

A proposal of an UNL Application Development Environment

Andre Bortolon, M. Eng

IJURIS, Federal University of Santa Catarina, Campus Universitário, Trindade, Florianópolis, Santa Catarina,
Brazil. CEP: 88040-900
bortolon@eps.ufsc.br

Hugo Cesar Hoeschl, Dr.

IJURIS, Federal University of Santa Catarina, Campus Universitário, Trindade, Florianópolis, Santa Catarina,
Brazil. CEP: 88040-900
metajur@eps.ufsc.br

Joel Ossamu Mitsui

IJURIS Collaborator, Federal University of Santa Catarina, Campus Universitário, Trindade, Florianópolis,
Santa Catarina, Brazil. CEP: 88040-900
joel@pta.com.br

Jaime Leonel de Paula Júnior, M. Eng.

IJURIS Collaborator, Federal University of Santa Catarina, Campus Universitário, Trindade, Florianópolis,
Santa Catarina, Brazil. CEP: 88040-900
jbj@pta.com.br

Ricardo Miranda Barcia, Ph.D

Production Engineering Pos-Graduation Coordinator at UFSC, Federal University of Santa Catarina, Campus
Universitário, Trindade, Florianópolis, Santa Catarina, Brazil. CEP: 88040-900
rbarcia@eps.ufsc.br

Abstract. The development of UNL applications may have a lot of problems. An experience occurred in Federal University of Santa Catarina showed that the occurrence of more problems than were foreseen in the start of the problem, may destroy the project. The construction of tools to help programmers to develop the applications increases the speed of development. Besides, these tools can popularize the usage of UNL, facilitating the access to the technology and the connection with other languages.

Introduction

The UNL purpose is to “provide a common communication environment for different language.” [1] Since UNL is a language to computers express the knowledge, it is necessary the existence of tools to support the creation of UNL projects and documents. These tools allow people who are interested to publish their documents, papers, and applications in UNL without knowing all the details of the language. In the UNL Project, there is the UNL Editor. The UNL Editor “is linked to a language server equipped with an ‘enconverter’ and a ‘deconverter’ for a natural language. As the author writes a document, e-mail or any other text, in his/her language, UNL editor ‘enconverts’ it into UNL documents. In this process, UNL expressions are produced automatically or interactively with the author.” [1]

Normally, applications work in a specific domain. So, it is possible to use just a smaller set of rules and universal words to represent this domain. It is also possible the inexistence of some words in the Master Dictionary, because this domain have not been included yet. These words must be included. But, the inclusion of new words is

preceded by a study to correctly contextualize them, finding the appropriated restrictions and so on.

Also, the application might need specific rules to “enconvert” or “deconvert” special cases. These rules should be applied just in this domain and cannot be included in the rules general dictionary.

This paper suggests an upgrade to the UNL Editor, the UNL Application Development Environment (UDE). This upgrade allows the management of UNL projects. An UNL project is the construction of an application that uses UNL to represent their knowledge. Using the UDE, the developer can create interfaces to link UNL with some other developing tools and programming languages, such as Java or C#. UDE can also manage the creation of transitory UWs and special rules, as if the creation of smaller word and rules dictionaries specific to a domain.

Motivation

A research team from Federal University of Santa Catarina is implementing a simulation on Environment Audit using UNL and Virtual Reality. This simulation intends to test a student to audit a chemical warehouse. All the textual elements from the simulation, such as signs and speeches, will be represented in UNL, allowing people from different countries to access the simulation and understand it in their own language. Five languages have been chosen: Portuguese, English, Spanish, Italian, and Japanese.

The texts were revised to avoid more complex structures and expressions only used in Portuguese. After that, it was started the conversion of the sentences to UNL. 164 Universal Words were extracted from 61 sentences. More than 80% of these words weren't in the knowledge base or in the UW Gate. So, these UWs needed to be created. But, the creation of new UWs needs a study to correctly contextualize them in the knowledge base, avoiding repetition or ambiguity. This study requires time and able people to do that. Also, the UWs must be shared between all the researchers. These problems delayed the work in 2 weeks.

The construction of the UNL sentences was affected by some other problems. Typing errors, absence of marks (such as periods before the attributes) and parenthesis occurred in almost a half of the sentences. Three revisions were made and errors were found in the next step, the construction of the dictionaries.

