GYMN@ZILLA: A browser-like repository for open learning resources

Oliver Streiter
Area Language and Law
European Academy of Bolzano, Italy
ostreiter@eurac.edu

Judith Knapp
Area Information and Communication Technology
European Academy of Bolzano, Italy
jknapp@eurac.edu

Leonhard Voltmer
Area Minorities and regional autonomies
European Academy of Bolzano, Italy
lvoltmer@eurac.edu

Abstract: This paper reports on recent developments of the LOGOS GAIAS Project. The LOGOS GAIAS project aims at providing material and technologies for developing language competence in the area of integrated landscape development in Europe. The authors introduce GYMN@ZILLA, a browser-like application which integrates existing educational and non-educational modules in a new didactic environment. GYMN@ZILLA allows to access documents from the Internet or the LOGOS GAIAS repository and to convert its text into an easy reader text, a glossary or a completion exercise. The authors show how the combination of content modules and procedural modules allows the rapid development of a platform with new pedagogic characteristics. The results are promising for the flexibility of the learning material, importance of the topics and for enhanced learner motivation.

Introduction

Nowadays it is widely accepted that learning can be supported efficiently by technology. Especially for livelong learners, i.e. professionals who wish to train a specific skill autonomously at any time or place, the availability of technical support is indispensable. Understanding and speaking a foreign language is one of these skills. The WWW with its huge amount of freely accessible authentic text is an ideal source for language course material.

In this paper we are reporting about an ongoing research project in the field of computer assisted learning of languages for special purposes. The LOGOS GAIAS project aims at providing material for developing language competence needed for working in the field of integrated landscape development in Europe. Didactic material for learning special purpose language in the field of ecology and nature-conservation is developed. Following a pragmatic approach, already existing material from different sources is combined with newly developed modules. We distinguish textual contents which are available on the Internet or to be compiled into the LOGOS GAIAS repository from content modules which are combined from different terminological systems. The atomic content modules and procedural modules in GYMN@ZILLA operate freely on the user-selected content much to the benefit of the individual internet learner and teacher.

The paper starts with an overview of the LOGOS GAIAS Project. Then we present two already developed systems of project partners, the learner dictionary ELDIT and the terminological system BISTRO and claim that the combination of these already existing resources is an efficient way to build a repository of learning resources. In the second part of the paper we report on a prototype that actually combines several

modules: ELDIT's and BISTRO's terminological resources, BISTRO's dynamic text-term linking tool and a cloze generator. In order to be maximally independent of specific input text, these modules are combined in a browser-like application called GYMN@ZILLA.

LOGOS GAIAS Project Overview

In the LOGOS GAIAS project 16 different organizations from 6 European countries are collaborating and developing a hypermedia learning platform for basic language learning material in the field of ecology and nature-conservation. The material trains interdisciplinary users in the relevant language for special purpose. A multilingual database and multilingual language modules are developed in an interdisciplinary approach. Biology experts are collaborating with computer scientists and teachers. The intensive collaboration among these experts shall result in a useful and user-oriented learning resource. The end users are landscape planners and people working in natural preservation as well as farmers and rangers. The project is co-funded by the EU project Leonardo (http://europa.eu.int/comm/education/leonardo/leonardo2_en.html) and is operative from December 2001 until November 2004.

Manual elaboration of the content material

The first step was to collect the basic didactic material, namely the domain specific terminology, several short technical texts which contain this terminology, and real live pictures and video sequences to contextualize the technical terminology. This step is already concluded.

The second step is carried out currently: Translation equivalents for the technical terminology in different languages are elaborated. The following languages are supported: German, English, Czech, Italian, Hungarian, and Slovakian. Translations and exercises for the texts are prepared. All material is collected in an already existing multilingual terminological database. Definitions, synonyms, illustrative examples and links to the texts containing the terminology are to develop.

Future steps are the realization of interactive exercises for the terminology and the texts. Interactive fill-in-the-blank exercises can be created from the texts. Additional audio and video content will illustrate the information and provide the facilities for an applied and communication centered learning environment. Some of the texts will be recorded as sound files and integrated into the exercise module. The video sequences are annotated in order to provide explanations and hints for the idiomatic use of the technical terminology.

In a further step didactic scenarios are developed to support language for special purpose learning activities. The terminology, text and video material will be combined with the interactive exercises, adding instructional modules and an electronic help feature for the learner.

Those more complex content resources remain within the LOGOS GAIAS repository, whilst GYMN@ZILLA shall add more 'atomic' information referring to terminological data categories like word definitions, translations or phrases.

