

# XML Database for Hadith and Narrators

Moath Mustafa Ahmad Najeeb

Department of Computer Science,  
College of Computer at Al-Qunfudah, Umm Al-Qura University, Makkah, Saudi Arabia

## Article history

Received: 19-09-2015

Revised: 26-12-2015

Accepted: 08-01-2016

Email: mnnajeeb@uqu.edu.sa

**Abstract:** Hadith is considered as the second legalization source in Islam after the Holy Quran, it is the sayings, actions and approval of prophet Mohammad (Peace Be Upon Him), many websites and programs introduced Databases to give the required support for solving many problems in Hadith processing, graduation, verification, etc. The main objective of this research is to build and implement a database for all Hadiths and narrators of "Sahih Al-Boukari" as a first step, this database will be extended to cover other Hadith books. This database contains two types of XML documents: Hadith document and narrators document. The tags of the XML documents are the features of the HPSG formalism. This formalism is characterized by rich lexical representations. The information in a lexical entry is not only about the nature of this input but also provides information for determining what elements that can be combined with this entry. To develop this application we used java programming language and XML to ensure the scalability, interoperability and the reusability of resources and developed contents. In its first version, our application use a large database that contains all Hadiths of "Sahih Al-Boukari" and all information about its narrators extracted from "Tahdheeb Altaahdeeb" book. This application permits decomposing the entered Isnad and obtain the chain of narrators, add Hadith or narrators in database using graphical user interface, make many types of search like Hadiths of any narrator or speaker, Hadiths of any entered book or part and so on. After decomposition of enter Isnad, the user obtains the chain of narrators in a list and it can visualize the information about any narrator in the chain. All extracted information will be later used in Hadith graduation or in judgment on the validity of the Hadith.

**Keywords:** Hadith Sciences, Database, XML, Isnad, Narrator

## Introduction

A database, usually abbreviated as DB, is a structured and organized set for storing large amounts of information in order to facilitate exploitation and operations (add, update, delete or search data). It is physically translated by a set of files on disk. Databases are basically containers for data. The system that contains databases is called a database management system or DBMS. This software supports all access to the database. Examples: Structuring, storage, maintenance (concurrent access, backup, data restoration), inserting, updating, consultation and deleting a database, etc.

Different Types of database included relational database model, hierarchical database model, network database model and so on.

One advantage of a database is the information that constitutes it can be accessed easily by several programs that use them simultaneously with different objectives.

A database may be local or distributed. It is called local when it is used on a machine by a user and distributed when the information is stored on remote machines (server) and accessed through network.

XML files are a modern reminiscence of hierarchical databases and so there are native XML databases. These databases are based on the structure provided by the XML to store and identify the data. They represent an important evolution of database concept to store large volumes of data or documents, including multimedia. The organization of data in the XML however remains hierarchical (Bourret, 2005).

Because its simplicity, openness, Extensibility and because it supports Unicode and multilingual documents and it can embed multiple data types and existing data, we adopt the XML format to store and transfer the necessary data to be used and processed in our applications in the Hadith science processing field.

Our main objective in this paper is to describe the design and the implementation of XML database that will contain all information about Hadith and Narrators.

The remainder of this work is organized as follows: The second section gives the main existing encyclopedia and application for Hadith sciences processing, we focus essentially on their databases. The third section gives an introduction into XML and its usage with Java. Section 4 describes the analysis and the design of the database for Hadith and narrators using the XML format. We discuss in section 5 the implementation of the database via some graphical interfaces created with java programming language. The last section concludes our work.

## Related Works

There are many efforts in computer society that try to build the necessary databases for "Hadith science" processing, graduation, verification, etc. Most of these efforts represented as websites (Turath, 2015; Dorar, 2015), applications (Harf, 2015) and digital encyclopedia (Shamela, 2015). The "Tarajim" encyclopedia (Ijikom, 2015) is one of the most famous encyclopedias in "Hadith science"; it contains the biographies of more than 150000 narrators of Hadiths, these biographies include the name of the narrator, his "Konia", his date of birth, his place of birth, his teachers, his students, his place of death, his date of death, etc., this encyclopedia produced by "Al-Turath" center for computer research, this center also developed a library called "Alfiya" which contains more than 1300 books in many "Hadith science" fields such as "sunna science" and "dictionaries science".

