

OBJECT ORIENTED MODELING APPLIED TO UNL

Adriana Gomes Alves, M.Eng.

adriana@inf.univali.br

Hugo Cesar Hoeschl, Dr. Eng.

metajur@eps.ufsc.br

Andre Bortolon, M.Eng.

bortolon@eps.ufsc.br

ABSTRACT

During the last few years, the Internet has emerged as one of the main channels for communication and access of information. The number of people getting connected and the interchange of experiences at all knowledge levels have grown at incredible speed, but yet, along with the advantages of a rapid access at long distances, comes the difficulty of communication due to the language barrier. The UNL (Universal Networking Language) has appeared as a way out to the minimization of this problem. Along with UNL also came the need for new projects that may turn feasible the use of this technology through the WEB. In this paper we present a way of using object oriented modeling in order to organize and better understand the basic elements of UNL from a computational point of view.

Keywords: UNL, Object Oriented Modeling.

I. INTRODUCTION

From among all emerging social problems, digital inclusion has stood out during the last few years as a topic for discussion among those who search for equal opportunities for all citizens. The final document issued at the Workshop for Digital Inclusion in 2001 [4], concludes that digital exclusion deepens social and cultural exclusion since those who have no access to the new technologies and means of knowledge, will find themselves less prepared to live with and get adapted to the new social processes. Digital inclusion should therefore guarantee equal opportunities for all.

The reality of computer technology in the world is still a privilege of a few. Statistics show that the gap between those who are connected and those who don't even have access to a computer will continue to grow. According to data from Internet Business [1], 41% of all Internet users in the world are concentrated in the USA. Asia, in spite of having the greatest part of the world population, represents only 20% of all internet access, and Latin America, ridiculous 4%.

Besides the problem of access to technology, there is also the problem of language, which constitutes one of the main barriers for a greater dissemination of web usage. This is due

to the fact that most of the information on the internet is only available in English. As an initiative to reduce this problem, UNL (Universal Networking Language) is presented as an artificial language whose goal is to allow translations between natural languages without the need to use an intermediate one, like English.

Until now, the main concerns with respect to UNL have been the questions regarding language and the definition of its structures, leaving the computational view of the problem more as a background study. In this paper, we intend to present some suggestions on how to model the basic structures of UNL as to better visualize and understand its elements, following an object oriented approach.

II. BASIC UNL CONCEPTS

UNL is an artificial language for computers in the form of semantic networks used to express any type of information. It is made of Universal Words (UWs), relations, attributes, and a Knowledge Base (KB). The Universal Words are themselves the vocabulary of the UNL, relations and attributes constitute the syntax, while the Knowledge Base constitutes the semantics [5].

A. Relations

Being that relations and attributes constitute the syntax of the UNL, it is by means of relations that complex ideas can be expressed, and hence, texts and sentences elaborated. Relations are represented by labels that bind two or more UWs. Labels are represented by strings of 3 characters or less. The following constitutes an example of a relation: agt(agent). Here, we define a “thing” which starts an action. Its syntax is agt(do,thing), where “do” is an action and “thing” the agent that executes the action. A practical example for the sentence “John breaks” is agt(break(agt>thing,obj>thing),John(icl>person)). The UNL presents a variety of around 42 types of relations which represent all possible associations between Universal Words. The UNL also tries to cover all the expressions necessary for any type of message that may be wished to be transmitted.

B. Universal Words

Universal Words may represent either simple or compound concepts. We have therefore two types of UWs: simple “UWs”, describing simple unit concepts, and “Compound UWs”, which are compound structures of binary relations grouped together. A UW is made up of a string of characters (in English) followed by a list of restrictions, or Constraint List. A basic UW corresponds to an English word, while a restricted UW represents a more restricted or specific concept. For example, if we take the basic word “state”, we may restrict it to more specific meanings by associating constraints to it, such as: state(icl>express), denoting an action that expresses something, or state(icl>country), which denotes a nation or country.

C. Attributes

Attributes are used with UWs to describe the subjectivity of sentences. They show what is said from the speaker’s point of view. This type of information may include things like feelings, attitudes, references or the moment in time when the sentence occurs. With attributes we can give more meaning and details to the message being presented. Attributes are always connected to a UW and we can have several attributes for one same word in one single UNL expression. For example, if we take the UW “eat” and associate to it the attribute @past, we will be expressing that the action of eating happened in the past. The time represented by attribute does not represent a verbal tense (like ate or was eating), only the moment in time with respect to the point of reference.

