iJET EXPLORE International Journal of Engineering & Technology
CENTRE OF PROFESSIONAL RESEARCH PUBLICATIONS
ISSN 2049-3444
International Journal of Engineering and Technology Volume 6 No.2, February, 2016

Processing Of "Hadith Isnad" Based On Hidden Markov Model

Moath Mustafa Ahmad Najeeb
Assistant Professor, Computer Science Department, College of computer at Al-Qunfudah, Umm Al-Qura University, KSA

ABSTRACT

Hadith is the sayings and actions of Prophet Mohammad (Peace Be upon Him), it is the second legalization source in Islam after the Holy Quran. Hadith consists of two parts: Isnad and Matn, Isnad is the sequence chain of narrators who narrate Hadith while Matn is the words or actions of the prophet. In this paper, we introduce a novel approach that recognizes the "Parts of Isnad" based on Hidden Markov Model (HMM), this approach contains three phases: preparation, training and testing. Our approach classifies the words and phrases of Isnad to its main categories such as: narrator name, prefix of narrator name, received method, prefix of received method, title, replacement and prophet name.

Keywords: *Hadith, Hidden Markov Model, Isnad, Narrator, Sahih Muslim, Natural Language Processing.*

INTRODUCTION

Hadith is the sayings and actions of Prophet Mohammad (Peace Be Upon Him), it consists of two parts: Matn and Isnad, Matn is the narration or the words of the prophet, while Isnad is the chain or the sequence of narrators who narrate Matn. Muslims scholars give a great attention to Hadiths because it considered as the second legalization source after the Holy Quran, they wrote hundreds of books to serve Hadith science, and they initiated many specialist sciences under Hadith science umbrella such as "Jarh & Ta'deel" and "Mustalah Al-Hadith" sciences [1]. Many books collect and classify Hadiths, the most two important and authentic books are "Sahih Al-Bukhari" and "Sahih Muslim", figure 1 shows the main parts of Hadith.

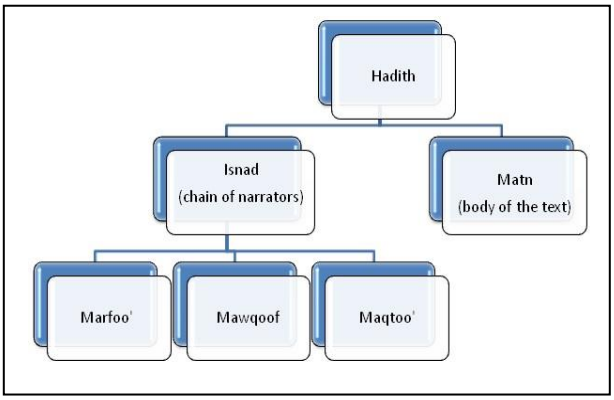

Figure 1: Main parts of Hadith

Hadith and the holy Quran are written in the Arabic language, so we had establish the Arabic Natural Language Processing (ANLP) laboratory in computer college at al Qunfudah, at Umm Al-Qura university in KSA [2], this laboratory aims to serve the Islamic sciences such as Quran science and Hadith science through developing the ANLP resources, algorithms and applications. Figure 2 shows the main components of the ANLP model.

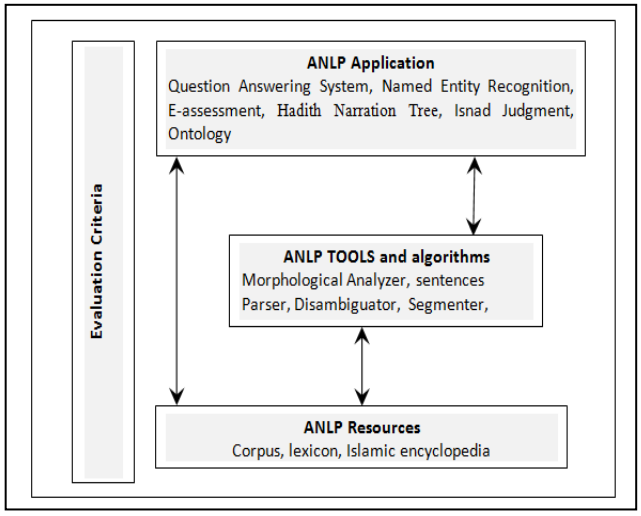

Figure 2: ANLP model

One of the most important techniques that can be used in natural language processing is Hidden Markov Model (HMM), HMM is a finite state machine that generate an observations with transmission probabilities at time (t), which depends on previous state at time (t-1) [3], so in HMM the states are not known (hidden), the transmission between states are depends on sets of probabilities called transmission probabilities, these transmissions could produce an observations (sometimes called results or outcomes) which depends on the probability distribution that relate on this state.

