

# Building an Arabic Semantic Lexicon for Hajj

Omar Batarfi  
Faculty of Computing and  
Information Technology  
King Abdulaziz University

Mohamed Yehia Dahab  
Faculty of Computing and  
Information Technology  
King Abdulaziz University

Ahmed Ezz  
Faculty of Science  
King Abdulaziz University

## ABSTRACT

Semantic lexicon is a lexicon augmented with information of lexical relationships among words. Although the semantic lexicon is the backbone of many intelligent applications, there is no serious effort has been done in developing an Arabic semantic lexicon. The main goal of this work is to build an automatic Arabic semantic lexicon. To achieve this goal, we select an Arabic dictionary and augment it with morphological information and semantic features such as Patterns, Meronymy, Holonymy and etc. The obtained results show that the objectives of this work are successfully accomplished, relations between different terms have been built and the glosses are automatically extracted for these terms.

## General Terms

Natural Languages Processing (NLP), Information Extraction.

## Keywords

Semantic Lexicon, Information Extraction, Morphology, Semantic Patterns, Lexical Relationships.

## 1. INTRODUCTION

The aim of this work is to provide a simple method to extract lexical entries and semantic lexical relations. The goal of semantic lexicon is to provide an automatic model capable of extracting words from text in a specific domain. A lexicon includes a wide array of information associated with entries. An entry in a lexicon is usually the base form of a word, the singular for a noun and the present tense for a verb [1]. As linguistic theories have progressed in modeling human language ability, the lexicon has become more central to those theories. The term computational applies in several senses for computational lexicons. Essentially, the lexicon is in an electronic form. With this new or renewed attention to the mental lexicon, two problems become evident. There is no generally accepted theory of how the lexicon is internally structured and how lexical information is represented in it [2].

### 1.1 Lexical analysis

A word can be thought of in two ways, either as a string in running text, for example, the verb delivers; or as a more abstract object that is the cover term for a set of strings. So the verb DELIVER names the set {delivers, deliver, delivering, delivered}. A basic task of lexical analysis is to relate morphological variants to their lemma that lies in a lemma dictionary bundled up with its invariant semantic and syntactic information. Lemmatization is used in different ways depending on the task of the natural language processing (NLP) system. In machine translation (MT), the lexical semantics of word strings can be accessed via the lemma dictionary [3].

### 1.2 Lexical Semantics

The study of what individual lexical items mean, why they mean, what they do, how can we represent all of these, and where the combined interpretation for an utterance comes

from. Lexical semantics is concerned with the identification and representation of the semantics of lexical items.

### 1.3 Major Sense Relations

In general linguistics, semantic analysis refers to analyzing the meanings of words, fixed expressions, whole sentences, and utterances in context. Lexical resources are commonly organized according to lexico-semantic relations such as antonymy, synonymy, hyponymy and meronymy. The term lexical relation is used here to indicate any paradigmatic relation among words, not just a semantic relation. So, lexical relations include phonetic relations (such as rhyme or alliteration), morphological relations (such as inflectional variation), and morpho-syntactic relations (such as co-membership in a grammatical category). Most lexical semantics texts claim that semantic relations are not really relations among words, but relations among word senses. Some of these texts call these relations sense relations [4] or meaning relations [5] rather than lexical relations. Many researchers have applied text- and web-mining algorithms for automatically creating lexical semantic resources like similarity lists [6], semantic lexicons [7], hyponymy lists ([8]; [9]), part whole lists [10], and verb relation graphs [11].

So far, the topic of study has been described as paradigmatic semantic relations among words. In the literature, these relations are usually called lexical relations or semantic relations, and sometimes those two terms are used contrastively [12]. WordNet makes the common accepted distinction between conceptual-semantic relations, which link concepts, and lexical relations, which link individual words [13]. Lexical semantics aims at precisely analyzing the meanings of lexical items, either by analyzing their internal structure and content or by representing their relations to other elements in the lexicon.

### 1.4 Sense Relations and Ontologies

Sense relations —semantic relations between lexical elements— form the basis for word nets, such as the electronic lexical database WordNet [14] and similar approaches for languages other than English, for example, the multilingual lexical database EuroWordNet [15] and [16]. Sense relations can be seen as revelatory of the semantic structure of the lexicon. There are both horizontal and vertical sense relations. Horizontal relations include synonymy (sameness of meaning of different linguistic forms, such as Orange and Apfelsine in German—both meaning ‘orange’) and various relations that can be subsumed under the general notion of opposition [17]. It should be emphasized that when talking about sense relations we are talking about meaning relations between lexical items. These relations have to be distinguished from ontological relations, notwithstanding that there is a close link between the two: “sense relations are relations between words (in a reading) based on ontological relations between the concepts that constitute the meanings of these words (in that reading)” [18]. Ontologies can be

conceived of as networks of cross-connected conceptualizations, with the relations holding between those conceptualizations being ontological relations. (This, admittedly, is a very broad definition, which in principle allows any conceptual relation to be subsumed under the label “ontological relation.”) Ontological relations might exist without being reflected in a language’s lexicon.

