Research paper   Open Access

Building a Bilingual Ontology Arabic-English by Protégé

Ahmed R. Elmahalawy*, Mostafa M. Aref**, A.A.Soliman***

*(Mathematics Department, Faculty of Science, Benha University, and Benha, Egypt
Email: aboreda92@gmail.com
** (Computer Science Department, Faculty of Computer and Information Sciences, Ain Shams University, and Cairo, Egypt
Email: aref_99@yahoo.com
***(Mathematics Department, Faculty of Science, Benha University, and Benha, Egypt
Email: a_a_soliman@hotmail.com

Abstract

Ontology is defined as the specification of concepts. Bilingual ontology describes the specifications of concepts in two languages. This paper proposes a description of the bilingual ontology (Arabic-English) by using a Protégé. This description contains the hierarchy of Arabic-English concepts such as noun and verb. The hierarchy is presented as an ontology graph of Arabic-English noun and verb concepts.

Keywords: Ontology, Bilingual Ontology (Arabic-English), Protégé, Hierarchy of noun and verb concepts.

1. INTRODUCTION

In recent years, a lot of research focus on ontology. Ontology is defined as the specification of concepts. Linguistics and philosophies were concern with Ontology for decades. In computer science, the topic Ontology gains a lot of attention. There are many applications such as semantic web, machine translation, and semantic representation are using Ontology. This paper presents a building of bilingual ontology (Arabic-English) by using Protégé. This bilingual ontology would be very useful in machine translation. Machine translation has a long history, complex somewhat [11]. Much has been said about translation as being one of the most efficacious. It is the only means of communication between different communities or different cultures. The concept of translation has been existed hundred years ago [10]. In the twentieth century, it has become a realistic fact, in the form of computer software's able to translate a wide range of texts from one natural language to any other languages [11]. But, during the second half of the twentieth century that it has use as an independent academic discipline called Translation Studies and taught at most local and international universities.

The paper is organized as follows: Section (II) gives a background on ontology, bilingual ontology and Machine Translation. Section (III) gives a more related work of bilingual ontology and machine translation. Section (IV) Building the bilingual (Arabic-English) ontology by using Protégé program to define a hierarchy of noun and verb concepts. From the noun and verb concepts we build the hierarchy. Section (V) gives a conclusion and a future work.

2. BACKGROUND

Machine translation is defined as an automated translation. This translation implemented by utilizing a computer software to transform a text from a naturalistic language (such as Arabic) to another language (such as English) without any human involution. The machine translation process is shown in Fig (1 [1]. One of the difficulties of machine translation is word ambiguous. The word is an ambiguous, if it can have one or more than meaning. For instance, the concept of "bank" which has two meaning in English the edge of a river, or a financial institution.

Fig (1): Machine Translation Process

Ontology is applied in the context of software and database engineering, yet it has a theoretical grounding as well. Ontology gives details of a range of words with which to make statements, which may be inputs or outputs of knowledge agents (such as a software program).

This ontology explains as an "explicit designation of a conceptualization". It is, "the things, ideas of a quality common to a group, other things that are took as true to

have existence in some area of interest and the relations that contract among them." While the expressions designation and conceptualization have caused much discussion, the most important points of these statements of ontology are:

1. The expression of ontology explains (designates) the ideas of a quality common to a group, relations, and other differences that are relevant for modelling a domain.

2. The designation takes the shape of the definitions of representational lexical (layers, cognitions, and others), which supply meanings for the lexicon and official restrictions on its cohesive use [2].

There are different numbers of people who talk in other languages outside houses non English language: not just those of us who have learned other languages in school or through living in other countries, but additionally employers who have learned enough Spanish to talk to their workers; employees in hospitals, law courts and do trade with general public stores who have picked up components of another language to make their jobs more comfortable; soldiers back from Afghanistan or Egypt with some something that a person is good in Arabic.

Bilingual is the most general expressions that are utilized when we speak about people who speak two languages. For example, a bilingual person might talk Arabic and English or any other two languages. How we make ability to speak two languages mostly depends on the person who works to find information and his make observations in the form of questions, or the policy maker and his statutory policy [3].

The word of Bilingual is divided into two parts: the first part is "Bi" which means (having two) and the second part is lingual which means (language), thus bilingual which means (having two languages). Bilingual is as well a noun, and a person can be called a bilingual, such as in the South American country like Canada, where the official languages are French and English, and where many of the citizens are bilingual [4].