HMM is a stochastic model, so it can do a statistical classification, therefore it is used in voice and handwriting recognition, HMM has a good ability in adaptation and handling the serial signals. HMM has also a wide usage in Bioinformatics research to find the DNA sequences. HMM has an important role in designing the detection systems of networks intrusion, these systems concern with prevent the attacks and data thefts from hackers [4]. HMM elements can be defined as follows:

N: It is the number of states in the model.
S: It is the space of the states
where $S= \{S_1, S_2 \dots S_N\}$
q_t: is the state at time (t)
M: It is the number of observations
V: It is the space of the observations
where $V= \{V_1, V_2 \dots V_M\}$
A: is the transition probability.
B: is the output probability.
π : is the start state probability

HMM use the values of (N, M, A, B, π) to generate the sequence or series of the observations O

$$O = O_1 O_2 \dots O_T$$

So, HMM model can be defined formally as:

$\lambda = (\pi, A, B)$
where:
π is the state probability.
A is the matrix of transition probabilities of states.
B is the matrix of observations probabilities at state i.

HMM used to handle many problems such as computing probability of certain observations sequence, this task accomplishes by sum all the probabilities of state sequences, HMM also handle the filtering problem in an efficient way, this task compute the probability of the last state if the model shows the sequence of the observations, this task can be solved by forward algorithm. The forward-backward algorithm is also used to compute the probabilities of a state located in the middle of the sequence. Moreover HMM used in to expect the most or the maximum probability sequence of states that lead to certain sequence of observations, this kind of problems can be solved by using the Viterbi algorithm, Part-of-Speech is a clear example in this situation.

In this paper we propose a new algorithm to detect the Part-of-Isnad (not Part-of-Speech) which classifies the main phrases and words of Isnad to categories like: narrator name, prefix of narrator name, received method, prefix of received method, title, replacement, prophet name.

The remainder of this work is prepared as follows: The second section explains the main existing researches and application for Hadith sciences processing. The third section describes the proposed system in details. Experiments and discussions will be in section four. The last section will conclude our work.

Related Works

Many efforts initiated in the literature that aims to serve Hadith science, Moath M. Najeeb [1] introduce a new multi agent system for Hadith processing composed of five agents which are: preprocessing agent, lexical agent, morphological agent, replacing agent and syntactical agent.

In [5], Moath tried to present an innovative system for Hadith Isnad processing based on an Artificial Intelligence technique called Associative Classification (AC), the new system help in judgment of Hadiths and distinguish between the accepted and rejected ones.

"Muhadith" is a cloud system for Hadith classification it is an expert system that tries to simulate the scholar of Hadith [6]. Harrag et al. [7] design a corpus for Hadiths based on a semi-structure format for "Sahih Al-Bukhari", this format divides each Hadith to nine sections, then they built a finite state transducer-based entity extractor to serve as an information retrieval system for Hadiths.

Some Islamic websites [8-13] introduce Hadith applications that do many tasks such as electronic books and searching techniques, some of these websites are listed below:

- Shamela library http://www.shamela.ws
- Dorar http://www.dorar.net
- Islamweb http://www.islamweb.net
- Harf www.harf.com
- Turath www.turath.com

"Rewaia" encyclopedia [8] is a well-known work in "Hadith science"; it contains more than fifty thousand Hadiths, it also provides many tools for analyzing and indexing these Hadiths.

"Golden Encyclopedia" is important and famous encyclopedia in "Hadith science" [9]. This encyclopedia contains the nine "Hadith books" which are: Sahih Al-Bukhari, Sahih Muslim, Sunan Abu Dawod, Sunan Al-Termidhi, Sunan Ibn Majah, Sunan Al-Darami, Musnad Ahmad Bin Hanbal, Muwatta' Malik and Sunan Al-Nasa'i, this encyclopedia contains more than six thousand Hadiths.

"Tarajim" encyclopedia is another well-known encyclopedia that serves Hadiths it contains the biographies of more than 150000 narrators of Hadiths, these biographies include the name of the narrator, his date of birth, his place of birth, his "Konia", his place of death, his teachers, his students his date of death, etc., [10].

Other efforts to serve Hadith science can be found in [14-17].

