IUG Journal of Natural and Engineering Studies
Peer-reviewed Journal of Islamic University-Gaza

ISSN 2409-4587

IUGNES
Vol 24, No 1, 2016, pp 39-53

An Ontology Based Approach to Enhance Information Retrieval from Al-Shamelah Digital Librar

Iyad M. AlAgha[1,*], Mohammed Gh. Al-Masri[2]

[1]Department Software Development, Faculty of Information Technology, Islamic University of Gaza, Gaza Strip, Palestine

[2]Palestine Technical College, Deir El-Balah

Received on (01-09-2015) Accepted on (23-11-2015)

Abstract

With the huge number of Islamic resources that emerged over hundreds of years, several difficulties were introduced when searching in this huge heritage. This has challenged developers to build computer applications to facilitate information retrieval from Islamic Resources. One of these applications is Al-Shamelah Digital Library (ADL), Al-maktabah Al-Shamelah, which is a huge database containing thousands of books in different disciplines. The search facility offered by ADL is mainly based on keyword matching, and does not provide semantic interpretations of Islamic texts. It also does not handle complex queries or extract implicit relations and meanings from text. Driven by these challenges, this work presents OntoADL, a system that supports semantic search over a section of Al-Shamelah digital library. At the core of OntoADL is our approach that leverages ontology-based annotations to produce highly relevant search results and to offer recommendations of related topics. The design and architecture of OntoADL is discussed, focusing on how ontology-based reasoning can result in intelligent results that meet the user's interests. The search service in OntoADL was evaluated by being compared with the search service in the conventional ADL. The OntoADL achieved 83% recall and 66% precision while the ADL system achieved 70% recall and 36% precision.

Keywords Semantic Web, Ontology, SPARQL, Semantic Annotation, Islamic Resources.

طريقة معتمدة على الأنطولوجيا لتحسين استخراج المعلومات من المكتبة الشاملة الرقمية

ملخص

مع وجود عدد كبير من المصادر الإسلامية التي ظهرت على مدى مئات السنين، ظهرت عدة صعوبات عند البحث في هذا التراث الضخم، وقد شكل هذا تحديا للمطورين لبناء تطبيقات حاسوبية لتسهيل استرجاع المعلومات من المصادر الإسلامية. واحد من هذه التطبيقات هي المكتبة الشاملة الرقمية، وهي قاعدة بيانات ضخمة تحتوي على آلاف الكتب في مختلف التخصصات. خدمة البحث التي يقدمها تطبيق المكتبة الشاملة يستند أساسا على مطابقة الكلمات الرئيسية، ولا يقدم تفسيرات دلالية للنصوص الإسلامية. كما أنه لا يتعامل مع الاستفسارات المعقدة أو يستخرج العلاقات الضمنية والمعاني من النصوص. انطلاقاً من التحديات المشار إليها، يقدم هذا العمل OntoADL، وهو نظام يدعم البحث الدلالي على قسم من المكتبة الشاملة الرقمية. يبنى OntoADL على طريقة معتمدة على الشروح الدلالية المستندة على أنطولوجيا لإعطاء نتائج بحث ملائمة وتقديم توصيات من المواضيع ذات الصلة. يتم شرح تصميم وهيكلية نظام OntoADL، مع التركيز على توضيح كيف أن الاستدلال على أساس الأنطولوجيا يمكن أن يؤدي إلى نتائج ذكية تلبي حاجات المستخدم. تم تقييم خدمة البحث في OntoADL عن طريق مقارنتها مع خدمة البحث التقليدية في المكتبة الشاملة الرقمية. حقق OntoADL نتيجة 83%استدعاء و66% دقة فيما حقق البحث التقليدي نتيجة 70% استدعاء و36% دقة.

كلمات مفتاحية: الويب الدلالي، الأنتولوجيا، SPARQL، الشرح الدلالي، المصادر الإسلامية.

* Corresponding author e-mail address: ialagha@iugaza.edu.ps

1. Introduction:

Al-Shamelah Digital Library (ADL) is a huge electronic database that contains thousands of books in different disciplines including doctrine, hadith interpretation, Prophet biography, jurisprudence (Fiqh), history, genealogy, literature and rhetoric. It is regarded as a fundamental source of Islamic Knowledge due to its comprehensiveness and diversity. ADL software has quickly become popular among students and researchers due to several reasons: (a) it contains a huge number of E-books, (b) it is structured into branches covering the various Islamic religious sciences, (c) it is free to use for academic purposes, and (d) it offers different tools for information search including searching for book titles, searching inside books as well as advanced searching capabilities [1]. However, the search tool offered by ADL is merely based on syntactic search which has many deficiencies: If the search keywords are not explicitly present in the text, no results will be retrieved. In addition, the ADL search service is not currently able to perform deep inference on the library content or recognize synonyms and semantically-related terms. For example, searching for the words “prophet” and “messenger” often gives different results although the two words are semantically related.

The need to support semantic search in ADL is highly demanded due to the unique characteristics of the Islamic resources, e.g. Quran and Sunna. In the following, these unique characteristics are discussed, considering how the current ADL search fails to address them:

a) Islamic texts comprise meanings, intents and metaphors that cannot be easily understood. The unique style and allegorical nature of Islamic texts needs special attention about searching and information retrieval issues [2]. The legacy keyword searching techniques in ADL are not suitable for searching in the Islamic resources. It is necessary to support information search at the semantic level so that ordinary users can reveal the underlying meanings and thoughts.
b) Islamic knowledge is known to be very diverse and interrelated: Information from one discipline, such as the Hadith of Prophet Mohammed, peace be upon him, cannot be fully understood without relating it to other disciplines such as Aqidah ”العقيدة” and Feqh "الفقه". The traditional search in ADL does not reveal the relationships or dependencies between search results.
c) The Islamic terminology includes legacy terms that are difficult to memorize or are not in common everyday usage. However, ordinary users may know familiar synonyms of the target term but without memorizing the term itself. Using these synonyms for searching may not return the desired results as the current search in ADL lacks the ability to recognize synonyms or semantically-related words.

The above characteristics make the traditional search techniques, which are widely used in other digital libraries, insufficient for the particular needs of the ADL users.

To illustrate how the lack of semantics may be problematic in ADL, assume that a user wants to search for all the sayings of the Prophet Mohammed that cover the topic of “backbiting” (الغيبة). By inputting the keyword “الغيبة”, the ADL search will only retrieve the texts that explicitly mention the word “الغيبة”. However, it fails to retrieve many other texts that cover the same concept but by using different words. Take the following saying as an example: "مَرَرْتُ بِقَوْمٍ لَهُمْ أَظَافِرُ مِنْ نُحَاسٍ يَخْمِشُونَ وُجُوهَهُمْ وَصُدُورَهُمْ فَقُلْتُ: مَنْ هَؤُلاءِ يَا جِبْرِيلُ؟ فَقَالَ: هَؤُلاءِ الَّذِينَ يَأْكُلُونَ لُحُومَ النَّاسِ وَيَقَعُونَ فِي أَعْرَاضِهِمْ", the core subject of this saying is backing “الغيبة”, even though it is not explicitly mentioned.

