Arabic Semantic Web Applications – A Survey

Aya M. Al-Zoghbya[a,*], Ahmed Sharaf Eldin Ahmed[b] and Taher T. Hamza[c]

[a]*Teaching Assistant, Computer Science Department, Faculty of Computers & Information Science, Mansoura University, Egypt*

[b]*Professor of Information Systems, Faculty of Computers & Information, Helwan University, Egypt*

[c]*Associate Professor of Computer sciences, Faculty of Computers & Information Science, Mansoura University, Egypt*

Abstract. Arabic Language is the mother tongue for 23 countries and more than 350 million persons. Moreover, since it is the language of the Holy Quran, many other Islamic countries, like Pakistan, teach Arabic as a second language. Nevertheless, it is noticed that the Arabic content on the web is less than what should be expected. The evolution of the semantic web (SW) added a new dimension to this problem. This paper is an attempt to figure out the problem, its causes, and to open avenues to think about the solutions. The survey presented in this paper is concerned with the SW applications regarding the Arabic Language in the domains of Ontology building and using, Arabic WordNet (AWN) exploiting and enrichment, Arabic Named Entities extraction, Holy Quran and Islamic Knowledge semantically representation, and, Arabic Semantic Search Engines. In fact, the study revealed significant deficiencies in the Arabic Language semantically treatment in many aspects. To mention few, most of the available tools don't support Arabic text. Moreover, very few resources are freely available. Hence, it is inevitable to put the Arabic Language in the category of the languages which machine can understand its meaning not just process its blocks.

Keywords: Semantic Web (Web 3.0), Arabic Language, Islamic Knowledge, Named Entity Extraction (NEE), Semantic Search Engine (SSE)

1. Introduction:

Citing Tim Berners-Lee, the inventor of the World Wide Web: "To a computer, the Web is a flat, boring world devoid of meaning. A new form of Web content that is meaningful to computers will unleash revolution of new possibilities .The Semantic Web (SW) is not a separate Web but an extension of the current one, in which information is given well-defined meaning.

Adding semantics to the Web involves allowing documents which have information in machine-readable forms, and allowing links to be created with relationship values" [*79*,*80*].

SW, so, can enhance the currently existing Web by the interposition of a machine interpretable layer that holds the metadata of the Web document. This metadata will permit computer software to understand what the Web page is about, and hence draw conclusions about it. This revolutionary advance can be utilized by software agents, so the information can easily be found, shared, integrated, and exchanged [*68*,*33*].

[*] Corresponding author. E-mail:elzoghby.aya@gmail.com

A significant predictable contribution of the SW, if it becomes satisfactorily actuated around the world, is the assistance in the knowledge augmentation. That's due to the utilization of the unification which will be used to express the world's concepts in a standardized logical computer language, so, anyone around the world can share these concepts using that unified language. This way, people can interact together easily and efficiently regardless of their tongue, dialect, or way of expressing [*58*,*1*,*16*]. It is, therefore, vital for all natural languages, especially those who have a broad base of utterers, to be supported by the SW tools and applications, so their users can smoothly interact, share, and exchange knowledge that is indicatively represented in a machine-readable format at the same time they dealing with the computer in their own mother tongue and their own style of expressing.

Among these natural languages that need to be supported in the SW environment, the Arabic Language. Besides that the Arabic is the language of a significant number of people who surely interested in utilizing that innovated technology of the Web; it is one of the strongest, richest and most languages able to articulate in the world. The SW, likewise, is the strongest technology has been emerged in the field of Web till now. And so, the integration of these two giants is a pursuit for the ability that cannot be emulated.

This survey reviews those studies that appreciate the Arabic Language and made efforts in order to support the Arabic Language in the environment of the SW. The reviewed studies are grouped thematically in four groups, each of which addressing an important aspect of the implementation of a comprehensive Arabic SW application. The survey addressed the researches related to the Arabic Ontologies development and usage, the Arabic Named Entity Extraction (NEE), the Islamic Knowledge Semantic Representation, and finally the Arabic Semantic Search Engines (SSEs).

As for the Ontologies, while the Ontology is the SW's backbone, and the success of any SW application depends on the design and development of Ontology [*41*], this review regards those applications that considered the Arabic Ontology as a standalone research or those depended on the Arabic Ontologies with the purpose of accomplish larger systems. Furthermore, the review considered the studies that based on the Arabic WordNet (AWN) as an Arabic Ontology or as an alternative to it, and also those tended to the development and enrichment of the current AWN.

And due to the importance of NEE tools as one of the pillars upon which many of SW applications can be built, we have dedicated a part to discuss the researches that focused on this area in the context of the Arabic Language.

Furthermore, it is well known that Arabic is the language of the Holy Quran; which honors the language itself. Therefore, it will be more valuable if the Semantic Islamic Knowledge representation researches involved. That’s why this survey contains a particularized section for them. Definitely, the exploiting of the abilities of the SW in this abundant domain will reflect a qualitative shift in the handling and treatment techniques of these significant and critical resources.

Whereas every topic from these mentioned above may represent one aspect of the SW application on the Arabic Language, the Arabic SSE is that application which may encompass all of them as sub-modules. It is the subject of the last part of the research.

Actually, this survey clarifies the inefficient handling of the Arabic Language quantitatively and qualitatively to the extent that caused its content on the Web to be unqualified yet to be used as a suitable environment for the application of the SW technology. However, this may be justified by that the strength of this language, which made it competent to carry the title of the 'Language the Quran' is the selfsame reason that makes it sophisticated to the extent that may hinder its treatment electronically in many cases.

The reviewed articles in this paper are selected according to its relevance, currency, and significance. As for the organization methodology, they are organized thematically, then chronologically. Also, the comparable results or methodologies are grouped. Some abbreviations are used in this research, so we will list them here for convenience.

Suggested Upper Merged Ontology (SUMO), Arabic WordNet (AWN), Princeton WordNet (PWN), Semantic Web (SW), Named Entity Recognition (NER), Named Entity Extraction (NEE), Semantic Search Engine (SSE), Google Translation API (GTA), Word Sense Disambiguation (WSD)

2. Semantic Web (Web 3.0)

The World Wide Web (WWW) dramatically changed the capability of accessing information electronically. Currently, it contains around 3 billion Web documents and continues to increase. These documents can be accessed internationally by over 300 million users. Certainly this voluminous of data is constantly increasing the difficulty of finding, accessing, and presenting the required information by a wide range of users. That's mainly because of the representation of the information in just a human-readable form, rather than to be machine-readable also. This means that, among a heap of information that able to be extracted by machine, the user himself who must do next filterization step. He must extract his own desired information or what he exactly looking for in this heap [*15*]. In fact, the Web can be at its full ability only when its data can be found, shared, processed and understand by means of human and machine alike [*40*,*25*].

This vision of the Web can be reached by the new generation of Web, means Web 3.0 or SW, which is the promising solution to reduce that gap between human-readable information form and that machine understandable. It represents the Web content in a more easily machine processable form and uses intelligent techniques to take advantages of that representation. That's by adding a semantic layer to the existing Web pages for holding their contents, so the Web page will contain the formatting information for human reader presentation, as well as the information about their content, metadata, for machine understanding [*27*]. Since the 'metadata' comprises chunk of the data meaning, the term 'semantic' hence arose, and 'SW' materialized. The approach used to adding that semantic layer is the Semantic Annotation [*20*], which is the process of labeling Web pages with the semantics of their contents through mapping the data instances to ontological concepts in a predefined Ontology [*23*,*78*].

The Ontology is one of the corner stones of SW applications. For a certain domain, the Ontology enumerates and gives semantic descriptions of concepts in that domain, defining domain-relevant attributes of concepts and various relationships among them. To this end, the World Wide Web Consortium (W3C) developed formal specifications such as RDF, RDFS, the Web Ontology Language (OWL), and the SW Rule Language (SWRL). These enable a formal description of the concepts, terms, and relationships within a knowledge domain [*3*,*10*].

In the context of Web, Ontologies give a shared understanding of a domain, which is necessary to overcome the terminological variations that emerge from the different cultures, tongues, or expressing ways of Web users by mapping terminologies that referred to a particular concept to that concept represented in the Ontology. Furthermore, the Ontologies are used to improve the Web search accuracy. That's by searching for Web pages which holding certain concept that represented semantically in an Ontology instead of search using ambiguous terminologies and keywords.

The SW has two more technologies over the 'metadata' and 'Ontologies'. They are 'Logic and inference' technology, by which automated reasoners can infer conclusions from a given knowledge, and 'Agents', which are independently work on behalf of a person computer programs as they can take tasks to accomplish, make certain choice and give answers [*43*,*22*]. Hence, the SW will enable the automatic collection and correlation of object's various information parts that available at various Web resources. Definitely, that will save the time spent on navigating the World Wide Web just to obtain particular information that exists somewhere out there.

Surely, multifarious applications can take advantage of considerable benefits of the SW. The SW applications can be defined as a software application that uses and produces information for the SW such as: Semantic Annotation, Semantic Communication, Semantic Search, Semantic Integration, Semantic Personalization, Semantic Proactivity, and Semantic Games [*60*,*8*,*26*].

Undoubtedly, there are many challenges faced by these applications that aim to leverage the advantages of the SW. Of these challenges: the Ontology development and evolution, the SW content availability and scalability, the visualization, the SW language standardization, and of course, the multilingualism [*42*].

As for English or French Languages, there exist large grounded environments that support the creation of their SW applications. For the Arabic Language, this is unfortunately not the case, as just some tools are adapted to support it. For instance, there is an obvious lack in Arabic editors for OWL language and RDF files, and even for tools that support Arabic character set; it does not fully support right-to-left script. Also, there is almost no Arabic

Language parser that supports RDF files in semantic editors. Moreover, there is no Arabic metadata definition similar to DCMI in English. Last but not least, there is no open-source for Arabic SW software's tools and Web services [*59*].

The main reason for that obvious inadequate support of the Arabic Language, in our point of view, is the difficulty of electronically dealing with the sophisticated particularities of such rich and strong language. That will be discussed in more details next sections.

3. Arabic Language and SW

Arabic Language is the official language of millions of people and is the religious language of all Muslims. It is a Semitic language of 28 alphabets, and it is one of the United Nations official languages [*48*] [*49*].

The Arabic Language has a set of specialties made it difficult language and may obstruct the development of SW tools for it. Among these specialties, its complex morphological, grammatical, and semantic aspects since it is a highly inflectional and derivational language. Therefore, the NLP tools that designed for the English Language can't directly accommodate the needs of the Arabic Language. Moreover, the Arabic Language has not capitalization property, which directly influences and complicates the identification of the Arabic Named Entities, as will be presented in a later section. Furthermore, the Arabic Language is highly ambiguous by several reasons. One of these is the vowelization feature of the Arabic Language, which causes the ambiguity when it absent, and this is a usual case. Other inducer of ambiguity in Arabic is the Polysemous, or multiple meaning words, which are words that share the same spelling and pronunciation but have different meanings [*66*,*18*]. Another issue that affects the SW tools processing for Arabic script is the problem of encoding, since different encodings for Arabic script exists on the Web [*11*]. Over and above, Arabic resources such as Corpora, Gazetteers, and NLP tools are either rare or not free, which leads to wasting a lot of time in the process of collecting and/or modeling and developing these resources. Consequently, SW essential techniques such as NER, which basically depend on such resources, could be affected [*52*]. All these factors collectively will certainly affect the availability of SW applications implementation.

