

Multilingual Universal Word Explanation Generation from UNL Ontology

Khan Md. Anwarus Salam^{1,3} *Hiroshi Uchida*^{1,2} *Tetsuro Nishino*³

(1) UNDL Foundation, Tokyo, Japan.

(2) United Nation University, Tokyo, Japan.

(3) The University of Electro-Communications, Tokyo, Japan.

salamkhan@uec.ac.jp, uchida@undl.org, nishino@uec.ac.jp

ABSTRACT

To develop a common language, it is essential to have enough vocabulary to express all the concepts contained in all the world languages. Those vocabularies can only be developed by native speakers and should be defined by formal ways. Considering the situation, at this moment Universal Networking Language (UNL) is the best solution as the common language, and Universal Words (UWs) are the most promising candidates to represent all the world concepts in different languages. However, UWs itself are formal and not always to be understandable for human. To ensure every language speakers can create the correct UWs dictionary entry, we need to provide the explanation of UWs in different natural languages for humans. As there are millions of UWs, it is very expensive to manually build the UWs explanation in all natural languages. To solve this problem, this research proposes the way to auto generate the UWs explanation in UNL, using the property inheritance based on UW System. Using UNL DeConverter from that UNL the system can generate the explanation in more than 40 languages.

KEYWORDS : UNL; Ontology; Word Semantics; NLP;

1 Introduction

To break the language barrier we need to have an artificial common language. For developing such a common language, it is essential to have enough vocabulary to express all the concepts contained in all the world languages. Because, human can understand the dictionary entries by reading the explanation (or meaning) of concepts in natural language. However, those concept dictionaries in different languages can only be developed by native speakers. Those universal concepts should be defined by formal ways.

Considering the situation, at this moment Universal Networking Language (UNL) is the best solution and Universal Words (UWs) are the most promising candidates. UNL represents natural language sentences as a semantic network with hyper nodes. In this semantic network, nodes represent concepts and arcs represent relations between concepts. These concepts are referred as UWs. UWs themselves are formal but not always to be understandable by human. As human should provide the dictionary entries in different language, it is essential to have UWs explanation in different natural language. As there are millions of UWs, it is very expensive to manually build all the UWs explanation in all natural languages.

UNL Ontology is a semantic network with hyper nodes. It contains UW System which describes the hierarchy of the UWs in lattice structure, all possible semantic co-occurrence relations between each UWs and UWs definition in UNL. With the property inheritance based on UW System, possible relations between UWs can be deductively inferred from their upper UWs and this inference mechanism reduces the number of binary relation descriptions of the UNL Ontology. In

the topmost level UWs are divided into four categories: adverbial concept, attributive concept, nominal concept and predicative concept.

Since UNL Ontology provides the semantic background of each UWs, the goal of this research is to auto generate the UWs meaning from UNL Ontology. Current UNL ontology contains around 1466598 unique concepts or UWs. So the goal of this research is to auto generate the natural language explanation for all these UWs. UNL ontology is developed in general domain.

Beside UNL Ontology there are other popular lexical resources available in general domain like WordNet (Miller, 1995), EDR dictionary etc. However, from other general ontologies currently it is not possible to auto generate the explanation for the concepts in different languages. To generate such explanation automatically, this research has been inspired from the unique architectural design of UNL (Uchida et. al. 1999). As the UNL systems are successfully implemented and became available online recently, it is possible to utilize UNL architecture now. The original idea of auto generating the explanation for different concepts in different languages is very new. There is no other existing ontology available which attempted to auto generate the explanation in different languages from the ontology itself.

This research proposes the way to auto generate the UWs explanation in UNL from the semantic background provided by UNL Ontology. The system first discover a graph the SemanticWordMap, which contains all direct and deductively inferred relations for one particular UW from the UNL Ontology. Using UNL DeConverter from that UNL the system can generate the explanation in more than 40 languages. This auto generated explanation will help the human to understand the UWs meaning to provide their corresponding dictionary entries. So beside the general users, this system is useful for the UWs dictionary builders and the editors.

With the property inheritance based on UW System, the system converts SemanticWordMap relations into UNL graph using rule-based approach. Finally from this UNL graph, UNL DeConverter generates the UWs meaning in different natural languages.