The Semantic Web can provide solutions to the previous problems by supporting ontology based reasoning. Since the emergence of Semantic Web, research in digital libraries has widely considered the use of ontologies to empower information retrieval with semantics [3]. This work builds on previous efforts on semantically-enhanced digital libraries, and aims to particularly enhance information search in ADL. It presents an ontology based approach to incorporate semantics into the ADL content and then support semantic search. This approach can be summarized as follows:

Of the entire content of ADL, the domain of the Prophetic Medicine (Tibb An-Nabawi) was chosen as a domain of knowledge to support semantic search. We constructed a domain ontology that models the concepts of the Prophetic Medicine as well as the relations between these concepts. Afterwards, semantic annotation was manually performed to link words and phrases from the sayings of Prophet Mohammed with relevant ontology terms. The resultant semantic annotations constitute the RDF store from which the search agents retrieve and infer about results. On top of our RDF store, we built OntoADL, our search application that takes keywords from the user as input and transforms them to SPARQL, the formal query language of the Semantic Web. The generated SPARQL queries are then executed to retrieve relevant results.

This paper is organized as follows: in the following section the related work is discussed. The processes of ontology construction and semantic annotation are then explained. The OntoADL system is then presented, focusing on how semantic annotations can be queried and utilized to provide intelligent results that go beyond the results obtained by traditional keyword-based search. Finally, the evaluation of the OntoADL system is discussed.

2. Related Work:

In recent years, the research on the processing of Arabic language has increasingly gained attention at both syntactic and semantic levels. However, the work on Islamic knowledge processing and retrieval has received little attention. At the syntactic level, some prominent efforts have been proposed including the development of Quranic dependency Treebank [4], morphological segmentation and part of speech tagging of Quran [5] and extraction of key-phrases [6].

A plenty of applications and websites are publically available to offer services such as Fatwa [7], Islamic e-book download [8], and interpretations of Quran and hadith [9, 10]. These applications are accompanied with search services to enable their users to access their content. However, retrieval methods in these applications lack semantic information, and are mostly based on keywords matching approach.

Few studies were conducted to link the Islamic Knowledge and Semantic values. Some efforts focused on the construction of ontologies covering specific topics of Quran. For example, Khan et al. [11] designed an ontology to represent living creatures including animals and birds mentioned in the Holy Quran. They also proposed a model for a semantic search system based on their ontology. Sharaf and Atwell [12] designed an ontology to model the Quran concepts. Then, they used the ontology to annotate named entities in verses, such as the names of historic people and places mentioned in the Quran. Al-Yahya et al. [13] proposed a computational model for representing Arabic lexicons using ontologies, and applied it on “Time nouns” vocabulary of the Holy Quran. Iqbal et al. [14] reported on the development of an ontology for Juz’ Amma (جزء عم) that encapsulates contextual information support, which are the translations, revelations place, tafsir, and hadiths. The work of Saad et al. [15, 16] focused on creating an ontology that covers verses related to the topic of salat (obligatory prayers). Al-Khalifa et al. [17] presented a framework for recognizing and identifying semantic opposition terms in Quran by using natural language processing armed with domain ontologies.

The previous efforts generally focused on the ontological representation of Quranic concepts but did not use the ontologies in practice for semantic search. In addition, they did not demonstrate how their ontologies can enhance search results or can be validated from the perspective of end users.

Some researchers explored ways to support information retrieval from Islamic Resources. Shoaib et al. [18] proposed and tested a model that exploited WordNet relationships to support semantic search in Quran. However, the advantage of their model was restricted to discovering WordNet relations such as synonym and hyponymy. They did not build an ontology to represent Quran concepts, thus were not able to capture the contextual relations in Quran. Salim et al. [19] presented a framework for the development of Islamic ontology that advocates the symbiosis of thesaurus and domain expert. They mapped knowledge sources such as documents, reports, etc. into the domain ontology to enable better knowledge indexing and searching process. However, they did not present any details related to the implementation and evaluation of their framework. Yauri et al. [20] proposed a system that makes use of an ontology to model the relations between verses in a part of Quran. Search results retrieved by the system allow to explore contextual relations between verses. They claimed that the users can query the Quran knowledge in natural language, although they did not explain the approach used to interpret user queries to formal semantics.

In the domain of Hadith, few efforts proposed ontology based approaches to judge the correctness of Hadith narrations. For example, Azmi and Bin Badia [21] proposed e-Narrator, an application that parses a plain Hadith text and automatically generates the tree of narration chain. Baraka and Dalloul [22] built an ontology-based system that automatically generates a suggested judgment of Hadith Isnad. It is based on the rules that Hadith scholars follow to produce a suggested judgment. This work also uses the hadith domain but with a different goal, which is supporting semantic search in Al-Shamela library.

3. Usage Scenarios:

Before illustrating the design and architecture of OntoADL in detail, an overview of its functionality is presented through a usage scenario. The intention is to show how the search service can be improved by the added semantics. Figure 1 shows the main window of the OntoADL search tool. The domain of search consists of sixty sayings of the Prophet Mohammed extracted from the book of “Al-tib Al-Nabawi”. Imagine a user wants to gather information about the topic of cupping (الحجامة). He typed the word “ الحجامة” in the search field and clicked the search button. The results window is split into two panes as shown in Figure 1. On the right pane, all the sayings that include the word “ الحجامة”, or any of its derivatives or synonyms, are displayed, and the matching text is highlighted. The left pane shows a list of recommended results that are inferred by the system based on the underlying semantics. For example, the list of recommended results includes:

1) references to all diseases that can be cured by the cupping (e.g. "وجع في البطن", "التهاب رئوي", "جرح").
2) Position of cupping ("مكان أو موضع الحجامة").
3) Cost of cupping ("أجرة الحجامة").
4) The category of cures to which the cupping belongs: ("الطب الوقائي").

associations between the term "cupping" and other related terms.

It is obvious from the above scenario that OntoADL enables the user to gain a broad overview of the topic of cupping. Through the offered links to recommended and related topics, the user will be able to explore the topic of cupping in detail, and understand how it relates to other topics in the domain, a thing that is not easily possible by using the conventional ADL search.

Another usage scenario that illustrates other benefits of the OntoADL search tool is shown in Figure 2. It shows the results of searching for the topic "cure of Pneumonia" (علاج الإلتهاب الرئوي). Retrieved result consists of a single saying which is " عليكم بهذا العود الهندي فإن فيه سبعة أشفية يستعط من العذرة ويلد به من ذات الجنب به ". The phrase "العود الهندي" (Indian lute) in the result is highlighted in a different color. Notice that the given result does not contain any of the searched keywords (علاج الإلتهاب الرئوي). In fact, none of the Prophetic sayings contain these keywords. However, the system retrieved a result referring to the "Indian Lute" which is used as remedy for Pneumonia according to the Prophetic Medicine. The list of recommendations of the left pane includes links to other diseases that can be cured by the Indian lute. These suggestions were built based on