Even so, the Arabic Language worthwhile sparing no effort to overcome these obstacles in order to transfer the success reached by the SW in multiple domains, such as Medicine, e-Commerce, e-Learning and Biology, to those of Arabic. What is promising is that the Arabic content on the Web is continuously increasing, so we must exploit this large amount of information to extend the success of SW applications to the Arabic Language. To that end, SW tools and applications that support Arabic have to be created to fulfill the requirements of the Arabic Language [*17*].

3.1. Arabic Ontology building applications

As stated above, Ontologies are one of the essential blocks in building a SW application since they present a well-defined and standardized form of the interoperable and machine understandable repositories. In general, there are two main kinds of Ontology, the 'domain specific Ontology, which represents the particular meanings of terms as they interpreted in that domain, and the 'upper Ontology' which is a representation of the common concepts that are applicable via a wide range of domain Ontologies [*56*].

As for the Arabic Language, a recent statistic made by OntoSelect Ontology library shows that there is a lack of Arabic Ontologies library, and about 49% of Ontologies are created for Latin character set [*72*]. Since each language has its own linguistic environment and cultured context, so each language needs its own Ontology. As most developed Ontologies are in English, there is an urgent need for the development of Arabic Ontologies to be used as the base of the Arabic SW applications.

Several researches concerned the building and population of Arabic Ontologies. Actually, the vast majority was domain specific. The subsequent section will review the most salient in a thematic chronological order.

In the legal domain, S.Zaidi et al. at [*64*] presented a Web based multilingual tool for Arabic Information Retrieval based on a legal domain Ontology as an attempt to improve both recall and precision of the search. The tool tries to minimize the level of noise in the results of search on the legal domain using an Ontology-based query expansion. The legal Arabic Ontology is constructed manually by Protégé 2000 using a top-down strategy in which the hierarchal concepts are related via is-a relationship. For the Ontology population, the United Nations Arabic articles are used as well as some Arabic newspapers.

The built Ontology is used as the base of a subsequent user-query expansion stage in the built search system. So, if the initial query was " قانون العقوبات", the ontological synonyms of the two legal concepts 'قانون' and 'العقوبات', will be used to extend the initial user query. Moreover, the ontological hyponyms and hyperonyms may be used in that extension. Note that a hyponym is a word or phrase whose semantic field is included within that of another word, its hypernym (sometimes spelled hyperonym). Using that way, the original user query will extended to be :" قانون العقوبات ، قوانين، عقوبة، العقوبة ، يعاقب، نص ، نصوص، تشريع ، حكم، أحكام، مخالفة، مخالفات، جنحة ، جنح، تخريب، اعتداء ، اعتداءات، قتل ، جريمة، جرائم ، قانون_عام_داخلي" . That's improved the system's results significantly as will clarify later in the Arabic Semantic Search Engines section.

The agricultural domain in Arabic is conceptualized in other domain Ontology that is presented at [*69*] **.** This article proposed a system that automates the process of constructing a taxonomic Ontology using a semi-structured domain specific Web documents. The system builds an agricultural domain Ontology using a set of 180 Arabic extension documents with 3817 HTML headings and 30 seed concepts representing the main concepts of the agricultural production. The Ontology is constructed using two complementary approaches. The first is to utilize the phrases structure that appears in the HTML headings of the used documents. The second is to utilize HTML headings' hierarchical structure for identifying new concepts and their taxonomical relationships between the seed concepts and between each other also.

The proposed system includes 7 models that are briefly acts as follows: the 'Heading Extractor' extracts heading from input HTML. A taxonomical Ontology is then extracted in the 'N-gram based Ontology Learner' module using the N-gram phrases in text headings. The 'Ontology Refiner' is a module filterizes the extracted Ontology in order to remove noisy or fake concepts. Using the heading structure of the input Web documents, the 'HTML Structure Based Ontology Learner' module learns an Ontology which will be extended by discovering new concepts that have sibling relations with previously learnt concepts in the ' HTML Ontology Refiner' module. Finally, the 'Ontology Merger' module takes the output of 'N-gram based Ontology Learner' and 'HTML Structure Based Ontology Learner' and merges them then adjusting the hierarchical structure of the resulted Ontology by discovering the right pattern and level of the concepts existing in the merged Ontologies.

According to the authors, the best obtained result for the lexical evaluation was 72.86 % for precision and 80.52% for recall. For the taxonomic evaluation, the F-measure was 75.66% as 69.08% for precision and 83.62% for recall.

Fatma Zohra Belkedim et al. in [*14*] and [*13*] addressed the Arabic Ontology development using verbs and roots, as the verbs are classified by their derivation rules in Arabic from roots, and that 85% of Arabic words are derived from tri-literal roots.

Beside the authors did not present an implemented model to support their hypothesis, the use of the derivational roots as a base for building an Ontology is imprecise since those derived words, despite they hold the same core meaning, but they could be categorized under different classes. The word 'عِلم', for instance, does not belong to the 'Human' class that encapsulates the word 'مُعلِّم', although they have the same root.

The research [*39*] propounded additional model for building and using domain specific Arabic Ontologies. This time, it was for the 'location domain'. It is a research that presents an Arabic semantic annotation tool called AraTation for semantically annotating Arabic News content on the Web documents.

The system accomplished on two stages, the first for the Arabic 'Information Extraction (IE)' to recognize Named Entities, and the second is the 'Semantic Annotation' that maps the extracted entities to the related ontological instances. The annotated documents saved as an RDF form, so it can be reused and machine processable on the Web.

As has been described earlier, the annotation process can't be accomplished without an Ontology used to mapping instances with its concepts. The authors of [*39*] found that the best suit is to build their own domain Ontology, which is built using protégé-OWL editor.

In fact, this study suffers from a gap between its Information Extraction (IE) module and the Annotation module. The IE module designed to extract four word types: Person, Organization, Company, and Location, while the Ontology is limited to just the Location domain.

The research results were evaluated by precision and recall measurement. However, the sample size was limited to the extent that the reliability of results can't be ensured; it was performed on just 25 documents. The results showed that the average of precision and recall is 67% and 82% respectively. The authors ascribed that rates to the exploiting of part-of property in the Ontology, and we thought that the exploitation of is-a property to addressing the issue of synonyms and derivations will enhance the results as much as possible.

For the Islamic domain, several researches have presented an ontological representation to the Islamic knowledge. For instance, the research [19] represents the field of 'opposition terms' in the Holy Quran, while [45] represents the 'Time Nouns' field. The research [2] on the other hand, depicts a conceptual graph as a wider Ontology for the Holy Quran as a whole. The authors of [29] regard the multilingual Ontology for Islamic portal. We assign a devoted section for discussing these researches due to their significance.

In [41], Lilac Al-Safadi et al provided a model that representing the Arabic Knowledge in computer technology domain, that presented in the Arabic blogs on the Web, in an Ontology. The research conducting an experimental study on a number of randomly selected Arabic blogs in the computer technology domain. This was to determine the nature of the frequent terminologies used in the blogs.

The study exhibited that the today's Arabic blogs' language is the modern rather than the traditional Arabic Language. Consequently, the researchers developed domain Ontology for combining both traditional and modern Arabic terminologies in the computer technology domain. The authors deemed that building a SSE for blogs using traditional Arabic Language is insufficient and it must utilize the Ontology that built for blogs.

The developed Ontology consists of 110 computer technology specific classes and 78 individuals as instances of predefined classes. 48 object relations were defined such as: has-logo:لها شعار , produced by: تنتجه شركة.

It is initially provided as a core, then enabled to be extended by the users' reuse and populate. The Ontology's relevant terms gathered by domain users, sources such as Computer Terms Dictionary and others, and the content of English Ontology that translated by Google English-Arabic translation tool. The Ontology is built using protégé 4.1 editor and tested by protégé 3.44 since protégé 4.1 capable of displaying Arabic text but does not include SPARQL query panel. Actually, that is one of the evidences on the weak support of Arabic Language on the SW technologies and tools.

The results of the Ontology testing show the Ontology's ability to bridge the semantic gap. We thought that it is a good core for building a SSE for the Arabic blogs.

On the other hand, [21] is another work that developed an Arabic Ontology in the computer technology domain. It, however, deals with the classic or standard traditional language rather than the blogs modern language. Also, the presented Ontology is much simpler and doesn't contain that amount of classes. The presented system will be discussed in more details as an Arabic SSE later.

As for [56], it is a project that stated at 2010 in Birzet University and still under construction. Its aim is to develop an Arabic Linguistic or Upper Ontology rather than particular domain Ontology, as most related researches done.

In fact, many articles, which are based on the use of Arabic Ontology in its systems implementation, treated the AWN as an Ontology. However, [56] differentiates between them. It indicated that AWN can sometimes be used as an Ontology due to its semantic beside lexical relationships. However, it is neither ideal nor optimal since it lacks formalization and suffers from many fundamental ontological problems. For instance, it, unlike the Ontology, lacks the Hyponymy subtype relation. Moreover, it contains just limited number of words comparing with its English counterpart (PWN). Also, it is created using the translation methodology which is ineffective in many cases. That's because different words from different languages may evoke the same concept. Also, not all concepts are shared cross languages since they are mainly dependent on the culture of that language's users.

According to these problems of AWN, the article [56] aim is to develop an Arabic Ontology that is logically and philosophically well founded. Its top level is defined from the known Top Level Ontologies: Suggested Upper Merged Ontology (SUMO) and DOLCE. The semantic relations of the Ontology are well defined mathematically. Moreover, the content and structure of the glosses is strictly based on ontological principles.

The Ontology is building via a four-steps approach. The first is to mine Arabic concepts from dictionaries. This step collects as much glosses /concepts as possible from specialized and general dictionaries. The selected dictionaries should focus on the semantic aspects where multiple meanings are mixed up. To conduct this step, manual mining via

Scan and OCR process is done first, then basic cleaning done automatically. Examples of that have been used dictionaries are: معجم البلدان، معجم الحاسبات، تعريف مصطلحات القانون الخاص، المعجم الوجيز and others. Step two is to automatically map these Arabic concepts and WordNet concepts using a smart algorithm that takes as an input the Arabic gloss and 117k English glosses in WordNet. It outputs the best matches and its rank with accuracy of +90%. The next step is to reformulate these glosses according to strict ontological guidelines. Step four then links all concepts with the Arabic Core Ontology. The top levels of the Arabic core Ontology is built manually based on DOLCE and SUMO upper level Ontologies. It takes into consideration the philosophical and historical aspects of the Arabic concepts terms.

The Core Arabic Ontology (Top 10 levels, ~420 concepts) is being evaluated, and the lower levels are evolving rapidly, but will never be complete. Many challenges and future work are considered as increasing the Ontology size and quality as automatic as possible, and to consider concepts from different Arab countries/communities/eras.

3.2. Arabic WordNet (AWN)

Concepts are the organizational units in the WordNet. They are more than just words, since they consist of compounds, collocations, idiomatic phrases and phrasal verbs. That extends the idea of storing just words to storing its conceptual information [*17*].

