A WEB Platform Using UNL: CELTA's Showcase

Lumar Bértoli Jr.,[1] Rodolfo Pinto da Luz,[2] and Rogério Cid Bastos [3]

[1] Instituto UNDL Brasil, Rodovia SC 401 - Km 1 –
ParqTec Alfa, Ed. CELTA - Bloco B - 3º pavimento - Módulo 2.10
João Paulo – 88030-000, Florianópolis– SC- Brasil
bertoli@undl.org.br

[2] Instituto UNDL Brasil, Rodovia SC 401 - Km 1 –
ParqTec Alfa,Ed. CELTA - Bloco B - 3º pavimento - Módulo 2.10
João Paulo – 88030-000, Florianópolis– SC- Brasil
luz@undl.org.br

[3] Rogério Cid Bastos, Departamento de Informática e Estatística, UFSC,
Campus Universitário - Trindade - 88040-900, Florianópolis - SC – Brasil
rogerio@inf.ufsc.br

Abstract. Economic globalization is changing the way companies communicate. The ease and speed of accessing information and taking decisions is better for everyone on the decision chain. To speed up access to information it is important to present information in one's native language and create a language-independent communications channel. To assist business-to-business operations, the development team at the Instituto UNDL Brasil designed the pilot project "CELTA's Showcase" to demonstrate that it is possible to create a multilingual business-to-business platform using UNL.

1 Introduction

The interconnection between producer and consumer is becoming extremely important. The expansion of markets from local to global influence requires the use of new technological resources in order to support the majority of these relationships. In addition, the specialization of markets requires the development of automated tools to facilitate the pairing of small groups of producers and consumers.

Due to the irreversible globalization of markets and the specialization of production areas that create high technology products, there is a growing need for perfect matching between producers and consumers, to allow maximum performance in efforts to connect both sides.

To increase the chances of matching the best producer-consumer pair, the *Instituto UNDL Brasil is* proposing a project in this field. The main objective of this project is the development of a multilingual Web platform that allows integration between producing companies and their customers. This project is being developed by the *Institute UNDL Brasil*, and was made possible by the creation of the *UNL Research and Development group (R&D)* in the year 2003 [1]. The R&D group has a highly trained IT team whose main objectives are:

Universal Network Language: Advances in Theory and Applications.
Research on Computing Science 12, 2005, pp. 276–285.

- UNL research; including the creation of tools and linguistic resources for Portuguese; and
- Application development using UNL for the Brazilian market.

The *Instituto UNDL Brasil* is located at the technology base company-incubator, CELTA [2]. CELTA is composed of startup companies that work with high technology and that are based in Santa Catarina State, Brazil. CELTA's objectives are to support startup companies in their administrative procedures, and to provide an atmosphere of technological interaction among them, the scientific community, and the market. CELTA is currently the largest high technology incubator in Latin America, incubating more than 30 companies in areas such as electronics, computer science and knowledge management [2][3].

The *Instituto UNDL Brasil has* presented a proposal that recognizes the concrete needs of the companies at CELTA. This proposal is the pilot project "CELTA's Showcase". The deliverable product of this project is a multilingual Web platform to serve startup companies incubated at CELTA. This project involves the development of a platform that will allow CELTA's companies to display their products and services, and to exchange information with their customers, partners, investors, suppliers, and potential customers. Initially it will support six languages defined as requirements by the incubated companies. More language options will be available as UNL evolves and UNL System integration becomes available. Product sales are often targeted only to markets of English and Portuguese speakers, reducing the chances of enlarging a company's market share.

The main objective of this project is to enhance the market interaction channels of CELTA's companies. The approach will increase the exchange between the companies' staff and potential foreign customers by allowing each party to use their native language.

2 Platform description

The pilot project "CELTA's Showcase" consists in the development of a platform to enlarge market share for companies incubated at CELTA.

CELTA's Website currently presents information only in Portuguese, impeding information access by foreign investors' and customers. In addition, there is no specific information about the companies; there is only a link to each company's website with no other description at all. This link allows the visitor to reach the company's website's if it has one.

Some companies display their products on their own website. In order to broaden the website's accessibility, some websites were translated manually for English and a few for Spanish. At this moment, two solutions have been used to accomplish this translation task. The cheaper solution is to leave this task in the hands of company staff (usually IT or administrative staff, and not professional translators). A more expensive and suitable solution is to contract a service or a professional to perform the task. In addition, these translations are only valid at the time of the translation, which creates a need for continuous contracting of translation services to keep the website's information updated. Unfortunately, incubated companies cannot implement these so-

lutions because in the first case they involve a misuse of highly qualified staff, and in the second because of high costs.