In the dictionaries, the errors mentioned above were multiplied. Since there is a lot of marks and tags as in the word dictionary as in the rules dictionary, the quantity of errors occurred have increased in the same proportion. Also, the inexperience of the UNL programmers contributed to that.

Because of these problems, the deadline weren't accomplished.

UNL Application Development Environment

The UNL Center's Universal Word Gate allows users to access the UNL's centralized UW Dictionary system through the Internet. This tool is useful to find UWs that correspond to meaning in many language like Japanese, Italian, Spanish, Russian, etc. It's only available to UNL Society members. However, this UW Dictionary System has only the words that have already been included in previous projects done by the UNL Center. In the project described in the previous section, a lot of UWs were not

found in the UW Gate, since this new UWs correspond to concepts that have not been studied in UNL yet. According to the normal UNL procedure, the UWs need to be requested to UNL Center and then included in UNL Center's UW Dictionary to be used. This procedure could waste more time than create this UWs locally only, probably delaying the delivering of the project.

So, the UWs had been created locally and restricted the attributes to the general form. To store and retrieve the locally created UWs and to improve and reduce the time of UNL sentences codification, the Brazilian research group of UFSC (*Federal University of Santa Catarina*) has developed a tool, based on UNL Center's Universal Word Gate, called UWGATEBR (<http://www.ijuris.org/uwgatebr/>) (Figure 1). Also, as verified on the development of Environment Audit application, this tool allows a progressive study of the UW restrictions, from the most general to the most specific, improving the best format to be used in the application without compromise the whole process because the UWs stored in the UWGATEBR database can be edited in any time. Consequently, the syntax errors of UNL code will be minimized before its use in *Deconverter*, reducing the rework and the development time.



Fig. 1. UW Gate Brazil Home page

The UWGATE.BR stands out for the easy way to add new *UWs* to the UW Dictionary used to codify UNL sentences. Everyone may include a new UW in the database. In the beginning, this new UW will be analyzed to be verified and corrected by the responsible administrators and they will decide approve the UW in the dictionary.

Public module

This module is available to anyone and works in the following way. The user can type all or part of the word in the *Palavra* field or/and a Head Word, and clicks on the ***Pesquisar*** (search) button. If the command is accepted, the UWGATEBR retrieves the

correspondent UW entries from the UW Dictionary database and shows the results in two columns: in the left column will appear the matched words (if the user input was a word), and in the right column the corresponding UW (or the matched UW from the user Head Word input). Clicking in any item of both columns, automatically the corresponding item in other column will be selected (Figure 2).

Address Go Link

UW GATE BR

[Consultar UW](#) | [Incluir nova UW](#) | [Log off](#)

Palavra: Head Word:

Pesquisar Limpar

acao acelerar acima ajuda algum ambiental ambiente andar animado ano	speed up(icl>do) above(icl>direction) help(icl>activity) some(mod>thing) environmental(mod>thing) environment(icl>place) walk(icl>do) animated(mod>thing) year(icl>date) previous(mod>thing)
--	---

Pesquisar Limpar

UW: environmental(mod>thing)

Fig. 2. Search page

To include a new word not found in the search, the user can click the hyperlink *Incluir nova UW* (add new UW) on the top of page. Four data fields are required in the input: palavra (word), radical (stem), HW (Head Word) and UW (Universal Word). Click in *Enviar* (send) button after fill out the data in the form (Figure 3).

Address Go Link

UW GATE BR

[Consultar UW](#) | [Incluir nova UW](#) | [Log off](#)

Palavra: Radical:

HW: UW:

Limpar Enviar

Fig. 3. Inclusion page

Administrative module

The access to this module is restricted to UWGATEBR administrators. Besides the public functions like search or add new *UWs*, the administrator module has additional functions such as approving new *UWs*, editing, remove and consulting all the *UWs* available on the *UW* dictionary database, and also manage the system users.

To approve a *UW*, just select the words and click in **homologar** (approve) (Figure 4). To remove a *UW*, just select the word(s) and click in **remover** (remove). Any necessary correction in the word can be done clicking the *UW* and will be loaded a page to edit the word, stem, HW and *UW*. After editing click **Gravar** (record) button (figure 5).

