

A META DESCRIPTION LOGICS KNOWLEDGE BASE FOR ARABIC LANGUAGE PROCESSING

Yasser YAHIAOUI¹, Ahmed LEHIRECH²

1. Centre Universitaire de Nâama, 2. Université Djillali Lyebes SBA.

1. alyahiaoui@gmail.com, 2. elhir@yahoo.com

ABSTRACT

Description Logics (DLs) are successful knowledge representation formalisms, which can be used to represent the terminological knowledge of an application domain in a structured and formally well-understood way. They are employed in various application domains, such as natural language processing, configuration, and databases .Sf.

In this work we try to use this performance of DL systems to describe a Meta knowledge base defining the classes of the word in Arabic language and relation between them. This can be useful for syntactic categorisation of sentences which is very important for automatic language processing.

KEYWORDS

Description logics, Arabic language processing, knowledge representation, Meta knowledge base.

INTRODUCTION

From the linguistic point of view, the language processing involves distinct levels namely: the lexical level, syntactic level, the semantic and pragmatic level. All these levels contribute to the representation of semantic

information. Thus, the role of lexicology becomes very productive since the meaningful representation of the lexicon is provided, and this helps the syntactic and semantic rules to become easier. Dixon 1991 in his work on syntax-based semantics, has proved that the irregularities and idiosyncrasies can be predicted from the semantics of words [01].

This aspect of idiosyncrasy (meaning non-regular behaviour) leads us to think of a heuristic treatment in order to manipulate these syntactic elements which can drive to a kind of classification. The latter facilitates the isolation and segmentation of the sentences in a text according to their meaning as a criterion of this classification.

The powerful knowledge representation formalism of description logics is providing an expressive tool which can be useful to create a Meta-knowledge base. This part of the knowledge base can be used as an add-on for ALP systems. It can also be considered as a tool for searchers on the Arabic language processing that provides a description for the totality of the Arabic words.

The description can be summarized in creating hierarchy classes of the Arabic word that provides a dependency graph based on the

subsumption relationship, and in this way we define the first part of the KB which we called, at this level, the Meta-Box and by the totality of words existing in the Arabic Word-Net ontology we define the second part including a terminological and assertional Box.

MOTIVATION

In the Beginning, it is important to understand the notion of the natural language processing (NLP) which is defined in literature as: “[...] a theoretically motivated range of computational technique for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications”[02]. This definition leads us to think about the levels of representation and especially on the syntactic categorisation which can carry the roles and definitions to construct meaningful representation of each word within the language.

For the Arabic language, the word has a role in a sentence and this role can be detected by the classification of the word syntactically, the fact which guides us to think about well-understood way of representations. Since the description logics are regarded as a structured and formally well-understood way that uses representation languages as the KL-One; “Since the days of the KL-One system, one of the main application of description logic

has been for the semantic interpretation in natural language processing” [03].

We aim in this representation at facilitating the semantic interpretation inspired, in this way, by the fact that “semantic interpretation is the derivation process from the syntactic analysis of utterance to its logical form”[03]. The lexical part becomes more and more important because we can begin the semantic integration from the lexical level by introducing a lexical semantic which means the specification of the semantic of each concept.

The use of DLs formalism to describe a Meta knowledge base using the syntactic roles of the Arabic word aims at constituting the lexical semantic part of the knowledge base which is not newly utilized; since we find, for instance, in the literature “a part of the knowledge base constitutes the lexical semantic knowledge, relating words and their syntactic properties to concept structures [...], giving a deep meaning to concept” [03].

THE ARABIC LANGUAGE PROCESSING (ALP) SYSTEM ARCHITECTURE

The architecture we propose here is basically constituted of two parts, the first one contains the high level knowledge giving the lexical and syntactical semantics. The second one gives the ordinary description of concepts (words) constructed as a KB built from the Word-Net Arabic ontology.