The conceptual information of the words can be fully detected via its both meaning and context. So, linking words to appropriate senses may help in figuring out that conceptual information. These senses hold the identical meaning of the word and can be linking by means of lexical relations between synsets or synonyms sets.

The WordNet is that lexical resource that offers wide range coverage of the general Conceptual Information, which make it a basic resource for many Information Retrieval tasks that facilitate the SW functions. WordNet is neither a traditional dictionary nor a thesaurus; it rather combines the features of both. As a thesaurus, the synsets involving all expressing words for a certain concept. As a traditional dictionary, it gives a definition and sample sentences for most of its synsets [*66*].

The successfulness of Princeton WordNet (PWN) opens the way for other promising projects such as the Global WordNet project that seeks to producing and linking of all the world languages. As for the Arabic Language, the AWN is a linguistic resource for Modern Standard Arabic with a semantic foundation. It based on the PWN and linked to SUMO [*65*,*9*,*32*].

As given before, we indicated the viewpoint of the author of [*56*] in the distinguishing AWN from Arabic Ontology and that AWN is neither an ideal nor alternative Arabic Ontology, in the case of its existence. The following two sections present the articles that dealt with the AWN different ways. Some tries to exploit it to create new systems, and others made its development and enrichment their objective.

As an attempt to emulating the AWN, AyaSpell-dic in [*77*] presents an Arabic thesaurus project[1] that aims to provide a list of Arabic synonyms for use in free software. It collects linguistic data from printed dictionary. Now, it is programming scripts to convert data into OpenOffice thesaurus format, and testing thesaurus in OpenOffice on Linux and Windows. It planned to get more data from Arabic dictionary, check data consistence more tests and search suitable applications for thesaurus. Unfortunately, it can't be used as Ontology since its synsets are not mapped to an upper level Ontology and it hasn't hyponym/hypernym relationships, which can be interpreted as specialization relations between conceptual categories.

[*6*], on the other hand, proposing a framework that can understand Arabic Web content using AWN Ontology as a core system that can be used in semantic applications such as: SSE, semantic encyclopedia, Arabic QA systems, semantic Dictionaries…etc.

The main goal of the built framework is to be able to convert any Arabic content into a conceptual structure that can be machine understandable. It needs in its main components a Word Sense Disambiguation (WSD) and the AWN as an Arabic Ontology. However, the authors found that the existing AWN doesn't meet their needs, and decided to perform a customization on it. One of the customizations is to use a specific stemming algorithm to store the stem. That ascribed to the diversity of the strategies used for storing stems in the original AWN so the authors need to make their

[1] http://groups.google.com/group/ayaspell-dic/browse_thread/thread/10e727f4c7a6dc0b?pli=1

own. Another customization is to merge the AWN with PWN to find interlingual translation. Actually this step is preventable since that is already available in the AWN; it is basically built upon the PWN. The third adjustment is to denormalize the AWN for the retrieval speeding up. The last is to paring the voweled word with that without vowels to maximize the findability, and that's a good idea that may maximize the utilization.

It used a Tokenization and Indexing module as a preprocessing before the WSD stage working out. Then the 'Micheal Lesks' algorithm is used as the basis for WSD process. The WSD module faced some problems in dealing with Arabic plurals and conjugations. Actually, those problems are solved perfectly in, e.g., the toolkit of RDI[2]. A further module is then used to measure the related similarity between two conceptual contents. A hierarchal clustering is performed after using the Bisecting K-means algorithm. An Encyclopedia named Arapedia was developed to test the framework. The results show that searching for "معدن الذهب" returns contents related to "ذهب", as the 'gold', not those related to "ذهب" by means of 'went'.

3.3. Arabic Named Entity Extraction

The main goal of the SW is to annotate the data on the Web with predefined Ontologies to get a machine readable, understandable, and processable Web. That will enable computers, software agents and human to work cooperatively through sharing knowledge and resources [*24*] [*4*] [*61*] [*5*]. Most of data that need to be annotated to concepts are in the form of Named Entities such as Person, Location, Organization, etc. These entities need to be extracted by means of NEE tools [*53*]. However, the sophisticated characteristics of the Arabic Language that addressed before may sometimes complicate the process of Arabic Information Extraction and Named Entities Extraction which basically based on the morphological, grammatical or semantic analysis of the Arabic Web documents. Nevertheless, respectable attempts have been performed to design and execute ANEE systems, and this section concerned them. Some of them used the ANEE as a tool for improving and enriching the AWN content, for instance [*38*], [*37*] and [*54*]. Others, on the contrary, used the AWN as the base of the ANEE tool execution as [*52*]. There is a third group targeted building ANEE tools based on other techniques as machine learning as [*7*], [*71*], [*34*], and [*76*]. The reviewed researches are organized according to the comparability of methodologies or results. Let's start with the first set:

The main idea of the authors of [*54*] and [*55*] is the automatic extraction of Arabic Named Entities (NEs) from the Arabic Wikipedia, automatically attaching them to AWN and automatically linking them to PWN.

The global architecture approach is to extract the candidate instances from PWN, removing the generic types that haven't Arabic counterparts, or adding them manually to the AWN. WSD process then performed depending on the 'Extracting Topic Signature' module. That will results in a set of English Named Entities with disambiguation information attached. The following process is the 'Filtering Candidates' which is the core of the approach. It is based on a local copy of Wikipedia for both Arabic and English loaded into a database. The process uses English NEs to lookup for a corresponding English Wikipedia page, once it founded, it looks for the occurrence of an 'interwiki-link' to an Arabic page. The title of the corresponding page is returned as the Arabic Named Entity. Since the originally built AWN has been voweled , the authors' thought that it is more appropriate to doing the last step in their approach which is the vowelization, so the produced English-Arabic NE pairs will be vowelized to be uniformed with the original AWN.

The system shows that 3,854 Arabic words corresponding to 2,589 English synsets were recovered. 3596 (93.3%) considered to be correct, 67 (1.7%) wrong, and 191(5%) are not known by the reviewer. Actually, it is an excellent work, at least for dealing with the Arabic Language and facing its problems and difficulties, as the polysemy, WSD, and vowelization, and trying to solve them that way.

However, the research is restricted as it just considers the Arabic Named Entities that have English counterparts in the English WordNet. Other Named Entities that have interwiki-links between Arabic and English Wikipedias are extracted the same way, but attached as direct hyponyms of the corresponding generic synsets. However, they will lack correspondence to the English Named Entities. That means that the Arabic Named Entities must have interwiki-links between Arabic and English Wikipedia to be extracted. So, if the Wikipedia article is originally constructed in Arabic and haven't

[2] http://www.rdi-eg.com/

interlinked to a corresponding English page, the Arabic Entities can't be extracted. The main reason of that shortcoming is the dependence on the English Language in the extraction of NE. That's for the sake of the capitalization feature that distinguishes the English and enabling it to determine entities so simply, while the Arabic Language haven't a similar feature and that is one of Arabic Language's difficulties and challenges.

As for the weakness caused by the limited number of documents in Arabic Wikipedia compared to the English Wikipedia, it were overlooked by the author out of the high growing ratio of Arabic Wikipedia, and that can be applied to progressively improve Named Entities coverage of AWN. But, while The Arabic Wikipedia grows constantly, the translation may be not always kept pace. Therefore, the same reason that justified by the author to be not bothered by the problem, could lead to another problem.

Other researches that addressed the problem of the AWN enrichment were presented by Lahsen Abouenour et al. in an Arabic Q/A system presented at [*38*] and [*37*]. In fact it was a mutual usefulness; the AWN was being developed at the same time of the user query expansion in the designed Arabic Q/A system.

The authors depend on the Yago Ontology in carrying out their research. The papers investigate the impact of the AWN enrichment using Yago Ontology in the context of Arabic Q/A systems. Yago is a large extendable Ontology of high quality, which contains about 3 million entities with 120 million facts. The authors explained their dependence on Yago Ontology by that the use of just NER system allows only the identification of NE, whereas the use of an Ontology like Yago enables the identification of the semantically related synsets.

While [*54*] used the English WordNet (EWN) as the ontological base of the enrichment of the AWN, and passing through the English /Arabic Wikipedia, [*38*] on the other hand used the Yago Ontology instead and overstep the use of English Wikipedia to be translated into Arabic Wikipedia and extract the Arabic entities from by the use of Google Translation APIs (GTA) directly to translate Yago English entities to the corresponding Arabic. Moreover, Yago offers not just a huge number of entities, but facts also. So, it is a rich resource. Indeed, the use of Yago was a smart choice, since Yago itself depends in its construction on the English WordNet. So, the use of Yago was as a start from where others end. It exploited the great effort that exerted in Yago and built upon it. In addition, Yago results were much better than those of [*54*]. Whereas [*54*] starts with just 16,873 English NEs, Yago has 3 millions. That's because the WordNet contains mainly concepts rather than entities or instances. Unfortunately, however, [*38*] did not take advantage of Yago to the fullest extent. It was not translate it as a whole to an equivalent Arabic NEs knowledge base, instead it just use it when facing an Arabic user query. Once get it, Arabic NEs in it are extracted, then translated into English using GTA. The translated entities can then extracted by Yago with their related facts. These then retranslated using GTA into Arabic NEs, and then mapped to their related entities in AWN according to synonymy, hypernymy, hyponymy and SUMO relations.

As to results, the system shows that the accuracy of question extending and answering is 23.53% using Yago, while it was 17.49% without Yago. Likewise, Mean Reciprocal Rank (MRR) is improved after using Yago to be 9.59 where it was 7.98 before. Furthermore, the number of the answered questions increased to be 31.37% using Yago, where it was 23.15% without. As we stated before, the adoption of both [*38*] and [*54*] on the basis that built upon the English Language caught in the trouble that they can't capture those entities that originally created in Arabic and haven't corresponding English. When the authors of [*38*] were asked about this problem, the question was, "What if the Arabic Named Entity in the user query does not exist in Yago KB?. Is it a potential case? Since Yago depends originally on the English version of Wikipedia, so is this sufficient for all Arabic Entities. This means that some Arabic entities may not appear in the English Wikipedia". The answer was:" Yes you are right! But in our case, since TREC and CLEF questions were used, using in the tests and since these questions come from European and American culture all NEs exist in principle in Yago. Now if we want use the system with questions that contain Arabic NEs we have to extend AWN relying on Arabic NE resources. Even though, we have a great number of Arabic NEs in Yago". While the answer of the second author was: "Yes that's right, not all Arabic entities appear in the English Wikipedia. The goal of the project was the looking for a full coverage of Arabic Named entities but extend as much as possible the AWN with available entities and see if this extension has an impact in the query expansion module of a QA system"

The impasse can be relatively overlooked in the case of [*38*] that is because of the enormous gains that can be obtained from the exploitation of the powerful rich resource Yago. Although it, in my own view, has not been exploited as it should.

With reference to [*52*], Mohammad Attia et al, demonstrate Arabic NEs repository construction system. The main difference between it and the aforementioned papers is that it depends on the use Arabic resources, AWN and Arabic Wikipedia, in the automatically NEs extraction process, while, on the contrary, those who start with English NEs collected from English resources, either PWN or Yago, so they can't capture Arabic NEs originally created in Arabic and have not English counterpart.

