

Technical Aspects of a Phraseology E-Learning Application

Denis Helic

Institute for Information Systems and Computer Media

University of Technology Graz

dhelic@iicm.edu

Abstract: This paper discusses technical aspects of a phraseology E-Learning application that is being developed in the scope of the Ephras project. The application includes a phraseology database containing phrases in four different European languages and a collection of interactive tests to selected phrases from the database. The first part of the paper describes an XML-based solution for the phraseology database. A discussion on different implications of XML technology on the architecture and implementation of the database concludes this section. In the second part of the paper implementation of interactive tests using IMS Question & Test Interoperability (QTI) standard is discussed. Amongst others, the paper presents a number of practical problems related to an application of QTI in phraseology, such as missing features or lack of didactics support. Finally, the paper introduces a possible solution to those problems in a form of a template-based authoring approach.

Keywords: Phraseology, E-Learning, XML, QTI, Templates, Authoring

1. Introduction

The Ephras project (Ephras, 2005) is a project funded by the European Commission under Socrates/Lingua2 programme. The goal of the project is to develop an E-Learning application for phraseology in four European languages - German, Slovak, Slovenian and

Hungarian language. The project aims to eliminate the lack of such phraseology computer-based learning material, as well as to meet the demands for multilingual learning material in the enlarged European Union. The Ephras E-Learning application is composed of a searchable database (Ephras Database) of 4x1000 phraseology data items in four languages (i.e., 1000 data items in each of the languages) accompanied with 150 interactive tests (Ephras Tests) to selected data items.

From the user point of view the application has the following properties. The user can search within the database using any of four languages as the starting language. The search results are presented in a list form where the user can click on a particular search result and obtain the full description of a particular data item. The links to the related data items (e.g., equivalent data items from other languages) are included in the data item description. Additionally, a link to an interactive test is provided if such a test exist for the selected data item. The interactive tests serve to check understanding and knowledge of the selected phrases, as well as to improve the skills for producing these phrases in written and spoken foreign language. As such they follow a sound didactic approach developed within the project in the form of an exercise typology. Moreover, the tests are aligned according to language, topic, knowledge, and skill level of the users.

2. Ephras Database

The database component of the Ephras application has the following properties. First, a special description model has been created to represent the phrases. This description model has been developed according to the latest phraseology principles and includes the following fields: basic form (i.e., the content of a phrase), meanings, style, grammar, collocation, pragmatics, examples, variants, keywords, synonyms, categories and multilingual comment. One single record in the database represented by this description model is called a data item.

Second, the source language is German with 1000 data items, where each of these data items is a single phrase in German with one or more meanings. Additionally, each German phrase is involved in a so-called equivalence relation with data items from other

three languages (target languages). The equivalence relation expresses rather complex phraseology relationships between different data items. For example, in the simplest case the equivalence relation might express a one-to-one relation between a German phrase and a phrase in one of the target languages, where both of the phrases have only a single meaning. However, in a more general case the equivalence relation might represent a one-to-many relation between a German phrase and a number of phrases in one of the target languages, where all of the phrases in question might have multiple meanings. Sometimes there is no direct phraseology equivalent for a German phrase in the target languages. The equivalents in that case are non-phraseology data items, i.e., a single word or a free collocation. Thus, the equivalence relation in such a case is a one-to-one relation between the German phrase and its corresponding free collocation.

Third, in addition to the primary direction of the equivalence relation (i.e., the direction from German to other three languages) the secondary direction of this relation can be established as well (see Figure 1). In some special cases (e.g., when only one-to-one relations are present) it is possible to infer the equivalence between data items from the target languages. For example, starting from a Hungarian data item it is possible to find its equivalent in Slovak by implicitly using the existing one-to-one relations between those two data items and their German counterpart.