## 1.5 Properties of Semantic Relations

Murphy in [12] portrayed the properties of semantic of semantic relations. These properties can be listed as following:

1. **Productivity**: New relational links among words can be generated.
2. **Binarity**: Some relations relate only pairs of words, although larger sets of words may be semantically available for the relation (e.g., black/white rather than black/gray/white).
3. **Variability**: Which word(s) a particular word is related to varies according to which sense of the word is used and the context in which it is used.
4. **Prototypicality** and canonicity: Some word sets better exemplify a relation than others, and some word sets (especially some antonym pairs) seem to have special status as canonical examples of a relation.
5. **Semi-semanticity**: Semantic properties of words are not the only factors at work in relating words and judging semantic relations.
6. **Uncountability**: The number of semantic relation types is not objectively determinable.
7. **Predictability**: Relations among words adhere to general patterns, indicating that semantic relations are rule governed.
8. **Universality**: The same semantic relations are relevant to the description of any language’s lexicon.

## 2. PROBLEM DEFINITION

A lexicon is a list of words in a language, a vocabulary, along with some knowledge of how each word is used. A lexicon may be general or domain-specific; we might have, for example, a lexicon of several thousand common words of Arabic. A lexicon may include multi-word expressions such as noun phrases, and other common expressions (‘أهلا وسهلا’, ‘عيد سعيد’).

Each word or phrase in a lexicon is described in a lexical entry. The lexical entry may include any of its properties of spelling or sound, morphological information, grammatical behavior and meaning. A lexical entry is therefore a potentially large record specifying many aspects of the linguistic behavior. Semantic lexicon is a lexicon augmented with information of lexical relationships among words. The lexical relationships include the following [17] and [19]: Synonymy and its inverse Antonymy.

Synonymy is the most common semantic relation between words and roughly expresses similarity between words (or expressions). It is based on equivalence of meaning such as ‘المسكين’ and ‘الفقير’. Antonymy: It expresses the opposite of the similarity relation such as ‘الحلال’ and ‘الحرام’.

- Hyponymy and its inverse hypernymy (hyperonymy): Hyponymy is based on similarity but central here is similarity within a class. Hyponymy depends on the

notion of inclusion, or more specifically on the inclusion of one class in another. For instance, the classes denoted by the words ‘جبرائيل’ and ‘ميكايل’ are both subclasses of the superclass that is denoted by ‘الملائكة’. We therefore say that ‘جبرائيل’ is a “hyponym” of ‘الملائكة’ and the other way around that ‘الملائكة’ is a “hypernym” of ‘جبرائيل’.

- Meronymy and its inverse Holonymy: Meronymy expresses a part-whole relation between objects (or classes of objects) that is denoted by words. For instance, ‘الانف’ is part of ‘الوجه’, which is part of ‘الجسم’. Holonymy is the opposite of meronymy. It defines the relationship between a term denoting the whole and a term denoting a part of, or a member of, the whole.

There are many efforts have been done in developing Arabic information retrieval and extraction applications, but these applications have lacks in understanding implicit knowledge. The main reason of these lacks is the absence of linguistic knowledge and semantic relations between words. To overcome these lacks, there is a need to develop a semantic lexicon which includes both linguistic knowledge and semantic relations between words. For English language, there are number of semantic lexicons are established such as: WordNet, FrameNet, and etc. There are little efforts have been done in establishing Arabic semantic lexicon such as [20] and [21]. Consequently, there is a need to develop an Arabic semantic lexicon that is the main goal of this research project.

## 3. RESEARCH OBJECTIVE

In this work we are trying to provide a simple method to extract lexical entries and semantic lexical relations. To extract lexical entries and semantic relations, we analyze an Arabic dictionary that is written in natural free text and process it without a need for predefined format or syntax rules. We process the gloss, text that is used to describe terms, to learn new terms and to learn new lexical relation. We believe that the gloss, which is a brief notation of the meaning of a word in a text, describes a term on term of another term(s). For example, the dictionary writers often use the Arabic proverb “things are identified by their opposites”. The application of this proverb can be found in the identification of the term ‘الحرام’, ‘forbidden’, by saying “والحرام: نقيض ‘الحلال’”, it is an antonym of ‘permissible’.

