



## Full length article

## IBRI-CASANTO: Ontology-based semantic search engine

Awny Sayed<sup>a,\*</sup>, Amal Al Muqrishi<sup>b</sup><sup>a</sup> Faculty of Science, Minia University, Egypt<sup>b</sup> Nizwa University, Oman

## ARTICLE INFO

## Article history:

Received 20 June 2016

Revised 24 September 2016

Accepted 2 January 2017

Available online 17 January 2017

## Keywords:

Ontological search engine

Keyword-based search

Semantics-based search

Resource Description Framework (RDF)

Ontological graph

## ABSTRACT

The vast availability of information, that added in a very fast pace, in the data repositories creates a challenge in extracting correct and accurate information. Which has increased the competition among developers in order to gain access to technology that seeks to understand the intent researcher and contextual meaning of terms. While the competition for developing an Arabic Semantic Search systems are still in their infancy, and the reason could be traced back to the complexity of Arabic Language. It has a complex morphological, grammatical and semantic aspects, as it is a highly inflectional and derivational language. In this paper, we try to highlight and present an *Ontological Search Engine* called IBRI-CASANTO for Colleges of Applied Sciences, Oman. Our proposed engine supports both Arabic and English language. It is also employed two types of search which are a *keyword-based search* and a *semantics-based search*. IBRI-CASANTO is based on different technologies such as *Resource Description Framework (RDF)* data and *Ontological graph*. The experiments represent in two sections, first it shows a comparison among Entity-Search and the Classical-Search inside the IBRI-CASANTO itself, second it compares the Entity-Search of IBRI-CASANTO with currently used search engines, such as Kngine, Wolfram Alpha and the most popular engine nowadays Google, in order to measure their performance and efficiency.

© 2017 Production and hosting by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

The World Wide Web data is growing rapidly in the data repositories because of various factors such as users, systems, sensors and applications. For example, millions of transactions that occur daily, and the social media tools such as Facebook, Twitter, LinkedIn, Google+, and Tumblr, add vast of information. These large data create several challenges that called V attributes : *Velocity, Volume and Variety*. Clearly, the velocity means the data comes at high speed, while volume focus on large and growing files and the variety means the files come in various formats (e.g. text, sound and video). These issues enable a competition among the developers to search about a technique that help to extract the accurate data

and overcome the current problems in order to reach a semantic search.

In a semantic, the data is stored in different levels as it is illustrated in Fig. 1, the hierarchy of layers to reach a proposed semantic search. It start from XML (Extensible Markup Language), RDF, RDFs (RDF Schema) and OWL (Ontology). Each concept is a complementary for the next and the last two concepts are the crucial to get semantic search. While the RDFS [3–5] suffers from many weaknesses, that leads to create a movement and extend it to the Ontology upper layer. For instance, RDFS has a weakness to describe resources in sufficient details because there is no localized range and domain constraints. In addition, it is difficult to provide reasoning support and has no existence/cardinality constraints and no transitive, inverse or symmetrical properties.

Ontology gets over from the issues of RDFs that makes this concept the nearest one to the semantic search. Actually, the term Ontology has been used for several years ago by the *artificial intelligence* and *knowledge representation* community. However, nowadays it is becoming a part of the standard terminology of a much wider community including information systems modeling [1]. The concept of Ontology is borrowed from philosophy, where it means a systematic account of existence [2], for instance ontological question like what are the fundamental parts of world and how

\* Corresponding author.

E-mail addresses: [awny.sayed@mu.edu.eg](mailto:awny.sayed@mu.edu.eg) (A. Sayed), [amalsyedsultan@gmail.com](mailto:amalsyedsultan@gmail.com) (A. Al Muqrishi).

Peer review under responsibility of Faculty of Computers and Information, Cairo University.



Production and hosting by Elsevier

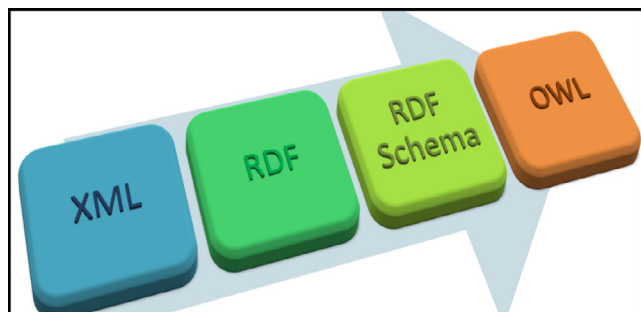


Figure 1. Hierarchy of Semantic Search.

they related to each other. Therefore, ontology helps philosopher to discuss challenging questions to build theories and modules. Our purpose in this research is to focus on non-philosophical ontology, which means the description of what exist within determined field.

Currently, Ontology is becoming very important because we have a lack of standards (shared knowledge) which are rich in semantics that represented in machine understandable form. Moreover, it has been proposed as a solution for the problems that arise from using different terminology to refer to the same concept or using the same term to refer to different concepts [6]. Ontology is built to develop the required *conceptualizations* and *knowledge representation* in order to meet various challenges. Actually, the Web has tremendous collection of useful information; however, extracting the accurate information from the web is extremely difficult, because the current search engines are restricted to the keyword-based search techniques. Thus, the interpretation of information contained in web documents is left to the human user to done manually. Therefore, all these obstacles lead to the first challenge which is the inability to use the abundant information resources on the web correctly. The second challenge is the difficulty of Information Integration from various sources because of the factor of synonyms and homonyms. Finally, the issues of Knowledge Management. Multi-actor scenario involved in distributed information production and management, for example, people and machines cannot share knowledge if they do not speak a common language.

There are three types of ontology based on the degree of conceptualization into three types, which are *Top-level Ontology*, *Domain Ontology* and *Application Ontology* [7]. Each type have its range and capacity of information. For instance, the Top-level Ontology depicts very general notions, in which they are independent of a particular problem or domain. In addition, they are applicable across domains and includes vocabulary related to things, events, time, space, etc.

*Domain Ontology* represented data in a particular domain, and provides vocabularies about concepts and their relationships or about the theories governing the domain. Moreover, it is rich of axiomatic theories whose focus is to clarify the intended meanings of terms used in specific domains. It is designed to not only fit the needs of specific community but also provides terminological structure that can share between different communities. Therefore, the reference ontology sometimes called the foundation ontology. It helps developer to avoid build ontology from the scratch by using other references of ontology that built before in order to implement minimal modifications on it. However, the application ontologies could be generating from the reference ontologies.

*Application Ontology* refers to knowledge pieces depending on both a particular domain and task. Therefore, it is related to problem solving methods and provide a minimal terminological structure to fit the needs of a specific domain and community, which

make it too specific. Therefore, it cannot be share or used by another community.

There are different types of search nowadays, the classical search and semantic search. Each type has its own view or technique of searching. The classical search is focused on popularized keywords, where it means the users can submit a set of keywords to the search engine and a ranked list of information is returned back to the user [8]. There are different sites and applications that support the keyword-based search engines such as *Google*, *Gmail* and *Yahoo*. The second type of search is a semantic search that clarifies the lack in the concept of keyword semantics in the previous examples and the classical search; because they give many irrelevant and inaccurate outcomes to the users [9]. It is so far from the concept of understanding searcher intent and the contextual meaning of the user query. Thus, it is a challenge that has been addressed and solved by many semantic search engines.