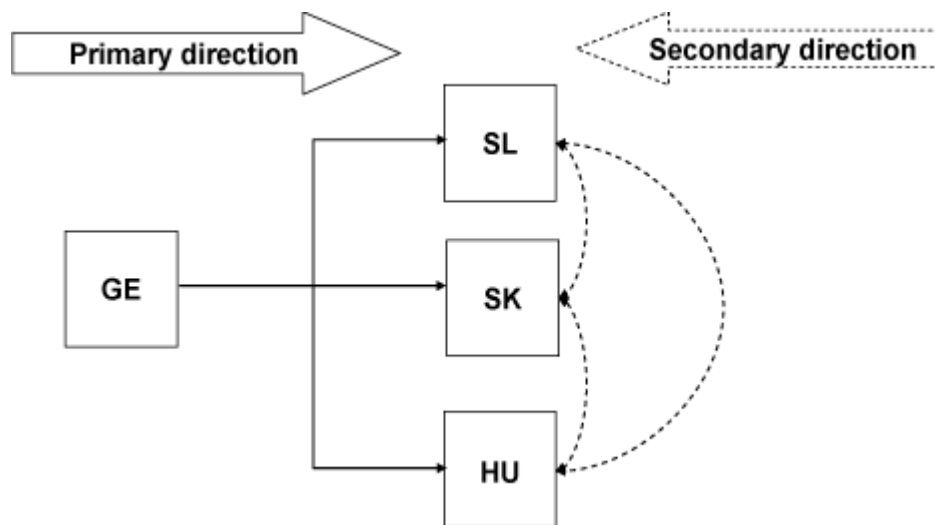


Figure 1: Directions of equivalence relation

3. Using XML to Implement Ephras Database

As shown by Helic and Durco (Helic and Durco, 2005) data with such properties like the phraseology data in question is referred to as semi-structured data (Abiteboul, 1997; Buneman, 1997). Recently, through the emergence of the Web and the related mark-up technologies, such as eXtensible Markup Language (XML), the latter evolved to a de-facto standard for managing semi-structured data (Goldman et al., 1999; Vianu, 2001). An XML document is a hierarchy of elements with ordered sub-elements. Each element has a name (also referred to as label or tag). The basic XML model is a labelled ordered tree where labels represent node names. Edges are always directed (to preserve the tree order) and do not have labels. Additionally, XML supports a referencing mechanism between nodes, which basically facilitates modelling of arbitrary graphs. Listing 1 shows an example of an XML document representing a couple of phrases.

```
<phrase id="1">
  <content>am/auf dem grünen Tisch</content>
  <meanings>
    <meaning>Etwas am ... </ meaning>
    <meaning>Ohne Kenntnis ... </ meaning>
    <meaning>Etwas nur ... </ meaning>
  </meanings>
  <style>...</style>
  <examples>
    <example>...</example>
  </examples>
  ...
</phrase>
<phrase id="11" variant="1">
  ...
</phrase>
<phrase id ="101" equivalent="1">
  ...
</phrase>
```

Listing 1: XML representation of phraseology data items

In the paper by Helic and Durco (Helic and Durco, 2005) the aspects of using of XML for the purposes of data representation, storage and persistence have been discussed. In this paper a special emphasis is given to the data presentation for the users.

Originally, XML is specified as a meta-document format that is used to define families of document formats. Definition of presentation instructions for such document families is not a part of XML specification and is defined elsewhere - namely by a number of so-called style-sheet specifications. Basically, a style-sheet is a separate document which defines how a certain XML document should be presented.

There exist three different approaches for such a definition. The first approach deals with specifying so-called formatting instructions, i.e., there exist a formatting instruction for each XML element that precisely specifies how that element should be presented. For example, one can specify that the <meanings> element should be presented as a numbered list or that the <content> element should be presented using bold font weight. So-called Cascading Style Sheets (CSS) represent a typical example of such an approach.

The second approach is based on a transformation of XML documents into target formats for which presentation semantics exists. Typical examples of such target formats include Hypertext Markup Language (HTML) and Portable Document Format (PDF). The definition of such a transformation includes a definition of how each particular XML element, together with its content and sub-elements are transformed into one or more elements in the target format. For example, one may specify that the <meanings> element and all of its <meaning> sub-elements are transformed into a combination of an HTML numbered list element and its list item elements. Note here, that the transformation approach is based on manipulation of the XML elements on-the-fly. Therefore, it is also possible to sort, filter or modify the content during a transformation.

Finally, the third approach is a hybrid approach which combines the other two approaches. For example, at the first step an XML document is transformed into an HTML document. At the final step a CSS style-sheet is applied to format the obtained HTML document.

