

# Extraction of Hadith Based on Semantic Annotation

<sup>1</sup> Ruqaiya J. Kadhim, <sup>2</sup> Dr. Norita M. Norwawi, <sup>3</sup> Dr. Adel M. Abdulaaziz, <sup>4</sup> Ashraf Al\_Omouh

<sup>1,2,3,4</sup> Science and Technology/Universiti Sains Islam/,Negeri Sembilan,Malasiay,0060

**Abstract** - The term hadith refers to the sources of basic law be referred and followed by Muslims after Holy Qur'an, however, the complexity of Arabic language including multi meanings and synonyms poses a lot of problem in term of extracting the actual context of the contents of the hadith fundamentally Allayat( Arabic verses) in the Qur'an and Hadiths which causes confusion and a lack of in-text retrieval .consequently, it becomes crucial to build the so-called semantic meaning of the content of Hadiths, which gives special characteristic of each hadith. In this paper we have presented a tool that provides semantic annotation to hadith document, which relates to the concept of prayer and all the hadith based on the concept of prayer content along with automatic information extraction (IE), suitable for the requirement of the Arabic language.

**Keywords** - Arabic Ontology, Hadith , Arabic Text Analysis ,Semantic Annotation.

## 1. Introduction

“Ref.[1]” For Muslims, Hadith is a second source of reference after The Holy Quran. Hadith (plural Hadith) are documented reference of the Words and deeds of the Prophet Muhammad by reliable narrator. “Ref.[3]” , “Ref.[4]” Hadiths are oral transmitted words and deeds of Prophet Muhammad. Hadith collections have been regarded as important tools for determining the Sunni (السنة), the way of life recommended as regulation for Muslims , based on the teachings and practices of prophet Muhammad and interpretation of Quran . Each Hadith is composed of two important components: the actual narrative text, known as *Matn* (المتن), and the chronological list of people (رواية) who were the transmitters of the Matn, traditionally known as *Isnad* (إسناد) ..

حدثنا قتيبة بن سعيد قال: حدثنا عباد، هو ابن عباد، عن أبي جمرة، عن ابن عباس قال:

"قدم وفد عبد القيس على رسول الله صلى الله عليه وسلم، فقالوا: إنا من هذا الحي من ربيعة، ولسنا نصل إليك إلا في الشهر الحرام، فمرنا بشيء نأخذه عنك، وندعو إليه من وراءنا، فقال: (أمركم بأربع، وأنهاكم عن أربع: الإيمان بالله). ثم فسر لها لهم: (شهادة أن لا إله إلا الله وأني رسول الله، وإقام الصلاة، وإيتاء الزكاة، وأن تؤدوا إلي خمس ما غنمتم، وأنهى عن الدباء، والحنتم، والمقير، والنقير " صدق رسول الله

“Ref.[1]” In terms of searching a Hadith from the huge volumes of Hadith Books available in the market, it is essential to seek the device of expert in the Islamic field to verify the Validity of special hadith. For example the reliability of the narrator of the Hadith reflects the standard of the Hadith. Nevertheless, the process of manually searching hadiths consumes a lot of. “Ref.[4]” Al-Quran and Hadiths articles that of available on internet have been considered as natural language text documents. However, the traditional human readable data resources (like electronic books or web sites) Showing real problems for achieving compatibility

The Hadith texts are quite huge and covers an entire full range of texts whose authenticity varies between bonafide to weak. According to IbnHajar: “Ref.[4]”, “Ref.[7]”, “Ref.[7]” the numbers of unique accurate Hadiths (الصحيحة الأحاديث) are 4400, and if westart to include the different Isnad(transmitters) from the Hadiths came through, then the runs into hundreds of thousands. From a veryearly stages, hadith scholars have carefully recorded and aggregated into six major collections; the bracketed numbers following the name of the collection refer to the number of Hadiths in the collection (fromwww.islamweb.net): Sahih of Bukhari (7397); Sahih of Muslim (12000); Sunan of Abu Daud (5274); Sunan of Tirmizi (3956); Sunan of Nasa’i (5758) and Sunan of IbnMajah (4341). There are other lesser known large collections, e.g., Muwatta’ of Imam Malik and Musnad of IbnHanbal. Two of the most revered workbook are that of Sahih of Bukhari and Sahih of Muslim .

