AIED 2009 Workshops Proceedings
Volume 2

SWEL'09: Ontologies and Social Semantic Web for Intelligent Educational Systems

Workshop Co-Chairs:

Darina Dicheva
Winston-Salem State University, USA

Riichiro Mizoguchi
University of Osaka, Japan

Niels Pinkwart
University of Technology, Germany

http://compsci.wssu.edu/iis/swel/SWEL09/

Preface

An important aim of researchers and developers engaged in educational Web applications today is to provide means to unite, as much as possible, the efforts in creating information and knowledge components that are easily accessible, shareable, reusable and modifiable by others. Within the educational field, this motivates efforts to achieve semantically rich, well-structured, standardised and verified learning content. Ontologies and Semantic Web standards allow achieving such reusability, shareability and interoperability of educational Web resources. Conceptualizations, ontologies and the available Web standards such as XML, XTM, RDF(S), OWL, OWL-S, RuleML, LOM, SCORM and IMS-LD allow specification of components in a standard way.

The notion of Social Semantic Web describes an emerging design approach for building Semantic Web applications which employs Social Software approaches. Social Semantic Web systems are usually characterized through their emphasis on collaborative creation, usage and continuous refinement of Semantic Web structures by groups of humans. Social Semantic Web systems typically elicit domain knowledge through semi-formal ontologies, taxonomies or folksonomies.

Ontologies, Semantic Web and Social Semantic Web techniques offer new perspectives on intelligent educational systems by supporting more adequate and accurate representations of learners, their learning goals, learning material and contexts of its use, as well as more efficient access and navigation through learning resources. The SWEL'09@AIED'09 workshop focuses on the best practices of using these technologies for knowledge representation, adaptation and personalization in educational settings.

The workshop topics include:

- Building ontologies for e-learning:
  - ontology development
  - theoretical issues in ontology engineering
- Using ontologies and Semantic Web standards in e-learning applications:
  - to represent learning content (knowledge)
  - to organize learning repositories / digital libraries
  - to enable sharable learning objects and learner models
  - to support authoring of intelligent Web-based educational systems
  - to support adaptive modularised and standardized architectures
  - to exchange user model information between Semantic Web applications
  - to facilitate the reuse of content and tools in different contexts and cultures
- Using Semantic Web and Social Web techniques for adaptation and personalization of e-learning applications:
  - to support personalized information retrieval

  - to support adaptive information filtering
  - to support mobile learning applications personalization
  - to support exchange of user model information between semantic web applications - reuse of content and tools in different contexts and cultures
  - to support intelligent learning group formation
  - to support collaborative learning

- Educational dimensions of the Social Semantic Web:
  - collaborative tagging of learning resources
  - semi-formal ontologies, taxonomies and folksonomies in education
  - social perspective: motivations and benefits of Social Semantic Web approaches in education

- Real-world systems, case studies and empirical research for semantics-based Web educational systems:
  - lessons learnt
  - best practices
  - case studies for improved learners, instructors and authors experience
  - case studies of successful integrations of Web2.0 applications as services

- Semantic Web applications for learning and teaching support in Higher Education:
  - pedagogical application and use-cases (existing and envisaged) of semantic technologies in education
  - applications of semantic technologies to support learners and teachers

We will discuss lessons learned, benefits and further steps to be undertaken.

SWEL'09@AIED'09 is eleventh in the series, following the workshop on Concepts and Ontologies in Web-based Educational Systems at ICCE'2002, the three sessions of SWEL'04 (in conjunction with ITS'04, AH'04, and ISWC'04), the three sessions in 2005 (in conjunction with AIED'05, ICALT'05, and K-CAP'05), the session in 2006 at AH'06, the session in 2007 in conjunction with AIED'07 and the session at ITS 2008. SWEL has an established audience and a Web portal hosting related resources – the Ontologies for Education (O4E) Portal (http://o4e.iiscs.wssu.edu/xwiki/bin/view/Blog/).

The workshop features a keynote talk "Learner Models as Metadata to Support E-Learning" by Gord McCalla, University of Saskatchewan, Canada.

July, 2009
Darina Dicheva, Riichiro Mizoguchi
and Niels Pinkwart

Program Committee

Table of Contents

Learner Models as Metadata to Support E-Learning: The Ecological Approach

Invited Speaker: Gord MCCALLA

ARIES Laboratory
Department of Computer Science
University of Saskatchewan
Saskatoon, Saskatchewan
CANADA

In this talk I will present an e-learning framework called the ecological approach that is based on the idea that learner models are a rich source of metadata that can be used to inform various activities of an environment that supports human learning. Learner models can provide insight into learning material that might be appropriate for a learner, guidance about other learners who might be able to help a learner who is at an impasse, useful information to create and support learning communities, or even data that can be used for empirical analysis of the effectiveness of an e-learning environment. I will discuss the characteristics and implications of the ecological approach, for e-learning and beyond. I will also mention some of the research currently underway in the ARIES laboratory that may lead to more ecological e-learning systems.

An Ontology-Based Approach for Sharing Digital Resources in Teacher Education

Serena ALVINO[a], Stefania BOCCONI[a], Pavel BOYTCHEV[b,1], Jeffrey EARP[a], Luigi SARTI[a]
[a] *Istituto per le Tecnologie Didattiche – C.N.R., Genova, Italy*
[b] *NIS-SU, Sofia University, Sofia, Bulgaria*

Abstract. Teacher Education (TE) is a dynamic, lifelong process that needs to fully embrace innovation and assume a broader European perspective. The EC-funded Share.TEC project aims to provide enhanced, culturally-aware access to TE-related resources across Europe by means of a federated resource brokerage system whose semantic core is the proposed Teacher Education Ontology (TEO). This paper describes the rationale for an ontology-driven approach, gives an overview of TEO's multi-layered structure for addressing multicultural and multilinguistic issues, and presents some aspects of the TEO implementation that allow for language-independent conceptualization and multidimensional hierarchal searching and filtering Other TEO features are also discussed, including the support for dynamically generated user interface and system stability against ontology modifications.

Keywords. teacher education, ontology, ontology-based Information System, multicultural semantics

Introduction

Teacher Education (TE) is a lifelong learning process that is central to Lisbon Strategy efforts towards the building of a European knowledge society. However, the TE field has been generally slow to embrace innovation, much less to generate it, and has yet to assume a broad European perspective. Major hurdles stand in the way: TE practice is usually geared to meet the specific requirements of national systems that are linguistically and culturally bound; TE communities (even virtual ones) tend to focus on the immediate locus; hesitancy persists in embracing digital culture, with only patchy adoption of ICT and scarce sharing of digital resources.

Providing impetus for innovation within initial and in-service TE is the goal of the EC-supported Share.TEC[2] project. Share.TEC has undertaken to build an advanced user-focused system dedicated specifically to fostering a stronger digital culture in the TE field. This system is to aggregate metadata describing TE-related digital resources located Europe-wide; provide personalized, culturally-sensitive brokerage for the retrieval of relevant digital content; support the development of a Europe-wide perspective among those working in and with the TE community.

[1] Corresponding Author: Pavel Boytchev, KIT, Faculty of Mathematics and Informatics, 5 James Bouchier blvd, Sofia, 1164, Bulgaria; E-mail: boytchev@fmi.uni-sofia.bg.
[2] Share.TEC - SHAring Digital REsources in the Teaching Education Community, eContentplus programme (ECP 2007 EDU 427015); http://www.sharetecproject.eu/.

In pursuit of these objectives, an ontology-driven approach has been adopted. The semantic core at the heart of the proposed Share.TEC system is a Teacher Education Ontology (TEO), which is currently being developed by partners in the Share.TEC project in collaboration with international experts. The scope of this ontology has been set on concepts relevant to the domain of Teacher Education, with particular regard for aspects considered pertinent to the sharing of digital resources and practice among potential members of the Share.TEC community, namely teacher educators, teachers, academic/educational publishers and content developers.

The purpose of TEO within the Share.TEC system is to provide:

- pedagogical characterization of digital content;
- representation of user profiles and competencies as a backbone for cultivating TE communities;
- a basis for multilingual and multicultural functionality;
- support for personalized interaction with adaptive user applications;
- support for the implementation of recommending functions.

There are a number of reasons why an ontology-based approach has been adopted for the Share.TEC system and platform. The chief among these is to permit the sharing of concepts among people. Share.TEC's European perspective necessarily means that its users will bring to the community different languages and cultures. What's more, the TE field includes people with very different backgrounds, ideas and assumptions. In such a situation, effective communication and shared understanding can be difficult to achieve. Accordingly, TEO seeks to reduce conceptual and terminological confusion by identifying and properly defining a set of concepts (and their relations) relevant to TE in Europe. The result should be a non-ambiguous and consistent vocabulary for identifying those concepts, and a framework on which culturally and linguistically diverse versions of that vocabulary can be mapped.

As an integral part of the Share.TEC system, TEO will support adaptive user interfaces, and will inform services that use reasoning techniques. This should lead to the implementation of inferential search engines, advanced ranking solutions and flexible representation of user profiles. Importantly, TEO has also provided the basis for the definition of a common metadata model for describing TE-relevant digital resources.

