith_Narrated_by> <Hadith_English>Narrated 'Umar bin Al-Khattab: I heard
Allah's Apostle saying, "The reward of deeds depends upon the intentions and every person will get
the reward according to what he has intended. So whoever emigrated for worldly benefits or for a
woman to marry, his emigration was for what he emigrated for
</Hadith_English><Hadith_French>Le Commandeur des Croyants, Aboû Hafç Omar ben El-
Kattâb (Que Dieu so it satisfied de Lui) a dit: J`ai extends l` Envoyé de Dieu, Salla Allah u alike was
slam , (à Lui, bénédiction et salut) dire: « Les actions ne valent que par leurs intentions ". Leurs
Niyates: « Chacun ne recevra la récompense </Hadith_French><Hadith_Russian>Сообщается, что
‘Умар бин аль-Хаттаб, да будет доволен им Аллах, сказал: - Я слышал, как посланник
Аллаха, , ска- зал: «Поистине, дела (оцениваются) только по намерениям и, поистине,
каждому человеку (достанется) лишь то, что он намеревался (об- рести), и (поэтому)
переселявшийся ради чего- нибудь мирского или ради </Hadith_Russian></Data>
```

7. CONCLUSION

The main goal of the current study is to design MHC. Thus, the building of the intended MHC will be a significant and worthy project. The proposed MHC has the potential to become a wonderfully useful source for the most Muslims around the world, as well as for other researchers from other religious backgrounds who are seeking information regarding the Hadith in different languages.

For the future work the researcher recommends the following the MHC can be extended to have Hadith in other languages, Hadith Explanation in other languages, and Hadith classification can be done.

REFERENCES

[1] Alfaifi, A.Y.G., and E. S. Atwell. "Arabic learner corpus v1: A new resource for Arabic language research." (2013).

[2] 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.

[3] 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.

[4] 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.

[5] Robert W.Sebesta. (2011). "*Programming The World Wide six edition*."Publication Addison-Wesley. Boston.

[6] https://en.wikipedia.org/wiki/."Hadith".Feb.17,2016[online].Available:https://en.wikipedia.org/wiki/.[Accessed :April.7,2016).(General Internet Site)