3. RELATED WORK

There are many related work depend on machine translation, ontology, corpus based and bilingual ontology.

In [5], the authors give a detail about the semi-automatic process of associating a Japanese word list with semantic concept taxonomy are called ontology, utilizing an English-Japanese bilingual dictionary. This problem focuses on how to connect the Japanese lexical things with the concepts in the ontology by automatic ways, so it is also hard to know many concepts manually. They have prepared three algorithms to connect the Japanese lexical things with the concepts such as: the equivalent-word match, the argument match, and the example match.

In [6], the researcher describes an alignment system that aligns Malayalam - English texts at word level in parallel sentences. A parallel corpus is a combination of texts in two various languages, one of whom language is translated to tantamount of the second language. So, the prime objective of this method is to construct word-aligned parallel corpus to be utilized in Malayalam and English machine translation (MT).

In [7], the authors developed the paper in [6]. Parallel corpus are assist in to create the statistical bilingual dictionary, in backing statistical machine translation and also in supporting as traineeship data for word meaning and translation disambiguation. Furthermore, the presentation of this approach can too be progressed by utilizing a listing of equations and morphological analysis.

In [8], the researchers describe the methodology to know the parallel Hindi-English sentences by utilizing a word alignment. This methodology is basis to improve the parallel Hindi-English word dictionary after syntactically and semantic analysis of the original text from Hindi-English. Develop this methodology is depend on two ways to solve this problem. The first way: is normalization of tagged Hindi-English sentences. The second way: is a mapping of Hindi-English sentence by utilizing parallel Hindi-English word dictionary.

4. BUILDING OF BILINGUAL ONTOLOGY

In this section the description of bilingual ontology is going to be built by a Protégé. To create a bilingual ontology by using a free, open source program is called a Protégé. The Protégé program will be found ready on the website [9]. After download and install a Protégé program on my computer, then we use the Protégé to create a new otology.

Protégé has the similar constituents of OWL (Ontology Web Language) such as classes, properties, individuals. There are three steps to create a new ontology as the following:

1. Start Menu and open Protégé.
2. Create an ontology (URI) to upload the ontology in my website.
3. Save a file on my computer with a file extension is (.owl).

The main building of ontology in Protégé are classes. The empty ontology contains a class one called Thing as shown in Figure 2.

Create a new owl classes as the following:

1. Choose the classes from Protégé as showing in Figure 2.
2. Choose add subclass from Figure 3.
3. Write a class name in the box as showing in Figure 4.
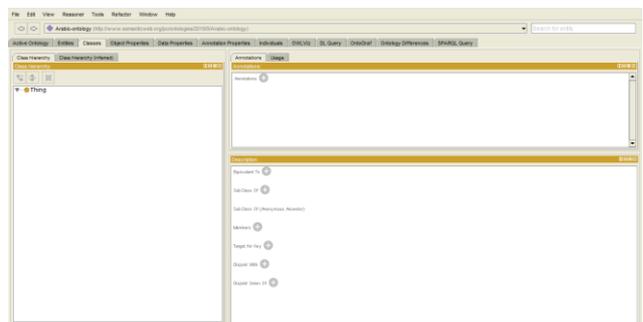4. Repeat the last three steps to add more than classes.
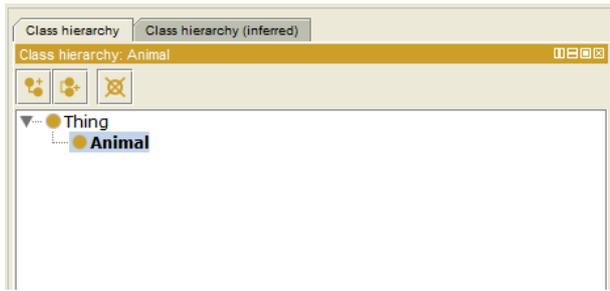

Fig (2): Empty Ontology
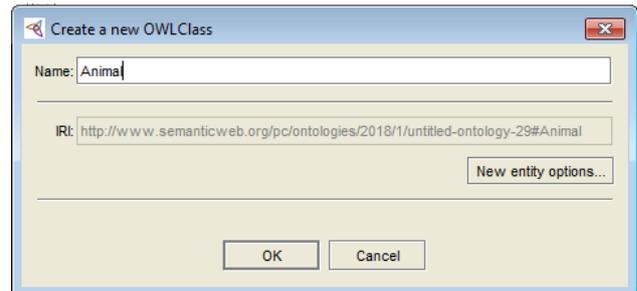
Fig (3): Add Subclasses

Fig (4)**:** Dialog Box

Now, let's add some classes to subclasses of the class Thing to create Arabic and English ontology. We choose some words to create an Arabic-English ontology. The hierarchy of Arabic-English as shown in Figure 5.