1. A Domain Ontology of Teacher Education

1.1. TEO Definition and Development

Domain ontologies, upon which Information Systems are subsequently based, can be considered as repositories of knowledge that allow accumulation and systematization of knowledge. According to Allard et al. [1], "*Domain ontologies should also be conceived as use-neutral, in the sense that they are meant to serve as a foundation. Building on this foundation, different problems can be tackled, various applications derived, knowledge bases built. Consequently, domain ontologies should be relatively stable and aim to be a long-lasting conceptual structure".*

As previously mentioned, the present work on domain ontology for Teacher Education seeks to capture those concepts of the TE world that are relevant for sharing digital resources among practitioners. It is also to provide a framework for mapping multicultural and multilinguistic semantics.

Figure 1 describes the process of TEO development.

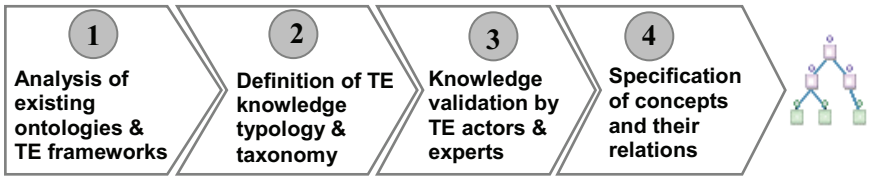


Figure 1. Phases in TEO's development process

As shown in **Figure 1**, TEO is grounded in existing research, especially that of Mizoguchi *et al*. [1], [2] and Guarino [3]. Specifically, it draws on three main models of reference: the OMNIBUS ontology[3], whose domain is education; the LORNET competency modelling ontology[4]; and the POEM learning object content model [4]. Other relevant sources that have influenced TEO's design include DOLCE [5], ONTOURAL [5], ALOCOM[6], PROTON[7] and user modeling ontologies [6], [7].

1.2. Basic Concepts of TEO

In its current version, TEO defines **162 classes** and **78 properties**. **Figure 2** presents the main concepts identified.

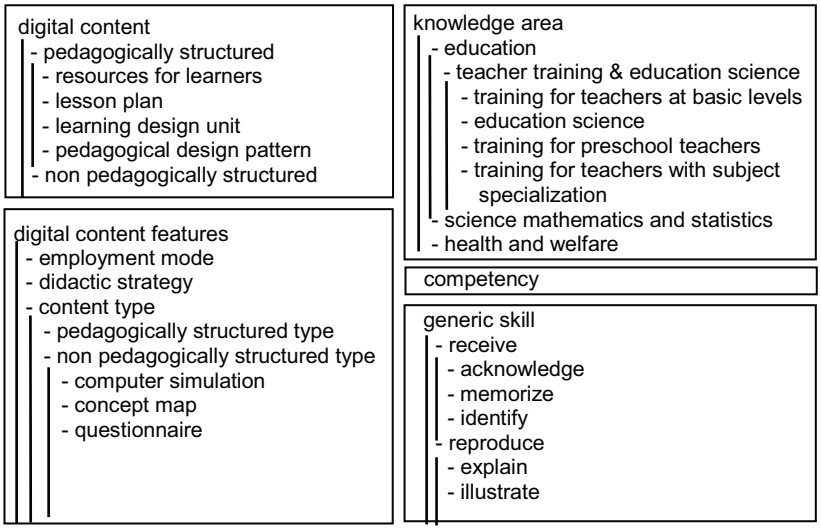


Figure 2. Overview of TEO's main concepts

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

Digital content refers to educational resources and artifacts closely related to the concept of the "learning object" [8]. Depending on its nature, a *digital content* artifact can be categorized as (i.e. "is-a") *pedagogically structured* or *non-pedagogically structured type*. *Resources for learners*, *lesson plans*, *learning design units* and *pedagogical design patterns* belong to the *pedagogically structured* category.

The *digital content* concept is also defined by other characteristics including *employment mode*, *didactic strategy* and *content type.* These act as *digital content features*, i.e. they are related to *digital content* instances via a *part of* relationship. A *computer simulation*, a *concept map* and a *questionnaire* are all examples of *non-pedagogically structured content type*.

Knowledge areas consist of topics drawn from the EUROSTAT [9] taxonomy of education and training. This classification was adopted as a reference due to its European perspective and pertinence to the TE domain[8]. The knowledge area hierarchy allows digital content to be described in terms of discipline and permits specification of the user's areas of interest. Association of a knowledge area and a *generic skill* generates a *competency* concept.

According to Paquette [10], competencies

> "[…] *link skills and attitudes to knowledge required from a group of persons and, more generally, from resources*".

In TEO, competencies are related with the *digital content* concept: this allows resources to be described and classified according to the specific competencies they address.

Following Paquette's approach to competency modeling [10], *generic skills* represent basic cognitive processes expressed as simple action verbs like *receive*, *reproduce*, *perceive*, *analyze*, *synthesize*. These concepts are arranged in a hierarchy: for instance, *acknowledge*, *memorize*, *identify* are lower levels of the *receive* action. At the top level, this hierarchy spans a continuum of "cognitive complexity".

Finally, we also considered the *role* concept, which draws on Mizoguchi's model [2].

Once basic concepts were defined in TEO, a knowledge validation process was carried out by TE actors & experts (see step 3 in **Figure 1**) to progressively improve TEO's conceptual framework and its technical implementation. Some important requirements emerged from development of Share.Tec's technical integration study and system architecture specification, so a new release of TEO has been produced to fulfill these needs.

2. TEO Implementation

TEO is designed to represent concepts and build relations between entities defining the domain of Teacher Education. This in itself is a positive step towards defining a consistent and complete picture of this domain, but TEO also contains important information that can be used by the Share.TEC software application.

The main contents of the Share.TEC repository are data, harvested from external repositories or provided by Share.TEC community members that are related to digital

[8] For interoperability purposes EUROSTAT has been mapped against Dewey's Decimal Classification: http://www.oclc.org/dewey/.

content. Along with these primary data, the main repository contains an online dynamic representation of TEO that supports core system services and features like:

- language-neutral concept-oriented data
- hierarchal searching and filtering
- dynamic multilingual user interface
- stability and system independence with respect to future changes in TEO.

2.1. Internal Logical Structure of TEO Representation

The internal logical structure of a TEO entity is designed with a minimalistic approach in mind – the simplest structure that facilitates all required functionality. Each TEO entity is represented as an individual node that is interconnected with other nodes through relations and that contains a list of translations of the concept represented.

The internal physical structure is more complex and is not discussed in this paper. It defines a wider spectrum of relations between ontology entities and contains additional data in order to allow (a) complete reconstruction of the ontology into a valid OWL file; and (b) support for extended functionalities like reasoning.

The structure of the "Medicine" node is shown in **Figure 3**. The node represents a language-neutral concept. It contains its verbal representation in three languages and is connected with other nodes through parent-child relations. This structure is sufficient to support conceptualization, hierarchies, multilingualism and ontology stability.

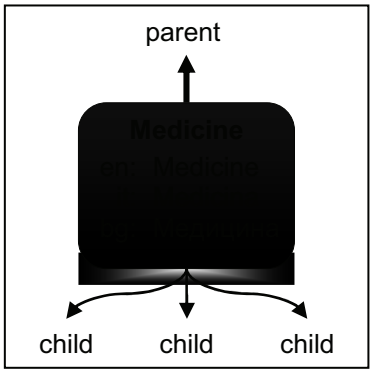


Figure 3. Internal structure of a TEO node

While this logical representation does not distinguish between classes, subclasses and instances, the distinction is made in the physical representation. TEO hierarchies are constructed by defining parent-child relationships between various nodes. This consolidates all TEO nodes into a single tree-like data structure that represents the domain knowledge of Teacher Education. Additional cross-branch relations are also represented, but these are not discussed here. **Figure 4** shows a vertical slice describing the complete path from the top concept, TEO, down to Cardiology. The path goes through Knowledge Area node, which is the root of the Knowledge Area hierarchy within TEO.

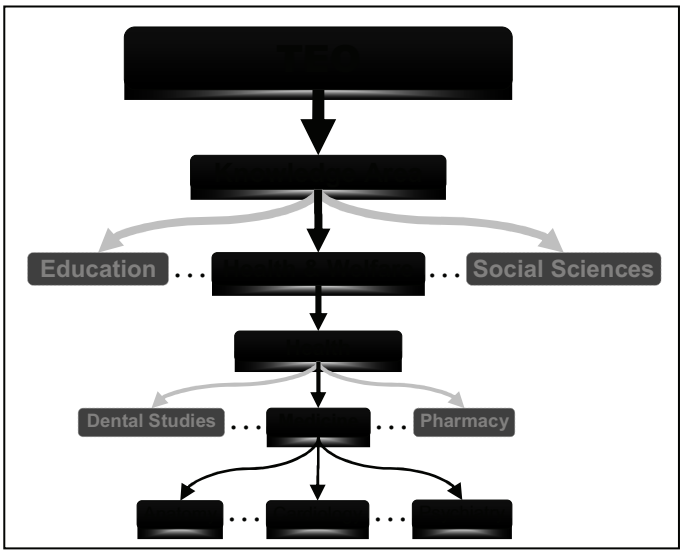


Figure 4. A vertical slice of TEO representing the full path down to Cardiology

2.2. Multilingual Support

Every node contains a set of translations of the node's concept into system-supported languages. This information is used when processing multilingual data from repositories across Europe. Whenever incoming data contains a concept expressed in a native language, Share.TEC scans TEO to find the corresponding node. When such data are processed, their texts are replaced by references to conceptual nodes. This makes the internal representation of data language-independent and links various translations of the same concept – **Figure 5**.