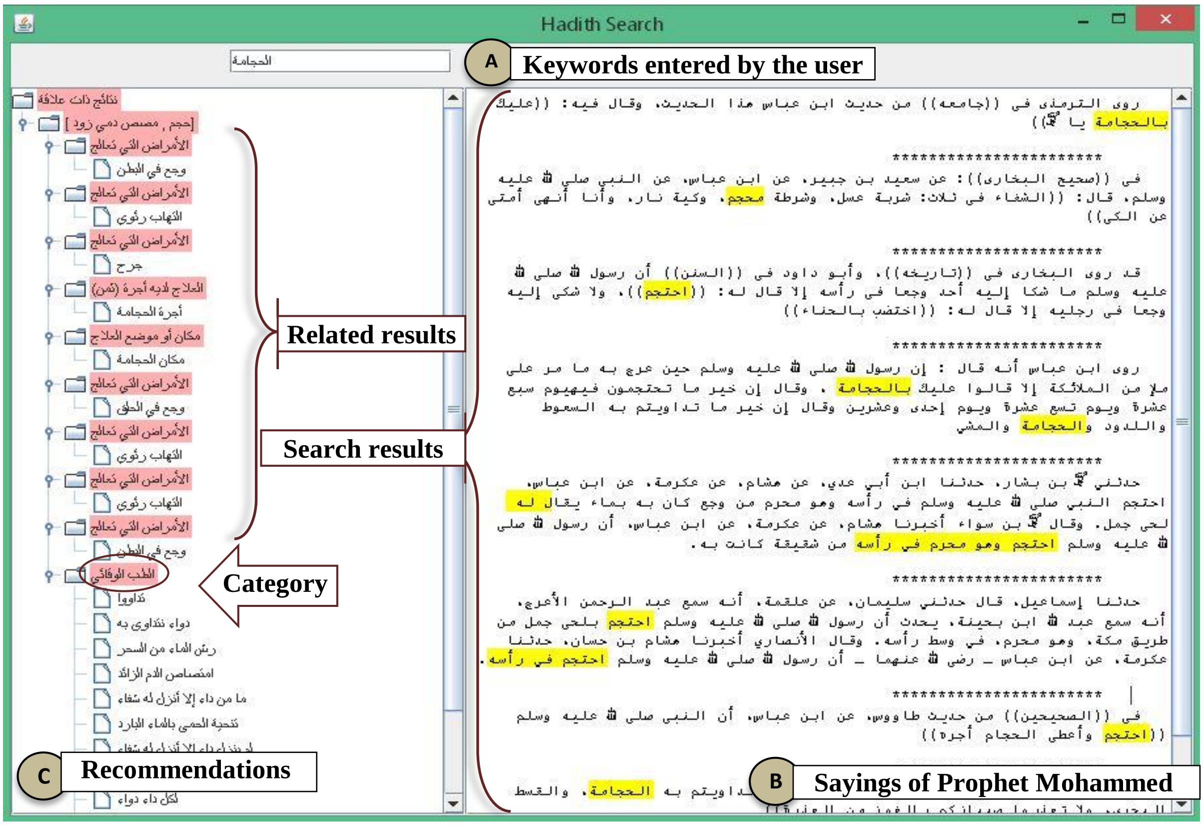

Figure 1 *The main window of the OntoADL showing the results of searching for the keyword: "الحجامة"*

Note that the recommended results are presented in a tree structure, and each result is an active link. By clicking on each link, the two panes of the window will be updated to display results related to the new topic. For example, clicking on the tree node labelled as "أجرة الحجامة" will cause the system to search for and display the sayings of the Prophet that refer to the cost of cupping. The left pane will be also updated to show a list of recommendations related to the new topic. These recommendations are selected based on the semantic annotations that link the text of the sayings with entities in the domain ontology. Based on these annotations, the system will infer

Figure 2 *The main window of OntoADL showing the results of searching for the keyword: "علاج الالتهاب الرئوي"*

the underlying semantics that link the term "Indian Lute" with other ontology entities. The list of suggestions also shows the general category of the term "Indian lute", which is the "Trees" category, and other related remedies that belong to the same category. The parent and sub categories of any term are captured based on the concept hierarchy defined in the ontology.

4. The OntoADL Architecture:

Figure 3 shows the architecture of the OntoADL system. The system process is briefly explained as the following: The end user interacts with the OntoADL system by inputting keywords through a dedicated user interface (see Figure 1 and2). User keywords are then handled by the Linguistic Module where they undergo a set of preprocessing steps. Keywords are then matched with the ontology content by the Matcher module. The output of the Matcher are the ontology entities, i.e. classes, properties, instances and literal values, that correspond to the user keywords. These entities are then transferred to the SPARQL query generator which is responsible for building SPARQL queries. It uses the input ontology entities to generate queries that, when executed, will retrieve results related to the user needs.

At the core of the system is the domain ontology that models the types, properties and interrelationships of the entities that exist in the domain of the Prophetic Medicine. The design of the domain ontology is briefly discussed in the next section.

The knowledge base component maintains all information that the OntoADL searches in when a search task is initiated. Information stored in the knowledge base are of two types: 1) Original content of ADL: this information refers to the Islamic content of ADL which we aim to enrich by means of semantic annotation. In our prototype system, we extracted and used only sixty Prophetic sayings from the book of "Al-Tib Al-Nabawi". 2) Semantic annotations: Semantic annotations are the ontology-based meta-data attached to the original content of ADL. They can be seen as the glue that ties the ontology into the ADL content. The annotation process in OntoADL is currently handled manually though the Annotation Module. A human expert should annotate in the texts of ADL all mentions of instances relating to concepts in the ontology. In the following sections, the components of the ADL system are explained in detail.

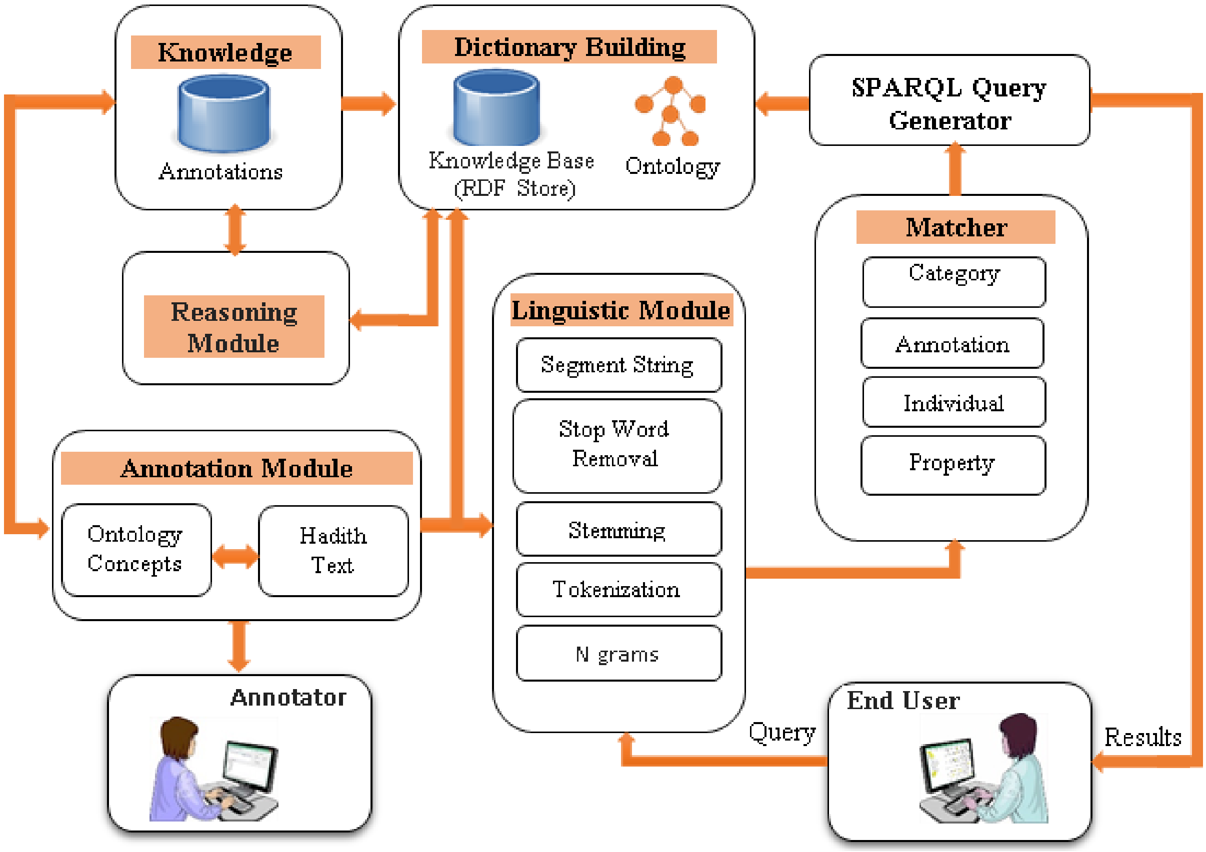

Figure 3 *The OntoADL Architecture*

5. Ontology of Prophetic Medicine:

The ontology we built to represent the domain of the Prophetic Medicine is briefly described in what follows. We think that the approach used to build the ontology can guide the modelling of other domains in the ADL. Note that only basic parts of the ontology that are essential to understand the rest of this paper are discussed. For more details about the ontology, the reader is referred to the full ontology which can be accessed online [23].