The dynamic inclusion of further material

The LOGOS GAIAS online learning platform will support terminology acquisition of six different languages. A huge amount of didactic material has to be collected and edited. This work can be greatly reduced by taking into account the possibilities of reusing and including already existing language learning and terminological material. Moreover the system can be enriched by additional reference sources such as electronic dictionaries and terminology databases if they are included wisely into the overall didactic concept. In fact, collecting and examining existing open language learning material that could be reused and included into the LOGOS GAIAS learning platform is one of the strongest focuses in project development.

The European Academy of Bolzano (http://www.eurac.edu), one of the 16 partner institutions in the LOGOS GAIAS project, houses several hypermedia projects. Two appropriate projects for inclusion in the

LOGOS GAIAS learning platform are the ELDIT learners' dictionary and the BISTRO terminology database. Both systems are freely accessible and composed of independent modules. First we briefly present the two systems and afterwards we describe the recombination of some of their modules into a new learning platform.
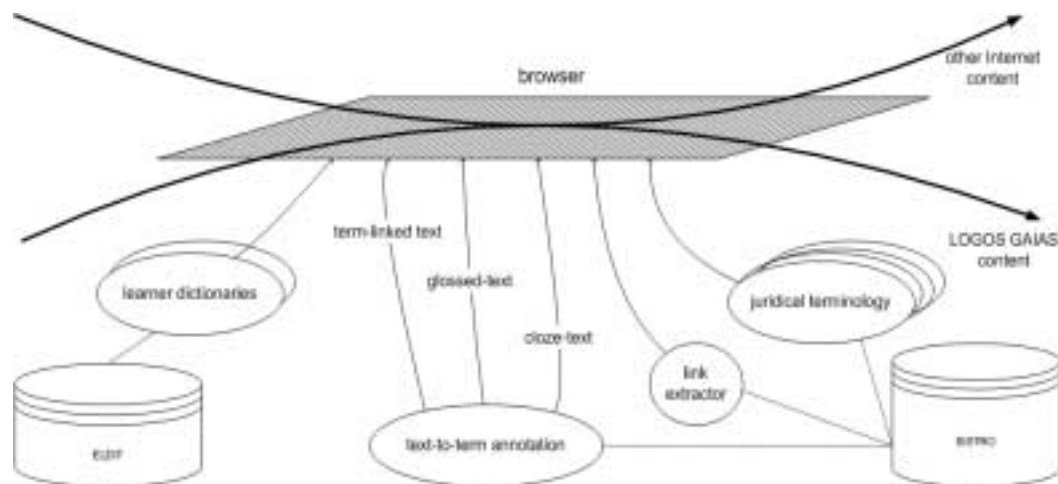

Figure 1: GYMN@ZILLA, the browser-like application which transforms textual content.

The ELDIT dictionary

The dictionary module is a union of two learners' dictionaries, namely a German and an Italian one. Each dictionary contains approximately 3000 word entries. The dictionaries are not meant as a pure information retrieval resource, but as a language learning tool. Therefore, every word entry is enriched with a huge amount of information which on one hand should help the learner to comprehend the right meaning of a word and on the other hand should help to use the word correctly. All these pieces of information are carefully selected and edited by language experts according to modern didactic and psycholinguistic criteria (Abel and Weber, 2000).

A dictionary entry is presented to the user in two frames (Fig. 1). The left-hand frame shows the **lemma** of the word and a list of different **word meanings**, each of which is described by a definition, an example sentence, and an optional translation equivalent in the other language. The right-hand frame is organized in several tabs and shows additional information such as **collocations, idiomatic expressions, compound words**, a **picture** and possibly **linguistic difficulties**. The linguistic difficulties can also be accessed directly on the place where they occur. They are indicated by a kind of footnote number and shown in a small window.