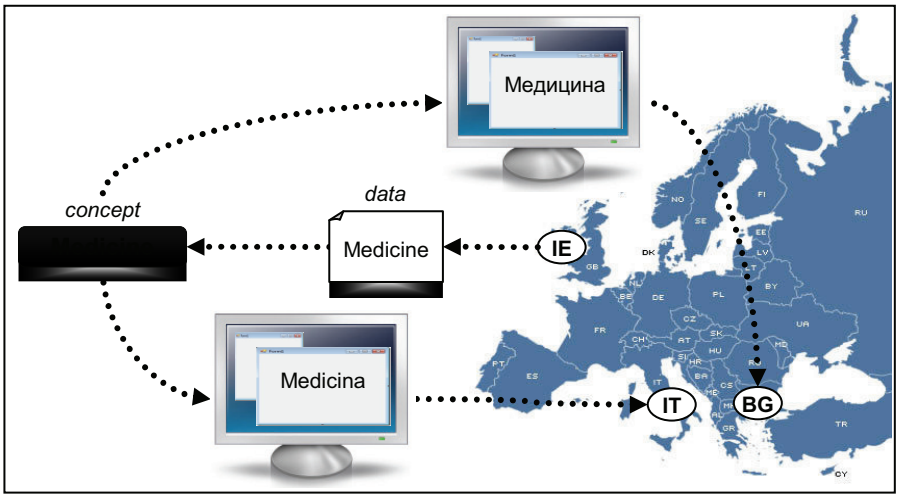


Figure 5. Native languages and language-neutral conceptualization

The same translations are used in the opposite direction, i.e. to translate concepts into users' native languages. The end-user system is multilingual and texts are translated on-the-fly while the screen forms are dynamically built. The original data harvested from, say, an Irish repository may contain the word "Medicine" but the same data viewed by Italian or Bulgarian users will be displayed as "Medicina" or "Медицина".

The current node structure can handle synonyms by providing more than one translation of the concept in a given language. Additionally, each node may have a description, which is available in all supported languages.

Translations are actually used not only for importing and displaying data, but in other activities, like searching and filtering. For example, a Bulgarian user may define a search criterion in Bulgarian (e.g. Knowledge Area="Медицина"): the system will match this to the node representing the Medicine concept and will find all data referring to that concept, irrespective of language.

2.3. Multidimensional Hierarchal Data Space

The hierarchal structure of TEO contributes to searching and filtering capabilities by expanding functionality. Instead of searching for a specific low-level concept like Cardiology, the user can set a higher level criterion such as Medicine. The result will be that all data referring to any of the 28 concrete concept instances under Medicine will match that criterion – from Anaesthesiology via Neurology and up to Psychiatry.

If the parent of Medicine (i.e. Health) is selected, then all medicines, dental studies, medical diagnostics, treatment technologies, nursing, caring, pharmacy, therapy and rehabilitations are selected.

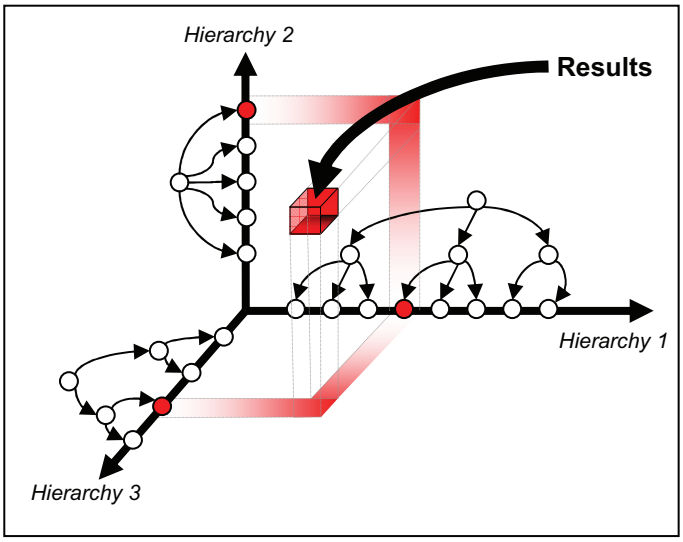


Figure 6. Multidimensional data searching

TEO hierarchies can be represented as axes in a multidimensional data space. Note that these axes are not mathematically continuous but are discrete sequences of concepts. **Figure 6** shows an example of a 3D data space defined by three hierarchies.

Solid circles represent selected entities. If all data are positioned in this 3D space, then applying filtering criteria will cut out a virtual box from all data.

All data elements inside this box are the user's search or filtering results. The multidimensional interpretation of Share.TEC data enriches the way users perceive TEO. They can "slide" along TEO axis by axis, slicing the data space in their preferred way.

Users are free to select any number of TEO axes and in any order (not just three as in **Figure 6**). Moreover, it is possible to select any non-leaf node from any hierarchy. This feature enables the user to limit the scope of the result not only to individual elements, but also to groups of related entities.

TEO has already captured and classified relationships between concepts. This information is used to define the boundaries of each group of related entities at any level of classification. Groups of related entities can be retrieved simply by accessing an upper-level node from the corresponding hierarchy – **Figure 7**.

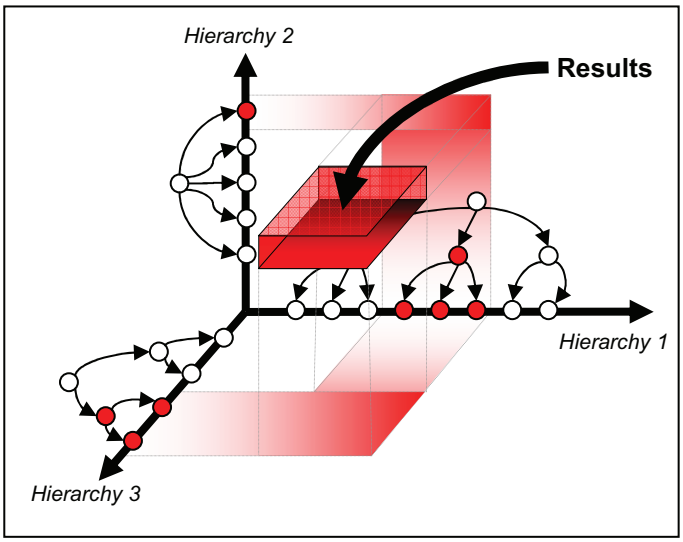


Figure 7. Hierarchal searching of data bound by a virtual searching box

The effect of this multidimensional and hierarchal approach is that the user can broaden or narrow the filtering criteria in a natural and domain-aware way. Axes corresponding to parameters which are not of interest to the user are totally ignored. For example if the user does not want to impose a filter on language, i.e. any language is acceptable, then instead of selecting the root node of all languages it is much easier just to ignore the language axis.

The multidimensional approach does not only utilize the taxonomies in TEO. A class together with its properties and the properties thereof (and so on) can also be treated as a hierarchy and thus be subject to multidimensional search and navigation.

2.4. System Sensitivity and TEO Modifications

An important element of the design of any data-driven system is to analyze its sensitivity with respect to modifications in the data. Our proposal is designed to be

non-sensitive or stable with respect to two major classes of modifications (stability meaning that if data are changed, then there is no need to modify the software).

The first class of modifications is adding a new language. Language-specific data is stored inside TEO nodes, so a new language will only add new properties and will not change the structure of TEO. The software application retrieves language information dynamically, so whenever a new language is added, it can be used right away. This means that while a new language is being added to the Share.TEC system, services can continue without interruption.

The second class of modifications affects the structure of TEO. If new instances or subclasses are added to any of the hierarchies, the system will use them right away, without any need for upgrading. However, if the modification goes beyond the boundaries of existing hierarchies, i.e. a new hierarchy is added, then the system will need to undergo major changes, because new data fields should be imported and/or analyzed.

3. Future Work

This paper has presented a domain-ontology for Teacher Education (TEO). Some of the major ideas underpinning the adopted approach were discussed, as were technical aspects regarding the implementation of TEO and related services.

Although minimalistic, the proposed implementation of TEO covers a wide range of features required for appropriate functioning of Share.TEC portal. It has been found that this implementation approach might be suitable for representing multicultural diversity in TEO. This is especially important where two or more cultures have their own specific views of TEO which are not compatible.