The Proposed Approach

Our approach consists of three phases: Preparation, Training and Testing.

Phase 1: Preparation
We collect our data form "Sahih Muslim" book, we separate the Matn of Hadith from the Isnad of Hadith, and then we consider the Isnad as a raw text and store it in the input file. After that we apply the "Preparation" procedure on this file as illustrated in the next steps:

Preparation procedure:
Procedure Input: Input file
Procedure Output: Hadith-Annotated text file

START
Step1: Divide Hadith to Matn & Isnad
Step2: Remove Matn
Step3: Divide Isnad to the Isnad-Parts or segments by scholar of Hadith (expert person)
Step4: our "Hadith-Annotated" text file is ready as input to the next phase.
END

Phase 2: Training
Training phase consists of four procedures as illustrated in the next steps:

Training procedure:
Procedure Input: Hadith-Annotated text file
Procedure Output: Parameters of HMM

START
Step1: call Determine-states procedure
Step2: call Fined-start-probabilities(π) procedure
Step3: call Fined-Transitions-probabilities(A) procedure
Step4: call Fined-Observations-probabilities(B) procedure
END

To determine the states, we call the following procedure:

Determine-states procedure:
Procedure Input: Hadith-Annotated text file
Procedure Output: States Array

START
For each Isnad-Part in Hadith-Annotated file
If the Isnad-Part exist in the array
Do nothing Else
Insert Isnad-Part into array
END

Now, we will describe the procedure that responsible for finding the start probability (π), this parameter represents the probability to start the Isnad with a certain Isnad-Part.

Fined-start-probabilities(π) procedure:
Procedure Input: Hadith-Annotated text file
Procedure Output: Start-Probabilities array

START
For each Isnad-Part in Hadith-Annotated file
Calculate how many times occurs at the beginning of the Isnad
Calculate start-probability array
END

The equation of start-probability (π) is as follows:

$$start - probability\ (\pi) = \frac{(Number\ of\ Hadith\ Isnad\ start\ with\ a\ certain\ Isnad - Part)}{(Total\ number\ of\ Hadith\ Isnad)}$$

The following procedure describes the Fined-Transitions-probabilities (A), which calculate the probability of sequence S_i S_j, i.e. S_j comes after S_i.

Fined-Transitions-probabilities(A) procedure:
Procedure Input: Hadith-Annotated text file
Procedure Output: Transition-Probability matrix

START

For each Isnad-Part in state Si
For each Isnad-Part in state S_j
Calculate number of times that Sj comes after Si

$$A = \frac{(\text{Total number of sequence Si Sj })}{(\text{Total number of Si})}$$

END

The following procedure describes the Fined-Observations-probabilities(B), which calculate the probability of assigning certain Isnad-Part to a phrase or word in the input file.

Fined- Observations-probabilities(B) procedure:
Procedure Input: Hadith-Annotated text file
Procedure Output: Observations-Probability matrix

START
For each word/phrases in Hadith-Annotated text file
X= Number of times that this word/phrase tied with a certain Isnad-Part
Y= Total number of this Isnad-Part
$$Observations - Probability = \frac{(X)}{(Y)}$$
END

Phase 3: Testing
We apply Viterbi algorithm using the previous parameters to Hadiths from "Sahih Muslim" book to find the Isnad-Parts of these Hadiths**.**

Experiments and Discussions
Let us check our system on the following Hadiths, table 1 shows 3 Hadiths from "Sahih Muslim" book:

