Abstract
The internet is a global mesh of interconnected web links and became essential for daily jobs. The management of data infrastructure for current web page is ill-suited for the huge developments in information technology and needs to be readily accessible. Semantic web is a framework of advanced current web based on an ontology that allows data to be meaningfully shared and used by humans and machines. Arabic language is one of the official languages of the United Nations which is spoken by more than 1.7 billion Muslims and in addition to Christians in 22 countries of the Arab world. Nahj Al-Balagha is the most important book after Holy Quran because it collects the speeches of Imam Ali Ibn Abi-Talib and consists of 241 sermons, 79 letters, and 489 wisdoms as guides in different topics like politics, science, ethics, history, social philosophy and others. Nahj Al-Balagha is written in the classical Arabic language which was used before 1400 years and for this reason Nahj Al-Balagha not fully understood by most of Muslims because of the different dialects.

This paper presents Nahj Al-Balagha Semantic Search Engine (NSSE) based on ontology and consists of two stages. The first stage is creating the Arabic lexical of the words in Nahj Al-Balagha by extraction the words and groups into set of synonyms and opposites. The second stage is creating ontology based on the first stage. The system was evaluated by using the average precision and recall of the experiments' results.

Keywords: Semantic web, Ontology, Nahj Al-Balagha, Arabic Ontology, Semantic Search Engine.
1. Introduction

Internet is an enormous and complex network of interconnected hyperlinks to allow users to browse web pages and share their data from database warehouses. Unlimited communication, social media, e-learning, online services, entertainment, easy sharing and others led to exponential growing in internet usage and may reach to 3.6 billion users by 2017 [1]. Current web contains billions of documents and there are many administrative problems and limitations related with links structures’ which led to information chaos. Even if the results have been found, Who can we check the accuracy of the results? The main problem in the current web is the query results specially when more than one document talking about the same thing. For example, the results of the word “Ajax” are: Ajax (a group of web development techniques), Ajax (sport club), Ajax (cleaner) and Ajax (dental chair). The current web is web of documents and design for direct human understanding. Those outcomes which mention the “Ajax” by names not by meaning which are ambiguous for machines’ understanding. The next generation of internet is the internet of things which impose a challenge how to restruct links in the web to be understandable by human and machine. The poor data interconnection in the current web caused lack of interaction between data and applications without human intervention. To solve these problems, we must restruct the links to the meaningful content of web pages. Tim Berners-Lee mentioned that “The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation”[2]. Semantic web is a mesh of web of data and data of data to be represented in structured manner.

Resource Description Framework (RDF) is an entity–relationship framework based upon entity(Resources)–predicate–value (object) model within the interaction between objects to create a description about resources " to link different applications and Web resources into a new global network" [3]. RDF is a syntax structure written in XML.

Ontology is an intelligent description for taxonomies to resemble data hierarchies in description logics based on the concept of object-oriented programming. The description of data and relationships between data provides knowledge representation for processing information in machines with efficiently interpretation [4]. “An ontology is an explicit specification of a conceptualization” [5].

Arabic language is the native language for the Arabic people in (22 countries) and is a religious language of all Muslims (more than 1.7 billion Muslims) in about 60 countries. It is a Semitic which consists of 28 alphabet letters and the alignment of writing from right-to-left. Arabic language has some unique characteristics such as:

- Highly inflectional because of the grammar is complex.
- There is a difference between masculine and feminine.
- There is no capitalization in Arabic.
- There are differences among singular, dual and plural.
- Richness of vocabulary.
- Short vowels in Arabic language is very important because the meaning of a word may change depending on them, for example: ٌ means (good), ٌ means (wheat) and ٌ means (land).

Nahj Al-Balagha is the most important book after Holy Quran because it’s the collection of speeches of Imam Ali Ibn Abi-Talib and consists of 241 sermons, 79 letters, and 489 wisdom as a guide in different topics from the ancient time of Adam to Imam Al-Mahdi. These topics deal with the most important issues for humans and give constitution in them like politics, science, ethics, social philosophy, history and others. These speeches were collected by Sharif Razi. Nahj Al-Balagha is the wealth of the Arab words in the singular and collected, in metaphors and idioms in elegant form. Nahj Al-Balagha is below of the Creator’s words and above of the creatures’ words as George Jordac said in the "Masterpieces of Nahj al-Balagha".

Section 1 of this paper introduces the Semantic Web. Section 2 gives an overview about some related works. Section 3 introduce NSSE in detail whereas section 4 evaluates the proposed system.
2. Related work