“Ref.[4]” With the advent of computer systems, many people have attempted to record these books in digital format. Therefore, many of the Hadith compilations now exist in computer readable format, major as plain text, web content, or in Characteristics faithful databases. It is possible to search these contents digitally. However, it will be primitive search which is incapable of analyzing and recognizing the context of Isnad. In addition, there exist some products commercial such as Hadith Encyclopedia by Sakhr (now Harf), software that has been meant for the laymen as well as Hadith jurisprudence. It does offer a limited option to render the narration tree of a Hadith. Nonetheless, these graphs have been manually pre-compiled and are hardwired into the Hadith database,† a tedious task which is prone to errors.

For many years, computer scientists have attempted to build automated tools to lexically, syntactically, and semantically “Ref.[8]” analyze text. One of the successful attempts has been accomplished through programming languages and compilers, the programming languages tend to have clear syntax and grammar, unlike human languages, which have very complex structures and rules. Semantic annotation is a basic method for smart content and is useful in a wide range of content oriented intelligent applications. “Ref.[9]” Given ontology, the annotation process usually starts with tagging one or more phrases in the document. The system selects the appropriate concept in the ontology. Creating metadata by annotating documents is one of the major techniques for putting machine understandable data on the web . “Ref.[10]” Even though several tools are readily available for annotating text, only few of them support Arabic and of which fewer of them fully support the creation of semantically interlinked metadata necessary for a truly semantic content, which in our case, is very crucial.

Performing semantic annotation is one approaches for adding a semantic layer to existing web pages. Technically speaking, the purpose of semantic annotation is to enable computers to understand human language so that they can perform more intelligent tasks. On the other hand Semantic annotation, is “the process of labeling Web Pages with the semantics of their contents” . “Ref.[11]” , “Ref.[12]” It can also be defined as the process of mapping data instances to ontological concepts. Where Ontology is the conceptualization of a domain that is typically represented using domain vocabulary. late of Many Semantic Web annotation tools have been developed which have proved their success in multiple languages, such as English, French, Korean and Spanish, to name a few. To move this success to the Arabic language, a set of tools needs to be created to deal with the special requirements of the Semantic Arabic language.

In this paper we have presented a tool that provides semantic annotation to hadith document which relates to the concept of prayer (salad) and all the Hadiths on the concept of prayer content along with automatic Information Extraction (IE) suitable for the requirements of the Arabic language. The paper has been organized as follows: section 2 presents the background about the characteristics of Arabic language and some information Structural hadith and Arabic ontology ,and challenges of computerizing it, section 3 presents the significance of our research, section 4 provides the system architecture of our tool, section 5 highlights the implementation of the proposed tool ;,section 6 presents the tool evaluation, and finally, section 7 summarizes our future vision for the improvement tool improvement.

## 2. Related Works

### 2.1 Hadith

“Ref.[13]” Most of the classical Hadith scholars were accustomed to doing this kind of thing by hand. One Mentioned exception was M. M.al-Azami, who was the first to use computers to study the Hadith. His leading work from the late 1970s to the mid-1990s resulted in the publication of generated computer indicators. “Ref.[14]” Though al-Azami has wide lectured about his methodology, techniques, and the many constraints he had to overcome, he not published any technical paper about his work other than reporting his progress .The above work was self-financing and purely academic so it in the end ran lost strength after more than a decade. Nonetheless, al-Azami’s work laid the ground for many institutions such as The Sunnah and Sierah Center in Qatar and commercial enterprises such as Sakhr, to enter into the field. These companies computer some aspects of the Hadith Most of the work was compiled and locked into computerized manually databases. The commercial enterprises were attention more in making money than deployment the knowhow, as this would have risk their trade secrets and, as a as a result few research papers have been published in the field .

### 2.2 Semantic Annotation

“Ref.[15]” , “Ref.[16]” , “Ref.[17]” , “Ref.[18]” , “Ref.[19]” In recent years, an increasing number of IR systems have started to use ontologies in order to reach for representations semantically of documents and to help the users explained their information needs. Most of approaches “Ref.[16]” used in these systems have focused on showing concept case in documents, on the based of concept descriptions (grammatical matches), and then to use the determination concepts for document annotation. However only few approaches try also to assess the

relevance/weight of the ontological annotations. In “Ref.[20]” the weights have been calculated for annotation on the base of frequency of occurrence of concept case in the document. The approach filed in “Ref.[16]” does not calculate annotation weights but calculates the weight of each relation instance that combined annotation instances, through analyzing the link structure of the knowledgebase. To the best of our knowledge, the issue of determining annotation weights based on ontology features has only been addressed in the approach presented in “Ref.[21]” .This approach extends traditional *tf-idf* method by taking into account the global usage of concepts, individuals and triples in the annotations.

