

# Semantic Hadith: Leveraging Linked Data Opportunities for Islamic Knowledge

Amna Basharat  
Dept. of Computer Science  
University of Georgia  
Athens, GA, 30605 USA  
amnabash@uga.edu

Bushra Abro  
Dept. of Computer Science  
Islamic International University  
Islamabad, Pakistan  
bushraabro@hotmail.com

I. Budak Arpinar  
Dept. of Computer Science  
University of Georgia  
Athens, GA, 30605 USA  
budak@uga.edu

Khaled Rasheed  
Dept. of Computer Science  
University of Georgia  
Athens, GA, 30605 USA  
khaled@uga.edu

## ABSTRACT

While the linked data paradigm has gathered much attention over the recent years, the domain of Islamic knowledge has yet to cache upon its full potential. The web-scale integration of Islamic texts and knowledge sources at large is currently not well facilitated. The two primary sources of the Islamic legislation are the Qur'an and the Hadith (collections of Prophetic Narrations) and form the basis of laying the foundation for anyone wanting to learn Islam. This paper presents ongoing design and development efforts to semantically model and publish the Hadith, which holds a primary position as the next most important knowledge source, after the Qur'an. We present the design of the linked data vocabulary for not only publishing these narrations as linked data, but also delineate upon the mechanism for linking these narrations with the verses of the Qur'an. We establish how the links between the Hadith and the Qur'anic verses may be captured and published using this vocabulary, as derived from the secondary and tertiary sources of knowledge. We present detailed insights into the potential, the design considerations and the use cases of publishing this wealth of knowledge as linked data.

## CCS Concepts

•Information systems → Multilingual and cross-lingual retrieval; Information extraction; •Computing methodologies → Ontology engineering;

## Keywords

linked data; hadith; Quran; Qur'an; semantic web; Islamic knowledge;

## 1. INTRODUCTION

The vast amount of Islamic Creed and legislation derives itself from and is based primarily on the two most fundamental sources of Islam: namely the *Qur'an* and the *Sun-*

*nah* (way of life) of the *Prophet Muhammad*. The later is contained with the vast body of *Hadith* literature [22]. Formally, the *Hadith* is defined as the (recorded) narrations of the sayings and deeds of the Prophet Muhammad.

Our research primarily is motivated to overcome the inherent knowledge acquisition bottleneck in creating semantic content in semantic applications. We have established how this is particularly true for knowledge intensive domains such as the the domain of Islamic Knowledge, which has failed to cache upon the promised potential of the semantic web and the linked data technology; standardized web-scale integration of the available knowledge resources is currently not facilitated at a large scale [7].

## 1.1 Background Context and Motivation

### 1.1.1 Importance of Hadith

To understand the important of Hadith, the principles of Qur'anic understanding and the science of *tafseer* or exegesis must be considered. The verses in the Qur'an cannot be understood in isolation. The Hadith are used to illustrate the Historical context, the reasons for revelation and elaboration of essential concepts that may not be directly evident. This important principle has been adopted by scholars across centuries to write scholarly commentaries and explanations. Infact, it is a necessary condition to produce an accurate *tafseer* of the Qur'an as explained in detail by Philips [30].

To explain this principle, as an example, consider the Figure 1, a derived snapshot taken from QuranComplex<sup>1</sup>, the official manuscript, with a translation and a commentary, provided by the Kingdom of Saudi Arabia. The snapshot shows two verses from the first chapter of the Qur'an. The translation is annotated with a commentary (given in the footnotes in this case) in order to provide additional details where important. It is worth noticing that most authentic and reliable commentaries would draw knowledge from the sources of Hadith. In the case of this snapshot, the verse 2 contains an annotation which provides an elaboration based on an authentic Hadith, from one of the many collections

Copyrights held by author/owner(s)  
WWW2016 Workshop: Linked Data on the Web (LDOW2016), Montreal, Canada

<sup>1</sup><http://qurancomplex.gov.sa/Quran/Targama/Targama.asp>

## Al-Fatiha (The Opening)

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

1. In the Name of Allāh, the Most Gracious, the Most Merciful

الْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ

Verse 1:2

elaborated/  
explained by:

2. All the praises and thanks be to Allāh, the Lord<sup>[1]</sup> of the 'Ālamīn (mankind, jinn and all that exists).<sup>[2]</sup>

### Commentary

[1] ...

[2] (V.1:2). Narrated Abu Sa'īd bin Al-Mu'alla: While I was praying in the mosque, Allāh's Messenger صلى الله عليه وسلم called me but I did not respond to him. Later I said, "O Allāh's Messenger, I was praying." He said, "Didn't Allāh say - Answer Allāh (by obeying Him) and His Messenger when he صلى الله عليه وسلم calls you." (V.8:24). He then said to me, "I will teach you a *Sūrah* which is the greatest *Sūrah* in the Qur'ān, before you leave the mosque." Then he got hold of my hand, and when he intended to leave (the mosque), I said to him, "Didn't you say to me, 'I will teach you a *Sūrah* which is the greatest *Sūrah* in the Qur'an?' He said, " *Al-Hamdu lillahī Rabbil-'ālamīn* [i.e. all the praises and thanks be to Allāh, the Lord of the 'Ālamīn (mankind, jinn and all that exists)], *Sūrat Al-Fātihah* which is *As-Sab' Al-Mathānī* (i.e. the seven repeatedly recited Verses) and the Grand Qur'ān which has been given to me." *Sahih Al-Bukhārī*, Vol.6, *Hadīth* No. 1)

Hadith: Bukhari-6-1

Figure 1: A snapshot of a typical Qur'anic Commentary

of Hadith, called *Sahih Bukhari*, which is known to be the most authentic and reliable Hadith collection.

### 1.1.2 Motivation: Potential for Knowledge Formalization and Linking

There are hundreds and thousands of Qur'anic commentaries produced over the last few centuries, in various languages that draw upon and rely heavily on the Hadith sources to provide an interpretation of the Qur'anic verses. Given this fact, the potential for knowledge formalization and linking is not only evident, rather it cannot be overemphasized. Formally modeling this wealth of knowledge and the links would enable new ways of research and knowledge discovery and synthesis - the very motivation for this research. However, realizing this vision to span across the plethora of Islamic resources is a mammoth task. We present some key challenges presented.

### 1.1.3 Challenges in Interlinking Islamic Knowledge Sources

There have been some recent efforts to publish Islamic knowledge as linked data on the Linked Open Data (LOD) cloud. The efforts primarily focus on the Qur'an. The two datasets that we consider in our research and attempt to link with our Semantic Hadith research include *SemanticQuran*<sup>2</sup> [34] and *QuranOntology*<sup>3</sup> [16].