The Prophetic Medicine refers to the words and actions of the Prophet with a bearing on diseases, treatments of diseases, and care of patients. The first step of the ontology construction is to elicit and structure the ontology classes. For this purpose, two domain experts were involved in the ontology acquisition process. The ontology was constructed by using Protégé [24], an open-source ontology editor, and was written in OWL 2 DL [25]. Ontology classes were mainly derived either from the names of chapters and sections of the book “Al-Teb Al-Nabawi” and from the sayings of the Prophet. Figure 4 depicts a part of the class hierarchy of the ontology: The top two classes of the ontology are the “Disease” and the “Cure” classes. The Disease class has three sub-classes that are: the “Spiritual Disease”, the “Curative disease” and the “Preventive disease” classes. The Cure class has the following subclasses: the “Preventive Medicine”, the “Curative Medicine” and the “Spiritual remedy” classes. These classes are briefly explained in Table 1.

An ontology concept can be referred by multiple names. For example, the term “مرض” and "داء" have the same meaning. Therefore, it is necessary to expand the system’s terminology so that it can recognize the variants of search terms. Synonyms of all ontology concepts were stored in the ontology. The rdfs:label property was used to associate each ontology class with all its synonyms. Each search keyword entered by the user will be matched with all the synonyms of ontology classes.

Table 1 *Top level classes of the ontology of the Prophetic Medicine*

Class	Description
Preventive diseases	Refers to psychosomatic diseases that could respond to spiritual cures.
Curative diseases	Refers to diseases that are treatable by natural remedies
Spiritual disease	Refers to diseases that could respond to spiritual remedies such as envy, magic.

Preventive medicine	Refers to preventive and therapeutic prescriptions
Curative medicine	Refers to natural remedies
Spiritual Cures	Refers to spiritual remedies such as Prayer and recitation of the Qur'an.

The second step of the ontology construction is to define properties and bind them to appropriate classes. Table 2 shows the main object properties we defined to model the relations between classes. The domain and range of each property are illustrated. For example, the property "cures" links instances of type "Cure" with instances of type "Disease". The property "hasType" links instances of type "Cure" or "Disease" with their formal categories. Categories of different cures and diseases are defined in the ontology by using the class "Category". For example, the term "Fever"(الحمي) has the category "Infectious Disease (مرض معدي)".

The final step in the ontology construction process is to create instances from the ontology classes and link them by using appropriate object properties. These instances were delimited after studying sixty sayings related to the topic of Prophetic Medicine.

A total number of ninety instances were created to represent different ontology classes. For example, the term "الحبة السوداء" (Black cumin) in the saying " عليكم بالحبة السوداء ..." is an instance of type "Seeds", which is a subclass of the top-level class "Cure". The term "العُذْرَةِ" (مرض يصيب الحلق) in the saying "وَلَا تُعَذِّبُوا صِبْيَانَكُمْ بِالْغَمْزِ مِنْ الْعُذْرَةِ" is an instance of the class "Throat_Disease". After creating the instances, we linked them using the appropriate properties from Table 2. For example, the instance "الحجامة" (Cupping) is linked to the instance "Headache" (الصداع) using the property "cures".

To give a glance of how ontology instances were created and interlinked, consider the Prophetic saying show in Figure 5: Circles in the figure denote some ontology classes, while the rounded rectangles denote ontology instances related to the given text. Properties used to link different instances, from the perspective of the human expert, were also depicted by using dashed lines. For example, the instance "Pleurisy" (ذات الجنب) is an instance of the class "Lung_Disease" (مرض رئوي) while the instance "Cupping" (الحجامة) is an instance of the class "Preventive_Medicine" (الطب الوقائي).

Note that some of the instances shown in Figure 5 were extracted from another saying and are not included in the given saying. For example, the instance "سبع أشفية" (Seven_Healings) was extracted from the saying " عليكم بهذا العود الهندي فإنّ فيه سبعة أشفية". However, it was included in Figure 5 because it is related to the instance "Costus" (القسط البحري). This example shows that the process of creating and interlinking instances demands a deep understanding of the conceptual relationships between the different Prophetic sayings.

Table 2 *The object properties of the ontology*

Object Property	Domain	Range	Description
Cures	Cure	Disease	يعالج
Cured By	Disease	Cure	يعالج_بـ
Causes	Disease	Disease	يسبب
Caused By	Disease	Disease	يسببه
Has Cure Method	Cures	Treatment_Method	طريقة العلاج
Hastime	Cures	Treatment_Time	وقت العلاج
Has Place	Cure	Treatment_Place	لديه مكان
Has Fees	Cure	Treatment_Fees	لديه أجره
Has Benefit	Cure	Benefit	لديه فائدة
Has Synonym	Cure Disease	Cure Disease	له مرادف