Zaidi et al. in “Ref.[22]” presented an ongoing work for information extraction from Arabic text using natural processing tool called GATE “Ref.[23]” .Crescent Quranic Corpus was used as input for the system. The system was built for named entities extraction using predefined patterns based on tokenized and tagged corpus with additional morphology and Part Of Speech (POS) features. They intended to use the extracted information for an automatic construction of Arabic domain ontology.

### 3. Significance of Study

“Ref.[24]” Arabic language is more complex and hardness compared with Latin languages. In terms of Natural Language Processing(NLP) tasks, Arabic language has five challenges that have been specific as shown in “Ref.[24]” . “Ref.[25]” Arabic language lacks case-sensitivity feature which is an important feature used by Latin languages to discover nouns. Single words in Arabic have been more than one affix and can be expressed as group of affix such as prefixes, lemma and suffixes. Next challenge is that Arabic words have different type of ambiguities related to writer models forms and spelling. Sources of information available in the Arabic language is one of the biggest challenges faced by researchers, such as corpora, name list, and mature NLP tools. Those resources are either in frequent and not free. Which makes gathering, analyzing and realization those resources as a time waste processes if the information extraction techniques depend on such resources .

“Ref.[12]” Ontology is the visualizing of a domain that typically is represented using domain terminology .Many tools have been developed deal with Semantic Web annotation recently. These tools have proved their success in several languages, such as English, French, Korean and Spanish. To move this success to the Arabic language, a set of Semantic Web tools needs to be created and

development to deal with the special requirements of the Arabic language .

## 4. Annotation Methodology

It is necessary to describe the general Design overview of our annotation tool and then check the details of this system. Figure1.Describes how interacts between the tool components with each other and shows the overcome of the tool architecture. The annotation process contains four separate steps

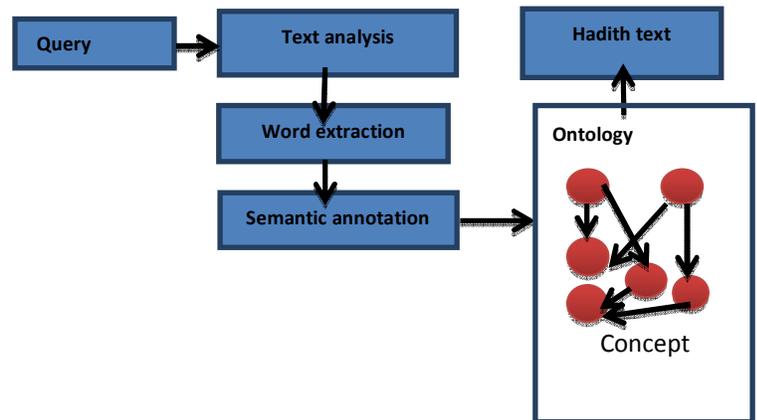


Figure1 Annotation Methodology

### 4.1 Text Analysis

- **Sentence Segmentation**

“Ref.[26]” First step in text processing is Sentence Segmentation. Division a text into sentences is generally based on punctuation and dotting In Arabic language, Guess boundary of sentence is relatively easy task approximately same as in English language. The rate of average number for the words per sentence is larger than the average in English word which will not affecting on segmentation process but on the parsing process. The sentences boundaries and phrase boundaries can be guess according to Arabic punctuation marks which are {‘ ,‘ ,. ,: ,... ,? ," , - , [ , =}.

- **Word Segmentation**

“Ref.[26]” Word segmentation is the process of getting words from text. The White Distance or interval between two words a good separator for this task but it will not work with special cases as words compound. Some words compounds are written with a space in the middle even though they are single phrase. Such cases must be solved at this stage. As

example the word “ عبد الله ” “allah | Abd” is a name of a man in Arabic language . It means that we must have knowledge base with like these words. After resolving this issue the word segmentation. Stage becomes relatively easy. however, There is another difficulty, when a few words are attached together without spaces (i.e. there are not spaces between two words when the first one ends with one of the letters “و”, “w”, “د”, “d”, “ر”, “r”, “ص”, “z”). It is formally a mistake, but may happen when dealing with non-formal text. we assume that , this mistake might not occur in the texts.