Since there is a few of the *Ontological Search Engines* that supports Arabic language. It could be traced back to the Natural Language Processing [10] and gaps/challenges to solve syntactic search and produce synonym meaning of words. Thus, this paper is focused on implementing the *Ontological Search Engine* based on the ontological graph that is called IBRI-CASONTO. Although, IBRI-CASONTO supports both Arabic and English languages, we shall put our attention to discuss the Arabic search in this paper. It uses both the keyword-based search as well as the semantics-based search which also known as the Ontological Search.

The rest of this paper is structured as follows. The second section introduces the researcher efforts in order to build the ontological search engines, their techniques, domain, languages support, for instance, Wolfram-Alpha, Kngine and Google. The third section discuss the Arabic Language and its related to the Ontology concept. Whereas the fourth section highlights the ontology components. Section five and six present our proposed engine, IBRI-CASONTO, in detail and the experimental evaluations that test the engine with simple and complex queries and compare the proposed engine with other common and popular semantic engines. At the end, section seven concludes the paper and gives some suggestions in order to improve the IBRI-CASONTO in the future.

## 2. Related works

Ontology is considered as a portal to make the engines more intelligent and powerful. It is a respectful mission for the current generation of the web which known as Web 3.0 and the future mission for Web 4.0. Ontology is powerful and has a correct and reliable data that stores in its repositories that called the ontological graphs. It enables user to get and retrieve a direct answer without any complexities.

There are several ontological graphs developed according to the developers' interest some of them serves one domain while others develop to involve multiple domains such as the electronic government. Our purpose focus on developing Arabic and English IBRI-CASOnto, which stands for Ibri College of Applied Sciences Engine. It is a domain specific that called a reference ontology. It is focused on the college information such as academic departments, academic staffs, students, where they live and so on. Developers already had been created some reference ontologies that focus on academic community for instance, *HERO ontology* [11], *Univ-Bench ontology* [11], *university ontology* [11,12] and *AIISO ontology* [12,13]. Currently, there are some engines that based on the concept of semantic such as *Kngine* [14], *Wolfram Alpha* [15] and the most popular engine nowadays *Google*.

*Kngine* [14] is the first multi-language question answering engine which supports around four languages and English, Arabic with them. Kngine stands for Knowledge Engine that is Web 3.0

Knowledge Engine. It is designed to provide customized and exact meaningful search results. For instance, semantic information about the keywords, user's queries, list things, find out the relations between the keywords. The exciting characteristics of this search engine, it gives precise results which links different kinds of related information together to present them to the user such as: *movies, photos, and prices and the users reviews*.

*Wolfram Alpha* [15] is a computational knowledge engine or -answer engine which developed by Wolfram Research. It is an online website that answers factual queries directly by computing the answer from externally sourced "curated data" or structured data, rather than providing a list of documents or web pages.

There are several techniques that used in semantic engines such as artificial intelligence, natural language processing [16] and machine learning. As shown in the Table 1, Kngine utilizes the efficiency of *Knowledge-Based approach* and the power of the *statistical approach* [17], whilst Google used its own search technology which called *Hummingbird algorithm* [18]. That means "precise and fast" of data or query's answer which are the powerful features for any search engine. On the other hand, all these engine have their own mobile application that facilitates them to be more popular and portable for the customers throughout the world. Furthermore, they have an advanced feature that called "voice recognition" which enables the operating system to convert spoken words into written text. Moreover, the Table 1 indicates that most of the search engines support English language, while there is few engines that support Arabic language such as Google and Kngine; however, these engines have a wide domain that not cover academic community. In addition, there are some weakness such giving incorrect outputs, ignoring Arabic diacritics and giving results in English while the searching process is done in Arabic. Therefore, according to the aim of this paper, our proposed IBRI-CASOnto search engine try to cover these issues.

### 3. Arabic language and ontological engines

Arabic language is considered as integral to the vast majority of the population of the Middle-East and the rituals of Muslims, because it is their mother tongue and the religious language of all Muslims of a variety of ethnicities throughout the world. It is also a Semitic language that has around 28 alphabets [19,20,29,21]. Moreover, Arabic is also one of the six official languages of the United Nations and the mother language of more than 330 million people in earth [22].

The Arabic Language has a collection of specialties that may obstruct the development of semantic web engines. The complexity in Arabic can be traced back to its complex morphological, grammatical and semantic aspects since it is a highly inflectional and derivational language. Because of these reasons, there are few ontological search engines available in the market and the current NLP tools can't directly accommodate the desires of the Arabic Language. Therefore, our IBRI-CASOnto tries to cater the user's needs and satieties the Arab nations based on the current approaches of developing ontological engines.

### 4. Ontology components

The ontology consists of different types of components, which could be divide into three types according to the ability to describe the entities of domain, such as Classes, Individual and Relation.

#### 4.1. Ontology classes

Classes are the core component of most ontologies. According to the different languages, which is used to implement ontologies, it is called a concept or a type. Classes represent a collection of individuals that share common characteristics. Sometime one class could be a subclass to another class. For example, if the Class

College is a subclass of the Class Organization. Then, every individual of the Class College is also be individual of Class Organization. In addition, classes could share relationships that will describe how the individual of one class relate to another.

#### 4.2. Ontology individuals

Individual represents the objects of domain of interest. It is called instance of class. Ontology is described the individual so that, it is considered as the base unit of ontology. Individual could represent concrete objects like people, machine, or abstract object like article or function.

#### 4.3. Ontology relations

Relation is often called property or slots in some system. It is describe how the individuals of classes are related to each other, or describe the way how each individual relate to specific class, or sometimes how the classes of specific domain relate to each other's. For example, the relation between classes, if we have a class person and a class country the relationship between them is lives in. That means every person lives in country. Besides, if we want to make relation between individuals related to classes. For instance, if we have individual called Ahmed in class person and in class country have Oman. If Ahmed lives in Oman then the relation will be between individuals Ahmed and Oman [23].

### 5. The proposed engine : IBRI-CASONTO

Our Semantic Search System (IBRI-CASONTO) was designed as a search engine for College of Applied Sciences (CAS), Sultanate of Oman. The system is based on the RDF dataset as well as Ontological graph. Moreover, this engine is developed for two languages Arabic and English. While, this paper is focused on designing the ontological graph more because we already mention the RDF on other paper [24]. In designing the ontological engines, there are different structures; however, most of them follow the same main steps which are designing, inference, storing, indexing, searching, query processing and the user friendly interface as it is illustrated in Fig. 2.