D. Dictionary

A dictionary of UNL words must be created for each native language which we may wish to make available on the web, for example: Portuguese, Japanese and German, among others. Linguists are responsible for the association of words (or radicals) and expressions of each language with the Universal Word (UWs). They must also inform all the grammatical attributes such as morphology, verbal inflexions, semantics, etc. which will make possible the translation of UNL expressions to a native language and vice versa.

E. Knowledge Base

The Knowledge Base (KB) is a hierarchic set of UNL expressions (UW, relations) that make up the UNL semantics. The categories of the KB allow for a better contextualization of words within a certain concept hierarchy. In this way, the word or expression that best fits to a

certain translation can be found. The present categories of the KB are: thing, do, occur, be, aoj>thing, mod<thing, how.

III. OBJECT ORIENTATION

The Object Oriented paradigm has been consolidating itself during the last few decades as an efficient technique used in the process of analysis and modeling of applications. Rumbaugh [3] defines object orientation as “a new way of thinking about problems with models organized around real world concepts. The main component is the object which combines structure and behavior in one single entity”.

The main concept in object orientation is the object itself, which represents one specific occurrence or instance of a class, corresponding to something abstract or from the real world. Objects are made up of attributes, that is, information or data. Objects are also made of the operations or methods that can be executed upon them. For example, considering the object “ball”, we may specify information (attributes) about it such as color, size, material, etc. On the other hand, we may also say that our object “ball” is capable of rolling, bouncing, being thrown, etc. (methods).

Objects can be classified considering their common characteristics (same attributes and methods). This represents the concept “class” of the object. Objects or classes may be associated with one another, forming networks known as Class Diagrams or Object Diagrams. A Class Diagram represents an abstraction of a system or concept which is desired to be modeled and understood. There are several types of associations between objects and/or classes, for example: specification, generalization, aggregation, simple association. We do not intend in this paper to enter into details of these concepts. These can all easily be found in the specialized literature.

Graphically, an object or class is usually represented with a rectangle divided into three sections: name of the class, list of attributes, and list of methods. The associations are defined with lines connecting these rectangles and, depending on the type of association, some symbol that may represent it. In this paper we shall use the UML notation for class diagrams [2].

IV. AN OBJECT MODELING PROPOSAL FOR UNL

The KB and the native words dictionary are both currently implemented in the computer in the form of sequential files [6], which is not an efficient method to access data. Knowing that the amount of gathered information will surely increase at amazing speed as applications for UNL are developed, the need appears to review the way in which this information is to be stored and retrieved by using more powerful computing tools. The use of an object oriented data base, for instance, is one of the possible solutions.

The object paradigm can be applied to any type of system or concept. It is thinking like this that we pretend in this chapter to present an initial proposal for the modeling of the basic UNL elements, seeking to obtain a diagram that may represent them graphically. This would surely facilitate its understanding and would also serve as a starting point for its implementation. Figure 1 shows a class diagram that represents the UWs, including the possible classifications of the KB, that is, how the UWs can be classified according to UNL definitions. The concept of inheritance applied to UWs, according to the KB, can be desirable when we come to give more details on relations, since these define the types of UWs that can be associated. UWs are, by definition, composed of a HeadWord and may or may not be restricted by means of Constraints, associating a more specific concept to it. Constraints, in turn, are composed of Relation Labels associated to other UWs.

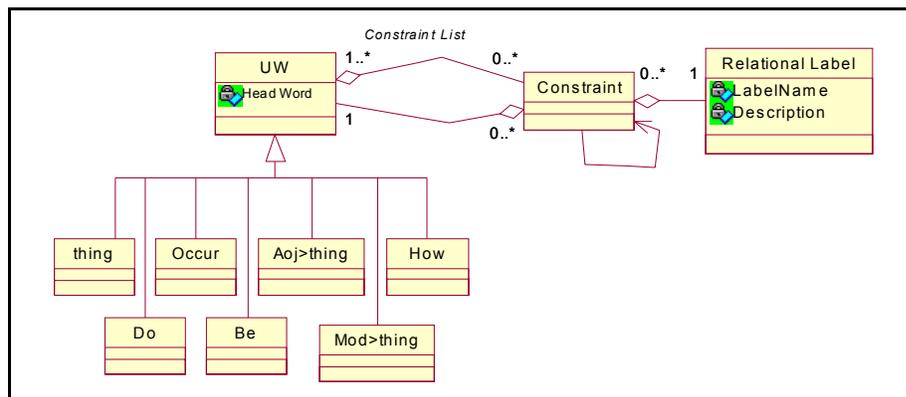


Figure 1 – Class Diagram of Universal Words (UW)