#### • Normalization

“Ref.[27]”The alphabet normalization on Arabic NLP is a basic task that researchers always apply with a common goal in mind: reducing noise in the data . This is true regardless of the task: preparation parallel text for machine translation, documents for information retrieval or text for language modeling. Normalization can be removal symbol, punctuation removal and Letter normalization (variant forms to one form conversion).Table 1 show letter normalization exampleTable1.

An example of Arabic letter normalization

Table 1 Arabic letter normalization

Normalized	D normalized
ض	D
أ-إ-آ-ا	A
هـ	H
و	W
ي	Y

This normalization will help us in searching used matching process after this stage, the normalization process will be more the ambiguity in tokenization process.

#### • Arabic Tokenization

“Ref.[28]” Tokenization is a necessary and extraordinary Process in natural language processing . “Ref.[29]” It is closely related with morphological analysis but usually it has been seen as an independent process . “Ref.[30]” Arabic words are often ambiguous in their morphological analysis. This is due to Majority of Arabic’s rich system of affixation and clitics( morphemes, which have sun tactic features of a word, however, phonologically depend on another word or phrase, and the omission of disambiguating short vowels and other orthographic

diacritics in standard orthography (“undiacritized orthography”). On average, a word form in the ATB has about 2 morphological analyses . Arabic word can be in the form [Proclitics] + [inflected word] +[Enclitics]. Then, tokenization here is equivalent to word segmentation in Chinese language where Arabic word is as a sentence in Chinese language.

#### 4.2 Word Extraction

Building the thesaurus is an essential first step in IE based on Several problems with Arabic language and the most important problems are:

1- Multiplicity of meanings ( polysemy ): one word has more than one meaning

بسم الله الرحمن الرحيم "واللاتي تخافون نشوزهن فعظوهن واحجروهن في المضاجع واضربوهن،فان أطعنكم فلا تبغوا عليهن سبيلا" صدق الله العظيم  
 امتنع:ضرب:Abstain:

بسم الله الرحمن الرحيم "ضربت عليهم الذلة والمسكنة" صدق الله العظيم

فرضت:ضرب:Impose:

بسم الله الرحمن الرحيم" ألم تر كيف ضرب الله مثلا كلمة طيبة مشجرة طيبة" صدق الله العظيم

اعطاء امثلة:ضرب:Give an example:

بسم الله الرحمن الرحيم " فضرب بينهم بسور له باب باطنه فيه الرحمة" صدق الله العظيم

ضرب:بناء:Bulid:

غرفة

مخدع:Room

مهجع:A dormitory

كمية من الماء:Amount of water

حدثنا خالد بن مخلد قال: حدثنا سليمان قال: حدثني عمرو بن يحيى، عن أبيه قال:

كان عمي يكثر من الوضوء، قال لعبد الله بن زيد

أخبرني كيف رأيت النبي صلى الله عليه وسلم يتوضأ؟ فدعا بتور من ماء،

فكفأ على يديه، فغسلهما ثلاث مرار، ثم أدخل يده في التور، فمضمض

واستنثر ثلاث مرات من غرفة واحدة، ثم أدخل يده فاغترف بها،

فغسل وجهه ثلاث مرات، ثم غسل يديه إلى المرفقين مرتين مرتين، ثم أخذ

بيده ماء فمسح رأسه، فأدبر به وأقبل، ثم غسل رجليه، فقال: هكذا رأيت النبي

صلى الله عليه وسلم يتوضأ

2 - Synonyms: a set of different words with the same meaning

1- Sunset غروب : الشفق الاحمر : الناحية : المساء

2- Afternoon الظهره : الزوال : الهجير

حدثنا محمد بن مقاتل قال: أخبرنا عبد الله قال: أخبرنا عوف، عن سيار بن سلامة قال

دخلت أنا وأبي على أبي برزة الأسلمي، فقال له أبي: كيف كان رسول الله صلى الله عليه وسلم يصلي المكتوبة؟  
فقال: كان يصلي الهجير، التي تدعوها الأولى، حين تدحض الشمس، ويصلي العصر، ثم يرجع أهدنا إلى رحله في أقصى المدينة، والشمس ونسيت ما قال في المغرب، وكان يستحب أن يؤخر العشاء، التي ندعوها العتمة، وكان يكره النوم قبلها والحديث بعدها، وكان ينقل من صلاة الغداة حين يعرف الرجل جليسه، ويقرأ بالسنتين إلى المائة.