2 BACKGROUND

2.1 Universal Networking Language (UNL)

UNL initiative was originally launched in 1996 as a project of the Institute of Advanced Studies of the United Nations University (UNU/IAS)¹. UNL was first introduced to public in 1999 (Uchida et. al. 1999). In 2001, the United Nation University set up the UNDL Foundation², to be responsible for the development and management of the UNL project. In 2005, a new technical manual of UNL was published (Uchida et. al. 2005), which defined UNL as an information and knowledge representation language for computer. UNL has all the components to represent knowledge described in natural languages. UWs constitute the vocabulary of UNL and each concept of natural languages has unique UW. A UW of UNL is defined in the following format:

<uw> ::= <headword>[<constraint list>]

Here, headword of a UW is an English expression which can be a word, a compound word, a phrase or a sentence. UWs are the basic elements for constructing one UNL expression of a

¹<http://www.ias.unu.edu/>

²<http://www.undl.org/>

sentence or a compound concept. So keys to the information in UNL database are UW. UWs are inter-linked with other UWs using “relations” to form the UNL expressions of sentences. These relations specify the role of each word in a sentence. Using “attributes” it can express the subjectivity of author. Currently, UWs are available for many languages such as Arabic, Bengali, Chinese, English, French, Indonesian, Italian, Japanese, Mongolian, Russian, Spanish, and so forth.

Each UWs are interlinked with each other through the UW System in the UNL Ontology. Master definitions for UWs describe all relations that a UW can hold. A minimum set of relations is used as constraints of UW for the purpose to make a UW distinguishable from sibling UWs.

2.2 UNL Ontology

UNL Ontology is a lattice structure where UWs are inter-connected through relations including hierarchical relations such as icl (a-kind-of) and iof (an-instance-of). UNL Ontology includes possible relations between UWs, UWs definition and UNL system hierarchy. In the UNL Ontology, all possible semantic co-occurrence relations, such as ‘agt’, ‘obj’, etc, between UWs are defined based on the UW System. Every possible semantic co-occurrence relation is defined between the two most general UWs in the hierarchy of the UW System that can have the relation. With the property inheritance characteristic of the UW System, possible relations between lower UWs are deductively inferred from their upper UWs and this inference mechanism reduces the number of binary relation descriptions of the UNL Ontology. In the topmost level UWs are divided into 4 categories adverbial concept, attributive concept, nominal concept and predicative concept.

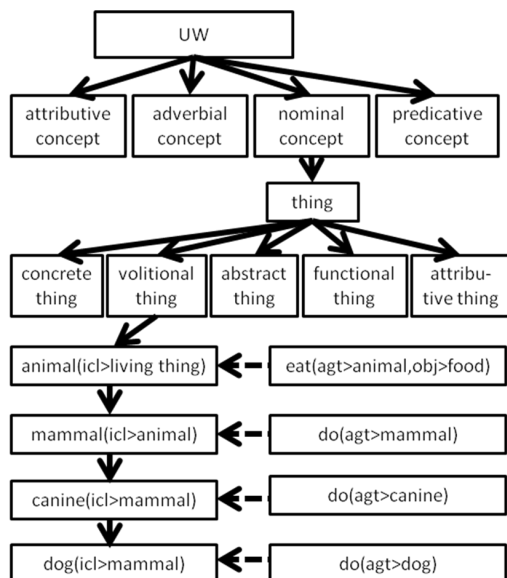


FIGURE 1 – UWs hierarchy in UNL Ontology.

Figure 1 shows the topmost level of partial UNL Ontology where the black directed lines represent “icl” relation and dotted directed lines represent “agt” relations. In Figure 1 we only

expanded partial “nominal concept” until “dog(icl>mammal)” to give a brief overview of the UNL Ontology. In UNL Ontology each UWs have incoming and outgoing relations with other UWs, which define the semantic background. For example in Figure 1 “animal(icl>living thing)” has two incoming relations, “agt” from “eat(agt>animal,obj>food)”, and “icl” from “volitional thing”. “animal(icl>living thing)” has only one outgoing relation “icl” to “mammal(icl>animal)”. As possible relations between lower UWs are deductively inferred from their upper UWs, we can infer that “mammal(icl>animal)”, “canine(icl>mammal)” and “dog (icl>mammal)” also has an incoming relation “agt” from “eat(agt>animal,obj>food)”.

2.3 UNL Explorer

UNL Explorer³ is a web based application, which combines all the components of UNL system to be accessible online. UNL Explorer users can translate the documents in various languages such as UNL, English, Japanese and Arabic etc. UNL Society members can add or edit information using UNL Explorer. It allows users to view the UNL Ontology which contains UWs hierarchy (a lattice structure) in a plain tree form. It can also display incoming and outgoing relationships for each UW.

