An Ontological Approach for Multicultural Digital Repositories

Pavel Boytchev[1], Stefania Bocconi[2], and Krassen Stefanov[1]

[1] FMI, Sofia University, Sofia, Bulgaria
{boytchev, krassen}@fmi.uni-sofia.bg
[2] Istituto per le Tecnologie Didattiche – C.N.R., Genova, Italy
bocconi@itd.cnr.it

Abstract. The EC funded Share.TEC project aims to provide culturally aware access to resources across Europe related to Teacher Education. The semantic core of the system is the Teacher Education Ontology (TEO), which not only supports the multilingual interface but also underpins fundamental functionalities like searching and browsing. This paper describes an approach that systematically reorganizes the structure of TEO to leverage the existing multicultural dimension. Special attention is focused on representing situations where the semantics of concepts do not map completely across cultures.

Keywords: teacher education, ontology, multilinguality, multiculturalism

1 Introduction

Share.TEC[1] is an ongoing international project aimed at promoting innovation in Teacher Education (TE). The project is building an advanced user-focused system fostering a stronger digital culture in the TE field by aggregating metadata for TErelated digital resources located Europe-wide; providing personalized, culturallysensitive brokerage for the retrieval of relevant digital content; and supporting the development of a Europe-wide perspective among those working in and with the TE community. The semantic backbone of the Share.TEC system is the Teacher Education Ontology (TEO), which has been set on concepts relevant to the domain of TE and the potential members of the Share.TEC community. The purpose of TEO is to provide:

- pedagogical characterization of digital content;
- representation of user profiles and competencies;
- a basis for multilingual and multicultural functionality.

One of the main expectations is that Share.TEC users will bring to the community different languages and cultures. The TE field includes people with very different backgrounds, ideas and assumptions. Instead of flattening out all cultural variations, Share.TEC's goal is to support this diversity by proving a system that has capabilities for dealing with different national languages and also with culturally-related variation across contexts. TEO also provides the

[1] Share.TEC - SHAring Digital REsources in the Teaching Education Community, eContentplus programme (ECP 2007 EDU 427015); http://www.sharetecproject.eu/.

basis for the definition of a common metadata model for describing TE-relevant digital resources and TE-relevant user skills, experiences and interests.

2 Teacher Education Ontology (TEO)

The rationale for adopting an ontology-based approach in Share.TEC is many-faceted. Firstly, is *sharing concepts amongst humans*. Potential users of Share.TEC come from a wide variety of TE-contexts and backgrounds, and bring with them different perspectives and assumptions. Even when referring to ostensibly "common" concepts, they adopt different terminology, expressed in a variety of languages. TEO is aimed at reducing conceptual and terminological confusion by identifying and properly defining a set of relevant concepts (and their relations) that characterize the TE domain in Europe. It also provides a reference framework for situating culturally and linguistically diverse versions thereof. Furthermore, TEO supports adaptive user interfaces, applications and services that make use of reasoning techniques, thus allowing the implementation of inferential search engines, flexible representation of user profiles, and advanced ranking solutions.

TEO has been built by consulting three main reference models, OMNIBUS[2], LORNET[3] and POEM [1], which provide vocabularies for concepts relevant to TE. Other relevant sources that have influenced TEO are DOLCE[4], ONTOURAL [2], ALOCOM[5], PROTON[6] and user modeling ontologies [3], [4].

TEO is organized in several branches, which define subdomains within the larger TE domain. The largest and most complex branches are: *digital content*, referring to educational resources and learning objects; *knowledge area* based on the EUROSTAT [5] taxonomy of education and training, the branch allowing description of digital content and user's interests; and *role*, drawn from Mizoguchi's model [6].

The internal structure of a TEO entity is designed as the simplest one that facilitates all required functionalities: (a) language-neutral concept-oriented data; (b) hierarchal searching and filtering; (c) dynamic multilingual user interface. Each TEO entity is represented as an interconnected node, which contains not only relations to other nodes, but also a list of translations of the concept represented in the node.

The actual internal structure is more complex. This defines a wider spectrum of relations between ontology entities and allows a complete reconstruction of the ontology into a valid OWL file and support for extended functionalities like reasoning.