However, there are no known publically available sources of data or vocabularies published as linked data for the Hadith. There are number of well known Hadith repositories available, which provide the provision of browsing and searching the hadith collections such as sunnah.com, dorar.net being the most prominent ones.

<sup>2</sup><http://datahub.io/dataset/semanticquran>

<sup>3</sup><http://www.quranontology.com>

We review some of the state of the art towards computational approaches applied to Hadith texts in Section 6. Here, we would like to emphasize that interlinking the Qur'anic verses and the Hadith is a non-trivial task. We summarize some factors that make this extremely challenging. Most of the classical sources do not use a standardized numbering scheme for the Hadith. This is contrary to the Qur'anic verses which have a standardized numbering scheme. There are multiple sources of the Hadith, which may have different levels of authenticity which is a matter of discussion beyond the scope of this paper. Despite the fact that most Hadith collections have now been classified into authentic categories, the mapping of this classification to the sources that cite them is only possible if the Hadith are extracted and linked in a formalized manner. In addition, to add to the challenge, the Hadith are of varying length, and oftentimes the commentator or the *tafsir* scholar will only quote a part of the Hadith or make a passing reference to it, making it extremely difficult to trace the original Hadith being cited. To add to the challenge, several Hadith may have common portions of narrations, therefore it makes it all the more challenging to identify, which exact Hadith is being quoted or referred to. We believe that a knowledge formalization and linking mechanism, using the linked data standards, is the way forward for solving some or more of these challenges.

## 1.2 Contributions of the Paper

In this paper we make the following contributions:

- We provide the first of its kind linked data model, called *Semantic Hadith* for publishing Hadith as Linked Data and for linking with other key knowledge sources in the Islamic domain, primarily the Qur'an.
- We present a classification of the various levels of links that may potentially be established between the Ha-

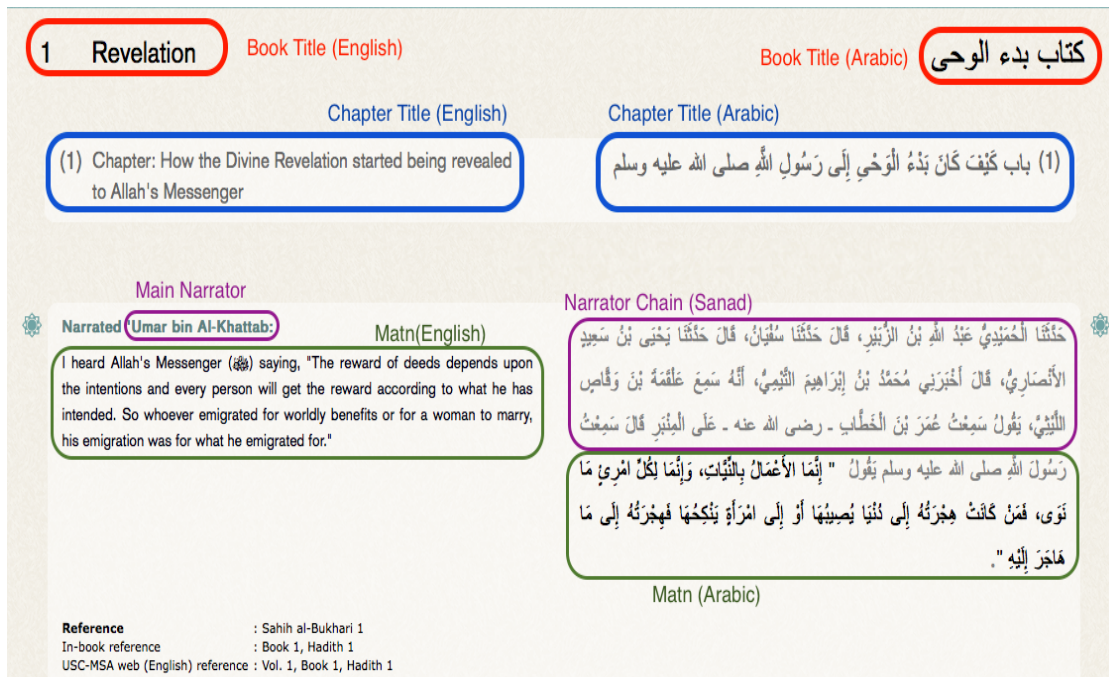


Figure 2: A Sample Hadith Snapshot

dith, the Qur'an and other data sets on the linked data cloud. This classification spans various levels of granularity. We highlight the linking challenges and design issues with each one and present potential modeling solutions.

- We provide a knowledge extraction, linking and publishing framework that may be reused for publishing similar knowledge and linked with the existing linked data cloud. We present our preliminary implementation of this framework.

## 2. ONTOLOGY FOR SEMANTIC HADITH

We first present an illustration of the structure of the Hadith, and then detail upon the design of the ontology for Semantic Hadith.

### 2.1 Hadith Structure

Figure 2 shows a sample of a Hadith taken from sunnah.com<sup>4</sup>. A given Hadith has two main parts: the actual narration or the content portion of the Hadith is called *Matan*, and the chain of narrators(reporters) through whom the narration has been transmitted and then recorded is traditionally known as the *Sanad* or simply the *chain of narrators*. The *Sanad* is a chronological chain of narrators, each mentioning the one from whom he heard the Hadith all the way to the prime narrator of the *Matan* followed by the *Matan* itself [32]. The *Sanad* plays the most important role in determining the authenticity of the Hadith, which is the most crucial indicator Scholars resort to when determining whether to accept or reject a Hadith.

### 2.2 Ontology Schema

Figure 3 shows the conceptual model for publishing Hadith data on the LOD cloud. Here we summarize the key entities and relations that we chose to include in the conceptual design model of the Semantic Hadith ontology schema.

