Birzeit University
Faculty of Information Technology
Master Program in Computing

خوارزمية ربط مفاهيمي
نحو ربط آلي بين المفاهيم العربية والانجليزية
Advanced Matching Function
Towards linking Arabic Ontology with the English WordNet

Prepared By
Mohammed M. Melhem

Supervisor
Dr. Mustafa Jarrar

M.Sc. Master Thesis

1433-2012

Advanced Matching Function towards linking Arabic Ontology with the English WordNet

Prepared By
Mohammed M. Melhem

Supervisor: Dr. Mustafa Jarrar

A thesis submitted in partial fulfillment of requirement for the degree of master of Computing - Faculty of Information Technology - Graduate Studies
Birzeit University

1433-2012

Birzeit University
Faculty of Information Technology
Master of Computing

Thesis Approval

Advanced Matching Function towards linking ArabicOntology with the English WordNet

Prepared by: Mohammed Majed Melhem
Registration No: 1095086

Supervisor: Dr. Mustafa Jarrar

Master thesis submitted and accepted, date:

The names and signatures of the examining committee members as follows:

1. Head of committee: Dr. Mustafa Jarra **Signature** ……...

2. Internal examiner: Dr. Bassem Sayrafi **Signature** .....

3. External examiner: Dr. Hashem Tamimi **Signature** .....

Birzeit-Palestine

1433-2012

Dedication

To my beloved ones

To true friends

Acknowledgements

Finally, I am writing this page. The relief and an achievement I feel proud to complete, another stage in my life towards new goals.

First, I wish to express my gratitude towards Dr. Mustafa Jarrar for his support and contribution to this thesis, particularly to his guidelines and continuous support to accomplish this research.

To Rana Rishmawi for the support and help she provided starting from data preparation, and other tasks like providing a good training set, I'm in debt to you.

To Sina institute members for your work to build the Arabic Ontology and to enrich the Arabic electronic content.

Abstract

The major challenge in constructing an Arabic ontology is that it needs many resources, manual work, and semi-manual techniques that require creating a sophisticated tool to ease the process. Such tool should provide the technique and ability to link and benefit from other existing linguistic resources. This tool will be used to link between concepts from different multilingual resources, based on a concept matching function.

The matching function is an algorithm that was developed In Birzeit University, by undergraduate students. It maps Arabic concepts in the Arabic Ontology with their equivalent concepts in the English WordNet. The existing function carries a certain level of tolerance, but suffers from some limitations in the matching process as it uses fixed parameters. These parameters are categorized into two main categories of weighing parameters that include; Keyword, Super Type, Sub-type, and Synonyms as its main parts. The second category is the expansion levels of processed data by this function.

This thesis introduces a novel approach to enhance the matching function, by increasing the accuracy of matching operation, minimizing the resulted errors, and increasing the performance of the algorithm. The presented solution utilizes machine-learning approach to configure and tune the mentioned parameters. The followed methodology resulted in an enhancement on the matching operation, where the overall accuracy for the top 15 matched concepts in the enhanced matching function represented in 55% compared to 41% achieved by old version of the matching function, showing a 14% of improvement.

الخلاصة

التحدي الرئيسي في بناء الأنطولوجيا العربية يكمن في حاجتها الى الكثير من الموارد، والعمل اليدوي، والطرق الشبه آلية للوصل الى انطولوجيا عربية. فمن هنا تبرز الحاجة الى التقنيات التي توجهنا الى خلق أداوات متطورة لتبسيط وتسهيل هذه العملية، ومثل هذه الحلول التقنية يجب ان تراعي الحاجة الى الاستفادة من الموارد اللغوية المتوفرة وأن تقوم بتوفير آلية ربط بين هذه المصادر والانطولوجيا العربية. هذه الاداة سوف تستخدم لخلق ربط بين مختلف المفاهيم ضمن اللغات على مبدأ التماثل في المعنى ضمن المفاهيم اللغوية.

الأداة المتوفرة حاليا (الخوارزيمة) التي تم تطويرها في جامعة بيرزيت من قبل مجموعة من طلاب البكالوريوس، تقوم على مبدأ ربط المفاهيم العربية ضمن الانطولوجيا العربية مع ما يعادلها من مفاهيم اللغة الانجليزية في قاعدة بيانات .WordNet حيث تقوم هذه الخوارزمية بتوفير مقدار من التفاوت المقبول في عملية الربط، ومن المحددات التي تعاني منها الخوارزمية هي استخدامها مجموعة من المعاملات الثابتة. حيث تنقسم هذه المعاملات المجموعتين رئيسيتين؛ هما متغيرات الاوزان والتي تشمل الاوزان التي تعطى لكل من العلاقات اللغوية جزء من، والمرادفات وعلاقة جنس من. المجموعة الثانية وتشمل المتغيرات التي تتحكم بحجم وكمية البيانات التي تتم معالجتها بواسطة الخوارزمية.

تقدم هذه الرسالة منهجا جديدا لتعزيز عملية الربط بين المصادر اللغوية، من خلال زيادة دقة عملية الربط والمطابقة بين المفاهيم ضمن المصادر اللغوية، وتسعى الدراسة ايضا الى التقليل من نسبة الأخطاء الناتجة عن عملية الربط و وتحسين أداء الخوارزمية. يقوم الحل المقترح على نهج التعلم الآلي الذي يقوم بالتعلم وضبط المعاملات السابقة الذكر، ونتيجة اتباع هذه المنهجية حصلنا على تحسن في مقدار دقة النتائج المطابقة الواقعة في مجال اعلى 15 مفهوم حيث ان النتيجة التي وصلنا اليها بواسطة التجربة المقترنة بهذه الرسالة تقدر بـ 55% مقارنة بالنتيجة الناتجة عن آلية الربط القديمة 41% من مجمل التطابقات الصحيحة، والتي تشكل تحسنا بمقدار 14% من مجمل النتائج.

Table of Contents

Abstract I

الخلاصة II

Chapter 1 : Introduction 1

1.1 Scope and motivation 1

1.2 Problem statement 3

1.3 Summary of contribution 4

1.4 Summary and structural outline 6

Chapter 2 : Background and Related work 8

2.1 Background of the WordNet 8

2.2 Short Overview of Ontology 12

2.3 The Arabic WordNet, and the Arabic Ontology 13

2.4 Concept Matching 17

2.5 Machine Learning and Ontology Matching 20

Chapter 3 : Introducing Machine Learning to the Matching Function 24

3.1 Illustrating the Matching Function by a Running Example 24

3.1.1 The Input and the Output of the Matching Process 25

3.1.2 Preparation and Translation Phase 26

3.1.3 Defining the Search Space 27

3.1.4 Defining the Query 28

3.1.5 The Matching Process 30

3.1.6 Centrality Calculation 31

3.1.7 The Learner 33

3.2 Structural Design 37

3.2.1 Core component 37

3.2.2 The Learner 45

3.3 Algorithm Complexity 52

3.3.1 The complexity of core component 52

3.3.2 The complexity of the learner 54

Chapter 4 Evaluation 56

4.1 Introduction 56

4.2 Data set 56

4.3 Evaluation setup 58

4.4 Evaluation Results 60

4.5 Discussion of Results 63

Chapter 5 Conclusions and Recommendations for Future Work 66

5.1 Conclusions 66

5.2 Recommendations for Future Work 67

References 68

Appendices 73

Appendix I Gloss construction instructions 74

Appendix II Training data sample 76

List of Tables

Table 3-1Learning item sample 34

Table 3-2 Gloss (G) matching operation hit against matched gloss (M) 34

Table 3-3 Neural network results of processing the example results 36

Table 4-1 Cross validation dataset distribution 59

Table 4-2 Initial Matching Function configured variables 59

Table 4-3 3-Cross validation accuracy comparison 61

Table 4-4 Expansion variables change between evaluation runs 63

Table 4-5 Accuracy comparison between with learning matching function and non-neural matching function 63

List of Figures

Figure 2.1 WordNet structure [8] 9

Figure 2.2 English WordNet relation database illustration 11

Figure 2.3 Arabic ontology structure [5] 16

Figure 2.4 Concept matching illustration 17

Figure 3.1 Matching function input and output 25

Figure 3.2 Search space overview 28

Figure 3.3 Illustration of query Q as an array 29

Figure 3.4 Structured illustration of the query Q 30

Figure 3.5 Snapshot of running Q against S concepts 30

Figure 3.6 Illustration of centrality calculation [8] 32

Figure 3.7 Search space(S) results after running the centrality 33

Figure 3.8 Matched gloss level (base synset) 35

Figure 3.9 Sample neural network input matrixes 36

Figure 3.10 Query block graphical representation 40

Figure 3.11 Calculating concept matching hits with Query elements algorithm 42

Figure 3.12 Illustration of eigenvector matrix based on a search space elements 44

Figure 3.13 Ranked Search Space centrality calculation algorithm 45

Figure 3.14 Compute Search Space Adjacency Matrix Algorithm 45

Figure 3.15 Search Space Item Iteration algorithm 48

Figure 3.16 Learner neural network structure 51

Figure 3.17 Learning operation algorithm 51

Figure 4.1 Training data set snapshot 57

Figure 4.2 Average weight tendency change between validation runs 62

List of Appendixes

Appendix I Gloss construction instructions.....74

Appendix II Training data sample.....76

Chapter 1 : Introduction

The main goal of this thesis is to develop a smart tool for concept matching to help build the Arabic Ontology. Section 1.1 introduces the scope of work and the motivation of the thesis. Section 1.2 presents the problem of this thesis and the research objectives, and section 1.3 summarizes the main contributions. The last section 1.4 of this chapter provides an overview of thesis structure and outline.

1.1 Scope and motivation

In the early stages of the internet, it was essential to develop a standard to share content and to create communication channels. Now with emerged technologies these communication channels are not enough for applications to interoperate between each other without having a predefined common understanding. Ontologies play a role in solving this problem by allowing a common understanding, where it provides possibility of knowledge sharing, reuse, interoperability, data integration, data retrieval, and search [1] [2].

Despite that, the majority of ontologies are application ontologies; which focus on a specific field of work or application's domain, a trend appeared toward building linguistic ontologies[3], mainly after the success of Princeton University WordNet project [3]. WordNet project is considered as a comprehensive linguistic ontology of

English language. WordNet resembles a thesaurus combined with a dictionary, giving a more intuitively usable resource that also supports automatic text analysis and Artificial Intelligence applications [3] [4]. Some of WordNet applications are word sense disambiguation, information retrieval, and automatic text classification. The success of WordNet yielded a motivation in the direction of similar linguistic projects (i.e. Euro WordNet), while there were few efforts toward building a comprehensive Arabic WordNet or Arabic ontology.

An initiative was established at Birziet University on 2010 to build an Arabic Ontology[6], the project's goal is to build a lexical Arabic resource that consists of a tree of concepts for all Arabic terms, with the semantic relations between these concepts, such relations are subtype of, part of, among others. Such project is considered as a big contribution for Arabic research field, as it provides a common understanding and lexical resource that leverage Arabic component in the IT field[1].

The process of constructing the Arabic Ontology involves a novel approach that is summarized into three major parts. Starting with constructing the top level Arabic concepts, which presents the core of the ontology project, the second part of the project involves a data collection and processing from the available Arabic lexical resources, where collected data are reformulated and reengineered to ensure it focus on the essential details of the concept without other, and not accidental details. [5] The third step presents a semi-automatic function that map between Arabic concepts

and its equivalent concepts in the WordNet using a concept-matching algorithm. This algorithm aims to hold several benefits to the Arabic Ontology project.

1.2 **Problem statement**

The main goal of this thesis is speeding up the process of building the Arabic Ontology, by providing smart and automated solutions, particularly in the mapping between Arabic concepts and other linguistic resources especially the English WordNet. As stated earlier, there exists a matching algorithm that maps between a set of given Arabic concepts and an existing concept in the English WordNet. The first version of this function was built by Mustafa Jarrar and three undergraduate students see [6]. In this research, we aim to reengineer and enhance the accuracy of this matching algorithm[1].

The first version of the matching function takes an Arabic concept (A concept here is described as a gloss which is an informal description of the meaning) and maps it to the English WordNet; however this function was not evaluated, and has low accuracy [7]. Moreover, its mapping operation was slow. The goal of this thesis is to

[1] By mapping a given set of Arabic concepts into the English tree, we can derive the subtype relations (the links) between these concepts. For example, if we have concepts A and B in the Arabic ontology, and concepts X, and Y from the WordNet, where there is a subtype relation between X and Y. Mapping between the Arabic concept A and the English concept X, and mapping between concepts B and Y, subsumes that there is a subtype relation between concepts A and B. Such mapping will help finding and defining the subtype relations between a given set of Arabic concepts.

reengineer and enhance the performance of this function and its internal configurable variables.

The Goal is to develop a self-learning algorithm; by enabling it to learn from its successful matching during its execution, as well as doing other improvements.

1.3 **Summary of contribution**

The main contribution of this thesis is to reengineer and enhance the existing matching function. We summarize the main contribution of this thesis of enhancing, rebuilding and redesigning the original matching function.

1. Reengineering the code: Since the existing matching function was designed in a poor and naïve way, a reengineering and redesign of this tool is very important, especially that the new design removes the hardcoded parameters, and separates it out into an external configuration file, aiming to ease the process of controlling and tuning these parameter values dynamically. Such parameters include ranking variables and their weights; such as synonyms, super-type, etc., in addition to the expansion variables that are used to construct processed data by the matching function.

2. Build a learning component: The learning component enables the matching algorithm to learn from its successful mapping, by changing and tuning values of the internal parameters during the matching process. The learning technique uses neural networks, and a statistical approach. The neural network mainly uses two activation functions[2]; sigmoid activation, and hyperbolic tangent activation function. Both functions are utilitarian for training data that produce positive numbers between zero and one, which present a percentage of each parameter processed by the neural network.

   The statistical approach trying to observe a pattern in the successfully matched concepts, and to what extent the expansion variables must be expanded or do we need these variables into the matching operation.

3. Evaluation and building a training dataset: to validate our enhancement to the matching function, we designed an experiment based on a training set that consists of thousands of tested and manually mapped concepts. The goal of this experiment is to observe the learning capabilities of our algorithm, and how the learning affects the accuracy of the matching function.

[2] mathematical functions used to scale numbers to a specific range

Our results showed that we were able to improve the overall accuracy of the algorithm approximated by a 14%, where 8% improvement was on the top 1 although this may not be seen as huge improvement but the consequence of this improvement on the ontology construction is very important.

1.4 **Summary and structural outline**

The contribution of this thesis is structured into four chapters that present and highlight our work as follows.

Chapter II Background and related work; we discuss related approach to our research. First we give an overview about WordNet. Next we preset the current state of the Arabic Ontology and the tools used to build it. In addition, we present related work to the matching function made in Birzeit University and other techniques used by various research groups such Euro WordNet, Russian WordNet, and Arabic WordNet project. The last section of this chapter introduces artificial intelligence topics that were utilized in the matching process by other researchers in similar fields.

Chapter III the matching function; Introduces machine learning to the matching function; this chapter introduces a new, revised and improved matching algorithm. First we present a detailed description of the components of the matching algorithm with its new reengineered design, and then we present the learner component, which

is the main contribution of this thesis. And last not least we discuss matching function complexity.

Chapter IV Evaluation and training; here we present how we evaluated our work, by mainly presenting the learning Data Set, and show how the learning process affected the accuracy and the performance of the matching function.

Chapter V Conclusions and future work; summarizes the main ideas of this thesis, and presents possible paths for future work.

Appendices:

Appendix 1: lists the steps required to construct a strong lexical gloss.

Appendix 2: lists the training set utilized in the research experiment

Chapter 2 : Background and Related work

This chapter gives the necessary background about some important related topics to this research, it also overviews related research and technologies to Ontology matching function. Introducing the English WordNet basics and introducing its different usage, next we presents the Arabic ontology project, which is being developed at Birzeit University and gives a background about the design of this ontology and its current state, in addition to the tools and methodologies used to build it. Afterwards we go through the related work and researches, revealing how our contribution is different from existing practice. Finally we give a background about artificial intelligence techniques and neural networks computational algorithms namely Sigmoid, and Hyperbolic Tangent activation functions.

2.1 Background of the WordNet

The WordNet is a lexical database for the most common words in English language. It groups English words into synonyms where each group of synonyms is called a synset; a synset is like a concept (i.e. meaning), and is given a synset ID (a unique identifier for each group of synonyms). Also there is a semantic relation between each synset, where the most important relation is a hyponymy (which is the subtype and super type relations between synset) [3].

Figure 2.1 presents an approximate representation of the English WordNet as a network of grouped English words under a shared concept in a rounded rectangle, as we can see in the figure, some words are duplicated under different synsets (e.g. Table) which means it is a polysemous term i.e. a term that has multiple concepts (Concepts in WordNet identify the contexts of word usage). In addition, the figure presents relations between synsets, where we have two types of relations; the one presented in the green line symbolizes the hyponym relation between synsets (called also is-a), and Meronymy relation (known as part-of).

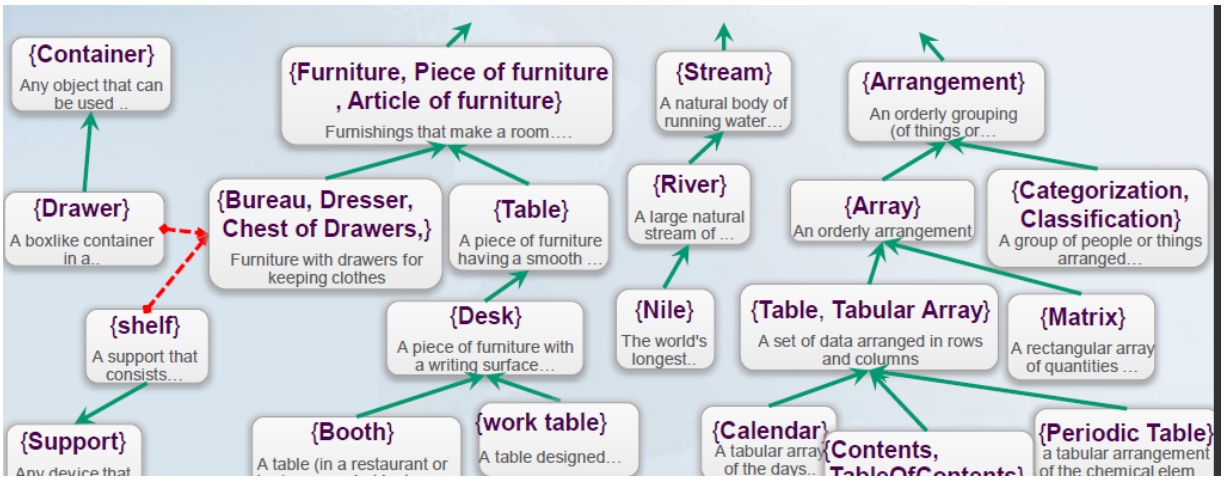

Figure 2.1 WordNet structure [8]

Hyponymy is transitive and asymmetric, and it forms a hierarchical semantic structure that inherits all the features of the more generic concept and adds at least one feature that distinguishes it from its super type. For example the word “Array” is defined as “an orderly arrangement” which is a super type of “Matrix” that is

distinguished from it by adding the rectangular feature. Meanings or concepts are organized into lexical units in the WordNet database, where words don not only represent single item but also include compounds, collections, idiomatic phrases, and phrase verbs acting as conceptual information to represent the meaning [9].

Next, we see illustration samples of WordNet relations; starting with antonym, which is a lexical relation between word-forms, not a semantic relation between meanings, e.g. ascending vs. descending. Hyponymy and hypernymy, which are also known as subtype and super type, represent relations between concepts. Meronymy described as a part of relation e.g. bone is a part of a skeleton. WordNet follows semantic networks principles where synsets represent lexicalized concepts and links represent conceptual relations [10].

To understand how English WordNet stores its data; figure 2.2 display the relational database model of the WorldNet database [11] that we transformed into MS SQL database diagram. The main tables of this model are Synsets, Words, Senses and relations. The synset table contains the synset ID, and the concept (Named glossary in the synset table). Words table contains all the English word forms (terms), which are linked to the synset table through Senses table to reflect the full structure of Synset. The last important table in this database is the relations table that includes both type of relations on word level, and synsets levels as described earlier.

The structure of the WordNet reflects both type of relations through relations table, which include word to word relations as described before and synsets semantic relations.