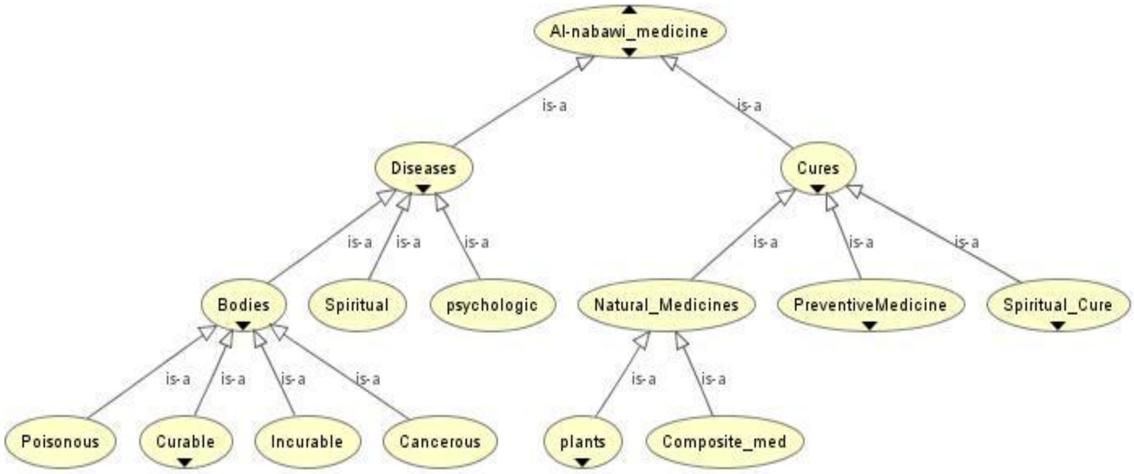

Figure 4 *Top level classes of the ontology of the Prophetic Medicine*

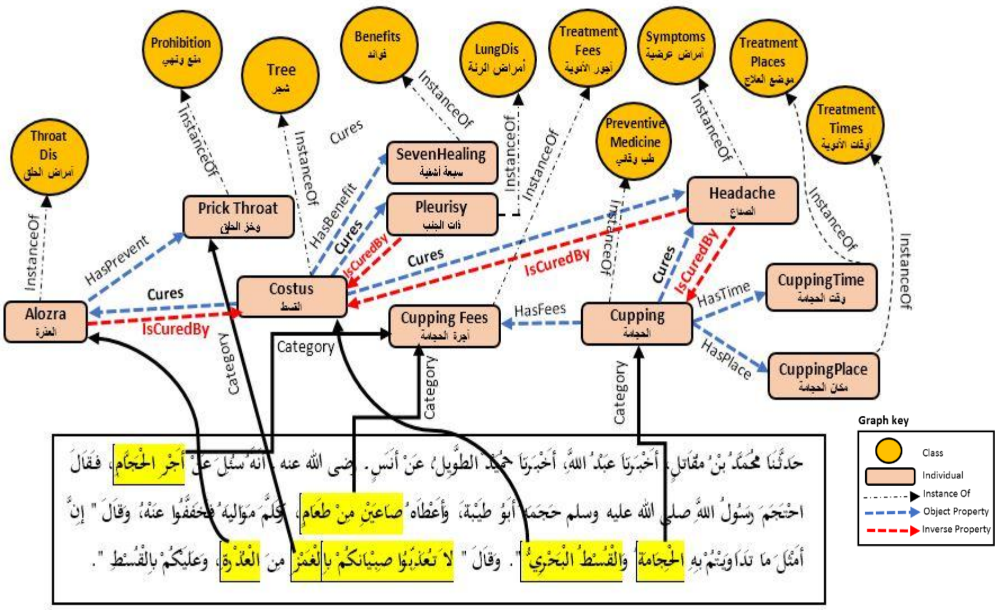

Figure 5 *An example showing the relationships between ontology classes, instances*

6. Annotation Module:

After constructing the ontology of Prophetic medicine and populating it with instances, the following step is to perform semantic annotation by linking the ADL content with the ontology concepts. The annotation process was performed on the sixty sayings from which we extracted the ontology instances. The aim of the annotation process is to enrich words and phrases in text with semantic information from the ontology so that search agents can infer new results.

The annotation process was performed manually. The reason for using manual rather than automatic annotation is due to the special capability required to identify the implicit relations and dependencies between different Islamic texts. For example, it may be necessary to link a disease mentioned in one saying with a remedy mentioned in another saying. This special capability, which goes beyond the capabilities of automatic annotation techniques, can be only offered by experts who can reliably devise the semantic relations between terms by virtue of profession and experience.

To facilitate the manual annotation, we built a tool that enables the user, i.e. domain expert, to easily link words, or sentences, of the text to the ontology terms. The main window of the annotation tool is shown in Figure 6: The whole ontology hierarchy is visually represented as a tree (see Figure 6.A) where inner nodes denote ontology classes while leaf nodes denotes instances/child classes. The annotation process is done as the following: 1) the user highlights the text segment to be annotated. 2) The user selects relevant ontology terms, i.e. classes and/or instances, from the ontology tree. 3) On clicking "Add annotation" button, the appropriate annotation is created and stored in the RDF store. The annotation tool also enables the user to view all stored annotations and delete them.

To illustrate how annotations are created and stored, consider the saying of Prophet Mohamed shown in Figure 5:

The highlighted phrases of this saying can be linked to concepts of the ontology as the following (see Figure 6):

- Annotation 1: the word "أجر الحجام" is linked to the ontology instance "Cupping_Fees" which is of the type "Treatment_Fees".
- Annotation 2: The word "صَاعَيْن مِن طَعَام" is linked to the ontology instance "Cupping_Fees" " أجر الحجامة" which is an instance of the class (concept) "Treatment_Fees""أجر العلاج" .
- Annotation 3: The word "الحِجَامَةُ"is linked to the instance "Cupping" "الحجامة" which is an instance of the class "Preventive_Medicine""الطب الوقائي" .
- Annotation 4: The word "القُسْطُ البَحْرِيُّ"is linked to the concept "Costus" "القسط" which is an instance of the class "Tree""أشجار" .
- Annotation 5: The word "العُذْرَةِ"is linked to the concept "Alozra" "العذرة" which is an instance of the class "Throat_Diseases "(أمراض الحلق).

Each annotation is represented in RDF and stored in the RDF store of the OntoADL. The ontology has a dedicated class named "Annotation" for representing annotations. For each annotation, an instance of type "Annotation" is created, and the annotation details are stored by using the data-type properties illustrated in Table 3. Values of these properties are set by the annotator through the annotation tool. Of these properties, the "hasCategory" property is the most important as it links the annotation to the URI of the ontology instance or class. Other data properties, i.e. hasSource, hasStartIndex and hasEndIndex, aim to maintain the source saying to which the annotated text belongs as well as the location of the text within the saying. Overall, the annotation of the sixty sayings resulted in 165 annotations at a rate of 2.75 annotations per each Prophetic saying.

7. Reasoning Module:

After the annotation process completes, we run the inference engine over the annotations in the RDF store in order to infer new facts from existing ones. This enables the declaration of derived classes or the additional statements by leveraging the property characteristics, e.g. transitivity, symmetry and inversion).

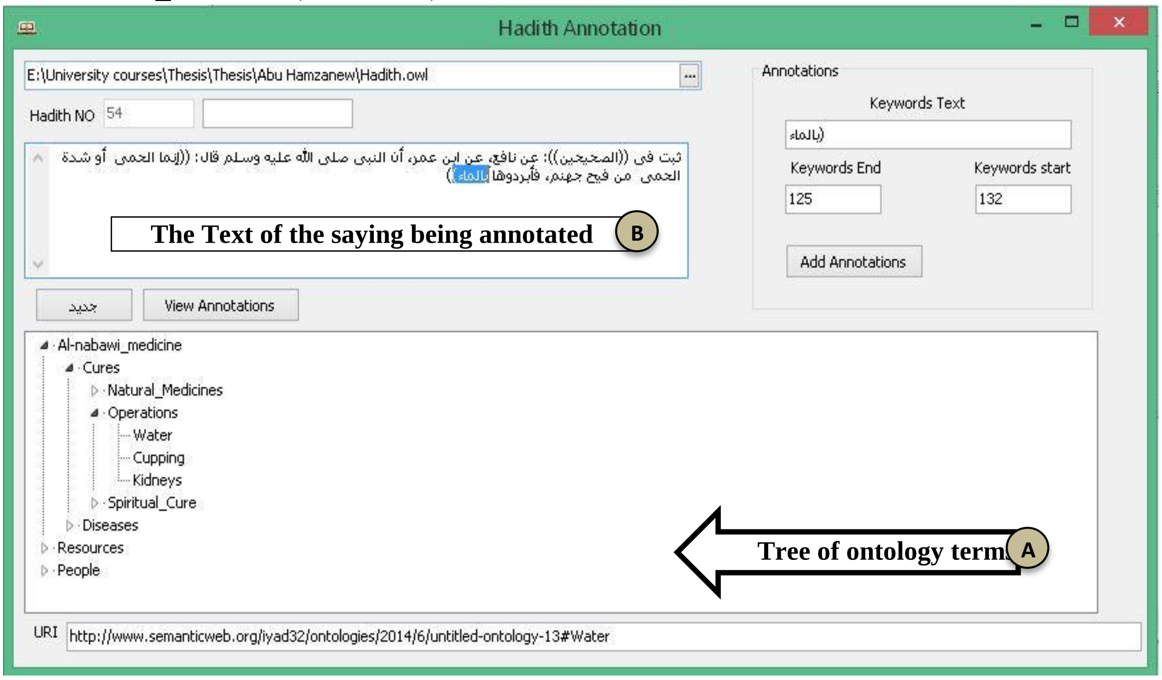

Figure 6 *OntoADL Annotation Tool*

Table 3 *Data properties associated with an annotation instance*

Data-type Properties	Domain	Range	Description
Has Category	Annotation	String	URI for any category whether classes or individuals
Has Source	Annotation	String	URI address for the hadith that contained specific text
Has Start Index	Annotation	Integer	Reflects the beginning of text in hadith
Has End Index	Annotation	Integer	Reflects the end of text in hadith

To illustrate how the inference engine can extend the system's search capabilities, consider the following fact that is defined in the ontology: <:Honey :cures :Diarrhea> (العسل يعالج الاسهال). The ontology also defines that :curedBy is the inverse property of :cures. Therefore, the inference engine will add the following statement without us doing anything: <:Diarrhea :curedBy :Honey> (الاسهال يعالج بالعسل). Given the predefined and inferred statements, the search agent will be able to answer any of the following queries "ماذا يعالج العسل؟" and "كيف يُعالج الاسهال ؟".