**Table 1**  
Ontological search engines.

Search engine	Specialty	Repository	Search approaches	Results	Voice recognition	Portability	Language support
Kngine	Knowledge Engine	Wikipedia and other sites	Knowledge-Based approach and the statistical approach	Direct answer or link to web pages	Yes	Yes	Multi-language (supports Arabic)
Google	Search Engine	Wikipedia	Hummingbird approach	Direct answer	Yes	Yes	Multi-language (support Arabic)
Wolfram Alpha	Computational Knowledge Engine	Curated data of other sites	It is own computational approaches	Direct computational Answer	Yes	Yes	Multi-language (doesn't support Arabic)

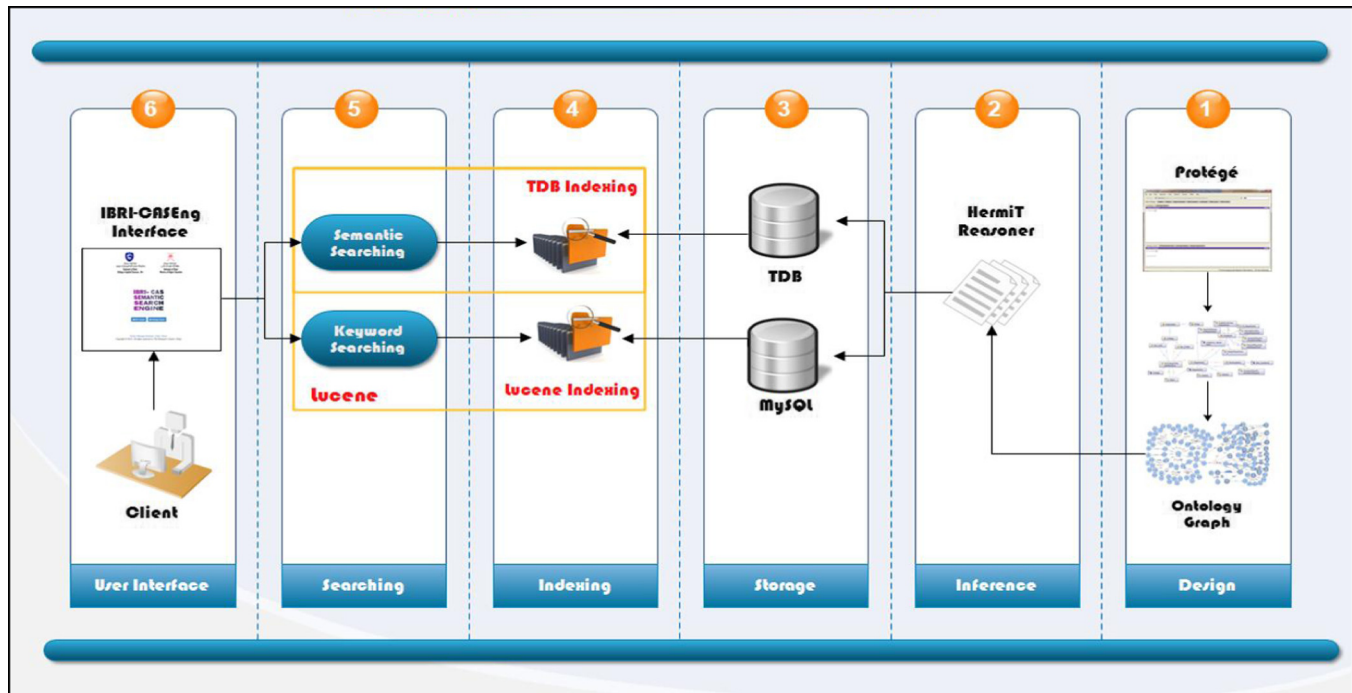


Figure 2. IBRI-CASOnto Structure.

### 5.1. IBRI-CASOnto design

Design is considered as a significant phase for developing any system. Our IBRI-CASOnto is designed based on different phases as it is illustrated in Fig. 2. In the following, we describe how each phase or step is implemented to generate our efficient and scalable ontological graph.

- *First step*, we determine the domain and scope of our ontology. We suggest the Ibri CAS (College of Applied Science) to be our domain of interest and highlight the academic department to serve our ontology specifically as a prototype of the system.
- *Second step*, determine the ontology representation language and the editor. We use the OWL to develop our ontology that is more compatible with the World Wide Web. In addition, OWL is based on the main elements of RDF in order to add more vocabularies to describe classes and properties.
- *Third step*, create the ontological graph of IBRI-CASOnto as it is illustrated in Fig. 3.
- *Fourth step*, we start the ontology by defining the Classes. Super-classes and sub-classes have been defined in protégé, each new class is sub-class from the general class that is called thing. Our IBRI-CASOnto have three main classes for English and Arabic (Person, Organization and Location) (الشخص-الموقع-المؤسسة) respectively. In addition, we define some relationships among different Classes. Some classes are equivalent to other classes. For example, in English ontology we find out that dean is equivalent to the classes AcademicAdministrator and HeadOfCollege. Moreover, in Arabic the Class (عميد\_الكلية) is equivalent to Class (رئيس\_الكلية).
- *Fifth step*, we define the instances for each class, which is called individuals. Individual is considered as a member of the class. For instance, the class Dean have only one individual that called Dean. Besides that, IBRI-CASOnto instances reach to more than 1000 individuals.
- *Sixth step*, we define the relationships or the object properties as they called in the protégé. There are different types of relations such as the relationship between classes or among the classes

and individuals. Besides, we define the domain and range for each property. Domain means the start edge of the relation while range means the end edge of the relation. For example, we define a relation, which is called headOf between the class dean and college. This relation is an inverse relation. The domain of it is dean and the range is the college. In addition, we define the equivalent for some property such as (headOf) is the equivalent to (manageOf). In Arabic, we define object property (رئيس) has an equivalent property that is called (مسؤول). IBRI-CASOnto contains more than 100 object properties among the classes and individuals.

- *Seventh step*, create the data property and define the construct, domain and range for each property. The range of the data property could be String, Number, Date, or Time act. In protégé, each new data property is sub property from the topDataProperty.

In conclusion, after we create all these thing *classes*, *properties* and *relationships* we need to interpret some things that is not understanding by ontology itself. Therefore, we move to the concept of inference. It is able to infer what owl:equivalentClass, owl:sameAs and rdf:subClassOf mean.

### 5.2. IBRI-CASOnto inference

The concept of inference is could be define as the deriving of conclusions from a given information via any suitable form of reasoning. In the Semantic Web, the inference is used to discover new relationships between the data that modeled as a set of defined relationships between the resources. It works as automatic procedures that deriving additional information by generating new relationships based on the ontology dataset. It also improves the quality of data by automatically analyzing the content of the data and discovering new relationships. In addition, inference is based on different techniques that are important in discovering possible inconsistencies of the data. Therefore, it plays a great role to reduce self-join issue among the triples. There are several automated reasoners, which can plugin inside the ontology environment such as protégé. For instance, Pallet [25], FaCT++ [26], HerMiT [27], etc. In