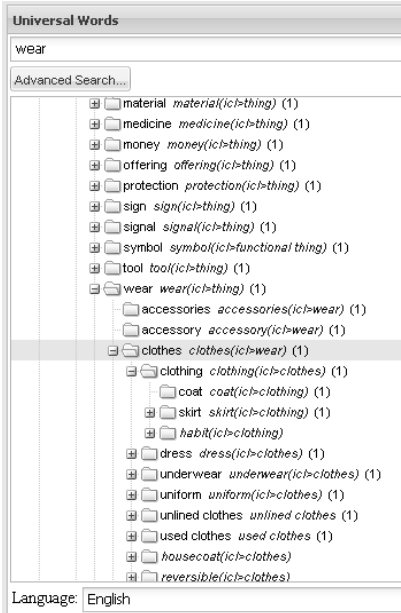


FIGURE 2 – UW search result for “wear” from UNL Ontology

UNL Explorer provides UNL Enconverter for natural language to UNL conversion. It also provides UNL Deconverter for UNL to natural language conversion. Both UNL EnConverter and Deconverter support different languages such as Chinese, English, Japanese and so forth. UNL

³<http://www.undl.org/unlexp/>

Explorer users can browse UNL Ontology from the Universal Words frame in the left side. Figure 2 shows sample UNL Ontology search result for the word “wear”.

UNL Explorer also provides an advanced search facility. Users can check incoming and outgoing relationships using this facility. Both UNL Ontology search mechanism is accessible for computer program using UNL Explorer API. However, to use this API, user need to be a UNL society member by signing an agreement with UNDL Foundation.

3 Multilingual Explanation Generation

The system framework for the multilingual explanation generation of the UWs is illustrated in Figure 3. The input of this system is one UW and the output of the system is the meaning of that UW in natural language such as English. For the given UW, the system first discover a SemanticWordMap, which contains all direct and deductively inferred relations for one particular UW from the UNL Ontology. So input of this step is one UW and output of this step is the WordMap graph. In next step we convert the WordMap graph into UNL using conversion rules. This conversion rules can generate “From UWs only” and “From UNL Ontology”, based on user’s requirement. So input of this step is the WordMap graph and Output is the UNL expression. In the final step we describe in natural language by converting the UNL expression using UNL DeConverter, provided by UNL Explorer.

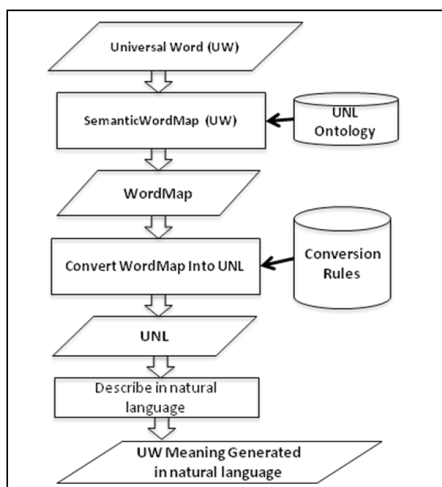


FIGURE 3 - System framework for multilingual explanation generation for UWs

3.1 SemanticWordMap

To discover inferred relationships, the system first discovers the SemanticWordMap (Salam et. el. 2011), which contains all direct and deductively inferred relations for one particular UW from the UNL Ontology. Edges of this graph are the relations of UNL Ontology. In UNL Ontology each relation is connected from “fromUW” to “toUW”. Starting from a given UW we discover the SemanticWordMap graph which includes deductively inferred relationships. A maximum search depth is established to limit the size of the graph.

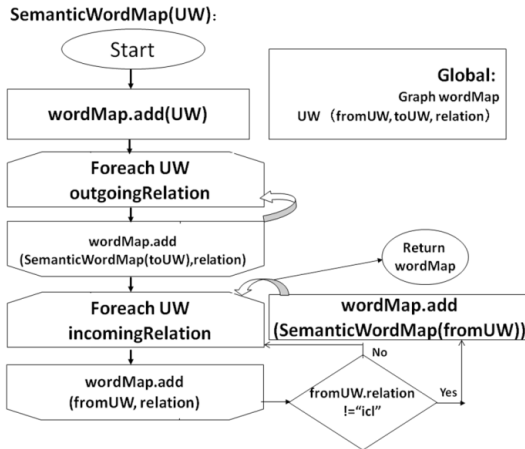


FIGURE 4- SemanticWordMap algorithm

To discover the SemanticWordMap graph from UNL Ontology user has to give a particular UW. First the algorithm adds that UW into the wordMap graph. For each outgoing relation from that UW, it add toUW into the wordMap and then recursively call SemanticWordMap(toUW) to discover the relations from toUW. Then for each incoming relationship it adds the fromUW with relation into the wordMap graph. If the relationship is not "icl", it adds the expanded graph by recursively calling SemanticWordMap(fromUW). As UNL Ontology contains a huge number of UWs and relationships, we have a heuristic approach to limit the SemanticWordMap graph to produce meaningful and specific information. So the algorithm keep discovering the graph until it reach maximum search depth or if it reach the topmost UW. Finally it returns the wordMap graph which contains all the UW relations.