The solution presented here was based on the analysis of the companies' requirements. It resulted in detailed specifications for the platform.

The platform will allow all of CELTA's incubated companies to have their information registered in a standardized multilingual website. This website will allow visitors, such as customers and investors, to access the companies' multilingual information. The principal content that will be available for visitors are:

- General company information;
- Products and services; and
- Customer lists

The content should be displayed using the chosen languages: English, Spanish, Chinese, German, and Arabic, in addition to Portuguese. These six languages encompass a high percentage of potential investors and consumers for CELTA companies., With the use of UNL, the insertion of more languages will not require complex changes and could be included at any phase of project implantation.

In addition to the content, an interface will be provided to allow interaction between visitors and CELTA companies. This interface will permit communication between companies and their customers and investors. This tool will allow language independent communication. The companies' staff will use the Portuguese interface, and the other party (customers and foreign investors) will use their native language, limited to the six languages supported by the platform.

The software development methodology adopted is the evolutionary prototyping model [4]. This methodology is based on software development using traditional cyclical software engineering phases, where all phases are repeated in the correct sequence until a deliverable prototype is reached. It was selected for two main reasons: 1) the UNL System is not yet fully operational and certified; 2) CELTA's company members are constantly changing and new requirements could emerge.

The platform is divided into three views: visitor, company, and administrator as presented in figure 1.

The visitor's view gives access to a website with information about all of CELTA's companies. Using this view, the user is able to navigate through each company's content. In addition to the navigation option, a search engine based on UNL will be available to help the visitor find specific information about the companies and theirs products and services. UNL will be used in this case not only for simple translation, but also to confirm that the correct meaning is being selected. This will be possible by integration with the UNL Knowledge Base (KB). Using the proper tools embedded in the UNL KB, the identification of the precise meaning of a concept is possible. For example, a visitor can search for a particular field, and all directly and indirectly related products will be displayed.

At each company's space, there also will be a tool that allows visitors to communicate with the company in his or her native language. This tool initially will be limited to the platform language support and later to UNL System support. As the platform development advances, this tool will transform from a fully restricted content solution to one with free text domain restrictions.

The UML use case [5] for the actor visitor is presented in figure 1(use case "a"). The visitor's view will be implemented as the *visitor's module*.

The company's view, allows a company's staff to manage the company's content and to interact with visitors that have sent messages through the visitor's view. The UML use case of the actor company is presented in figure 1(use case "b"). The company's view will be implemented as the *company's module*.

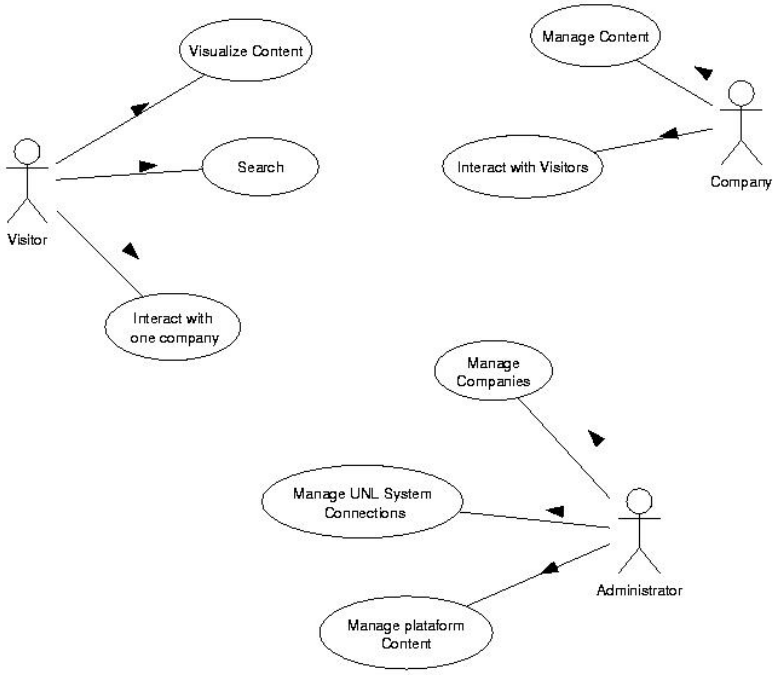

Fig. 1. UML use cases

The administrator's view allows CELTA administration to manage registered companies, to manage the connections with the UNL system and to administrate the platform's overall content. The UML use case of the actor administrator is presented in figure 1(use case "c"). The administrator's view will be implemented as the *administrator's module.*