3 Providing multilingual support

One of the goals that TEO has been designed to pursue is to allow multilingual support in the easiest possible way. Every node contains a set of translations of

[2] http://edont.qee.jp/omnibus/doku.php
[3] http://www.lornet.org
[4] http://www.loa-cnr.it/DOLCE.html
[5] http://ariadne.cs.kuleuven.be/alocom/
[6] http://proton.semanticweb.org/

the node's concept into several European languages. These translations are used when multilingual data from repositories across Europe are being processed.

The process of writing data to the Share.TEC repository, as well as reading it, always passes through the concept node, which is multilingual by construction. Metadata for which there is a corresponding concept in the ontology derive their linguistic representation from the ontology itself. This makes the internal representation of data language-independent and connects various translations of the same concept. Consequently, an Irish content provider for example can define the topic of a document as "Medicine", see Fig. 1. When the description of this document is harvested, it will be bound to the concept "Medicine" from TEO. When the same description is retrieved by an Italian or a Bulgarian user, they will see the text "Medicina" or "Медицина".

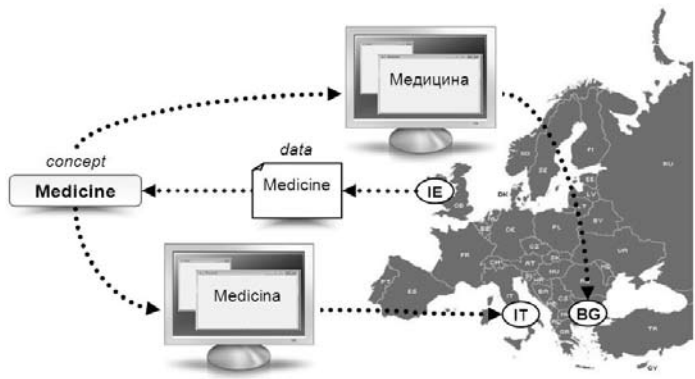


Fig. 1. Native languages and language-neutral conceptualization.

Translations are actually used not only for importing and displaying data, but also for searching and filtering. A Bulgarian user may define a search criterion in Bulgarian; the system will match it to the corresponding concept node and will find all data referring to that concept node, irrespective of the data language.

4 Multicultural approach and challenges

Unfortunately, multilingual systems do not provide enough level of adaptation to user expectations. The domain of teacher education contains several branches of taxonomies which cannot be directly translated into all other languages, because the education systems and the culture behind them are different.

The initial approach adopted in Share.TEC was to treat these differences as special cases where differences are represented in disjoint subclasses and commonalities are captured in superclasses, as described in [7]. A more systematic approach could start from the fact that two concepts from two cultures may be related in one of several different ways: they may fit each other exactly; one of them may be a subset/superset of the other; they may intersect; or they may be non-intersecting.

Multilinguality provides an adequate support only for the first relation, when both concepts have the same interpretation (at least ostensibly). The other three relations are either impossible to support or can only be represented at the cost

of reduced accuracy.

Table 1 represents two *imaginary* sets of concepts about the education systems in two cultures: S and O (named after the last letter of words in the corresponding languages). The education systems of both cultures have three levels: *primary*, *secondary* and *tertiary*. However, there are differences in the interpretations of these levels. For example, the tertiary level in culture S (called *tertius*) comprises grades 9 to 11, while the corresponding level in culture O (called *tertio*) comprises grades 9 to 12. Also, culture S has a pre-primary level, which is nonexistent in culture O.

A multilingual system cannot correctly handle searching for *secundus*, because it will never find *post-secundo*, even if the document refers to the same age and grade.

Table 1. Imaginary example for culture discrepancy.

Educational level	Culture S		Culture O	
	Grades	Translation	Grades	Translation
Pre-primary	-	preprimus	-	-
Primary	1...4	primus	1...4	Primo
Secondary	5...8	secundus	5...7	secundo
Post-secondary	-	-	8	post-secundo
Tertiary	9...11	tertius	9...12	Tertio

To enhance the effectiveness with which a system handles cultural differences, the main challenge is to provide systematic, formalised mapping between two or more notquite compatible cultures. While TEO accommodates multiculturalism to a degree within the constraints of the Share.TEC project, a more systematic approach to mapping would enhance multicultural responsiveness. However, planning the implementation of such a thorough revision would need to be carefully evaluated with respect to the project's time constraints and practical limitations.