In the Ephras application we have chosen to apply the hybrid approach because of the following advantages:

- The learning material presented to the users should be aligned according to the language, knowledge level and the skills of the users. The possibilities for dynamic manipulation of the content during the transformation phase of the data items prove very valuable in implementing the needed functionality. For

example, for users with a lower knowledge level certain sophisticated information might be filtered out.

- The transformation style-sheet can be easily exchanged (even in the run-time) to produce further output formats. Currently, the Ephras application produces only an HTML-based presentation for the user screens. However, the application can be easily extended to produce a PDF-based output for printing purposes.
- The transformation style-sheets produce raw HTML pages without any formatting specifications. To improve the presentation of such HTML pages CSS style-sheets should be used. Moreover, design, accessibility and usability issues are best solved through the use of CSS style-sheets. For example, one might specify an alternative CSS style-sheet that formats the HTML content using a larger font or alternative colours to support visually impaired users.

An example of an HTML document that contains information of a German phrase might be seen in Figure 2.

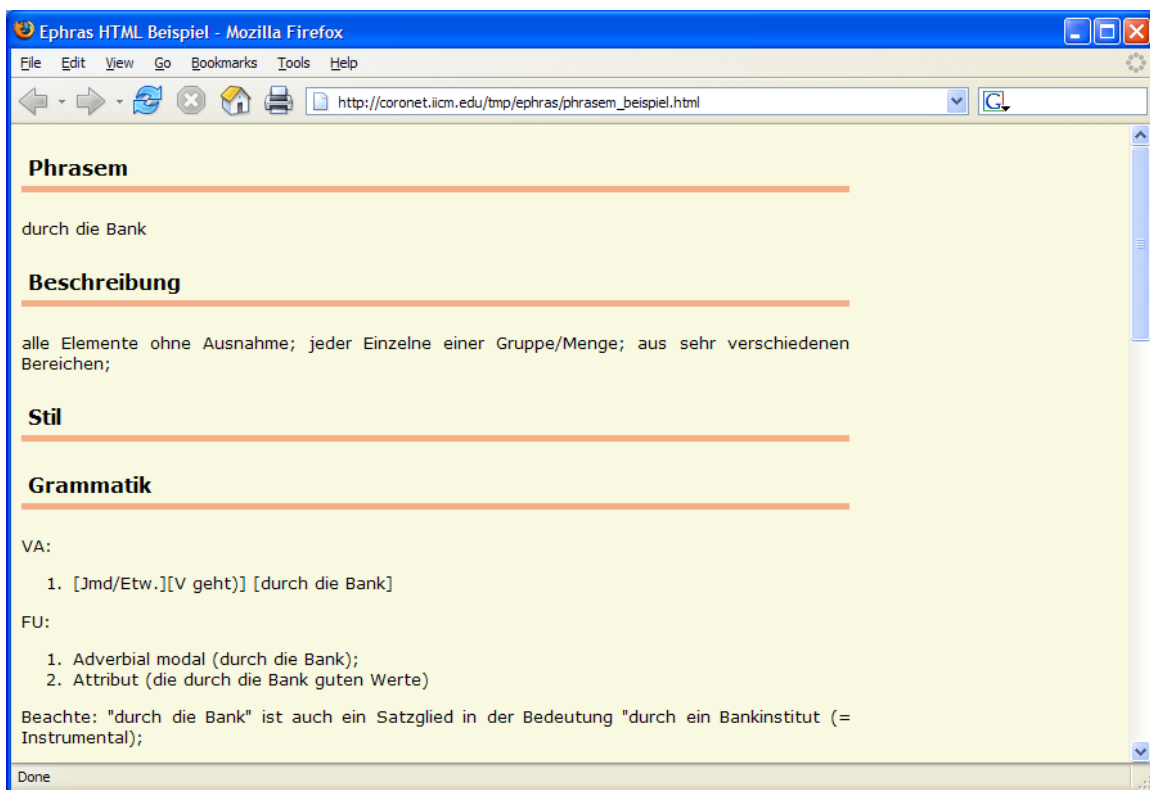


Figure 2: Example of phrase presentation in HTML

4. Ephras Tests

The basis for development of the interactive tests in the Ephras project is a sound and successful didactical approach for teaching phrases in foreign languages. This approach includes learning activities such as phrase recognition, phrase meaning or phrase pragmatics identification, phrase form and grammar understanding, consolidation and reflection. Each of these activities is typically represented by a number of exercise types which are used in accordance with the current context, as well as the user's knowledge and skill level. For example, identifying of phrase meaning can be realized with a multiple choice question where the student needs to select the correct meaning of a phrase from a number of possible answers, with a short text essay where the student needs to write down the meaning of the phrase, or with a combination of these two question types. The exercise types for the Ephras tests include:

- Multiple choice exercises where the users need to select a correct answer among a number of answers.
- Select text exercises where the users are supposed to select the correct sentence containing a particular phrase from a larger text.
- Free text essay exercises where the users are supposed to explain the meaning of a phrase using their own words.
- Ordering exercises where the users are supposed to match a number of phrases with their respective meanings by dragging and dropping the phrases on the screen.
- Fill-the-gap exercises where the users need to type in the missing words or characters in a larger text.
- Crossword-puzzle exercises where the users are supposed to fill in a crossword puzzle to obtain a particular solution.
- Phrase creating exercises where the users need to create on their own a number of phrases by dragging and dropping arbitrary words presented on the screen.
- Phrase reshuffling exercises where the users need to reshuffle the words from a phrase to obtain its proper form.

- Equivalent guessing exercises where the users need to guess a proper equivalent of a particular phrase in another language.
- Equivalent translation exercises where the users need to translate a phrase in another language.

A typical Ephras test is a sequence of exercises of different types that takes into account the language, knowledge and skill level of the users.

5. Using IMS QTI to Implement Ephras Tests

IMS Question & Test Interoperability (QTI) (IMS QTI, 2005) is a de facto standard for cross-platform representation of questions and tests in E-Learning systems. The benefits of such a common representation are manifold. First, it provides a question/test format for authoring tools therewith allowing the developers of these tools to concentrate on the usability issues instead of the question/test features. Second, it supports development of question/test databases that are easily integrated and shared among a variety of systems because of a common question/test database schema. Last but not least, it provides a clearly defined interface for retrieving of test results which can be very important for management and adaptation of the learning process in accordance with the test results (Davis and Davis, 2005).

Because of such advantages of the QTI standard we decided to apply this format for development of the Ephras tests. Even though the QTI standard offers the above mentioned benefits its application in the Ephras project is related with a number of difficulties. These difficulties are caused by an impedance mismatch between the QTI standard and its concrete application in a specific subject area such as foreign language teaching. This impedance mismatch is visible at two levels. First, the QTI standard is a technical specification which supports development of questions and tests for various subject domains where the questions and tests are interoperable at the level of authoring tools, question/test databases or E-Learning systems As such the QTI standard defines a number of general question types that might be applied in a number of areas and does not take into account specific question and test types of a particular domain (Milligan, 2003).

However, there exist a number of specific question types which are commonly used in foreign language teaching but are not reflected in the QTI standard, with crossword puzzles or phrase reshuffling exercises being typical examples. Second, the QTI standard is not concerned with didactical issues and, as a matter of fact, it tries to be as didactically neutral as possible (Smythe and Roberts, 2000).

To resolve the impedance mismatch and tackle the problems caused by that mismatch in the Ephras application a template-based solution has been applied. Thus, a number of question and test templates have been developed. Each of these templates bridges a single gap between the features from the Ephras tests and the features offered by the QTI standard.

For example, a crossword puzzle template has been developed to support creation of typical crossword tests. To create a crossword puzzle a number of parameters need to be defined including the number of rows, the number of columns, the row questions, and the solution of the crossword puzzle. These parameters are defined by operating a QTI compliant editor (see Figure 3). In the Ephras project we used the Author from the Canvas Learning (<http://www.canvaslearning.com/>) (Canvas Learning, 2005).