The system has two sets that are term (T) and gloss (G) sets.

$T = \{t_1, t_2, t_3, \dots, t_{nt}\}$ , Where  $nt$  is the number of terms, and

$G = \{g_1, g_2, g_3, \dots, g_{ng}\}$ , where  $ng$  is the number of glosses. In simple case, each term has only one gloss, this mean that  $nt = ng$ . There are two types of terms: original terms (OT) that has been mentioned explicitly in the dictionary text and while candidate terms (CT) that can be processed from OT, consequently,  $T = OT \cup CT$ . Such that OT is the set of original terms that has been mentioned explicitly in the dictionary text, and CT is the set of candidate terms that can be processed from OT.

$OT \cap CT$  does not necessary equals  $\{\}$ . The candidate terms can be learned from compound terms by shallow natural language processing. Most of Arabic compound terms are in the form of <noun> <modifier>+, from right to left, i.e. one modifier or more. For example, ‘طواف الوداع’, ‘الوداع’ is the modifier and ‘طواف’ is the noun. We mean by a modifier here is a word that limits or qualifies the sense of another word. We consider in this case, ‘طواف الوداع’ is a special case

of 'طواف' <noun> <modifier> is a special case of <noun>. According to this consideration, we add 'طواف' to the set of candidate terms. Also, the candidate terms can be extracted from simple terms, i.e. unigram, by extracting the root from the simple term and searching for different derivations and inflections of the extracted root in the text that describe the term. For example, the term "الحرم", the extracted root from this term is "حرم". By searching for all different derivations and inflections of the extracted root "حرم" that are found in the its gloss we can suggest the following terms: "الحرم", "الحرام", "أحرام", "حرم", "أحرم", "حراماً", "الحريم", "حرم", "الحرم".

"حرام", "أحرام", "الحرم", "المحرم", "المحرمة", "الحرم", "أحرام", and "أحرام".

After building the set of candidate terms, CT, we search in the glosses for sentences that hold a term or two terms, terms from  $OT \cup CT$ . We believe that these sentences most often represent lexical relations among terms.

It is important to assign a value to each word in the all glosses. The words that have high frequent, depending on thresholds, will have a low value because these words represent stop words that do not have weight that affect on the meaning. Terms will not be included in both the assigned values and frequency.

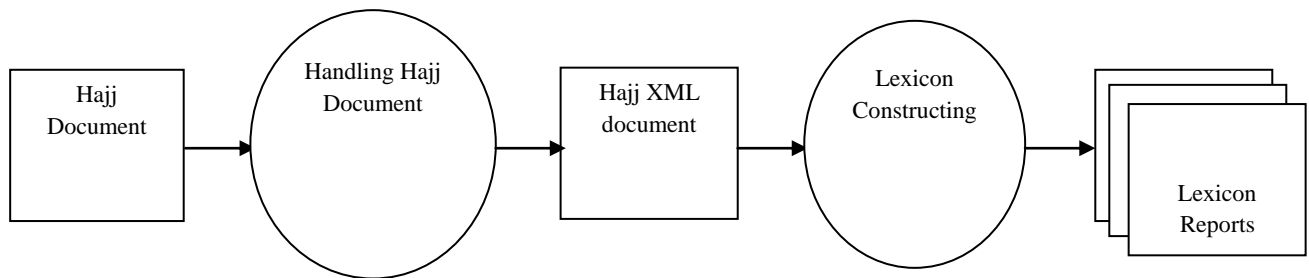


Figure 1: System Design

#### 4. RELATED WORK

Lexical semantic is the study to explore the knowledge. Exploring the meaning of the words is considered as a type of knowledge. Based on this, lexical semantic has been used to facilitate in exploring query meaning to extract data between different languages. Cross-Language Information Retrieval uses lexical semantic to do matching between queries and documents that do not necessarily share the same language. Translation is required to do the matching between distinct vocabularies. Diekema used a lexical framework to build single lexical knowledge base to be used to do query translation and also to eliminate any ambiguity in Cross-Language Information Retrieval [22]. He built this preliminary framework to be used as a methodology to perform the automatic combination of various semantic lexical resources such as machine readable dictionaries, ontologies, encyclopedias and machine translation lexicons. The preliminary framework uses resource matrix to provide an overview of the types of information in the collected resources and of certain resource characteristics. Dictionary entries base form was included in the matrix to show if the entries are singular form for noun, indefinite form for verbs, roots and stems. These features of different resources assisted in forming an intelligent resource combination after mapping different resources, threshold has been set to show the certainty of translation among some of the combinations. The created preliminary framework was applied on five English and Arabic language resources include: Word net 2.0, the lexicon of buckwalter stemmer, Ajeeb translation, Arabeyes' worldlist and Gigaword corpus. These five resources was filled the feature matrix. The evaluation of the preliminary framework showed a successful translation with none not fixed problem that related to part-of-speech differs between the Arabic term and the translation.