- *Hadith*: This is the central entity in the domain model. Since there had been no standardized numbering scheme for the Hadith since the beginning, a few alternate numbering schemes may be encountered, therefore the provision to include alternate numberings is made.
- *Matn*: This is primarily a textual entity, which contains the main narration of the Hadith, without the chain or narrators or the Sanad.
- *Narrator*: A Narrator is essentially a Person, with the special role of a narrator of the Hadith. One narrator may have many Hadith attributed to him or her. If a narrator is the root narrator of the Hadith, then a Hadith is usually *attributedTo* him/her. This is shown by the relation between the *Hadith* and *Narrator*. Notice in the Figure 2, the english translation does not provide the entire NarratorChain, rather it only provides the name of the narrator to whom the Hadith is attributed to. However, this is not the case for the Arabic (original) version of the Hadith, which usually contains the entire chain of narrators. The chain is often omitted in the books for simplifying the hadith text for the reader and making it more meaningful and relevant. However, the NarratorChain is considered indispensable for determining the validity and authenticity of the Hadith, especially if no other validation source is mentioned.
- *Sanad(NarratorChain)*: This is an entity which will contain reference to a Narrator entity, and a level,

<sup>4</sup><http://sunnah.com/bukhari/1>

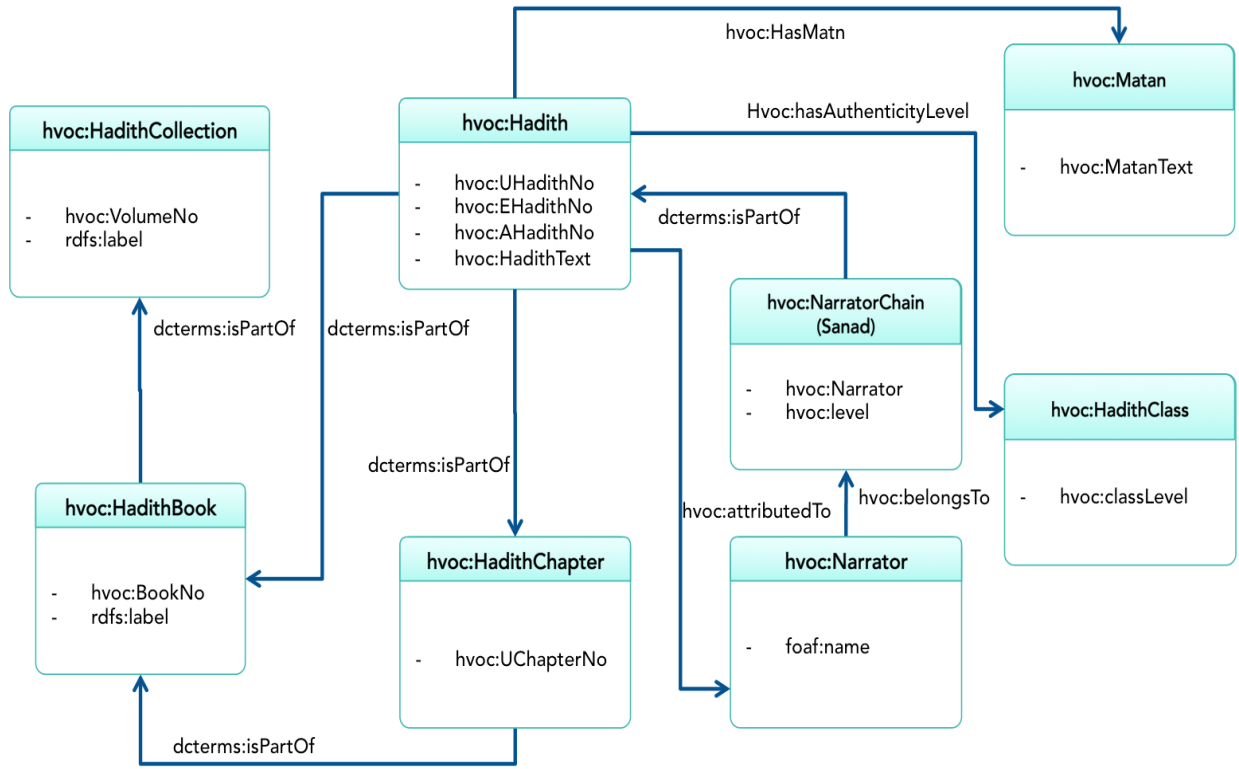


Figure 3: High Level Design of Semantic Hadith Ontology

which will indicate the sequence of the narrator in the chain. Same narrators may appear in many chains.

- *HadithClass*: This indicates the authenticity level of the Hadith. These are detailed in [32].
- *HadithChapter, HadithBook and HadithCollection*: These are entities meant for structural organization of the Hadith. A Hadith is a part of a Chapter, which usually contains thematically co-related collections of Hadith. Chapters are collected in Books and Books are compiled as Collection or Volume.

### 2.3 Vocabulary Design

We choose the *hvoc* prefix for the SemanticHadith vocabulary, as in the domain model. We also ensure reuse of well established linked data vocabularies such as FOAF<sup>5</sup> [10], SKOS<sup>6</sup> [27], and DublinCore<sup>7</sup> [35]. We also provide equivalence relations where applicable. Some of the most relevant equivalence relations are with the *bibo* ontology<sup>8</sup>.

## 3. LINK MODELING AND DESIGN ISSUES

One of the most important constituents in the design of Semantic Hadith, is the aspect of facilitating the interlinking of knowledge at various levels. We have earlier described the Macro-Structure for Islamic Knowledge in [7]. We distinguish between the nature of links based on the level of

granularity at which they are modeled. A *Macro-Level Link* is considered to be one where the source entity is either at the level of a *Verse* in the Qur'an or a *Hadith* in a Hadith Collection. If a link is established for a group of Verses or Hadith, then it will also be considered at the Macro-level. A Micro-level link will be at a sub-verse, sub-Hadith or word or phrase level. For the scope of this paper, we would detail upon only the Macro-level links of the most essential types.

### 3.1 Hadith-to-Hadith Links

As essential type of links to be established are those links, where by one Hadith is linked to or related to another Hadith. This could be done for Hadith which may be part of the same collection; or it may be between Hadith that are part of different collections. These relations may be of the following primary types: 1) Two Hadith may be considered to be related if they have the same 'sanad'. 2) Two Hadith may be considered to be related if they have the same 'matan'. Note that two Hadiths may occur in the same collection, in two different chapters, under different thematic categorizations, however, they may be enumerated or numbered differently. Therefore, by asserting this Hadith as similar/related or identical, we aim to make these links explicit. Oftentimes, the same Hadith may be made part of a different collection and therefore, asserting an identity link would become crucial. This is illustrated in Figure 4. To handle the annotations between two Hadith, we define an entity called *HadithRelation*, for which the *source* and *destination* represent the two ends of the relation. The relation would often have a common *Theme*. The *RelationType* indicates whether the two Hadith are similar, indicated by *Identity* as the *RelationType*, or one Hadith may elab-

<sup>5</sup><http://xmlns.com/foaf/spec/>

<sup>6</sup><https://www.w3.org/2004/02/skos/>

<sup>7</sup><http://dublincore.org/>

<sup>8</sup><http://bibliontology.com>

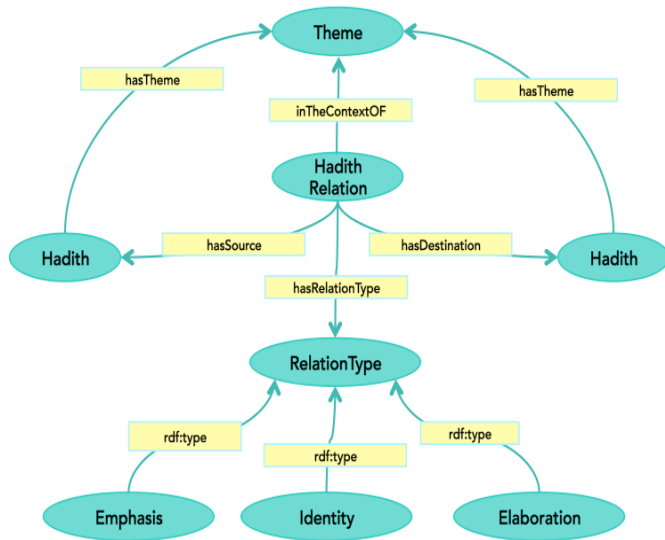


Figure 4: Conceptual Design model for Hadith-Hadith Relationship

orate another indicated by **Elaboration** and so forth. These relation types are not exhaustive and may be iteratively refined.