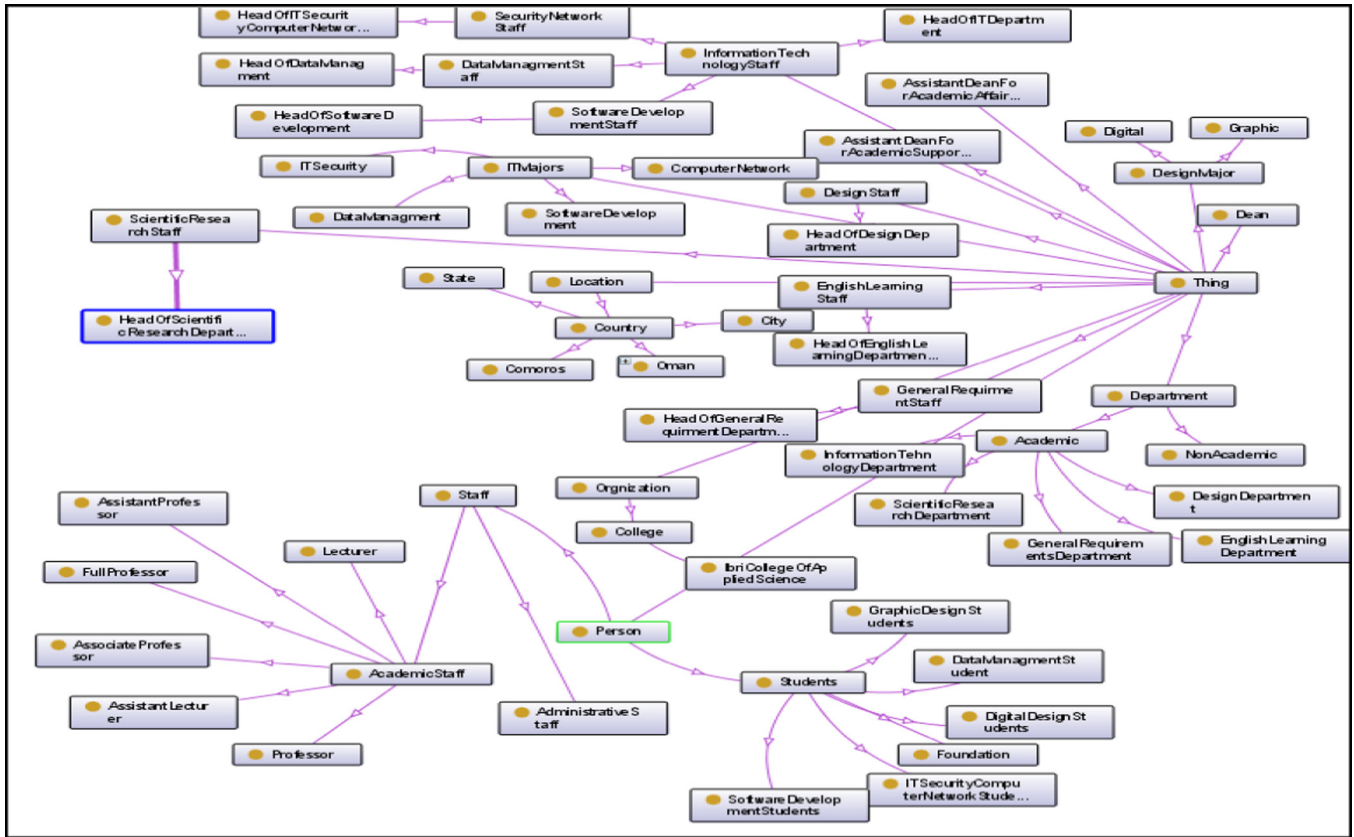


Figure 3. English IBRI-CASOnto.

our IBRI-CASOnto, we use protégé 4.3 with the plugin reasoner that called HerMiT reasoned as it is shown in Fig. 2.

HerMiT is an open source that is already plugin in protégé 4.3 and it is a perfect reasoner for ontologies, which is written in OWL. This reasoner is based on a novel “hypertableau” calculus that delivers efficient reasoning than any known algorithm. Thus, the use of HerMiT inference could help to save time and effort for developing the ontology. Moreover, it is the first reasoner that able to classify a number of ontologies which had considering as too complex for any available system to handle.

Our IBRI-CASOnto uses the HerMiT reasoner by following some of the steps which are as follows: First, open the OWL file inside the protégé environment. Second, click the reasoner tap to show the list of available reasoners. The third step is selected the HerMiT and then starts reasoned as it is shown in Fig. 4. Finally, after clicking the start reasoner wait for some time while the inference is finished. It will gives inferred SubClasses, equivalentClasses, disjointClasses, SubObjectProperty, equivalentObjectProperty, SubDataProperty, equivalentDataProperty, ObjectPropertyCharacteristic, DataPropertyCharacteristic, inverseObjectProperty, Class assertion (Individual) and Properties assertion (value) [28]. For example, the class “department” have two equivalent classes, which are division and section. Moreover, two asserted individual “Assistant Dean for Academic Affairs & Scientific Research” and “Assistant Dean for Academic Support Affairs”. In addition, the department individual is inferred all data and object property values.

### 5.3. IBRI-CASOnto storage

There are different mechanisms to store the ontology dataset. In our search engine, we use two directions the Relational Database

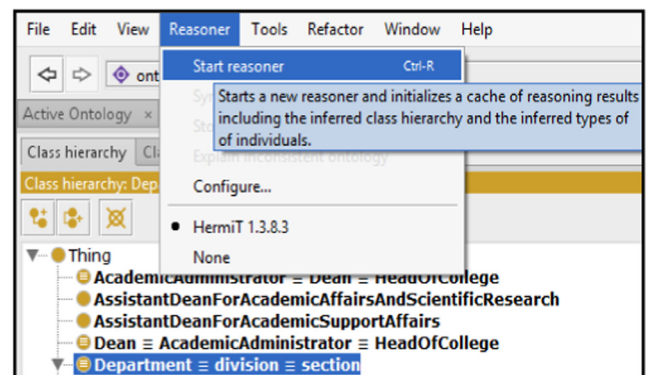


Figure 4. Inference steps.

and Triple-store as it is illustrated in Fig. 2. Triple-store means the Database Management Systems (DBMS) for data modeled using RDF. It is unlike the Relational Database Management Systems (RDBMS), which store data in relations (or tables). Moreover, the RDBMS are queried using SQL, while the triple-store stores RDF triples and are queried using SPARQL.

A key feature of many triple-stores is the capability to do inference. It is essential to note that a DBMS typically presents the capacity to deal with concurrency, security, logging, recovery, and updates, in addition to loading and storing data. However, some of the triple-stores offer all these capabilities while others not.

Inside these storage systems (RDBMS and triple-store), we choose a specific one according to our needs. Hence that the triple-store is defined as a purpose-built database for the storage

and retrieval of triples through semantic queries. A triple is a data entity composed of subject-predicate-object. There are different types of triple-stores [29] such as Jena SDB, Jena TDB, OWLLIM, Sesame and others. In IBRI-CASOnto, we decide to use the Jena TDB because it is a component of Jena for RDF storage and query. It supports the full range of Jena APIs. Besides, TDB can be used as a high performance of RDF store on a single machine. It also includes automatic protection against multi-JVM usage, which prevents this under most circumstances. On the other hand, we decide to use the MySQL as RDBMS for the keyword searching purpose.

#### 5.4. IBRI-CASOnto indexing process

Indexing is a high-level concept among the developers of the search engines, in order to retrieve the data from the ontology dataset faster as well as efficient. In our search engine, we use two ways for indexing the ontology datasets, which store in the Jena TDB and the RDBMS MySQL as it is shown in Fig. 2.

##### 5.4.1. TDB-Indexing technique

We use the TDB indexing which is built on the Fuseki for Jena TDB dataset. Many of the persistent datasets in the TDB triple-store use a custom implementation of threaded B+ Trees. The TDB triple-store implementation only provides for fixed length key and fixed length value. In addition, there is no use of the value part in triple indexes. The threaded nature is referred to the meaning of the long scans of indexes proceeds without needing to traverse the branches of the tree.

##### 5.4.2. Lucene-Indexing technique

Lucene is used for indexing the MySQL database. The indexing process in Lucene consists of a chain of logical steps after gain access to the original content you need to search. The steps are acquire the content, build content, analyze documents and index documents.

#### 5.5. IBRI-CASOnto searching process

The apparent purpose of searching is to find different mechanisms that facilitate people to extract a multitude of things that satisfies their needs. The initial search results relating consciously to users are always limited to the time and accuracy of the rendered results. In addition, the quality of a search is typically described using precision and recall metrics as we shall discuss later in the experimental results. The searching process inside the IBRI-CASOnto Ontology is implemented via two types of searching which are the Keyword Searching and Semantic Searching as it is shown in Fig. 2. In following, we give a clear background how these types of searching is worked in our system.