The used methodology composed of several steps beginning by the 'Mapping' step, which maps the identified nouns of AWN that can instantiate NEs to the corresponding categories and hyponym subcategories in Arabic Wikipedia. Next, the 'NE identification' step identifies which of the articles of these categories are NEs, which then extracted, connected to AWN and inserted in the NEs repository. This is done by exploiting Wikipedia inter-lingual links to locate correspondence between articles in ten different languages to identify the NEs. As mentioned before in [*54*], this method is commonly used due to the lack of something like the English capitalization feature in Arabic. This feature is the magical tool that can resolve many problems in the NER systems. For Arabic articles that don't have a correspondence in any of other languages, and they are with ratio of about 37% of the Arabic articles, other two heuristics used which are: 'Keyword Searching' in that article's abstract and using the 'Geonames' for looking up entities. Subsequently, the 'Post-processing' step requires further NEs from Arabic Wikipedia which are not reachable through AWN. The diacritization, in which the acquired NEs are diacritized, is then takes place. Finally, the vowelization stage, as that of [*54*], is performed to uniform the extracted NEs with those originally in the AWN. Here it is done by means of matching with Geonames.

While the system presented in [*54*] extracted 3854 Arabic NEs, about 45000 NEs extracted by that in [*52*]. Also, it overcomes the problem of originally compiled NE in Arabic and has no English equivalent, as it starts from the Arabic Wikipedia directly. However, the research has some weak points, such as it just deals with the extracted NEs as a lexicon repository, and there are neither ontological facts nor relationships. As for the restriction caused due to the limitation of Arabic Wikipedia articles number, and hence the abundance of extracted NEs, it didn't take a lot of author's attention since, as the author, the process is automated, so the NEs number will grow as the growth of Arabic Wikipedia itself. Eventually, there is an obvious discrepancy in the number of available categories here, which are limited to Person, Location, and Organization, and those presented in a knowledge base like Yago. This is another factor that affects the extracted Arabic NEs abundance.

As for both 'machine translation' approach, that is implemented at both [*38*] and [*52*], and 'rule-based' approach, that will be discussed later in this section, they work effectively only if abundant large size corpora is used as the base of extraction. In the case of machine learning techniques, on the other hand, they exploit the use of only a set of language features as well as a small set of training data to get accurate results [*7*]. Here, [*7*] and [*71*] are considered as other 'machine- learning' approach researches.

Regarding to [*7*], the authors considered NER system that based on a self-training and semi-supervised learning approach. The idea of the research emerged from the author's conviction of the importance of the existence of more other categorizations to the entities rather than the common categories, means: Person, Organization, and Location (POL). As he said: "Non-POL entities that stand out in the history- domain name, e.g., important events (wars, famines); cultural movements (romanticism); and political, religious, scientific, and literary texts. Ignoring such domain-critical entities in some sense misses the point".

Actually, the documentation and organization of the classes strictly requires intensive efforts, especially with the continuously growing of the number of domains. Based on that, the authors considered an NER system that based on a self-training and semi-supervised learning approach. The system identifies only mentions without associating it to predefined categories. The system also developed a small corpus of Arabic Wikipedia articles annotated for named entities. That to facilitate an annotation schemes that allow annotators to define entity classes more freely. The corpus or dataset is made to provide a test bed for the new NER models evaluation.

Since the machine-learning is not the topic of this survey, we will not go into the deep details of the system nor its approach of conducting. However, we thought that it is a good methodology to determine and categorize the huge amount of daily produced NEs in circumstances that make pre-defined categories restriction is a disappointing attitude.

We thought that the quality of those systems that depend on Yago may be in the middle between the two extremes; 'machine learning' approach and 'WordNet-based' approach.

As to [*71*], it is another research that exploits the 'machine learning' approach in tackling the Arabic NER problem. It integrates two machine learning techniques, namely, 'Bootstrapping semi-supervised pattern recognizer' and 'Condition Random Fields (CRF) classifier' as a supervised learner.

The pattern recognizer extracts all expected patterns to let CRF identify more Named Entities. The system supported by RDI toolkit[3] to deal with some difficulties of the Arabic Language.

The NER system in this research identifies ten NE classes or categories which are Person, Location, Organization, Job, Device, Car, Cell Phone, Currency, Date, and Time. The system consists of three main modules. The first is the 'CRF classifier', which is used for segmenting and labeling the sequential data. To that purpose, it used 15 features for the word. Actually, the RDI toolkit is used significantly in the specification of these features values. By way of example, not exhaustive listing, the use of ArabSemantic RDI tool in the identification of the word's semantic field. In fact, it acts as the synonyms relationship of the WordNet. So, this is the first distinction between [*71*] and other Arabic NER systems that based on the AWN, such as [*38*], [*54*], and [*52*]. Another powerful contribution of the RDI tools in this system is the use of its ArabMorpho-POS tagger in the extraction of the word's lexical features.

The second component is the 'Bootstrapping pattern recognizer'. It bootstraps gigantic set of Web pages to iteratively find all occurrences of the relation instances in 150 MB News corpus coupled with their pattern representation. Thus, another difference appears, which is the media used to extracting NEs from. Here it is a News corpus rather than Wikipedia in the others. The last component is the 'Matcher module' which also based on the RDI toolkit which consists of Arabic RDI-AraMorpho-POS tagger and RDI-ArabSemantic tool mentioned before.

Since NE keywords are frequently occurred within contexts. So using patterns, the identification of NE will depend on some indicators or phrases. This way will replace the most common way in the NEs identification, which is exploiting the capitalization feature of the English Language, and must be used indirectly via the translation to and from English. In addition to wasting time, that has other disadvantages that are mentioned earlier.

The proposed system runs the three modules sequentially till no new NE occurrences extracted. For each NE class, the three modules are run and fed up with the collected data set, by ANERCorp and ANERGazet, and a subset of the feature set. CRF classifier yields some NE occurrences as the best seeds for the pattern recognizer, and then the recognizer uses the matcher module to produce good patterns for boosting CRF classifier.

The research results show that using only little size of gazetteers datasets, the technique may generate good patterns and may work cooperatively with CRF. The technique also finds new NE occurrences/contexts easily without the need of extraction work. These features give preference to the use of such machine learning techniques over the other techniques.

The feature that exceeds [*7*] on [*71*] is that its ability to accommodate the increasing number of the NEs and hence classes or categories they belong to, without the restriction of predefined range.

Definitely, great benefits will be gain if bestow ontological characteristics are bestowed upon such systems to be a comprehensive semantic Named Entity knowledge base.

On the other side, regarding to the implementation of the 'rule-based' approach in the ANEE, the articles [*34*] and its improvement or supplement researches [*35*] and [*36*] have been addressed.

About [*34*], it is the 'rule-based' research that aimed to extract Just Person NEs in Arabic corpus. It developed a system named PERA in which various Arabic text corpora have been analyzed to get the best rules of Person NEs recognition. The recognition system has two components which are the 'Gazetteer name list' and 'Grammar rules'. Person names that are extracted from the available corpora and other resources are used to build up a lexicon in the form of gazetteers, then the learned patterns and person indicators used to derive a fine grammar rules in order to give high-quality recognition of Arabic person names. A further filtration mechanism is then employed for enabling revision capabilities.

The system developed to be incorporated in various applications, language independently.

[3] http://www.rdi-eg.com

Therefore, the dictionary includes the corresponding English translation of the Arabic names as a metadata.

The system results are evaluated by calculating the precision, recall, and F-measure over 46 sets. Their average in the recognition of person names was 85.5%, 89% and 87.5% respectively. Actually the size of these 46 groups was not specified, so the reliability of the results can't certainly be determined. By the authors, this work has several ongoing extending activities, such as recognition and categorization of other Arabic NEs as locations and organizations. That's indeed accomplished in the authors' subsequent researches [*35*] and [*36*] in which a system called NERA is developed. In fact, it is a respectable effort due to its addressing to a lot of the Arabic Language's challenges such as the complex orthographic system, the ambiguity, and the lack of resources in order to get satisfactory results in terms of Precision, Recall, and F-measure.

The NERA system able to recognize 10 categories of Arabic Named Entities rather than just Person names in PERA. These categories are Person name, Location, Company, Date, Time, Price, ISBN, Measurement, Phone number, and File name.

The system's results are summarized in the following table that shows the accumulative recognition accuracy achieved by each category against the reference corpora

Table 1: Accumulated accuracy of the 10 Arabic NEs [*36*]

No.	Entity type	Precision (%)	Recall (%)	F-measure (%)
1	Person	86.3	89.2	87.7
2	Location	77.4	96.8	85.9
3	Company	81.45	84.95	83.15
4	Date	91.2	92.3	91.6
5	Time	97.25	94.5	95.4
6	Price	100	99.45	98.6
7	Measurement	97.8	97.3	97.2
8	Phone no.	94.9	87.9	91.3
9	ISBN	94.8	95.8	95.3
10	File name	95.7	97.1	96.4

Moreover, the authors have more ideas for the results to be improved more and more.

With regard to [*84*], it is a research that used the n-gram and maximum entropy method as a base for directing an Arabic Named Entity system named ANERsys. To that end, the authors developed their own training set corpora ANERcorp and gazetteers ANERgazet. The main reason for building these resources is that there is no freely available other on the Web. So, they were used during the research process for the training and then they released to be used as free resources. The ANERcorp consists of 316 articles from various domains, with 150,286 tokens and 32,114 types by 4.67 ratio of tokens to types. The ANERgazet consists of three sub-gazetteers which are: Location, Person, and Organization gazetteers.

The 'maximum entropy classifier' built by means of the observation and experiments to determine a list of characteristics about the context in which Named Entities usually appear, then estimated the different weights using the General Iterative Scaling (GIS) algorithm. Finally, built a classifier which basically computes the probabilities for each word to be assigned to each of the considered classes.

The results of implementing the ANERsys on the built ANERcorp test corpus using the developed ANERgazet is improved from 62.72 % precision without ANERgazet to be 63.21% and the recall increased to be 49.04% rather than 47.58%. The F-measure also increased to 55.23 from 54.11 without the ANERgazet.

The authors of [*83*] continued the work in the research and issued the upgraded version that named ANERsys 2.0 presented in [*83*]. It differs from [*84*] in that it combined the 'maximum entropy' method used at [*84*] with the use of 'POS-tagger' in order to improve the NER process in the case of longer proper names. That increases the F-measure by 10 points over the first version of the system.

It worth mentioning that there is yet another attempt which closely resemblance to that in [*83*], in terms of its interest of the extraction of Arabic collocation terms. That attempt is presented in the article [*73*]. The presented research depended on the use of GATE, which is a software tool kit written in Java and widely used worldwide for NLP and IE especially those of Named Entities.

This time, it is adopted in the context of ANER system for extracting Arabic terms in the form of collocations by writing new 'Java Annotation Pattern Engine (Jape)' rules and using them on the 'Crescent Quranic Corpus' developed at Leeds University[4] which is tagged per word, verse, and chapter and contains binary or ternary additional information about morphology and POS. The extracted collocations in [*73*] are prepositional collocations in forms such as Noun-Noun, Adjective-Noun, Verb-Noun, Noun-Proposition-Noun, and so on. The

[4] The university that presented the articles [*2*], [*31*], and [*12*]

extraction pattern is firstly defined, and then used to write the appropriate (Jape) rule and passing it as a parameter to ANNIE transducer, which is the main component of the IE in GATE. The Arabic collocations in the corpora are detected, displayed and the annotated documents can be saved to data store with the new tags to be exploited for other tasks such as, by the authors, the automatic construction of domain Ontology.