For example, Figure 5 shows the partial SemanticWordMap for dog(icl>mammal). The output of this first step is the SemanticWordMap discovered from UNL Ontology. Here dotted arrows represent "agt" relations and black arrows are "icl" relations.

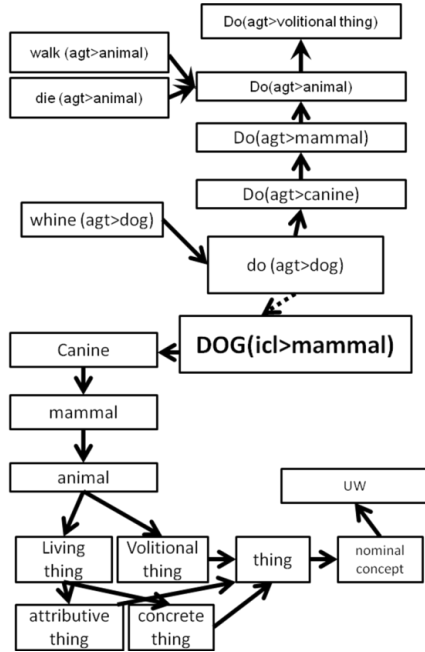


FIGURE 5- Partial SemanticWordMap for dog(icl>mammal)

3.2 Convert SemanticWordMap Into UNL

In this step we convert SemanticWordMap relations into UNL using some generalized rules. We first categorize the UWs into several categories such as “do”, “is-a”, “occur” and “be”. In general “do” categories represent actions, “is-a” represent features, “occur” represents changes and “be” represents status. Due to the property inheritance characteristic of the UNL Ontology, possible relations between lower UWs are deductively inferred from their upper UWs. Using SemanticWordMap we deductively infer the relationship with dog(icl>mammal). For example from Figure 3 we can say that UWs walk(agt>animal) and die(agt>animal) are related with dog(icl>mammal) as well.

TABLE I. CATEGORIZED RELATIONS FOR DOG(ICL>MAMMAL)

UW	Categorized from SemanticWordMap	
	UW Categories	Description
DOG (icl>mammal)	do	whine, walk, die....
	Is-a	canine, mammal, animal, ..

Table I shows the categorized relations from SemanticWordMap for the UW `dog(icl>mammal)`, and the generated description categorized into several UW relationship types. Steven Pinker pointed out that there are specified connections between verbs and object types in (Pinker, 2007). In this direction, we have manually identified such rules. After categorization we can convert the relations into UNL expression using different rules for each category. All these rules are currently designed by human.

After categorization we can convert the relations into UNL expression using different rules for each category. For example Figure 6 shows UWs relation derived from SemanticWordMap. To convert these category UWs relations into UNL, we use the following “Rule 1”:

(Rule 1: do) If (isaKindof(UW2, "do")) agt(UW3:08.@entry.@ability,UW1:00.@topic)

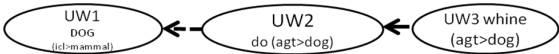


FIGURE 6: do relations derived from SemanticWordMap

Rule 1 check whether UW2 is related with “do” by using “isaKindof(UW2, ”do”)”. For example if `isaKindof(“do (agt>dog)”, ”do”)` = TRUE, then the generated UNL is: `agt(whine(agt>dog):08.@entry.@ability,dog(icl>mammal):00.@topic)`

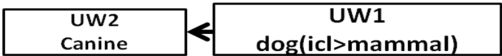


FIGURE 7: icl relations derived from SemanticWordMap

Figure 7 shows “icl” relation derived from SemanticWordMap. To convert this UWs into UNL we use following “Rule 2”:

(Rule 2: is-a) If (isaKindof(UW1,UW2)) icl(uw1:09, uw2:0F)

Rule 2 check whether UW1 has “icl” relationship by using “isaKindof (UW1,UW2)”. For example `isaKindof(“dog (icl>mammal)”, ”canine(icl>mammal)”)` = TRUE, so the generated UNL is: `icl(dog(icl>mammal):09, canine(icl>mammal): 0F)`

The above mechanism works for UWs under “nominal concepts”. For other types of UWs such as “attributive concepts” we need to use different set of rules. For the UW `write(agt>person,obj>report)`, we can get the partial SemanticWordMap as shown in Figure 7.