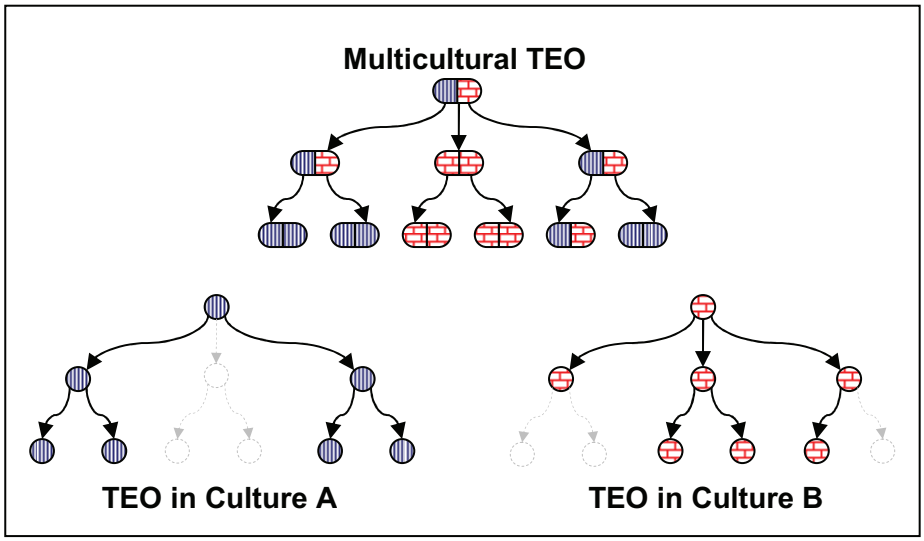


Figure 8. Multicultural coverage

Multicultural support goes beyond multilingual support, as it deals with those elements for which translation is problematic, i.e. it cannot be resolved with direct correspondence. The basic assumption is that when a node contains at least one

translation in a given language, then the corresponding concept exists in the culture based on that language. Thus the same data space will be seen in different ways by different users – **Figure 8**.

Although the projection of TEO onto a given culture may hide some of the concepts (especially those which do not exist in that culture), users are still able to view and work with the complete data space if they remove the language/culture filter. However, in this case they risk encountering unfamiliar and untranslatable concepts.

Acknowledgements

All the authors contributed on an equal basis to the conceptual design of this paper. ITD authors were mainly responsible for the introduction and first section, while Pavel Boytchev of NIS-SU was mainly responsible for the second and third sections.

References

[1] D. Allard, J. Bourdeau, R. Mizoguchi, Towards modeling knowledge of cultural differences and cross-linguistic influence in Computer-Assisted Language Learning (CALL). In: E. Blanchard & D. Allard, *Proceedings of the Culturally Aware Tutoring Systems (CATS) Workshop,* ITS Conference 2008, Montreal, Canada, June (2008), 11-22.
Last retrieved from http://www.iro.umontreal.ca/~blanchae/CATS2008/CATS2008.pdf April 10, 2009.

[2] R. Mizoguchi, E. Sunagawa, K. Kozaki, & Y. Kitamura, A model of roles within an ontology development tool: HOZO. *Journal of Applied Ontology* 2(2) (2007), 159-179.

[3] N. Guarino, C. Masolo, A. Oltramari, L. Schneider, Sweetening Ontologies with DOLCE. In Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW02), *Lecture Notes in Computer Science* 473, (2002).

[4] 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).

[5] 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). Last retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=04281129 on April 10, 2009.

[6] 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 (2006), 17-25.

[7] 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).

[8] D.A. Wiley, Connecting learning objects to instructional design theory: a definition, a metaphor, and a taxonomy. In: D.A. Wiley (ed.), *The Instructional Use of Learning Objects*, Association for Instructional Technology, (2000). Last retrieved from http://reusability.org/read/chapters/wiley.doc on April 10, 2009.

[9] R. Andersson, A.K.Olsson, Fields of Education and Training, *EUROSTAT* (1999).

[10] G. Paquette, An Ontology and a Software Framework for Competency Modeling and Management, *Educational Technology & Society* 10(3) (2007), 1-21.

Comped, a Web-based Competency Ontology Editor for Dynamic Geometry

Cyrille DESMOULINS[a,1] and Paul LIBBRECHT[b]
[a] *Laboratoire d'Informatique de Grenoble, France*
[b] *DFKI Saarbrücken, Germany*

Abstract. The Intergeo project addresses the issue of sharing interactive geometry across Europe upon a competency ontology enabling semantic annotation of resources. Whereas ontology engineer and platform administrator can manage this ontology using standard ontology editing tools, it is not the case for other roles, usually taken by average teacher and curriculum experts. This paper describes CompEd, a dedicated tool providing online means integrated to the Intergeo platform to access and edit this competency ontology. It presents its functionalities and how it is linked and synchronised with other Intergeo components.

Keywords. Competencies, Topics, Ontologies, Learning-resources, Authoring.

1. Introduction

The Intergeo project, funded by the European Community, aims at providing teachers with means to share dynamic geometry resources across Europe. It has developed a platform (http://i2geo.net/) where they can add a new resource and search for existing one by subject, level, instructional type, etc.

A core element of that platform is a competency ontology, named GeoSkills, that provides a shared semantic to resources. This ontology contains as of this writing about 600 classes and 2,500 instances representing competencies from the mathematics education standards of Spain, Germany, France and United Kingdom.

After a first phase where curriculum experts of the project have edited GeoSkills using Protégé with the help of ontology experts [6], a web-based tool was needed to enable the wide community of mathematic teachers and curriculum experts to edit this ontology across Europe.

This article is focused on this web-based competency ontology editor, called CompEd. It starts with a brief description of GeoSkills and the roles that need to edit it. CompEd features are then presented followed by the specification of its architecture and how it synchronises with other software components. We conclude with the implementation status and perspectives on users manipulation of GeoSkills in Comped.

2. Editing GeoSkills ontology with roles.

In this section, we present the GeoSkills ontology and its rationale, and then the roles that need to edit it.

[1] Corresponding Author.

2.1. The GeoSkills ontology and its rationale.

The GeoSkills ontology objective is to encode both a fine-grained mathematical semantics as well as names taken from various contexts (educational regions and languages). OWL-DL has been chosen to express GeoSkills for its well-defined semantics, its decidable knowledge representation and its interoperability over the WEB [3].

In order to provide to users names and descriptions of competencies they are used to, each elements of the OWL ontology (classes, instances and properties) can be described by names for each language. This is made using for instances, dedicated datatype properties, and for classes and properties, rdfs:label annotations.

GeoSkills essential ingredients are topics, competencies, pathways, levels and programs.

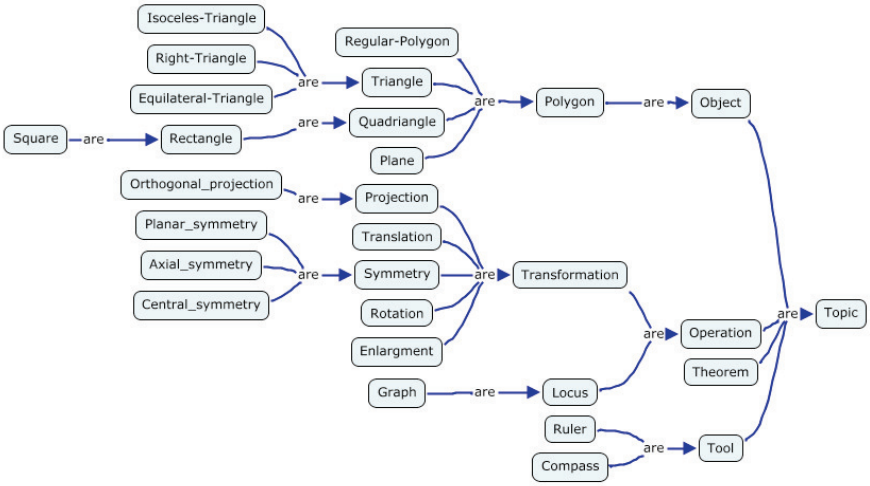


Figure 1. Extract of the topics hierarchy of GeoSkills

Topics: are made as a taxonomy (see figure 1), that is, a hierarchy of abstract classes each representing mathematical topics and objects. Multiple inheritance is possible thanks to OWL and is of great use in this case. Because OWL-DL properties only relate on instances, each class has a single representative individual. Properties on Topics are used to annotate resources with a topic or to relate a competency to topics

Examples of topics include isoceles triangle or ApproximationProcess_for_roots.

Competencies: are becoming the major entity of assessment and learning-plans. In GeoSkills, just as in [7] or [4], competencies are made of a verb and a set of topics. The class hierarchy of competencies represents the specialisation hierarchy of verbs. Examples of competencies include Calculate_trigonometric_ratio, Reproduce_an isosceles_triangle, or Identify_square_numbers. In the first case, calculate trigonometric ratio, the OWL individual is of the class Calculate and contains the topic trigonometric ratio.

Pathways: are a series of educational contexts such as elementary-school, or *Secondaire_de_Qualification_Technique_Artistique.*

Levels: are elements of a pathway, for example one of its year. For example *Gymnasium_Saarland_7te,* or Bachillerato_Ciencias_y_Tecnologia_2.

Programmes: a programme is the concrete plan of a level within a pathway; it is bound to curriculum standards. A programme is more a document than an individual, containing links to competencies where they are referenced.