Address: <http://www.ijuris.org/uwgatebr/homologar.asp> Go Link

UW GATE BR

[Consultar UW](#) | [Incluir nova UW](#) | [Homologar UWs](#) | [Listar UW](#) | [Usuários](#) | [Log off](#)

Palavra	Radical	Head word	Universal word	Todos
data	data	date	date(icl>time{>abstract thing})	<input type="checkbox"/>
decimal	decima	decimal	decimal(icl>quantity)	<input type="checkbox"/>
enviar	envi	send	send(icl>do)	<input type="checkbox"/>
inicial	inicia	initial	initial(aoj)	<input type="checkbox"/>
melhor	melhor	best	best(aoj>abstract thing)	<input type="checkbox"/>
ponto	pont	dot	dot(icl>simbol)	<input type="checkbox"/>
valor	valor	value	value(icl>abstract thing)	<input type="checkbox"/>
vento	vent	wind	wind(icl>natural fenomenon)	<input type="checkbox"/>

Fig. 4. List page

The UWAGATEBR is an effort to create the **UNL Application Development Environment (UDE)** capable to organize, manage and create the necessary elements, such as Word Dictionaries, UNL Sentences and UNL rules dictionary (language grammar), to build an application using UNL. The tool uses all the UNL conventions, syntax and rules described in the UNL Specifications. The importance to build knowledge base libraries specific to each application project is the high performance and precision of the conversion from UNL code to the target language. This modularity of knowledge can be improved individually, and then gradually will reach the entire language.

UNL Application Development Environment will works like a Visual Studio. Each project created on UNL Application Development Environment will be associated to a database, which contains all the information necessary to build and manage the UNL rules dictionary files and word dictionary files grouped by language. Even aid on application sentences files and UNL sentences files creation.

The UDE databases will be linked with the UNL Center and UNL Language Center databases, using the dictionaries already existing.

Address

UW GATE BR

[Consultar UW](#) | [Incluir nova UW](#) | [Homologar UWs](#) | [Listar UW](#) | [Usuários](#) | [Log off](#)

Palavra:	<input type="text" value="enviar"/>
Radical:	<input type="text" value="envi"/>
HW:	<input type="text" value="send"/>
UW:	<input type="text" value="send(icl>do)"/>

Fig. 5. Edition page

The UW Dictionary structure table will store the UWs and the respective meanings grouped by language. These UWs will be extracted from the user sentences. Once identified the UW, the user will complete the dictionary entry with the appropriated restrictions. Also, the user has to insert the grammatical attributes of the word. The goal is to provide a tool that supports the building of all elements of a word dictionary entry. The sentences will be displayed in the screen as a hyper graph with nodes and edges. According to the node selected by user, all the relations, attributes set to this node will be displayed, permitting to user changes them in any time.

Untitled Document - Microsoft Internet Explorer

File Edit View Favorites Tools Help

← Back → Search Favorites History

Address

select language:

UW:

All matched results of search

[\[software\] {} software "software\(icl>tool\)" \(NOU, MCL, ^NBR\) <P,0,0>](#)

...

NEW

UW

Stream

Description

Attributes

Priority

Fig. 6. Interface of word dictionary builder tool prototype

The tool interface will provide more agility on UNL Rules Dictionary, Word Dictionary and UNL code building because the users, that must have a basic knowledge of programming concepts, will not worry about the UNL syntax. All the information, explanation, and fields of the syntax will be displayed on screen. To add a new entry in word dictionary, the users just have to set the entry priority, describe the UW, stem and add all necessary attributes, saving this information in the project database (Figure 6).

The same concepts can be used applied to create rules. The interface will display a list of rule types (attribute changing, left shift, right shift, and so on). Left insertion and right insertion will be indicated as different rule types to facilitate user work. To complete the rule is necessary set the conditions, actions, relations and roles of the nodes on the Left and Right Generation Window (Figure 7).

Different databases are used to enconversion and deconversion process. The user can indicate the same word dictionary for both process, but the rule databases need to be different.