##### 5.5.1. Keyword-Based search

It is done by the support of Apache Lucene, which provides with the access to the Lucene indexes. This type of searching gets the matched keywords as a full-text query without understand the concept behind it.

##### 5.5.2. Semantic-based search

Semantic Searching of IBRI-CASOnto is supported by Apache Jena Fuseki. It provides a SPARQL server that can use the Jena TDB for persistent storage. In addition, it provides with the SPARQL protocols for query, update and rest update over the HTTP. Moreover, the SPARQL query offers the searching over the triple-store and retrieve the needed results.

Table 2 is illustrated that our engine is tested based on different SPARQL queries from our domain ontology. After getting, the original query from the user; the ontology is provides with equivalent

queries, which is called reformulated queries. These queries have been tested with the help of SPARQL in the SPARQL Expert Interface. They retrieve equivalent results with the original query. That means our concepts in the ontology holds different meaning based on the domain ontological area.

#### 5.6. IBRI-CASOnto interface

The user interface is one of the most important parts of any system. Our IBRI-CASOnto system provides with a usable interface that enables users to interact with the engine easily. Thus, a powerful system with a poorly designed user interface has little value that could put the system in the trap.

Fig. 5 is illustrated that our IBRI-CASOnto Ontology offers three parts of searching which are Keyword Searching, SPARQL Expert and CAS Queries. Each one of them provide with a guide that helps the user to search probably. Whereas the CAS Queries includes a set of predefined queries based on our Arabic and English ontology. The next one is the SPARQL Expert, which requires an expert of writing SPARQL Query because it forces the user to write a manual query. The last part is the Keyword Searching that retrieves the results based on the full-text matching of the query.

## 6. Experimental results

Our IBRI-CASOnto is based on two types of searching which are the classical search (Keyword-based search) and the semantic search (Entity-based search). The purpose of the classical search is to measure the matching of the keywords with the RDF dataset as well as the ontological graph. It is arranged based on the high score of matching. While the aim of semantic search is to get the exact answer from the ontological graph. In addition, it is built to understand the context of the searching text and retrieve the coherent answers without going on a maze as the classical search.

We conduct two experiments to measure the performance of our proposed search engine. First, compare the (Keyword-based search and Entity-based search) of RDF and the ontology based on simple and complex queries. Second, compare our proposed engine IBRI-CASOnto with other engines such as Wolfram-Alpha, Kngine and Google. As mentioned above, the data set used is an ontological graph that holds information about departments, staff, faculty and students for College of Applied Sciences, Ibri, Oman. As it is shown in the Table 3, the CAS-Ontology dataset contains around 31,279, which classified into 2159 subjects, 132 predicates and 5575 objects, whereas the English CAS-Ontology dataset contains around 32,322, which categorized into 3035 subjects, 150 predicates and 6507 objects.

#### 6.1. Evaluation metrics

The analysis evaluation of search engine is measure based on different metrics to get a quality model that is presented based on ISO 9126 standards for system quality. In this section, it distinguishes between three varieties of evaluation measurements which called *Recall*, *Precision* and *Accuracy*.

- **Recall:** It is referred to the fraction of the documents that are relevant to the query which are successfully retrieved (i.e. sum of all true positives and false negatives). It is known as a lexical recall or a correct recall (Rc):

$$\text{Recall} = \text{Number of retrieved relevant} / \text{Number of possible}$$

- **Precision:** This measure (Pc) is defined as the fraction of the documents retrieved that are relevant to the user's information need. It is called lexical precision or a correct precision:

**Table 2**  
SPARQL test queries.

Original Queries	Reformulated Queries	SPARQL Queries
من هو عميد الكلية؟	• من هو رئيس الكلية؟	<pre>PREFIX rdf: &lt;http://www.w3.org/1999/02/22-rdf-syntax-ns#&gt; PREFIX rdfs: &lt;http://www.w3.org/2000/01/rdf-schema#&gt; PREFIX xsd: &lt;http://www.w3.org/2001/XMLSchema#&gt; PREFIX owl: &lt;http://www.w3.org/2002/07/owl#&gt; PREFIX ibri: &lt;http://www.ibri.cas.edu.om/&gt; SELECT ?y WHERE {   ?x rdf:type الكلية   ?y } </pre>
من هم رؤساء الأقسام الأكاديمية؟	• من هم المسؤولين عن الأقسام الأكاديمية؟	<pre>PREFIX rdf: &lt;http://www.w3.org/1999/02/22-rdf-syntax-ns#&gt; PREFIX rdfs: &lt;http://www.w3.org/2000/01/rdf-schema#&gt; PREFIX xsd: &lt;http://www.w3.org/2001/XMLSchema#&gt; PREFIX owl: &lt;http://www.w3.org/2002/07/owl#&gt; PREFIX ibri: &lt;http://www.ibri.cas.edu.om/&gt; SELECT DISTINCT ?y WHERE {   ?x rdf:type قسم تقنيه المعلومات   ?y } UNION {   ?x rdf:type قسم المتطلبات العامه   ?y } UNION {   ?x rdf:type قسم اللغة الانجليزية   ?y } UNION {   ?x rdf:type قسم التصميم   ?y } UNION {   ?x rdf:type قسم البحث العلمي   ?y } </pre>
أسماء الموظفين الأكاديميين الذين حققوا مؤهل الدكتوراه؟	• أسماء أعضاء هيئة التدريس الذين حققوا مؤهل الدكتوراه؟	<pre>PREFIX rdf: &lt;http://www.w3.org/1999/02/22-rdf-syntax-ns#&gt; PREFIX rdfs: &lt;http://www.w3.org/2000/01/rdf-schema#&gt; PREFIX xsd: &lt;http://www.w3.org/2001/XMLSchema#&gt; PREFIX owl: &lt;http://www.w3.org/2002/07/owl#&gt; PREFIX ibri: &lt;http://www.ibri.cas.edu.om/&gt; SELECT ?y WHERE {   ?x rdf:type موظف اكاديمي   ?y } </pre>
من هم طلاب تخصص تطوير البرمجيات؟	<ul style="list-style-type: none"> <li>• من هم تلاميذ تخصص تطوير البرمجيات؟</li> <li>• من هم متعلمي تخصص تطوير البرمجيات؟</li> <li>• من هم دارسي تخصص تطوير البرمجيات؟</li> </ul>	<pre>PREFIX rdf: &lt;http://www.w3.org/1999/02/22-rdf-syntax-ns#&gt; PREFIX rdfs: &lt;http://www.w3.org/2000/01/rdf-schema#&gt; PREFIX xsd: &lt;http://www.w3.org/2001/XMLSchema#&gt; PREFIX owl: &lt;http://www.w3.org/2002/07/owl#&gt; PREFIX ibri: &lt;http://www.ibri.cas.edu.om/&gt; SELECT DISTINCT ?y WHERE {   ?x rdf:type طلاب تخصص تطوير البرمجيات   ?y } </pre>



**Figure 5.** IBRI-CASOnto Interface.

**Table 3**  
Dataset descriptions.

Dataset	Object	Predicate	Subject	Triples
Arabic CAS_ Ontology	5575	132	2159	31,279
English CAS_ Ontology	6507	150	3035	32,322

$$\text{Precision} = \text{Number of total relevant} / \text{Number of total retrieved}$$

- Accuracy: This metric gives a good overall view of the competency of a search engine and how accurate it is. It is computed by dividing the number of correct outputs (i.e. the sum of true positives and true negatives) by the total number of queries.