### 3.2 Linking the Qur'an and Hadith

One of the most significant aspects of linking the Hadith dataset is with the verses in the existing Qur'an datasets. We distinguish two types of relationships that may occur between the Qur'an and the Hadith: 1) There may be Verses, entire of which or part of which may be 'Cited' or quoted in a Hadith. This is the most direct kind of relation that exists between a Hadith and a Verse. 2) The other relations are based on those that can only be derived from Scholarly commentaries. The design and modeling issues for both these types are delineated further.

#### 3.2.1 Verse to Hadith Links based on Direct Citations

A direct link between a Hadith and Verse is characterized as one whereby a Hadith contains within its main body a complete verse or a meaningful portion of it. This is modeled in the Figure 5. A **Citation** entity is created, which is specific reference to a relation with its **source** as a Hadith and the **destination** as the Verse, indicating that its the Hadith that is encapsulating the Verse. It is considered important that we characterize the **CitationType** as either **Complete**, **Partial** or **In-Direct**. A **Complete** Citation will include the entire verse in the body of the Hadith and the Verse will be quoted as such. A **Partial** citation may only contain part of the Verse in the body of the Hadith. To indicate this, the **sub-verse** entity is introduced, which will identify the part of the Verse **citedIn** the Hadith. This is indicated by the relation **characterizedBy**. It is important to note that it is important to annotate and capture the **sub-verse**, since there may be portions of the same verse that may be linked to different Hadith.

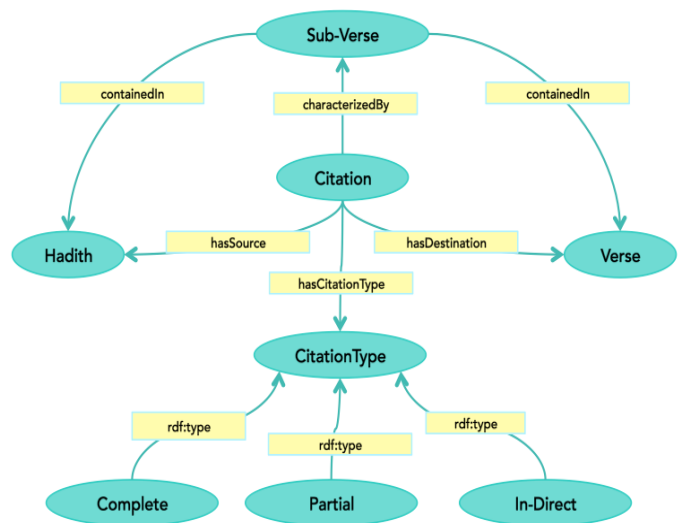


Figure 5: Conceptual Design model for Hadith-Verse Relationship

#### 3.2.2 Verse to Hadith Links based on Scholarly Commentaries

Another important type of links to be established between the Hadith and the Qur'anic verses are shown in the model as conceptualized in Figure 6. This is based on the earlier motivation, provided on the basis of Figure 1. In this type of relation, we create an entity **Verse-Hadith-Relation**. In this case, the source is a **Verse** and the destination is a **Hadith**. The reason is that the Hadith will always be used to elaborate or provide the context for the verse in any given commentary or book of exegesis. The **RelationType** may be provided. In this relation type, the most important aspect is establishing the source of the authority of the relation. This is established by the relation **uponAuthorityOf** with a **Scholar** and a **relationEstablishedIn** with a **Book**. The **Book** is naturally **authoredBy** the **Scholar** to whom the relation is attributed.

### 3.3 Linking Hadith with other Datasources

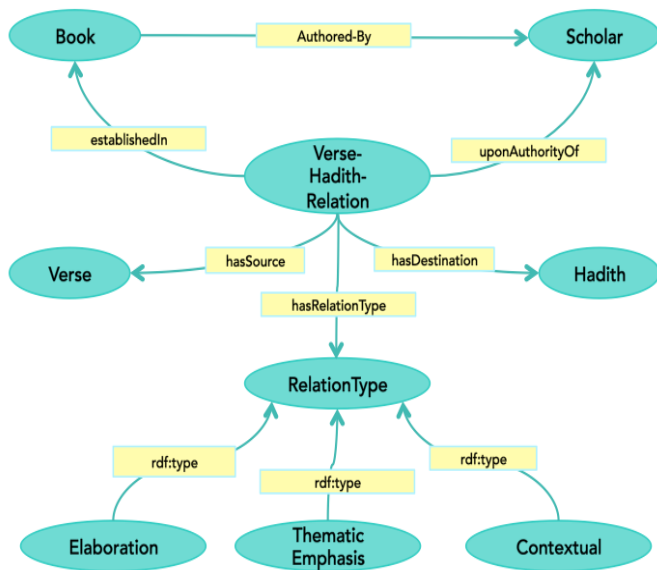
We aim to provide the provision of linking the Semantic Hadith with other available datasources in the LOD cloud. We present a high level view of the linked cloud model for Islamic knowledge in Figure 7. We also mention those datasources, which although are not directly available on the LOD, present potential for linking.

#### 3.3.1 Linking with Existing Datasources in the LOD Cloud

The two available datasources to which the Semantic Hadith is linked to are the **QuranOntology** and **SemanticQuran**. Semantic Quran links itself to DBPedia<sup>9</sup> and Wiktionary<sup>10</sup>. Links would be established between entities in the Hadith to the ones in these two datasets to begin with. For this infor-

<sup>9</sup><http://dbpedia.org>

<sup>10</sup><http://wiktionary.dbpedia.org/>



**Figure 6: Conceptual Design model for Verse to Hadith Relationship based on Scholarly Commentaries**

information extraction would be carried out. There are some important datasources which are not directly part of the linked data cloud but have been made available through QuranOntology and SemanticQuran. These are shown in the Figure 7 namely: QuranyTopics<sup>11</sup><http://quranytopics.appspot.com>, QuranCorpus<sup>11</sup>, and Tanzil<sup>12</sup>.