Ruiz also uses semantic lexical based on the conceptual-semantic and lexical relations between synsets (sets of cognitive synonyms) in the WordNet to do machine translation between English and Chinese [23]. WordNet synsets are terms grouped together because they share the same meaning [24]. Ruiz mainly work is to built bilingual

mapping English-Chinese lexicon into WordNet (lexical triangulation). The mapping process uses the translation evidence from multiple sources to be used as thresholds. Thresholds are used to create a frequency-ranked list of translation which selects the most likely correct translation.

Ting Qian et al. built their semantic lexicon for finding the contrasting semantic categories from a source lexicon [25]. The proposed method, called bootstrapping, is applied on English nouns to discover the distinct semantic of these nouns; therefore, syntactically similar phrases and sentences will clarify any ambiguity which will help in providing accurate semantic interpretations. A small set of examples is used by Bootstrapping as a seed to discover words with semantic properties. These examples are changed their semantic properties to discover different semantic classes. Feature of each word is examined by a perceptron-based model to extract two kinds of features: morphological (affix and word length), and contextual [26]. Affixes play important role in exhibiting meaningful distinctions when they are modified. World length indicates if the word has more syllables or not because longer words are more likely to contain affixes. Accordingly, words have classified to four groups: shortest (fewer than 5 letters), short (5-7), medium (8-12), long (13-19), and longest (> 19). KNEXT system produced what is called "factoids" that logically extracted from text [27]. These factoids imply classifier to distinguish event nouns from others. Perceptron-based linear classifier is trained iteratively by a bootstrapping procedure. The examples will be used in the training the perceptron algorithm to determine whether the active features of the nouns have similarity with the categories of examples. In the next iteration, the size of the training examples will include the new discovered instances. The lexicon that is chosen for implementation was the WordNet noun-list which contains 21,512 nouns, and the seed set includes 15 event nouns and 215 non-event nouns [28]. The implementation of training a perceptron to discover Event Noun (e.g. birth, celebration) is divided on two Trials: Trial 1 must show features at least 10 times, and Trial 2 must show features at least 15 times. The result of the experiment shows that for both manual and

automatic classification, the nouns are classified successfully in spite of the initial training set that very small.

The research [29] stated that: “Many unilingual dictionaries exist for many languages; Arabic has no such an electronic one”. A pioneer work has been done in HIAST where the data from "Al Wasseet Dictionary “ has been computerized in a form of a Database. The dictionary is enriched by adding other fields by Arabic linguists [30]. The database includes 16 tables: verb tables, noun tables, idiom tables, particles tables and additional two tables.

The research [31] described how to build a large comprehensive, integrated Arabic lexicon by automatic parsing of newspaper text. They have built a parser system to read Arabic newspaper articles, isolate the tokens from them, find the part of speech, and the features for each token. The system develops and enriches the dictionary every time it analyzes a new text.

The research [32] designed a methodology to build an automatic Arabic lexicon that contains morphological information, part of speech tag, linguistics attributes, patterns and affixes for all lexicons entries.

The researchers in [33] designed an ontology that focused on Arabic language vocabulary (nouns only) associated with the semantic field of “Time”. They limited the vocabulary to those words which exists in the Holy Quran. The basis of their theoretical foundation is based on the field theory of semantics.

An important Arabic corpus with syntax and semantic multilabel organized in a hierarchy fashion found in the research[34] as well as how to generate different Arabic words found also in [35].

The research [36] was the first attempt to build an Arabic sentiment lexicon for sentiment analysis on a basis of Arabic WordNet.

## 5. SYSTEM DESIGN

Figure 1 shows the high level design of the proposed system.

The system has two main processes:

1. Handling hajj document process
2. Constructing Lexicon process

The Handling hajj document process receives the hajj document and reads it through word automation and then generates XML format file. The structure of each XML file contains four tags: root, term, name, and gloss. Figure 2 shows the structure of the XML files.