### 6.2. RDF and ontology evaluation

RDF and Ontology are the main two backbones of IBRI-CASOnto system. In our engine, RDF is designed to be a keyword-based search while the Ontology is considered as a classical search and a semantic search. In our experimental of RDF and Ontology, we classify approximately 80 queries (which exists in the paper's appendix) into two categories, which are the simple, and complex queries that based on the number of self-joins. As we have been seen in Table 4 and Fig. 6, the comparison of simple and complex queries under two types of searching Keyword-based search and Entity-based search. In the case of comparison, we depend on first answer if it is true or not, and ignoring whole the retrieved answers. As it is illustrated in Tables 4 and 5 (the important information is shown as a part of paper and the rest of details the paper's appendix), some of the queries have identical answers with different symbols, for instance, the retrieved queries that is relevant defined by the symbol  $\checkmark$  and the irrelevant queries which are not retrieved defined by the symbol X. However, the irrelevant answers that is retrieved define by the symbol – and the no responses queries are defined by symbol 0 (See the significant results that is shown in the grey rows at the bottom of the table). Systematically, it seems that the semantic search is better than the classical search in both simple and complex queries. Clearly, the accuracy equals to 100% in both types of queries. While the classical search is better with the simple queries as it is illustrated in Fig. 6 and Table 4, the accuracy has 45% however; it equals 0% in complex. We retrieved only 18 relevant results, because this searching is based on the full-text that means all the keywords should be exist in the same triple to get the result. The total relevant of queries is efficient with the semantic search; therefore our next experimental take this to compare our IBRI-CASOnto with other semantic search engines.

### 6.3. Semantic search engines comparison

“The Semantic Web is the representation of data on the World Wide Web. It is a collaborative effort led by W3C with participation

from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming”. W3C Semantic Web. The Semantic Web is a charter that allows publishing, sharing, and reusing data and knowledge on the Web and across applications, enterprises, and community boundaries.

Our experiment of semantic search engines is compared with four types of famous semantic engines, which are IBRI-CASOnto, Wolfram Alpha, Kngine and Google. We submitted 40 different queries against the tested engines that exists in the paper's appendix. As it has been shown in Table A2 in the appendix, our engine is retrieved 23 relevant queries out of 40 answers and 17 irrelevant queries are not. However, the rest of engines have irrelevant answers. The ratios of precision are comparable between the rests of the engines where Wolfram Alpha, Kngine and Google have 30, 25 and 42.5, respectively, as it is illustrated in Fig. 7 and Table 5. In addition, the accuracy of our engine is also high compared to other engine; it has 100 per cent, while Wolfram Alpha, Kngine and Google have 30, 25 and 42.5, respectively. Consequently, it seems that our engine retrieved better and efficient results than other engines. Thus, it is built according to the ontological domain-specific, highly scalable performance and handles the complex queries well by understanding the context behind the query.

## 7. Conclusion and future work

In conclusion, though new improved keyword-based technologies for searching the WWW are evolving constantly, the growth rate of these improvements is likely to be slight. Problems of imprecise and irrelevant results will continue to hinder Web searchers, especially with the continued expansion of the Web. Search engines based on a new concept as the semantic Web technology, are effectively able to handle the above-mentioned problems. A domain specific ontology based semantic search engine as ours is advantageous in several ways. Firstly, our approach has been able to successfully eliminate the problem of irrelevant results, which is one of the main problems encountered by the users of a regular search engine. By using the mapping technique between instances and classes, the search engine effectively fetches the exact information. Secondly, by producing exact information as the result, the search engine eliminates the need to go through numerous results as in case of a regular search engine. Lastly, our design although based on the IBRI-CASOnto domain, is

**Table 4**  
Performance of IBRI-CASOnto engine.

	Simple query				Complex query		
	Keyword-based search		Entity-based search		Keyword-based search		Entity-based search
	RDF	Ontology	Ontology		RDF	Ontology	Ontology
Retrieved Relevant	18	18	39	Retrieved Relevant	0	0	29
Retrieved Irrelevant	16	22	0	Retrieved Irrelevant	40	40	0
Not Retrieved Relevant	6	0	0	Not Retrieved Relevant	0	0	0
Not Retrieved Irrelevant	0	0	1	Not Retrieved Irrelevant	0	0	11
% Precision	52.49	100	100	% Precision	0	0	100
% Recall	75	100	100	% Recall	0	0	100
% Accuracy	45	45	100	% Accuracy	0	0	100



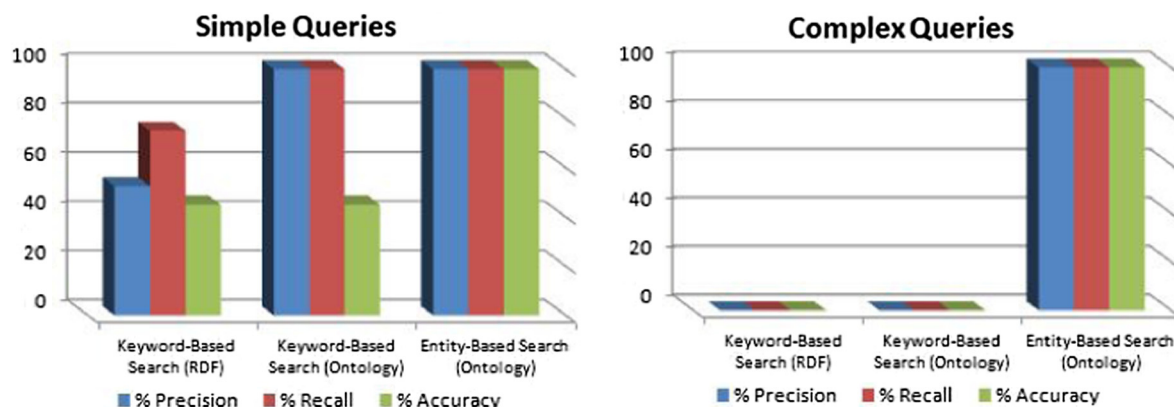


Figure 6. Classical and Semantic Search of IBRI-CASOnto.

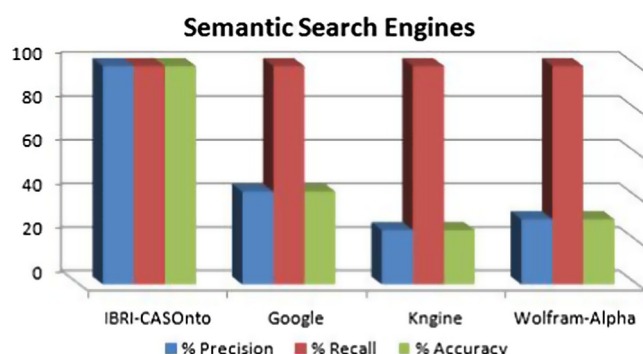


Figure 7. Performance of semantic search engines.

highly scalable and can be easily adopted by other enterprises as their site search tool. This would only require the enterprise to feed in the relevant RDF codes based on the ontology of the domain. As a result, the page containing the site search (ontology based semantic search) would be automatically generated. In the future work, we shall extend the RDF graph to contain all information about MoHE. In addition, we try to demonstrate a good indexing mechanism, which is suitable to deal with the large dataset. With the consideration of the time, store and IR, which are important in order to retrieve data, fast, scalable and efficient.