To solve these problems, we proposed the creation of thesaurus, the process of manually constructing a thesaurus is both an art and a science. Nevertheless, here we have presented only a brief overview of this complex process. First, one has to define the boundaries of the subject area (Domain-Concept) it is (Salat). (In automatic construction, this step is simple, since the boundaries are taken to be those specified by the area covered by the document database). Domain-Concept definition includes identifying central thematic areas and sub thematic ones since it is unlikely that all topics included are of equal importance. Once this is completed, the domain is generally partitioned into divisions or subareas. Once the domain, with its subareas, has been sufficiently defined, then the desired characteristics of the thesaurus have to be identified. Since manual thesauri are structurally more complex than automatic ones Now, the collection of terms for each sub-region may begin. A variety of sources may be used for this including indexes, encyclopaedias, handbooks, textbooks, journal titles and abstracts, catalogues, as well as any existing and relevant thesauri or terminology systems. After the initial terminology has been identified, each term is analyzed for its related terminology including synonyms, wider and narrower terms, and sometimes also definitions and scope notes. These terms and their relationships are then organized into structures such as hierarchies, possibly.

#### 4.4 Ontology Construction

“Ref.[18]” With development of Internet, users find it more difficult to get their necessary information on the web. How and effectively searching appropriate information on the web has become an important problem. The best way to improve the quality of information by designing a better search and adding the semantics to web resources is. Ontology is a specific formal of a shared conceptualization. Ontology is formal explicit representation of concepts in a particular domain, concepts characteristics shows properties and attributes of the concept known as salat and limited on this Domain Islamic knowledge. Ontology is a shared conceptualization with a clear hierarchy Sequence and a strong support for logical consequences . “Ref.[19]” Ontology contains a set of specific classes or concepts Has been described clearly, property of the concepts, slot, Adhere, facet and a chain of instance have related to one class, which combines to

form the knowledge Repository. Class is the pulp of ontology, which describes the concepts in some domain. Whereas Slot, describes the property of the class and the instance.

When creating the ontology, some availability of basic principles in the design of the ontology. According to “Ref.[3]” other things that need to considered are:

- No stander methods to Design model a Special domain- there is always a Surrogate Convenient. The best solutions to focus on the development of applications available for a particular area.
- Ontology is the development of processes refined because they use the same principles
- Concepts in the ontology of a particular domain have a direct relationship with the domain, relations between the concept sex plain domain.

#### 4.5 Semantic Annotation

In this research was to determine the domain of research(prayer) and the definition of classes and related concepts in a database designed for this purpose, to link the texts of hadith with the concept requires an understanding of the relationship between the text and the concept is to identify the semantic annotation of each concept and text Basically annotation includes two main properties such as: instance and document, to defines the concepts and documents that are linked to each other. Dependence the same Technical, in this study Domain-Concept and Document have been allocated with a multi valued annotation property. Annotations can be defined as the process is to give the feature or characteristic of a specific domain knowledge base by selecting a special domain where a certain Group of concepts and the process is done in two ways manually or semi-automatically techniques using natural language processing. Usually in information retrieval systems, are added some subclasses such as Automatic-Annotation or Manual-Annotation for discrimination each case. To understand the behavior of the system researcher has been developed a distinguish technique between automatic annotations and manual annotations, and proved that manual annotations are more reliable.

“Ref.[12]” Most semantic annotation systems require two major components,: an Information Extraction (IE) module and a semantic annotation module using ontologies. Core of using IE is to identify named entities (NE) within a knowledge domain. Then, the semantic annotation module is responsible for Explain the implication, and also the semantic relationships of the context. In the following section we have presented the

system architecture of our Arabic (hadith) semantic annotation tool.

## 5. Implementation

### 5.1 Designing User Interface

The User Interface (UI) provides a ontology tree of concepts about the salat( الصلاة )it have 15 main concepts and have 82 sub-concepts in Figure 1 . Also, the Arabic Word Net is currently under construction and it does not have much related concepts to (salat) Therefore, we have focused on building our own ontology and use thesaurus ; where the (salat)has been used as the candidate domain ontology. We have developed an Arabic user interface which allows users to choose one sub-concept that belong to salat (الصلاة) and then display the hadiths that are related with this concept. We have another process when the query of salat is input then the concept will be displayed related with this query When the user chooses one of the concepts it will display the hadith texts related with query. On the other hand, there is the possibility to use semantic annotation by typing any word.