The project was divided into 3 main phases (prototypes). In the first phase, to be developed within a window of six months, information about each company will be collected in Portuguese, using a web application at the end of the first phase. A UNL specialist will manually transform this content to UNL [6]. As the information becomes represented in UNL, the translation to the visitor's natural language will be conducted automatically and transparently, depending only on the language supported by the platform.

The second phase of the project will have a planned duration of 18 months and will allow the companies to describe their activities, services, and products using free text,

instead of the limited words and text used in the first phase. In this phase, companies will be allowed to introduce more accurate and extended information.

The third phase will take 36 months, for a total project time of 60 months. The goal for the end of this phase is for companies to be able to interact with their customers using their native language in real time within the scope of the business transaction. This feature depends exclusively on UNL development, at least for the six intended languages.

3 Development Status

The first and present phase began to be developed in August of 2003 and is expected to be concluded in March 2004. All systems involved in the development are open source, as is the platform [7].?? The platform was planned to be operational system and hardware independent. Its requirements were identified through interviews and forms analysis of pre-compiled data from CELTA company websites. These interviews were accomplished in a sample of the universe of companies that considered various technological specializations.

The technical solution for the interaction between the visitor and the company in the native language of each one is limited in this first phase to the use of predefined forms based on UNL that limit the questions and answers. This technical restriction will not affect the ability to exchange the most important information on each company's field. This communication solution comprises one of the initial requirements of the platform.

After the implantation of the system at the end of the first phase, feedback from the companies will provide the information needed to broaden the inclusion range of this communication tool, including more users' needs. At the end of the project, it is expected that all needs will be covered. In the following phases of the project, this communication tool gradually will begin to support free real-time conversation within a limited scope.

The first phase of the project was distributed in the following steps:

1. Requirements analysis and data acquisition

In this already completed step, a questionnaire targeting the companies was prepared with generic questions. It was distributed to five companies as a way of identifying common institutional information, the characteristics of products and services, as well as terms and concepts commonly used in customer contacts. This questionnaire was based on the first set of interviews and website analysis.

After the characteristics were identified, construction began of the *company module*. The implementation was accomplished through a web form (Figure 2) allowing the CELTA companies to register their information on the platform's database. The module was developed using PHP [8] and the database was generated using MySQL [9]. The objective of this step was to collect the content for conversion in UWs and UNL sentences. This content will be part of the corpus of the *visitor's module*.

Fig. 2. Company Module – web form

A support tool for dealing with UWs and UNL sentences was developed by *Instituto UNDL Brasil*. This tool is helping UNL specialists to perform the *enconversion* [6] activity, transforming the content in Portuguese to UNL. This support tool allows the specialist to browse all content entered by the companies. It marks content already in UNL, and presents content to be *enconverted* [6]. This tool is necessary only during the time that a *UNL Encoverter* for Portuguese is not available. It is expected that this tool will be removed at the end of the project, in the third phase.

2. Visitor's module development

To create the visitor's module, two items should be developed: software structure and content.

The structure for the visitor's view contents was developed using PHP in order to present the content in the visitors' native language. Using one of PHP's capabilities - the generation of dynamic content for web navigators - the visitor's view was created, reserving unique spaces for each company.

When the visitor accesses the website of the CELTA's Showcase platform, the visitor's web navigator will visualize the content in the visitor's language. If the language set on the web navigator is not already supported by the platform, a message is displayed requesting the selection of one of the supported languages. This message is displayed in all the supported languages of the platform at the same time. After the selection, all the content is displayed using the target language.

The accomplishment of the content display in different languages is possible due to the visitor's module. The visitor's module will access UNL System to present the dynamic generated pages. This is possible since all content is stored in a database using UNL.

The content, gathered in the last step will be converted to UNL using the tool developed in the first step. The other activities of this step have already been accomplished. This step is still being developed mainly due to the *enconversion* of the content to UNL.

3. Development of the communication tool

In this third step, a communication tool will be developed. Two classes of users, the visitor, and the company will use the communication tool. The implementation will be quite similar in both cases.