## Acknowledgments

This work is founded by TRC (The Research Council) Sultanate of Oman from 2012 to 2015.

## Appendix A

### (1) Simple Test Queries:

1. Academic staff names

2. Information Technology Faculty
3. IT Staff
4. Design majors
5. Staff nationality
6. Academic staff emails
7. Foundation students names
8. Information Technology majors
9. Head of Data Management email
10. Head of Network and Security
11. Software Development Students
12. Digital Students emails
13. Head of Information Technology department email
14. Head of English department qualification
15. Head of Design Department name
16. Head of General Requirement Department
17. Academic departments in college
18. Assistant professors emails
19. Assistant lecturers names
20. Awny sayed nationality
21. Mohamed Kayed email
22. Dean email
23. Dean nationality
24. Dean major
25. Head of Computer Sciences
26. Graphic design students
27. mayyadha qualification
28. lecturer emails
29. Governorates in Sultanate of Oman
30. Regions in Sultanate of Oman
31. Cities in Muscat
32. Cities of ALBuraymi
33. Information Technology staff majors
34. Academic majors
35. Academic departments
36. Batch 2010 names
37. Batch 2012 emails

Table 5

Performance of different ontological search engines.

	IBRI-CASOnto	Google	Kngine	Wolfram-Alpha
Retrieved Relevant	23	17	10	12
Retrieved IrRelevant	0	23	30	28
Not Retrieved Relevant	0	0	0	0
Not Retrieved IrRelevant	17	0	0	0
% Precision	100	42.5	25	30
% Recall	100	100	100	100
% Accuracy	100	42.5	25	30

**Table A1**

Performance of IBRI-CASANTO engine.

Query number	Simple query			Query number	Complex query		
	Keyword-based search		Entity-based search		Keyword-based search		Entity-based search
	RDF	Ontology			RDF	Ontology	
Q1	-	-	✓	Q1	-	-	✓
Q2	-	-	✓	Q2	-	-	✓
Q3	0	-	✓	Q3	-	-	✓
Q4	-	-	✓	Q4	-	-	X
Q5	✓	✓	✓	Q5	-	-	✓
Q6	-	-	✓	Q6	-	-	✓
Q7	-	-	✓	Q7	-	-	✓
Q8	-	-	✓	Q8	-	-	✓
Q9	✓	✓	✓	Q9	-	-	✓
Q10	✓	✓	✓	Q10	-	-	✓
Q11	✓	✓	✓	Q11	-	-	✓
Q12	-	-	✓	Q12	-	-	X
Q13	-	-	✓	Q13	-	-	X
Q14	✓	✓	✓	Q14	-	-	✓
Q15	✓	✓	✓	Q15	-	-	✓
Q16	✓	✓	✓	Q16	-	-	✓
Q17	✓	✓	✓	Q17	-	-	✓
Q18	-	-	✓	Q18	-	-	✓
Q19	-	-	✓	Q19	-	-	X
Q20	✓	✓	✓	Q20	-	-	X
Q21	✓	✓	✓	Q21	-	-	✓
Q22	✓	✓	✓	Q22	-	-	✓
Q23	✓	✓	✓	Q23	-	-	✓
Q24	✓	✓	✓	Q24	-	-	✓
Q25	X	X	X	Q25	-	-	X
Q26	-	-	✓	Q26	-	-	X
Q27	✓	✓	✓	Q27	-	-	X
Q28	0	-	✓	Q28	-	-	✓
Q29	0	-	✓	Q29	-	-	✓
Q30	0	-	✓	Q30	-	-	✓
Q31	0	-	✓	Q31	-	-	✓
Q32	0	-	✓	Q32	-	-	✓
Q33	0	-	✓	Q33	-	-	X
Q34	-	-	✓	Q34	-	-	X
Q35	✓	✓	✓	Q35	-	-	X
Q36	-	-	✓	Q36	-	-	✓
Q37	-	-	✓	Q37	-	-	✓
Q38	✓	✓	✓	Q38	-	-	✓
Q39	✓	✓	✓	Q39	-	-	✓
Q40	✓	✓	✓	Q40	-	-	✓
Retrieved Relevant	18	18	39	Retrieved Relevant	0	0	29
Retrieved Irrelevant	16	22	0	Retrieved Irrelevant	40	40	0
Not Retrieved Relevant	6	0	0	Not Retrieved Relevant	0	0	0
Not Retrieved Irrelevant	0	0	1	Not Retrieved Irrelevant	0	0	11
% Precision	52.49	100	100	% Precision	0	0	100
% Recall	75	100	100	% Recall	0	0	100
% Accuracy	45	45	100	% Accuracy	0	0	100

38. Head of Scientific Research Department major
39. Head of Information Technology nationality
40. Dean qualification
- (2) Complex test queries:
  1. Academic employees who have Phd degree
  2. Academic staff who have Bachelor degree
  3. Students who live in Al Dhahirah and study Digital design major
  4. Male Students who live in Ibri that located in South Batinah
  5. Male students from Batch 20,112 and Al Dhahirah
  6. Academic staff emails from Design Departement who have Phd degree.
  7. Male students who study Software development major
  8. Egyptian academic staff who have PhD
  9. Iraqi academic staff who have PhD degree
  10. Non-Omanis Female students from IT

11. Non-Omanis Female students from Information Technology
12. Full professors from English department and their nationality is British
13. All Female Head of academic department from Comoros
14. Omani academic staff emails from Design Department
15. Number of Omanis in the academic departments who have achieved the Phd degree
16. Female Students of Graphic Design from Muscat
17. IT staff who are lecturers and their nationality is Indian
18. IT faculty who are lecturers and their nationality is Indian
19. Egyptian student who study Network
20. Female student who study in Computer Science department
21. Omani students from batch 2012 studies Network
22. Network students from ALBuraymi and their gender is Male

**Table A2**

Performance of different ontological search engines.

Query number	IBRI-CASOnto	Google	Engine	Wolfram-Alpha
Q1	✓	✓	-	-
Q2	X	✓	-	-
Q3	X	-	-	-
Q4	✓	-	-	-
Q5	X	-	-	-
Q6	✓	-	-	-
Q7	✓	-	-	-
Q8	X	✓	✓	✓
Q9	X	✓	✓	-
Q10	X	-	-	-
Q11	X	-	-	-
Q12	✓	-	-	-
Q13	X	✓	✓	✓
Q14	✓	-	-	-
Q15	✓	-	-	-
Q16	✓	-	-	-
Q17	X	-	-	-
Q18	✓	-	-	-
Q19	✓	-	-	-
Q20	X	✓	✓	✓
Q21	✓	-	-	-
Q22	✓	✓	✓	✓
Q23	✓	✓	-	✓
Q24	✓	✓	✓	✓
Q25	✓	✓	✓	✓
Q26	✓	✓	✓	✓
Q27	✓	✓	-	✓
Q28	✓	-	-	-
Q29	X	✓	-	✓
Q30	X	✓	-	-
Q31	✓	-	-	-
Q32	✓	-	-	-
Q33	X	-	-	-
Q34	X	-	-	-
Q35	X	✓	✓	✓
Q36	✓	-	-	-
Q37	✓	-	-	-
Q38	X	✓	-	-
Q39	X	✓	✓	✓
Q40	✓	-	-	-
<b>Retrieved Relevant</b>	23	17	10	12
<b>Retrieved IrRelevant</b>	0	23	30	28
<b>Not Retrieved Relevant</b>	0	0	0	0
<b>Not Retrieved IrRelevant</b>	17	0	0	0
<b>% Precision</b>	100	42.5	25	30
<b>% Recall</b>	100	100	100	100
<b>% Accuracy</b>	100	42.5	25	30