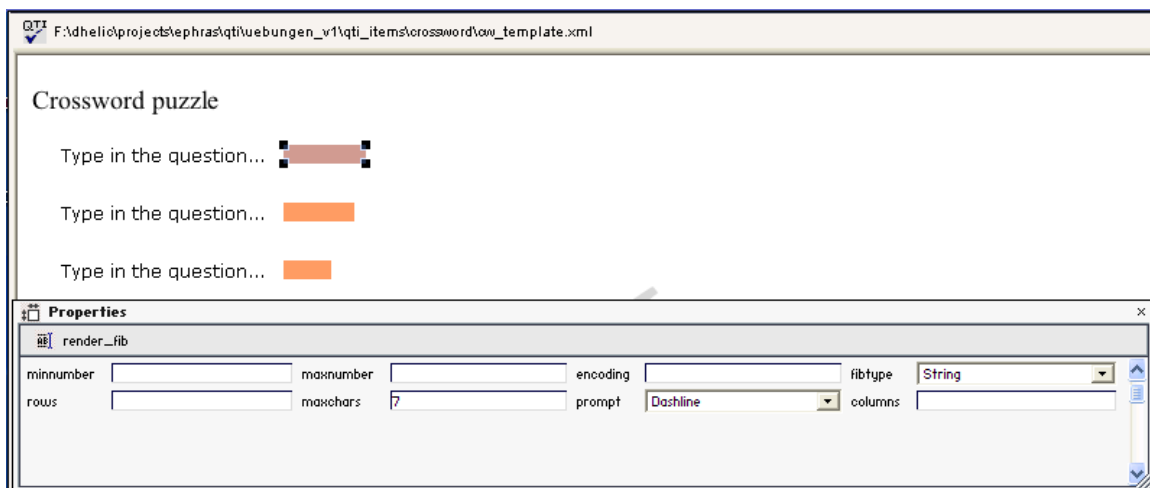


Figure 3: Editing the crossword puzzle template

Internally, the crossword puzzle is mapped onto QTI questions, such as single and multiple fill-in-gap questions with appropriate layout. However, through the abstraction offered by the crossword puzzle template the authors do not need to deal with the underlying low-level QTI solution.

The same template-based solution has been applied to resolve the impedance mismatch problem at the didactical level. Thus, for each learning activity mentioned above several templates have been developed. Each of these templates reflects a single possibility for supporting a particular activity. For example, to support phrase meaning identification activity three templates have been developed. The first template comprises a multiple choice question and a number of possible answers where only one single answer is correct. The second template includes a short essay question, where students need to type in their own textual explanation of the phrase meaning. The third template combines the first two, i.e., it includes a multiple choice question together with a short essay question.

Again, the authors do not need to be concerned with the internals of the QTI standard. Rather they work with the QTI editor and only define a couple of parameters to create a working template instance (see Figure 4).