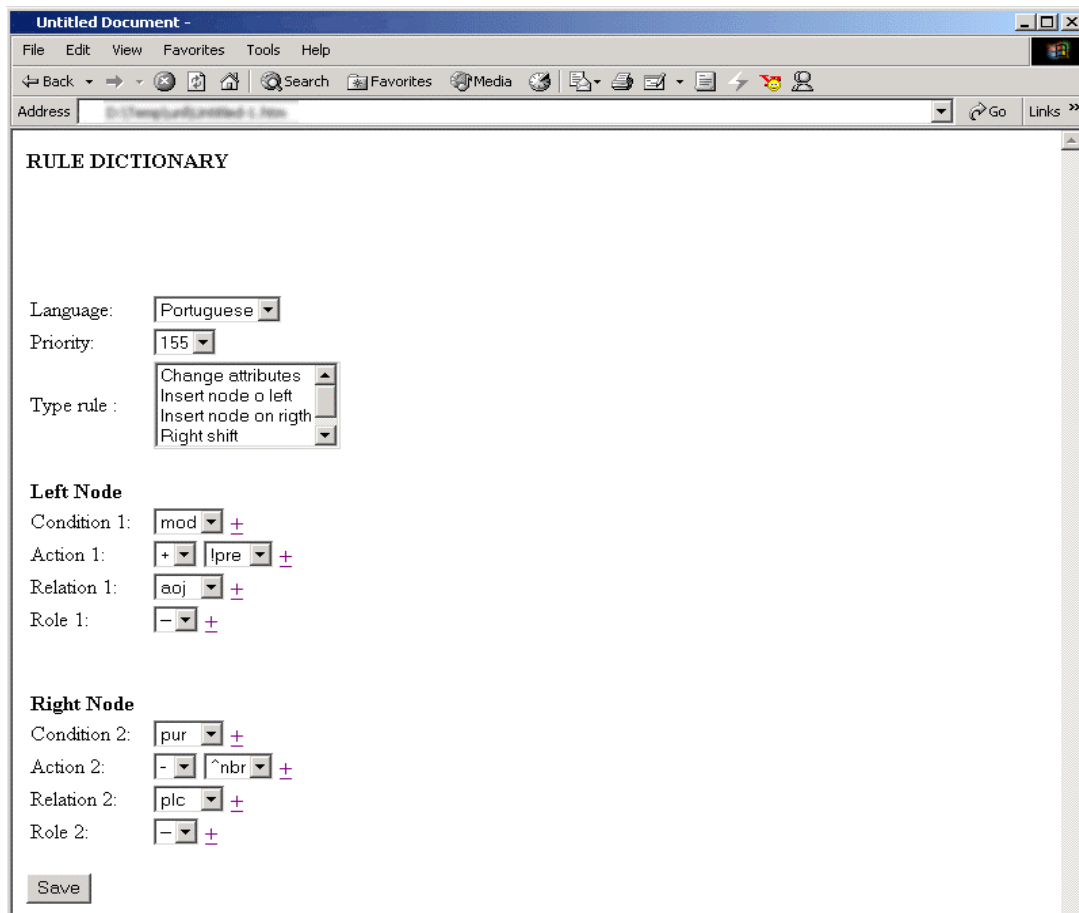


Fig. 7. Interface of rules dictionary builder tool prototype

Conclusion

The construction of tools to facilitate the integration of UNL with other applications and the creation of UNL documents can help the dissemination and the popularization

of UNL. Allowing people quickly create and develop small applications using UNL, even in specific domains, will generate a lot of new words and the continuous increasing of rules to solve the deconversion. This process has already happened with other languages, such as Java. Collecting the best sets of words, the most efficient rules, will generate better and better dictionaries. Also, the easy access to the mechanisms of UNL can motivate people from countries that don't still have UNL Centers to produce the dictionaries to enconvert and deconvert.

References

[1] Uchida, Hiroshi; Zhu, Meying; Senta, Tarcisio Della. **A Gift for a Millennium**. Tokyo, 1999.

Bibliography

Uchida, Hiroshi; Zhu, Meying. **The Universal Networking Language beyond the Machine Translation**. in: International Symposium on Language in Cyberspace. Seoul, 2001.
UNL Centre. **Deconverter Specifications. Version 2.5**. Tokyo, 2001.