At the *visitor's module*, the communication tool will offer the visitor the opportunity to exchange information with CELTA companies using the visitor's native language. In order to make this possible, interactive forms will be used that allow limited communication among the parties in this phase. With these forms, the visitor can compose different questions and answers that will be automatically converted to UNL. These forms are restricted to a set of questions and answers. This set of questions and answers will be preprocessed for both customer and company.

The communication tool will convert the questions from the visitor's language to UNL automatically. This solution eliminates the need to use ENCO. This is only possible due to the simplicity and the restrictions at this phase.

After the UNL sentences are composed, the questions will be sent to the addressee, the CELTA company's responsible staff person, who will receive the message in his or her, native language, Portuguese, after the UNL System decodes the sentences.

This message will then be presented at the platform company's module. This application also will be developed using PHP. It will allow the company representative to answer the questions. This application will be built in a way similar to the one that the visitor used. The answer given by the company's representative will be sent to the visitor's e-mail in his or her native language, with a link attached to the platform allowing further conversation.

At this phase, the mapping from the language of the language set, including Portuguese, to UNL will be accomplished using the same methodology presented before to create the UNL content for the platform. The words and sentences used will be previously chosen and the same *enconversion* method will be applied to convert them to UNL. If necessary, some concepts will be generated as temporary UWs.

4. UNL System link

For the immediate operation of the platform, the current version of the UNL System [10] will be utilized. However, the platform will be loosely coupled to the UNL System, allowing future modifications of both the UNL System and the UNL without interfering in the platform directly.

The dynamically generated content for the platform (by PHP) will be sent to the UNL System Interface, UNL Web Service, before being returned to the visitor. This interface was planned as a web service [11] responsible for connecting the platform with the UNL System. The UNL Web Service mainly will be responsible for the following operations: processing the UNL requests, and selecting the most appropriate server for the required conversion. All these processes are transparent to the visitors and companies.

4 How does it Work?

There are two possible scenarios for accessing the platform as a visitor. Both scenarios integrate the same path up to a point where a division occurs. The visitor accesses CELTA's Showcase website through his or her web navigator where one can find information about the participating companies.

The content of the website is stored in UNL and in native languages already cached in previous access.

In the first scenario, the visitor's module will prepare the pages in the native language and simply display them to the visitor.

In the second scenario (figure 3), the visitor's module will prepare the pages based on the answer that the UNL System gives to the platform. If the user requests the information and it is not yet cached on the platform, the visitor's module sends a request to the UNL Web Service. This request contains the UNL sentences or UWs and the intended target native language.

The UNL Web Service is responsible for handling the data exchanged between the platform and the UNL System. It has a dynamic table that helps locate the intended native language UNL Language Server. It will communicate with the platform in a standard protocol for web service. The interface with the UNL System follows the UNL Language Server protocol.

The UNL System converts the content using the DECO and the linguistic resources that correspond to the native language requested, such as dictionary and grammar.

After it receives the native language content corresponding to the requested UNL content from the UNL System (UNL Language Server), the UNL Web Service returns it to the platform.

As the content arrives, it is stored on the corresponding language cache allowing the completion of the page construction to be presented to the visitor, as a web page.

This whole process will be fast, even though it is complex, because the information once transformed to the visitor's natural language will be stored locally at the platform's database language cache.

5 Final Considerations

This platform could be based entirely on translations performed by humans and all of the companies' content could be converted to the five other languages, providing them with the type of solution that big companies use. But these services would be very expensive for incubated companies and would represent a large share of their monthly expenses.

The implementation of this platform adopted an incremental method of using UNL resources to avoid interference from a technology that is not yet mature. In order to isolate the platform from problems related to UNL development, an interface will be provided by UNL Web Service.