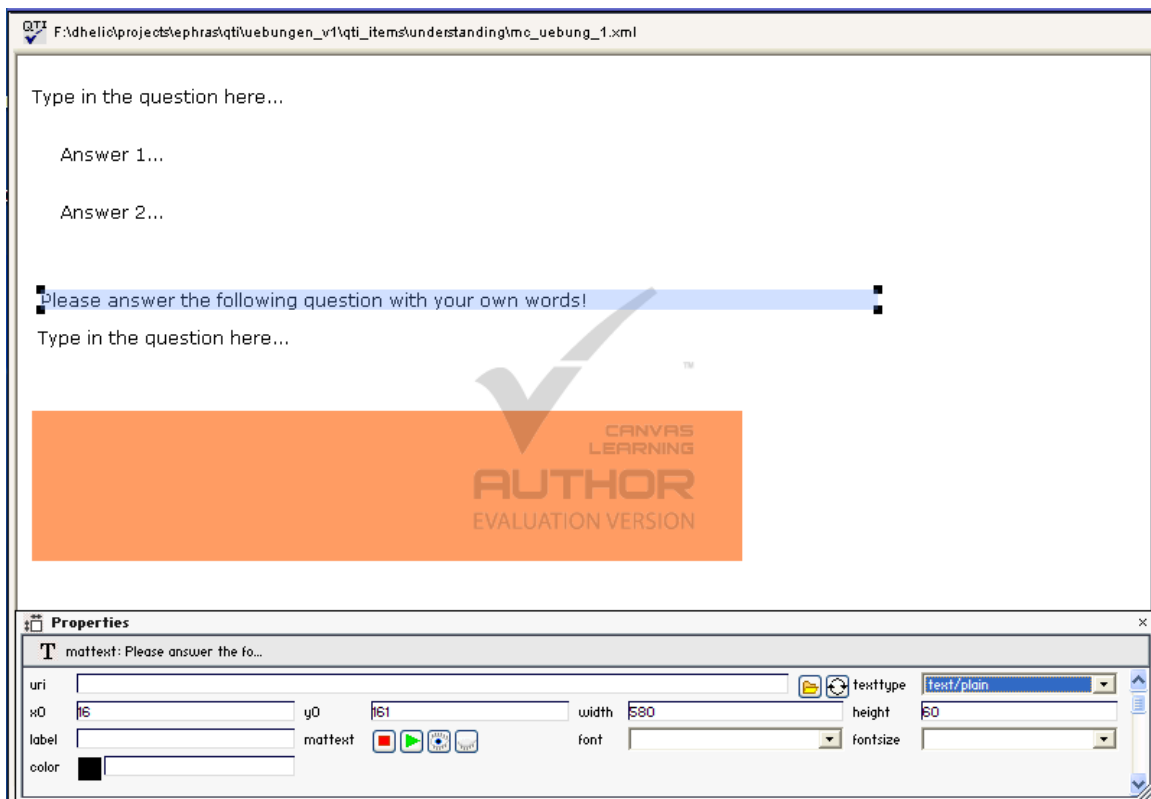


Figure 4: Editing the phrase understanding template

6. Conclusion

In this paper the technical aspects of a phraseology E-Learning application called Ephras have been discussed. Since the phraseology data managed in the Ephras application exhibits the properties of semi-structured data an XML-based format has been applied to represent it. Aspects of an application of style-sheets technology for presentation of such phraseology data have been discussed. The presented solution is general enough to allow data presentation in different formats and for different purposes, i.e., on-screen presentation or printing purposes.

The second component of Ephras application – Ephras tests – follows a sound didactical approach to improve knowledge and skills of the users in understanding, recognizing, writing or using phrases in spoken foreign language. The implementation of the tests has been based on the recent E-Learning standard for development of interactive questions and tests – IMS QTI. To bridge the gap between the sophisticated features of the tests used in the Ephras application and the QTI features a template-based approach to authoring and development of the tests has been applied on the top of IMS QTI.

Acknowledgments

This work has been supported by the European Commission Socrates/Lingua2 Programme project called Ephras.

References

Abiteboul S., Querying Semi-structured Data, In *Proceedings of the 6th International Conference on Database Theory - ICDT '97*, pages 1–18, 1997.

Buneman P., Semistructured Data, In *PODS '97: Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, pages 117–121, New York, NY, USA, 1997.

Canvas Learning, <http://www.canvaslearning.com/>, 2005

Davies, W. M. and Davis, H. C., Designing Assessment Tools in a Service Oriented Architecture, In *Proceedings of 1st International ELeGI Conference on Advanced Technology for Enhanced Learning* (in press), Napoli, Italy, 2005.

Ephras, EU Socrates/Lingua2 programme, <http://www.ephras.org/>, 2005.

Goldman R., McHugh J., and Widom J., From Semistructured Data to XML: Migrating the Lore Data Model and Query Language, In *Proceedings of the 2nd International Workshop on the Web and Databases (WebDB '99)*, 1999.

Helic D., and Durco P., Aspects of an XML-Based Phraseology Database Application, To be published in *Proceedings of Slovko 2005 - Computer Treatment of Slavic and East European Languages*, 2005.

IMS QTI, *IMS Global Learning Consortium Question & Test Interoperability Specification (QTI)*, <http://www.imsglobal.org/question/>, 2005.

Milligan C., Question and Test Interoperability (QTI): Extending the specification for Mathematics and Numerical Disciplines, *Maths CAA Series*, The Maths, Stats & OR Network, University of Birmingham, UK, Nov. 2003.

Smythe, C & Roberts, P., An Overview of the IMS Question & Test Interoperability Specification, *Computer Aided Assessment (CAA'2000)*, Leicestershire, UK, 2000.

Vianu V., A Web Odyssey: from Codd to XML, In PODS '01: *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 1–15, New York, NY, USA, 2001.