The ELDIT dictionary is one module of the ELDIT language learning system. Several other modules are currently developed, namely interactive **quizzes** which allow to practice the information presented in the dictionary, a learner **text corpus** consisting of 800 texts, an electronic **tandem feature** for collaborative language learning, a **user modeling** component which allows to adapt the system to the individual user and an **electronic tutor** which will guide the user through the learning process (Gamper and Knapp, 2002). Since approximately one year the first version of the dictionary module is online and can be freely accessed at http://www.eurac.edu/Eldit.
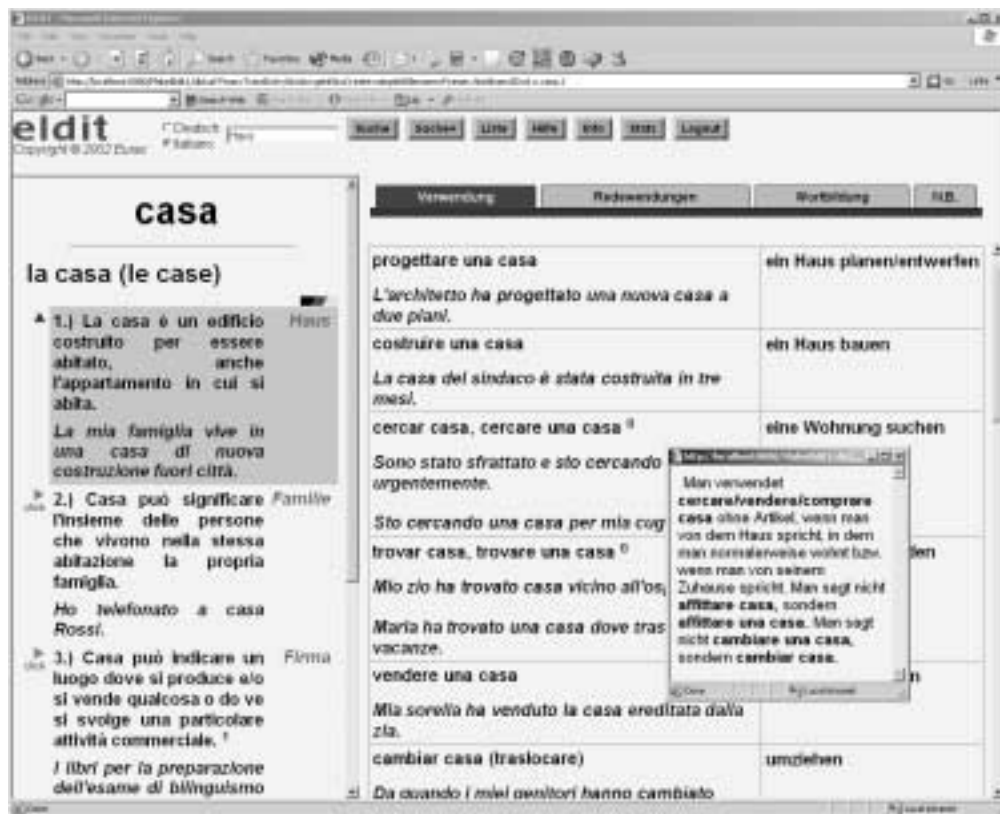
Figure 2: Screenshot of the Italian word "casa" (house) in ELDIT

The BISTRO terminology tool

BISTRO (http://www.eurac.edu/bistro) stands for „Juridical Terminology Information System Bolzano" and provides comparative information on 14.000 legal concepts for terminographers, lawyers and translators from Italy, Austria, Germany and Switzerland. Each concept is described in these four legal systems with a definition and a context from reliable sources and eventually by a short explanatory text. Yet in the Italian legal system not only Italian, but also German and two Rhaeto-Romance languages (Gherdëina, Badiot) have official status. BISTRO informs also about the correct expressions in these languages. The resources are elaborated by an interdisciplinary and multilingual team of professional terminographers and comparative lawyers of the concerned legal systems. BISTRO's huge amount of complex data requires a sophisticated user interface.

The user interface is characterized by a query piloting bar on the left, the actual information output in the middle and relevant legend about the used colors and signs at the right side (c.f. screenshot below).

The five modules term search, corpus search, text annotation, term extraction and SQL select shall guide the user to find the appropriate information. In the picture, the **term search** window is pulled down and the query for "arresto" (arrest) gives a hit list with an entry in penal law and another in penal procedure law, each with various German translations. Additional information on definition, context and recommended translations can be accessed through an internal link on the concept number. Further information on morphology and English translations is linked externally (c.f. screenshot: links "MOR" and "LEO" in parenthesis after the German term).

Also a query in the second module **corpus search** gives a hit list of internal and external links. Internally one of the largest special language corpora is searched: The 5 million words legal corpus Italian and German parallel CATEX corpus (Gamper 99). The external link is a meta-search hit list in query-pertinent legal resources in the internet.

The **text annotation** tool allows the dynamic linking of web-pages or ASCII-text with the BISTRO terminology. The output screen is the text selected by its URL. The original links contained in the text are removed and made available in a separate window. Instead, new in-line links are generated which relate the terms in the text to terminological entries.