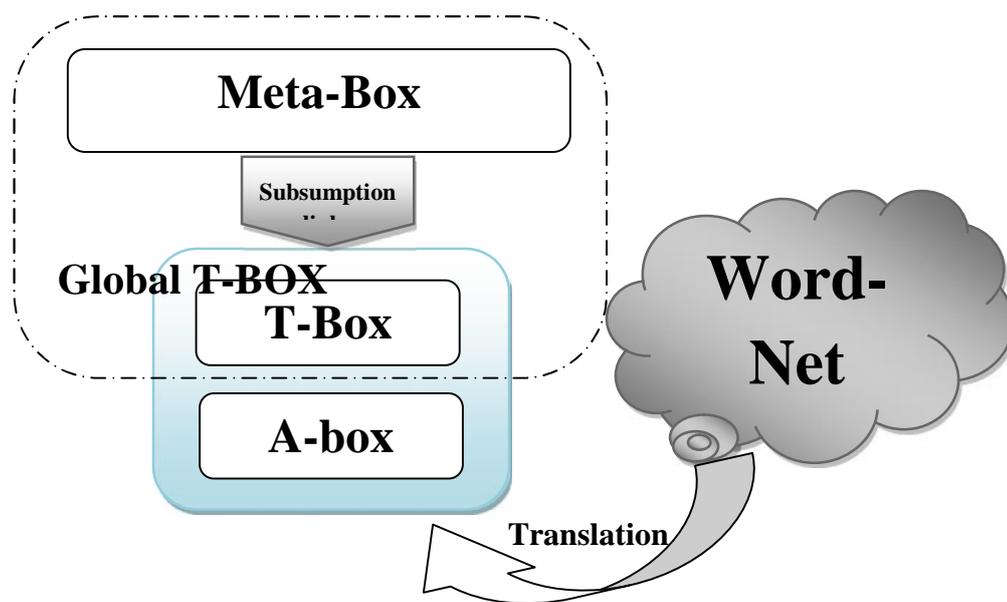


Fig 01: System Architecture

In the traditional DLs systems “A knowledge base (KB) comprises two components, the TBox and the ABox. The TBox introduces the terminology, i.e., the vocabulary of an application domain, while the ABox contains assertions about named individuals in terms of this vocabulary” [04].

However; the addition of the Meta-Box in the present system (fig 01) can be considered as a consolidation for the traditional system. This fact makes the introduction of the syntactic role of the term (concept) with its definition possible and it is implicitly defined in the class of the word which is integrated by establishment of the link between the Meta-Box and the T-Box using the relation of the subsumption. Since all the terms existing in Word-Net Arabic ontology can be subsumed by a concept among the Meta-Box concepts either directly or indirectly.

The T-Box / A-Box are traditionally constructed by the translation of definitions in the Ontology of concepts and their relations to DLs knowledge base defined using the KL-One description language.

THE WORD IN THE ARABIC LANGUAGE

The word in Arabic language can be viewed as an occurrence of a node of a dependency graph representing a hierarchical organization of the classes existing for the syntactic roles as following : the word is called in Arabic “Kalima”, this model includes the three nodes which are “Fiil” (verb) , “Ism” (noun) , and “harf” (propositions, conjunctions and so on). And all of these terms have dependency with others to create the hierarchy of concepts. (see fig 02)


```

kalima≡T
F'il ⊆ kalima
Ism ⊆ kalima
harf ⊆ kalima
F'il ≡ motassaref ∪ djamid
motasarref ≡ Tam ∪ nakis
tam ≡ motaaddi ∪ ellazim
nakis ≡ istemrar ∪ elmokaraba
djamid ≡ amr ∪ madhi
nakis d ⊆ madhi
choroè ⊆ madhi
madeh ⊆ madhi
eddam ⊆ madhi
etaâdjoub ⊆ madhi
F'il ≡ ¬Ism ∩ ¬harf ≡ ¬(Ism ∪ harf)
Ism ≡ ¬F'il ∩ ¬harf ≡ ¬(F'il ∪ harf)
harf ≡ ¬Ism ∩ ¬F'il ≡ ¬(Ism ∪ F'il)
Ism ≡ ism f'il ∪ ichara ∪ alla ∪ âalam ∪ maoussoul ∪ dhamir ∪ achiae ∪ mouchtak
âalam ≡ arabi ∪ aâdjami
maoussoul ≡ âakil ∪ ghiralâakil
dhamir ≡ mouttassil ∪ mounfasil ∪ moustatir
mouchtakat ≡ makan ∪ sifa ∪ f'ail ∪ maf'oul ∪ moutlak
chaie ≡ djamad ∪ hayaouane ∪ insane
harf ≡ ouhadi ∪ thounaie ∪ tholathi ∪ roba'î ∪ khomassi
Ism ≡ ∃do.F'il ∪ ∃doenwith.F'il ∪ ∃haskhabar.Ism
elmotâddi ≡ ∃hasmaf'oul.Ism ∪ ∃has2maf'oul.Ism ∪ ∃has3maf'oul.Ism
F'il ≡ ∃hasf'ail.Ism

```

fig 03: A Sample of the Meta-Box Described in KL-One

CONSTRUCTION OF THE SECOND LEVEL

The relations existing in the linguistic Word-Net Arabic are known and limited in an ultimate number. We find eleven (11) relations namely: Antonym, synonym, Meronym (inverse of holonym), hyponym, implication (entailment), causality, value, has as value, affinity, derived from, similar to. All of these relations can be interpreted By DLs language using the known constructors and quantifiers and we propose these interpretations as following: [6, 7, 8]:

Antonym: in DLs it becomes disjunction

Synonym: in DLs we replace it by equivalent

Meronym: interpreted as a restriction or role.