In Figure 2 we show a proposal for the modeling of a complete text in UNL. A text is made up of sentences, which in turn is composed of UNL binary relations. Relations associate two UWs, which in turn may have a series of attributes that will give them more meaning in the sentence.

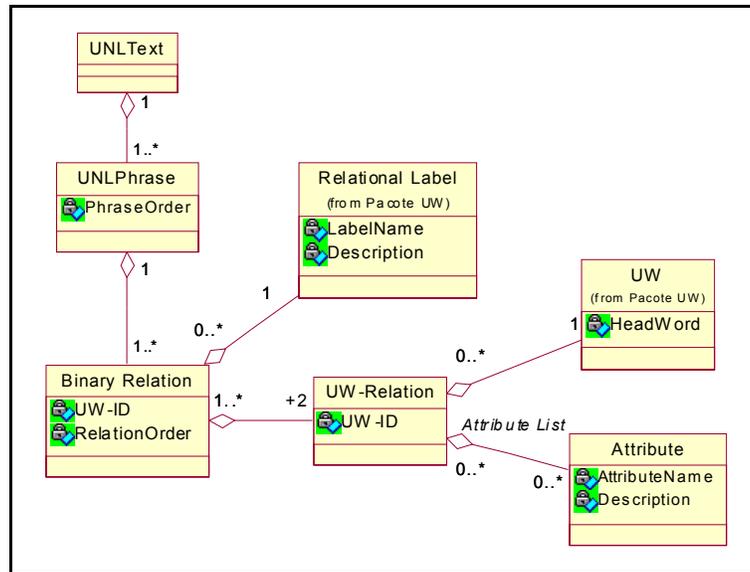


Figure 2 - Text Modeling with UNL

But, to use this model, we still have to include some restrictions between the Relational Labels and the UWs. This is due to the fact that for each type of relation the types of UWs that can be related together are well defined in UNL. This problem and other details like this one will remain to be treated in a revision of the models we have just presented. We suggest for this matter the collaboration of researchers on UNL that possess knowledge on object oriented programming.

V. CONCLUSION

The UNL appears as an artificial language that pretends to substantially increase as well as to improve the quality of communications among people from different places and of different tongues. It is not only concerned with serving as a mere tool for translations. It also seeks to put built-in knowledge into the process of interpretation of texts, preserving the original ideas.

There is still a lot to be done so that UNL may effectively be used at large scale. In the area of computer sciences, many projects will still appear that may come to use the technology behind UNL or to improve the quality of the information processing. This paper presented some ideas on how the basic elements of UNL could be modeled by means of the object oriented

paradigm. Nevertheless, there is still a lot left to detail and specify so that the model may be considered robust and stable.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Damián Rodríguez Sánchez for the paper revision and translation to English.

REFERENCES

1. DIEGUEZ, Flávio. Analfabetismo Digital. Revista Educação, São Paulo, year 28, n. 248, pp 28-36, dec. 2001.
2. FOWLER, Martin, SCOTT, Kendall. UML Essencial: Um breve guia para a linguagem-padrão de modelagem de objetos. 2.ed. Porto Alegre: Bookman, 2000.
3. FURLAN, José Davi. **Modelagem de objetos através da UML**: the Unified Modeling Language. São Paulo: Makron Books, 1998.
4. OFICINA PARA INCLUSÃO DIGITAL. **Relatório final**. Brasília, 2001. Available from: <<http://www.governoeletronico.gov.br/default2.cfm?idarea=10>> Accesed on: April 04, 2002.
5. UNL Center. The Universal Networking Language (UNL): Specifications. Version 3 Edition 1. Available from: <http://www.unl.ias.unu.edu/unlsys/unl/UNL%20Specifications.htm>. Accessed on: March 27, 2002.
6. HOESCHL, Hugo C. Aplicações em Engenharia do Conhecimento: UNL. Florianópolis: UFSC. 2002. (Class Notes)