The GeoSkills ontology is available under the Creative Commons Attribution ShareAlike and Apache Public Licenses; the current version can be downloaded from http://i2geo.net/ontologies/current/GeoSkills.owl.

With the Intergeo approach based on his ontology, searching through "Thales" competencies across European curricula provides also two types of competencies, obviously competencies having in one of their name or referring to topics whose name contain "Thales" but also to competency related to the "Intercept theorem" which is the English name of the French/Italian/Spanish "Thales Theorem". And inferred knowledge can be used, for example to match use binomial identities to solve equations[1] with queries using "equality", because *mathematically* an equality is a kind of identity.

2.2. Roles editing GeoSkills ontology

The Intergeo platform's main goal is to allow sharing of interactive geometry constructions and related materials. Overall, the usage of the platform is the execution of the following roles that access or edit the GeoSkills ontology:

- **The annotator** uses the editing front-end of the community platform in order to annotate resources as referencing the given competencies or topics, and a given educational-level, as well with many other information fields (such as authorship or license). Annotator needs a read-only access to the ontology, to check if the competencies chosen are the proper ones with the correct semantics.
- **The searcher** uses text-search, the ontology or curriculum-text browsing to identify the correct term so as to search through the platform's database to find relevant resources to use in teaching, to edit, or to evaluate. It also need a simple read-only access to GeoSkills, allowing her/him to browse through curricula, classes and instances of competencies and topics.
- **The curriculum encoder** identifies a curriculum-text of interest that could be shared among platform users, obtains an appropriate electronic version, browses through it and creates, in the ontology, the needed competencies and topics.
- **The competency translator** adds or edits competency or topic names or descriptions in one's own language. This does not require editing competency and topic classes or instance but only their denominations.
- **The ontology engineer**, together with the **platform administrator**, operates changes on the ontology for any facet, such as edition of the axioms or educational levels.

3. CompEd features

Whereas ontology engineer and platform administrator are able to us generic ontology editor such as Protégé, it is not the case of average curriculum encoders, competency translator furthermore annotators or searchers, roles usually played by average teacher or education experts but usually not semantic web experts.

1 Throughout this paper, we provide hyperlinks to the CompEd user-readable representation of the GeoSkills node when they are referenced.

We first tried to use of the Protégé client-server[1] which allowed team members to work synchronously on the ontology from remote places provided they are equipped with a very good network connection; only Universities met this challenge thus far. For other members, in particular companies involved in the Intergeo project, it was necessary to allow exclusive work on a local copy. Another limitation was met in the generic ontology-editor nature of Protégé, which makes it able to perform all sorts of changes, many of which should be reserved to ontology experts.

Then a platform to edit the ontology was needed, a web-based editing tool that allows every people from the dynamic geometry community to contribute and use aside of the Intergeo platform. As explained in [5], the GeoSkills ontology has been developed using an approach close to that described in MI2O. We propose a web-based tool that corresponds to the last phase of that methodology, the deployment, with iterations through the validation and refinement phases. This editing tool is called CompEd (Competency Editor). Its objective is to edit topic and competency individuals of GeoSkills as well as the topic and competency sub-classes and individuals. Editing includes altering names and relation properties (such as the generalisation/specialisation, instantiation relationships, or the involvement relationship of a topic in a competency).

3.1. Web based navigation

Even before the editing actions, a first important aspect is to allow web-based navigation of nodes of the ontology to allow the annotation of curriculum texts and textbooks: both of these features are to be done by having topics, competencies, and levels addressable through URLs which can also be presented in a browser. The annotations edited in the Intergeo platforms use these links as part of their presentation as in the figure 2 aside from this paragraph.

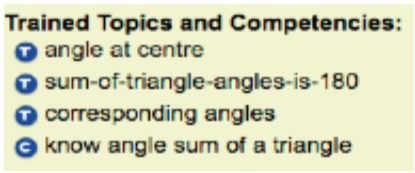


Figure 2. Rendered annotations of a resource.

3.2. CompEd Features

CompEd offers the browsing and editing of individual topics, competencies, and competency processes. Individuals can be reached by tracking recent activity; by browsing the alphabetic list view or hierarchical tree view; by navigating the relationships; by keyword searching; or by an external URL.

Items are displayed in a consistent way. As depicted by figure 3, which is an example for the "solve similarity problems" competency individual, the display is divided into three parts:

- The first part provides general information, which includes the name of the URI, the URI itself, the created and the modified dates. Below, the names in the user's language for the particular item are displayed. Names are grouped by type (common, uncommon, rare, false-friend). If wished, the user can click on the "more languages" link to get the other languages names.

1 The Protégé client-server setting is based on Java RMI and is documented at http://protegewiki.stanford.edu/index.php/Protege_Client-Server_Tutorial}

- The topic part just provides a list of topics that are connected to the competency item. The list items are links, which simplifies the navigation to the topic. Note there is no topic part in the view for topic items (only competencies are linked to topics).

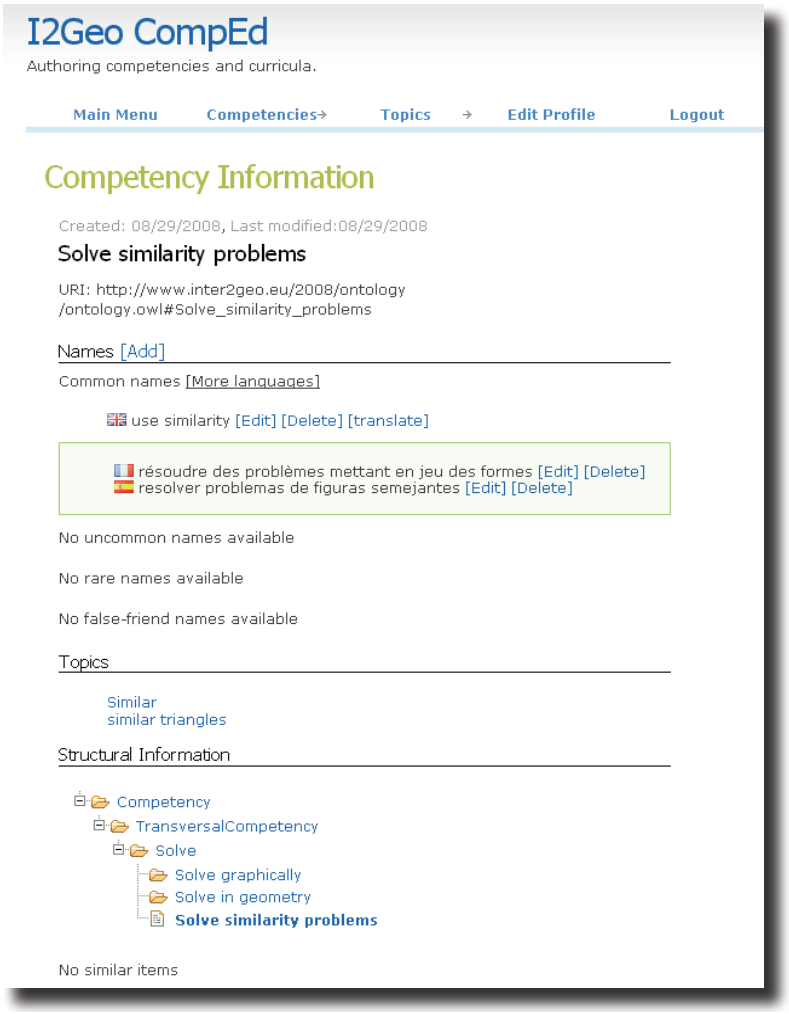


Figure 3. Presentation of a competency in CompEd (for curriculum encoder role)

- Finally, the structural part shows a hierarchical tree, which represents the generalisation/specialisation/instantiation path down to the competency item. In the case of competency classes (called *Competency processes* in the English GUI and *Catégories de compétences* in the French one), the tree will have all super-classes, subclasses, and individuals that are on the path through the competency process node.

Adding and editing of names as in the picture above includes the provision of a textual name, a language, and a type. The type can be one of: common, uncommon,

rare, or false friend. While the latter pieces of information have a default value to be displayed in Intergeo tools (common name and the native language of the user), the validation through OWL axioms guarantees that a name is provided.

Editing of competencies includes:

- changes, additions, and deletions of competencies,
- alterations on the competencies' URI,
- making connections to competency processes,
- referencing to topics.
- provision of a default common-name in any language

Editing of competency classes is very similar except that connections are established to other competency classes (which denotes a subclass relation) and to competencies instances (which denotes a membership relation). CompEd supports the user in altering data as much as possible, i.e., it suggests default values and signals errors in a user-friendly way.

The remaining input that is not covered in the CompEd usage is left for the ontology experts which includes adding or deleting extra properties, defining a class with a necessary and sufficient restriction, adding or deleting axioms about the ontology. Currently, edition of educational levels is also left to them, basically by using Protégé editor. They work informed by the curriculum encoding community based on a public forum where users of the curriculum knowledge, curriculum encoders, and ontology experts discuss.[1]

4. CompEd Architecture

The CompEd server software has been designed with high-usability in mind based on web-technologies that are widely spread. Thus the AppFuse framework[2] is at its core and its memory management is supported by the RDBMS persistence engine MySQL through the widespread java persistence framework Hibernate.[3]. These choices make CompEd a long-lasting responsive edition framework.