The AnnotationDiff tool is used in the system's results evaluation. It compares the results of system annotation with other manually annotated with Noun-Adjective annotation. The AnnotationDiff calculates the F-measure as 0.66. Actually it is a very weak result, but the authors stated that they working on improving it.

The article [*39*] , which mentioned before in section 2.1, developed a tool called AraTation for semantically annotating the Arabic News Web documents. The system carried out an Information Extraction module that works as the first stage of the designed system. It is developed upon a dictionary that contains about 5000 predefined unique Arabic words belong to the news domain and fall in one of the four types: Person, Organization, Location, or Company.

The research, unfortunately, has a defect in dealing with the inflection issue in the Arabic words. As an example, the uniqueness gauge yields 5 unique words if it invoked with the words: عراقي عراقيين، العراق، عراق، عراقيون. That of course will affect the overall performance and reduces the rates of passing the precision and recall. Of fairness, the authors were aware of this problem and stated it to be a future concern.

The Arabic Named Entity Extractor (ANEE) project that presented in [*76*] is a noteworthy work although it is a commercial project done in the domain of Arabic NER. It can extract ANEs from both structured and unstructured data using semantic concepts, by means of analyzing the concept and content to deliver meaning based results. It has the ability to solve linking issues such as finding the Arab descent in US cities. To perform this task, the user just select "US cities" & "Arabic Names" then every mention of “New York City, Washington DC, Los Angeles, and Boston” found with “Mohamed, Ahmed, Usama, and Tarek” would be found Just select the category and ANEE finds every mention, comprehensively and automatically. So, it is not just NER system, it is also works as a query answer system. It also finds alternate names and aliases. For instance, it can recognize that 'Abu Ammar' (أبو عمار) is the same entity as 'Yaser Arafat' (ياسر عرفات), and that is not just great, but it also so important as a NER task. Also, it can deal with the prefixes and suffixes issues. The ANEE provides over 25 predefined categories and 100 subcategories. Moreover, ANEE's taxonomy can be customized in order to adding taxonomies or modifying entity concepts. A tool called WORDCON is used to easily Romanize the extracted lists of entities. That is a great feature that enables further search, discovery and analysis capabilities. E.g., the Arabic entity "محمد" can be translated into English as: “Mohamed”, “Mohamad”, “Mohamid” or “Mohammad”. Another feature of ANEE is the ability to adding relationships between entities. So the related entities just extracted if they appear together.

In fact, the methodology that is applied in this work is not certainly known. However, something like that may be based on Ontology. That is ensured through a telephone conversation I had with the head of the producing company. Also, the application's poster mentioned that linguistic and statistical methods are used, as well internationally recognized Ontologies.

As a whole, it seems a great application and worth to be checked.

Further detailed discussion provided in [*70*] which presents a comprehensive survey on the information extraction approaches applied on Arabic in comparison with Latin languages. It gives a detailed explanation of the Information Extraction techniques. It distinct the NER and Ontology based extraction methods, gives examples for both Latin and Arabic. Also, it lists the available and needed Arabic/English resources for doing that. Furthermore, it demonstrates the strength, weaknesses, opportunities and threads of applying NER on Arabic Language.

3.4. SW and the Islamic Knowledge

As explained earlier, Ontology is an effective conceptualism used for the SW. Therefore, constructing an Ontology for the Islamic knowledge which consists of the Holy Quran, Hadith, and explanatory books will extend the great functionalities and advantages of the SW to the domain of the Islamic Knowledge. That is very important for both Muslims and non-Muslims who looking for a trustworthy source of exhaustive and scholarly knowledge on the Holy Quran and related resources. That's because it will enable a semantic meaning in order to understand the Islamic Message as described in the Holy Quran and Hadith, since most of available information resources on Islamic Knowledge considered the natural language text processing rather than semantic manipulation and machine interoperability.

The most prominent articles that considering the representation of the Islamic Knowledge and Holy Quran semantically were accomplished around 2009 and 2010.

At 2009, [19] is directed. It is one of the rear articles that try to represent the Islamic Knowledge using the semantic technology. The research built a framework for recognizing and identifying semantic opposition terms using Natural Language Processing and domain Ontologies. At the time of publication, the work was in progress and to our best knowledge; there is no subsequent version that presents a complete system. The authors stated that the semantic opposition, which is one of the powerful rhetorical techniques of classic Arabic exhibited in the Holy Quran, can be represented by a SW technique as Ontologies. The authors' reasoning line used in the argument was logical convincing, and justified the hypothesis. The proposed framework contains two components which are the domain Ontology, limited to deal just with Women field, and SemQ tool. The system's input is a Quranic verse and the output is the list of semantically opposed words in the verse with the degree of opposition. However, it is just a prototype so there are no results to be evaluated nor evidence to support the validity or reliability of the system. Also, it has an excessively narrow subject focus as the domain of the Ontology is limited to the narrow field 'Women'.

As for the work presented in both [44] and [45] , they considered the problem of structuring a computational model that can represent the real word entities meaning transparently. That's since the author's found that there are two important language characteristics that are not addressed by the NLP community, and tried to address them in that researches. The first characteristic was the definition of word's meaning which defined through a collection effort of the language users. The second is the dynamical behavior of language's words. Some adapt their meaning, some obsoleted and others originated. Therefore, the research tries to provide a solution based on an ontological model for representing a dynamic and collaboration computational lexicon that can accommodate these two characteristics. The model implemented on the 'Time Nouns' semantic field vocabulary of the Holy Quran which is tended to be extended to other fields in the future work. The ontological structure utilizes the word features, as semantic units, in its semantic representation. So, the words 'woman' & 'girl' can be discriminated by the atomic components (features) gender, and adulthood. While the gender feature will be female for both, the adulthood will be the definition, so woman is adult and girl is a child. This way of meaning representation is implemented over the in the Holy Quran 'Time Nouns' scope. The Ontology developed using UPON (Unified Process for ONtology) ontological engineering approach, and OWL standard language representation in order to enabling the shared and open access to that resource. The Ontology is structured according to the hyponymy/hypernymy relationship between the concepts. The conceptual classification forms the main classes of the Ontology which is divided into two main categories. The first is the 'top level classes' that represent major concepts from the field theory of semantics, and they are reusable across different semantic fields and different languages. The other category is the 'lexical classes' which represent the actual words within a specific semantic domain.

The structure of the ontological lexicon implies that: the more going deeply into the hierarchy, the more argument will gain by the computational formula, and so, the meaning narrows. So, using that approach in representing the language semantics, the meaning of the word can be described by the combination of all features from the top level down until the word itself. For example, in a try to answer the question "what is the meaning of the word 'Ghadat – غداة'?" the final formula that representing the word's meaning will be:

Ghadat = Time + Specific + Tangible + Day +Light+ from dawn till Sunrise + Beginning of day. Moreover, the represented Ontology is capable of naming a new phenomenon or concept. That is achieved by viewing the ontological hierarchical classification, while the Ontology is a set of features. So the lexicographers can either choose from or add to this feature to produce the componential formula. A set of closed words are then displayed, so the lexicographer can select a target word and add a new sense to represent the new phenomena, or may coin a new word based on the meaning of closely related words.

Actually, it is a good utilization of the ontological capabilities in the representation of the word's meaning transparently which may open new prospects for the use of the Ontologies in the Arabic Language. However, the authors encountered several problems such as what is the appropriate atomization feature for each concept and to which depth level. Another problem is the difference in the meaning description of the same concepts from one person to the other.

Saidah et al., on the other hand, attempt to build a framework for Islamic Knowledge that aims to develop a method for extracting the Islamic concepts, then exploiting it in the automatically building of Islamic Knowledge Ontology. The authors published two papers in this regard, [*63*] and [*62*]. Of course, the development of Islamic Knowledge Ontology is not that trivial work; at least it needs a compilation of Ontologies in different fields of the Islamic resources. Thus, the authors took just a step towards a hard work. They have presented an approach for the automatic generation of Ontology instances from the Holy Quran which based on the combination of Natural Language Processing, Information Extraction, and Text Mining techniques. The system has been developed to generate Ontology for that Islamic knowledge only involves verses from the Quran. It also focuses on verses that involve the *solat* or prayer phrases. During the pattern extraction process, only the Quran verses containing keyword *solah* or prayer will be extracted. Following that, the related Quran verses will be verified with the contents of the surahs for elaborating the hidden meaning, since the obtained verse might be hanging in which the super-concept cannot be identified. It gives an essence to that particular verse. This will assist in generating the concept and relation. The verse then will be extracted according to the identified pattern.

The whole system still under construction and the presented results are just a sample that does not give accurate indicators of the performance. However, it is noted that the dealing with Islamic concepts and Quranic verses was in transliteration form rather than pure Arabic characters. At least, that transliteration could be paired with the corresponding Arabic. That will facilitate the use of Arabic users to the system.

In Malaysia 2010, Juhana Salim et al. in [*29*] treat the issue of multilingual Ontologies for Islamic Portal. Their research aims to identify an Islamic Semantic Retrieval system to retrieve Islamic knowledge in different languages (Malay, English, and Arabic). The research firstly developed the Islamic Ontology that then used to annotate the documents of the domain of the interest. To build that multilingual Islamic Ontology, the developed system first used Islamic Extraction system to extract the content of authoritative Web pages. To that end, the system depended on resources and thesaurus, which are the Library of Congress Subject Heading (LCSH), Library of Congress Classification (LCC), and Index Islamicus.

After words extraction and stop words discarding, the system used (LCSH) and (LCC) for expanding the extracted words. All terms of the thesaurus are connected to concepts in the Ontology and are given a hierarchical taxonomy using Border and Narrower Term relationships, and they can get further relationships and properties via the BP 77.5 class of (LCC). The 'Index Islamicus' resource is also used for more expansion to the developed domain Ontology.

The Ontology, to be multilingual, is then translated using 'Bahasa Melayu-Bahasa InggerisArab-Urdu/Hindi' and several Web tools such as 'Google Translate'. The developed Ontology is then used in the annotation process. Actually, the research doesn't discuss the results sufficiently, it just presented as a noisy figure. So, the reliability of the system can't be evaluated basing on that figure.

Other noticeable works are those presented at [*75*], [*2*], and [*72*]. With respect to [*75*], the authors aimed to exploit the important role of SW technologies in the distributed knowledge sources of the Holy Quran. That by means of flexible and efficient knowledge modeling, storage, publishing, reasoning and retrieval. The paper presents a semantic based knowledge representation model for the Holy Quran and its relevant resources in order to supplement Quranic learning and research. The presented framework enables a reasoning - capable knowledge base for Quranic and related textbooks knowledge depending on the SW tools as Ontologies and semantic reasoning. The importance of this study lies in its distinguished attempt to deal with critical resources which is the Holy Quran and books of Ahadith comprehensively. Actually, an abundant set of Web sites have designed to retrieve Islamic knowledge on the level of keywords, but very rare, if exist, those which provide a machine processable form on the level of information in this field.