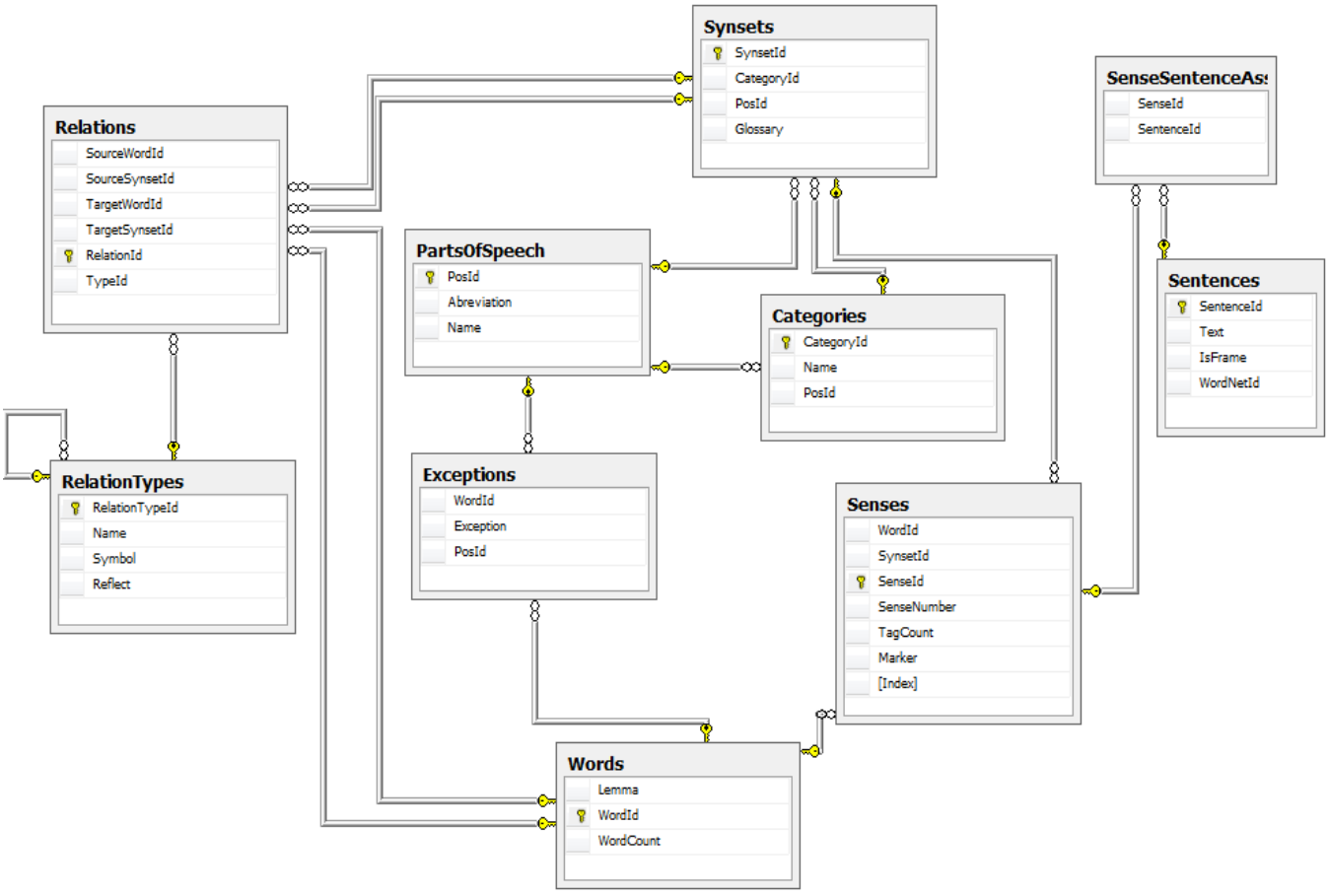

Figure 2.2 English WordNet relation database illustration

The structure and type of contents of the English WordNet made it an important resource for a wide range of applications in the Natural Language field, same as in automatic developing of lexical dictionaries. Moreover, WordNet plays a major role

in constructing other global WordNet databases or linguistic ontologies as in the Arabic Ontology project [7].

2.2 **Short Overview of Ontology**

The term Ontology is derived from the Greek word “onto” that presents a field of studying the analytical and philosophical nature of existence, or reality [2]. It was adopted and utilized in computer science after the spread of internet usage, and the increased need for data sharing in different fields e.g. e-commerce, e-governments, etc. Data sharing raised a need for system data integration to enable sharing parties to inter-communicate and perform different operations [12]; for example in the health field, health centers need to exchange patients’ medical profile over the internet, in a meaningful way using health ontology.

Data interoperability is not limited to technical issues only i.e. speed and security; it should also tackle the common understanding of transferred data, its structure, and the most crucial part; the data semantics. For instance, a health center looks into patients’ medical files with a general view while other medical centers may require detailed medical files based on the provided services. To solve such ambiguity, researchers suggested utilizing the Ontology as a reference of common understanding of data meaning [2]. Ontology helps understanding the exchanged data; since ontology is used by computerized systems – not by humans – it must be represented

in a way that systems can interoperate and process [12]. In other words, Ontology should be presented in a formal way using logic; allowing computerized systems to extract and conclude the meaning in an automated manner from the logical expression.

Most of existing ontologies are typically application ontologies, that focus on particular domains, however there is a recent interest in developing linguistic ontologies like (WordNet) for variety of tasks, including conceptual indexing, word sense disambiguation and cross-language information retrieval [13]. Application ontology is intended to represent the semantic of a certain domain or application [14]. For example, in the Palestinian e-government ontology each word convey one concept, represents application knowledge and data structure, and is used only by a certain type of applications [15]. While Linguistic Ontology is intended to represent the semantic of all words of human language, in an application independent way, and in more comprehensive view, but also focus on identifying and distinguishing multiple meaning of each word [16].

2.3 **The Arabic WordNet, and the Arabic Ontology**

English WordNet can be seen as a linguistic Ontology [3], there are also WordNet for other languages that follow the same steps as the English WordNet (i.e. French, German, Italian, Hebrew …).

There was a small effort in this direction to build an Arabic WordNet, which was a project supported and funded by the American Central Intelligence Agency (CIA) [17]. The research team in this project managed only to come up with six thousands concepts only and based on translating the English WordNet entries [13]. The used translation model cannot provide effectiveness, because conceptualization represents line of thoughts in different cultures, which cannot be matched literally between languages. For example, in Arabic language some concepts linked to regional bases such as "وضوء", or to unique cultural content such as word "لبنه" that fall between "Yogourt" and "cheese". This project was terminated without success.

Sina institute in Birzeit University started a long term project in 2010 to build an Arabic Ontology using different methodologies and techniques that bypass other projects' flaws, these methodologies are summarized below:

1. Constructing the top levels of the Arabic ontology "by identifying the top-level concepts", which form the core of the Arabic Ontology. This step is being done in close cooperation with department of philosophy as it require deep and analytical understanding of the top level, and most abstract concepts in the Arabic language.

2. The second step in the Arabic Ontology project involves mining and extrapolating Arabic concepts from available resources (around 30 thousand

concepts) dictionaries, thesaurus or any other type of linguistic resource. Following, the collected data is reformulated and restructuring into definitions that ensure focusing on the essential and distinctive qualities of the entities [5].

3. Third step involves developing an advanced algorithm to map between Arabic concepts and its equivalent concepts in the English WordNet, which allow inheriting semantic relations from the English WordNet and to ease the process of enrichment of the Arabic Ontology.

It should be noted that the Arabic Ontology is built as an ontology not as WordNet, however used as a WordNet but for Arabic. Arabic ontology is differentiated from the WordNet by using formal logic within it, which helps to construct a good formalization and philosophical foundation. In addition, Arabic ontology follows strict rules in gloss formalization.

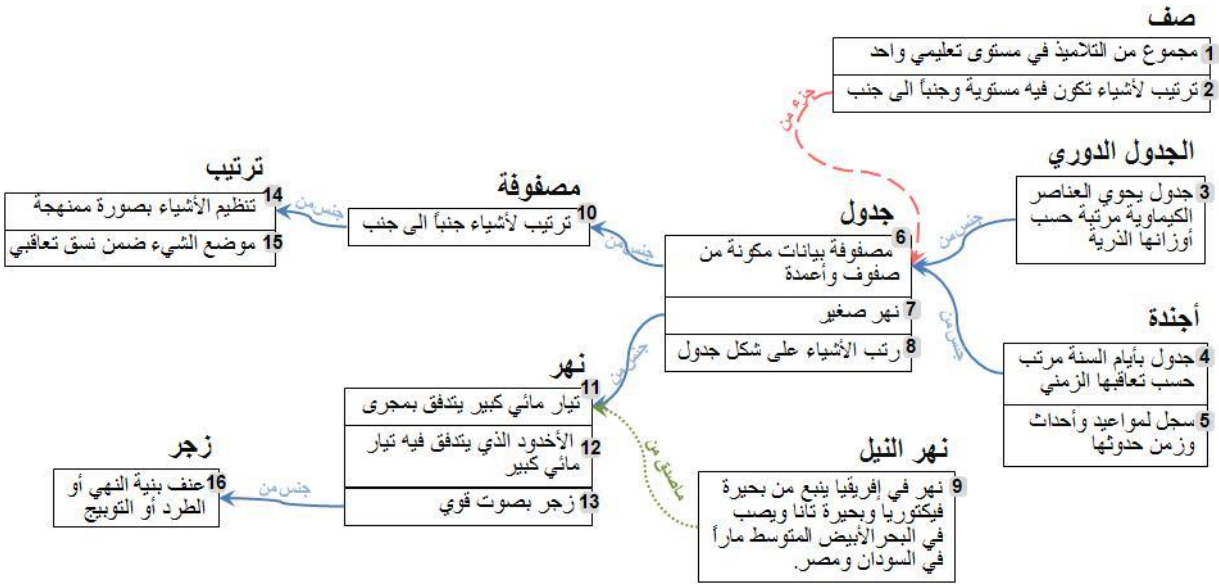

Figure 2.3 Arabic ontology structure [5]

Figure 2.3 shows an example of the Arabic Ontology structure. The sample word (نهر) has three different meanings (Glosses) associated with a unique identifier, looking at identifier number (11); this meaning is considered a sub-type of concept (7), noting that the relations here are between concepts not the words; thus such relations does not apply to other word meaning.

From the above figure we can observe other types of relations such as instance of (ما صدق من), which represents an instance of specific concept such as concept number (9). This figure is an illustrative example only, the Arabic ontology is supposed to include all concepts of each Arabic language. The results are a tree/graph of all meaning in the Arabic language.

2.4 **Concept Matching**

The matching function represents a concept-matching algorithm that links between different concepts based on analyses of the similarity between their meanings. The matching process could be between different lexical linguistic resources (Arabic Ontology and English WordNet) or within the same linguistic resource i.e. natural langue processing to find similarity between two English sentences. As shown in figure 2.4,

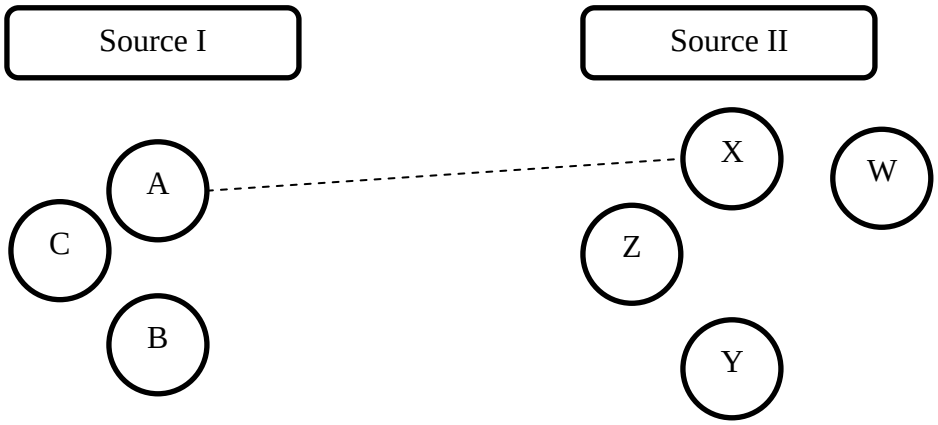

Figure 2.4 Concept matching illustration

The matching operation involves mapping different resources to each other – mainly English WordNet and targeted language ontology. In the process of constructing linguistic ontologies, different mechanisms and algorithms were developed form simple ideas like those that map based on the translated terms, or by finding the similarity between translated concepts and WordNet concepts.

To illustrate the concept-matching let us consider that we have Arabic concepts A identified by a term: جدول and gloss: مصفوف بيانات مكونة من صفوف وأعمدة. In addition, we have an entity C from an English resource, which is identified by a term: Table and gloss: a set of data arranged in rows and columns. The matching operation can be divided into different types as described in following researches.

The Russian WordNet is sample on a mapping done utilizing a bilingual dictionary. Where the research group assumes that each term in one language must hold the same conceptualization in another language [18], this mechanics done by performing mapping between A and C based on translated terms excluding their concepts. However this approach is proven as improper as explained earlier under Arabic WordNet project. Hence another approach was developed to overcome the failure of previous technique, by finding concepts similarity based on the concept meaning similarity by performing an analysis on concepts of A and C and identifying how close are the meaning of two concepts. Using a statistical approach to find percentage of similarity between matched concepts is based on the number of similar words between matched concepts.

A combination between previous techniques was introduced under the matching function, which is used in the Arabic Ontology project, this function was developed

by [6]. This function uses a bilingual dictionary to construct the search space[3], to initialize an optimized lookup space from the WordNet.

The other technique used within the matching function is finding concept similarity between concepts inside the search space and Arabic concepts. This techniques uses a list of predefined variables with given fixed weights to calculate the similarity between the matched elements based on the number of similar word-hits during the process. A hit represent a similar word found between matched concepts words, and overall value calculated by multiplying number of matched words with its corresponding predefined variables i.e. synonym or hyponym. In addition, a search engine technique [19] used to enhance achieved matching by approximating the closest match from the processed search space tree or a graph. This technique relies on centrality calculation [12].

Since the previous approaches were built using a fixed weight and fixed values, which does not offer a fair weight distribution between matching equation parameters. Moreover, it opens doors to find new mechanisms to improve the matching results based on fair weight distribution between the Matching Function variables.

[3]Set of concepts from the WordNet

2.5 **Machine Learning and Ontology Matching**

Artificial Intelligence (AI) is defined as the study and design of intelligent agents [20], where an intelligent agent is a system that perceives its environment and takes actions that maximizes its chances of success [21]. AI research is a specialized technical field, divided into subfields that are often linked to each other [22]. Main topics of AI include reasoning, knowledge, planning, learning, communication, perception and object manipulation and others.

This research uses machine learning techniques that are considered central to AI research. One of the methods we use is concerned with the design and development of algorithms that allow computers to learn from their behavior based on empirical data. Learning can take advantage of data to capture characteristics of intersect of their underlying probability [21]. The main focus of machine learning is to automatically learn to recognize complex patterns and make intelligent decisions based on data; the problem lies in the fact that the set of all possible behaviors given all possible inputs is too large to be covered by the set of training data.

Considering the challenge in ontology matching problem, several solutions introduced, which apply machine learning techniques to create a semantic mappings [23]. These algorithms can be categorized into two dimensions schema-based, and instance based matching [24]. Where the first type covers different concept features

and relations in the ontology in addition, uses some similarity measures. The second diminution compares concept instances within ontologies to find the similarity align them.

Several algorithms used different neural models, some based on unsupervised neural model like in [25], and their approach stands on classifying the ontology structure of matched ontologies against each other taxonomic structure. This type of neural networks used to find corresponds among matched ontologies attribute via categorization, and classification, such model cannot be applied in our work here since we assume the Arabic ontology taxonomy structure is not ready yet. That is, we assume only the Arabic concepts are only represented with a term and gloss (a sentence written in natural language to describe the concept), unlike others' work that are based on graph of concepts.

Other ontology matching algorithms addressed supervised learning techniques, using different approaches to match between ontologies. Some of these researches used neural network based approach such as in [26], this approach consists from two phases the first phase takes an ontology-mapped pair to identify measurement metrics under different categories. Using a feed forward network and identified metrics the algorithm defines a combined metric to be used under each category.

Another concept matching techniques presented a framework for concept similarity measures, which uses a Support Vector Machines (SVM) [27] to perform the mapping. Using four types of similarity measures word similarity[4], word list similarity[5], concept hierarchy similarity that represent the path from the root of the ontology to the concept, and structure similarity between the ontologies [28], based on the provided similarity parameters SVM decides whether concepts can be mapped.

Comparing available ontology machine based matching we observe, that most of them targeting application ontologies, or two different ontologies with known structure from both sides. Since the Arabic ontology does not have a defined structure (to the date of this research) none of these approaches can be used in our work, thus we are going into a different approach using supervised machine learning. Our assumption, that the concepts in the Arabic ontology are not connected (i.e., not a graph) brings more value and complexity to our research. By this, our approaches can be used to find similarity (and thus detect redundancies) within same ontology.

In our research we are looking for weight variables that are continuous, this lead to use a regression function using a neural network to analyze the data and recognize patterns under regression analysis. This analysis is a statistical technique for estimating, and finding the relation between a dependent variable and independent

[4] Word formation prefix, suffix, edit distance
[5] Multipart words like ‘monitoring_system’ usually used as concept label. [4]

variables. More specifically, it helps to identify and understand the typical values change of variable when other values are varied.

This statistical approaches and supervised learning mechanisms object to improve the control of matching function variables and final results as it will be explained further in the next chapters.

Comparing different approaches to design a neural network for example SVM, which is a supervised model that analyze data and recognize patterns by taking an input and predicts the output, making it a non-probabilistic binary linear classifier [27]. We decided to use a back propagation neural, which is a multilayered network and work by approximation non-linear relation between input and output by adjusting the internal weights, and such network suit our purpose in the current research to predict the best weight variables should be used in the matching operation.

Chapter 3 : Introducing Machine Learning to the Matching Function

This chapter describes the matching function components, starting with an illustration example to explain how the matching operation is achieved. In the illustration part we will go through all parts of matching function and learner component, starting with a description of the input, processing, and presenting the final results. In the second section of this chapter we go through a detailed structure description and matching function component formalization from design perspectives. The last section describes matching function complexity.

3.1 Illustrating the Matching Function by a Running Example

In this section we present the matching function using a running example. The details of the matching function will not be displayed in formal rather in an illustrative manner.

3.1.1 The Input and the Output of the Matching Process

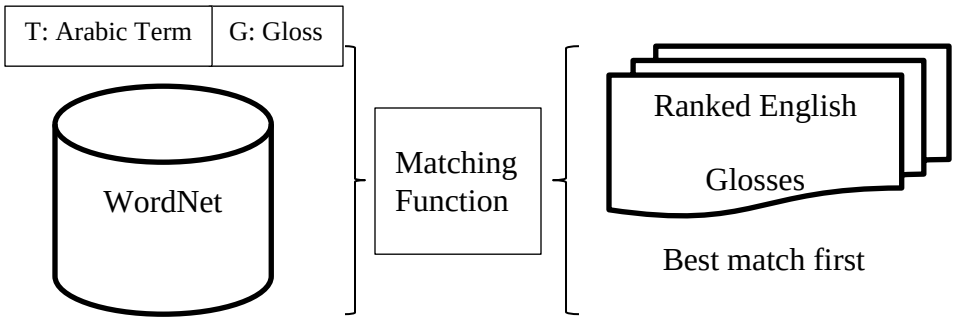

Figure 3.1 Matching function input and output

The input of the algorithm consist of two parts; a gloss in Arabic *gloss G* with a term *T* as shown in figure 3.1 (a), such input is typically taken from Arabic dictionaries and written by humans for example *G "الناس الذين يسكنون دولة ما وإن لم يكن يربط بينهم نسب حديث والعامة قد تقول شَعَب"* and *T* is *"شعب"*. The second part of the input is the WordNet itself, which consist of 117 thousand English glosses, and for each of these glosses there are one or more associated terms. The gloss represents an auxiliary informal controlled conceptualization of the projected meaning of a linguistic term, for the commonsense perception of humans [7].

Given the previous input; we expect an output as shown in figure 3.1 that shows a list of best equivalent meanings of English glosses to the Arabic gloss. These glosses are ranked on percentage of meaning similarity between the input *G* and the set of

English concepts. Typically, the top ranking in the list represent the most equivalent meaning among the WordNet concepts.