The decision not to use an OWL persistence engine is due to the apparently still lacking persistence framework for this technologies which scale long term and the ongoing need to load the complete ontology in RAM for most forms of reasoning.

5. CompEd synchronisation with other components

5.1. Editing tools

Two tools, CompEd and Protégé, can edit the GeoSkills ontology. Protégé 3.3.1 has been the first editing tool for creating a GeoSkills first version, used by two curriculum experts. It offers all the possible OWL expressivity. The normal tool to be used by curriculum experts is CompEd, but it offers an expressivity reduced to instances, hyponymy (is-a relation), links between competency and topics, and names. Because CompEd is unaware of axioms that have been expressed in OWL with Protégé, violations and new statements appear once the reasoning is invoked, nightly.

1 The curriculum encoders' online community is being built at http://curriculum.i2geo.net/

2 See http://www.appfuse.org/

3 See http://www.hibernate.org/

The Pellet classifier [1], on a dedicated server, makes these ontological consistency checks. This classifier provides also automatic classification of Competency classes between them, and of Competency instances into classes.

5.2. Accessing tools: SkillsTextBox

To allow users to identify the competencies, topics, or levels they mean, we extend the familiar auto-completion paradigm: users can type a few words in the search field, these are matched to the terms of the names of the tokens; the auto-completion pop-up

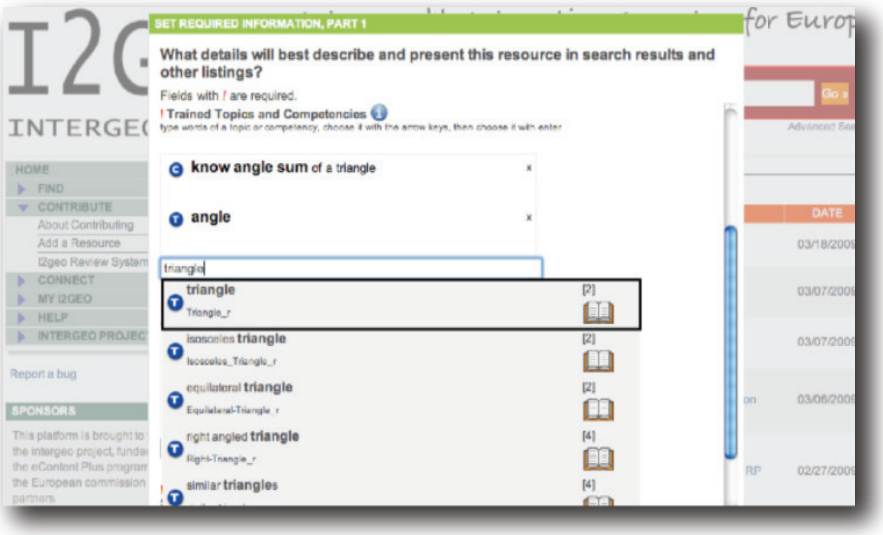


Figure 4: skilltextbox matching.

presents, as the user types, a list of matching tokens as seen on figure 4. This list presents, for each candidate, the default-common-name, the name found to match the user's input, the number of related resources, an icon of the type, and a link to browse about the ontology at the node and around it. When chosen using either a click, or a few presses of the down key followed by the return key, the choice action either triggers a search or the addition of the node in a list, or for annotations.

SkillsTextBox uses a simple HTML form equipped with a GWT script [2]. This script submits the fragments typed to the index on the server, which uses all the retrieval matching capabilities (stemming, fuzziness through edit distance or phonetic matching) to whose names start with the typed input, first in the languages supported by the user than in any language. The index returns the 20 best matching tokens and the script renders as an auto-completion list. More information about it is at http://www.activemath.org/projects/SkillsTextBox/.

5.3. CompEd, OWL, and the Term Index: Synchronisation

The competency-editor, the Protégé editor, the Skills-text-box' term index all are places which store a representation of the GeoSkills' ontology; in this section we explain how the OWL ontology file is at the centre of the synchronisation with incremental updates and regular resets.

The architecture of these pieces is depicted in figure 5: CompEd stores the contents of GeoSkills in a way made for massive collaborative edition; it cannot allow edition of all facets of the ontology; on the other side, Protégé allows full edition of the OWL ontology but is not suitable for such collaborative edition; the ontology server stores the ontology in RAM and performs the reasoning but it only receives the updates done by the CompEd users through update XML documents which are then incorporated into the OWL file. Finally, the term index contains an index of the names, ready for retrieval in the auto-completion and search functions.

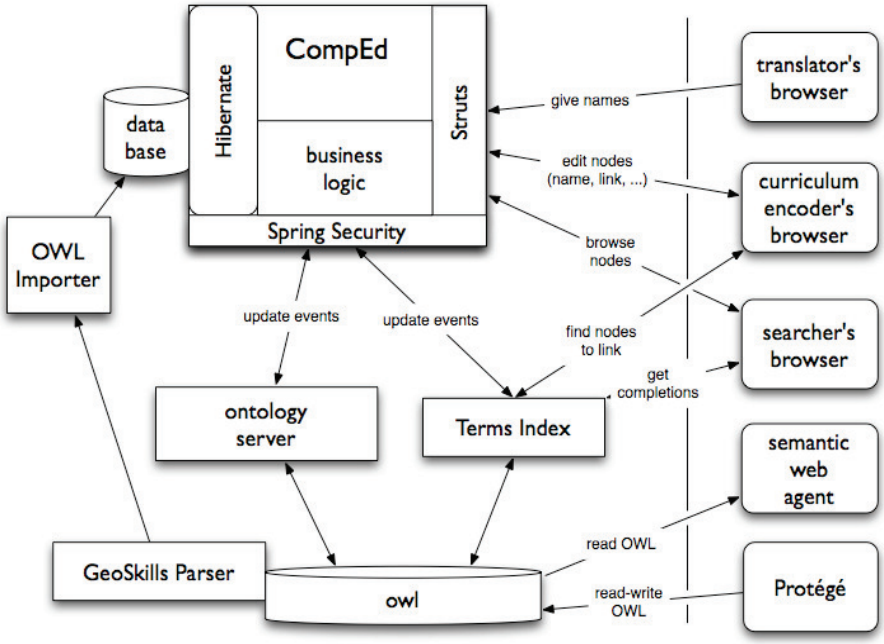


Figure 5. Architecture of the usages of the GeoSkills ontology

The communication flows between the pieces are as follows:

CompEd updates: following the actions of a curriculum-encoder or curriculum-translator, CompEd modifies his RDBMS storage and also sends an update document to the ontology server and to the term index. The latter update their representations following these updates.

Regular resets: because the intent of the competency editing process is the GeoSkills ontology, the ontology is used to replace the contents in RDBMS. This is done through a conversion from the OWL file, read through the reasoner, to the tabular format. These resets are applied every night and are the key to receive the reasoner results (such as the axioms that add properties or classes).

Ontology adaptation: from time to time, having concerted themselves, the ontology engineers will request to work at the ontology level, for example to add axioms, to add particular new classes or to perform clean-up operations. In order to do so, the CompEd server is taken read-only and the work on the OWL file in a text-editor or using Protégé can happen. It is followed by the regular reset, which re-imports the OWL file in the RDBMS.

Conclusion

The CompEd ontology editing tool has been developed to help standard users to collaboratively edit the GeoSkills ontology. It is linked with the other Intergeo platform tools and thus offers a standardised way of accessing and editing this ontology on the web. A synchronisation mechanism is the basis the enables that the ontology is consistently handled.

The CompEd ontology editing tool has been developed to help standard users to collaboratively edit the GeoSkills ontology. It is linked with the other Intergeo platform tools and thus offers a standardised way of accessing and editing this ontology on the web. A synchronisation mechanism is the basis the enables that the ontology is consistently handled.

Because of its collaborative aspect, CompEd seems to be one of the sole tools in the world to undertake the encoding of a multilingual and multicultural pool of educational competencies and topics.

The public deployment of CompEd and the opening of the curriculum-encoders group to users which have never seen Protégé has happened early 2009: the group now contains also encoders of the Czech, German, Dutch, Russian mathematics curriculum standards.

Perspectives

The commitment to encode the curriculum standards of mathematics of many European countries seems to be novel at least by its great diversity and start on the strong basis of a usable editing tool and internationalisation infrastructure. The perspective of such a large coverage may uncover new cross-lingual issues, which such an enterprise as Academic Benchmarks[1] seems not to have met yet.

The curriculum encoders' work includes the annotations of curriculum standards, or other texts for this purpose, by the additions of hyperlinks from sentences of the texts till the nodes of the ontology. Currently, encoders' are requested simply insert links to CompEd URLs. However this curriculum-linking task is only easy for HTML documents which are also the easiest documents to post-process to make actionable, allowing a reader to click on sentences to choose annotations. Most educational ministries, however, deliver the curriculum standards in PDF form, which has the advantage to be very close in appearance to the paper form, which a reader may well be used to. It is not yet clear whether this format will be acceptable for curriculum linking or to become actionable.