Related to hadith that mentions mentioned about salat in the book of Sahih Bukhari, such as a word “hot”, what is the relationship between warm and pray? There semantic meaning to some hadith that rose in prayer for the hot weather at the time of noon and hadith is:

حدثنا محمد بن مقاتل قال: أخبرنا عبد الله قال: أخبرنا عوف، عن سيار بن سلامة قال  
دخلت أنا وأبي على أبي برزة الأسلمي، فقال له أبي: كيف كان رسول الله صلى الله عليه وسلم يصلي المكتوبة؟  
فقال: كان يصلي الهجير، التي تدعوها الأولى، حين تدحض الشمس، ويصلي العصر، ثم يرجع أهدنا إلى رحله في أقصى المدينة، والشمس ونسبت ما قال في المغرب، وكان يستحب أن يؤخر العشاء، التي ندعوها العتمة، وكان يكره النوم قبلها والحديث بعدها، وكان ينفلت من صلاة الغداة حين يعرف الرجل جلسه، ويقراً بالسنتين إلى المائة

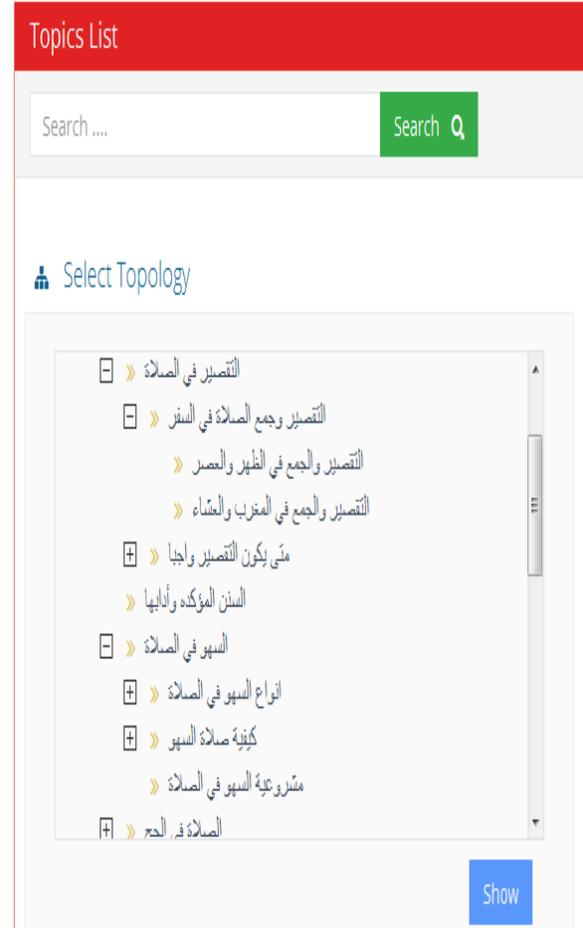


Figure 1 tree for salat

## 6. Evaluation

These aspects include the creation of valid metadata and the proper use of the search function. In this part, we have presented the results of our experiments conducted for assessing the quality of our annotation tool. The work presented here aims to evaluate the quality of the obtained results from the semantic annotation tool. We have used 10 different words to test the search function integrated in the semantic annotation tool. the query not required to be actually found in hadith but the relationship between query and hadith is semantically, even if the relationship was not clear like,

Water with prayer  
Heat with prayer  
War with prayer  
Guard with prayer.