If we map graphically the education system of both cultures it would be much easier to formalize the level of discrepancy. Fig. 2 represents various combinations of possible mappings between concepts in both cultures. Here are some of these combinations:

- *Case 1*: A concept exists only in one culture, the other culture has no corresponding concept (see pre-primary)
- *Case 2*: A concept exists in both cultures and has exactly the same meaning (see primary)
- *Case 3*: A concept in a culture is a subset of a similar concept in the other culture (see secondary and tertiary)
- *Case 4*: A concept in a culture maps completely to several concepts in the other culture (see secondary)
- *Case 5*: A concept in one culture corresponds to a similar concept in another culture, but also has a further meaning that does not find correspondence in the other (see tertiary)

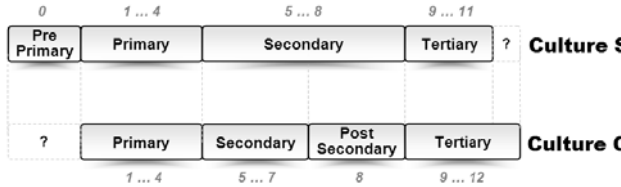


Fig. 2. Graphical representation of cultural discrepancy.

5 Enhancing TEO's capabilities for tackling multiculturalism

This section describes a possible alternative solution for handling incompatibilities between cultures that reuses the current logical and physical structure of TEO. This solution has the added advantage of not imposing new relations between nodes or new structures in the underlying repository. Therefore it does not jeopardize existing metadata assets. Currently, this proposed solution has not yet been implemented in Share.TEC and possible implications and requirements are discussed in the following sections.

5.1 A new organizational model for the ontology

The backbone of this approach to multiculturalism is to provide a unified conceptual ontology which covers all cultures and all peculiarities in each of them (we will also call this ontology *culturally independent* or *neutral*). This unified conceptual ontology is built bottom up: once the culturally dependent ontologies are set, it is relatively straightforward to unite them and define the neutral concepts. The idea is to find the largest concept fragments that fit completely in any of the relevant cultural concepts from all cultures.

In the example shown in Fig. 3 below, these concept fragments are represented as a middle layer between both cultures and they build the multicultural ontology itself. This conceptualization not only achieves multicultural representation but, as a matter of fact, a culturally independent conceptualization.

A key feature of the conceptualization is to reduce the diversity of relations in a way that each neutral concept maps to only one concept in any of the cultures. For example, concept *C3* maps only to concept *secondary* in culture S.

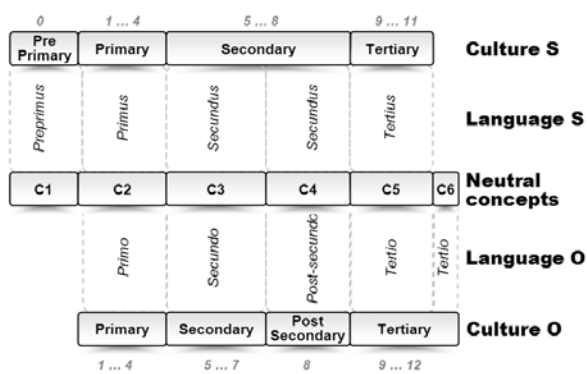


Fig. 3. Intermediate layer of culturally neutral conceptualization.

The mapping of two cultures can be used in a variety of situations, like *searching* for concepts in one culture using keywords from another culture, *translating* words in one language into another considering the possible cultural differences, *comparing* sets of concepts from two cultures to see whether they represent the same topic or not, etc. The rest of this section will only illustrate *searching* in a multicultural repository, because the other situations are handled in the same way.

If we search for a concept that does not exist in all cultures, the system will naturally exclude some of the cultures. For example, if we search for *pre-primus*, the system will translate the query from "*search for pre-primus*" into "*search for C1*". As a result, the documents found will only be those which are linked to *C1*, and these are documents only from culture S.