One of an essential linking aspects would be to thematically map the QuranyTopics to those of HadithTopics.

### 3.3.2 Linking with other sources

There are other datasources that we plan to link with in the future. Scholars database from a source such as MuslimScholarsDatabase<sup>13</sup> or eNarrator (Hadith Isnad Ontology) [32] [4]. The major limitation is that these sources are not currently available in Linked Data format. However, they present huge potential for linking.

## 4. LINKED DATA PUBLISHING FRAMEWORK AND IMPLEMENTATION

In order for Semantic Hadith to become a defacto standard and an integral part of the emerging Semantic Web and the LOD cloud for the Islamic Knowledge domain, we also aimed at providing a reusable framework for publishing available Hadith based knowledge sources as linked data. This is shown in Figure 8. As elaborated in Section 1.1.3, there are multiple hadith repositories available. Therefore, this reusable framework will benefit multiple hadith publishers to not only expose their data, but also to establish equivalence links with other repositories. This would be essential towards realizing the vision of linked Islamic knowledge as presented in [7].

<sup>11</sup>[corpus.quran.com](http://corpus.quran.com)

<sup>12</sup>[tanzil.net](http://tanzil.net)

<sup>13</sup><http://muslimscholars.info>

## 4.1 Overview of the Framework

The key stages of the framework shown in the Figure 8 include: 1) Data Selection, where the data source is selected; 2) Vocabulary Design and Selection, where conceptual and formal knowledge modelling is carried out; 3) Knowledge Extraction, where the process of information and knowledge extraction is carried out; 4) RDF Generation, where the extracted knowledge is converted into the RDF format; 5) Publishing, Linking and Validation is done to make the converted RDF data available via a SPARQL endpoint; and 6) Consumption, is the last stage where the dataset now available as linked data may be consumed into applications.

## 4.2 Implementation Details

We provide some key details of the ongoing implementation process, about the dataset used for publishing as linked data, the knowledge extraction and linking mechanism. We summarize some key results and also highlight some challenges and limitations faced in the implementation process.

### 4.2.1 Data Sources

As the first Hadith repository to be annotated using the Semantic Hadith Model, we have taken the data of Sunnah.com, which is a structured data repository of some of the most well known and authentic collections of Hadith. The foremost collections are those of *Sahih Bukhari* and *Sahih Muslim*. Altogether, there are 11 collections in this dataset, with over 25,000 Hadith.

### 4.2.2 Knowledge Extraction and Linking

For the initial implementation, we focused on extracting some of the key relations explained earlier.

We extracted *Verse-Hadith* Links from QComplex Commentary<sup>14</sup>. This is one of the only datasource through which we were able to extract numbered hadith references, which could be automatically mapped to the hadith collections available with us. An example of such a reference is shown in Figure 1. A pattern extraction module was designed to parse the contents of the commentary. The content of the verses, translation and the footnotes were segmented. The mapping between the verses and the corresponding footnotes was easy, given the direct correlation. Pattern matching was then applied to extract the collection name, volume number and the hadith number. This was then mapped to the numbers in our hadith collection. This can be challenging at times, because not all hadith collections use the same type of numbering convention. In such a case, it is non-trivial to map the hadith citation to the corresponding hadith in the repository. Human intervention will be required for validation. We were able to obtain and validate some 300 verse to hadith relations. Since the commentary is not a detailed one, rather comments are only sparingly included as footnotes to the verse translations, it was expected that this number would be small.

We also performed text mining on the arabic text of the hadith data to obtain the *Hadith-Verse* citations, as described in Section 3.2.1. For this, we developed a verse-extraction component, which implements a sub-string matching problem, in order to detect complete or partial verses that may be cited in a given hadith. This is not trivial for several reasons. Different verses span different lengths in

<sup>14</sup><http://qurancomplex.gov.sa/Quran/Targama/Targama.asp>

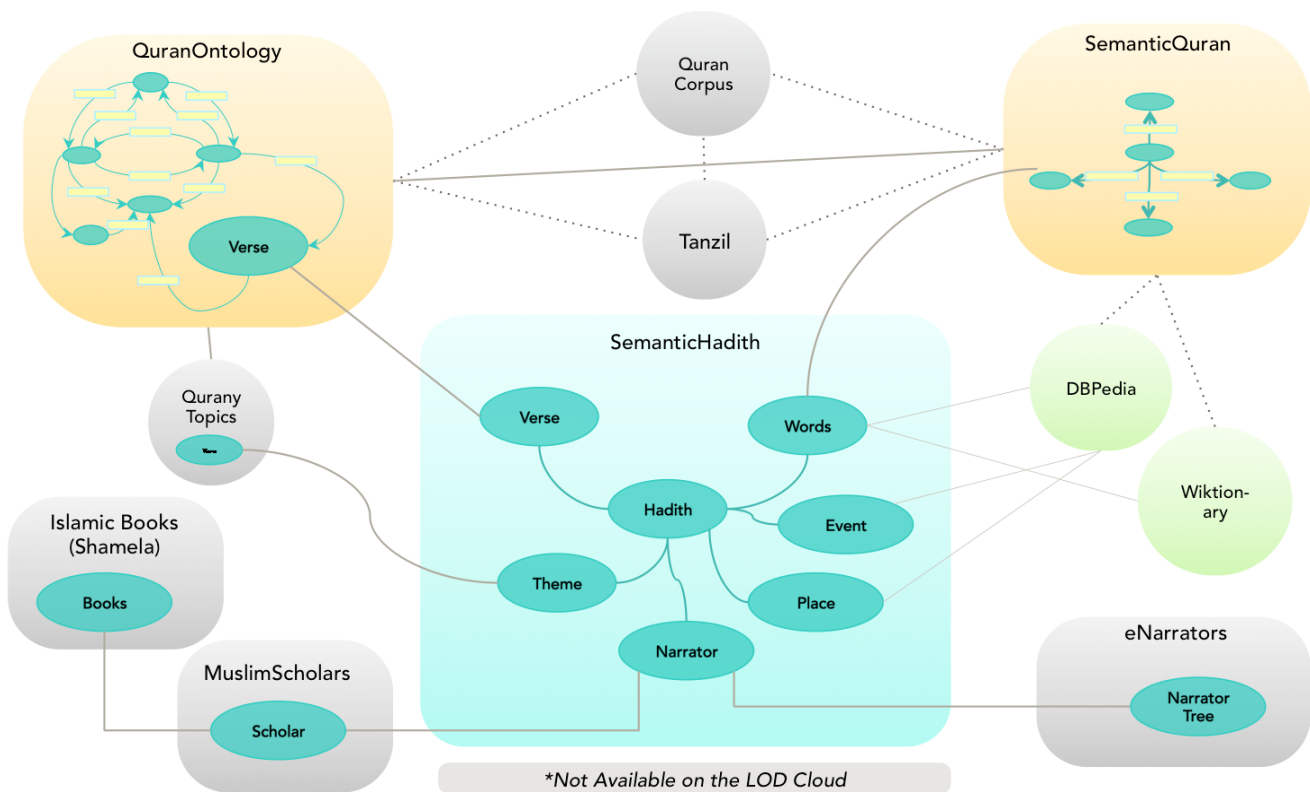


Figure 7: A view of the proposed and available Linked Data Cloud for Islamic Knowledge Sources