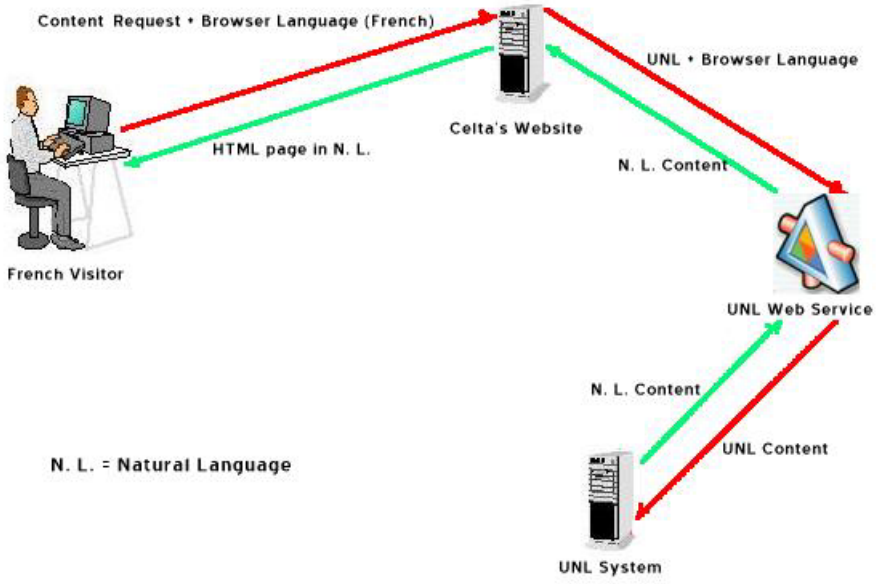

Fig. 3. System works

In addition to isolating the proposed platform, the creation of the UNL Web service in the future will allow any application to have access to a MT system and to knowledge management tools such as UNL KB and UNL Encyclopedia.

It is likely that the UNL Web Service, responsible for communication between the platform and the UNL System, can contribute in some way to the enhancement of the UNL System.

One of the main advantages of using UNL, instead of conventional translation, is the ability to write the content just once, which will then be replicated automatically to several languages through the UNL System. This advantage reduces deviation when updating content and reduces translation costs.

Another important advantage of using a platform like the one proposed here is that the companies can exchange simple information with customers, partners and foreign suppliers, with each one writing in their native languages, without a need for human translators'. This will already be possible at the end of the first development phase.

It is expected that even at the end of the first phase the platform will begin to achieve the planned goals and will increase the efficiency in communication among the companies at CELTA and their customers, partners, suppliers, and investors. It is also expected that this application will broaden the possibilities for using UNL, especially those related to commercial applications, and based on this platform, similar projects and products can be developed. The users of such a platform will not have any contact with UNL after its completion.

As the second development phase begins, the developers will receive feedback from the users. This feedback will begin after the results of the first phase are achieved. The continuation of the phases, will allow users to enjoy the capabilities of the UNL System to a larger degree.

Acknowledgements The authors would like to thank the Instituto UNDL Brasil, Fundação de Ciência e Tecnologia do Estado de Santa Catarina, and Universidade Federal de Santa Catarina for their support

References

1. Instituto UNDL Brasil. Conferência Sobre Tecnologia UNL. Available at: <http://www.undl.org.br/>. Accessed on: 15 oct.2003.
2. CELTA. CELTA On Line. Available at: <http://www.celta.org.br/>. Accessed on: 15 oct.2003.
3. Castiñeira, M. Inés;Mülbert, Ana Luisa; Schuhmacher, Vera; Madeira, Mauro. "Projeto Pedagógico em Ciência da Computação: Como Atender a Diversidade Regional?", Weimig2003, II Workshop de Educação em Computação e Informática, Poços de Caldas, May, 2003, 9 pp.
4. RUMBAUGH, J. Object-Oriented Modeling and Design. NJ: Prentice Hall, 1991. 500 pp.
5. RUMBAUGH, J.; JACOBSON, I.; BOOCH, G. The Unified Modeling Language Reference Manual. Reading, MA, USA: Addison-Wesley, 1999. 550 pp.
6. Uchida, Hiroshi; Zhu, Meiying; and Della Senta, Tarcisio. "The UNL, a Gift for a Millennium", Institute of Advanced Studies, United Nations University, 1999, 64 pp
7. Open Source Initiative. "Frequently Asked Questions". Available at: < http://www.opensource.org/>. Accessed on: 15 oct.2003.
8. PHP project. "What is PHP?". Available at: <http://www.php.net/>. Accessed on: 15 oct.2003.
9. MySQL. "MySQL Reference Manual". Available on: www.mysql.com/doc/en/What-is.html. Accessed on: 15 oct.2003.
10. Uchida, Hiroshi; Zhu, Meiying. "The Universal Networking Language beyond Machine Translation", International Symposium on Language in Cyberspace, Seoul, Sept 2001, 13 pp.
11. World Wide Web Consortium. "Web Services". Available on www.w3.org/2002/ws/. Accessed on: 15 oct.2003.