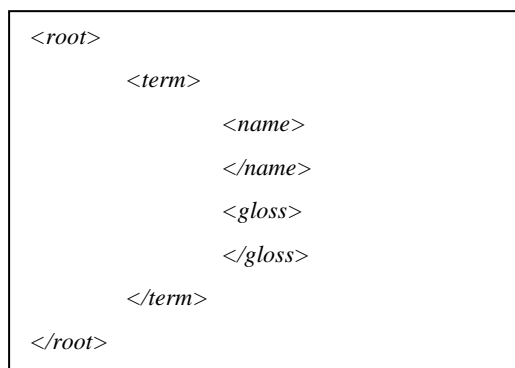


Figure 2: Structure of XML file

The root tag is used by XML parser to indicate the starting and ending of the XML file. The term tag contains the data of each term in Hajj. Each term has two data items: name, and gloss. The name tag contains the term’s name, and the gloss tag contains the details data of the term as written in the original Hajj document. The hajj document contains 16307 words. The document has three parts:

1. Time Terms
2. Place Terms
3. Hajj Terms

Consequently, we will construct three XML files: “time.xml”, “place.xml”, and “hajj.xml”. The three files have same structure.

The Constructing Lexicon process receives the XML files and then constructs the Hajj lexicon. Figure 3 shows the phases of the Constructing Lexicon process.

To develop a semantic lexicon of nouns for hajj we select a dictionary for hajj (the dictionary is domain dictionary and it should be a machine readable, it is not necessary to be structured but if the dictionary is structured it would save time and helpful) then we followed up the following phases.

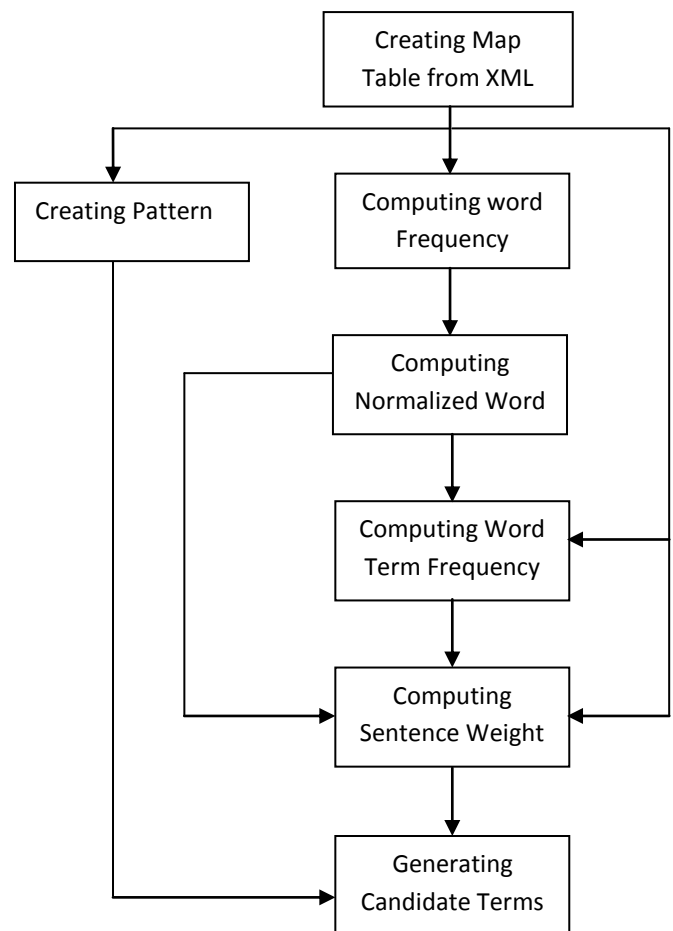


Figure 3: phases of the Constructing Lexicon process

## 5.1 Dictionary Pre-processing

In this phase, we used unstructured dictionary namely “كتاب الحج” by Amer Anwar Alzobadany [36]. The dictionary contains terms and glosses.

- **Terms:** Are the terms or nouns that can be used in hajj e.g. أيام التشريق, يوم التروية, مكة. A term maybe
  - Simple concept, contains a single noun such as مكة, مني
  - Compound concept, contains more than a noun such as أيام التشريق.
- **Glosses:** A gloss is a free text of brief explanatory, note and/or relations between other terms, or may include all. A gloss may include the original linguistic meaning and linguistic features and/or domain professional meaning.

Both terms and glosses have different appearance such as:

- They may appear with diacritics and may not e.g. كِتَابِ اللَّهِ, كِتَابِ اللَّهِ. The diacritics also may differ in different positions e.g. كِتَابِ اللَّهِ, كِتَابِ اللَّهِ.