The "golden encyclopedia in Hadith" is another famous encyclopedias in "Hadith science" which produced by "Harf" company (Harf, 2015), its database contains the nine "Hadith books" which are: Sahih Al-Bukhari, Sahih Muslim, Sunan Al-Termidhi, Sunan Ibn Majah, Sunan Abu Dawod, Musnad Ahmad Bin Hanbal, Sunan Al-Nasa'i, Muwatta' Malik and Sunan Al-Darami, this encyclopedia contains more than six thousand Hadiths. Another application produced by "Harf" company is "Albian" which contains all Hadiths mentioned in "Sahih Al-Bukhari" and "Sahih Muslim" as approved, correct and authenticated Hadiths.

"Rewaia" encyclopedia (Ijikom, 2015) is another noted effort in "Hadith science"; it contains more than fifty thousand Hadiths, it also provides many tools for analyzing and indexing these Hadiths.

Other well-known websites that give the user many options in searching Hadiths and quick retrieving the

electronic books are Islamweb (2015), Al-Durar Al-Sania (Dorar, 2015) and Al-Shamla library (Shamela, 2015).

A cloud system for Hadith called "Muhadith" (Bilal and Mohsin, 2012) was introduced for Hadith classification; it is an expert system that tries to simulate the scholar of Hadith, so it checks if a certain Hadith is "correct" or "weak". "Muhadith" uses a Service Oriented Architecture (SOA) to solve the communication problems between old distributed systems.

Harrag *et al.* (2011) 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.

Hyder and Ghazanfer (2008) propose Hadith-oriented database using relational and data-warehousing techniques; they represent the chain of narrators as a graph and give a weight for each arc between the nodes in this graph, there database store narrators' information in a way that can support the searching process of the biographical and historical events.

Other efforts that suggest suitable applications to serve "Hadith sciences" based on particular databases can be found in (Moath, 2015; Moath *et al.*, 2015; Abdelrahman *et al.*, 2015; Moath *et al.*, 2014; Moath, 2014; Aqil and Badia, 2011).

In spite of importance of all previous works, it has some shortcomings:

- Each application use different databases for narrators, Hadiths, books, etc
- Each database has different format and structure
- The cooperation between the companies is missing
- All these efforts start from scratch every time they built a new application
- Most databases in the previous works cannot be reused
- There is a lack in scalability and interoperability
- Many companies convert Hadiths books to digital format without any processing or analyzing
- Many applications do not have a clear process for continuity and updating
- There is no automatic enrich for these databases

To cope with these issues, we propose an XML database for Hadiths and narrators that ensure the reusability, scalability and interoperability of resources, algorithms, tools and applications.

## Analysis and Design of our Database

The Hadith reports the sayings, actions and approval of Prophet Mohammad (Peace Be Upon Him). All Hadiths represents a corpus of traditions that constitute the Sunna (the way) of the Prophet.

The Hadiths have been in writing well after the death of the Prophet (Peace Be Upon Him). As verses of the Qur'an, they were stored and disseminated orally. Some traditions relate that (probably) for fear of confusion between the Qur'an and the Hadiths, the Prophet forbade the write-up of Hadiths. The companions were therefore recorded the text of the Qur'an in writing and it is only later, when he had become familiar to many, that the authorization to put the Hadith writing was given by the prophet. Each Hadith is composed of two main parts as is shown in Fig. 1.

The first component is the chain of narrators (reporters), also called silsila isnad or sanad (سند). This chain includes the start narrator or the 'originator' and the final narrator of the Hadith. Between the start and the final narrator, there are any numbers of transmitters (narrators) who had passed on the Hadith orally from one to the other. Figure 2 shows the components of sanad.