the Qur'an. While some may be as long as an entire page's length of a standard book size, others may be as short as one or two words. Therefore, in order to determine, whether the verse is actually being quoted or cited in a hadith requires further validation. Even applying a threshold, relative to the length of the verse, is not an optimal solution. Setting a substantial minimal length was considered, but this may not guarantee a comprehensive coverage. For the first prototype, only 1,325 expert validated links were asserted. In the sunnah.com data, these links may be found as hyperlinks to the verses on the site quran.com.

In addition, similarity computation algorithms were devised to extract *Hadith-Hadith* similarity relations. The 4,973 relations, listed in Table 2 are strongly similar Hadith that have at least 60% of text in common. However, the challenge with this approach is that, it cannot be distinguished if the similarity is in the *Sanad* or the *Matan* or both. The more meaningful similarities that are of interest are in the *Matan* of the hadith. In future experiments, we aim to segment the *Sanad* and the *Matan* and extract respective similarity relations. While the similarity threshold for the current approach only took into consideration the common substring, we plan to conduct experiments with more meaningful similarity measures such as Cosine, Jaccard and Pearson correlation coefficient, as done in our work for Qur'anic verses [8].

#### 4.2.3 Results

Based on the dataset and experiments carried out, we summarize some of the dataset and link statistics in the

Tables 1 and 2. Table 1 summarizes the statistics for some of the key entities present in the dataset.

Table 2 provides the raw count for the candidate relations extracted under the different categories mentioned. It must be noted however, that the relations are not classified according to any of the parameters mentioned in the design. It is also worth mentioning that some of these relations may actually be symmetric.

Table 1: Entity Statistics in the Semantic Hadith Dataset(Sunnah.com)

Entity	Count
No of Collections	11
No of Books	311
No of Hadith(Arabic)	25,934
No of Hadith(English)	18,040
No of Chapters	8,968

Table 2: Link Statistics in the Semantic Hadith Dataset

Link Type	Count
Hadith-Hadith Relation	4,973
Hadith-Verse Relation(Citations)	1,325
Verse-Hadith Relation(Scholarly)	313

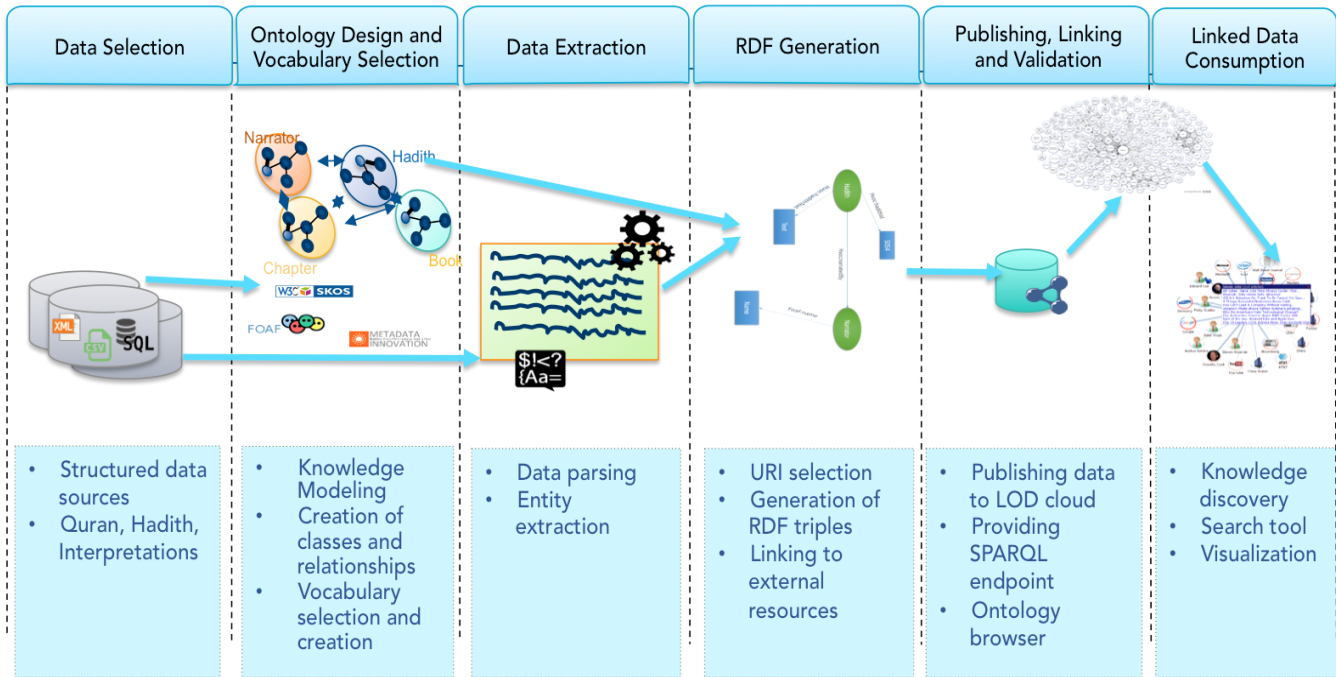


Figure 8: Linked Data Generation and Publishing Framework for Semantic Hadith

### 4.3 Existing Limitations and Proposed Solutions

Some of the key limitations we face are with respect to link extraction and validation. There is an obvious lack of structured knowledge sources, with well marked citations. Therefore, the Verse-Hadith links are extremely difficult to be extracted using mere computational means. Human contribution is a must. For this purpose, we intend to pursue a crowdsourcing approach, based on our prior work[6]. We not only intend to use crowdsourcing and human computation methods for the purpose of knowledge acquisition, but also for knowledge validation. Infact, we believe a hybrid human-machine computation methodology to be the only indispensable means of being able to fulfill the vision for linked Islamic knowledge at scale, while ensuring the desired reliability and authenticity.