- They may have determiner (ال) and may not
- They may be plural, dual or single
- The term may appear in masculine form like مؤمن or may appear in feminine form like مؤمنة

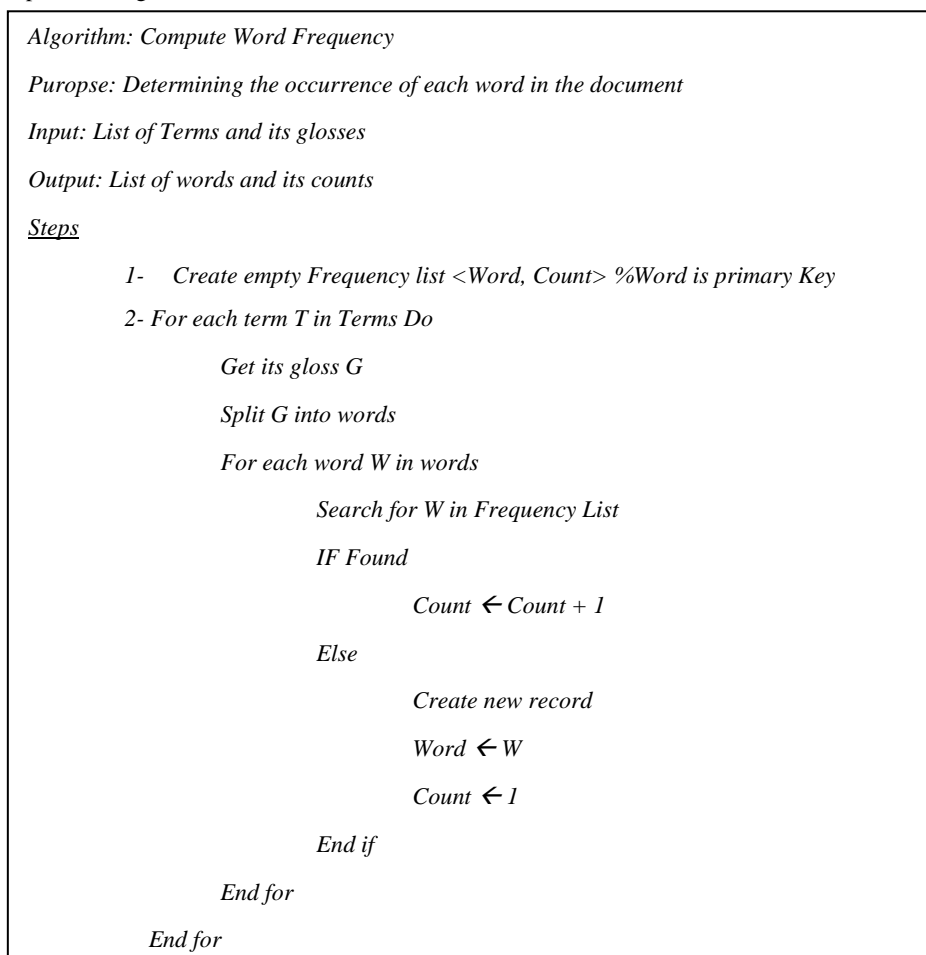
In addition to previous linguistics prefixes and postfixes, the following prefixes and postfixes should be removed before any process:

- They may attach with a symbol or more that may not represent a linguistic aspect such as “: , \* -- -- >”

The main objective from this phase is to create, from natural text; an XML file contains a structured database from term and gloss.

## 5.2 Compute Word Frequency

The main idea behind this phase is to discover the importance of each term analyzing the gloss of each term. Analyzing the glosses by counting the terms in whole dictionary to know how much each term has occurred to describe other terms. The input of this phase is the structured dictionary while the output is frequency list that contains word and count. Figure 4 shows the Compute Word Frequency Algorithm



**Figure 4: Compute Word Frequency Algorithm**

The proposed enhancement in this phase

- The first proposed enhancement in this phase is save word without neither prefixes nor postfixes.

- Search in Frequency List by all possible forms of the word. We mean by possible forms the following:

- possible diacritics for each word
- the existence of determiner (ال)
- the form of single (masculine or feminine), dual (masculine or feminine) and plural (masculine and feminine)

### 5.3 Normalize Words

The input of this phase is MapTable <String, Integer> which represents terms and each occurrence in the different glosses. The output also in this phase is another MapRootTable <String , Array\_Of\_String > which map the normalized term to integer or count of occurrence of the normalized term which represent a global data member and wordNormalizedCountMap that represents each normalized term and its occurrence

### 5.4 Compute Word and Term Frequency

The main idea behind this phase is to discover the importance of each word of each term by analyzing the gloss. Analyzing the term gloss by counting the word to know how much each word has been occurred in describing the current term.

The input of this phase is MapTable <String, Integer> which represents terms and each occurrence in the different glosses also Map <String, String> which represent each term and its gloss. The output also in this phase is another wordTermWeightMap <String-String , float > which map the normalized term concatenated with word as a key map to float which represent the weight of the word in this term.

### 5.5 Compute Sentence Weight

The main objective for this phase is to compute the weight of each sentence or gloss of each term. We analyze the term gloss by counting the weight of each word that is used in describing the current term. We normalize the weight by removing the maximum word weight from sum and also by calculating the average of weight.