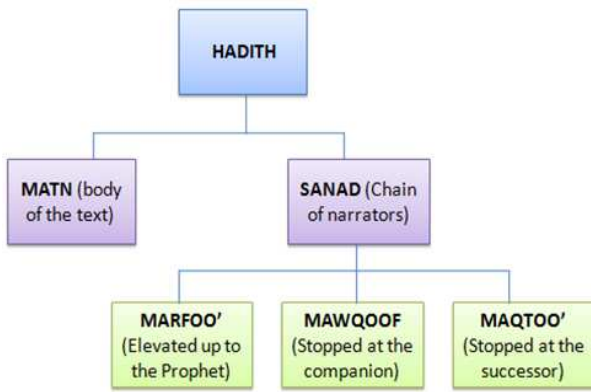


Fig. 1. Components of Hadith

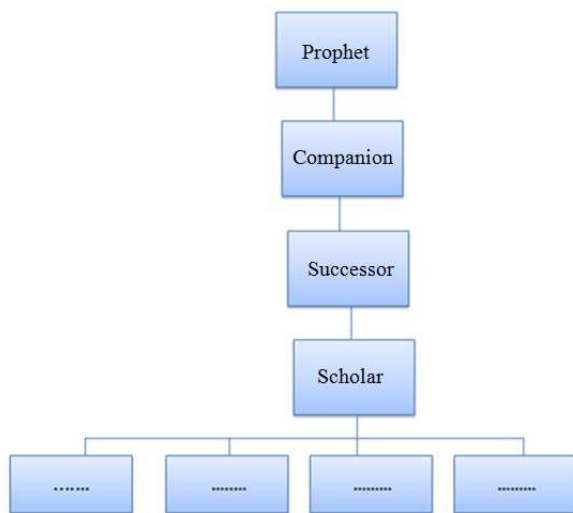


Fig. 2. Components of sanad

The second essential component is the text known as the 'matn' and is carried from the originator (Moath *et al.*, 2015). The subject of the matn is the speaker of the Hadith, there are:

- *Qudsi Hadith*: These Hadiths are transmitted by the Prophet but inspired by God. They differ from the Qur'an since the Qur'an the meaning and the letter come from God, while in the Hadith qudsi only the meaning comes from God formulation belonging to the Prophet: It starts with: the prophet said: God said
- *Marfoo' Hadith 'or shareef or Nabawi*: Reports the words, deeds, approvals of the Prophet...From him we are talking about when we say without qualification or Hadith qualifier
- *Hadith mawqoof* is assigned to a companion and regards his acts or words
- *Maqtoo' Hadith* 'is assigned to a successor (generation after the companions)

To design our database, we used The Unified Modeling Language (UML). UML is a standard visual modeling language intended to be used for modeling business and similar processes, analysis, design and implementation of software-based systems. It is a common language used to describe, specify, design and document existing or new business processes, structure and behavior of artifacts of software systems. UML can be applied to diverse application domains. It can be used with all major object and component software development methods and for various implementation platforms (e.g., J2EE, NET) (Donald, 2003). To model the system, use case diagram is used; it allows capturing the total or a particular functionality of the system. Figure 3 shows the use case diagram of our application.

The static view of an application is represented by the class diagram which is known as a structural diagram. Class diagram is not only used for visualizing, describing and documenting different aspects of a system but also for constructing executable code of the software application. The class diagram describes the attributes and operations of a class and also the constraints imposed on the system. The class diagrams are widely used in the modeling of object oriented systems. The class diagram shows a collection of classes, interfaces, associations, collaborations and constraints (Donald, 2003). The class diagrams are the only diagrams which can be directly mapped with object oriented languages and thus widely used at the time of construction. The UML diagrams like activity diagram, sequence diagram can only give the sequence flow of the application but class diagram is a bit different. So it is the most popular UML diagram used for construction of software applications. Figure 4 shows the class diagram of our database.