The system's framework starts by the 'Data Collection' process which collects the Holy Quran, Ahadith books and scholarly texts related. They then standardized through the 'Metadata Generation' process, which parses text to extract data about data, and 'Tag Generation' which extract various data that used later in an annotation process. After that, Metadata and Tags are combined to formalize data representation in XML format. Subsequently, the 'Knowledge Modeling' process builds Ontology models using Ontology schemas for Quran and related text books. The Ontology then populated by the XML representation documents, then the populated Ontology stored in Ontology repository. A contextual association between the Holy Quran and books of Ahadith is then performed in the 'Contextual Modeling' process. Finally, the system processes the given user query in the 'Knowledge Retrieval' process and the results passed to the user.

In fact, the research faces various challenges due to the complexity and richness of the Islamic knowledge content. The structural organization and thematic complexity of the Quran is one of them. The organic unity and coherence in the structure of Quran is a moreover challenge. However, from my point of view, the most difficult challenge is the contextual linking between Quran and other related knowledge resources. In my own perspective, it is a great respectable effort. That's due to its dealing with very sensitive, critical, affluent and complicated resources. Such system may be a base for a new Muslims to learn and understand their religion. That is because the Quranic KB provides possibilities to improve the understanding of the Quran's themes, structure, and context with the association of the Ahadith and explanatory books. However, despite of these advantages, it has the shortcoming of the use of the transliteration rather than the sheer Arabic characters to represent the Arabic words. As with [*63*] and [*62*], that will definitely affect and may hamper the use of such great system by a broad sector of the native Arabs. A simple solution is to pair the originally Arabic words with their transliteration in order to maximize the benefits.

Concerning this, the Quranic Ontology made in [*2*,*31*] was great in the sense that it provides the pronunciation both in Arabic and English letters. Nonetheless it did not handle the related knowledge resources issue as done in [*75*] .

The article [*12*] concentrated on the applications presented at the Website http://corpus.quran.com/. One of these applications is the Ontology of Quranic concepts[5] . The Quranic Ontology defines the concepts in the Quran using knowledge representation, and it uses the predicate logic to show the relationships between these concepts. The Named Entities such as the name of historic people and places mentioned in the Quran, are linked to concepts in the Ontology, while [*2*] sheds light upon the other applications related to the Arabic Language and the Holy Quran and have been carried out in the Leeds University.

3.5. Arabic Semantic Search Engines

Search engine is the most important tool used in discovering information published on the Web. In fact, the Web is the biggest global unstructured database, which makes it a difficult resource to be machine understandable and processable. That means that finding the right information amongst this massive data overload is a challenging task. Moreover, the user query words sometimes are semantically ambiguous. That's because different people may use different terminologies for the same concept as Synonymous, while the same user, on the other hand, may use the same word for different concepts as Polysemous [*39*,*57*].

Therefore, most search engines face a problem to understand the meaning of query words. Accordingly, the main challenge of traditional search engines is how to accurately understand users' needs, process the relevant semantic knowledge of query information source, and automatically and individually return the accurate relevant information to each user [*28*,*50*,*81*].

In recall/precision interpretation, the traditional search engines can be described as: with high-recall, they have low precision, and with low recall, the precision becomes none. That mainly caused due to: the sensitivity of results to the keywords and the misinterpretation of the Synonymous and Polysemous words [*74*]. So, even if the main relevant pages are retrieved, there are slightly-relevant or irrelevant documents also retrieved, and that is the factor that affected the precision parameter. If those important pages are missed or there is no of those relevant have been retrieved, that is the case of low or no recall. Consequently, the traditional keyword-based search engine is not appropriate to be used anymore.

The alternative of keyword-based search engines is the SSEs that generate information more relevant to the user's needs. The SSE uses the Ontologies rather than the ordinarily used lexicons that are used in the traditional search engines in the indexing process. The SSE can, therefore, extract triples from

[5] corpus.quran.com/ontology.jsp

RDF files, that hold the Web document's metadata, to provide semantic information about the search keywords or hence concepts, answer user's query, and discover the relations between keywords and concepts in an attempt to uncover the meaning of the Web contents [*59*].

So, the use of semantic search will move the search model from the document level to that of entities and knowledge [*67*]**.**

Using Ontologies, the search engines can look for pages that refer to a precise concept in an Ontology instead of collecting all pages in which ambiguous keywords occur. In this way, differences in terminology between Web pages and the queries can be overcome [*51*,*82*]. In addition, Web searches can exploit generalization/specialization properties of Ontologies. So, if a query fails to find any relevant documents, the search engine may suggest to the user a more general query. Or if too many answers are retrieved, the search engine may suggest to the user some specializations.

By understanding the meaning of a query and its possible dimensions, it is likely that results returned to the user will be more relevant, and that resources that have been missed, will be retrieved, which means higher recall with more precision [*30*].

Swoogle[6], Hakia[7], SenseBot[8] and DeepDyve[9] are among the top SSEs. But they, unfortunately, have weak to no support of Arabic Language. Arabic is still not well supported through SSEs [*41*]. As more Arabic Web sites are increasing in the World Wide Web daily, semantic search systems that treat the semantics for Arabic Language are essential. The need for exploring relevant information quickly and accurately is the challenge for the Arabic Web sites. Little work has been done on Arabic semantic search. The following section presented some.

In [*64*], S.Zaidi et al. described a Web based multilingual tool for Arabic Information Retrieval based on a legal domain Ontology as an attempt to improve both recall and precision of the search. The tool tries to minimize the level of noise in the results of search on the legal domain using an Ontology-based query expansion. The Arabic Ontology is constructed manually by Protégé 2000 using a top-down strategy in which the hierarch concepts are related via is-a relationship. For the population, the United Nations Arabic articles are used as well as some Arabic newspapers. As mentioned earlier section, [*64*] designed an Arabic legal domain Ontology and used it as a resource for the user query expansion. The user query analyzed then expanded through navigation over the built Arabic legal Ontology. The query is translated and expanded through the PWN rather than the Arabic legal Ontology in the case that the user wishes to get the results in English or French documents. Note that the AWN was not yet been accomplished.

The relevance of returned documents calculated on the first returned ten documents. The author shows that the use of the tool improved the results as that while the recall was 115 and the precision is 2 of the first 10 before applying the tool, they are improved to be 1230 for the recall and 7 for precision of the first 10 after. The authors have respectable perspective which is the use of the cross-language retrieval is a useful way in order to share and distribute the information independently of the used language. Actually, another contradictive perspective, presented at [*56*], stated that the interlingual translation sometimes being inconsistent and has conflicts.

The aim of the research presented at [*17*] is to improve Arabic search engine's results by query expansion based on synonyms and stems. That's by implementing a model for new IR systems using components like Arabic stemmer and word synonyms structure for improving the retrieval process of Arabic search engine. The authors assumed that a search keyword expansion based on synonyms can potentially improve the system recall since the question's answers may contain the basic keywords synonyms.

The proposed system contains two service sides, namely client and server. Client side acquires the search query, removes stop words, views synonyms tree which based on Arabic synonyms database. It shows all the related synonyms located in the same keyword synonyms ring. Next, the search query will be expanded on the bases of the selected synonyms. Lastly, Arabic light stemmer is implemented on that expanded Arabic query. As for the server side, there are the indexed tokens wildcard cursor, wildcard cursor stemmer, and the indexed articles database.

Two evaluation approaches are made to evaluate results. The first based on measuring the performance of the system against related synonymous words and the total number of retrieved relevant words after selecting the synonym(s). As for the other, it focused

[6] www.swoogle.com
[7] www.hakia.com
[8] www.SenseBot.com
[9] www.deepdyve.com

on comparing the total number or retrieved sense-related words with the result of single query word results.

Although the work in the AWN was already standing at the publication time of this research, but it had not been accomplished yet. Nevertheless, this research had been discriminated with the construction of an Arabic WordNet core that is fully based on the Arabic Language, while the AWN, on the other hand, based on the English PWN, which makes it vulnerable to criticism of the heterogeneity arises from the difference in languages characteristics such as structure, semantic, and syntax.

In spite of this, the research tries to simulate the idea of WordNet using just the synonyms and indexed tokens database that could serve as lexicon vocabulary storage. It lacks to the concepts relationships that distinguish the WordNet and the Ontologies in general. And so, the research results will not amount the level which can be obtained if an Arabic Ontology is used, or even AWN, instead of just Arabic synonyms DB. This is because the existence of relationships allows exploitation of the real meaning of the word as a concept - not just an alternate word – has relationships with other concepts with levels of hierarchy in specialization and globalization.

The work at [*57*] proposes a search engine for the Arabic Language, with some treatment to the semantic level. As described before, [*17*] developed an Arabic synonyms DB, since the AWN was not published yet, while in the case of [*57*], it is published at 2007, i.e., after the producing of AWN. However, it uses a little like approach as [*17*]. It developed a terminological dictionary of Arabic words as a treatment to the semantic level of the document retrieval process.

The research focuses on the indexing step of the document retrieval process that considered as a composite of three steps: 'document indexing', 'searching from user request', and 'results presenting'. The used terminological dictionaries are developed as follows:

Each term x_i belongs to the set of all existing terms V in the index of a collection of documents can be represented as:

$$\mathrm{Dic}\ (x_i) = \{t_{i1}, t_{i2}, \ldots, t_{ik} \}, \qquad \text{Eq.(1)}$$

where t_{ij} is a term that is related or similar to the term x_i with a corresponding degree of importance. So, the word "طب" is represented as“الأمراض، الأوبئة، التشريح، أدوية”.

As it is obvious, the research restricts the semantic functionality in the set of synonyms or related words. The semantic level may be treatment more beneficially if an Ontological level is considered rather than just lexical. So, it suffers the same limitations as [*17*].

Another work, [*85*], introduces a method for ranking Arabic Web sites using Ontology concepts. In this paper, an Arabic Ontology concepts for the electronic commerce domain (التجارة الإلكترونية) in Arabic Language had been built, which then used for ranking Arabic documents. The proposed method ranks the document according to the frequency of Ontology concepts in the document, by means of the document that contains more Ontology concepts takes a higher rank. The system is implemented by Visual Basic .Net and its performance is compared on three different search engines: AltaVista, Google, and Yahoo. The system's results in ranking the first 30 documents show that the used ranking methodology is better than AltaVista 4.2 times, better than Yahoo 4.5 times, and than Google 2 times. The author uses ranking over just 30 documents to be comparable with an expert ranking, and he indicated that the number of considered documents don't affect the algorithm's performance nor results.