World Wide Web is a huge mesh of data links warehouse and this network is growing rapidly. Because it is hard for machines to integrate the information, the retrieval of related data is one of the challenges that face world wide web. The results of research in the current web depends on textual keywords and the search algorithms. This is not an efficient solution because the user may get an inaccurate information or may be ambiguous bout the meaning. For this reason, there are many researches and studies were carried out in order to get a more intelligent structure to reduce chaotic results. Semantic search integrates technologies to evolves to next generation of search engines built on store meaningful information to solve complex queries. C. Mangold created a comparison between ten of semantic search approaches based on architecture, coupling between documents and ontologies, combination between interactive and capabilities of the system to be invisible to the user, capabilities learning system, query recall and precision and structure and technological of ontology [6]. W. Wei, P. M. Barnaghi and others studied the aims of semantic search through number of systems and divided them according to their methodologies and goals into document-oriented search, attribute values and relations of knowledge-oriented search, multimedia information, relation-centered semantic and mining-based search[7]. Majdi Beseiso and Roslan Ismail evaluated the tools that supported Arabic language in a survey of Arabic language support in semantic web. The evaluation based on the layers (RDF, OWL and Query) that tools (protégé, Jena, Sesame and KAON2) supported them and there is a real need for new tools in supporting NLP for Arabic [8]. M. Al-Yahya and others presented a proposal computational model for representing Arabic language lexicons. The model used time vocabulary in the Holy Quran to relate Arabic language vocabulary. The ontology consists of 18 classes and despite the limitations, the model is capable of representing word semantics and gives an approach to facilitate semantic analysis [9]. In his paper [10] R. Sujatha introduced a survey of Semantic Search Engine, and showed that semantic search is more timely than traditional search engine. G. Pandey gave a brief overview comparison factors of semantic web with the current web contents, conceptual perception, scope, environment and resource utilization. Semantic web in constantly evolving and this imposes challenges and problems like ontology development, defining a formal semantic, proof and trust, availability of content, scalability, multilingualism, visualization and stability of semantic web languages[11]. A. Azman Ta’a and others tried to facilitate Quran knowledge because Quran knowledge requires special knowledge base and there are many ways to present the Al-Quran knowledge. The thematic approach is one of these approaches which aims to help users to understand the Al-Quran knowledge in a systematic way[12]. A. Malve and P. P. M. Chawan presented a comparative study of keyword and semantic based search engine and concluded that semantic search engines have more accurate results over keyword search through the meaning of the query [13].

3. The Nahj Al-Balagha Semantic Search Engine System

This system aims to create a knowledge base that can be used in future to provide efficient search to reach accurate information from the Nahj al Balagha. The NSSE is a prototype for the semantic search engine in the Nahj Al-Balagha not only for daily search but, researchers and scientists can rely upon to prepare books and lectures.

The system is a Nahj Al-Balagha Semantic Search Engine (NSSE) which is based on Nahj al Balagha ontology, uses lexicon of Arabic words in Nahj al Balagha for the Arabic language. The core function of this system is to create knowledge based on the relation between words, senses, connotations and their locations. For example, the word heart "القلب" means a member of the human body, the connotation of heart in the Arabic language sometimes referred to member of the human body and other to reasoning and comprehension. This system consists of two main stages shown in figure (1):
3.1 First stage: lexical ontology

Lexical ontology is used for automatically text analysis. It consists of Arabic words in Nahj Al-Balagha, their synonyms and opposites. The lexical ontology architecture consists of: The Words Extraction, the relations building and the ontology building.

3.1.1 The Words’ Extraction

The first step in the system is the words extraction which is an infrastructure to start the ontology. The essential function in the words extraction is to cut the words in Nahj Al-Balagha, storing them in Excel spreadsheet (approximately 69389 words) and remove the duplicate words (approximately 1140 duplicated words). The Excel sheet will consist of unique words and relations in tabular form.

3.1.2 The Relations Building

This step is done manually by using Arabic/Arabic dictionaries to find words, synonyms, opposites and relations between them. The relations between words are based on means of semantic relations.

1. Means property: If A and B has the same meaning: $A \rightarrow B$
   For example: (خير) (good) has the same meaning of (بَر) and means good

2. Anti-property: if A is an (inverse) of B: $A \rightarrow !B$
   For example: (خير) (good) is inverse of (شر) (bad)

3.1.3 The Lexical Ontology Building

This process reads the words and relations by using an open source java API which is called JExcelApi. The purpose of this step is to extract data from Excel sheet and create ontology. For example

```xml
<rdf:Description rdf:about="http://www.NSSE.com#قلب”>
  <a:anti rdf:resource="http://www.NSSE.com#حَمَال”/>
  <a:anti rdf:resource="http://www.NSSE.com#جسد”/>
  <a:anti rdf:resource="http://www.NSSE.com#شكل”/>
  <a:anti rdf:resource="http://www.NSSE.com#هيئة”/>
  <a:anti rdf:resource="http://www.NSSE.com#طرف”/>
  <a:means rdf:resource="http://www.NSSE.com#إفادة”>
  <a:means rdf:resource="http://www.NSSE.com#مهجة”>
  <a:means rdf:resource="http://www.NSSE.com#كَبَد”>
  <a:means rdf:resource="http://www.NSSE.com#لَب”>
  <a:means rdf:resource="http://www.NSSE.com#عقل”>
  <a:means rdf:resource="http://www.NSSE.com#وعدان”>
  <a:means rdf:resource="http://www.NSSE.com#روح”>
  <a:means rdf:resource="http://www.NSSE.com#عاطفة”>
</rdf:Description>
```
Figure 1: NSSE architecture.
The expert person can use the ontology above but GUI helps user to use the lexical as shown in figure (2).