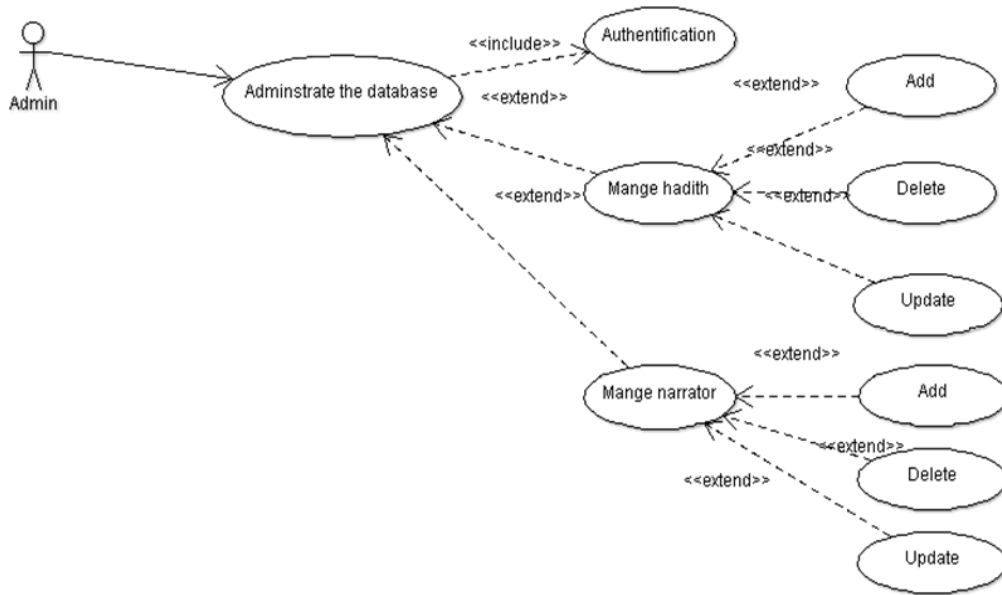


Fig. 3. Admin use cases diagram

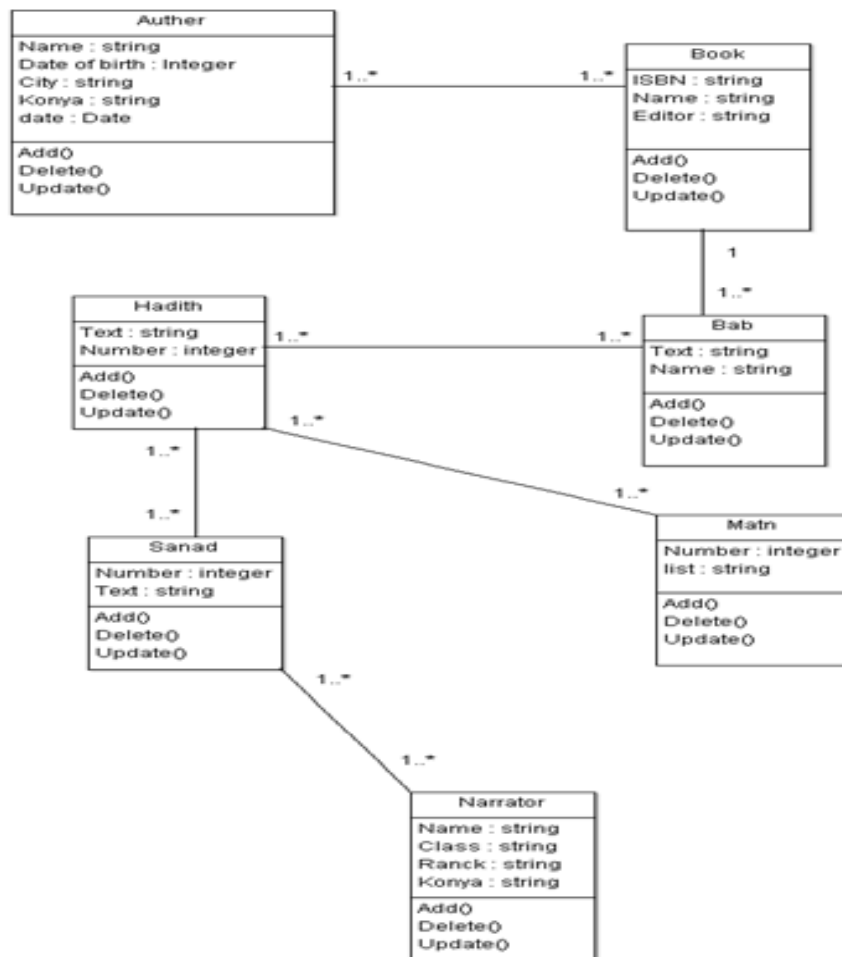


Fig. 4. Class diagram

According to the previous class diagram we can see:

- Each book is written by at least one author
- An author can participate in several books
- Each book can contain multiple parts
- Each part contains at least one Hadith
- Hadith can be in several parts
- Each Hadith is composed matn and sanad
- Each sanad is composed of several narrators

### XML and HPSG Features Overview

XML stands for Extensible Markup Language and was defined by the World Wide Web Consortium (W3C). XML is an established data exchange format. An XML document consists of elements; each element has a start tag, content and an end tag. There is no fixed set of tags, any XML tag can possess an unlimited number of attributes. Elements, tags, attributes and structure provide context information and open up new possibilities for highly efficient search engines, intelligent data mining, agents, etc. An XML document must have exactly one root element. The data coded in XML is easy to read and understand, it also can be processed easily via standard parsers. XML represents data without defining how the data should be displayed but data can be transformed into other formats via XSL or CSS. The tree structure of XML documents allows documents to be compared and aggregated efficiently element by element. The creation of documents in XML format is important for the internationalization of applications. All the above benefits support our decision to chose the XML technology in parallel with the java programming language. This object oriented language supports Unicode and contains several methods for processing and writing XML. Older Java versions supported only the DOM API (Document Object Model) and the SAX (Simple API for XML) API. Stax (Streaming API for XML) is an API for reading and writing XML Documents. In our case we used JDOM API (Java Document Object Model).

The used tags of our XML documents are the features of HPSG formalism which is a theory that analyzes the relationship between the elements of a linguistic structure not using movement or processing (as in theory of Chomsky) but in terms of information sharing. So there is only one level of representation. In fact, HPSG is a system of constraints on built structures (feature structures). Thus, the constraint satisfaction indicates the correct structure formation.

Another essential and original feature of HPSG is the integration of knowledge. The objective of grouping a homogeneous knowledge that represents various linguistic information is clearly announced. HPSG provides the integration of phonological knowledge, lexical, syntactic, semantic and pragmatic. In the current

state of the theory, it is essentially the aspects lexical, syntactic and semantic those were taken.