If there is a complete correspondence between both cultures, "*search for primus*" will find all documents about *primus* and *primo*.

If we search for a concept that is a subset of a related concept in another culture, we may get some additional (but still related) documents. For example, "*search for secundo*" will be translated into "*search for C3*" and the result will contain documents about *secundo* and *secundus*. Some *secundus* documents may refer to students from 8th grade, which are not in the secondary school level of culture O. However, they refer to the secondary school level of culture S and in that culture there is no distinction between 8th graders and 5-to-7th graders.

When a concept maps completely to two or more other concepts, then the search returns accurate results. A "*search for secundus*" will be translated into neutral concepts as "*search for C3 or C4*". The list of results will contain documents about *secundus*, *secundo* and *post-secundo*. Note that *post-secundo* will not be found by a multilingual system, because it is not a translation of *secundus*.

5.2 Implementing the model in TEO

As described above, one of the possible ways to support multiculturalism with ontologies is to build a separate ontology for each culture, and then link these together by identifying appropriate relationships linking related concepts. However, given the practical difficulties in implementing such an approach in the current phase of the Share.TEC project, an alternative could be to merge various ontologies into one.

With a concept-neutral ontology (the middle layer in Fig. 3), we can attach translations for each neutral concept drawn from all corresponding cultural concepts. Then we may remove the separate ontologies and keep only the neutral ontology as shown in Fig. 4.

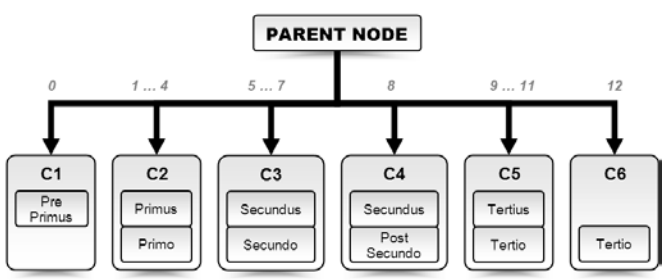


Fig. 4. TEO-like representation of a multicultural ontology.

Every conceptual node of the neutral ontology contains a list of the names of the concepts in all cultures that have this concept. This node structure is actually the same as the current TEO node structure.

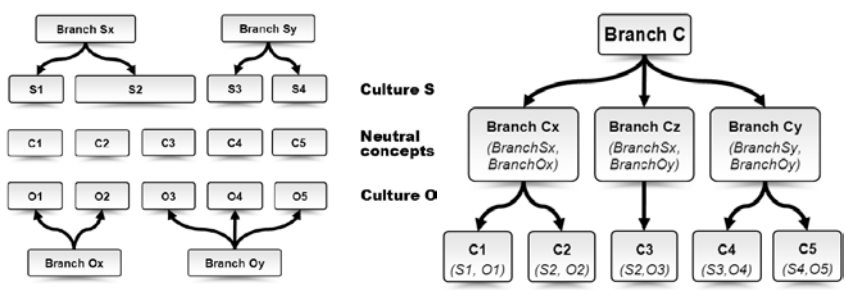
6 TEO's systematic multiculturalization

Given that the structures of the existing TEO and the neutral ontology are the same, it is possible to enhance TEO's multicultural features by performing two steps. The first step is to adjust the granularity of concepts. If there is a concept that is not atomic in all cultures, then this concept must be split. This step requires examination of all nodes in the ontology from representatives from all cultures.

The second step is to remove all translations of concepts that do not exist in a given culture. This insures that nodes are mapped correctly during multicultural searching. Removing unrelated translations is essential for multicultural support, and will not affect the user experience; users would still be able to see the description of the concept in any of the languages, because only concept names are liable for removal, while descriptions are kept intact.

A multicultural TEO will unite all culturally dependent TEOs. It will also allow the run-time "extraction" of individual cultural-dependent TEOs, which is not possible in the current TEO implementation. It is possible to treat local TEOs as projections of the global TEO onto the specific cultural domain.