The evaluation has been performed on 10 queries in table (1) where we have predicted that the selected words are visible in the associated text. The recall and precision of the semantic annotation search function have been calculated by applying the precision and recall formulas.

$$\text{Precision} = \frac{|(\text{relevant documents}) \cap (\text{documents retrieved})|}{|(\text{documents retrieved})|}$$

$$\text{Recall} = \frac{|(\text{relevant documents}) \cap (\text{documents retrieved})|}{|(\text{relevant documents})|}$$

Table 2. Ten words and the corresponding number of retrieved texts

word	The total number of hadiths of the prophetic	Retrieved text	Number hadith recovered and have a direct relationship with the query	Number hadith recovered and not related to with the query	Precision	recall
الماء	27	13	10	3	76%	37%
الحراره	8	6	5	1	83%	62%
الشفق	8	7	5	2	71%	62%
بزوغ	9	7	6	1	85%	66%
لحد	10	8	8	0	100%	80%
بابل	1	1	1	0	100%	100%
مخالطه	9	4	3	1	75%	33%
مسافه	9	6	4	2	66%	44%
الايام	5	3	2	1	66%	40%
حراس	5	3	2	1	66%	40%

Table2.illustrates the result of the retrieved documents. From the results we can see that the search function retrieved most related documents to the entered words. For example, when we searched for “dawn/بزوغ”, the semantic annotation search retrieved most of the documents about “sun/بزوغ” and “warmth/الحراره”.

Although the two concepts were not mentioned explicitly in the query they were related with “الحراره” in the ontology about “weather in noon/حر الظهر”, therefore, the search function retrieved the documents that contains these two concepts since they have the ontological relationship (part-of) with the instance “hot/الحراره”. Overall, the average precision and recall obtained in this experiment was: 79% for the precision and 56% for the recall. The measure of recall indicates that only 44% of the required correct annotation is not performed by the tool. In other words, in 79% of the cases, our tool has managed to map the instances existing in the text to the appropriate concepts of the ontology. On the other hand, the measure of precision value indicates that 21% of annotation performed by the tool was incorrect, or an instance is mapped to a wrong concept.

## 7. Conclusion

This paper had presented the development of a semantic annotation tool with integrated information extraction for the domain of hadith. The novelty of this work is that the proposed tool had semantically produced annotated documents using ontology for salat. Even if the tool presented in this paper has achieved its intended goals, there are many potential extensions that can enhance the its performance and output, which includes:

- Increasing the performance and the time response of the tool. Improving the Information Extraction module.
- a new indexing for hadith related with the concept ontology for salat
- increasing the performance and the time response of the tool

## References

- [1] Rahman, N.A .Bakar, Z.A. , Mohd, T.T.S.(2010). Query Expansion using Thesaurus in Improving Malay Hadith Retrieval System. Information Technology (ITSim) Information Technology Symposium (ITSim). (Volume:3 ). ISSN :2155-897 Page(s): 1404 – 1409.
- [2] Saad,S .Salim, N. Zainal , H.(2010).Towards Context-Sensitive Domain of Islamic Knowledge Ontology Extraction. International Journal for Infonomics (IJ), Volume 3, Issue 1, March
- [3] Sh , Lahore. M.A . 1972Sahih Muslim, A. H. Siddiqi (trans.), 5 vols,
- [4] Aqil, M,A. bin Badia,N.(2010). e-NARRATOR – AN APPLICATION FOR CREATING AN
- [5] ONTOLOGY OF HADITHS NARRATION TREESEMANTICALLY AND GRAPHICALLY. The

Arabian Journal for Science and Engineering, Volume 35, Number 2C. Page(S): 50 – 68.