In another example, the ontology defines the following two separate facts: <:Tumor :causes :Fever> (الورم يسبب الحمى) and <:Fever :causes :High_Temperature> (الحمى تسبب إرتفاع درجة الحرارة). Since the property :causes is defined as a transitive property, the inference engine will add the following statement: <:Tumor :causes :High_Temperature> (الورم يسبب ارتفاع درجة الحرارة). With the new statement, the search agent can expose the high temperature as a consequence of tumors although this fact is not explicitly included in the ontology. After the activating the inference engine, newly inferred statements are automatically added to the RDF store so that the search agent can benefit from the new knowledge.

8. Processing of User Queries:

The focus in the previous sections was on issues related to the knowledge representation in OntoADL including the ontology construction, the semantic annotation and the semantic inference. The following sections explain how the user query is handled from the point the user inputs the keywords until the search results are retrieved and presented to the user.

The keywords entered by the user are preprocessed by the Linguistic Module (see Figure 3). The aim of text preprocessing is to prepare the keywords so that they can be properly matched with the ontology content. The following preprocessing steps are applied on the keywords:

- Orthographic normalization (e.g. replacing "أ"with "ا"and "ه"with "ة"). Stanford Arabic Word Segmenter [26] is used to apply normalization.
- Removal of stopwords and special characters such as "_" which often occurs in ontology text.
- Part of speech tagging: Stanford Arabic POS [27] is used for this purpose. Part of Speech tagging is necessary to identify verbs, which often correspond to ontology properties, and nouns, which often refer to ontology classes and instances.
- Word Stemming by using the Arabic stemmer proposed by Khoja [28]. Stemming aims to make the Arabic words comparable with the ontology terms regardless of the different formats.

User keywords are then matched with the ontology terms. The aim of ontology matching is to identify ontology terms that best describe the user keywords. These ontology terms will be used later to build SPARQL queries that can retrieve relevant results and recommendations.

The ontology matching process is handled by the Ontology Matcher module as the following: First, all possible N-grams are generated from the user keywords (n is limited to 3 in our prototype). Generating N-grams is essential to consider different combinations of the keywords when matching with the ontology. Second, the stem of each N-gram is matched with the stems of all ontology terms (classes, instances and properties). We used Jaccard metric to compute the similarity between the N-gram and the ontology term. The Jaccard similarity between the word sets S and T is simply $\frac{|S \cap T|}{|S \cup T|}$. Only ontology entities that achieve highest similarity results are considered. It should be noticed that the user keywords are matched with the names of ontology entities as well as all their synonyms to ensure the best matching results.

9. Translating Ontology Terms to SPARQL:

The aim of the SPARQL Query Generator (see Figure 3) is to generate SPARQL queries. The input to this component is the ontology terms that correspond to the user keywords. These terms should have been identified from the matching process described in the previous section. The generation of SPARQL queries is done as the following:

The first step is to find the keywords that refer to the user's interest. Target keywords will be interpreted as the variables following the "SELECT" in the resultant SPARQL query. The first noun in the list of user keywords often refers to the user's main interest. For example, in the queries: "فوائد السواك (The benefits of Sewak)" and " أعراض الحمى (The symptoms of fever)", the words "فوائد (benefits)" and "أعراض (Symptoms)" are the target keywords while the following words often modify the meaning. If the user query starts with a question word (e.g. “ما، كيف،..”) or a command word (e.g. "أذكر، عدد"), we take the following noun as a target.

Target words are interpreted as variables after “SELECT” based on the following rules: (1) if the target word refers to an ontology class, e.g. “:SomeClass”, interpret it as the variable “?someVariable”, and add the following triple <?someVariable rdf:type :SomeClass> to the WHERE clause. (2) if the target word refers to an ontology instance or property, interpret them as their URIs in the ontology.

After interpreting the target words, the next step is to generate the WHERE clause by linking the query words. For this purpose, we used the rules defined by AlAgha and Abu-Taha [29] who proposed a set of handcrafted rules to translate Arabic natural language queries to SPARQL. Due to the limited space, we only present two sample rules and show how they are applied. For more details about the translation rules please refer to AlAgha and Abu-Taha [29]:

Rule 1: If the user keywords refer to an ontology class C followed by an ontology instance I, we find the property P that links the class C with the instance I so that we get a complete RDF triple. This can be done by querying the ontology to find the property P that has the class C in its domain and the class of instance I in its range.

To illustrate how the above rule is applied, take the query “علاج الحمى” (Cure of fever). This query has the word “علاج”(Cure) as a target which also refers to the ontology class “:Cure”. The word "الحمى" (Fever) matches with the ontology instance :Fever. Based on the above rule, we refer to the ontology to find the property that links the class :Cure with the class of the instance :Fever. The property :cures is captured from the ontology statement: <:Cure :cures :Disease>. This generates the following SPARQL query: “SELECT ?someVariable WHERE {?someVariable rdf:type :Cure . ?someVariable :cures :Fever}”.

Rule 2: If the user keywords refer to an object property P followed by an ontology instance I, we find the ontology class C that is linked with the class of instance I through the property P. This can be done by querying the ontology for statements that have P as a predicate and the class of I as an object. The identified class C is interpreted as a variable, e.g. “?someVariable”, and add the following triple <?someVariable rdf:type C> to the WHERE clause.

The following example illustrates how the above rule is applied: Given the user input: “أسباب الصداع” (causes of headache). This query has the word “أسباب”(causes) as a target which also matches with ontology property “:causes”. The word "الصداع" (Headache) matches with the ontology class :Headache. On applying Rule 2 we have the property :causes followed by the instance :Headache. A subject is required to generate a complete triple in the resultant SPARQL. The missing subject can be determined by referring to the ontology statements that have the pattern <?, :causes, :Disease> where the object :Disease is the class of the instance :Headache. By referring to Table 2, the class :Disease belongs to the domain of the property :causes, and thus is used to as a subject of RDF triple. Afterwards, the class :Disease is replaced with a variable, i.e. ?someDisease, and the triple <?someDisease rdf:type :Disease> is added to the WHERE clause. The following SPARQL is finally generated:”SELECT ?Disease WHERE { ?Disease rdf:type :Disease . ?Disease :causes :Headache}”.

Notice that OntoADL is not a question answering system. Therefore, it is not currently able to correctly handle complex user queries that require deep processing and linguistic analysis. For example, queries consisting of multiple clauses, e.g. “ما أعراض الأمراض التي تعالج بالحجامة” or queries that contain superlative/comparative words, e.g. "ما أكثر/أهم ..." are not supported.