The visualisation of the English and Arabic ontology by using the OntoGraph tab of the Protégé program to create a hierarchy of Arabic-English ontology. The graph of Arabic-English ontology as shown in two figures Figure 6, Figure 7 respectively.
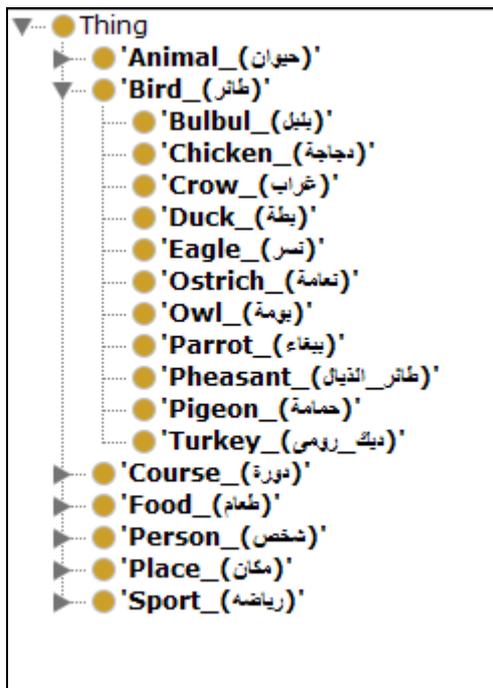
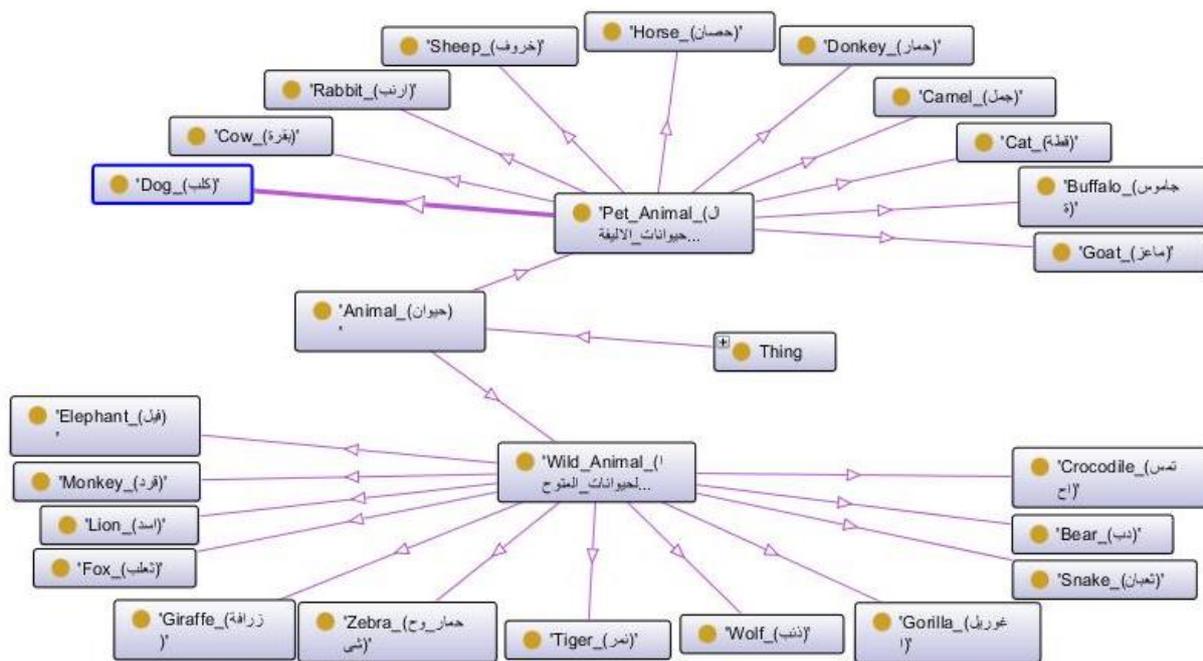Fig (5)**:** Hierarchy of Arabic-English Ontology
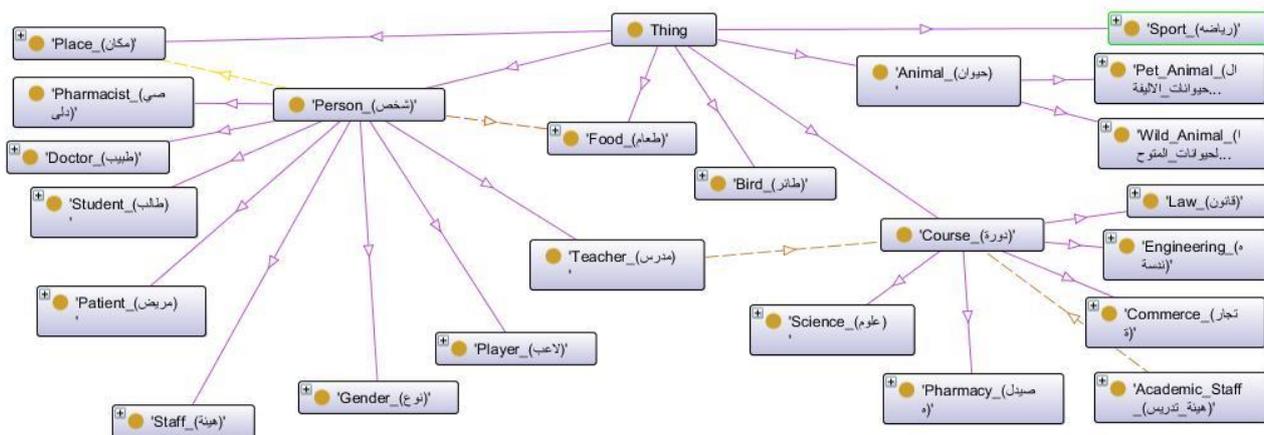
Fig (6)**:** Graph of Arabic-English ontology