- 6- "صحيح البخاري"  
للإمام الحافظ أبي عبد الله محمد بن إسماعيل البخاري، مراجعة الشيخ محمد علي القبطي والشيخ هشام البخاري، طبعة 1424 هـ
- 7 - "صحيح مسلم"  
لأبي الحسن مسلم بن الحجاج القشيري النيسابوري، دار السلام للنشر والطباعة، 1404 هـ
- [8] D, Brickley and R, V, Guha. (2010). RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation, 2004. URL: <http://www.w3.org/TR/rdf-schema/>. Last visited: 27 Jun, 2010.
- [9] 9- M. Erdmann, A. Maedche, H.-P. Schnurr, and S. Staab, "From Manual to Semi-Automatic Semantic Annotation: About Ontology-Based Text Annotation Tools", *Elec. Trans. on AI J.*, Linköping Univ., Sweden, **6(2)**(2001). URL: [http://www.aifb.kit.edu/portal/index.php?title=Spezial:Hochladen&wpDestFile=2001\\_474\\_Erdmann\\_From\\_Manualto\\_1.pdf](http://www.aifb.kit.edu/portal/index.php?title=Spezial:Hochladen&wpDestFile=2001_474_Erdmann_From_Manualto_1.pdf). Last visited: 27 Jun, 2010.
- [10] Li, D. and Huan, L. 2008. The Ontology Relation Extraction for Semantic Web Annotation. C cgrid. 2008 Eighth IEEE International Symposium on Cluster Computing and the Grid (CCGRID). 534-541.
- [11] Taniar, D., and J. Rahayu. 2006. Web Semantic Ontology. Hershey, USA.
- [12] Bin Saleh, Al-Khalifa, H, S. L, M. 2009. AraTation: An Arabic Semantic Annotation Tool. ACM *ijWAS '09* Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services. Page(S): 447 – 451. doi>10.1145/1806338.1806421
- [13] "سنتنايماجمة معالفهارس"  
للإمام عبد الله محمد بن يزيد المعروف بابن ماجة القرظي، تحقيق محمد مصطفى الأعمش، مطبعة دار للطباعة، العمارة، الرياض، مجلدات، الشركة العربية، طم، م 1984
- [14] M. M. Al-Azami, "A Note on Work in Progress on Computerization of Hadith", *J. of Islamic Studies*, Oxford, **2(1)**(1991), pp. 86–91.
- [15] Song, J, F, Weidong, W, Z, Li, X, G. and Xu Z, N. (2005) Ontology-Based Information Retrieval Model for the Semantic Web. In *EEE '05: Proceedings of the 2005 IEEE International Conference on e-Technology, e-Commerce and e-Service*, pages 152–155, 2005.
- [16] Rocha, C. Schwabe, D. Marcus. Aragão, P, D. (2004). A hybrid approach for searching in the semantic web. In *WWW'04: Proceedings of the 13th international conference on World Wide Web*, pages 374–383.
- [17] Kiryakov, A. Popov, B. Terziev, I. Manov, D. and Ognyan off, D. (2004). Semantic annotation, indexing, and retrieval. *J. Web Sem.*, **2(1)**:49–79.
- [18] Regina M. M. Braga, W, C, M, L., and Mattoso, M. (2000). Using Ontologies for Domain Information Retrieval. In *DEXA Workshop*, pages 836–840.
- [19] Ozcan, R and Aslandogan, Y, A. (2005) Concept-Based Information Access. In *ITCC '05: Proceedings of the International Conference on Information Technology: Coding and Computing*, pages 794–799.
- [20] Vallet D. Fernández, M, and Castells, P. (2005). An Ontology-Based Information Retrieval Model. In *ESWC, pages* 455–470.
- [21] Ruotsalo, T and Hyvönen, E. (2007). A Method for Determining Ontology-Based Semantic Relevance. In *the 18<sup>th</sup> Database and Expert Systems Applications Conference, DEXA*, pages 680–688.
- [22] S. Zaidi, M.-T. Laskri, and A. bdelali. (2010). Arabic Collocations Extraction using Gate. International Conference on Machine and Web Intelligence (ICMWI), pp. 473–475.
- [23] H. Cunningham, et al. (2011). Text processing with GATE (Version 6). University of Sheffield Department of Computer Science. ISBN 0956599311.
- [24] S. Abdelrahman, M. Elarnaoty, and M. Magdy. 2010. Integrated MACHINE learning Techniques for Arabic Named Entity Recognition. *International Journal of Computer Science*. vol. 7, no. 4, Page(s). 27–36.
- [25] Albukhitan, S. Helmy, T. (2013). Automatic Ontology-Based Annotation of Food, Nutrition and Health Arabic Web Content. The 4th International Conference on Ambient Systems, Networks and Technologies Volume. Pages 461–469.
- [26] Jurafsky, D and Martin, H, J. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing*. Computational Linguistics and Speech Recognition. Prentice Hall. (2<sup>nd</sup> ED).
- [27] Nizar, Y, H. 2010. *Introduction to Arabic Natural Language Processing Synthesis Lectures on Human Language Technologies*. Morgan and Claypool Publishers
- [28] S. Bird, E. Klein, and E. Loper. (2009). *Natural Language Processing with Python*. O'Reilly Media.
- [29] Chanod, J, P and Tapanainen, P. 1996. A non-deterministic tokenizer for finite-state parsing. *ECAI 96*. 12th European Conference on Artificial Intelligence.
- [30] Habash, N. and Rambow, O. (2005). Arabic tokenization, part-of-speech tagging and morphological disambiguation in one fell swoop. *Proceedings of the 43rd Annual Meeting of the ACL*. Page(S) 573–580.