One of the possible implications of this approach is the existence of discrepancies not only in the end nodes of the ontology, but also in nodes higher up in the hierarchy. It may happen that a whole sub-branch of a given taxonomy is missing from some culture. Fortunately, this is not a problem, because this branch should not contain translations of its nodes in this culture. Another implication is discrepancy in the relationships of the nodes in the ontologies. For example, post-secondary education may be branched under secondary education in one culture, and under higher education in another. Such a situation should not impose problems on the approach, because mapping between a culture and the neutral concepts may cross branch boundaries.



(a) (b)

Fig. 5. Cross-branch cultural discrepancy and conceptualization of higher levels.

One such example is shown in Fig. 5 (a), where two levels of both cultures are shown. If the user query is "*search for O2*", the system will translate it to "*search for C3*" and will find documents about S2 and O3. They belong to two incompatible branches – Sx and Oy. As this case illustrates, granulation

and conceptualization must be applied to all nodes of the ontology, eventually affecting all levels in all taxonomies.

Fig. 5 (b) shows concept-neutral representation of the case in (a). Branch Cz hosts the intersection of branches Sx and Oy, i.e. only concept C3. If a higher-level query is executed, e.g. "*search for all documents in branch Sx*", the system will treat this query as "*search for all documents in branch Cx or branch Cz*". Note that target branches are all branches that have *BranchSx* as a concept translation.

7 Conclusion

As designed and implemented, the existing Teacher Education Ontology in the Share.TEC system provides multilingual and some degree of multicultural support. This paper outlines a possible solution for enhancing the existing ontology into one that allows systemic multicultural support.

The core idea is to reorganize the set of concept-neutral nodes by granulating the existing nodes, and then shifting the names in different languages to the concept neutral nodes. Several cases of discrepancy caused by cultural differences have been discussed.

The described process is largely bottom up but should also rely on a top down process aimed at guaranteeing the uniformity and compatibility of the culturaldependent ontologies.

The proposed approach is due to be implemented in the existing prototype of the Share.TEC system (available at http://sharetec.it.fmi.uni-sofia.bg). This would require a minimal change in TEO , keeping its physical and logical representation intact. The same approach could also be recursively applied to higher nodes in the ontology hierarchies.

One of the two major benefits of this proposal is that it builds on the current structure of TEO. The other benefit is that once a concept has been identified as being granulated into its "atomic" components (i.e. those that share a common meaning in all cultures), then differences can be handled systematically at the linguistic level.

A first step in applying this approach would be to determine the scope and scale of the required granulation. Concepts that can be considered universal and intercultural do not need any granulation at all. However, if a concept belongs to just a subset of considered cultures, or is even unique to a given culture, a careful and collaborative process is needed to identify the related neutral component.

References

1. S. Alvino, P. Forcheri, M.G. Ierardi, & L. Sarti, A general and flexible model for the pedagogical description of learning objects, Proceedings of WCC2008, Milano, Italy, September (2008)
2. M. Grandbastien, F. Azouaou, C. Desmoulins, R. Faerber, D. Leclet, & C. Quénu-Joiron, Sharing an ontology in education: lessons learnt from the OURAL project, Proceedings of Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007), Niigata, Japan, July (2007)

3. D. Paneva, Ontology-based student modeling, Proceedings of the Fourth CHIRON Open Workshop Ubiquitous Learning Challenges: Design, Experiments and Context Aware Ubiquitous Learning, Turin, Italy, September, 17-25. (2006)
4. L. Razmerita, A. Angehrn, T. Nabeth, On the role of Actor models and Actor modelling in Knowledge Management Systems, in Proceedings of HCI International Conference, Greece (2003)
5. R. Andersson, A.K.Olsson, Fields of Education and Training, EUROSTAT (1999)
6. R. Mizoguchi, E. Sunagawa, K. Kozaki, & Y. Kitamura, A model of roles within an ontology development tool: HOZO. Journal of Applied Ontology 2(2) 159-179. (2007)
7. S. Alvino, S. Bocconi, P. Boytchev, J. Earp and L. Sarti, Capturing the Semantic Foundations of an Application Domain: an Ontology-Based Approach. In Proceedings of IV Workshop of the AI*IA Working Group on "Artificial Intelligence & E-Learning", Reggio Emilia (IT), 12 December 2009, ISBN- 978-88-903581-1-1, pp. 37-49. (2009)