Among the avenues to be explored deeper is a more synthesised and complete exploitation of the conclusions of the reasoner. While inherited property values are easily handled by the parsing infrastructure which uses the reasoner, the automated classification results have been ignored thus far because it would make any parent class a direct subclass of the node: at least in the competency editing process, this is wrong as it would flatten the whole tree of inheritance (e.g. as in figure 3). We have to explore

[1] Academic Benchmarks Inc. is an american corporation providing services of matching curriculum standards to content resources.
See http://www.academicbenchmarks.com/.

such avenues as taking parent-classes inferred by the reasoners and removing the asserted ancestor parent-classes.

An issue we encountered together with the curriculum encoders of the Intergeo project is the readability of URIs. On the one hand, this characteristic is good, creating, for example, URLs and annotations that are more readable hence easier to manage. On the other hand a readable URI carries a textual semantic and several times we encountered the wish to adjust that URI to resembles better the semantic of the node. Changing a URI, however, would need a richer infrastructure e.g. resulting in redirects or the adjustment of all the links. One of the safest approaches could be to have unreadable URIs, as safe randoms. More experience is required to decide on best practice.

Beyond parsing, there should also be the possibility of the ontology server to feedback on changes done in the curriculum editing process, including indicate inconsistencies that have appeared. The XML encoding of the updates could be of use for this purposes displaying errors.

Acknowledgements

This work has been realised within Intergeo eContentPlus project, which was partially funded by the European Community and by participating institutions.

We wish to thank in particular Martin Homik who has been the main implementer of CompEd helped by Arndt Faulhaber (DFKI). We also thank the members of the Intergeo project's Work-package 2: Colette Laborde (CabriLog), Maxim Hendriks (TU/e), and Albert Creus-Mir (Maths4More).

References

[1] Clark & Parsia, *Pellet*, 2009, http://clarkparsia.com/pellet.

[2] Google Inc., *Google Web Toolkit (GWT), a java to javascript compiler and toolkit*, 2009, http://code.google.com/webtoolkit/.

[3] McGuinness, D. L., van Harmelen, F., *OWL Web Ontology Language Overview*, 2004, W3C.

[4] Melis, E., Faulhaber, A., Eichelmann, A., Narciss, S., "Interoperable Competencies Characterizing Learning Objects in Mathematics", in *Intelligent Tutoring Systems*, Springer, 2008, pp. 416-425, ISBN 978-3-540-69130-3.

[5] Laborde, C., Dietrich, M., Creus-Mir, A., Egido, S.,Homik, M., Libbrecht, P., *Deliverable D2.5: Curricula Categorisation into Ontology*, Deliverable, 2008, *http://svn.activemath.org/intergeo/Deliverables/WP2/D2.5-Curricula-Categorisation/D2.5-Curricula-Categorisation.pdf*

[6] National Library of Medecine, “Protégé Editor version 3.3.1”, 2008, http://protege.stanford.edu/.

[7] van Asche, F., "Linking Learning Resources to Curricula by using Competencies", in *First International Workshop on Learning Object Discovery & Exchange*, 2007.

Sharing Open-Content Learning Resources in Emerging Disciplines

Darina DICHEVA[1], Christo DICHEV and Yi ZHU

Winston-Salem State University
Winston-Salem, NC 27110, USA

Abstract. This paper proposes a new type of collaborative learning repositories based on a hybrid organizational structure that leverages the potential of domain specific collaborative tagging in combination with conventional digital libraries classifications. It is exemplified in LinkedCourse - a learning repository prototype for rapid collaborative development, sharing and reuse of resources for emerging disciplines. The focus of the paper is on the collaborative semantic annotation and community formation, setting the socio-ontological framework of LinkedCourse. The implemented approach is aimed at keeping the flexibility of collaborative tagging and annotations within a domain-specific ontological framework.

Keywords. Participatory learning repositories, open content, collaborative semantic annotation, community formation

Introduction

Learning repositories were introduced as enablers for storing, sharing, reuse and repurposing of learning resources. However, there has been little sharing and reuse of educational materials through public repositories. One important aspect in the repository design that has been neglected is that the resources should be reusable and modifiable, keeping track of all contributors. Another factor limiting the widespread use of learning repositories is that they don't address adequately the specific needs of individual communities, which are typically formed around a shared domain of interest. Learning repositories are more likely to be successful if they are developed to meet genuine needs of a community. For example, in emerging disciplines the domain is evolving. The classification of the learning content (a form of a light-weight ontology), being domain dependent is also evolving. So is the domain vocabulary. This means that shared conceptualizations within not well bounded and evolving domains demonstrate a "work in progress" tendency. This suggests a hybrid classification framework that combines participatory processes and traditional classification approaches.

In parallel, we are witnessing a growing popularity of online communities that rely on mass participation and constant update strategies, such as social bookmarking. Many applications support building communities by empowering users to directly participate in a transparent collaborative process of content development. Typically

[1] Corresponding Author: Darina Dicheva, Winston-Salem State University, 601 S. Martin Luther King Jr. Drive, Winston Salem, NC 27110; E-mail: dichevad@wssu.edu.

they provide users with the ability to tag resources and to publish or share their tagging with other users. What makes the social tagging systems more attractive in repository context compared to conventional indexing approaches is that they support social interactions allowing users to connect to other users and to their resources and tags.

We believe that in a community of practice (a group of people connected by shared interests) tagging still has unexploited potential. A supporting hypothesis is that community-produced tags will be of higher semantic quality since they will reflect a vocabulary of community specific terms. This fact suggests creation of domain specific tagging mechanism able to leverage the implicit semantics emerging from the evolving tag structure and vocabulary. For example, community generated tags can be used as a source of terms to augment the evolving domain taxonomy, as they tend to represent the most current and natural domain terminology. The beauty of folksonomy is that the users do not have to learn any formal mechanisms; instead, they can tag content using freely chosen words. We believe that this basic freedom should not be sacrificed. Our approach is not to limit the tagging, but rather to set taggers in an appropriate context, namely, within a community formed around a particular course or topic.

Other challenges addressed in this work include stimulating participation and tracking content ownership. There are many incentives for publishing research publications in the academic community: academic reputation, promotion, institutional policy, etc. However, there are few such incentives to publish teaching materials unless in the form of an officially published textbook. This is one of the reasons why many lecturers tend to restrict the access to their materials to themselves or their close colleagues. The acceptance of learning content sharing on a community level requires adequate incentives implemented at a repository level. The learning repository discussed in this paper serves as a platform for content generation and reuse possibly by re-purposing and adapting of the original material. This presumes that a content unit can have many contributors, which leads to questions about ownership. We address this issue by using a Creative Commons License and keeping track of unit contributors.

The above considerations imply, that the new generation of learning repositories should depend on people's participation not only in content evolution but also in repository structure evolution and should address factors such as community formation and crediting authorship. To address existing needs related to emerging disciplines, we propose a community driven framework for rapid collaborative development, sharing and reuse of learning resources exemplified in *LinkedCourse* – a web application which we are currently developing.

The paper is organized as follows. We discuss the main ideas underpinning the proposed infrastructure in Section 1, introduce the infrastructure focusing on the socio-ontological aspects in Section 2, and present *LinkedCourse* in Section 3.

1. Towards Community-oriented Sharing of Learning Resources

Web 2.0 systems provide a medium for sharing and exchange of resources such as bookmarks, photos, videos, files, etc. Currently available "folksonomies" are geared toward casual social networking and prove successful for sharing and collaboration. Apparently, Web 2.0 offers a fresh approach that can be also used for sharing educational materials in emerging disciplines. It can foster the development of diverse communities of authors who are willing to share their material. Inspired by Web 2.0 and successful social bookmarking practices, we propose a novel community-oriented

framework for rapid knowledge exchange. We aim at infrastructure that supports participatory learning repositories, instead of the push models that traditional repositories provide, as users' participation can create content and keep it vibrant.

The key difference of our approach compared to the conventional learning repositories is that (1) it is focused on the connectivity of resources and users, and (2) it enables lawful modification of repository resources. Differently from social bookmarking systems that typically allow sharing links to someone else's resources, the goal here is (1) to support sharing of resources created by the participating members while addressing the corresponding intellectual property concerns, (2) users are expected to share not arbitrary bookmarks but links to learning content in a particular subject area, (3) the tagging process is based on a mix of controlled, semi-controlled and uncontrolled vocabularies (however taggers are still not limited in their choices).

Such an environment would be of interest to resource users who need learning material, possibly willing to endorse or critique the content and give credit to content providers, to resource providers who seek self-expression, community endorsement, and critique for the provided work and possibly attention from publishers and open-book supporters, to users who are interested in lawful reuse/modification of registered resources and/or seek collaboration for resource development, and to publishers looking for promising textbook authors based on demonstrated interest from the public.