3.1.2 Preparation and Translation Phase

In this preparation phase, we take the Arabic term T and lookup its closest English terms E from existing dictionaries[6] by creating an array of English terms where each $\tau \in E$ $\{\tau\}$ for example word "شعب" expansion "people, nation, country, public, branch, rank and file, fork, and ramify". Sometimes there are few terms found for example "أَعْرَبَ" does not have direct English match, thus we use to find a related Arabic to lookup their English matches so we have more items in E. Next in the preparation phase we translate Arabic gloss G into English language G_t.Taking input previous input T and using online translation APIs such as Google APIs or Bing translate services we retrieve respectively "People who live in a country not linked by a modern proportions" and "People who live in a State that was not linked by a modern, and general rates says people" which present the base data for matching function operation.

[6] We use a comprehensive bilingual Arabic English dictionaries built by Sina institute at BirZeit University.

3.1.3 Defining the Search Space

Instead of comparing the translated gloss G_t with the 117k WordNet glosses, which is time consuming process, we need to define a search space *S'*. The search space *S'* is a set of glosses from WordNet, typically consists of several hundred or several thousands of glosses taken from the English WordNet. The search space is constructed as the following: for each term in the set τ defined earlier, we extract its synsets form the WordNet. But this might not be enough, thus we need to expand *S'* with more relevant and related glosses. We call the expanded set *S*.

The difference between S' and S is that in addition to synsets, we include the super type synsets[7], sub types[8], and synonyms[9] of each term in τ. Figure 3.2 shows the term "شعب" and it's related English terms which are constructed in the previous step; the right hand list shows the set of glosses founded for each of the English terms.

[7]L1: search space super type expansion level.
[8]L2: search space sub type expansion level.
[9]L3: search space synonyms expansion level.

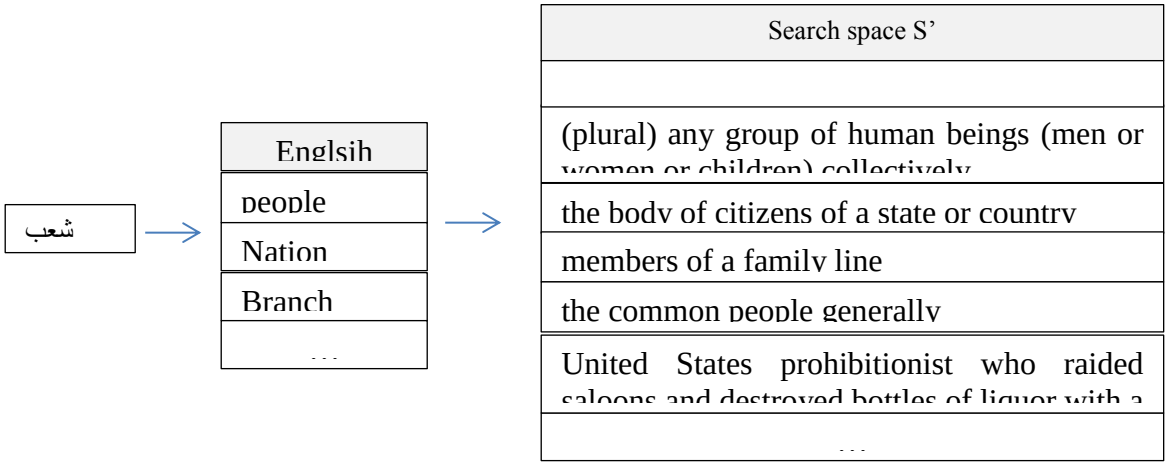

Figure 3.2 Search space overview

3.1.4 Defining the Query

Before starting matching the input (translated Arabic gloss) with each gloss in the search space, the matching function performs an additional step that involves G_t toward creating the query *Q*. The query *Q* is seen as a set, where each element in this set is a word in the translated Arabic gloss, there are called the original keywords. Seeing the words 1-m in figure 3.3 this set also includes other things super type[10], sub type[11], and synonyms[12].

[10] L4: query super type expansion level.
[11] L5: query sub type expansion level.
[12] L6: query synonym expansion level.

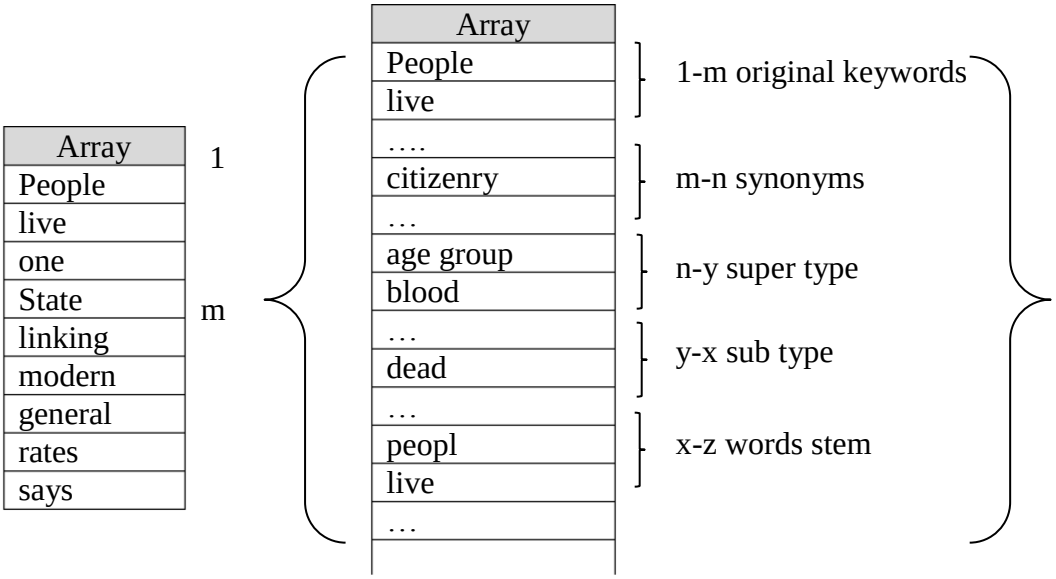

Figure 3.3 Illustration of query Q as an array

Noting that in our implementation, we used a structured approach for the query Q as shown in figure 3.4, as a part of matching function reengineering and improvements. The presented data structure of query block help us to prevent multiple hit calculations, which minimize ranking errors. Moreover current structure help us to easily inspect the exact match hit position and level. Such information considered crucial for statically data collection and identifying correct query word matching.

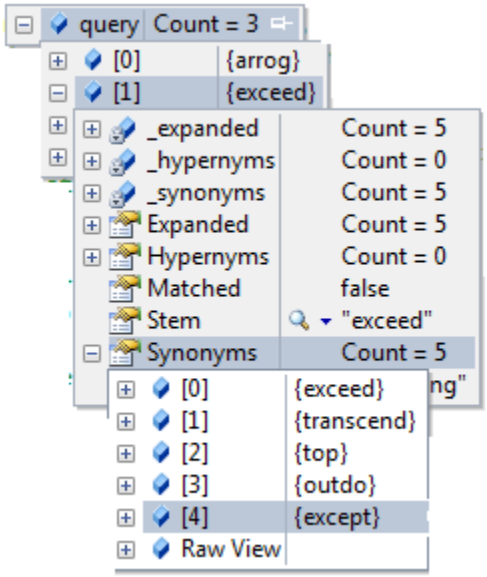

Figure 3.4 Structured illustration of the query Q

3.1.5 The Matching Process

After preparing Q as determined above, the matching operation runs against each element in the search space S, and compares elements' keywords with query elements. The final results of the matched operation gives a set of ranked English glosses; based on the number of hits found in the process as shown in the figure 3.5.

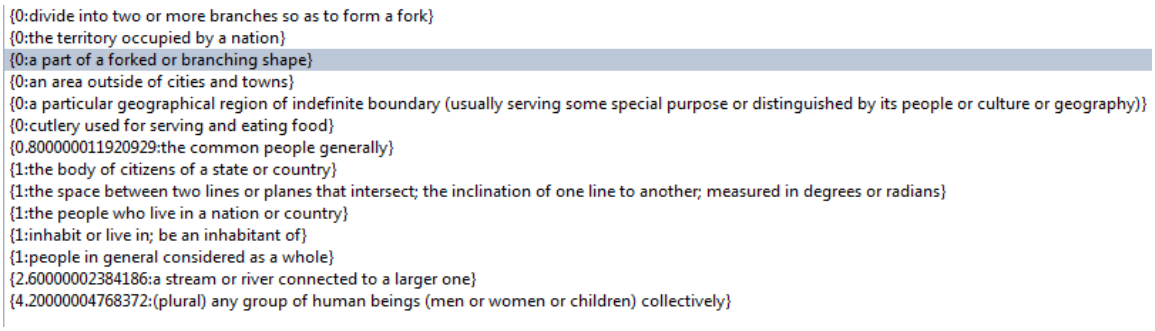

Figure 3.5 Snapshot of running Q against S concepts

Ranking operation represent the number of hits or matches on the level of four main weighing variables; w1: number of hits found on the keyword level, w2: number of hits under super type, w3: the number of hits under subtype level, and w4: the total number of hits under synonyms. In addition, a special weight variable included for successful hits done on the word level itself instead of its steam. The overall rank calculated by multiplying each weight variable with a predefined fixed value and summed to each other giving us the matching process ranking value$Rank = \sum_{i=1}^{4} w_i * v_i$.

3.1.6 Centrality Calculation

Looking at the figure 3.5, we notice matching process ranking on the left side of each gloss, also we can see that some glosses have similar ranks or too close values, thus raise a need to identify which gloss is the most closest gloss to the matched gloss. Since WordNet is a tree/graph concepts, the generated search space S typically will be a tree or collection of trees, see figure 3.6 as an example. This makes us reconsider the resulted ranks since we are talking about graphs, where a node with the highest rank does not necessary mean it is the closest equivalent meaning as shown in figure 3.6. To resolve this issue the matching function employs the eigenvector Centrality that is used in the same manner in search engines to find the most central or closest element within a scattered tree of elements [19].

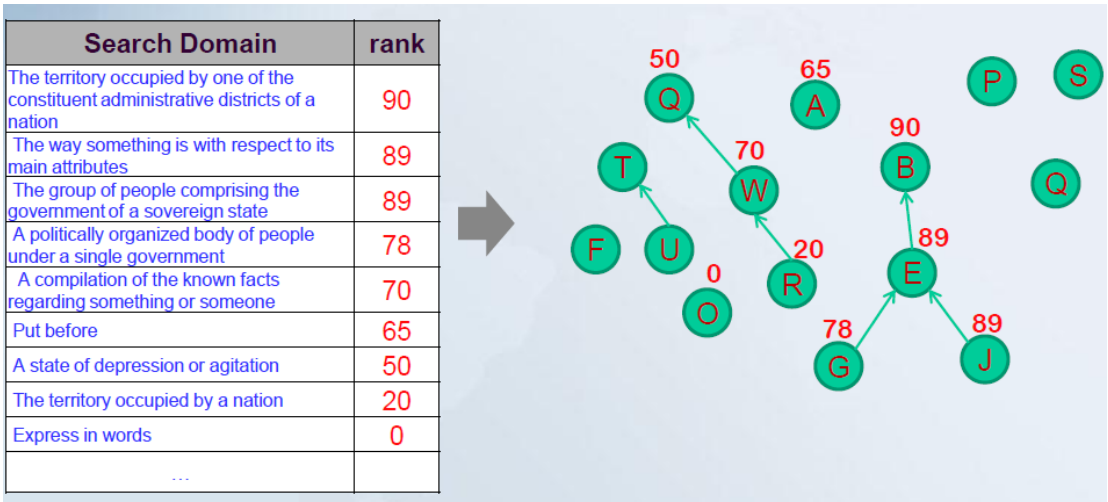

Figure 3.6 Illustration of centrality calculation [8]

Under the centrality calculations, the scattered collection of trees transformed into a square matrix[13] of ranks, where first element under each matrix row represents a search space element rank, and the remaining row elements filled with item Si related synsets weight. If the related synset exist in the search space we use its rank or one for zero ranked elements, and a zero value used for none search space elements

By running the Eigen centrality against the generated matrix we find the most affecting points in matrix graph, which represent the closest result to the best match. Referring to the figure 3.5 shows a group of possible matched glosses which is inconsistent to input, however the centrality changes this results as shown in figure 3.7 where this ambiguity are resolved.

[13] Number of rows and columns equal to number of elements in the search space S

{100:(plural) any group of human beings (men or women or children) collectively}
{77.3:the people who live in a nation or country}
{77.3:people in general considered as a whole}
{68.77:any number of entities (members) considered as a unit}
{12.66:the body of citizens of a state or country}
{12.21:the common people generally}
{0:the angle formed by the inner sides of the legs where they join the human trunk}
{0:place under attack with one's own pieces, of two enemy pieces}
{0:shape like a fork}
{0:have or develop complicating consequences}
{0:people descended from a common ancestor}
{0:fill with inhabitants}
{0:a unit with political responsibilities}

Figure 3.7 Search space(S) results after running the centrality

In the summary of the matching function process, we find that output set consist from English WordNet glosses ranked from highest match to the lowest match. Even though the highest ranked gloss in figure 3.7 close to the Arabic gloss, however the exact match had lower rank making it fall as the second choice for the matching operation. This leads us to the next component of the matching function components.

3.1.7 The Learner

In the described previously matching algorithm we observe it is dependency on a fixed weight variables (w1-w5), and predefined expansion variables for both Search space S and query Q, the effect of these variables can observed from previous section results where the best match didn’t came first. The learner component tries to answer what is the best weight value that must be associated to parameters w1 – w5? And what is the best value for expansion level? The learner is a program that starts from

chosen arbitrary values of previous parameters; it keeps changing and tuning these parameters to achieve better results.

Changes on the parameters are performed based on successful matches achieved in the experiment, which is affected by a statistical approach and supervised learning that utilizes neural networks to achieve the desired goal. To explain how the learner functions, suppose that we have successful matches from table 3.1, which are the main inputs of the learner component.

Table 3-1Learning item sample

Term	**Gloss**	**Matched Concept**
شعب	الناس الذين يسكنون دولة ما وإن لم يكن يربط بينهم نسب حديث والعامة قد تقول شَعَب	The people who live in a nation or country

In the learning process the learner performs one gigantic step that was described in the regular matching operation in sections 3.1.1, 3.1.2, 3.13, and 3.1.3. The aim of this operation is to collect statistics about the matching operation as shown in table 3.2. That presents general information about hits between “inputs” matched concept and Q of provided Arabic gloss.

Table 3-2 Gloss (G) matching operation hit against matched gloss (M)

Parameter	**Number of hits**
Word Form	0
Synonym	2
Keywords (Definition words)	0
Hypernym	0
Hyponym	0

Next, the learning operations answers our previously asked questions related to P variables. P parameters values change based on the observation done in the expansion of matched glosses as shown in figure 3.8 where our example is found directly under the base synsets of search space.

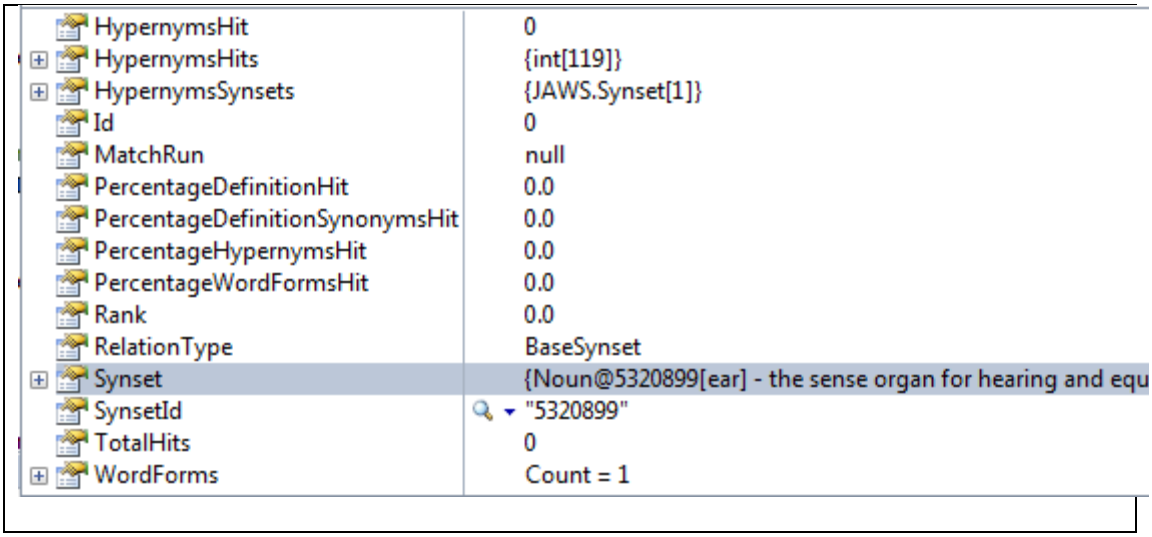

Figure 3.8 Matched gloss level (base synset)

The most important step of the learner component involves a neural network, to tune and balance W parameter values based the information collected from analyzing successfully mapped concepts. The back-propagation neural network is constructed from three layers[14] where the first layer uses a Sigmund activation function and the second one uses hyperbolic tangent activation function to tune the w values. Both functions are considered as best solution for data learning technique as both of them generate positive values (skipping the negative part of the hyperbolic tangent) that

[14]The neural network that process the input matrixes consist from three layers, where both the input and the output layers consist from four neurons (to correspond to number columns in the matrix). The hidden layer consists of seven neurons (number of input matrix column multiplied by two minus one).

ranges between the 0 and 1 that could be interpreted as positive weight values. Figure 3.9 presents a sample on how the arbitrary initial weights in part (a) changed to the values presented under part (b) of the same figure.

a) The ideal weight matrix b) System weight matrix

Figure 3.9 Sample neural network input matrixes

Table 3.3 represents the final output of the processed example, where we can notice the processed example affected with a small change toward better match in the next operations. Hence, we did several tests over the weight elements by fixing parts of the original matrix and changing others to observe the effect of the exact changes over system variables. The results showed that the least affecting element in the equation is the subtype element (more details will be discussed in chapter 4).

Table 3-3 Neural network results of processing the example results

	Keyword	**SuperType**	**Synonym**	**WordForm**
Used weight	1	0.4	0.3	0.3
Ideal Weight	0	0	1	0
Result weight	0.9806	0.4001	0.2999	0.2999

3.2 **Structural Design**

This section describes the internal component of the matching function from the technical point of view, which adds a detailed description to the previous section.

3.2.1 Core component

In the previous section, we described the matching function and how it includes four main items; the search space *S*, query *Q*, ranking, and the centrality calculation.

3.2.1.1 Build the Search Space (S)

The search space *S* as described previously is a subset of the set of all glosses in the WordNet (around 117,000 concepts). Our goal is to have a minimum number of glosses in *S*, but on the same time make sure that the equivalent (which is our target) gloss is included in *S*. The following are the steps we use to building *S*:

1. First, we find the synonyms of the input Arabic term *T* using an existing Arabic dictionary[15], for example if *T* = "أذن" the set of results of Arabic synonyms will be {سنابل, إذن, سنبلة, حاسةالسمع, سبلة, شيءمماثللأذن, سنبل, كوز} we call this set *T'*.

[15] This Arabic dictionary is a huge Arabic – Arabic thesaurus built at Sina Institute in Birzeit University.

2. Using a bilingual dictionary Arabic – English dictionary we translate all *T'* elements creating a new set of English words called *B*. This set might consist from one to hundreds of English terms.

3. To constructing the search space *S'*, we take each English term in *B* and extract its glosses from the WordNet. We call the set of glosses of all terms in *B* a space baseS_b.

4. To maximize the chance that the target gloss is included in S_b we use WordNet semantic relations to extendS_b, by including synonyms, subtypes, and super types of each element inside set S_b. The expansion depth identified based on matching function configuration parameters. In other words, the question is how many levels of synonyms, subtypes, and super-types should we consider when expandingS_b?! Having more levels might disturb the accuracy of the results, by having fewer levels; we may miss the target gloss. However, we will show later how our learning component searches for the suitable levels. Nevertheless we call the expanded S_b as *S'* search space.

3.2.1.2 Build Query Block (Q)

The query *Q* is a tree of words extracted from the input gloss *G*. The query block constructed as follows:

1. The input gloss translated from Arabic to English using translation services for example *G* "في الحيوان عضو السمع" , become *Gt* "In animals: a member of hearing"
2. After cleaning none primary words from *Gt* by keeping the keywords this present the base information of *Q* for example the keywords of *Gt* are animals, member, and hearing.