10. Presentation of Results:

The execution of the generated SPARQL will retrieve one or more ontology terms, i.e. classes, instances, literal values. The ontology terms cannot be directly presented to end users as they may not be meaningful. Results to be presented to the end user are constructed from the retrieved ontology terms as the following:

The Prophet sayings that are annotated with the ontology terms are retrieved and displayed in the results area. The annotated part is highlighted in different color since it should refer to the desired result. In the example shown in Figure 2, the execution of the SPARQL query generated from the keywords “علاج الالتهاب الرئوي” retrieves the ontology instance :Indian_Lute (العود الهندي). All texts that are annotated with this instance are displayed in the results area.

The list of recommendations displayed on the left side of the search window is constructed by retrieving ontology statements that contain the resultant ontology term. These statements include all other ontology terms that are related to the term of interest. These statements are then mapped to Arabic, grouped and structured in tree format. Given the example in Figure 6, all ontology instances or classes that are related with the instance :Indian_Lute are retrieved. Among many things, this include all diseases that are cured by the indian lute. These diseases are retrieved from the statements that match the pattern: <:Indian_lute :cures ?>.

11. Evaluation:

The aim of this part was to explore the potential of our ontology-based approach to improve information search as compared to the conventional approach. OntoADL system was tested using a set of search queries formulated by a domain expert. Search results were assessed in terms of precision and recall, and compared with the results obtained from the traditional search in ADL.

11.1. Experimental Settings:

The dataset we used for evaluating the search performance was restricted to sixty sayings of Prophet Mohammed from the domain of Prophetic Medicine. These are the sayings that were semantically annotated using the ontology we constructed. Since the same sayings are also included in the traditional ADL, our intention was to execute the search queries over the same dataset and compare results across the two systems. To make OntoADL comparable with ADL, we configured the ADL to search only in the same reference book, i.e. “Atteb Al-Nabawi”. We also excluded all results that were not contained in the sixty sayings under consideration.

The Prophetic sayings and the ontology content were introduced to two human experts who were specialized in the Science of Hadith and who have used the ADL software for several years. Experts were then asked to formulate forty queries that can be answered using the knowledge contained in the OntoADL system. These forty queries were then reduced to twenty queries after excluding redundant, similar and unsupported ones. These queries are shown in Table 4. Queries were then inputted to both ADL and OntoADL systems, and search results were presented to the expert to judge their correctness and completeness.

Although the query set was limited to twenty queries only, it took a significant effort to validate the search results. This is because there was no gold reference to validate the results. For each query, we had to inspect all information items, both retrieved and non-retrieved, and assess their relevance in order to calculate the precision and recall.

11.2. Evaluation Metrics

Precision measures the number of correctly identified items as a percentage of the number of items identified. Recall measures the number of correctly identified items as a percentage of the total number of correct items[30]. They are calculated using the following equations:

$$\text{Precision} = \frac{Number\ of\ relevant\ items\ retieved}{Number\ of\ retrieved\ items}$$

$$\text{Recall} = \frac{Number\ of\ relevant\ \ items\ retrieved}{Number\ of\ relevant\ items}$$

11.3. Results and Discussion:

Table 4 shows the evaluation results: For each query, the precision and recall values are calculated for the keyword-based search in ADL as well as the ontology-based search in OntoADL. On average, the OntoADL semantic search achieved higher precision and recall values (Avg. Precision = 0.83, avg. recall = 0.9) when compared to the ADL keyword based search (Avg. precision =0.3 and avg. recall = 0.69).

By looking closely at individual results, the semantic search also outperformed the conventional search for seventeen of the twenty queries in hand. Three queries numbered as 9, 17 and 20 gave comparable results across the two search methods, indicating that the semantics did not present any improvement. This can be attributed to the lack of ontology statements or annotations related to topics such as ”سب الحمى” or “تقرحات الجلد”. When there are no semantics associated with searched term, search results will be limited to keyword-based search.

Table 4 *Recall and precision values for twenty queries executed over ADL and OntoADL*

#	Query	ADL keyword-based search		OntoADL (Ontology-based Search)	
		Recall	**Pres.**	**Recall**	**Pres.**
1.	الحجامة	0.92	0.79	0.92	0.82
2.	الحث على المداوة	0.43	0.50	0.67	0.40
3.	علاج الحمى	0.14	0.20	0.70	0.60
4.	أمراض تعالج بالحناء	1.00	0.33	1.00	0.75
5.	التداوي بأمر الله	1.00	0.10	1.00	0.55
6.	فوائد العسل	0.67	0.40	0.80	0.60
7.	الأجرة على الحجامة	1.00	0.25	1.00	0.65
8.	علاج الصداع	0.50	0.25	0.75	0.75
9.	نهي سب الحمى	1.00	0.14	1.00	0.25
10.	علاج الالتهاب الرئوي	0.00	0.00	1.00	0.67
11.	أهمية السواك	1.00	0.50	1.00	1.00
12.	نهي تعذيب الأولاد	0.00	0.45	0.044	0.75
13.	الرقية من الوجع	0.80	0.22	0.50	0.50
14.	علاج الإسهال	0.29	0.29	0.89	0.80
15.	الوقاية من السحر	0.77	0.14	1.00	0.46
16.	الأدوية الوقائية	0.50	0.45	0.88	0.78
17.	أمراض الجسم	0.85	0.68	0.85	0.69
18.	العلاج بالماء	1.00	0.50	0.83	0.92
19.	أنواع الأدوية الطبيعية	0.83	0.56	0.83	0.83
20.	تقرحات الجلد	1.00	0.50	1.00	0.50
Average precision and recall		**0.70**	**0.36**	**0.83**	**0.66**

The difference in precision between the two systems was remarkably large. This is because the search with no semantics often gave a large number of irrelevant results that did not match with the query needs. For example, searching for "علاج الحمى" (Query No. 3) in ADL resulted in 97 result, most of which revolved around the topic of fever (الحمى). However, only few results fulfilled the intent of the query which is the cure of fever (علاج الحمى). In fact, the keyword-based search ignored the conceptual relations between terms in the query phrase, and retrieve results based on the word co-occurrence in text. For example, any chapter or section containing the word "علاج" or the word "الحمى" was retrieved.

In contrast, the semantic search could often handle the conceptual relations between words in search phrases and interpret these relations in RDF. For example, searching for " أمراض تعالج بالحناء " (Query No. 4), the OntoADL only retrieved diseases that are linked with the ontology instance :Henna (الحناء) via the property :cures (or its inverse property :curedBy).

Another limitation we revealed in ADL is the inability to recognize synonyms of keywords, a thing that degraded the search performance. Terms such as "الالتهاب الرئوي" (Query No. 10) or "الإسهال" (Query No. 14) exist in the contemporary medical terminology but do not literally exist in the Prophetic terminology. In fact, the Prophetic medicine refers to these terms by using alternative terms such as "ذَاتُ الجَنْبِ" or "الاستطلاق". In contrast, the OntoADL search overcomes this limitation by matching with the synonyms stored in the ontology.

Beside the above results, experts provided positive feedback on the system performance. For example, one expert indicated that the list of recommendations offered by OntoADL helped him to narrow the scope of search results and explore salient topics related to the topic of interest. The other expert said that OntoADL offers solutions to common problems of ADL by expanding the search terminology to support synonyms and derivatives of supplied words.

Experts also gave suggestions to improve the OntoAD. For example, they indicated that the search interface is simple and should be improved to support advanced search tools such as searching in specific books, chapters or searching in particular category. In fact, these suggestions have not been considered in the current prototype version of OntoADL since our priority at this stage is to demonstrate the potential of semantic search in Islamic resources.

12. Conclusion and Future Work:

This work presents an integrated solution to enhance searching in Islamic resources in ADL by using semantic technologies. Our solution provides insight into the different stages of developing a semantic search service including the ontology construction, the semantic annotation of text, the handling of search keywords and translating them to SPARQL, and the presentation of results to end users. The user scenarios and evaluation results have showed the potential of semantic search to retrieve more accurate, concise and informative results as compared to results retrieved by traditional ADL. Due to the massive size of ADL and the variety of disciplines it covers, it was difficult to support semantic search in the entire content of ADL. As a proof of concept, we only selected a section of ADL and augmented it with ontology-based annotations.