Another Arabic Semantic Search tool that based on an e-commerce domain Ontology is SemArab that is existing in [*46*] and [*47*]. In fact, the e-commerce Ontology that has been built is a tree-like that contains just 5 commercial concepts, each with at most 10 branched related concepts without any interrelationships between them. The presented tool, SemArab, employs the use of the Ontology in order to improve the search process in a semantic manner. The tool has an easy to use GUI that enables the user to just type his query in a form of keywords and select the related Ontological concept it belongs to. The system then searches the Web using a general search engine to get the related documents based on the combination of the user's keyword and its Ontological concept plus all its related concepts. The first 100 search results are then used to extract their content, and then they are filtered in order to looking for any occurrences for any of user's query keyword or its related concepts. The found results are kept to be ranked after the measurement of the concepts similarity. The results are ranking according to the frequency of the Ontology concepts exist in the extracted documents. The research results are summarized in a form of table that shows the results of SemArab vs. those of traditional Web search engines as follows:

Table 2: SemArab vs. Web Search Engines [46]

Parameter name	SemArab		Google		Bing	
	Relevant	Irrelevant	Relevant	Irrelevant	Relevant	Irrelevant
Ahmad as a person	68 %	32 %	35 %	65 %	32 %	68 %
HP laptop as a sales	90 %	10 %	81 %	19 %	80 %	20 %
AlOthaim as an organization	74 %	26%	33 %	67 %	33 %	67 %
1999 SR as a various	65%	35%	37 %	63 %	29 %	71 %

Actually, the system uses the Ontology just to get the function of the word's synonyms rather than to exploit the Ontology as its full powerful. Also, the used methodology, even if it improves the precision factor, it will not affect the recall, since it can't find those absent related URIs which are not fetched ordinarily in the traditional search engines.

Ibrahim Fathy Moawad et al., in [21], show an Arabic SSE that is based on an Arabic Ontology and coupled with an existing Arabic syntactic search engine. It is semantically reason using the Arabic Ontology that provides a semantically understanding for the user's query and so improves the search results.

As the construction of an Ontology of the whole Arabic Language is a very exhausting work, and need to be a standalone project, as that presented in [56], the authors considered just the domain of 'Computer Technology' and use its limited vocabulary in constructing the system's Ontology. Besides building that Arabic domain specific Ontology, the system architecture achieved through the conducting of three more modules which are an 'Interactive Semantic Query Analyzer', 'Semantic Ranker', and the 'Interface with Syntactical Search Engine'.

As for the 'Interactive Semantic Query Analyzer', it recommends the end user with extra semantic search criteria by accessing the computer domain Arabic Ontology. It then takes the user query as an input and extending it semantically through the match with the Ontology concepts, and then uses the extracted query by means of the Google search engine to find the related documents. The search results are then ranked semantically in the 'Semantic Ranker' module before presented to the user. That's using the 'concept frequency' technique rather than the 'term frequency' used by the syntactic search engine. So, possible words of the same concept have to be generated using the Ontology and the ranking process then takes place. The 'Interface with Syntactical Search Engine' module is used for coupling the 'Semantic Ranker' and 'Semantic Query Analyzer' modules with the Google search engine.

The system is evaluated with two experiments based on the computer domain. The results of this SSE compared with those of Google syntactic search engine. The number of pages using semantic query were far less than that of Google. The authors just state that 'this result helps in a great deal with more accurate search results', without any interpretation into measurements such precision or recall.

Finally, the research [59] describes a method for extracting semantic RDF triples from Arabic Languages Web pages. According to this article, the development of the SW technology is to adopt an intelligent algorithm that utilize the four technologies NLP, NN, Multi Agent Systems, and Fuzzy Logic Controllers as a related technologies. This research presented a SSE model and referred to it as (543) semantic model, as an abbreviation to the use of 3 ideas with 4 technologies and 5 meaning theories. The article proposed a detailed overview for some concepts like RDF triples, OWL, OWL/RDF and Arabic Language characteristics and challenges. However, it did not present any framework nor prototype to implement the suggested (543) semantic model.

A brief outline for more articles related to Arabic SSEs and more can be found in a survey presented at [67]. It concludes that very few Arabic semantic search systems have been developed, and that the emergence of one may dominate the market.

4. Conclusion

Our survey considered most of the applications, to the best of our knowledge, which are concerned with the activation of the use of Arabic Web documents in a semantic manner rather than just syntactically. It can be deduced that the Arabic Language is still far away from the environment of the Web's third generation, or the SW in terms of both quantity and quality.

Regarding to Ontologies, which are the cornerstone of the SW implementation, they are very rare in Arabic and those that exist are limited to close domains which can't be used in wider ranges. Moreover, the only Upper Level Arabic Ontology research is still under construction.

Although various researches tried to overcome that shortage by the use of the AWN instead of the Arabic Ontology, it has been experimentally proven that this is not appropriate in many cases for several reasons. It is, for example, not as efficient as its English counterpart in terms of the count of the available items. Moreover, the dependence on the translation approach in its creation is not the optimal solution due to the inconsistency of some concepts across languages because of the different cultures of their users.

Accordingly, it is necessary to develop and enrich the AWN to catch up with its English counterpart PWN in order to achieve its maximum benefit when used in those applications that require only the WordNet capabilities. As to those wider applications that need a more comprehensive representation of the concepts meaning and interrelationships, a more powerful and richer resources for both Upper-Level and Domain Ontologies are certainly needed.

To get an entire overview, it was necessary to review the applications that concerned the ANEE, since the Web explosion motivates the need to recycle the included knowledge via the development of applications that can catch knowledge from texts and give it back to the user in the form that achieves his requirements. That can be reached only by extracting the information held in Web documents and accessing its meaning by annotating it to the predefined Ontologies. The majority of these information units are in the form of Named Entities. So, they can't be extracted without efficient NEE tools. In the case of Arabic Language, these tools are limited since they are mainly founded or grounded on morphological and grammatical tools and other utilities and resources as Arabic Corpora and Gazetteers, and each has its hindrances. The morphological tools are scarce and those freely available are mostly insufficient. The other resources as Corpora and Gazetteers weren't better off, and there scarcity and inefficiency exacerbated the problem.

The Arabic Ontologies, AWN, and ANEE can all be considered as the infrastructure of larger applications such as those of Semantic Islamic Knowledge Representation and Retrieval, or those for Arabic SSEs in general. The performance of these applications is significantly affected by the complications suffered by its infrastructure.

Definitely, the essential reason for all of these difficulties is the nature of the Arabic Language itself which often leads to the inability of dealing with it except for human mind. Nevertheless, this should not prevent further effective efforts to try to reach the maximum possible solutions that enable the Arabic Language and its users to take advantage of the new electronic technologies generally and the SW in particular. This will not be achieved unless the Arabic electronic infrastructure is built. That infrastructure includes the basic resources such as the Arabic Ontologies, Arabic Gazetteers, AWN, and so on. It also contains other accompaniment utilities such as the Morphological analyzers, NEE. The Arabic support SW tools as OWL/RDF editors, reasoners and so on must also be considered in it. If that infrastructure is actualized and is sufficiently powerful, the development of wider applications as ASSEs will be more facilitated and its results will be more reliable and trustworthy.

Literature Cited.

[1] A. Crapo, X. Wang, J. Lizzi, and R. Larson, "The Semantically Enabled Smart Grid," November 17-19, 2009.

[2] Abdul-Baquee Sharaf, Eric Atwell, Kais Dukes, Majdi Sawalha, Amal Al-Saif, Serge Sharoff , Katja Markert, Latifa Al-Sulaiti, Bayan Abu Shawar, Nora Abbas and Andy Roberts, " المشاريع الحاسوبية على اللغة العربية والقرآن بجامعة ليدز "Arabic and Quranic Computational Linguistics Projects at the University of Leeds"," 16-19 October 2010.

[3] Aidan Hogan, Andreas Harth, Juergen Umrich, Sheila Kinsella, Axel Polleres, Stefan Decker, "Searching and Browsing Linked Data with SWSE: the Semantic Web Search Engine," vol. 9, No. 4, 2012.

[4] Aldo Gangemi, Sean Bechhofer, Asunción Gómez-Pérez, and Jim Hendler , "Introduction to the Introduction to the Semantic Web," , Karlsruhe, Germany, Oct 2008.

[5] Barry Norton, "Semantic Technologies: Origins, Linked Data and Beyond," , Accra, Ghana, March, 2011.

[6] Bassel AlKhatib, Mouhamad Kawas, Wajdi Bshara, and Mhd. Talal Kallas, "Ontology-Based Semantic Context Framework (OBSC) Framework for Arabic Web Contents," , Dubai, April 9-10, 2008.

[7] Behrang Mohit, Nathan Schneider, Rishav Bhowmick, Kemal Oflazer, and Noah A. Smith, "Recall-oriented learning for named entity recognition in Wikipedia," Pittsburgh, Pennsylvania, August 2011.

[8] Benjamin Heitmann, Conor Hayes, and Eyal Oren, "Towards a reference architecture for Semantic

Web applications," , 2009.

[9] Christiane Fellbaum, Musa Alkhalifa, William J. Black, Sabri Elkateb, Adam Pease, Horacio Rodríguez, and Piek Vossen, "Building a WordNet for Arabic," , May, 2006.

[10] Dean Allemang & James Hendler, *Semantic Web for the Working Ontologist, 2nd Edition: Effective Modeling in RDFS and OWL.*: Morgan Kaufmann; 1 edition, March 1, 2011.

[11] Didouh Omar, "الأنطولوجيا العربية و الويب الدلالي،" ، جاكرتا، يوليو ٢٠١٠.

[12] Eric Atwell, Claire Brierley, Kais Dukes, Majdi Sawalha, and Abdul-Baquee Sharaf, "A An Artificial Intelligence Approach to Arabic and Islamic Content on the Internet," 2011.

[13] F. Belkridem, and A. El Sebai, "An Ontology Based Formalism for the Arabic Language Using Verbs and Derivatives," , 2009.

[14] Fatma Zohra Belkredim and Farid Meziane, "DEAR-ONTO: a derivational Arabic ontology based on verbs," vol. 21, No.3, 2008.

[15] G. Madhu , A. Govardhan, and T.V. Rajinikanth, "Intelligent Semantic Web Search Engines: A Brief Survey," vol. 2, No.1, January 2011.

[16] Grigoris Antoniou and Frank van Harmelen, *A Semantic Web Primer*. London, England: The MIT Press; second edition edition, March 21, 2008.

[17] Hayder K. Al Ameed, Shaikha O. Al Ketbi, Amna A. Al Kaabi, Khadija S. Al Shebli,Naila F. Al Shamsi, Noura H. Al Nuaimi, Shaikha S. Al Muhairi, "Arabic Search Engines Improvement: A New Approach using Search Key Expansion Derived from Arabic Synonyms Structure," , 2006.

[18] Hend S. Al-Khalifa and Areej S. Al-Wabil, "The Arabic Language and the Semantic Web: Challenges and Opportunities. ," November 2007.

[19] Hend S. Al-Khalifa, Maha M. Al-Yahya, Alia Bahanshal, and Iman Al-Odah, "SemQ: A proposed framework for representing semantic opposition in the Holy Quran using Semantic Web technologies," , 01 March 2010.

[20] Ian Horrocks, "Ontologies and the semantic web," vol. 51 Issue 12, December 2008.

[21] Ibrahim Fathy Moawad, Mohammad Abdeen, and Mostafa Mahmoud Aref, "Ontology-based Architecture for an Arabic Semantic Search Engine," December 15-16, 2010.

[22] J Davies, Rudi Studer, and Paul Warren, *Semantic Web technologies : trends and research in ontology-based systems.*: Hoboken, N.J. : John Wiley & Sons, 1 edition, July 10, 2006.