### 5.6 Generate the Candidate Terms from Pattern

Some patterns that we discovered are very useful to discover new terms or more precisely candidate terms. Examples of these pattern ويسمى ايضا, سميت, تسمى.

This phase includes many tasks such as:

- CleanStr is a method that remove any special character such as % \$, ^, -, etc. from term and gloss
- Using XML file that stores the pattern list
- Regular expression to search for patterns in any given gloss sentence

### 5.7 Create Patterns

In previous phase we used predefined patterns which are very difficult to discover. In this phase we are trying to discover the patterns automatically without any human involvement.

In this phase we try to discover patterns that used between different terms and also we give to these patterns weights according to how much these patterns have been occurred between different terms and how much these patterns can discover different relation between terms.

## 6. RESULTS

For evaluation we selected one part of the dictionary for this task and the results were as follow: We extracted 2814 words which have been normalized, by removing prefixes and postfixes, into 2265 words. We applied the predefined patterns into the selected part only to know if these patterns can be used to extract relations between terms. The number of these patterns was eight.

Items	Number
Number of right relations	17
Number of right relations with simple error	31
Number of wrong relations	20

We extracted 10 patterns automatically that can be used after that to discover new relations.

## 7. CONCLUSION AND FUTURE WORK

In this work, automatic Arabic semantic lexicon was built. The outcome for this semantic lexicon based on unstructured dictionary namely “كتاب معجم مصطلحات الحج” by Amer Anwar Alzobadany. We augment the dictionary with morphological information and semantic features. Patterns were discovered automatically and determined their weights according to their occurrence. These patterns are used to generate candidate terms. The extracted terms assisted built glosses automatically.

In the future work we suggest using different dictionaries in the same domain to validate the results and enhance the output of the system. Further enhancement includes augmenting a well stemmer to extract more valid terms and semantic relations. This work could be applied to cover different specialized dictionary such as Quran dictionary, Sunnah dictionary and Feqah dictionary to enrich Islamic lexicons. For better presentation, the system could exhibit the terms and their relations in user friendly interface to will be understandable by different users.

## 8. ACKNOWLEDGMENTS

This work was supported by King Abdulaziz University, Deanship of Scientific Research Project No. (430/14-007).

## 9. REFERENCES

- [1] Singleton D., “Language and the Lexicon”, Oxford University Press Inc., New York, 2000.
- [2] Dirk Geeraerts, “Theories of Lexical Semantics”, Published in the United States by Oxford University Press Inc., New York, 2010.
- [3] Nitin I., Fred J., “Handbook of Natural Language Processing”, Second edition, Chapman & Hall/CRC, ISBN: 978-1-4200-8593-8, 2010.
- [4] Lyons, J., “Semantics” (2 vols.). Cambridge: Cambridge University Press, 1977.
- [5] Allan, K., “Linguistic meaning” (2 vols.). London: Routledge, 1986.
- [6] Lin, D. “Automatic retrieval and clustering of similar words”, In Proceedings of COLING/ACL-98. pp. 768-774. Montreal, Canada, 1998.