We think that the research in this field can be extended in many directions: First, we will try to expand the knowledge base to cover additional domains of ADL. This will allow us to evaluate the system in a wider context and to examine how it generalizes to new domains. Second, the process of matching keywords to the ontology can be improved by integrating Arabic dictionaries and semantic similarity measures. This can reduce the gap between the ontology and the user's terminology, and eliminate the burden of defining all synonyms of concepts inside the ontology. Third, the user interaction can be د. بكر الريس4 improved by providing a natural language interface to the system. The intention is to allow the users to query the Islamic knowledge by using natural language queries expressed in Arabic. For this purpose, approach as such [31] and [29] can be integrated to support the interpretation of natural language queries into SPARQL. Fourth, we will explore approaches to handle the annotation process automatically or semi-automatically by exploiting Arabic named entity recognition.

References:

[1] Official Website of Ak-Maktaba Al-Shamela. Available from: http://shamela.ws/, [Accessed on 20-8-2015]

[2] Iqbal, R., Mustapha, A., and Yusoff, Z. M. An Experience of Developing Quran Ontology with Contextual Information Support. *Multicultural Education & Technology Journal*, ***7(4)***, (2013) 333-343.

[3] Zhao, Y. and Hu, C. An Ontology Based Framework for Knowledge Service in Digital Library. *Wireless Communications, Networking and Mobile Computing (WiCom), International Conference on*, Shanghai, China, IEEE, (2007) 5345-5348.

[4] Dukes, K. and Buckwalter, T. A Dependency Treebank of the Quran Using Traditional Arabic Grammar. *Informatics and Systems (INFOS), The 7th International Conference on*, Cairo, Egypt, IEEE, (2010) 1-7.
[5] Dukes, K., Atwell, E., and Habash, N. Supervised Collaboration for Syntactic Annotation of Quranic Arabic. *Language resources and evaluation*, ***47(1)***, (2013) 33-62.
[6] Saad, S., Salim, N., and Omar, N. Keyphrase Extraction for Islamic Knowledge Ontology. *Information Technology (ITSim), International Symposium on*, IEEE, (2008) 1-6.
[7] Fatwa-Online. Fatwa-Online. 2015; Available from: http://www.fatwa-online.com/, [Accessed on 30/8/2015]
[8] Books, I. Islamic Books. 2015; Available from: http://www.download-islamic-ebooks.com/, [Accessed on 1/9/2015]
[9] Interpretations, Q. Quran Interpretations. 2015; Available from: http://holyquran.net/tafseer/index.html, [Accessed on 30/8/2015]
[10] Site, I. Islam site. 2015; Available from: http://hadith.al-islam.com/, [Accessed on 30/8/2015]
[11] Khan, H.U., Saqlain, S.M., Shoaib, M., and Sher, M. Ontology Based Semantic Search in Holy Quran. *International Journal of Future Computer and Communication*, ***2(6)***, (2013) 570-575.
[12] Atwell, E. A Corpus-Based Computational Model for Knowledge Representation of the Quran. *Proceedings of the Corpus Linguistics Conference (CL)*, Liverpool, UK, (2009) 169.
[13] Al-Yahya, M., Al-Khalifa, H., Bahanshal, A., Al-Odah, I., and Al-Helwah, N. An Ontological Model for Representing Semantic Lexicons: An Application on Time Nouns in The Holy Quran. *Arabian Journal for Science and Engineering*, ***35(2)***, (2010) 21-35.
[14] Iqbal, R., Mustapha, A., and Mohd. Yusoff, Z. An Experience of Developing Quran Ontology with Contextual Information Support. *Multicultural Education & Technology Journal*, ***7(4)***, (2013) 333-343.
[15] Saad, S., Salim, N., and Zainal, H. Islamic Knowledge Ontology Creation. *Internet Technology and Secured Transactions (ICITST), International Conference for*, London, UK, IEEE, (2009) 1-6.
[17] Al-Khalifa, H.S., Al-Yahya, M., Bahanshal, A., and Al-Odah, I. SemQ: A Proposed Framework for Representing Semantic Opposition in the Holy Quran Using Semantic Web Technologies. *Current Trends in Information Technology (CTIT), International Conference on the*, Dubai, IEEE, (2009) 1-4.
[18] Shoaib, M., Yasin, M.N., Hikmat, U., Saeed, M.I., and Khiyal, M.S.H. Relational WordNet Model for Semantic Search in Holy Quran. *Emerging Technologies (ICET), International Conference on*, Mallorca, Spain, IEEE, (2009) 29-34.
[19] Salim, J., Hashim, S.F.M., and Noah, S.A.M. A Framework to the Development of Islamic Ontology: Symbiosis of Thesaurus and Domain Expert in Ontology Development. *Electrical Engineering and Informatics (ICEEI), International Conference on*, Bandung, Indonesia, IEEE, (2011) 1-6.
[20] Yauri, A.R., Kadir, R.A., Azman, A., and Murad, M.A.A. Quranic-Based Concepts: Verse Relations Extraction Using Manchester OWL Syntax. *Information Retrieval & Knowledge Management (CAMP), International Conference on*, Kuala Lumpur, Malaysia, IEEE, (2012) 317-321.
[21] Azmi, A. and bin Badia, N. e-Narrator: An Application for Creating an Ontology of Hadiths Narration Tree Semantically and Graphically.*Arabian Journal for Science and Engineering*, ***35(2c)***, (2010) 51-68.
[22] Baraka, R.S. and Dalloul, Y.M. Building Hadith Ontology to Support the Authenticity of Isnad. *International Journal on Islamic Applications in Computer Science and Technology*, ***2(1)***, (2014) 25-39.
[23] AlAgha, I. and AlMasri, M. Ontology of Prophetic Medicine. 2015; Available from: https://goo.gl/wKp2LO, [Accessed on 1/9/2015]
[24] Noy, N.F., Sintek, M., Decker, S., Crubézy, M., Fergerson, R.W., and Musen, M. A. Creating Semantic Web Contents with Protege-2000. *IEEE Intelligent Systems*, ***(2)***, (2001) 60-71.
[25] Motik, B., Grau, B.C., Horrocks, I., Wu, Z., Fokoue, A., and Lutz, C. Owl 2 Web Ontology Language: Profiles. 2009; Available from: http://www.w3.org/2007/OWL/draft/ED-owl2-profiles-20090420/all.pdf, [Accessed on 1/9/2015].
[26] NLP, S. Stanford Arabic Segmenter. 2015; Available from: http://nlp.stanford.edu/software/segmenter.shtml, [Accessed on 1/9/2015].
[28] Khoja, S. APT: Arabic Part-of-Speech Tagger. 2001; Available from: http://zeus.cs.pacificu.edu/shereen/NAACL.pdf, [Accessed on 1/9/2015].
[29] AlAgha, I. and Abu-Taha, A. AR2SPARQL: An Arabic Natural Language Interface for the Semantic Web. *International Journal of Computer Applications*, ***124(18)***, (2015) 1-7.

[30] Maynard, D., Peters, W., and Li, Y. Metrics for evaluation of ontology-based information extraction. *International world wide web conference*, Edinburgh, UK, (2006).

[31] AlAgha, I. Using Linguistic Analysis to Translate Arabic Natural Language Queries to SPARQL. *International Journal of Web & Semantic Technology*, ***6(3)***, (2015) 25-39.