23. Administrator of IT department who is Female and have PhD
24. Network staff and Security faculty who are Jordanian and have PhD
25. Design student names from Muscat and their nationality Comoros
26. Female Information system student emails from Information Technology department
27. Female student from Languages department who live in UAE
28. Male IT staff who have PhD degree
29. Academic staff who is head of major and have Master degree
30. Lecturers from English department who have Master
31. Information Technology major that has more than 2 PhD staff from Egypt
32. IT faculty who are lecturers from data management
33. IT employees emails from Morocco in Sultan Qaboos University
34. Male students emails from batch 2012 in Nizwa University

35. Female academic staff who have PhD degree from Oman Universities
36. Design faculty who are assistant professors and their major graphic design
37. Female Information Technology students who study software development
38. Academic department that have more than 50 student who lives in Ibri from Al Dhahirah region
39. Omani students from batch 2010 who are male and their major network
40. Omani students emails from batch 2013 who are female and their major digital design

### (3) Queries to compare Semantic Search Engines

1. Cities in Al Dhahirah
2. Oman Universities
3. Villages in ibri
4. Student nationalities in College Of Applied Science ibri
5. Student names in SQU

6. Student names in College Of Applied Science ibri
7. Staff emails in College Of Applied Science ibri
8. Leader of Sultanate of Oman
9. Sultanate of Oman President
10. Road of College of Applied Science ibri
11. Dean of Nizwa University
12. Dean of College Of Applied Science ibri
13. Capital of Oman
14. Omani students from batch 2012 who are male and their major Digital from College Of Applied Science ibri
15. Female academic staff in College Of Applied Science ibri
16. Male Lecturers from English department, in College Of Applied Science ibri, who have Phd degree
17. Colleges in Sultan Qaboos University
18. Head of academic departments in College Of Applied Science ibri
19. Student majors in College Of Applied Science ibri
20. Sultanate of Oman colleges
21. IT staff nationality in College Of Applied Science ibri
22. Regions in Oman
23. Regions in Sultanate of Oman
24. Governorates in Oman
25. Governorates in Sultanate of Oman
26. Cities in Oman
27. Cities in Sultanate of Oman
28. Academic staff who have Master degree from College Of Applied Science ibri
29. Weather in Ibri
30. Location of College Of Applied Science ibri
31. Students, from College Of Applied Science ibri, who live in South Batinah and study Network major
32. Female Student majors in College Of Applied Science ibri, who live in Ibri
33. Non-Omanis nationalities from Nizwa University
34. IT majors in Sultan Qaboos University
35. President of USA
36. Design majors in College Of Applied Science ibri
37. Administrator of academic department, in College Of Applied Science ibri, who have PhD degree
38. Location of Nizwa University
39. Currency of Oman
40. Academic staff from Egypt in College Of Applied Science ibri

See [Tables A1 and A2](#).

## References

- [1] Tom Gruber. A translation approach to portable ontology specifications. In: Knowledge Acquisition: 5; 1993. p. 199–220.
- [2] Nicola Guarino. Formal ontology and information systems. In: Guarino N, editors. Formal Ontology in Information Systems. Proceedings of the First International Conference, Trento, Italy, 6–8 June 1998. IOS Press; 1998. p. 4.
- [3] Manola F, Miller E, McBride B. “RDF primer”, W3C Recommendation; 10 February 2004.
- [4] Klyne G, Carroll JJ, McBride B. Resource Description Framework (RDF): concepts and abstract syntax, W3C Recommendation; 10 February 2004.
- [5] Hayes P, McBride B. “RDF semantics”, W3C Recommendation; 10 February 2004.
- [6] Guarino Nicola. Ontologies and Knowledge Bases. Towards a terminological clarification; 1995. p. 1.
- [7] Ontogenesis; 2010. [Retrieved 3 19, 2015, from Reference and application ontologies]: <<http://ontogenesis.knowledgblog.org/295>>.
- [8] Agrawal S, Chaudhuri S, Das G. DBXplorer: A System for Keyword-Based Search over Relational Databases. ICDE Conf.; 2002.
- [9] Antoniou G, van Harmelen F. The MIT Press Cambridge, Massachusetts London, England a Semantic Web Primer; 2008.
- [10] Giunchiglia F, Kharkevich U, Zaihrayeu I. Concept search. In: ESWC; 2009.
- [11] Ghomari LZ-G. Process of Building Reference Ontology for higher education. London: Proceedings of the World Congress on Engineering; 2013.
- [12] L, G.; 2013. Higher education reference ontology. Retrieved 5 24, 2015, from datahub: <[http://datahub.io/dataset/higher\\_education\\_reference\\_ontology](http://datahub.io/dataset/higher_education_reference_ontology)>.
- [13] Mesaric J, Dukic B. An Approach to Creating Domain Ontologies for Higher Education in Economics. In: Proc. of 29th International Conference on Information Technology Interfaces, Cavtat, Croatia; 2007. p. 75–80.
- [14] Ramachandran A, Sujatha R. Semantic search engine: A survey, IJCTA-Volume 2 Issue 6; 2011.
- [15] Alpha search engine available at: <<http://www.wolframalpha.com/>>.
- [16] Guo, Ren. Towards the Relationship Between Semantic Web and NLP; 2009.
- [17] <http://www.kngine.com/Technology.html>.
- [18] Sullivan, Danny. “FAQ: All About the New Google “Hummingbird” Algorithm | Why is it called Hummingbird?”; 2013.
- [19] Rodriguez, Horacio, et al. Introducing the Arabic Wordnet.
- [20] Beseiso Majdi, Rahim Ahmad Abdul, Ismail Roslan. A Survey of Arabic Language Support in Semantic Web, vol. 9– No. 1; 2010.
- [21] Beseiso Majdi, Rahim Ahmad Abdul, Ismail Roslan. An Arabic language framework for semantic web. In: 2011 International Conference on Semantic Technology and Information Retrieval, Putrajaya, Malaysia; 28–29 June 2011.
- [22] Saleh L, Al-Khalifa, H. AraTation: An Arabic Semantic Annotation Tool; 2009.
- [23] Horridge M. A Practical guide to building OWL ontologies using Protege 4 and CO-ODE tool. The university of Manchester; 2009.
- [24] Almuqrishi A, Sayed A, Kayed M. CASENG: ARABIC SEMANTIC SEARCH ENGINE. Published on: Journal of Theoretical and Applied Information Technology, 20th May 2015. Vol. 75. No. 2; 2015.
- [25] <http://clarkparsia.com/pellet/>, logical reasoner Pellet [accessed on march 1st 2013].
- [26] <http://owl.man.ac.uk/factplusplus/>, logical reasoner Fact++ [accessed on march 1st 2013].
- [27] <http://hermit-reasoner.com/>, logical reasoner Hermit [accessed on march 1st 2013].
- [28] W3C wiki. (2010, January 14). Retrieved 3 2015, 19, from Semantic Web tools: <<http://www.w3.org/2001/sw/wiki/tool>>.
- [29] Wikipedia; 2014, December 20. Retrieved April 20, 2015, from Triplestore: <[http://en.m.wikipedia.org/wiki/Triple\\_store](http://en.m.wikipedia.org/wiki/Triple_store)>.