Table 1: Three Hadiths from "Sahih Muslim" book

Hadith Number	Hadith
1	حَدَّثَنَا يَحْيَى بْنُ أَيُّوبَ، وَقُتَيْبَةُ بْنُ سَعِيدٍ، وَعَلِيُّ بْنُ حُجْرٍ، جَمِيعًا عَنْ إِسْمَاعِيلَ بْنِ جَعْفَرٍ، قَالَ ابْنُ أَيُّوبَ: حَدَّثَنَا إِسْمَاعِيلُ، قَالَ: أَخْبَرَنِي الْعَلَاءُ، عَنْ أَبِيهِ، عَنْ أَبِي هُرَيْرَةَ، أَنَّ رَسُولَ اللهِ صلَّى الله عَلَيْهِ وَسَلَّمَ قَالَ: «لَا يَدْخُلُ الْجَنَّةَ مَنْ لَا يَأْمَنُ جَارُهُ بَوَائِقَهُ»
2	حَدَّثَنَا يَحْيَى بْنُ يَحْيَى، قَالَ: قَرَأْتُ عَلَى مَالِكٍ، عَنْ نَافِعٍ، عَنِ ابْنِ عُمَرَ، أَنَّ حَفْصَةَ أُمَّ الْمُؤْمِنِينَ، أَخْبَرَتْهُ أَنَّ رَسُولَ اللهِ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ، «كَانَ إِذَا سَكَتَ الْمُؤَذِّنُ مِنَ الْأَذَانِ لِصَلَاةِ الصُّبْحِ، وَبَدَا الصُّبْحُ، رَكَعَ رَكْعَتَيْنِ خَفِيفَتَيْنِ قَبْلَ أَنْ تُقَامَ الصَّلَاةُ»
3	عَنْ أَبِي إِسْحَاقَ، عَنْ عَبْدِ اللهِ بْنِ مَعْقِلٍ، عَنْ عَدِيِّ بْنِ حَاتِمٍ، قَالَ: سَمِعْتُ النَّبِيَّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ يَقُولُ: «مَنِ اسْتَطَاعَ مِنْكُمْ أَنْ يَسْتَتِرَ مِنَ النَّارِ وَلَوْ بِشِقِّ تَمْرَةٍ، فَلْيَفْعَلْ»

Tables 2, 3 and 4 are representing the Hadith-Annotated text files for Hadiths 1, 2 and 3 respectively.

Table 2: Isnad-Parts for Hadith Number 1

Word/Phrase	Isnad-Part
حَدَّثَنَا	Received method
يَحْيَى بْنُ أَيُّوبَ	Narrator name
وَقُتَيْبَةُ بْنُ سَعِيدٍ	Narrator name
وَعَلِيُّ بْنُ حُجْرٍ	Narrator name
جَمِيعًا	Other
عَنْ	Received method
إِسْمَاعِيلَ بْنِ جَعْفَرٍ	Narrator name
قَالَ	Received method
ابْنُ	Prefix of narrator name
أَيُّوبَ	Narrator name
حَدَّثَنَا	Received method
إِسْمَاعِيلُ	Narrator name
قَالَ	Prefix of received method
أَخْبَرَنِي	Received method
الْعَلَاءُ	Narrator name
عَنْ	Received method
أَبِيهِ	Replacement
عَنْ	Received method
أَبِي	Prefix of narrator name
هُرَيْرَةَ	Narrator name
أَنَّ	Other
رَسُولَ اللهِ صلَّى الله عَلَيْهِ وَسَلَّمَ	Prophet name
قَالَ	Other

Table 3: Isnad-Parts for Hadith Number 2

Word/Phrase	Isnad-Part
حَدَّثَنَا	Received method
يَحْيَى بْنُ يَحْيَى	Narrator name
قَالَ	Prefix of received method
قَرَأْتُ عَلَى	Received method
مَالِكٍ	Narrator name
عَنْ	Received method
نَافِعٍ	Narrator name
عَنِ	Received method
ابْنِ	Prefix of narrator name
عُمَرَ	Narrator name
أَنَّ	Other
حَفْصَةَ	Narrator name
أُمَّ الْمُؤْمِنِينَ	Title
أَخْبَرَتْهُ	Received method
أَنَّ	Other
رَسُولَ اللهِ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ	Prophet name

Table 4: Isnad-Parts for Hadith Number 3

Word/Phrase	Isnad-Part
عَنْ	Received method
أَبِي	Prefix of narrator name
إِسْحَاقَ	Narrator name
عَنْ	Received method
عَبْدِ اللهِ بْنِ مَعْقِلٍ	Narrator name
عَنْ	Received method
عَدِيِّ بْنِ حَاتِمٍ	Narrator name
قَالَ	Prefix of received method
سَمِعْتُ	Received method
النَّبِيَّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ	Prophet name
يَقُولُ	Other

Tables 2, 3 and 4 considered together as the Hadith-Annotated text file. Now, we have to calculate all HMM parameters as follows:

States = {narrator name, prefix of narrator name, received method, prefix of received method, title, replacement, prophet name, other}

After we determined states, we have to calculate the start probabilities (π) as shown in table 5.

Table 5: start probabilities (π)

Isnad-Part	Start Probability
Narrator name	0/3
Prefix of narrator name	0/3
Received method	3/3
Prefix of received method	0/3
Title	0/3
Replacement	0/3
Prophet name	0/3
Other	0/3

Then we can calculate Transitions probabilities (A) as shown in table 6.