Hyponym: X is a hyponym Y interpreted X is subset of Y or X is subsumed by Y.

Implication: Y imply X means logically disjunction between negation(Y) and X (Neg(Y) ∪ X)

Causality: using roles

Value: assertion

Has As Value: assertions

Affinity: restrictions

Derived From: Y derived from X means logically disjunction between Negation (X) and Y (Neg(x) ∪ Y).

Similar To: equivalence

We note here, in this second level, that we have to define two parts: the first is terminological and the second is assertional but we can say that there is causality between the interpretations of the relation existing and the position of each possible word defined in Word-Net Arabic in each one of the two parts. And this is clear because the word is well

defined if we can represent all its relations linking the current word with all the others, in another way, in DLs terminology the concept is well defined if we reach a level of representation in which all relationship with others concepts are represented that means the encapsulation of all the elements of the set of individuals concerned by the definition of the concept.

LINKS BETWEEN THE TWO LEVELS

The classes defined in the Meta-Box are the collection of what can an Arabic word means because it is a high level categorization of the terms and in this way, we can be sure that all concepts existing in the TBox of the second level are included in one at least of the bottom concepts of the Meta-Box and it can be considered as an assertion in this level; but globally, the Global Terminological Box GT-Box is divided into two part the Meta-Box and the T-Box. However; the Global Assertional Box is similar to the assertional Box (see fig 01). We mathematically represent all this as:

$$\forall C \in TBox, \exists D \in MetaBox; C \subseteq D$$

and we have

$$GTBox \equiv MetaBox \cup TBox$$

also

$$GKB \equiv GTBox \cup ABox$$

CONCLUSION AND PERSPECTIVES

In this work, we arrived to evaluate an idea which is new in the domain of Arabic language processing and especially for the DLs based systems. This idea consist to create an add-on which we named the Meta-Box for the purpose of getting richer lexicology integrating

syntactic roles with the definition of every word existing in the Arabic linguistic ontology named Word-Net.

As results we define a new architecture for ALP description logics based system with enlarged knowledge base in where we have the traditional representation consolidate with other knowledge integrated to enhance efficiency, expressivity and the use cases of this Global KB. This letter is designed to be a useful tool for all kinds of treatments and manipulations existing in the ALP domain as the question answer systems, the automatic translation, the text summarization, SF.

We propose to evaluate the GKB an automated sentences segmentation which can provide a lot of efficiency for the ALP applications especially for the case of automated translation and text segmentation.

BIBLIOGRAPHY

1. Dixon,; *"a new approach to English grammar on semantic principles"* OXFORD : Clarendon press (1991).
2. R. Brachman, R Bobrow, P. Cohen, J. Kovstad, B. Webber and W Woods,; *"Research in Natrel language understanding, annual report"*. Technical report 4274, Bolt Beranek and Newman, Cambredge, MA (USA), (1979).
3. E. Franconi, *"Natural Language Processing"*. Description Logic Handbook. pp.450~461 Cambredge University press (2003).
4. D. Nardi and R. J. Brachman, *"An introduction to description logics"* handbook of description logics Cambredge university press (2003).
5. Varol Akman; *"Modeling and using context"*, third international and interdisciplinary conference, context 2001, Dundee UK July 27-30 , (2001).
6. Miller G. A., Beckwidth, R., Fellbaum, C., Gross, D. and Miller, K. J.,. *"Introduction to WordNet: An On-line Lexical Database"*, International Journal of Lexicography, Vol 3, No.4 (Winter 1990), pp. 235-244. (1999)

7. Fellbaum (2004), WordNet 2.0. Available from:
<http://wordnet.princeton.edu/oldversions>
(accessed 14 August 2009).
8. Euro WordNet Web Site:
<http://www.illc.uva.nl/EuroWordNet/>, (accessed 1
August 2009)

Additional references

1. عبد الله بن عبد الرحمن العقيلي الهمداني المصري »
شرح ابن عقيل على ألفية ابن مالك « المحقق : محمد
محيي الدين عبد الحميد مكتبة مشكاة (2000)
2. عبيد ربه الشنقيطي «نظم الأجرومية» مطبعة عيسى
البابي الحلبي في مصر سنة (1344هـ)