3. Utilizing WordNet APIs we expand the query block *Q* with synonym, hypernym, and hyponyms based on the existing semantic relations within the WordNet as shown in figure 3.2. The expansion involves extracting the word forms of different synsets linked to keyword concepts under different semantic relations.

Figure 3.10 illustrate the query block structure[16] where each query element[17]hosts its own expanded elements for example keyword "member" has its stem "member" and

[16] Improved queries block structure that minimize matching errors, and ease the process of matching operation by minimizing the multiple hits count and other traditional array limitations. Moreover, this structure eases the tracking of matched Q elements, and ranking calculation under each individual element.

an expansions of 4 synonyms i.e. term phallus, 52 expanded super-type terms, and 56 subtype terms from the WordNet.

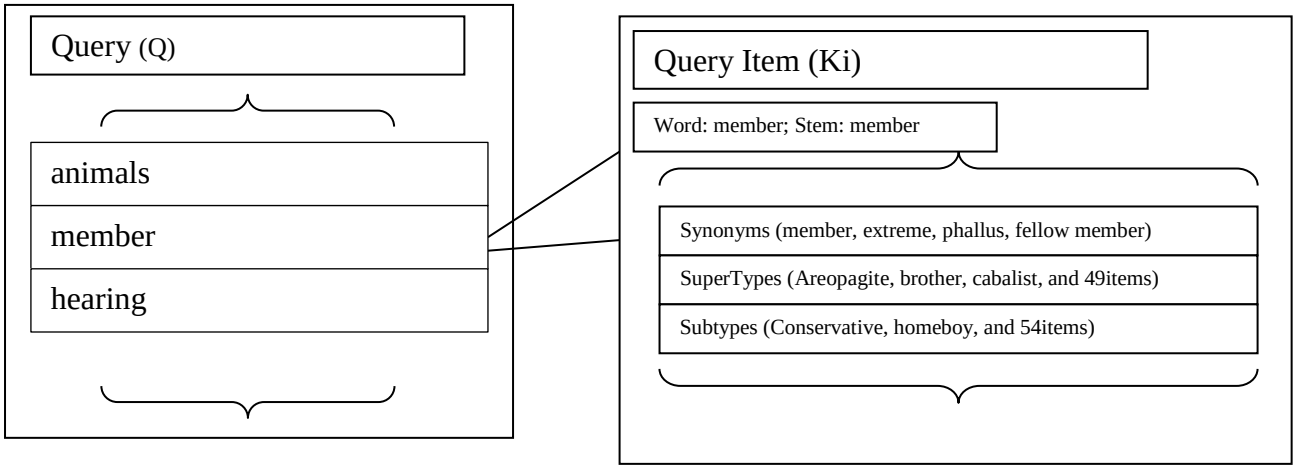

Figure 3.10 Query block graphical representation

3.2.1.3 Matching Process and Ranking

In this step, we rank every gloss in the search space *S*, by comparing its words with every element/keyword in our query *Q*. That's, we want to find the target gloss, which has the highest rank. The ranking operation calculates number of hits/matches between query block *Q* elements and search space *S* elements – a hit occur when a word or stemmed word of *Q* match a word or stemmed word of S as shown in Table 3.11lines3 to 11.

[17] Query element represents the keyword, and its stem.

Following hit calculation, the algorithm performs a mathematical equation to identify the rank of matched element as shown here $Si\ Rank = KeywordHits * w_1 + SynoymHits * w_2 + SuperTypeHits * w_3 + SubTypeHits * w_4$, where w_i is the weight for the hit i. Weight variables present a decimal digit that illustrates the overall effect of specific variable in the gloss rank. Weight variables divided into four main parts keyword, synonym, super-type, and subtype. To reflect the exact effect of each matched element between the glosses and query *G*, the matching function calculates number matched words under each category which shows number of words matched for example under super-type.

Next after finding total words matched under each specific category each value multiplied with its corresponding weight variable and summed to represent the final rank of matched gloss.

Calculate Query element hits within a concept algorithm	
Input:	Concept C, Query Q
Output:	Ranked Concept(Cr)

1. For each query element Qi in Query Q
2. IF Qi Found in Cr Keyword elements Then
Update Cr Keyword hit and mark Qi as matched
3. End IF
4. IF Qi Not found AND Qi Found in Cr Keyword synonyms Then
Update Cr Synonyms hit and mark Qi as matched
5. END IF
6. IF Qi Not Found and Qi Found in Cr Keywords super types Then
Update Cr Super type hits and mark Qi as matched
7. End IF
8. IF Qi Not Found AND Qi Found in Cr Keywords sub types Then
Update Cr Sub types hits and mark Qi as Matched
9. End IF
10. IF Qi Not foud Then

Run Ranking on Qi Subelements

11. **END IF**
12. **End FOR**
13. **Cr Rank = KeywordHits * KeywordWeight + SynoymHits * SynonymWeight + SuperTypeHits * SuperTypeWeight + SubTypeHits * SubTypeWeight;**

Figure 3.11 Calculating concept matching hits with Query elements algorithm

3.2.1.4 Centrality Calculation

There are many times when matching outputs found carrying close ranks for example 89% and 90%, or in similar ranks as shown in figure 3.7. Sum situation create confusion especially for an automated solution, where not necessary the highest ranked element mean it is the equivalent meaning between available close ranked concepts.

Matching operation occur over a linguistic resource "WordNet" that presents a tree/graph of interlinked linguistic concepts. Concluding by that, the search space *S* represents a collection of trees/graphs that can be presented as in figure 3.6. Since S presents a collection of trees/graphs we found our self within scopes of graph theory[18], in our case we need to find the most influencing node in the search space *S* that involves a mathematical calculation called eigenvector[19] centrality, such

[18] Graph theory in math and computer since is the study of graph, which is a mathematical structure used to modal pairwise relations between object.

[19] Eigenvector centrality is a measure of the influence of a node in a network. It assigns relative ranks to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the nodes.

technique used also by search engines to find most relevant result in the network of linked networks [19].

Figure 3.14 represents eigenvector centrality calculation performed by matching operation toward resolve and finds the equivalent meaning inside the search space *S*. In purpose to perform the required calculations, the algorithm converts the scattered search space ranked concepts into a matrix. In reference to the example in figure 3.7 each single row of the generated matrix presents search space elements *Si* including its adjacent synsets from the WordNet, where *Si* ranks presents the diagonal of the column of the matrix rows and the remaining columns takes either synset rank if it is part of search space[20], or zero for none search space synsets figure 3.12 illustrate a matrix based on referenced example from figure 3.6.

[20] A value of 1 placed in the matrix for synsets that found in the search space S with a zero rank, while for diagonal matrix elements it takes the original rank.

Columns

Rows

4.2	0	0	0	0	0	0	1	– · –	0
0	0	0	0	0	0	0	0	– · –	0
0	0	0	0	0	0	0	0	– · –	0
0	0	0	0.8	0	0	0	0	– · –	0
0	0	0	0	0	0	0	0	– · –	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮		⋮
0	0	0	0	0	0	0	0	– · –	0

Figure 3.12 Illustration of eigenvector matrix based on a search space elements

Following creation the matrix that represent S ranks the system iterate through its element applying the following equation $Si = \lambda x$, where λ represent a constant and x matrix value at that point; figure 3.13 presents a pseudo code of centrality calculation algorithm. The λ constant calculated based on the dot product of matrix x into a vector matrix generated based on the multiplication of randomly generated victor and x diagonal elements victor as shown in figure 3.14.

The resulted matrix resulted vector contains the strength or closeness of diagonal search space elements to the matched concepts as showed earlier in figure 3.7, where the confusion removed between close ranked concept and gloss “(plural) any group of human beings (men or women or children) collectively” shown to be the best equivalent to the input Arabic Gloss.

Calculate Search Space Centrality Algorithm	
Input:	Search Space S
Output:	Most contributing concepts to the graph score (S')

```
1. Adjacent Matrix [,] = Compute Adjacency Matrix For (S)
2. newRanks [] = Power Iteration (Adjacent Matrix)
   maxRank = 0;
3. For each rank in newRanks
4. maxRank = max(maxRank, rank)
5. End For
   For each rank in newRanks
      Si rank = (rank / maxRank)* 100.0
6. End For
```

Figure 3.13 Ranked Search Space centrality calculation algorithm

Compute Adjacency Matrix Algorithm	
Input:	Search Space S
Output:	Adjacent Matrix A

```
1.  Initiate squared array A with number of elements in S ^ 2
2.  For each item S(i) h in S
3.      Get All Related Concepts to S(i) As AdjC
4.      For j less than matrix length
5.          Get S(j) From S
6.          If i equals j Then
7.              Set A[i, j] to S(i) Rank
8.          Else If( AdjC Contains S(j)) Then
9.              If S(j) rank equals to zero Then
10.                 Set A[i, j] to one
11.             Else
12.                 Set A[I,J] to S(j) rank
13.             End If
14.         Else
15.             Set A[i, j] to zero
16.         End If
17.     End For j
18. End For i
```

Figure 3.14 Compute Search Space Adjacency Matrix Algorithm

3.2.2 The Learner

Matching function is a concept similarity method that match two concepts from two different languages (Arabic and English) by measuring the degree similarity between

matched elements, where similarity based on number of matched words between the translated gloss and target gloss (English WordNet concept).

This operation involves finding number of matched words under a specific sematic relation type, where a weight given for each matched element and the overall value indicates the similarity rank between matched concepts. Though this technique uses a predefined fixed value it gave a satisfactory results for a certain level, however this results seem insufficient and considered limited since ranking depends on a fixed parameters and weight values.

In our solution we introduce the learner component, which is goal to improve matching output by controlling and modifying matching internal configurations. The modification done through an observation and learning successfully matched elements. In order to operate a successful learning operation we had to identify all of the involved parameters and weighting variables, which may affect ranking operation. Since the matching operation involves linguistic resources with semantic relations, which allows us to identify main ranking variables in following categories: original words “keywords”, synonyms, Hypernym “super-type”, hyponym “subtypes”, in addition words stem “word definition”, where these variables presents main weight variables in the ranking process.

In addition, there are additional parameters that affects matching operation not the ranking it self but it affects matching operation parts search space *S* and the query block *Q*. These parameters are called expansion parameters since they affect the depth of search space and number of concepts it include, and the depth matching done over the query block. Expansion variables have following categories: Synonym, super-types, and sub-types.

The learner component tries to answer what is the best weight variables and how each category affect the overall rank of matched concepts, and to observe what is the best expansion variables to be used based on observing successfully matched concepts.

3.2.2.1 Identify Search Space levels

The search space *S*is one of the important factors in the matching process; especially that its content affects the final outputs. Whenever there is any problem in the content of the search space it leads into an inaccurate matching results or no match at all, for that reason we decided to find the maximum possible expansion over various expansion groups, which is built based on search space base S_b content.

Expansion variables are identified under following categories synonyms that contains all concepts with the same meaning, super-type, and subtype relations. These expansion levels were selected as it contains either similar meaning concepts, more

generalized meaning, or more specialized meaning to the lookup terms based on what search space *S* built.

In figure 3.15 shows the detailed description of identifying search space expansion levels under each of mentioned category. The identification process done by looking for manually matched gloss in the WordNet and the find its relation to Arabic term related concepts, and based on observation and statics of how common match found selected category an update happens to the search space element, this update ensure in the coming matching operation the equivalent search space will be included.

Find Best Search Expansion Levels	
Input:	Source Language Term T, Matched gloss M
Output:	Level of the match occurrence L, and at what relation type match found R

```
1.  spaceBase = dictionary lookup results for T, including T
2.  transaltedSpaceBase = translation of spaceBase items
3.  For level = 0 and less than maximum allowed level
4.      For each item in translatedSpaceBase
5.          IF M found in item related concepts at level value
6.              Set L equal to Max(level value, realtedLevel)
7.              Set R equal to related
8.          Else IF M found in item synonym concepts at level value
9.              Set L equal to Max(level value, synoymLevel)
10.             Set R equal to synonym
11.         Else IF M found in item SuperType concepts at level value
12.             Set L equal to Max(level value, superTypeLevel)
13.             Set R equal to SuperType
14.         Else IF M found in item Sub Type Concepts at level value
15.             Set L equal to Max(level value, subTypeLevel)
16.             Set R equal to SubType
17.         End IF
18.     End For Each
19. End For level
```

Figure 3.15 Search Space Item Iteration algorithm

3.2.2.2 Identify query block and concept matching expansion level

The query block Q that build during matching operation plays a major part in the matching ranking process, since it controls the possible words to use in the similarity matching operation. Same categories as search space S expansion categories identified that include only related meaning only. The only difference between search space expansion and query block expansion is here we only take the word forms under each category of semantic relation while in the search space we take the concepts.

Query block *Q* expansion variables are managed based on an observation made from the successful match between input gloss, and manually mapped gloss. The information collected on the level of each expansion category (synonym, super-type, and subtype), where the learner looks at what is the trending depth of matched words between query elements and equivalent gloss. This value will be reflected into matching function configurations for example the most trending level for synonyms expansion depth is 3, while for super-types is 2.

3.2.2.3 Tuning weight variables

Weighting variables of the matching process as described before are keyword weight, synonym weight, super-type weight, subtype weight, and stem weight. These

variables play a major role in similarity matching between matched concepts. In reference to figure 3.3 that illustrate the query block we can observe the division of mentioned categories, where each matched word under a category increase the hit count associated to a specific category i.e. matched 3 keywords , 1 super-type, 2 subtype, and 0 stem. Each weighting variable then multiplied with its corresponding hit count, and collectively presents gloss rank.

Any incorrect weight value allocation to any of the categories will turn the results into the incorrect direction, for example if a small weight set for the keyword weight and higher weight for the super type it will devote results into a more general meaning. Thus we need a smart solution that will help to distribute the correct weights between these categories to achieve better matching results.

Since we try to predict the best weight distribution which is a a group of attributes with a continuous value, this lead us for using a regression supervised learning based on a neural networks, which is going to resolve current issue by learning from a will defined training set, and successful mapping results. In purpose of solving this problem we designed a neural network that consist from three layers where first layer consist form four input neurons, 7 hidden neurons to and four output neurons that reflects output categories figure 3.16.

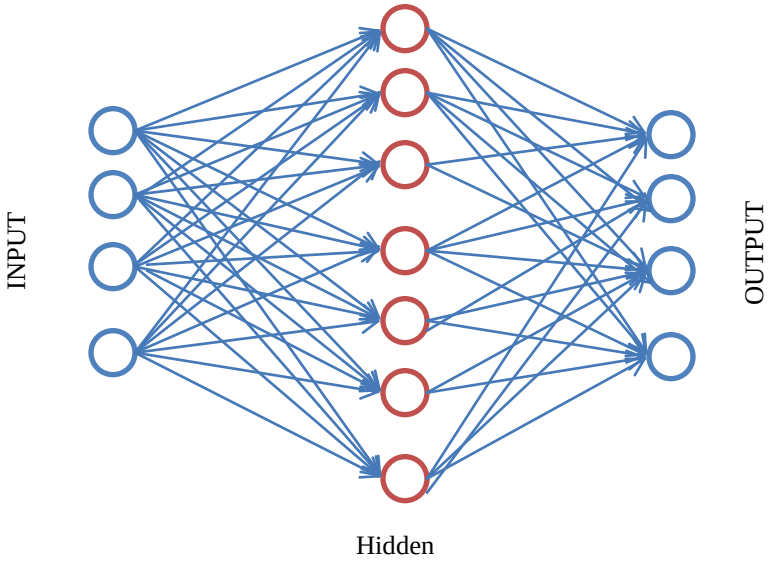

Figure 3.16 Learner neural network structure

.

The implemented solution starts with initial random small weights values near zero as a startup stage, then following we used a forward propagation against input vector to find the heuristic function H. Using function H we compute the cost function J that used by back propagation algorithm using both hyperbolic activation function on the input layer, and sigmoid activation function to iterate through input weight variables.

Learning system weights operations algorithm	
Input:	Ideal weight matrix [Keyword Weight, Synonym Weight, SuperType Weight, SubType Weight], I
Output:	Results matrix R[Keyword Weight, Synonym Weight, SuperType Weight, SubType Weight]
1. **Initialize initial matrix X;** 2. **IF first run fill X with Random values otherwise use define system variables** 3. **Initialize network layers (input layer with 4 neurons, internal layer with 7 (4 input variable + 3 variants) neurons, out layer with 4 neurons) Set learning mechanism to Resilient propagation with input I and X** 4. **Do network learning until error range reach to defined acceptable error rate** 5. **Set R with network processed results**	

Figure 3.17 Learning operation algorithm

Followed this operation we use a numerical method of checking derivate *J'* that computed by previous step and compare it to *J* until we had a successful difference less than 1 percent, where we stop back propagation to perform an optimization. The optimization purpose is to minimalize *J* using gradient optimization function since we are dealing with a linear equation, by completing the operation we reach to the best possible weight values to use it within the matching function pseudo code available in Figure 3.17.

3.3 **Algorithm Complexity**

This section discusses the complexity of the matching function, namely its scalability, covering space, time, and memory complexity. As the function consist of several components, the following sub section present each component's complexity individually, and then identify the overall complexity.

3.3.1 The complexity of core component

The matching function's complexity defined by a group of internal procedures that can be simplified into four major equations f1, f2, f3, and f4. f1 represents the complexity of building the search space S, f2 shows the complexity for query block construction. While f3 and f4 represent the mathematical calculation matching process hit calculation, and centrality calculation respectively.

The time complexity of f1 represented as follows $f_1 = 4t + 3ts_1 + l_1(6 + 2s_2) + l_2(6 + 2s_3) \in \Theta(\max(ts1, l1s2, l2, s3))$ where t represent number of words translated, $s1, s2, and\ s3$ represent expanded search space elements synonyms, super type and subtype synsets; l presents the possible expansion levels of each type. The worst case is to build a search space reached when number of expansion levels are too height that may return the whole WordNet synsets.

The complexity of the query-builder function f2 is calculated using $f_2 = 5 + 13wn + 5wl_3 + 5wl_4 \in \Theta(\max(wn, wl_3, wl_4))$ where w represents the number of translated gloss keywords, and l represent expansion variables under super type, and subtype relations. The n variable represents number of word synonyms. In the query block the worst case occurs when the expansion levels are too high and number of keywords too that may result in retrieving all of the word forms from the WordNet database.

The hit count calculation function f_3 represents an improved version compared to the original function where its complexity is $\Theta = \text{nm}$; where n is the number of query words and m is the number of words forms under specific relation type. In our case we used a depth first search which has $\Theta = \text{l}$, where l is the longest path to find the matched word.

The centrality calculation f_4 represents a high risk function it generates large set of matrixes that equal twice the screech space S size, which is in the worst case twice the number of WordNet synsets, the complexity of this function calculated as $f_4 = 8 + 13n + 7n^2 \in \Theta(n^2)$, where n represents the number of matched synsets.

In short, in this section we can observe that the matching operation space and time affected by number of elements in the search space S which increase the time and space required in the centrality calculation.

3.3.2 The complexity of the learner

As mentioned before the learner component realties on two major parts: the matching operation from the core component excluding the centrality in the learning process. Since the analyses done between the matched gloss and Arabic gloss, these analyses performed by three functions that we described earlier, f_1 that identifies the search space expansion levels, f_2 performs the lookup for query expansion levels, and the weight tuning function f_3.

Both $f_1$and f_2 uses the same approach to find the expansion levels but on different datasets, their function complexity presented as follows $f_2 = l + 20lt \in \Theta(lt)$ where l represent the maximum limit the expansion can be reached, and t is the

number of items either words or glosses retrieved from WordNet. The worst possible case us limited to the number WordNet concepts.

The weight tuning operation complexity can be identified simply as follows $f_3 = 14 + 3n + 3e$ where n represents the number of training elements that been converted into a matrix form. While e represents the cost value that is calculated based on the sigmoid function that ranges between zero and one. Space complexity of the learning operation increases when the number of training elements are high - where we have two 1x4 matrixes for each training element.

We can summarize the complexity of the matching function is affected by the length of the Arabic gloss and the number of related synsets linked to Arabic term English translation. In the worst cases the matching function is limited to the size of the WordNet content.