![Lexical GUI](image)

**Figure 2:** Lexical GUI.

### 3.2 Second Stage: NSSE building

This stage uses the lexical ontology for the Arabic language in the first stage and based on Nahj Al-Balagha ontology.

#### 3.2.1 The Ontology Building Process

Ontology building describes classes and the relation between them. The knowledge management system and ontology building are manually done by using protégé depending on examination of the sources of Nahj Al-Balagha such Nahj Al-Balagha, and explained Nahj Al-Balagha.

The outcomes from this system is the Nahj El-Balagha ontology which consists of Imam Ali's concepts in 20 hierarchal classes form as a prototype ontology and it can be extended to contain all Imam Ali's speech. The main classes of the ontology are Religion concepts (ﺎﺻٕل انذٍٚ), Ethics (الأخلاق), Promotion of Virtue and Prevention of Vice (الامر بالمعروف والنهي عن المنكر), Belief (الاسلاو الاعب الابن), Family (الاسشح), History (التاريخ), Piety (التقوى), Repentance (انزٕثخ), Economy (الاقزصبد), Jihad and war (انغٓبص انؾشة), Governor (انؾبكى), Pray (انذػبء), Worship (انؼجبداد), Reason and science (انؼقم انؼهى), Quran and Sunnah (انقشاٌ انسُخ), Judiciary (القضاء), Hypocrisy (انُفبق), Asceticism (انضْذ), Counsel and advice (انًٕػظخ انُصٛؾخ) as shown in figure (3).

This system includes super classes and subclasses, equivalent classes (because the system is a set of relations and shared properties between classes), composite classes and disjoint classes (to prevent repetition in the individuals).
Figure 3: Nahj Al-Balagha Ontology Main classes
3.2.2 Instances of the Ontology Classes

The outcomes of this step are a RDF which contains subject(entity), predicate and object (value). The subject is a Nahj Al-Balagha ontology individuals, the predicate is the relation called IS_IN (ٕعذ فٙ) and the object is the text’s number (sermons, letters and wisdoms), for example:

```
<انًٛشاس rdf:about="http://www.NSSE.com/# انفشائض
_فٙ_انٕسصخ_ثٍٛ_فقسًٓب_انًسهًٍٛ_أيٕال
>فٕٙٚعذ rdf:datatype="http://www.w3.org/2001/XMLSchema#string">

3.2.3 Search & Query

This step sim to receive the query requested by user and returns the results of the search operations by using an open source java API which is called Apache Jena.

3.2.4 Graphical User Interface (GUI)

GUI is used to create interaction between users and the system with better accessibility and ease of use. The main interface of GUI is shown in figure (4)

![GUI Main Page](image)

Figure 4: GUI Main Page.
3. The experimental results

This paper took Nahj Al-Balagha as a case study and applied the concepts on it. The experiment is executed on 241 sermons, 79 letters, and 489 wisdoms. The NSSE GUI of the experiment is shown in figure (5).

The average recall and precision of the proposed system are 93% and 88.5% respectively for 10 executed queries shown in the Table (1)

<table>
<thead>
<tr>
<th>Object</th>
<th>Word Search</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>الأيمان</td>
<td>الإسلام</td>
<td>0.8</td>
<td>0.7</td>
</tr>
<tr>
<td>القرآن</td>
<td>شفاء</td>
<td>0.85</td>
<td>0.75</td>
</tr>
<tr>
<td>القرآن</td>
<td>كتاب</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>العمل الألهي</td>
<td>قسط</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>الإمام</td>
<td>يعسوب</td>
<td>0.95</td>
<td>0.9</td>
</tr>
<tr>
<td>الصلاة</td>
<td>ذي لب</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>الدار الأخرى</td>
<td>صافات الفائد</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>صفا الدين</td>
<td>زينة الدنيا</td>
<td>0.8</td>
<td>0.9</td>
</tr>
</tbody>
</table>
5. Criticism and Comparison

This system depends on the first part (lexical ontology) which consists of words, synonyms and opposites. When the lexical ontology contains a variety of words and their relationships, the system will be better and stronger. Arabic wordnet (AWN) is the largest Arabic lexicon which consists of words and synonyms. AWN has a weakness such

1. There are errors in some meanings
   Example: the results of the search "غاب" which are refer to "jungle".
2. Associated characters with words were ignored in AWN
   Example: there are no results when search about "Like his eyes"
3. There are limited capacity in words
   Example: No results found about "زبرع" which means "ornament"

The above weakness points were noted and find the solutions in the proposed system.

6. Conclusions and Recommendations for future work

This paper proposed Nahj Al-Balagha Semantic Search Engine (NSSE) which is built on Imam Ali's speeches. NSSE consists of two stages: the first stage groups Nahj Al-Balaghas' words, their synonyms and opposites. The second stage creates the ontology based on the first stage. The evolution of the system was done by using precision and recall as a performance metrics. The end results of this paper were the semantic web which is an efficient way to provide variety of related results based on the meanings not syntax because of the detailed information of words and this will reduce the cycle time for search.

The future work will attempt to improve the results query by using planning to receive more specific results and give suggestions about equivalent query which have the same meanings.

References