- [7] Riloff, E. and Shepherd, J., “A corpus-based approach for building semantic lexicons”, In Proceedings of EMNLP-1997, 1997.
- [8] Shinzato, K. and Torisawa, K., “Acquiring hyponymy relations from web documents”, In Proceedings of HLT-NAACL-2004. pp. 73-80. Boston, MA.,2004.
- [9] Pantel, P. and Lin, D., “Discovering Word Senses from Text”, In Proceedings of SIGKDD-02. pp. 613-619. Edmonton, Canada, 2002.
- [10] Girju, R.; Badulescu, A.; and Moldovan, D., “Learning semantic constraints for the automatic discovery of part-whole relations”, In Proceedings of HLT/NAACL-03. pp. 80-87. Edmonton, Canada, 2003.
- [11] Chklovski, T., and Pantel, P., “VERBOCEAN: Mining the Web for Fine-Grained Semantic Verb Relations”, In Proceedings of EMNLP-2004. pp. 33-40. Barcelona, Spain, 2004.
- [12] Lynne Murphy, “Semantic Relations and the Lexicon, Antonymy, Synonymy, and Other Paradigms”, Cambridge University Press, ISBN -13: 978-0-511-06899-7, 2003.
- [13] Evens M., “Relational models of the lexicon”, Cambridge University Press, Cambridge, England, 1988.
- [14] Fellbaum C., “A semantic network of English: The mother of all WordNets” In EuroWordNet: A Multilingual Database with Lexical Semantic Networks, P. Vossen (ed.), pp. 209–220. Dordrecht, the Netherlands: Kluwer. 1998.
- [15] Vossen, P., “Condensed meaning in EuroWordNet.” In The Language of Word Meaning, P. Bouillon and F. Busa (eds.), pp. 363–383. Cambridge, U.K.: Cambridge University Press, 2001.
- [16] Vossen, P. (ed.)”EuroWordNet: A Multilingual Database With Lexical Semantic Networks.” Dordrecht, the Netherlands: Kluwer. [Reprinted from Computers and the Humanities 32(2/3)], 1998.
- [17] Cruse, D. Alan, “Lexical Semantic”, Cambridge: Cambridge University Press, Published by the Press Syndicate of the University of Cambridge, 1986.
- [18] Nickles, M., A. Pease, A. C. Schalley, and D. Zaefferer, “Ontologies across disciplines.” In Ontolinguistics: How Ontological Status Shapes the Linguistic Coding of Concepts, A. C. Schalley and D. Zaefferer (eds.), pp. 23–67. Berlin, Germany: Mouton de Gruyter, 2007.
- [19] Green, Rebecca; Bean, Carol A.; and Myaeng, Sung Hyon (eds.), “The Semantics of Relationships: An interdisciplinary perspective”, Kluwer Academic Publishers, Dordrecht. 2002.
- [20] S Elkateb and W Black and P Vossen and D Farwell and et al., “Arabic WordNet and the Challenges of Arabic”, proceedings of Arabic NLP/MT Conference, London, 2006.
- [21] Elkateb, S., Farreres, J., Fellbaum, C., Pease, A., and Black, W., “Arabic WordNet: A Linguistic Resource with a Deep Formal Semantic Foundation”, in Proceedings of the 6th International conference on Language Engineering, Cairo. 2006.
- [22] Anne R. Diekema, Center for Natural Language Processing, Syracuse, NY. Proceeding Semitic '04 Proceedings of the Workshop on Computational Approaches to Arabic Script-based Languages Association for Computational Linguistics Stroudsburg, PA, USA ©2004
- [23] M.E. Ruiz, et al. 2001. CINDOR TREC-9 English-Chinese Evaluation. In “Proceedings of the 9th Text REtrieval Conference (TREC-9)”, E.M. Voorhees and D.K. Harman ed., pages 379-388, NIST, Gaithersburg, MD.
- [24] Miller. 1990. WordNet: An On-line Lexical Database. International Journal of Lexicography, 3(4), Special Issue.
- [25] Benjamin Van Durme, Ting Qian, and Lenhart Schubert. 2008. Class-driven Attribute Extraction. In COLING.
- [26] Frank Rosenblatt. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, 65(6):386–408.
- [27] Lenhart K. Schubert. 2002. Can we derive general world knowledge from text? In HLT.
- [28] Christiane Fellbaum, editor. 1998. WordNet: An Electronic Lexical Database. MIT Press.
- [29] Al Dakkak, O., and Zein, A., “Towards Arabic Electronic Dictionary”, 3rd International Conference on Information and Communication Technologies: From Theory to Applications, 2008. ICTTA 2008.
- [30] S. Attar, M. Bawab, and O. Al Dakkak, "Arabic Lexical Database", Workshop on Arabic Natural Language Processing, ICTIS 2007, April 2007.
- [31] Ben Halima M., and Alimi M., “An interactive system for Extracting Arabic Lexicon from Arabic Newspaper text”, Fifth International Conference on Innovations in Information Technology, 2008.
- [32] Al-Shalabi R., and Kanaan G., “Constructing An Automatic Lexicon for Arabic Language”, International Journal of Computing and Information Sciences, Volume 2, Number 2, August 2004.
- [33] Al-Yahya, M.; Alkhalifa, H.; Bahanshal, A.; Alodah, I.; Al-Helwah, N.; “An Ontological Model for Representing Computational Lexicons, A Componential Based Approach”, International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE), 2010.
- [34] Siddiqui, Muazzam Ahmed, Mohamed Yehia Dahab, and Omar Abdullah Batarfi. "Building a sentiment analysis corpus with multifaceted hierarchical annotation." International Journal of Computational Linguistics 6.2: 11-25, 2015.
- [35] Aqel, Afnan, Sahar Alwadei, and Mohammad Dahab. "Building an Arabic Words Generator." International Journal of Computer Applications 112.14,2015.
- [36] Mahyoub, Fawaz HH, Muazzam A. Siddiqui, and Mohamed Y. Dahab. "Building an Arabic sentiment lexicon using semi-supervised learning." Journal of King Saud University-Computer and Information Sciences 26.4: 417-424, 2014.