Chapter 4 Evaluation

4.1 Introduction

This chapter explains our evaluation which we did to examine our work in improving the Matching Function Tool. Particularly we discuss the means by which we did an assessment to our work, in addition to provide clarifying examples. The evaluation process goes through evaluating both parts/components of the Matching Tool starting with the reengineered version of the matching function, and the learner component as explained later.

4.2 Data set

The data set that was utilized in the evaluation process was obtained from Sina Institute, which is part of the manual work done by the institute toward enrichment of the Arabic Ontology. The data set is a set of Arabic concepts extracted from different resources i.e. Arabic dictionaries, Thesauruses, and other sources.

Our data set consist from a 975 concepts that were verified and validated manually to meet minimum acceptable number of conditions to meet the standard gloss structure[21]. These Arabic concepts (term, and its gloss) were manually mapped to

[21] Detailed description found in Appendix I.

their most close equivalent concept in the English WordNet. In addition to the standard mentioned inputs (Arabic Term, Arabic Gloss, and English gloss) the training set contains additional items as shown in figure 4.1.

The first item of information is the Concept Id which is a unique identifier associated to the Arabic that comes from the Arabic Ontology database, this identifier used to create a link between mapped concepts. The matched percentage represents a satisfactory ranking on the manually matched elements, which is informative information only that could to be compared to automated ranking results. The last additional information is the English synset identifier extracted from the WordNet, this identifier used by the learner component to extract WordNet Synset directly to be process during the learning process.

	A	B	C	D	E	F
1	Concept_ID	Arabic Term	Arabic Concept	Match Percentage	English Term	English Concept
2	9	كَلْب	حَيَوان مِنْ فَصيلَة الكَلْبِيَّات من الحَيَواري، يوجد منه أنواع كثيرة بعضها يستخدمه البشر مثل كلب الصَيْد وكلب الحِرَاسَة وغيرها	100	Dog	A member of the genus Canis (probab descended from the common wolf) tha has been domesticated by man since prehistoric times; occurs in many bree
3	27	عَرَبِيّ	البَحْر العَرَبِيّ هو البَحر الذي تطل عليه اليَمن وامتداده هو المحيط الهندي	100	Arabian Sea	A northwestern arm of the Indian Ocea between India and Arabia
4	36	أَبُو	الأَبّ الرّوحي هو من يقوم مقام الأَبّ في النّصيحة والقدوة لشخص بالغ إستخدام حديث	100	Godfather	Someone having a relation analogous to that of a male sponsor to his godchi
5	54	طَرَقَ	يحمل بالاجمال معنى الضرب على الشيء، كالحديد او الباب، باداة او باليد، او تحمل معنى التتابع وغيرها	100	Knock	Make light, repeated taps on a surface
6	63	دَقّ	الشيءُ صار صغيرًا جدًا	100	Reduce	Reduce in size; reduce physically
7	83	بَشِع	الدميم الذي لا تستسيغه العيون مجاز	100	Ugly	Displeasing to the senses
8	110	أَكَلَ	أَفْنَى الشيء ولم يبق منه مجاز	100	Eat Up	Use (resources or materials) over time in order to function
9	129	طَائِرَة	الطّائرة الورقية هي الطّيّارة التي يلعب بها الصبيان، عودين من خشب يربطان متعامدين يوضع عليهما الورق أو القماش أو غيره مما خف وزنه ويربط بخيط ثم يترك ليطير في الهواء	100	Kite	Plaything consisting of a light frame covered with tissue paper; flown in win at end of a string

Figure 4.1 Training data set snapshot

4.3 **Evaluation setup**

To evaluate the advance matching function, we conducted 3-level cross validation [29] to measure the accuracy of matching process. To evaluate this, we did the matching using old version function of the function (without learning), where we can identify the exact improvement over the matched results.

The accuracy of the function is defined as the following: given an Arabic Concept, and given the English concepts in WordNet, these English concepts are ranked according to their meaningful-closeness to the Arabic concept, the accuracy of the function is then measured based on what is the probability for the matching function to give the highest rank to be really the meaningfully-equivalent to the Arabic concept.

The accuracy levels are divided into five groups; the Top 1 accuracy represents concepts where the exact equivalent match has the highest rank in the search space, the Top 5 accuracy represents the concepts where the exact matched concept exist within top five ranked concepts. Top 10 denotes that the equivalent matched concept found between the top ten highest ranked concepts in the search space, while Top 15 indicates that the equivalent matching concepts exits within top fifteen highest ranked concepts from the search space concepts.

To perform the 3-corss validation process the training dataset was divided into three groups A, B, and C – each group contains around 325 Arabic glosses. Flowing we designate each cross validation run with a training dataset and test dataset as show in table 4.1, where each run involves the following exercises (1) running the matching function without learning against the test set, (2) training the matching function using a training set, and last (3) running the matching function with learning again against the same training set.

Table 4-1 Cross validation dataset distribution

Cross Validation Run	Training Dataset	Testing Dataset
1	B, C	A
2	A, C	B
3	A, B	C

In each scenario of the 3-cross validation, the neural network and matching function re-configured to the initial values shown in table 4.2, which allows to observe the behavior of the matching function with learning component under different training set, and testing set. In addition to compare matching results with matching function without learning component that works based on the initial configurations by default.

Table 4-2 Initial Matching Function configured variables

Variable	Value
Keyword weight	1
Synonym weight	0.8
Sub-type weight	0.6
Super-type weight	0.9
Super-type space expansion level	0
Sub-type space expansion level	0
Synonym space expansion level	1
Term dictionary expansion level	3

Term super-type expansion level	0
Term sub-type expansion level	0
Term synonym expansion level	1

4.4 **Evaluation Results**

In each run of the cross validation run mentioned above, we have evaluated three different learning settings as one baseline[22] setting per each run.

As shown in table 4.3, under run 1 the matching function without learning has a 19% of its retrieved matched concept score as with Top 1, compared to a 27% Top 1 accuracy using the matching function with learning, which indicates 8% improvements in Top 1 accuracy.

Comparing top 5 accuracy between different matching techniques, and different cross validation runs we observe a slight improvement in the number of concepts that has exact within specified range using learning, where we have a 10%, 11%, and 8% improved to run 1, run 2, run 3 respectively after the learning operation. Similar observation found for Top 10, and Top 15 accuracy ranges that indicate to the improvement that learning component provides to the matching function mentioned in table 4.3.

[22] As the baseline setting does not involve learning, the base line is given based on the matching function without learning component and manually predefined configurations.

Table 4-3 3-Cross validation accuracy comparison

RUN	Matching Function (without learning)			Matching Function (with learning)			Improvement		
	1	2	3	1	2	3	1	2	3
Top 1	19%	16%	22%	27%	25%	28%	8%	9%	7%
Top 5	37%	35%	43%	48%	46%	51%	10%	11%	8%
Top 10	39%	37%	44%	55%	53%	55%	16%	16%	11%
Top 15	40%	38%	45%	57%	53%	56%	16%	15%	12%
Retrieved	72%	78%	74%	82%	84%	84%	10%	6%	10%

.

Worth mentioning the retrieved factor presented in table 4.3, which represents number of matched search space concepts that contains the equivalent concept even if it is not included in within the top 15 ranked elements. The matching function without learning in run 1 only retrieved 72% of the testing set concepts, while only 78% of the second training set compared to 74% of the third testing set. Even there is a small improvement in the number of retrieved concepts, however this issue returns to the fact that there are some wrong mapped manually matched concepts that mapped to the wrong version of the WordNet database, or due to search building parameters that did not include terms linked to the matched concept.

In addition to the previous observation, we tracked the tendency in the weight values and how they affect the ranking as explained in section 3.2.13. Figure 4.2 shows a trend chart of weigh variable change, where keyword associated weight value that keep bouncing between 1 and 0.8 indicating its importance in the matching operation. The second important variable is the super-type associated weight value that has a

range between 0.8 and 0.9. The remaining other two weight values associated to the synonym and subtype have less influence on the ranking operation by having a tendency to values between 0.3 to 0.4 for synonyms, and 0 to 0.2 range for the subtype weight value.

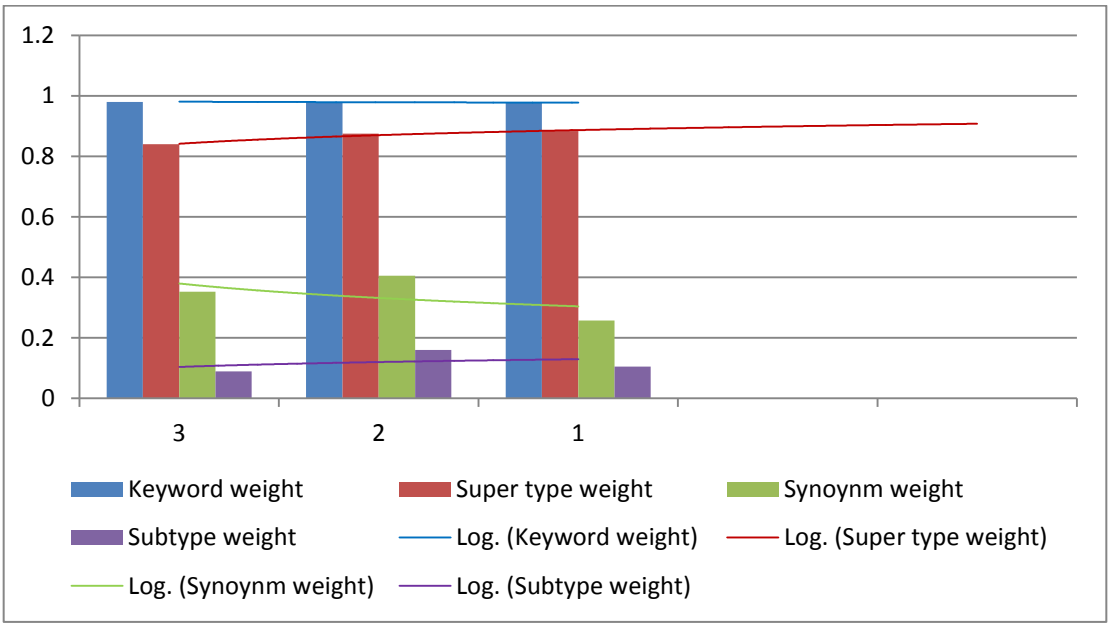

Figure 4.2 Average weight tendency change between validation runs

The expansion variables showed kind of stability in their values between different runs as shown in table 4.4, while there are changes compared to the initial values, such as the super-type search expansion variables moved from 0 as initial value into 2 through different evaluation runs, while other search space subtype expansion variable didn't change at all.

Table 4-4 Expansion variables change between evaluation runs

Variable	1	2	3	Initial
Super-type space expansion level	2	2	2	0
Sub-type space expansion level	0	0	0	0
Synonym space expansion level	1	2	2	1
Term dictionary expansion level	5	6	6	3
Term super-type expansion level	2	1	1	0
Term sub-type expansion level	0	1	1	0
Term synonym expansion level	2	2	2	1

4.5 **Discussion of Results**

The impact of the learner component indicates that the accuracy of the matching function was improved with an average of 8% for top 1 matched concepts between different validation runs, an average of 10% for those concepts that found within the top 5 matched concepts, and an average of 14%, and 15% improvement for both concepts found either between the top 10, or the top 15 matched concepts respectively.

Table 4-5 Accuracy comparison between with learning matching function and non-neural matching function

Accuracy	Matching Function (without learning)	Matching Function (with learning)	Improvement
Top 1	19%	27%	8%
Within Top 5	38%	48%	10%
Within Top 10	40%	54%	14%
Within Top 15	41%	56%	15%
Retrieved	72%	83%	7%

As shown in the previous table 4.5, the maximum retrieved percentage of matched concepts is 81% from the testing sets, which referees to a certain error in the manually matched and other external factors as follows:

1. Dictionaries and treasures may return none rich results for Arabic terms, which affect the constructed search space content.
2. Week automated translation services; where most of online translation services that uses mainly the modern language and does not reflect all Arabic possible word forms.
3. The structure of the manually matched concepts didn't necessary followed the formalization structure[23], and in some cases the Arabic gloss written to reflect a more general meaning that does not reflect the particularity of mapped concept.

In addition, the experiment showed that initial analysis and identified configuration variables and weight values were reasonable to a certain level to perform a successful matching; however, these values did not reflect the best possible setup. The learning operation leads us to identify the importance of ranking variables in following:

1. Keyword level matching considered the most influencing variable on the overall ranking elements.

[23] For more information you may refer to Appendix I.

2. Super type word come in the second level of affecting ranking operations, because they contain the general meaning of matched word.
3. While synonyms and subtype words has least effect on the final ranking of matched gloss.

In addition to main weighing variables, the learning operation highlighted other parameters, these parameters affects the matching operation since these variables affects number and quality of mapping data. These variables are number of retrieved elements from the dictionary, levels of search space expansion from the base term and dictionary output. In addition, we need to include the expansion levels of different synset relations to build up the search space.

Despise of the above issues, we have achieved an acceptable result toward improving the matching function, this improvement observed in the obtained results and that may be more accurate with additional training and enhancement over mentioned points above.

Chapter 5 Conclusions and Recommendations for Future Work

5.1 Conclusions

The process of enriching Arabic Ontology presents a challenge by itself that requires lots of manual work, which by itself is not enough to reach the desired goals in acceptable time frames. The Matching Function should provide a solution of automated mapping between ontologies, also to validate manually generated records.

In our work, we presented a new feature to the matching function that increases its accuracy. Using neural networks and machine learning techniques we advocated the use of successfully matched concepts as a base for the matching function setting reconfiguration, toward reaching best mapping. Our experiment shown an improvement in the matching processed by our approach, compared to the regular matching outputs of the same set in the experiments.

As result of this thesis work we have developed improved the matching function, which can be used in the process of mapping Arabic ontology to the English WordNet, with high accuracy rates. It can also be used to validate manually mapped records to ensure the quality of data. In addition, the Matching Function can be used as a tool to measure meaning similarity between two different sentences, either from

the same language or from different languages, which presents a helpful tool for researchers.

5.2 **Recommendations for Future Work**

Future direction of our work will involve extending the Matching Function to target other languages and ontologies, as current enhanced structure and architecture of the matching function will ease this process. Extra research and enhancement should be included to the matching function toward having a supporting tool in building Ontologies, or lexical resources. These enhancements include algorithms to examine concepts structure to the required concept standards, also this algorithm may provide suggestion to fix issues to achieve required goal.

Another important feature we need to include here is the ability to inherit relations from mapped ontologies, in addition to provide information about percentage of familiarity between different languages based on mapped information. Last and not least, we could include an additional enhancement on the learner component that utilizes word similarity to measure matching ranking.

References

[1] Mustafa Abusalah, John Tait, Michael Oakes, "Arabic-English Automatic Ontology Mapping Based on Machine Readable Dictionary," in *Mexican International Conference on Computer Science*, Sunderland SR6 0DD, UK, 2009.

[2] M. Jarrar, "Towards Methodological Principles for Ontology Engineering," Vrije Universiteit Brussel Faculty of science, May 2005.

[3] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller, "Introduction to WordNet: An On-line Lexical Database," *Int J Lexicography*, pp. 235-244, Aug 1993.

[4] Richard Beckwith, George A. Miller, and Randee Tengi, "Design and Implementation of the WordNet Lexical Database," in *WordNet pdf*, 1993, pp. 62-77.

[5] M. Jarrar, "Building a Formal Arabic Ontology (Invited Paper). In proceedings of the Experts Meeting on Arabic Ontologies and Semantic Networks. Alecso, Arab League," Tunis, 2011.

[6] Samar Ayesh,Muath Al-Badawi, Hammam Samara, "TOWARDS BUILDING AN ARABIC ONTOLOGY," BIRZEIT, June 2010.

[7] M. Jarrar, "Arabic Ontology," BirZeit Unierstiy, 2011. [Online]. Available: http://sites.birzeit.edu/comp/ArabicOntology/. [Accessed 2012].

[8] M. Jarrar, "Online Courses by Mustafa Jarrar - Advanced Artificial Intelligence Course- AAI-Spring2012," Apr 2012. [Online]. Available: http://jarrar-courses.blogspot.com/2012/05/test.html. [Accessed 2012].

[9] G. A. Miller, "Nouns in WordNet: A Lexical Inheritance System," August 1993.

[10] G. A. Miller, "Adjectives in WordNet," in *WordNet.pdf*, August 1993, pp. 10-25.

[11] "WordNet Sql Server Database," [Online]. Available: http://opensource.ebswift.com/WordNetSQLServer/.

[12] Anton Deik, Bilal Faraj, Ala Hawash, Mustafa Jarrar, "Towards Query Optimization for the Data Web Disk-Based Algorithms: Trace Equivalence and Bisimilarity," in *In proceedings of the International Conference on Intelligent Semantic Web – Applications and Services,* June 2010.

[13] Horacio Rodríguez1, David Farwell1, Javi Farreres1, Manuel Bertran1, Musa, "Arabic WordNet: Current State and Future Extensions," in *The 2nd International Conference on Arabic Language Resources and Tools*, Cairo, Egypt, 2008.

[14] R´gis CLOUARD, Arnaud RENOUF, Marinette REVENU, "An Ontology-Based Model for Representing Image Processing Application Objectives," *International Journal of Pattern Recognition and Artificial Intelligence,* vol. 24, no. 8, 2012.

[15] M. Jarrar, "Jarrar's Courses AAI: Lexical Semantics and Multilingualism," Apr

2012. [Online]. Available: http://jarrar-courses.blogspot.com/2012/05/aai-lexical-semantics_08.html. [Accessed 2012].

[16] Hamdy Mubarak, Mostafa Ramadan, "Arabic WordNet Enrichment Using Sakhr Semantic Features and Lexical Database," in *Arabic WordNet Conference, the Arab League Educational, Cultural and Scientific Organization (ALECSO),* Tunisia, July 2011.

[17] P. Vossen, "EuroWordNet: A Multilingual Database with Lexical Semantic Networks," *Computational Linguistics,* pp. 628-630, 1998.

[18] Sergey Yablonsky, Andrey Sukhonogov, "Semi-Automated English-Russian WordNet Construction: Initial Resources, Software and Methods of Translation," Petersburg Transport University, Moscow, 2006.

[19] Wikipedia, "Centrality," Jun 2011. [Online]. Available: http://en.wikipedia.org/wiki/Centrality. [Accessed June 2012].

[20] M. NEGNEVITSKY, Artificial Intelligence, A Guide to Intelligent Systems, Second Edition, Addison Wesley, 2005.

[21] T. M. Mitchell, Machine Learning, McGraw-Hill Science/Engineering/Math, March 1, 1997.

[22] WikiPedia, "Artificial intelligence," 2012. [Online]. Available: http://en.wikipedia.org/wiki/Artificial_intelligence. [Accessed Jun 2012].

[23] M.L. Caliusco and G. Stegmayer, "Semantic Web Technologies and Artificial

Neural Networks for Intelligent Web Knowledge Source Discovery," in *Semantic Web Technologies and ANNs*, pp. 17-36.

[24] Davies, J., Studer, R., Warren, P, "Semantic Web Technologies: Trends and Research in Ontology-Based Systems," Wiley,, London, 2007.

[25] emil.chifu, "A Neural Model for Ontology Matching," 2011.

[26] B. Bagheri Hariri, H. Abolhassani, and H. Sayyadi, "A Neural-Networks-Based Approach for Ontology," Jappan, 2006.

[27] N. Cristianini and J. Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.," Cambridge Univ Press, 2000.

[28] R. Ichise, "Machine Learning Approach for Ontology Mapping using Multiple Concept Similarity Measures," Principles of Informatics Research Division, National Institute of Informatics, Tokyo,, 2008.

[29] J. Schneider, "Cross Validation," Carnegie Mellon University, 1997. [Online]. Available: http://www.cs.cmu.edu/~schneide/tut5/node42.html. [Accessed 2013].

[30] C. Fellbaum, "English Verbs as a Semantic Net," *Int J Lexicography,* pp. 278-301, 1990.

[31] Valentina Balkova, Andrey Sukhonogov, and Sergey Yablonsky, "Russian WordNet From UML-notation to Internet/Intranet Database Implementation," in

Second International WordNet Conference, GWC 2004, Jan 2004.

[32] Nora I. Al- Rajebah, Hend S. Al-Khalifa, AbdulMalik S. Al-Salman, "Exploiting Arabic Wikipedia for Automatic Ontology Generation: A Proposed Approach," in *International Conference on Semantic Technology and Information Retrieval*, Putrajaya, Malaysia, 2011.

[33] S. Yablonsky, "English-Russian WordNet for Multilingual Mappings," St. Petersburg State University, Volkhovsky, Russia, 2010.