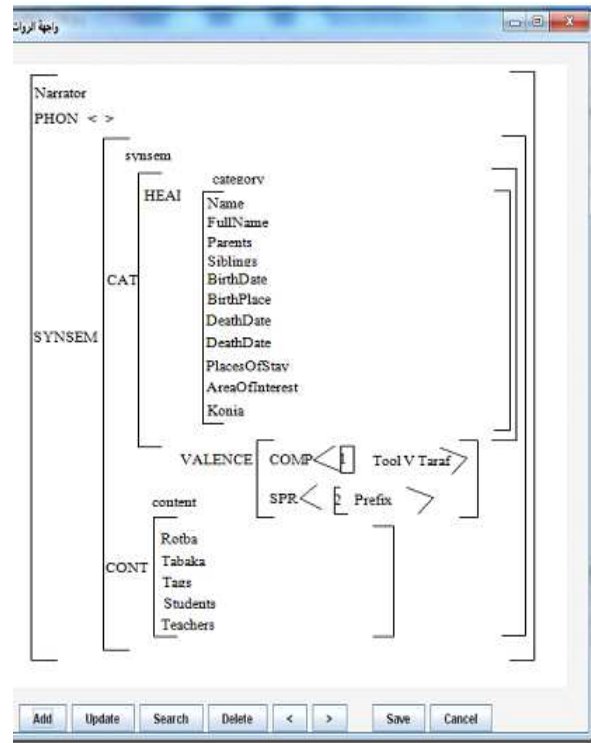


Fig. 5. HPSG AVM Structure of Narrator

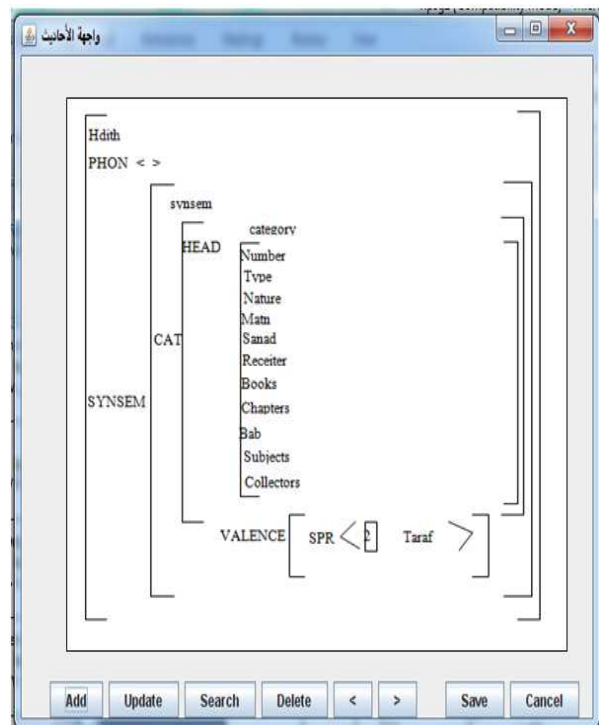


Fig. 6. HPSG AVM Structure of Hadith

In addition, HPSG is based on the formal framework of logic attribute/value. The formal properties are described in a useful and consistent framework both from a strictly theoretical point of view and implementation. Lexical entries, phrases and principles, which can be seen as additional constraints, have the same form: Attribute Value Matrix (AVM). This feature makes HPSG the most linguistic theory now used for natural language processing (Moath *et al.*, 2015).

The features are the basic element of HPSG structures. They must be suitable to a type. The features described here (the list is not exhaustive) are most frequently used:

- PHON: Phonological aspects are not taken in the presentation of HPSG. We generally use the feature phon to indicate following the words of the constituent
- SYNSEM: Includes all the syntactic and semantic features describing the component
- HEAD: Head features give the characteristics on the part of speech itself: Its type, form, optionally components selected outside the framework of the valence, etc
- CONTENT: Describes the semantic content of a constituent. This description is done primarily through a structure specifying the type of semantic relationship and its arguments, which corresponds to a predicative structure.
- COMPS: Indicates the complements categorized by the constituent

- SPR: Indicates specifiers categorized by the constituent

Figure 5 shows the HPSG AVM structure for narrator and Fig. 6 shows the HPSG AVM structure for Hadith.

## Implementation

In this section we will present our prototype. The most important interfaces will be displayed, these interfaces allow the user to access different modules like lexicon update, decompose “asanid” and search.

The user to enter the information about Hadith: number, matn, sanad, book, part, receiver and speaker, the information will be saved in the XML document named Hadithdb.xml as shown in Fig. 7.

Figure 8 shows the Result of “sanad” decomposition, using this interface; the user can obtain the list of narrators of any entered sanad.

Figure 9 shows the Search narrator interface using the full name, through this interface; user can visualize the information like the name, the adjective, the city, the rank, the year of death of any narrator.