From Figure 8 using from UWs only we can get UNL expression for “Person write a report”. However in the meaning of the UW we should not use that concept. Instead we can use immediate higher UW concept. So in this case instead of “write” we can use “produce”. By replacing person with someone we can get the UNL expression for “Someone produce a report”. In this way, using different rules the system can convert SemanticWordMap relations into UNL.

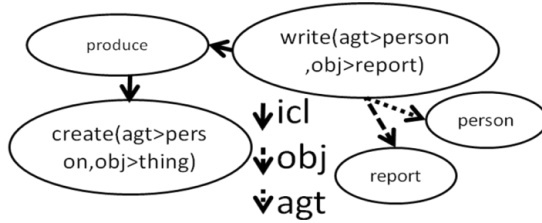


FIGURE 8: UNL Relations from UWs only

3.3 Describe in natural language

Finally, we used UNL DeConverter to convert the UNL expressions into natural languages. UNL DeConverter is a language independent generator that provides a framework for syntactic and morphological generation as well as co-occurrence-based word selection for natural collocation. It can deconvert UNL expressions into a variety of native languages, using a number of linguistic data such as Word Dictionary, Grammatical Rules and Co-occurrence Dictionary of each language. We used UNL DeConverter to convert the UNL into several natural languages such as English, Japanese etc.

3.4 Implementation in UNL Explorer

Finally the explanation can be expressed in different natural languages such as in English, Japanese or other languages using UNL DeConverter. Determining this kind of relationship can be very useful for knowledge engineering. For experiment we implemented the proposed method in UNL Explorer. Our implementation could successfully produce 1466598 UWs explanation in UNL. Table II shows some sample UWs explanation generated by the proposed method. Here, we only reported sample explanations in English and Japanese, together with the UNL expression.

TABLE II. SAMPLE UWs MEANINGS AUTO GENERATED USING OUR PROPOSED MECHANISM

Universal Word	Explanation Generated from UNL Ontology in Different Languages		
	English	Japanese	UNL
write(agt>person,obj>report)	Someone produce a report	誰かが報告書を作成する	agt(produce(icl>manufacture(agt>thing,obj>thing)):08.@entry, someone:00.@topic) obj(produce(icl>manufacture(agt>thing,obj>thing)):08, report(icl>account):0I)
Dog (icl>mammal)	Dog is a canine, mammal and animal. Dog can eat, whine, walk and die.	犬は犬、哺乳類や動物です。犬は、食べて駄々をこねる、歩いて、死ぬことができます。	aoj(:01.@entry,dog(icl>mammal):00) and:01(animal(icl>living thing):0S.@entry mammal(icl>animal):0H) and:01(mammal(icl>animal):0H,canine(icl>t ooth):09.@indef)

Using UNL expressions and UNL DeConverter it is possible to generate the explanation in more than 40 languages as well. However, the quality of the explanation depends on the quality of that language DeConverter. Therefore precision of the system highly relies on UNL DeConverter and the semantic background provided by UNL Ontology. As the users of this system are the editors of UNL Ontology, it helps them to improve the quality of manually built UNL ontology. The UNL dictionary builders can also differentiate the UWs from the natural language explanation without understanding the UNL language.

Conclusion

In this research we proposed the way to auto generate the meaning of each UWs using UNL Ontology. However, UNL Ontology by nature is a growing resource with millions of UWs. As UWs are not always understandable by human, the explanatory sentences are needed to develop necessary UWs for every language. For explaining UWs meaning it is necessary to auto generate from the same representation. Using our proposed solution computer can auto generate the meaning of UWs in more than 40 natural languages.

References

- George A. Miller. 1995. WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41.
- H. Uchida, M. Zhu, and T. Della Senta. “The Universal Networking Language”, 2nd ed. UNDL Foundation, 2005.
- H. Uchida, M. Zhu, T. Della Senta. “A gift for a millenium”. Tokyo: IAS/UNU. 1999..
- Khan Md. Anwarus Salam, Hiroshi Uchida and Tetsuro Nishino. “How to Develop Universal Vocabularies Using Automatic Generation of the Meaning of Each Word”, 7th International Conference on Natural Language Processing and Knowledge Engineering (NLPKE'11), Tokushima, Japan. ISBN: 978-1-61284-729-0. Page 243 – 246. 2011.
- Steven Pinker. The Stuff of Thought: Language As a Window Into Human Nature. USA. 2007.