Table 6: Transitions probabilities (A)

	Narrator name	Prefix of narrator name	Received method	Prefix of received method	Title	Replacement	Prophet name	Other
Narrator name	2/16	0/16	7/16	3/16	1/16	0/16	0/16	3/16
Prefix of narrator name	4/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4
Received method	9/16	4/16	0/16	0/16	0/16	1/16	1/16	1/16
Prefix of received method	0/3	0/3	3/3	0/3	0/3	0/3	0/3	0/3
Title	0/1	0/1	1/1	0/1	0/1	0/1	0/1	0/1
Replacement	0/1	0/1	1/1	0/1	0/1	0/1	0/1	0/1
Prophet name	0/3	0/3	0/3	0/3	0/3	0/3	0/3	2/3
Other	1/6	0/6	1/6	0/6	0/6	0/6	2/6	0/6

Table7: Observations probabilities (B)

	حَدَّثَنَا	يَحْيَى بْنُ أَيُّوبَ	وَقُتَيْبَةُ بْنُ سَعِيدٍ	وَعَلِيُّ بْنُ حُجْرٍ	جَمِيعًا	عَنْ	إِسْمَاعِيلَ بْنِ جَعْفَرٍ	قَالَ	…
Narrator name	0/16	1/16	1/16	1/16	0/16	0/16	1/16	0/16	
Prefix of narrator name	0/4	0/4	0/4	0/4	0/4	0/4	0/4	0/4	
Received method	3/16	0/16	0/16	0/16	0/16	7/16	0/16	1/16	
Prefix of received method	0/3	0/3	0/3	0/3	0/3	0/3	0/3	3/3	
Title	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
Replacement	0/1	0/1	0/1	0/1	0/1	0/1	0/1	0/1	
Prophet name	0/3	0/3	0/3	0/3	0/3	0/3	0/3	0/3	
Other	0/6	0/6	0/6	0/6	1/6	0/6	0/6	1/6	

Then we can calculate Observations probabilities (B) as shown in table 7, note that this table has to contain all words and phrases in the Hadith Input File, because this will produce a huge table, so we just show a part of the first Hadith's Isnad, the rest of Hadiths can be calculated in the same way.

Testing:

Consider the following example Hadith form "Sahih Muslim" book, this Hadith will be tested against the proposed system.

حَدَّثَنِي أَبُو الطَّاهِرِ، أَخْبَرَنَا عَبْدُ اللهِ بْنُ وَهْبٍ، أَخْبَرَنِي حَيْوَةُ بْنُ شُرَيْحٍ، عَنِ ابْنِ الْهَادِ، عَنْ عَبْدِ اللهِ بْنِ دِينَارٍ، عَنْ عَبْدِ اللهِ بْنِ عُمَرَ، أَنَّ النَّبِيَّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ، قَالَ: «أَبَرُّ الْبِرِّ أَنْ يَصِلَ الرَّجُلُ وُدَّ أَبِيهِ»

The results will be as shown in table 8.

Table 8: Result of testing the example Hadith

Word/Phrase	Isnad-Part
حَدَّثَنِي	Received method
أَبُو	Prefix of narrator name
الطَّاهِرِ	Narrator name
أَخْبَرَنَا	Received method
عَبْدُ اللهِ بْنُ وَهْبٍ	Narrator name
أَخْبَرَنِي	Received method
حَيْوَةُ بْنُ شُرَيْحٍ	Narrator name
عَنِ	Received method
ابْنِ	Prefix of narrator name
الْهَادِ	Narrator name
عَنْ	Received method
عَبْدِ اللهِ بْنِ دِينَارٍ	Narrator name
عَنْ	Received method
عَبْدِ اللهِ بْنِ عُمَرَ	Narrator name
أَنَّ	Other
النَّبِيَّ صَلَّى اللهُ عَلَيْهِ وَسَلَّمَ	Prophet name
قَالَ	Other

CONCLUSION

In this paper, we present a new approach for Hadith Isnad processing based on Hidden Markov Model (HMM), this approach aims to determine the main parts of Isnad (Isnad-Parts). The proposed approach contains three phases, the first phase is the preparation phase which treats the input file of the system, this file is a raw text file that contains Hadiths from "Sahih Muslim" book, and the file will converted to Hadith-Annotated text file. The second phase is the training phase, it is use the Hadith-Annotated text file to calculate the parameters of the HMM. The last phase is the testing of the system using Viterbi algorithm. In the future work, we will use this approach to build a new "Hadith-Judgment" system that aims to distinguish between the correct or authentic Hadiths (Sahih) and the fake Hadiths (Da'eef).