Fig (7)**:** Graph of Arabic-English ontology

Finally, From Arabic and English Protégé that hold more information as illustrated in Table 1.

Table (1): Information about Arabic-English Protégé

	Classes	Nodes	Edges
Arabic-English Protégé	260	252	255

5. CONCLUSIONS

In this paper, we presented a description of the bilingual (Arabic-English) ontology by using a protégé. The description contains a hierarchy of Arabic-English concepts by using noun and verb. This hierarchy is given

in a form of a graph of Arabic-English ontology for noun and verb. Illustrate the result about the graph of bilingual

ontology such as number of classes, nodes, and edges is given in a table.

REFERENCES

[1] C. Stern and A. Dufournet. What is machine translation? systran translation technologies, Machine translation, http://www.systransoft.com/systran/translation-technology/what-is-machine-translation/, August 2011. (Accessed 22-October-2015).

[2] T. Gruber. Ontology (computer science) - definition in encyclopedia of database systems. Ontology, http://tomgruber.org/writing/ontology-definition-2007.htm, September 2007. (Accessed 14-October-2015).

[3] N. Takaya. What do we mean when we say bilingual? — psychology in action. Bilingual, http://www.psychologyinaction.org/2012/01/17/what-do-we-mean-when-we-say-bilingual/, January 2012. (Accessed 31-October-2015).

[4] I. Thinkmap. bilingual–dictionary definition: Vocabulary.com. Bilingual,

[9] Protégé. https://protege.stanford.edu/, November 1999. (Acessed 15-January-2018).

[10] A. Jebbar. Machinetranslation vs human translation. Machine translation, http://www.translationdirectory.com/

http://www.vocabulary.com/dictionary/bilingual, June 2013. (Accessed 31-October-2015).

[5] A. Okumura, E. Hovy. Building Japanese-English Dictionary based on Ontology for Machine Translation. In proceedings of ARPA Workshop on Human Language Technology, pages 236-241, 1994.

[6] K. T. Nwet. Building Bilingual Corpus based on Hybrid Approach for Myanmar - English Machine Translation. International Journal of Scientific & Engineering Research, 2 (9), 2011.

[7] K. T. Nwet, K. M. Soe, and N. L. Thein. Developing Word-aligned Myanmar-English Parallel Corpus based on the IBM Models. International Journal of Computer Applications, 27(8), 2011.

[8] S. Dubey, T. D. Diwan. Supporting Large English-Hindi Parallel Corpus using Word Alignment. International Journal of Computer Applications, 49, No.6, (7), 2012. articles/article1326.php, 2002. (Acessed 15-November-2015).

[11] A. Yanishevsky. The emerging role of machine translation. Local- isation guide, pages 12–13, 2009.