[34] William BLACK, Sabri ELKATEB, Piek VOSSEN, "Introducing the Arabic WordNet Project," in *Proceedings of the third International WordNet Conference*, 2006.

[35] M. Watson, Practical Artificial Intelligence Programming With Java, Third Edition, November 11, 2008.

[36] WikiPedi, "Natural language processing," 2012. [Online]. Available: http://en.wikipedia.org/wiki/Natural_language_processing. [Accessed Jun 2012].

Appendices

Appendix I Gloss construction instructions

Source: http://jarrar-courses.blogspot.com/2011/12/arabic-ontology.html

What should and what should not be provided in a gloss:

1. Start with the *principal/super type* of the concept being defined.

   E.g. ‘Search engine’: “A computer program that …”, ‘Invoice’: “A business document that…”, ‘University’: “An institution of …”.

2. Focus on distinguishing characteristics and intrinsic prosperities that differentiate the concept out of other concepts.

   E.g. Compare, ‘Laptop computer’: “A computer that is designed to do pretty much anything a desktop computer can do, it runs for a short time (usually two to five hours) on batteries”. “A portable computer small enough to use in your lap…”.

3. Written in a form of propositions, offering the reader *inferential knowledge* that help him to construct the image of the concept.

   E.g. Compare ‘Search engine’: “A computer program for searching the internet, it can be defined as one of the most useful aspects of the World Wide Web. Some of the major ones are Google, ….”; A computer program that enables users to search and retrieves documents or data from a database or from a computer network…”.

4. Use supportive examples :

   - To clarify cases that are commonly known to be false but they are true, or that are known to be true but they are false;

- To strengthen and illustrate distinguishing characteristics (e.g. define by examples, counter-examples). Examples can be types and/or instances of the concept being defined.

5. Be consistent with formal definitions/axioms.

6. Be sufficient, clear, and easy to understand

Appendix II Training data sample

Arabic Term	Arabic Gloss	English Term	English Concept
اختصاص المحاكم	هو الولاية القضائية الممنوحة لها لتحقق في الدعاوى وتحكم فيها.	jurisdiction	(law) the right and power to interpret and apply the law
عَيْن	ذات الشيء ونفسه	same	identical
التعريف	وهو قول يشرح المعنى الذي يدل عليه اللفظ فيزيل ما تنطوي عليه الألفاظ من غموض ويقابل التعريف الحقيقي الذي هو أساس التعريف	definition	a concise explanation of the meaning of a word or phrase or symbol
صَرْف	انْفَقَ المَال	pay	give money, usually in exchange for goods or services
اِشْتَرَكَ	اشترك الرجلُ في أمر أو شيء أي دخل فيه وشَارَكَ من كان فيه قبله	Participate	become a participant; be involved in
تغرير	أن يخدع أحد الأشخاص آخر بوسائل احتيالية قولية أو فعلية تحمله على الرضا بما لم يكن ليرضى به بغيرها.	lure	provoke someone to do something through (often false or exaggerated) promises or persuasion
جَبَّبَ	فلانٌ المواشي سقاها	Water	provide with water
فهرس	كِتَاب تجمع فيه أسماء الكتب بترتيب يسهل منه إجاد العنوان المراد البحث عنه	index	an alphabetical listing of names and topics along with page numbers where they are discussed
الحلول الاتفاقي	حلول الموفي لدين غيره محل الدائن باتفاق مع الدائن أو المدين.	subrogate	substitute one creditor for another, as in the case where an insurance company sues the person who caused an accident for the insured
الطّثرة	الدّسم الذي يعلو اللبن	Cream	the part of milk containing the butterfat
المصادفة	لا تعني أن الأحداث تخرج من العدم بل تعني فقط أن الأحداث تحدث بفعل علل غير مطردة ولا منسقة فيما بينها	spontaneous	happening or arising without apparent external cause
عقار	الدَوَاء أو المَادَّة الطِبّيَّة المستخرجة من النَبَات، الجمع عَقَاقِير	herbal medicine	a medicine made from plants and used to prevent or treat disease or promote health
الشّرف	الارتفاع والانحدار	Downslope	a downward slope or bend

ترست مجموعة أو اتحاد شركات	انضمام عدة مؤسسات أو اندماجها في مؤسسة تحت إدارة واحدة بحيث تذوب شخصية كل منها والهدف الرئيسي من مثل هذا التنظيم هو السيطرة على فرع معين من فروع النشاط الاقتصادي وخاصة في مجالات الصناعة والتجارة والخدمات.	cartel	a consortium of independent organizations formed to limit competition by controlling the production and distribution of a product or service
أَكَلَ	وضع الطَعَام في فمه ثم مضغه ثم بلع؛ تَنَاوَلَه	eat	take in solid food
مِقْصَف	ما يشبه المطعم إلا أن الخدمة فيه قليلة والطعام عادي غير مزين، يكثر عادة في الأماكن التي يحتاج فيها الناس إلى الطعام بعجلة مثل المدارس وأماكن العمل حديث	canteen	a recreation room in an institution
طَاسَة	هو الطَاس والطَاسَة أخصّ من الطَاس وهو الإناء من المعدن لا يد له يُشرب به الماء أو الحليب أو غيره من الشراب البارد؛ تقليديًا يزيّن بالنقوش ويصنع من الصُفْر أو الفِضَّة أو الذَهَب وفي العصر الحديث شاع صنعه من الصُلْب الغير قابل للصدأ	Chalice	a bowl-shaped drinking vessel; especially the Eucharistic cup
تَرَدَّدَ	الرجوع أو الذهاب مرّة بعد أخرى إلى مكان ما	frequent	be a regular or frequent visitor to a certain place
حَفِيَ	كان حافيا، أي لا يلبس في قدمية شيئا	unshod	not shod
وجه	الشيَ نحو كذا أدار وجه الشيء حتى قابل كذا	face	turn so as to face; turn the face in a certain direction
تجويف	حيز أجوف في جسم ما مثل التجويف البطني.	Cavity	(anatomy) a natural hollow or sinus within the body
بِدْعَة	أي شيء 'يختلق' في أي مجال من المجالات ولم يكن أحد قد فعل مثله من قبل ليقاس عليه ولا ليأخذ منه	heresy	any opinions or doctrines at variance with the official or orthodox position
اِحْمَرَّ	ازداد حَمَارُهُ	Redden	turn red or redder
وَيلْزِيَّة	المنسوبة إلى منطقة وَيلْز في بَرِيطَانِيَا	Welsh	of or relating to or characteristic of Wales or its people or their language
أزَل	دوام الوُجُود في المَاضِي	immemorial	long past; beyond the limits of memory or tradition or recorded history
انخلع	قلبه من الخوف أو الحزن زادت عليه المشاعر بشدة حتى شعر وكأن قلبه قد انتزع من صدره مجاز	afflict	cause great unhappiness for; distress
فَقْر	في طب فَقْر الدم ويقال أيضا فُقْر الدم مرض يصيب الدم سببه نقص فيه واضطراب في تكوينه يصحبه شحوب وبهر وخَفَقان	Anemia	a deficiency of red blood cells
غَدَرَ	أخل بالعهد ولم يفِ به	breach	a failure to perform some promised act or obligation
كُلِّيَّة	مَعْهَد أو مَدْرَسَة للدِرَاسَات العُلْيَا والمُتَخَصِّصَة	college	an institution of higher education created to educate

			and grant degrees; often a part of a university
البيع الجبري	بيع يتم تنفيذاً لحكم قضائي.	forced sale	a sale of property by the sheriff under authority of a court's writ of execution in order satisfy and unpaid obligation
متعرج	وصف لما له انعطافات مثل بعض الخيوط الفطرية.	flexuous	having turns or windings
رابطة الدم	هو الانحدار من أصل مشترك ودرجة رابطة الدم بالنسبة للذكر والأنثى وهي العامل الأكبر في تحديد حقهما في الزواج.	parentage	the kinship relation of an offspring to the parents
ثَلَاثَاء	الشيء الذي يجيء ثالثا	third	in the third place
حَلَقَ	أزال شعر رأسه أو ذقنه كله بالمُوسَى	Shave	remove body hair with a razor
مُخْلِص	الشخص الذي يفي ولا يخون	loyal	unwavering in devotion to friend or vow or cause
الأعمال التجارية	أعمال تتعلق بالتجارة وتحدد حسب طبيعة العمل أو صفة القائم به.	commercial activity	activity undertaken as part of a commercial enterprise
سِدَارَة	غطاء للرأس كان يُلبس في العراق مع الزيّ الغربيّ يعادل الطَرْبُوش في مصر	Fez	a felt cap (usually red) for a man; shaped like a flat-topped cone with a tassel that hangs from the crown
اِنْحَلَّ	في طب جسم الإنسان ضعف من مرض أو تعب	weaken	become weaker
حَقّ	ضد البَاطِل، العَدْل والحَقِيْقَة والوُجُوْد الثَابِت والصِدْق والذي لا جدال فيه لا جمع له	correct	free from error; especially conforming to fact or truth
أَوْسَعَ	الشيءَ صيّره أَوْسَع متعد	Enlarge	make larger
مصفي الشركة	مصفٍ يعين لتصفية الشركة.	liquidator	(law) a person (usually appointed by a court of law) who liquidates assets or preserves them for the benefit of affected parties
جَمَعَ	الشيء ضَمَّه إل بعضه	gather	assemble or get together
دينار	عملة تونس 1 دينار = 1000 مليم.	dinar	the basic unit of money in Tunisia
الضرر	أذى مادي أو أدبي يتعرض له شخص بفعل الغير.	Damage	any harm or injury resulting from a violation of a legal right
تردد	خاف أو قلق ولم يرغب في عمل شيء ما رغبة كاملة	puzzle	be uncertain about; think about without fully understanding or being able to decide
خف	الحذاء الخفيف المصنوع من الجلد الطري أو القماش ونعله طريّ ويغطي الكعب، سميّ بالخُفّ لخفة وزنه الجمع خِفَاف	shoe	footwear shaped to fit the foot (below the ankle) with a flexible upper of leather or plastic and a sole and heel of heavier material
اِعْتَادَ	الشيءُ فلانًا عاد إليه مرارًا	Used to	in the habit

مَرَّ	الحدثُ أو الأمرُ حدث وانتهى	happen	come to pass
جِنْس	في احياء هو أخص درجة من درجات التصنيف العلمي	Genus	(biology) taxonomic group containing one or more species
الملّاح	كل شخص استخدم على متن سفينة للعمل فيها أثناء رحلة بحرية.	sailor	any member of a ship's crew
أَيِّم	من ليس له زوج، يستخدم للرجل والمرأة، إلا أنه أكثر ما يستخدم لمن كان متزوجًا ثم خلا بموت أو طلاق	widowed	single because of death of the spouse
المقاول من الباطن	مقاول يتفق مع المقاول الذي تعهد بتنفيذ العمل في عقد المقاولة على أن يقوم بدلاً منه بالأعمال التي تعهد بها لرب العمل كلاً أو جزءاً.	Subcontractor	someone who enters into a subcontract with the primary contractor
أم	القوم تقدّمهم واتبعوه وكان لهم إماما المصدر أمًّا وإمامةً	head	travel in front of; go in advance of others
حاجز أمواج	بنية صعبة هندسيا تقام في البحر وتحمي الميناء او المرسى او الشاملئ او منعلقة الساحل، بكسر الامواج. ويمكن ربط حاجز الامواج بالشاملئ او تركه على المياه الساحلية	Breakwater	a protective structure of stone or concrete; extends from shore into the water to prevent a beach from washing away
أَزْرَق	أحد الأَلْوَان الرئيسية (الأَحْمَر والأَزْرَق والأَصْفَر) وهو لون السَمَاء نهارا حين تكون صافية انظر الصورة	blue	blue color or pigment; resembling the color of the clear sky in the daytime
مَدَّ	أبعد أو أطال أو ارتفع	extend	extend one's limbs or muscles, or the entire body
فلان	فلان كلمة تستخدم وصف لشخص ما بدون ذكر من هو، غالبًا يقصد به أي شخص من الناس ، أو شخص لا يريد المتحدث ذكره	anonymous	having no known name or identity or known source
التأقلم	التواؤم الفيزيولوجي مع التغيّرات المناخية	adaptation	the process of adapting to something (such as environmental conditions)
سَقَطَ	الجَنِين أَجْهِضَ، وغالبا ما يقصد بذلك أن الاجهاض كان من غير عَمْد	Miscarry	suffer a miscarriage
بَشِع	الدميم الذي لا تستسيغه العيون مجاز	Ugly	displeasing to the senses
الضرر الحال	الضرر الذي وقع بالفعل.	detriment	a damage or loss
الأقط	الرّجل الثقيل الوخم	heavy	slow and laborious because of weight
سَمِعَ اليه	الإدراك مع الإصغاء والإهتمام	Listen	pay close attention to; give heed to

عُقَاب	طَيْر من الجَوَارِح، أكبر التي تصيد منها يكنى بأبي الهَيْثَم ومعروف بحِدَّة البصر	Eagle	any of various large keen-sighted diurnal birds of prey noted for their broad wings and strong soaring flight
حلول الأجل	هو التاريخ الذي يصبح فيه الالتزام واجب التنفيذ وقد يكون عيناً أو مبلغ النقود.	maturity date	the date on which a financial obligation must be repaid
الغريزة	نمط معقد من النشاط أو الاستجابة، موروث أو غير متعلم، شائع في نوع أحيائي أو بيولوجي معين	Instinct	inborn pattern of behavior often responsive to specific stimuli
اِنْتِحَار	قتل النفس عمدا	Suicide	the act of killing yourself
حَلَقَ	أزال شعر رأسه أو ذقنه كله بالمُوسَى	Shave	the act of removing hair with a razor
كَلْب	حَيَوان مِنْ فَصِيلَة الكَلْبِيَّات من الضَوَارِي، يوجد منه أنواع كثيرة بعضها يستخدمه البشر مثل كلب الصَيْد وكلب الحِرَاسَة وغيرها	Dog	a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds
شهادة الإيداع	شهادة يصدرها مصرف لشخص أودع لديه مبلغاً معيناً قابلاً للدفع لحامله.	certificate of deposit	a debt instrument issued by a bank; usually pays interest
دِيسَمْبَر	الشهر الثاني عشر في السنة الروميّة	december	the last (12th) month of the year
مُقَلْنس	وصف لما له قلنسوة.	Calyptrate	having a calyptra
بَعِيْد	ضد القَرِيب، الذي يفصل بينك وبينه مسافة أو زمان	away	at a distance in space or time
أتول	جزيرات في المحيطات الدافئة تنتظم في نطاق يتوسطه متسع مائي.	Atoll	an island consisting of a circular coral reef surrounding a lagoon
السوق	مكان تتم فيه مبادلة سلعة معينة على نطاق تجاري.	marketplace	the world of commercial activity where goods and services are bought and sold
الملاحظة	ما يؤخذ على الرأى أو الكتاب من هنات	criticize	find fault with; express criticism of; point out real or perceived flaws
مُذَبْذَب	المتردد بين أمرين أو شيئين	hesitant	lacking decisiveness of character; unable to act or decide quickly or firmly
عَمْيَاء	من كانت فاقدة البصر منذ الولادة تحديدًا	Blind	unable to see
تنحى	عن الأمر أبعد نفسه عنه ولم يعد يتدخل فيه	remove	go away or leave
الطبن	الحطب وغيره الذي تأتي به الريح	firewood	wood used for fuel
الحكم	امر قانوني مسطور او شرط مدوّن في عقد قانوني وعلى سبيل المثال في عقد الوصية او عقد الهبة.	judgement	the legal document stating the reasons for a judicial decision

عدل	عَدَلَ الرجل عن جَادَّة الصواب، أي مال عنها	damage	the occurrence of a change for the worse
الخيار	حق طليق لا يسقطه التقادم لأنه شروع في استعمال حق الملكية أو في استعمال الحرية الطبيعية التي خولها الانسان وهذا الحق الطليق يقوم على سبب يتجدد أبداً.	be given	have a tendency or disposition to do or be something; be inclined
بِيَانُو	آلة وترية يتم العزف عليها بالضغط على مفاتيح أو أزرار تجعل المطارق تضرب على أوتار مشدودة لحدث أصواتا	piano	a stringed instrument that is played by depressing keys that cause hammers to strike tuned strings and produce sounds
حَيْض	هو سيلان الدم من رحم المرأة مرة في كل شهر	Menstruation	the monthly discharge of blood from the uterus of nonpregnant women from puberty to menopause
التلاع	مجاري الماء من أعالي الوادي وما انهبط من الأرض	ravine	a deep narrow steep-sided valley (especially one formed by running water)
البيع العاجِل	بيع ينفذ فوراً بدفع الثمن وتسليم المبيع.	sell	be sold at a certain price or in a certain way
العرف	ما تعارف عليه الناس في عاداتهم ومعاملاتهم	custom	accepted or habitual practice
عَامِل	هو المُتَسَبب بحدوث شيء ما بسب ما يعمله الجمع عَوَامل	Factor	anything that contributes causally to a result
كَاتِب	وظيفة صغيرة يقوم الموظف فيها بكتابة بعض الأمور لصاحب العمل أو لمديره الجمع كُتَّاب وكَتَبَة	copyist	someone employed to make written copies of documents and manuscripts
الأَطْلَسِيَّة(الفِهْقُة)	الفقرة العنقية الأولى التي تتمفصل مع الجمجمة.	atlas	the 1st cervical vertebra
عَمَّة	هي أُخْت الأَبّ	aunt	the sister of your father or mother; the wife of your uncle
خِمَار	ما تغطي المرأة بها شعرها ورقبتها	scarf	a garment worn around the head or neck or shoulders for warmth or decoration
شباط	الشهر الثاني في السنة الشمسية بعد كَانُونُ الأَوَّل وقبل آذَار عِدّته 28 يوما	february	the month following January and preceding March
البسل	أخذ القليل من أي شيء	nibble	eat intermittently; take small bites of
البَشَرة	الطبقة الخارجية الخلوية التي تكسو جسم الكائن الحي.	epidermis	the outer layer of the skin covering the exterior body surface of vertebrates
اِعْتَبَرَ	بالشيء اتّعظ واخذ منه عِبْرَة	moralize	interpret the moral meaning of
اِتَّكَلَ	عليه استسلم له وترك الأمر بيديه	depend	be contingent upon (something that is elided)
تأمين على الحياة	تأمين تكون فيه حياة المؤمن عليه محلاً للعقد.	Life Insurance	insurance paid to named beneficiaries when the insured person dies

المعاوضات	تصرفات قانونية يتبادل فيها الطرفان الأخذ والعطاء.	Tradeoff	an exchange that occurs as a compromise
أَصْفَر	أحد الأَلْوَان الرئيسية (الأَحْمَر والأَصْفَر والأَزْرَق)، وهو لون الزُعْفُرَان والذَهَب والعُصْفُر	yellow	of the color intermediate between green and orange in the color spectrum; of something resembling the color of an egg yolk
مَنْزِل	مكان نزول القافلة	landing place	structure providing a place where boats can land people or goods
النصاب	عدد الأعضاء الذي يتم به انعقاد مجلس أو اجتماع لجنة معينة.	quorum	a gathering of the minimal number of members of an organization to conduct business
مباعد للمحور	وصف للسطح أو الوجه البعيد عن المحور لعضو نباتي كورقة أو زهرة أو فرع ينشأ على جانب المحور أو الساق.	abaxial	facing away from the axis of an organ or organism
سرمد	الزَمَان الذي لا بداية له ولا نهاية أي أنه يشمل الأبَد والأزَل	aeonian	continuing forever or indefinitely
عَرَصَة	الأَرْض حول الدار أو جنبه تابعة له واسعة وخالية ليس فيها بناء	backyard	the grounds in back of a house
الرّتوة	مكان مرتفع من الأرض مثل الرّبوة	Knoll	a small natural hill
إِمْلَاء	قراءة الشيء على أحد ما حتى يكتبه كما قُرِأ	dictation	speech intended for reproduction in writing
قُلْنُسُوَة	لباس داخلي للرأس يُلبس غالبا داخليا تحت العَمَامَة أو الكُوفِيَّة أو غيره من أغطية الرأس للرجال والنساء، يسمى أيضا الطَاقِيَّة	beret	a cap with no brim or bill; made of soft cloth
وَحْي	في اسلام هو جِبْرِيْل لأنه ينزل بالوحي من الله	Gabriel	(Bible) the archangel who was the messenger of God
بوربو	نوع من الخَمْر تصنع من الذرة	bourbon	whiskey distilled from a mash of corn and malt and rye and aged in charred oak barrels
زَانِيَة	المرأة التي تمارس الجنس مع غير زوجها أو بدون زواج	adulterer	someone who commits adultery or fornication
الرهن	تخصيص مال معين لضمان الوفاء بالدين.	Mortgage	a conditional conveyance of property as security for the repayment of a loan
اسْتِقَامَة	اتخاذ الطريق السَوِيّ وعدم المَيْل إلى سُوء الخُلُق أو الأخْلاق	Righteousness	adhering to moral principles
عاب	عليه كذا قال له أن كذا من عُيُوبه	reproach	express criticism towards
ارتفاق	هو تكليف مرتب على عقار للاستعمال أو لمنفعة عقار يملكه مالك آخر فإذا كان الارتفاق تكليفاً بعدم الفعل فهو سلبي أو بالفعل فهو إيجابي.	trust	something (as property) held by one party (the trustee) for the benefit of another (the beneficiary)