Acknowledgement

The author would like to thank Deanship of Scientific Research at Umm Al-Qura university (Project ID: 43508026) for the financial support.

REFERENCES

[1] Moath, N., 2015. Muti-Agent system for Hadith Processing, International Journal of Software Engineering and Its Applications Vol. 9, No. 9 (2015), pp. 153-166 http://dx.doi.org/10.14257/ijseia.2015.9.9.13

[2] Moath, N., A. Abdelkarim and A. Musab, 2014. Arabic natural language processing laboratory serving Islamic sciences. Int. J. Adv. Comput. Sci. Applic., 5: 114-117. DOI: 10.14569/IJACSA.2014.050316

[3] Lawrence R., 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, In Proceedings of the IEEE, Vol.77, No.2, 257-286. http://www.cs.ubc.ca/~murphyk/Bayes/rabiner.pdf

[4] Valeria F., Filippo A. and Valerio P., 2007. Hidden Markov Models in Bioinformatics. Current Bioinformatics, Vol. 2, 49-61. DOI: 10.2174/157489307779314348.

[5] Moath, N., 2014. Towards innovative system for hadith isnad processing. Int. J. Comput. Trends Technol., 18: 257-259. DOI: 10.14445/22312803/IJCTT-V18P154.

[6] Bilal, K. and S. Mohsin, 2012. Muhadith: A cloud based distributed expert system for classification of ahadith. Proceedings of the IEEE 10th International Conference on Frontiers of Information Technology, Dec. 17-19, IEEE Xplore Press, Islamabad, pp: 73-78. DOI: 10.1109/FIT.2012.22

[7] Harrag, F., E. El-Qawasmeh and A. Al-Salman, 2011. Extracting named entities from prophetic narration texts (Hadith). Proceedings of the 2nd International Conference on Software Engineering and Computer Systems, Jun. 27-29, Kuantan, Pahang, Malaysia, pp: 289-297. DOI: 10.1007/978-3-642-22191-0_26

[8] Ijikom, 2015. شركة رواية-ايجيكوم http://rewayaegycom.com/rewaya_products.html (Accessed on Oct 18, 2015)

[9] Harf, 2015. موقع شركة حرف لتقنية المعلومات على الشبكة العنكبوتية www.harf.com (Accessed on Sep 19, 2015)

[10] Turath, 2015. مركز التراث لأبحاث الحاسب على الشبكة العنكبوتية www.turath.com (Accessed on Oct 12, 2015)

[11] Shamela, 2015. موقع المكتبة الشاملة على الشبكة العنكبوتية http://shamela.ws (Accessed on Aug 25, 2015)

[12] Dorar, 2015. موقع الدرر السنية على الشبكة العنكبوتية http://www.dorar.net (Accessed on Jul 3, 2015)

[13] Islamweb, 2015. موقع إسلام ويب على الشبكة العنكبوتية http://www.islamweb.net (Accessed on Jul 18, 2015)

[14] Abdelrahman, O., N. Moath, A. Abdelkarim and A. Musab, 2015. Hadith graduation as a service on cloud computing. Proceedings of the International Conference on Cloud Computing, Apr. 27-28, The College of Computer and Information Sciences at Princess, Nourah Bint Abdulrahman University, Riyadh. http://www.pnu.edu.sa/ICCC/Documents/Proceedings%20book.pdf

[15] Aqil, A. and N.B. Badia, 2011. E-narrator-an application for creating an ontology of Hadiths narration tree semantically and graphicall. Department of Computer Science, King Saud University, Riyadh, Saudi Arabia. http://ajse.kfupm.edu.sa/articles/352c_p.04.pdf

[16] Hyder, M. and S. Ghazanfer, 2008. Towards a database oriented Hadith research using relational, algorithmic and data-warehousing techniques. The Islamic Culture, Research Quarterly Journal of Shaikh Zayed Islamic Center for Islamic and Arabic Studies: ISSN 1813-775X, Shaikh Zayed Islamic Center for Islamic and Arabic Studies, University of Karachi, 19, pp. 14-32.

[17] Moath, N., A. Abdelkarim, A. Musab and O. Abdelrahman, 2015. A lexicon for hadith science based on a corpus. Int. J. Comput. Sci. Inform. Technol., 6: 1336-1340. http://www.ijcsit.com/docs/Volume%206/vol6issue02/ijcsit2015060293.pdf