[23] Jiaqiang Dong, Yajun Du, and Mingli Feng, "A Novel Strategy for Constructing User Ontology," vol. 5 No.5, May 2011.

[24] Jim Hendler, ""Why the Semantic Web will never work" (note the quote marks)," , Heraklion, Greece, June, 2011.

[25] John Davies, Dieter Fensel , and Frank van Harmelen, *Towards the Semantic Web: Ontology-driven Knowledge Management.*: Wiley; 1 edition, January 21, 2003.

[26] Jörg Wurzer, "Application Challenges that may be amenable to Semantic technology solutions," , Riga, Latvia, July, 2011.

[27] Jorge Cardoso, "On the Move to Semantic Web Services," , 2005.

[28] Jorge Cardoso, *Semantic Web services: theory, tools, and applications.*: IGI Global, Mar 30, 2007.

[29] Juhana Salim, Siti Farhana Mohamad Hashim, and Akmal Aris , "A framework for building multilingual ontologies for Islamic portal," vol. 3, pp. 1302 - 1307 , 15-17 June 2010.

[30] Junaidah Mohamed Kassim and Mahathir Rahmany, "Introduction to Semantic Search Engine," , Selangor, 2009.

[31] Kais Dukes, Eric Atwell and Nizar Habash, "Supervised Collaboration for Syntactic Annotation of Quranic Arabic," November 2011.

[32] Karim Bouzoubaa. (17 Oct 2010) ArabicWordnet Use and Enrichment. pdf Document.

[33] Karin Breitman, Marco Antonio Casanova, and Walt Truszkowski, *Semantic Web: Concepts, Technologies and Applications.*: Springer London Ltd, 28 October 2010.

[34] Khaled Shaalan and Hafsa Raza, "Person Name Entity Recognition for Arabic," June, 2007.

[35] Khaled Shaalan, and Hafsa Raza, "Arabic Named Entity Recognition from Diverse Text Types," , Berlin, 2008.

[36] Khaled Shaalan, and Hafsa Raza, "NERA: Named Entity Recognition for Arabic," , NJ, USA, July, 2009.

[37] Lahsen Abouenour, Karim Bouzoubaa, and Paolo Rosso, "On the Extension of Arabic Wordnet Named Entities and Its Impact on Question / Answering," , Valencia, Spain, October 2010.

[38] Lahsen Abouenour, Karim Bouzoubaa, and Paolo Rosso, "Using the Yago ontology as a resource for the enrichment of Named Entities in Arabic WordNet," May 2010.

[39] Layan M. Bin Saleh and Hend S. Al-Khalifa, "AraTation: An Arabic Semantic Annotation Tool," 2009.

[40] Li Ding, Deborah L. McGuinness, Tim Finin and Anupam Joshi, "Semantic Web Technologies: A Tutorial," July 18, 2006.

[41] Lilac Al-Safadi, Mai Al-Badrani, and Meshael Al-Junidey, , vol. 19 No. 4, April 2011.

[42] Lyndon Nixon, Elena Paslaru, Michal Zaremba, Enrica Dente, Ruben Lara, Walter Binder, Ion Constantinescu, Radu Jurga, Vincent Schickel-Zuber, Vlad Tanasescu, Mark Carman, Loris Penserini, Marco Pistore, "State of the Art of Current Semantic Web Services Initiatives," , July 2004.

[43] M. Hildebrand, "Search-based user interaction on the semantic web, a survey of existing systems. Technical report," The Netherlands, 2007.

[44] Maha Al-Yahya, Hend S. Al-Khalifa, Alia

Bahanshal, Iman Al-Oud and Nawal Al-Helwa, "An Ontological Model for Representing Computational Lexicons: A Componential Based Approach," , Beijing, China, Aug.21-23, 2010.

[45] Maha Al-Yahya, Hend S. Al-Khalifa, Alia Bahanshal, Iman Al-Oud and Nawal Al-Helwa, "An Ontological Model for Representing Semantic Lexicons: An Application on Time Nouns in the Holy Quran," vol. 35, No. 2C, 2010.

[46] Majdi Beseiso, Abdul Rahim Ahmad , and Jamilin Jais, "Semantic Arabic Search Tool," , Kuching, Sarawak, July 2010.

[47] Majdi Beseiso, Abdul Rahim Ahmad , Jamilin Jais, "A New Architecture for Semantic Arabic Search Tool," , Bangalore, India, July 2010.

[48] Majdi Beseiso, Abdul Rahim Ahmad, and Roslan Ismail, "A Survey of Arabic Language Support in Semantic Web," vol. 9– No.1, November 2010.

[49] Majdi Beseiso, Abdul Rahim Ahmad, and Roslan Ismail, "An Arabic language framework for semantic web ," in *2011 International Conference on Semantic Technology and Information Retrieval*, Putrajaya, Malaysia, 28-29 June 2011.

[50] Martin Hepp, Pieter De Leenheer, and Aldo de Moor, *Ontology management: semantic web, semantic web services, and business applications*. New York ; [London]: Springer, 2008.

[51] Meena Unni , K. Baskaran, "OVERVIEW OF APPROACHES TO SEMANTIC WEB SEARCH," vol. 2, No. 2, pp. 345-349, July-December 2011.

[52] Mohammed Attia, Antonio Toral, Lamia Tounsi, Monica Monachini and Josef van Genabith, "An automatically built Named Entity lexicon for Arabic ," 2010.

[53] Mohit Behrang, Kemal Oflazer, and Noah Smith, "Named entity recognition from Arabic Wikipedia," 2010.

[54] Musa Alkhalifa and Horacio Rodríguez, "Automatically Extending Named Entities coverage of Arabic WordNet using Wikipedia ," vol. 3, No.3, June, 2010.

[55] Musa Alkhalifa and Horacio Rodríguez, "Automatically Extending NE coverage of Arabic WordNet using Wikipedia," , Rabat, Morocco, May 4-5, 2009.

[56] Mustafa Jarrar, "Building a Formal Arabic Ontology (Invited Paper).," , Alecso, Arab League. Tunis, April 26-28, 2011.

[57] Naima Tazit, El Houssine Bouyakhf, Souad Sabri, Abdellah Yousfi, Karim Bouzouba, "Semantic internet search engine with focus on Arabic language," , 2007.

[58] Natasha F. Noy, "What do we need for ontology integration on the Semantic Web (Position Statement)," vol. 25 No.1, October 20, 2003.

[59] Omar Isbaitan, Huda Al-Wahidi, "Arabic model for semantic web 3.0," 2011.

[60] Raúl G. Castro, Asunción G. Pérez and Muñoz-García Óscar, "The Semantic Web Framework: A Component-Based Framework for the Development of Semantic Web Applications ," Turin , 12 September 2008.

[61] Rudi Studer, "Ontologies and Linked Data," , Riga, Lativa, 2011.

[62] S. Saad, N. Salim, and H. Zaina, "Islamic Knowledge Ontology Creation," , London, 2009.

[63] S. Saad, N. Salim, Z. Ismail and H. Zainal, "A framework for Islamic knowledge via ontology representation," , Shahalam, Malaysia, 2010.

[64] S. Zaidi, M.T. Laskri, and K. Bechkoum, "A Cross-language Information Retrieval Based on an Arabic Ontology in the Legal Domain," , Morocco, 2005.

[65] Sabri Elkateb , William Black , Horacio Rodríguez , Musa Alkhalifa , Piek Vossen , Adam Pease , and Christiane Fellbaum , "Introducing the Arabic WordNet Project," , South Jeju Island, Korea, January 22-26, 2006.

[66] Sabri Elkateb, William Black, Piek Vossen, David Farwell, Adam Pease, & Christiane Fellbaum, "Arabic WordNet and the challenges of Arabic. The Challenge of Arabic for NLP/MT," , London, 23 October 2006.

[67] Samhaa R. El-Beltagy, "Technology : Semantic Search," Feb 2010.

[68] Samhaa R. El-Beltagy, Maryam Hazman, and Ahmed Rafea, "Ontology Based Annotation of Text Segments," , Seoul, Korea, March 11-15, 2007.

[69] Samhaa R. El-Beltagy, Maryam Hazman, and Ahmed Rafea, "Ontology learning from domain specific web documents," vol. 4, No. 1/2, May 2009.

[70] Samir AbdelRahman, Maryam Hazman, Marwa Magdy, and Aly Fahmy, "Information Extraction,".

[71] Samir AbdelRahman, Mohamed Elarnaoty, Marwa Magdy and Aly Fahmy, "Integrated Machine Learning Techniques for Arabic Named Entity Recognition," vol. 7, Issue 4, No 3, July 2010.

[72] Soraya Zaidi and M. T. Laskr, "Review of Arabic Textual Terminology Tools for Ontologies Building," , Sousse, Tunisia, 25–28 November 2009.

[73] Soraya Zaidi, M. Laskri, and A. Abdelali, "Arabic collocations extraction using Gate," , Algiers , 3-5 Oct. 2010.

[74] Soumyarashmi Panigrahi and Sitanath Biswas, "Next Generation Semantic Web and Its Application," vol. 8, Issue 2, March 2011.

[75] Sumayya Baqai, Amna Basharat, Hira Khalid, Amna Hassan, and Shehneela Zafar, "Leveraging semantic web technologies for standardized knowledge modeling and retrieval from the Holy Qur'an and religious texts," , 2009.

[76] Taghride Anbar. (2011, Dec.) ANEE Product : COLTEC Computer Language Technology. [Online]. http://www.coltec.net/default.aspx?tabid=221

[77] Taha Zerrouki. (2012, Jan.) Arabic thesaurus project. [Online]. http://groups.google.com/group/ayaspell-dic/msg/4bf02344837b16af

[78] Thomas B. Passin, *Explorer's Guide to the Semantic Web*.: Manning Publications, March 1, 2004.

[79] Tim Berners-Lee. (2010, Mar.) from the Semantic Web to the Web of Data. [Online]. http://www.slideshare.net/dpalmisano/from-the-semantic-web-to-the-web-of-data-ten-years-of-linking-up

[80] Tim Berners-Lee, "W3 future directions, Keynote speech," in *First International Conference on the World*, Geneva, May,1994.

[81] Vipul Kashyap, Christoph Bussler, and Matthew Moran, *The Semantic Web: Semantics for Data and Services on the Web (Data-Centric Systems and Applications)*.: Springer , 15 Aug 2008.

[82] Walter Renteria-Agualimpia, Francisco J. López-Pellicer,Pedro R. Muro-Medrano, Javier Nogueras-Iso, and F.Javier Zarazaga-Soria1, "Exploring the Advances in Semantic Search," 2010.

[83] Yassine Benajiba, and Paolo Rosso, "ANERsys 2.0 : Conquering the NER task for the Arabic language by combining the Maximum Entropy with POS-tag information," , Pune, India, December 17-19, 2007, pp. 1814-1823.

[84] Yassine Benajiba, Paolo Rosso, and José Benedíruiz, "ANERsys: An Arabic Named Entity Recognition. System Based on Maximum Entropy," , 2007, pp. 143-153.

[85] Zakaryia Qawaqneh, Eyas El-Qawasmeh, and Ahmad Kayed, "New Method for Ranking Arabic Web Sites Using Ontology Concepts," , Oct. 2007.