جماعة	طَائِفَة من الناس التي يجمعها مكان و بيئة واحدة أو تتشارك اهداف و غايات واحدة	assembly	a group of persons gathered together for a common purpose
الطعن	الالتجاء إلى سلطة إدارية أو قضائية للحصول على إلغاء قرار أو حكم أو تعديله أو تفسيره وتفيد أيضاً طلب إعادة بحث النزاع المحكوم فيه من قبل.	appeal	(law) a legal proceeding in which the appellant resorts to a higher court for the purpose of obtaining a review of a lower court decision and a reversal of the lower court's judgment or the granting of a new trial
علم الأحياء	العلم الذي يبحث في دراسة الكائنات الحية.	Biology	the science that studies living organisms
النظرية	بناء أو نسق متدرج من الأفكار، يتم فيه الانتقال من المقدمات إلى النتائج	Theory	a well-substantiated explanation of some aspect of the natural world; an organized system of accepted knowledge that applies in a variety of circumstances to explain a specific set of phenomena
الطُحْلُب (ج الطَّحَالِب)	من النباتات الدنيئة التي لا يتميز جسمها إلى جذور و سيقان وأوراق وهي غنية بالكلوروفيل.	alga	primitive chlorophyll-containing mainly aquatic eukaryotic organisms lacking true stems and roots and leaves
ارتفاق	هو تكليف مرتب على عقار للاستعمال أو لمنفعة عقار يملكه مالك آخر فإذا كان الارتفاق تكليفاً بعدم الفعل فهو سلبي أو بالفعل فهو إيجابي.	easement	(law) the privilege of using something that is not your own (as using another's land as a right of way to your own land)
نام	الشيء مُدَّدَ على الأرض وهو عادة غير ذلك مجاز	lie	be lying, be prostrate; be in a horizontal position
مَوْطِن	الموضع أو الموقع بصورة مجازية، أي قد لا يكون له بعد مكاني حقيقي	limbo	an imaginary place for lost or neglected things
الربح	هو المكسب الذي يتحقق إما من عمل تجاري منفرد أو من مجموعة أعمال خلال سنة حسابية أو مالية ويقابلون بين الربح القائم أو الاجمالي وهو مجموع المبالغ المقبوضة وبين الربح الصافي الذي يحصل من الفرق بين الربح القائم والمصروفات التي أنفقت على العمل.	lucre	the excess of revenues over outlays in a given period of time (including depreciation and other non-cash expenses)
الإجرائي الصناعي	مجموعة القواعد القانونية المتعلقة بتنظيم النشاط الصناعي وحمايته وتطويره.	industrial	having highly developed industries

الهندوسية,الهندوكية	أسلوب فى الحياة أكثر مما هى مجموعة من العقائد والمعتقدات، وتاريخها يوضح استيعابها لشتى المعتقدات والسنن، وليست لها صيغ محددة المعالم، ولذا تشمل من العقائد ما يهبط بها إلى عبادة الأحجار والأشجار والحيوان، وما يرتفع إلى التجريدات الفلسفية الدقيقة	hinduism	a body of religious and philosophical beliefs and cultural practices native to India and characterized by a belief in reincarnation and a supreme being of many forms and natures, by the view that opposing theories are aspects of one eternal truth, and by a desire for liberation from earthly evils
المقاومة	رد فعل سياسي أو عسكري يعبر عن رفض التدخل الأجنبي	resistance	the military action of resisting the enemy's advance
التهجين	يعني اسفاد حيوانين من نوع واحد ولكن من سلالتين مختلفتين أو عِرقين مختلفتين كإنزاء أو نزو حصان عربي أصيل على حِجْر من الاكاديش وكانزاء ثور بلدي على بقرة من سلالة العَكْش الناتج من هذه العمليّة والتلقيح يقال له الهجين والهجينة	Hybridization	(genetics) the act of mixing different species or varieties of animals or plants and thus to produce hybrids
عَلَّمَ	التعويد تدريجيا على أداء بعض الأفعال والمواقف	accustom	make psychologically or physically used (to something)
الكاهن	الذى يتعاطى الخبر عن الكائنات فى مستقبل الزمان، ويدعى معرفة الأسرار	forecaster	someone who makes predictions of the future (usually on the basis of special knowledge)
نَظَرَ	في الشيء تفحصّه جيدا كي يتخذ رأيًا فيه	Look into	examine so as to determine accuracy, quality, or condition
جِدَار	الحَائِط، بناء صلد طويل وعالي إلا أنه قليل العرض، يستخدم للتطويق والحماية قد يبنى من أي نوع من مواد البِنَاء	wall	an architectural partition with a height and length greater than its thickness; used to divide or enclose an area or to support another structure
العدل	تحقيق المساواة بين المواطنين أمام القانون.	Equality before the law	the right to equal protection of the laws
الإمر	الأحمق الضعيف الذي لا رأي له من الرجال	fool	a person who lacks good judgment
طَيْر	لحم الطيور البرية أو المنزلية المستخدم كغذاء	poultry	flesh of chickens or turkeys or ducks or geese raised for food
سَمِنَ	زاد حجمه من الأكل أو قلة الحركة، ازدادت كمية السَمْن تحت جلده	fatten	make fat or plump
أَبَّنَ	أثْنَى على الشخص بعد موته	eulogise	praise formally and eloquently
قِطَّة	حيوان مستأنس من فصيلة السنوريات	cat	feline mammal usually having thick soft fur and no ability to roar: domestic cats; wildcats
بكتيري	وصف لما له علاقة بالبكتيريا.	Bacterial	relating to or caused by bacteria

اسْتِفْسَار	عند أهل المناظرة طلب بيان معنى اللفظ وهو إنما يكون إذا كان في اللفظ إجمال أو غرابة وإلا فهو تعنُّتٌ تفوت بهِ فائدة المناظرة	challenge	questioning a statement and demanding an explanation
فُنُوْن	فنون هي كل الإبداعات الفنية التي ترتقي إلى الكمال والجمال وتسمو بالخيال إلى الخلق والإبداع كالشعر والموسيقى والرسم والنحت والزخرف والبناء	Art	the creation of beautiful or significant things
فلسفة	في صوفية التشبُّه بالإله حسب الطاقة البشريَّة لتحصيل السعادة الأبديَّة	humanism	the doctrine that people's duty is to promote human welfare
منثنية – منثن	وصف لما يبدو منحنيا.	flex	form a curve
سهم التمتع	سهم يعطى لمساهم استهلكت أسهمه وفق نظام الشركة يخوله حقوقاً أقل من تلك التي تخولها الأسهم غير المستهلكة لأصحابها.	dole	a share of money or food or clothing that has been charitably given
جرح	الشئ خَدَشَه خدشا غير زائل	injure	cause injuries or bodily harm to
بَرْبَر	أقوام في المغرب العربي وشَمَال إفْرِيقيَة بعضهم من البَدُو وبعضهم من الحَضَر	berber	an ethnic minority descended from Berbers and Arabs and living in northern Africa
الإدارة القانونية	صلاحية تمنح لشخص لحفظ واستغلال مال معين بمقتضى القانون.	administration	the persons (or committees or departments etc.) who make up a body for the purpose of administering something
تأجيل	هو عدم سريان مفعول الشيء وتأجيله إلى أجل غير محدد.	delay	the act of delaying; inactivity resulting in something being put off until a later time
فصيلة	مجموعة من الأجناس تربطها خواص مشتركة.	Family	(biology) a taxonomic group containing one or more genera
السنَّوريات	فصيلة القط من رتبة اللواحم ومنها القط والأسد والنمر وغيرها.	Felidae	cats; wildcats; lions; leopards; cheetahs; saber-toothed tigers
الرمس	الصوت الخفي	pierce	sound sharply or shrilly
الجريدة الرسمية	هي التي تصدرها الدولة فتنشر فيها القوانين والقرارات بقوانين والعقود والوثائق الإدارية ومحاضر المناقشات التي تدور في المجلس النيابي، كما تنشر فيها الإعلانات الرسمية.	gazette	a newspaper or official journal

أراغونيت	معدن من كربونات الكالسيوم (جيري) تستعمله كائنات التكليس المشكلة للقشرة أو الهيكل العظمي، مثل المرجان وبعض الطحالب الكبيرة، ويتأثر الاراغونيت أكثر من الكالسيم بتحمض المحيط، ويستعمله العديد من الكائنات البحرية	Aragonite	a mineral form of crystalline calcium carbonate; dimorphic with calcite
أيم	من ليس له زوج، يستخدم للرجل والمرأة، إلا أنه أكثر ما يستخدم لمن كان متزوجًا ثم خلا بموت أو طلاق	divorced	of someone whose marriage has been legally dissolved
مِجْنَقَة	آلة لإطلاق الطائرة من على سطح سفينة	catapult	a device that launches aircraft from a warship
أَدَارَ	الشيء جعله يدور	Spin	cause to spin
شُبَّاك	النَافِذَة المُشَبّكة بالحديد أو الخشب	Window	a framework of wood or metal that contains a glass windowpane and is built into a wall or roof to admit light or air
حِذَّاء	ما وطيء عليه البعير من خُفِّه والفرس من حافره	Shoe	U-shaped plate nailed to underside of horse's hoof
اِسْتِرَاحَة	مبنى خارج المدن يتوقف فيه المسافرون للراحة والتَزَوُّد الجمع اِسْتِرَاحَات	Layby	designated paved area beside a main road where cars can stop temporarily
شركة أشخاص	منشأة تجارية يملكها شخصين أو أكثر بمقتضى إتفاق تعاقدي	partnership	a contract between two or more persons who agree to pool talent and money and share profits or losses
ثمن البخس	ثمن يقل كثيراً عن قيمة المبيع في السوق.	Undervaluation	too low a value or price assigned to something
إزالة الشيوع	قسمة المال الشائع بين الشركاء فيه أو بيعه كما يتفقون عليه أو ينص عليها القانون.	allocation	the act of distributing by allotting or apportioning; distribution according to a plan
الإجرائي البحري	مجموعة القواعد القانونية التي تتعلق بالملاحة البحرية ونقل المسافرين والبضائع في البحر.	marine law	relating to or involving ships or shipping or navigation or seamen
تَجَبَّرَ	المريض تحسنت صحته	Heal	get healthy again
حَرَض	تحول النسيج أو العضو إلى ما هو أدنى منه تركيبا .	degeneration	the process of declining from a higher to a lower level of effective power or vitality or essential quality
مُحَمَّد	مُحَمَّد بن عبد الله هو نَبِيّ الإسْلام	mohammad	the Arab prophet who, according to Islam, was the last messenger of Allah (570-632)
أعد	الشيءَ والعُدَّةَ جَهَّزَهُ وهيَّأه	fix	make ready or suitable or equip in advance for a particular purpose or for some use, event, etc
قُبَّعَة	غطاء للرأس يكون مُخَيَّط على عكس العَمَامَة التي لا يكون فيها خِيَاطَة قُلْنُسُوَة	hat	headdress that protects the head from bad weather; has shaped crown and usually a brim

العنشط	السيء الخُلُق	naughty	badly behaved
إهْتَمَّ	لديه حب أو إعجاب أو تذوق لشيء ما	like	feel about or towards; consider, evaluate, or regard
مُعَارَضَة	عدم القبول برأي الآخر	Opposition	the action of opposing something that you disapprove or disagree with
ذهب	هذا رجل ذَهَبا أي رجل يساوي ذهبا في قيمته وقدره	gold	something likened to the metal in brightness or preciousness or superiority etc.
حق الملكية	سلطة المالك في أن يتصرف في ملكه تصرفاً مطلقاً عينياً ومنفعة واستغلالاً.	ownership	the act of having and controlling property
الطعن	وسيلة يمنحها القانون للأشخاص لإلغاء القرار القضائي المخالف للقانون إذا كان ماساً بحقوقهم.	appeal	(law) a legal proceeding in which the appellant resorts to a higher court for the purpose of obtaining a review of a lower court decision and a reversal of the lower court's judgment or the granting of a new trial
بَقَرَة	صفة؛ لفظ تحقيري يطلق على امرأة بدينة	cow	a large unpleasant woman
الشّاهد	الذي رأى الأمر وشهد عليه	Witness	someone who sees an event and reports what happened
شبكة غذائية	شبكة علاقات غذائيه ضمن مجموعه إيكولوجيه تضم سلسلات غذائيه عدة مترابطة في ما بينها	Food web	(ecology) a community of organisms where there are several interrelated food chains
جعله شرعيا	في شرع وجده حَلالا، عده حلالا	capacitate	make legally capable or qualify in law
جَدّ	أحد الذين يعود النسل إليهم عن طريق الأب، فهو أبو الأب وأبو أبوه وهكذا	ancestor	someone from whom you are descended (but usually more remote than a grandparent)
أَسْوَد	هو انعدام اللَوْن	Black	total absence of light
اِقْتِصَاد	تَوَسَّط في الانفاق فلم يسرف ولم يقتر	economy	frugality in the expenditure of money or resources
طيار	الذي يقود الطَائِرَة سواء كانت مهنته أم لا	aeronaut	someone who operates an aircraft
سلوك	مصطلح تقني يعبر عن الاستجابة لمثير ما	Behavior	the action or reaction of something (as a machine or substance) under specified circumstances
الحكم	هو ما يقضي به من القضاء.	judicial decision	(law) the determination by a court of competent jurisdiction on matters submitted to it
بيئة	المُحِيْط، مجموعة العناصر الطبيعية أو الاصطناعية التي تحيط بالإنْسَان أو الحَيْوَان أو النَبَات أو غيره وتشكل المحيط الذي يوجد فيه حديث	medium	the surrounding environment

مُحَاسَبَة	المهنة التي يقوم أصحابها بحساب الأرباح والخسائر وغيرها في الأعْمَال	accounting	the occupation of maintaining and auditing records and preparing financial reports for a business
مؤشر ِدليل	كمية قابلة للقياس مرتبطة بإحكام بهذه الظاهرة	Index	a number or ratio (a value on a scale of measurement) derived from a series of observed facts; can reveal relative changes as a function of time
حَفِيَ	كان حافيا، أي لا يلبس في قدمية شيئا	barefoot	without shoes on
اِسْتَجْرَى	وكّل وكيلا أو أرسل رسولا	Appoint	assign a duty, responsibility or obligation to
الفأرة	الجهاز اليدوي الذي تحركه على الطاولة لتشير إلى أو تختار من البنود الموجودة على شاشة الحاسوب	mouse	a hand-operated electronic device that controls the coordinates of a cursor on your computer screen as you move it around on a pad; on the bottom of the device is a ball that rolls on the surface of the pad
القيمة المتبقية	قيمة السوق المقدرة للاصول المستهلكة والتي يمكن استردادها عند سحب الاصول و عزلها من الخدمة	residuary	entitled to the residue of an estate (after payment of debts and specific gifts)
حق	ضد البَاطِل، العَدْل والحَقِيْقَة والوُجُوْد الثَّابِت والصِدْق والذي لا جدال فيه لا جمع له	correct	free from error; especially conforming to fact or truth
إندوبلازم	السيتوبلازم الموجود داخل غشاء endoplast: في النبات (?) البلازما (?) في الحيوان: السيتوبلازم الداخلي في كثير من الخلايا مثل الأوليات الحيوانية والبيض، ويختلف عن الإكتوبلازم في كونه أكثر سيولة، وفي وجود حبيبات كثيرة فيه.	endoplasm	the inner portion of the cytoplasm of a cell
الصلح	هو اتفاق يبرمه المفلس مع غرمائه العاديين ليوقف إجراءات التفليسة ، فيحدد شروط دفع ما عليه من الديون بعد إقالته عادة من جزء منها.	deal	an agreement between parties (usually arrived at after discussion) fixing obligations of each
القاضي	شخص مخوّل بإصدار حكم قضائي سواء كان ممتهناً القضاء أم لا.	Judge	a public official authorized to decide questions brought before a court of justice
الحكر	عقد يكسب المحتكر بمقتضاه حقاً عينياً يخوله الانتفاع بأرض موقوفة بإقامة مبان أو استعمالها بما لا يضرّ بالوقف لقاء أجر محدود.	royalty	payment to the holder of a patent or copyright or resource for the right to use their property
مُسْتَقِيم	ما ليس به عِوَج	straight	having no deviations
المخنة	نساء القبيلة	materfamilias	a female head of a family or tribe

أَلْمَانِيَا	دولة أوروبية في وسط أوروبا	germany	a republic in central Europe; split into East German and West Germany after World War II and reunited in 1990
الموهوب له	شخص ينتقل إليه المال الموهوب في عقد الهبة.	Donee	the recipient of funds or other benefits
كَذِب	نقيض الصِدْق، قَوْل ما يخالف ما يعلم القَائِل أنه حَقِيقَة	lie	tell an untruth; pretend with intent to deceive
طَبِيبُ أَسْنَان	الطَبِيب المختص بعلاج الفَمّ والأَسْنَان	dentist	a person qualified to practice dentistry
العمليات المصرفية	مجموع العمليات التي يؤديها المصرف مقابل عمولة.	Banking	engaging in the business of keeping money for savings and checking accounts or for exchange or for issuing loans and credit etc.
حكم	الذي يحكم بين اثنين أو أكثر، مثل القاضي إذا كان نِزاعا بين طرفين لا حدًّا يقام	arbiter	someone chosen to judge and decide a disputed issue
كَنَبَة	أَرِيكَة مُنَجَّدَة وَثِيرَة تتسع لأكثر من جالس الجمع كَنَبَات وكَنَب	Sofa	an upholstered seat for more than one person
حصل	الدَيْن أو الثَمَن جَمَعَه	raise	raise the level or amount of something
نزع الكلسيوم	عملية إزالة المحتوى الكلسي من العظام أو غيرها من المواد.	Decalcification	loss of calcium from bones or teeth
امراض محمولة بالنواقل	أمراض تحملها من ثوي إلى ثوي اخر كائنات تسمى النواقل (مثل البعوض أو القراد)، مثل الملاريا وحمى الضنك وداء الليشمانيات	vector-borne transmission	indirect transmission of an infectious agent that occurs when a vector bites or touches a person
قَصْر	مسكن كبير واسع فخم يسكنه الخاصة ذوو الثراء والسلطة سمي كذلك لقصور العامة عن الوصول إليه	Palace	a large and stately mansion
الخيال	إبداع أو القدرة على إبداع الصور الذهنية عن أشياء غير ماثلة أمام الحواس أو عن أشياء لم تشاهد من قبل في عالم الحقيقة والواقع	Imagination	the formation of a mental image of something that is not perceived as real and is not present to the senses
قلّة التغذية	حالة مؤقتة أو مزمنة تنتج عن تناول كميات أقل من الكميات التي ينصح بها لتلبية الإحتياجات اليومية من الطاقة المستخلصة من الغذاء و/أو البروتينات، وذلك إما بسبب تناول كميات غير كافية من الغذاء او بسبب سوء الإمتصاص أو سوء الإستخدام البيولوجي للمفذيات المستهلكة	malnutrition	a state of poor nutrition; can result from insufficient or excessive or unbalanced diet or from inability to absorb foods
وَارِث	هو الشخص الذي يَرِثْ	Inheritor	a person who is entitled by law or by the terms of a will to inherit the estate of another