Figure 10 shows the interface that allows user to display information of any narrator who knows a part of his name. By clicking on the information button, the Hadith’s collectors will be displayed as shown in Fig. 11.

Figure 12 shows the interface used by user to visualize all Hadith of any narrator. The result can be stored in a word document.

```

<?xml version="1.0" encoding="UTF-8" ?>
- <ahadiths>
+ <hadith>
+ <hadith>
+ <hadith>
- <hadith>
  <Number>9</Number>
  <Matn>الإيمان يضع وستون شعبة، والحياء شعبة من الإيمان</Matn>
  <sanad>حدثنا عبد الله بن محمد قال: حدثنا أبو عمار الطفي قال: حدثنا سليمان بن بلال، عن عبد الله بن دينار، عن أبي صالح، عن أبي هريرة رضي الله عنه</sanad>
  <book>الإيمان</book>
  <part>الإيمان</part>
  <narrator>أبي هريرة رضي الله عنه</narrator>
  <speaker>النبي صلى الله عليه وسلم</speaker>
</hadith>
+ <hadith>
+ <hadith>
+ <hadith>
+ <hadith>
+ <hadith>
+ <hadith>
+ <hadith>
+ <hadith>
+ <hadith>
+ <hadith>
+ <hadith>
    
```

Fig. 7. Hadithdb.xml



Fig. 8. Result of “sanad” decomposition



Fig. 9. Search narrator interface using the full name



Fig. 10. Search narrator interface using a part of the name



Fig. 11. Search narrator books interface



Fig. 12. Search Hadith interface using the narrator name

## Conclusion

This work is an effort to collect and store only authentic information about Hadith and narrators from the most famous collection “Sahih Al-Bukhari”, in common format. The first phase of the project includes the biographical information of the first 4 generations of narrators, covering the first 300 years after the migration of Prophet Mohammad (Peace Be upon Him).

As future work, we intend to continue the development of the first version of our Hadith “Takhrij” environment by adding new modules.

## Acknowledgement

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

## Ethics

This article is original and contains unpublished material. The author confirms that no ethical issues expected to arise after the publication of this manuscript.

## References

- 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.
- Aqil, A. and N.B. Badia, 2011. E-narrator-an application for creating an ontology of Hadiths narration tree semantically and graphical. King Saud University, Saudi Arabia.



- 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
- Bourret, R., 2005. XML and Databases. Felton, CA 95018 USA.
- Donald, B., 2003. UML basics: An introduction to the unified modeling language. IBM Global Services.
- Dorar, 2015. موقع الدرر السنية على الشبكة العنكبوتية <http://www.dorar.net>
- Harf, 2015. موقع شركة حرف لتقنية المعلومات على الشبكة العنكبوتية [www.harf.com](http://www.harf.com)
- 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
- Hyder, M. and S. Ghazanfer, 2008. Towards a database oriented Hadith research using relational, algorithmic and data-warehousing techniques. Islamic Culture Res. Q J., 19: 14-32.
- Ijikom, 2015. شركة رواية-ايجيكوم [http://rewayaegycom.com/rewaya\\_products.html](http://rewayaegycom.com/rewaya_products.html)
- Islamweb, 2015. موقع إسلام ويب على الشبكة العنكبوتية <http://www.islamweb.net>
- Moath, N., 2015. Muti-agent system for hadith processing. Int. J. Software Eng. Applic., 9: 153-166. DOI: 10.14257/ijseia.2015.9.9.13
- Moath, N., 2014. Towards innovative system for hadith isnad processing. Int. J. Comput. Trends Technol., 18: 257-259. DOI: 10.14445/22312803/IJCTT-V18P154.
- 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
- 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.
- Shamela, 2015. موقع المكتبة الشاملة على الشبكة العنكبوتية <http://shamela.ws>
- Turath, 2015. مركز التراث لأبحاث الحاسب على الشبكة العنكبوتية [www.turath.com](http://www.turath.com)