## 5. PROSPECTIVE APPLICATIONS

The most significant benefit of realizing the linked data vision for Islamic knowledge sources will be towards enabling semantics driven distributed knowledge search and retrieval. Most current applications in the Islamic domain only provide limited provision for semantic and conceptual search and retrieval beyond the traditional keyword based searches, upon a single repository. With the Semantic Hadith model, the first of its kind tools will now be possible that would let Qur'an and Hadith repositories to be queried and searched in a federated manner.

The given listing illustrates a federated query between the Semantic Hadith and Semantic Quran datasets. Given that a Verse-Hadith relation exists with the Semantic Hadith dataset, this query retrieves the arabic and english texts for the respective verse.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX hvoc: <http://www.hadith.islamicinformatics.org/SemanticHadith#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX qvoc: <http://mlode.nlp2rdf.org/quranvocab#>
```

```
select ?hadith_text ?surahNo ?verseNo
      ?ayahText ?ayahEng
WHERE {
  ?verse hvoc:isRelatedTo ?hadith;
    hvoc:verseNo ?verseNo ;
  hvoc:surahNo ??surahNo .
  ?hadith hvoc:hadithId ?hId;
  hvoc:hadithText ?hadith_text .
```

```
SERVICE <http://mlode.nlp2rdf.org/sparql> {
  ?s qvoc:chapterIndex ?surahNo;
    qvoc:verseIndex ?verseNo;
  rdfs:label ?ayahText;
  rdfs:label ?ayahEng.
  FILTER (lang(?ayahEng) ="en" &&
    lang(?ayahText) ="ar")
}}
```

This could be taken to another level, by adding another level of federation, and querying the themes of the verse from the QuranOntology.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX hvoc: <http://www.hadith.quraninformatics.org/SemanticHadith#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX qur: <http://quranontology.com/Resource/>
```

```
select ?hadith_text ?surahNo ?verseNo
```



```

?tname
WHERE {
?verse hvoc:isRelatedTo ?hadith;
  hvoc:verseNo ?verseNo ;
hvoc:surahNo ?surahNo .
?hadith hvoc:hadithId ?hId;
hvoc:hadithText ?hadith_text .

SERVICE <http://quranontology.com/Query> {
?verse qur:DiscussTopic ?t.
?t rdfs:label ?tname.
FILTER(LANGMATCHES(LANG(?tname), "ar"))
}}

```

This could be further enhanced by automated interlinking with other available datasources on the linked data cloud, as envisioned in Figure 7. For instance, once the available hadith are annotated with mentioned events, place or people, they may be linked to the available entities in dbpedia. This would enable richer knowledge discovery and retrieval for a range of applications.

We expect that using this model, more hadith and Qur’anic exegesis repositories, that also rely on and cite heavily the hadith sources, will be published in the linked data format. This will enable the design and development of enhanced learning tools for the Islamic domain, which will provide efficient and personalized access to primary sources of knowledge, ensuring reliability and authenticity. Given that these tools will give better access to meaningfully interlinked knowledge, it will require less effort to find resources and access knowledge beyond books. More content, both classical and contemporary, would become discoverable.

## 6. RELATED WORK

The linked data approach has emerged as the de facto standard for sharing the data on the web. It provides a set of best practices for publishing and connecting structured data on the web [9]. The linked data design issues provide guidelines on how to use standardized web technologies to set data-level links between data from different sources [23]. Increased interest in the LOD has been seen in various sectors e.g. Education [11], [31], Scientific research [3], libraries [28], [25], Government [12], [24], [33], Cultural heritage [26] and many others, however, the religious sector has yet to cache upon the power of the linked open data.

Research in computational informatics applied to the Islamic knowledge has primarily centered around Morphological annotation of the Qur’an [13], [14], Ontology modeling of the Qur’an [1], [5], [15], [36], [37], and Arabic Natural language processing [15]. The LOD take-up in the area of Islamic knowledge has been particularly extremely limited. As mentioned earlier, there have been some recent efforts to publish Islamic knowledge as linked data on the Linked Open Data (LOD) cloud. The efforts primarily focus on the Qur’an. The two datasets that we consider in our research and attempt to link with our Semantic Hadith research include *SemanticQuran*<sup>15</sup> [34] and *QuranOntology*<sup>16</sup> [16].

Much of the work in the Hadith sciences has focused automating the extraction of the Chain of Narrators. Some

works in this regard include [20], [18], [17], [19], [21]. There are also work references with respect to mining the hadith for indexing and classification [2], [29]. Some recent efforts have attempted to model the hadith as semantic ontologies [4] [32]. However, the efforts have focused on annotating the different constituents of the hadith. None of these datasources are available as open source.

Our work is the first of its kind to propose the linked data based model to propose the linking of hadith with the Qur’an. This linked knowledge forms a vital backbone to enable better integration and discovery of knowledge sources.

## 7. CONCLUSIONS AND FUTURE WORK

In this paper we presented the design and development of our Semantic Hadith framework, which aims to provide the foundation for semantically interlinking the most important Islamic knowledge sources using the linked data standards. We presented the design of the Semantic Hadith Ontology and explained the nature of links with other data sources. The implementation still needs to be matured. The validation of the links and extracted knowledge is a huge challenge we are looking into. We are investigating into crowdsourcing models for knowledge acquisition and validation at scale.

## 8. ACKNOWLEDGEMENTS

We are thankful to the authors of sunnah.com for providing us with the valuable datasource to carry out this research. We would also like to acknowledge the efforts of Mr. Muhammad Shoaib (Jeju National University, South Korea) in extending his help with some of the experiments.

## 9. REFERENCES

- [1] H. S. Al-Khalifa, M. Al-Yahya, A. Bahanshal, I. Al-Odah, and N. Al-Helwah. An approach to compare two ontological models for representing quranic words. In *Proceedings of the 12th International Conference on Information Integration and Web-based Applications and Services*, pages 674–678. ACM.
- [2] K. A. Aldhlan and A. M. Zeki. Datamining and islamic knowledge extraction: alhadith as a knowledge resource. In *Information and Communication Technology for the Muslim World (ICT4M), 2010 International Conference on*, pages 21–25. IEEE, 2010.
- [3] T. K. Attwood, D. B. Kell, P. McDermott, J. Marsh, S. Pettifer, and D. Thorne. Utopia documents: linking scholarly literature with research data. *Bioinformatics*, 26(18):568–574, 2010.
- [4] A. Azmi and N. B. Badia. itree-automating the construction of the narration tree of hadiths (prophetic traditions). pages 1–7, 2010.
- [5] S. Baqai, A. Basharat, H. Khalid, A. Hassan, and S. Zafar. Leveraging semantic web technologies for standardized knowledge modeling and retrieval from the holy qur’an and religious texts. In *Proceedings of the 7th International Conference on Frontiers of Information Technology*, FIT ’09, pages 42:1–42:6, New York, NY, USA, 2009. ACM.
- [6] A. Basharat, I. B. Arpinar, S. Dastgheib, U. Kursuncu, K. Kochut, and E. Dogdu. Semantically enriched task and workflow automation in crowdsourcing for linked data management.