قَلَّلَ	قدم الشئ على أنه أقل مغزى أو أهمية	downplay	understate the importance or quality of
نَزْع الكلور	عملية إزالة عنصر الكلور من المواد.	dechlorinate	remove chlorine from (water)
الميناء	موقع على شاطئ تابع لدولة معينة معد لاستقبال السفن.	seaport	a sheltered port where ships can take on or discharge cargo
عِلْمُ الحَاسُوب	في حاسوب هو العلم الذي يُعنى بدراسة كل ما يتعلق بالحَاسُوب سواء بالبرمجة أو المعدات	computing	the branch of engineering science that studies (with the aid of computers) computable processes and structures
الثكنة	الجماعة من الناس	club	a formal association of people with similar interests
طَيْر	الطَيَرَان، التَحْلِيق أو الإرْتِفَاع في الجو	flying	an instance of traveling by air
بِيْئَة	المُحِيْط، مجموعة العناصر الطبيعية أو الاصطناعية التي تحيط بالإنْسَان أو الحَيْوَان أو النَبَات أو غيره وتشكل المحيط الذي يوجد فيه	medium	the surrounding environment
المنطق الرياضي	منطق يعتمد على طائفة من الرموز والإشارات لأداء المعاني والأحكام بدلاً من الألفاظ والعبارات اتقاء لغموضها والتباسها ويخضع لقوانين معينة	Mathematical Logic	any logical system that abstracts the form of statements away from their content in order to establish abstract criteria of consistency and validity
مَائِدَة	وجبات الطعام بشكل عام	Table	food or meals in general
المنفعة	قدرة الشيء على إشباع حاجة معينة للإنسان.	satisfy	fulfil the requirements or expectations of
التوطن	حالة الكائن الحي الذي ينتقل إلى بيئة جديدة فيجد فيها مهادا صالحا لنموه وتكاثره واستقراره.	establishment	(ecology) the process by which a plant or animal becomes established in a new habitat
غَادَرَ	تَرَكَ الشيء سواء كان غَدْرَا أم لا	Leave	go and leave behind, either intentionally or by neglect or forgetfulness
انتروكيناز	إنزيم تفرزه الأمعاء الدقيقة في الفقاريات يحول مولدة التربسين (تربسينوجن) إلى تربسين فعال نشيط.	enterokinase	enzyme in the intestinal juice that converts inactive trypsinogen into active trypsin
أَفَقَ الرّجل في العطاء	فَضّلَ وأعطى بعضاً أكثر من بعض	prefer	like better; value more highly
اثنولوجيا	علم الأعراق البشرية و هو فرع من الأنثروبولوجيا يبحث في أصول الشعوب المختلفة وخصائصها وتوزعها وعلاقاتها بعضها ببعض	Ethnology	the branch of anthropology that deals with the division of humankind into races and with their origins and distribution and distinctive characteristics
تَعَرَّضَ	الشيءَ وللشيء تصَدَّى له	Stand up	defend against attack or criticism
بكتيريا	كَائِن حَيّ مِجْهَرِيّ وحيد الخلية	monad	(biology) a single-celled microorganism (especially a flagellate protozoan)

تَوَقُّف	حدوث شئ مفاجئ يتسبب في عرقلة أو إعاقة سير شئ ما	blocking	the act of obstructing or deflecting someone's movements
استنفاد	شدة إنقاص مادة ما من مخزونها في الكائن الحي.	Depletion	the act of decreasing something markedly
الطائرة	كل واسطة للنقل مهيأة للارتفاع والتحليق في الفضاء.	craft	a vehicle designed for navigation in or on water or air or through outer space
أثَر	ما يترك الشيء أو الأمر وراءه يدل على مروره أو حصوله	trace	an indication that something has been present
دودة برنامج	دودة برنامج يكرر نفسه ولكنه يتكاثر في الشبكة بشكل مقصود بعكس الفيروسات و دود الانترنت الذي حدث عام 1988 ربما يكون الأشهر فقد استطاع الدود أن يتكاثر في أكثر من 6000 نظام	Worm	a software program capable of reproducing itself that can spread from one computer to the next over a network
أَمَرَ	ضد نَهَى، كَلَّفَ أحد ما أن يفعل شيئا ما	assign	give an assignment to (a person) to a post, or assign a task to (a person)
صدق	الفعل أو العمل كان مُخْلِصا فيه فلم يُقَصِّر او ينقلب	sincerely	with sincerity; without pretense
الفاشيّة	حركة وطنيّة سياسيّة في ايطاليا أنشأها موسوليني في مدينة ميلان عام 9191 م عمادها الشباب وهدفها مناهضة الشيوعيّة وحب النظام ومحاربة الفوضى والازدراء بالنظم البرلمانيّة	Fascism	a political theory advocating an authoritarian hierarchical government (as opposed to democracy or liberalism)
حَكَم	الذي يحكم بين اثنين أو أكثر، مثل القاضي إذا كان نِزاعا بين طرفين لا حدًا يقام	arbiter	someone chosen to judge and decide a disputed issue
مَارِس	إله الحرب في الأساطير الرُومِيَّة، وهو المَرِيِّخ، لا يؤنث ولا يجمع	Mars	(Roman mythology) Roman god of war and agriculture; father of Romulus and Remus; counterpart of Greek Ares
وَطَن	الوطن الأصلي ووطن الفطرة في الشرع هو محل مولد الإنسان،	Birthplace	the place where someone was born
ذاكرة فورية	دائرة التحكم تراقب البيانات المتنقلة بين CPU و الـ RAM و يوضع فيها شيء يقرأ أو يكتب فإذا طلبت نفس البيانات مرة أخرى تحصل عليها بسرعة عالية من الكاش و عندما تمتلئ تستبدل القديم بالجديد و تستخدم طريقة LRU لتحديد البيانات	cache	(computer science) RAM memory that is set aside as a specialized buffer storage that is continually updated; used to optimize data transfers between system elements with different characteristics
التجويف الحقي	تجويف مستدير في عظم الورك يستقر فيه رأس عظم الفخذ .	Acetabulum	the cup-shaped hollow in the hipbone into which the head of the femur fits to form a ball-and-socket joint
الوجود	تحقق الشيء في الذهن أو في الخارج ومنه الوجود المادي أو في التجربة والوجود العقلي أو المنطقي	actualization	making real or giving the appearance of reality

ظَهَّرَ	الفلم أو الصورة وضعها في الحامض الخاص كي يُظْهِرَها على ورق الصور الخاص، ويقال أيضا حَمَّضَ لأنها توضع في الحامض متعد حديث	develop	make visible by means of chemical solutions
مؤل	مكان أو مبيت طبيعي لنبات أو حيوان أو مجموعة خاصة من الكائنات المرتبطة ببعضها البعض بشكل وثيق	Habitat	the type of environment in which an organism or group normally lives or occurs
شعب	الناس الذين يسكنون دولة ما وإن لم يكن يربط بينهم نسب حديث والعامة قد تقول شَعَب	nation	the people who live in a nation or country
تَارِيْخ	هو تحديد الزمن باليَوْم والشَهْر والسَنَة أو ببعضهما الأجَل أو الوَقْت	date	the particular day, month, or year (usually according to the Gregorian calendar) that an event occurred
الايثرنت	أشهر الطُرق الشائعة لتبادل المعلومات بين أجهزة الكمبيوتر في الشبكة ويسمح الإثرنت للأفراد أن يتبادلوا الملفات ويتشاركوا الملحقات كالطابعة	ethernet	a type of networking technology for local area networks; coaxial cable carries radio frequency signals between computers at a rate of 10 megabits per second
صَفْرَاء	الغُدّة الصفراء غدة في الكبد، تسمى أيضا الغدّة الصفراوية والمَرَارَة	gallbladder	a muscular sac attached to the liver that secretes bile and stores it until needed for digestion
كتابة	ما يُكْتَبُ عموما	written	set down in writing in any of various ways
المجوس	قريب من المانوية ويذهبون إلى أن الله تعالى هو النور الأعلى وهو يزدان، وان الشيطان من جنس الظلمة وهو "أهرمن"	zoroastrianism	system of religion founded in Persia in the 6th century BC by Zoroaster; set forth in the Zend-Avesta; based on concept of struggle between light (good) and dark (evil)
حل الشراكة	هو انقضاء أجل العقد الذي تتحدد فيه مصالح المتعاقدين المالية والأدبية أو فسخه قبل حلول هذا الأجل وينحل عقد الشركة بانحلال رابطة المشاركة بين الشركاء.	Dissolution	the termination or disintegration of a relationship (between persons or nations)
أبر	لَدَغَ بالإبْرَة (كالعَقْرَب)	needle	prick with a needle
لَبِسَ	ارتدى ملابسه	get dressed	put on clothes
لطيف	ذو اللطف وذو اللطافة	nice	pleasant or pleasing or agreeable in nature or appearance
حَكَمَ	الدولةَ أو البلادَ كان أميرا أو سيدًا عليها يقوم على أمورها ويقضي بين الناس فيها	Rule	exercise authority over; as of nations

الثمن	نسبة سلعة إلى النقد عند المبادلة وثمن السوق أو الثمن الجاري هو الذي يتعين في المنافسة الحرّة بنقطة توازن العرض والطلب	Price	the property of having material worth (often indicated by the amount of money something would bring if sold)
ثَابِت	المستقر الذي لا يتغير، ضد المُتَغَيِّر	Static	showing little if any change
أيم	من ليس له زوج، يستخدم للرجل والمرأة، إلا أنه أكثر ما يستخدم لمن كان متزوجًا ثم خلا بموت أو طلاق	single	not married or related to the unmarried state
خَبِيْر	الخَابِر الذي يعرف خَبَر الشيء	knowledgeable	alert and fully informed
شقة	نِصْف الشيء إذا شَقَّ	half	one of two equal parts of a divisible whole
نَجْمَة	في علم الهَنْدَسَة هي شكل مغلق يتكون من خمسة زوايا بارزة أو أكثر ومثلها غير بارزةحديث	Star	a plane figure with 5 or more points; often used as an emblem
عوامل بِيْئَة	هي العوامل التي تسيطر على أحوال البيئة الطبيعية	Environmental Condition	the state of the environment
طَوَّقَ	وَضَعَ عليه طَوْقا	Collar	furnish with a collar
هَرِيْس	اسم طعام مصنوع من القَمْح المدقوق بقوة مخلوطا باللحم	Mash	a mixture of mashed malt grains and hot water; used in brewing
استشارة	رأي يقدّم من خبير في مسألة يعينها القاضي.	consultation	a conference (usually with someone important)
الأبّ	والد الابن على قول من اجاز التشديد	Father	a male parent (also used as a term of address to your father)
تَفَتَّحَ	فلانٌ في الكلام جاهر به مفتخرًا بما عنده من العلم والأدب	Orate	talk pompously
دفتر الأوراق التجارية	دفتر تسجل فيه أوراق القبض وأوراق الدفع وتذكر فيه تواريخ استحقاق الأوراق التجارية التي يكون التاجر دائناً أو مديناً بمبالغها.	account book	a record in which commercial accounts are recorded
استوقد	النارُ هاجت وارتفعت نارها	ignite	cause to start burning; subject to fire or great heat
تَبَحَّثَ	يستخدم بمعنى بَحَثَ بجدّ بغرض التوكيد، خصوصا عند قصد البحث بمعنى الدراسة	research	attempt to find out in a systematically and scientific manner
رخصة	فعل يبيحه القانون ويطلقه.	Permit	the act of giving a formal (usually written) authorization
صُومَال	دولة عربية تقع في القَرْن الإفْرِيقِي	Somalia	a republic in extreme eastern Africa on the Somali peninsula; subject to tribal warfare
عَجِيْن	ما عُجِنَ من الدَقِيْق أو الطَحِيْن بالماء، فَعِيل بمعنى مَفْعُول	pastry	a dough of flour and water and shortening

احتاط	اتخذ الوسائل لحائجة غير حاضرة ولكن هناك احتمال لحدوثها	anticipate	regard something as probable or likely
تفرع	عملية تؤدي إلى تفرع أو خروج الكمبيوتر عن التسلسل الطبيعي لتنفيذ برنامج	Branching	the act of branching out or dividing into branches
اِسْتَعْرَضَ	فلانا قال له أعرض عليّ ما عندك	show	show or demonstrate something to an interested audience
قَرْن	الذؤابة من شعر النساء حين ترفع وتربط من خلف عند الرأس؛ وقد تكون اثنتين على طرفي الرأس	ponytail	a hair style that draws the hair back so that it hangs down in back of the head like a pony's tail
سدارة	غطاء للرأس كان يُلبس في العراق مع الزيّ الغربيّ يعادل الطَرْبُوش في مصر	kerchief	a square scarf that is folded into a triangle and worn over the head or about the neck
جَدَّدَ	الكلام أعاده من جديد	Repeat	to say, state, or perform again
آذَار	الشهر الثالث في السنة الشمسية بعد شُبَاط وقبل نَيْسَان	March	the month following February and preceding April
شغف	الشيءُ أو الشخصُ فلانًا جعله يُوْلِعُ به	love	have a great affection or liking for
حلم اليقضة	تفكير متمنى يحقق من خلاله النائم إشباعا لرغباته ودوافعه التي يعجز عن تحقيقها في الواقع	daydream	have dreamlike musings or fantasies while awake
أصحاب الشأن	شخص او منظمة، له او لها مصلحة مشروعة في مشروع او كيان ما، او قابل للتأثر بإجراء او بسياسة ما	Stakeholder	someone entrusted to hold the stakes for two or more persons betting against one another; must deliver the stakes to the winner
لَيْلَة	واحِدَة اللَّيْل، وهو الزمن في اليوم من مَغْرِب الشمس حتى طلوع الفَجْر	night	the time after sunset and before sunrise while it is dark outside
الصّرد	الخطأ في الرمي	miss	a failure to hit (or meet or find etc)
شرط عدم الضمان	شرط يقضي بإعفاء البائع من التزام الضمان.	condition	an assumption on which rests the validity or effect of something else
قَصْر	مبنى كبير واسع فخم يستخدم لبعض الأعمال العامة سمي كذلك تشبيها له بالقصور التي يسكنها الخاصة	Palace	a large and stately mansion
تاع الماء	سال على الأرض	hush	run water over the ground to erode (soil), revealing the underlying strata and valuable minerals
حَمَّرَ	الشيء صبغه بالأحمر	Redden	make red
قَدَّرَ	الله الأمر قضاه وحكم به في ميعاد وقَدَر	predestine	foreordain by divine will or decree
الاعتراف	هو إقرار من شخص طبيعي أو معنوي بأمر معين.	Admission	an acknowledgment of the truth of something

الحق المساطحة	حق عينى يعطي صاحبه الحق في إقامة بناء أوغراس على أرض إلغير.	deed	a legal document signed and sealed and delivered to effect a transfer of property and to show the legal right to possess it
مَلِكَة	الأنثى الوحيدة الخصبة في مستعمرة من الحشرات الاجتماعية مثل النمل والنحل والنمل الأبيض ، وتتمثل مهمتها في وضع البيض	Queen	the only fertile female in a colony of social insects such as bees and ants and termites; its function is to lay eggs
بالوعة	عملية أو أنشملة أو آلية تزيل غازات الدفيئة أو الأهباه الجوية أو سلائف غازات الدفيئة من الغلاف الجوي	Sink	(technology) a process that acts to absorb or remove energy or a substance from a system
جهاز ادخال	مدخل يمكنك استخدامه لوصل اجهزة الادخال إلى الحاسوب مثل لوحة المفاتيح أو الفأرة أو الميكرفون ....الخ	INPUT Device	a device that can be used to insert data into a computer or other computational device
الشرط عدم الضمان	شرط يقضي بإعفاء البائع من التزام الضمان.	condition	an assumption on which rests the validity or effect of something else
مَنْزِل	الدار التي فيها بضعة بيوت، المبنى فيه أكثر من بيت للعوائل	House	a dwelling that serves as living quarters for one or more families
مِقْصَف	ما يشبه المطعم إلا أن الخدمة فيه قليلة والطعام عادي غير مزين، يكثر عادة في الأماكن التي يحتاج فيها الناس إلى الطعام بعجلة مثل المدارس وأماكن العمل حديث	cafeteria	a restaurant where you serve yourself and pay a cashier
الساعة الرملية	الساعة الرملية قَنِّينَتَان من زجاج يتصل رأس كل منهما بالأخرى وبينهما ثقب في أحدهما رمل يتسرب الرمل من الثقب ويستغرق سَاعَة كاملة للمرور	sandglass	timepiece in which the passage of time is indicated by the flow of sand from one transparent container to another through a narrow passage
بَدَوِيّ	الذي يعيش في البَادِيَة؛ الأعرابيّ	bedouin	a member of a nomadic tribe of Arabs
أكل	وضع الطَعَام في فمه ثم مضغه ثم بلع؛ تَنَاوَلَه	eat	take in solid food
سَهْل	ضد الصَعْب، اليَسِير لا جمع له	simple	easy and not involved or complicated
الشّك	نقيض اليقين	doubt	uncertainty about the truth or factuality of existence of something
الالتماس المنقول	هو انضمام او تمازج عينين منقولتين يملكهما مالكان مختلفان أو هو تحويل شيء عن وضعه الأول إلى وضع آخر بالاستصناع.	transfer	the act of transfering something from one form to another
متعدد الوسائط	يحتوي على وسائل الاتصال الصوتي ونقل الصورة الثابتة أو متعدد الوسائط المتحركة والوثائق المطبوعة أو أية تشكيلة منها	Multimedia	transmission that combine media of communication (text and graphics and sound etc.)

صَوْت	هو كناية عن ما يجب الأخذ بنظر الاعتبار	Vote	the opinion of a group as determined by voting
الطّبّ	السحر والمداوي من السحر وغيره	witchery	the art of sorcery
نمطي	نص يستخدم لجميع الأغراض حيث يطبع مرة واحدة،ويخزن ثم يعاد استخدامه مرة ثانية في المستندات الأخرى فمثلاً ،يقوم الشخص المتلاعب بصياغة رسالة غرام نموذجية ثم يرسلها إلى كل فتاة جديدة يقابلها وكل ما يفعله هو كتابة اسم المرسل إليها فقط	Boilerplate	standard formulations uniformly found in certain types of legal documents or news stories
نَصْل	حديدة السهم والرمح والسيف والسكين وغيرها من الأسلحة الحادّة الجارحة	Blade	the flat part of a tool or weapon that (usually) has a cutting edge
رَبَعَ	بالمكان أقام واطمأن	reside	live (in a certain place)
سَاعَة	ساعة جيب هي نوع من الساعات صغيرة توضع في الجيب ولها سلسلة تربط بالملابس	pocket watch	a watch that is carried in a small watch pocket
حُبُوب	طعام الفطور المعد من الحبوب	cereal	a breakfast food prepared from grain
عزيز	القوي المنيع، ضد الذليل	firm	strong and sure
انحل	المشكلة ظهر لها حل، انقضت	resolution	finding a solution to a problem
سِجِلّ	في قانون أو فقه كتابٌ يكتب به كاتب القاضي صورة الدعاوى والحكم فيها وصكوك المبايعات ونحوها لتبقى محفوظةً عندهُ	registry	an official written record of names or events or transactions
ظَهْر	خلاف البَطْن، الجزء الظَاهِر من الشيئ وغير البَاطِن	Front	the side that is seen or that goes first
مُسْتَطِير	مُتَفَرّق ومُنْتَشِرَ كتفرق وانتشار الطَيْر في السماء	spread	become distributed or widespread
نواكشوط	عاصمة موريتانيا. مدينة موريتانيا.	Nouakchott	capital of Mauritania; located in western Mauritania near the Atlantic coast
مقصف	ما يشبه المطعم إلا أن الخدمة فيه قليلة والطعام عادي غير مزين، يكثر عادة في الأماكن التي يحتاج فيها الناس إلى الطعام بعجلة مثل المدارس وأماكن العمل حديث	cafeteria	a restaurant where you serve yourself and pay a cashier
تَعاطُف	مشاركة أحاسيس الآخرين	sympathy	sharing the feelings of others (especially feelings of sorrow or anguish)
دَقِيْق	الأمر الغَامِض ذو الخفايا	vague	lacking clarity or distinctness
ثَعْلَبَة	اسم لمَرَض يصيب فروة الرأس ويسبب سقوط الشعر والصَلَع	Alopecia	loss of hair (especially on the head) or loss of wool or feathers; in humans it can result from heredity or hormonal imbalance or certain diseases or drugs and treatments (chemotherapy for cancer)

بَنْشَرَ	هي الفعل من أن تُثْقَب عَجَلَة السَيَّارَة أو الدَرَّاجَة أو المركبة	puncture	the act of puncturing or perforating