The fourth module **term extraction** does what its name promises: Bistro fetches a given web page and extracts term candidates. This tool is useful for terminographers, but it is also a learner's tool, for example an interpreter can recapitulate and prepare vocabulary in a special language field. The Bistro team is especially proud that term extraction works well even for the lesser used Rhaeto-Romance languages Gherdëina, Badiot and Fashan (Streiter et al. 2002).
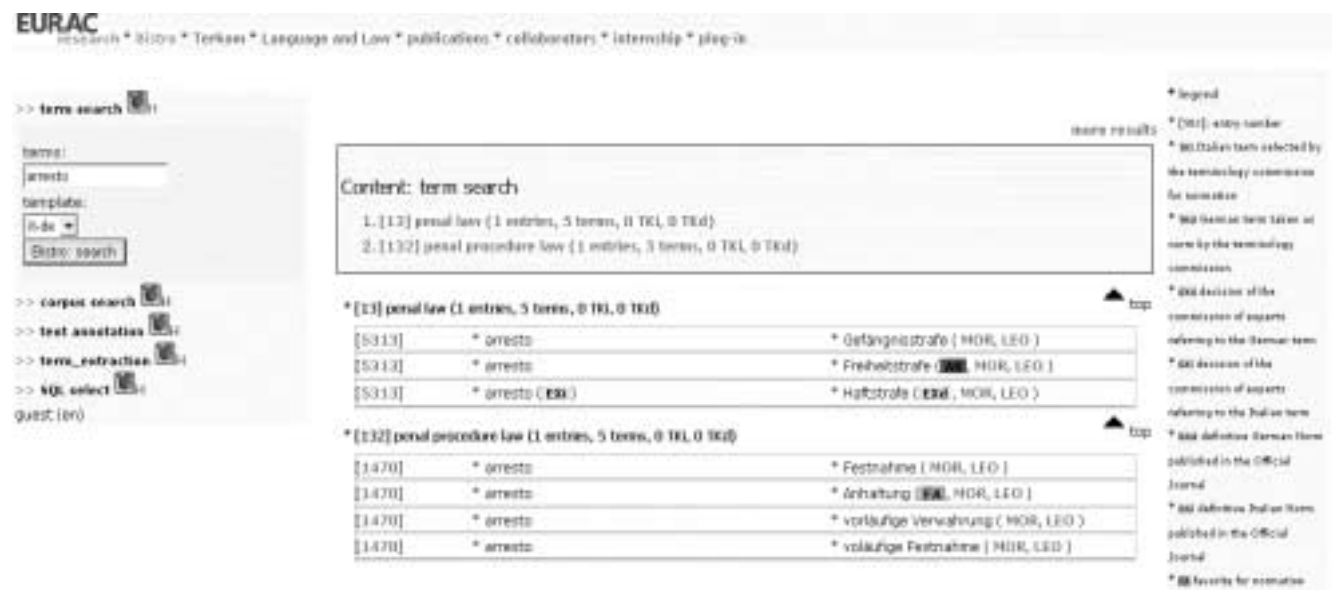

Figure 3: Screenshot of the Italian word "arresto" (arrest) in BISTRO.

GYMN@ZILLA: Integration of ELDIT and BISTRO modules in LOGOS GAIAS

Our goal is to combine different content modules with procedural modules for the creation of a new learning environment. Presupposition was the modular concept of both ELDIT and BISTRO: Data elaboration process and software development are independent of each other and each module of the two systems can be accessed independently.

Until now we have integrated the dictionaries of ELDIT and BISTRO as content modules and the text-to-term linking tool of BISTRO as procedural module. Special output formats of this module within GYMN@ZILLA are responsible for the creation of cloze texts and glossaries of various types. The system we created is a browser which dynamically processes text (submitted ASCII text or text on the Web) in the following ways:

(1) linking basic vocabulary to the ELDIT learners dictionary vocabulary,
(2) linking language for special purposes to the BISTRO terminology database (Fig. 4),
(3) creating completion exercises on the basis of BISTRO terminology
(4) creating completion exercises on the basis of ELDIT terminology
(5) creating glossaries with BISTRO's terminological content
(6) creating glossaries with ELDIT's terminological content
(7) extraction of new terms from the text

The browser has some interesting aspects we would like to point out:

* The links which are contained in a text can be followed (Fig. 4) with the processing mode maintained for the following texts: If a user chooses the option to annotate an internet page with ELDIT annotations, this operation will be applied automatically to the following pages which are accessed through the browser unless specified differently.
* A language recognition feature has been implemented to guess the language of the accessed document for proper processing.
* With several dictionaries integrated, the user can choose between special or common language annotations or opt for both of them (BISTRO annotation vs. ELDIT annotation).
* Since we have rich data both in ELDIT and BISTRO, it is not only possible to annotate the terms with their translations but also with synonyms, with a definition, with a context or with an example.
* The annotated terms can be extracted and viewed separately. They form a special glossary to the input text which may be printed or read by text-to-speech programs. Of course also the term's translation or other information (grammar, morphology …) can be collected in this glossary.
* Applying the above mentioned possibilities we can also provide an exercise generation tool which dynamically creates various types of completion exercises. With a technically very simple style variation the software substitutes the terms and dynamically creates a cloze. The correct answer is linked. Alternatively, the cloze's gaps can be filled with the initial or last letters of the omitted term.

Figure 4: Screenshot of GYMN@ZILLA and a BISTRO annotated text

Learning with technology has been proven to be very motivating especially for young students (Sentance 1997, Goodfellow & Lamy 1998, Leh 1999, Egert 2000). The positive effect on learning with glossed texts and

with the other above mentioned possibilities has also been shown in several studies (Nerbonne et al.1998, Laufer 2000, De Ridder 2000). Our browser provides all these features. Moreover, a learner can both, maintain a certain study concept and individualize learning, by browsing the Internet, because the browser sticks with a once chosen processing style (term-linking, glossing or gap-generation). The didactic implications of this approach will be subject to research within the LOGOS GAIAS project.

Conclusions

For us several other interesting facts turned out during the development of GYMN@ZILLA: The first and most interesting point is that the **combination of** pure **content** modules (ELDIT's and BISTRO's dictionaries) **and** a non learner-oriented **software** module (BISTRO's term annotation tool) **resulted in** a highly **didactic learner's environment**. With GYMN@ZILLA we have demonstrated that ELDIT and BISTRO modules are open and ready to work in combination with each other's, such that typical learning material can be produced with little effort.

Secondly, we have the possibility to switch swiftly between languages and subject areas, but also between prepared content, the user's private texts and the Internet. This language learning module provides hence a degree of flexibility and actuality that is second to none.

Thirdly, the successful implementation of GYMN@ZILLA opens a **large room for further experiments**. As to content, resources of other authors, other vocabulary, other languages and language specializations can be connected and thus be extended to all partner languages (Hungarian, Slovakian...) As to processing modules, BISTRO offers further possibilities such as corpus-search, text-classification and the corpus collection with a web-robot. We shall carefully examine how these modules might contribute further to a dynamic learner platform.

Figure 5: Screenshot of GYMN@ZILLA and a cloze text

References

Abel, A. & Weber, V. (2000). *ELDIT, prototype of an innovative dictionary.* In Proceedings of the 9th EURALEX International Congress on Lexicography (EURALEX'00), Stuttgart, Germany.

Egert, C. (2000). Language learning across campuses. *CALL,* Vol. 13(2).

Gamper, J., & Knapp, J. (2002). *A Web-based Language Learning System.* In Proceedings of the 1st International Conference on Web-based Learning (ICWL2002), Springer, Lecture Notes in Computer Science.

Goodfellow, R. & Lamy, M.N. (1998). Learning to learn a language - at home and on the Web. ReCALL 10(1).

Laufer, B. (2000) Electronic dictionaries and incidental vocabulary acquisition: Does technology make a difference? In *Proceedings of EURALEX'00.*

Leh, A.S.C (1999). Computer-mediated communication and foreign language learning via electronic mail. *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, 1(2), October 1999.

Nerbonne, J & Dokter, D. & Smit, P. (1998) Morphological processing and computer-assisted language learning. *CALL*, 11(5).

De Ridder, I. (2000) Are we conditioned to follow links? Highlights in CALL materials and their impacts on the reading process. *CALL.*

Sentance, S. (1997). A rule network for English article usage within an intelligent language tutoring system. *CALL*, 10(2):173–200.

Streiter, O. & Zielinski, D. & Ties, I. & Voltmer, L. (2002) *Similarity-based Term Extraction for Minority Languages: A case-study on Ladin*, In Proceedings Soziolinguistica y Language Planning, Ortisei, Italy.

Gamper, J. & Dongilli, P. (1999). Primary Data Encoding of a Bilingual Corpus. *Proceedings of the 11th Annual Meeting of the GLDV,* Frankfurt a/M, Germany.