<sup>15</sup><http://datahub.io/dataset/semanticquran>

<sup>16</sup><http://www.quranontology.com>

- International Journal of Semantic Computing*, 8(04):415–439, 2014.
- [7] A. Basharat, K. Rasheed, and I. B. Arpinar. Towards linked open islamic knowledge using human computation and crowdsourcing. In *Proceedings of the International Conference on Islamic Applications in Computer Science And Technology*, 2015.
- [8] A. Basharat, D. Yasdansepas, and K. Rasheed. Comparative study of verse similarity for multi-lingual representations of the qur’an. In *Proc. on the Int. Conference on Artificial Intelligence (ICAI)*, pages 336–343, 2015.
- [9] C. Bizer, T. Heath, and T. Berners-Lee. Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3):1–22, 2009.
- [10] D. Brickley and L. Miller. Foaf vocabulary specification 0.98. *Namespace document*, 9, 2012.
- [11] S. Dietze, S. Sanchez-Alonso, H. Ebner, H. Q. Yu, D. Giordano, I. Marenzi, and B. P. Nunes. Interlinking educational resources and the web of data a survey of challenges and approaches. *Program-Electronic Library and Information Systems*, 47(1):60–91, 2013.
- [12] L. Ding, T. Lebo, J. S. Erickson, D. DiFranzo, G. T. Williams, X. Li, J. Michaelis, A. Graves, J. G. Zheng, Z. Shangguan, J. Flores, D. L. McGuinness, and J. A. Hendler. Twc logd: A portal for linked open government data ecosystems. *Journal of Web Semantics*, 9(3):325–333, 2011.
- [13] K. Dukes, E. Atwell, and A. M. Sharaf. Syntactic annotation guidelines for the quranic arabic dependency treebank. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta*, 2010.
- [14] K. Dukes and N. Habash. Morphological annotation of quranic arabic. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta*, 2010.
- [15] A. Farghaly and K. Shaalan. Arabic natural language processing: Challenges and solutions. *ACM Transactions on Asian Language Information Processing*, 8:1–22, 2009.
- [16] A. Hakkoum and S. Raghay. Ontological approach for semantic modeling and querying the qur’an. In *Proceedings of the International Conference on Islamic Applications in Computer Science And Technology*, 2015.
- [17] F. Harrag. Text mining approach for knowledge extraction in sahih al-bukhari. *Computers in Human Behavior*, 30:558–566, 2014.
- [18] F. Harrag, E. El-Qawasmeh, and A. M. S. Al-Salman. *Extracting Named Entities from Prophetic Narration Texts (Hadith)*, pages 289–297. Springer, 2011.
- [19] F. Harrag and A. Hamdi-Cherif. Uml modeling of text mining in arabic language and application to the prophetic traditions hadiths. pages 11–20, 2007.
- [20] F. Harrag, A. Hamdi-Cherif, and E. El-Qawasmeh. Information retrieval architecture for hadith text mining. *Journal of Digital Information Management*, 6(6), 2008.
- [21] F. Harrag, A. Hamdi-Cherif, and E. El-Qawasmeh. Vector space model for arabic information retrieval – application to hadith indexing. In *Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First International Conference on the*, pages 107–112. IEEE, 2008.
- [22] S. Hasan. *An introduction to the science of Hadith*. Al-Quran Society, 1994.
- [23] T. Heath and C. Bizer. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1):1–136, 2011.
- [24] J. Hendler, J. Holm, C. Musialek, and G. Thomas. Us government linked open data: semantic. data. gov. *IEEE Intelligent Systems*, 27(3):0025–31, 2012.
- [25] L. C. Howarth. Frbr and linked data: Connecting frbr and linked data. *Cataloging and Classification Quarterly*, 50(5-7):763–776, 2012.
- [26] J. Marden, C. Li-Madeo, N. Whysel, and J. Edelstein. Linked open data for cultural heritage: evolution of an information technology. pages 107–112, 2013.
- [27] A. Miles, B. Matthews, M. Wilson, and D. Brickley. Skos core: Simple knowledge organisation for the web. In *Proceedings of the 2005 International Conference on Dublin Core and Metadata Applications: Vocabularies in Practice*, DCMI ’05, pages 1:1–1:9. Dublin Core Metadata Initiative, 2005.
- [28] E. Miller and M. Westfall. Linked data and libraries. *The Serials Librarian*, 60(1-4):17–22, 2011.
- [29] M. Naji Al-Kabi, G. Kanaan, R. Al-Shalabi, S. I. Al-Sinjilawi, and R. S. Al-Mustafa. Al-hadith text classifier. *Journal of Applied Sciences*, 5:584–587, 2005.
- [30] A. A. B. Philips. *Usool at-Tafseer: The Methodology of Qur’aanic Explanation*. AS Noordeen, 2002.
- [31] N. Piedra, E. Tovar, R. Colomo-Palacios, J. Lopez-Vargas, and J. A. Chicaiza. Consuming and producing linked open data: the case of opencourseware. *Program: electronic library and information systems*, 48(1):16–40, 2014.
- [32] Y. M. D. Rebhi S. Baraka. Building hadith ontology to support the authenticity of isnad. *International Journal on Islamic Applications in Computer Science And Technology*, 2(1):25–39, 2014.
- [33] N. Shadbolt, K. O’Hara, T. Berners-Lee, N. Gibbins, H. Glaser, and W. Hall. Linked open government data: Lessons from data. gov. uk. *IEEE Intelligent Systems*, 27(3):16–24, 2012.
- [34] M. A. Sherif and A.-C. N. Ngomo. Semantic Quran - a multilingual resource for natural-language processing. *Semantic Web*, 6(4):339–345, 2015.
- [35] S. Weibel, J. Kunze, C. Lagoze, and M. Wolf. Dublin core metadata for resource discovery. Technical report, 1998.
- [36] A. R. Yauri, R. A. Kadir, A. Azman, and M. A. A. Murad. Quranic-based concepts: Verse relations extraction using manchester owl syntax. In *Information Retrieval and Knowledge Management (CAMP), 2012 International Conference on*, pages 317–321. IEEE.
- [37] A. R. Yauri, R. A. Kadir, A. Azman, and M. A. A. Murad. Quranic verse extraction base on concepts using owl-dl ontology. *Research Journal of Applied Sciences, Engineering and Technology*, 6(23):4492–4498, 2013.