1.1. Lawful modification of online learning materials

In emerging disciplines, courses are not well established and some can share multiple commonalities. Relatively minor modifications of some materials could effectively exploit the commonality of the bulk of shared content. The issue here is that there is no lawful way for instructors to modify learning materials found on the web even if they are willing to give proper credit to the authors. Thus, a mechanism for declaring that certain learning material is open and freely available for modification, extension, and reuse—as long as the authors are properly credited—is urgently needed.

1.2. Collaborative authoring

The availability of infrastructure, supporting the reuse of open licensed learning material in a subject domain, could make course content creation a much simpler task. Rather than writing a complete set of course materials, instructors can work on single topics, which are not covered and in which they feel experts. Authors can share their lecture notes, slides, assignments, problem sets, syllabi, reading lists, etc. They can form ad hoc working groups to collaboratively develop and adapt existing units. Through reuse of shareable units, a complete set of material for a specific course can evolve without waiting for the 'definitive' textbook to be published. This is especially important in emerging disciples, where there are additional barriers for textbook writing: the initial market is relatively small and typically fragmented, and the lifetime of publications is often short (due to the rapid evolution of technologies).

1.3. Content contributors vs. content consumers

In a resource-sharing community, some members act more as contributors and others more as consumers. Various studies report that in participatory media (including wikis, photo-sharing sites, etc.), 5-10% of the users contribute half to all of the content [1].

Resource users however also play an important role through community filtering that ensures the promotion of good quality content. The contributor-consumer interaction offers richer opportunities though. If someone finds an open-content learning resource that is not an exact match of what they need, the potential places to look for a better match is in its "consumer" resources or in resources for which it plays a role of a "consumer." Indeed, the content of a given resource might not match a particular instructional goal, however, someone may have adapted it appropriately. Therefore, we exploit a *resource connectivity*-based strategy for exploration.

2. *LinkedCourse* Framework

The proposed framework aims to support rapid, community-based development and sharing of learning resources while acknowledging and preserving the copyright of the authors. The keystones of the proposed framework are presented below.

2.1. Distributed content and intellectual property

The learning material registered in the repository is distributed and resides on authors' websites. The repository contains only records with metadata for the original resources and their authors.

The registered resources are licensed under the Creative Commons license [2] that allows the content to be copied and redistributed, with or without modifying, and used for commercial or noncommercial purposes, provided the authors receive attribution throughout the use of the module, even when modified. This promotes the greatest possible sharing of materials.

2.2. Collaborative semantic annotation

The advantages and disadvantages of ontologies and folksonomies are well known [3, 4]. Ontologies can make content well organized, but they require time and expertise. Studies have shown that there is an ongoing reluctance among both users and institutions to create ontologies and metadata [5]. On the other hand, user-generated folksonomies can be more relevant and inspire discovery, but users lack discipline and expertise. Controlled tagging brings discipline but can create a gap between resource providers and users of learning collections, making the retrieval process tricky [6]. Yet, the existing approaches of combining ontologies and folksonomies have not demonstrated convincing results (see for example [4, 7, 8]).

Our approach is not based on simple coexistence of folksonomies and taxonomies as two different and complementary approaches for semantic annotation; instead, the idea is to mix them in an approach that lies somewhere in the middle. The suggested approach for sharing learning content is an attempt to combine some aspects from both worlds: conventional digital libraries and ad hoc classification. It is based on two observations:

- Based on their experience with personal folders, instructors are used to classify their material under courses, and subdivide it by course topics.
- Tags are inseparable from the context of the community in which they are

created and used [9].

Based on the first observation, the learning resources in our repository are divided into course collections. The course structure is employed not only as a predetermined classification framework but also to narrow down the user base to a particular community and thus to limit their tagging vocabulary by limiting the domain vocabulary as a source of tag choices.

We view tagging in three dimensions:

- as a process (from the viewpoint of the user's choice of terminology),
- as folksonomy (from the viewpoint of the generated collective vocabularies and resulting knowledge organization), and
- as a social activity (from the viewpoint of the social context of interactions and the resulting formation of communities).

The learning material resides in course collections. This is the place where storing, tagging and searching resources takes place. Thus courses are used as both an upper organizational infrastructure of learning resources and social infrastructure for user interactions and forming course level communities. Tagging in such communities will generate a conceptual structure as perceived by the corresponding community.

The *LinkedCourse* platform is aimed at aggregating community generated tags within semi-controlled vocabularies, metadata and domain-specific ontologies. The challenge here is in striking a balance between the open user-generated tags and the semi-controlled vocabulary. Our strategy is to constrain not the tag choices but instead the user base through limiting the domain, which serves as a common point of interest.

We are also extending the folksonomy model with upper ontologies, including WordNet, Dublin Core, and FOAF, in order to ensure some reusability and interoperability of information. More specifically, we use Dublin Core for annotating resources and FOAF for presenting authors' profiles. Entering complete Dublin Core data is optional. A minimal DC subset - the title, the author and the descriptions are derived from the resource submission process. FOAF information is obtained in two ways: retrieved by the FOAF RDF file of an user if they submit their FOAF URL, or generated from personal data entered by the user.

We envisage the use of three semi-controlled vocabularies. The first one comes from course names. Though course names put some boundaries on the tags variations, the choice of the descriptive terms is left open. The purpose here is to capture the domain specific terminology clustered around related concepts. For example, similar learning content can be found in courses named "Internet Systems", "Internet Technology", "Web Programming", "Web Design", "Scripting Languages", etc.

Another semi-controlled vocabulary comes from the resource types, e.g., lecture notes, code examples, assignments, free software, test samples, problem sets and solutions, syllabi, reading lists, etc. A third source of "controlled classification" comes from the automatic tagging of resources with contributors' information, e.g., username, institution, home page, etc. We plan to use *LinkedCourses* as an experimental environment for examining our hypothesis that collaborative tagging converges to controlled vocabularies that can be used as sources of terms for augmenting the evolving taxonomies of emerging domains.

The uncontrolled part of the tagging leaves users the freedom to pick arbitrary

categories for classifying learning resources besides the course and resource type classification. Such a feature will enable users to group resources by additional properties, including content-related, instructional, presentational, etc. Instead of having users haphazardly entering in tags to describe the resources they bookmark, *LinkedCourse* suggests tags used by the members of the corresponding community. In addition, using a tag cloud as a categorization system allows visualizing the power of the ad hoc classification. The tag cloud will allow users to navigate the collection by all properties used for grouping the resources and to discover interrelationships between groups that may not be apparent when navigating through courses.

2.3. Community building

The core idea driving *LinkedCourse* architecture was to build communities of instructors through collaboration and social tagging. One of our goals is to explore the feasibility and the potential of supporting the creation of sustainable communities of practice. In this aspect we aim at creating repositories that provide platform for discovering not only resources but also people. In contrast to traditional repositories, we provide a richer view on resources enabling users to see how they are used and who interacts with them.

The framework enables users to bookmark and vote on the quality of registered resources, to subscribe to receive information (through RSS and Atom presentations) when a new resource in a particular course, of a particular type, from a particular author, or tagged with a particular tag, is registered or updated in the system, etc. Several strategies are used to create incentives for reuse, including measuring and rewarding the contribution and use of content, combined with technical support that facilitates and encourages reuse. Besides, the framework supports the connection of users to encourage networking and help with further collaboration (for example, based on individuals' bookmarks). It also enables users' involvement in maintaining the website, implements a reward policy to encourage members' 'housekeeping' work and uses it to rank the involvement of community members, etc.

3. *LinkedCourse* Implementation

LinkdCourse design requirements include support for registering and tagging of learning resources, maintaining references between resources, users' reviewing and ranking resources, 'housekeeping' for maintaining good structure and content, intuitive navigation and searching throughout the collection of resources for finding courses or resources of interest, or other users with similar interests, provision of personalized resource spaces, and community building and communication.

3.1. Services

To enable this functionality we are implementing a service-oriented architecture. The main envisioned services are listed below.

- Registering resources: for each registered resource only a *resource entry* is maintained containing information, such as name, type, description, URL, etc.

- Tagging, reviewing, commenting, or voting for resources.
- Exploring resources: a combination of browsing paradigms are envisaged to support the exploration of resources: *facet-based browsing*, providing a five-dimensional view on the content (based on facets); *pivot browsing* [10], providing a lightweight mechanism to navigate an aggregated collection, *attribution & credit reference map*, and *tag clouds*.
- Community building and communication: services for support communication and collaboration between registered users, such as forums for discussing courses, resources, and tags and RSS and Atom presentations related to courses, resources, tags and authors; services to help involving new members and contacting existing members; statistics on the blogs' use, services for importing and exporting bookmarks, services for maintaining tags (e.g. reporting overlapping tags (having similar names), non-used tags, resources/tags with little metadata, or such that members voted as low quality/not useful).

3.2. Interface

Currently, the *LinkedCourse* interface contains two main spaces: global and private space (see Fig. 1). The private space contains the following tabs: *My Courses, My Resources, My Bookmarks, My Community*, and *My News.* The global space is the space where users can browse all information submitted to *LinkedCource*: courses, resources, people, and tags. This is the space to which unregistered users have access. The private space is envisaged as a projection of the global space on a particular user. Therefore, it contains all courses, resources, and tags created or bookmarked by that user.

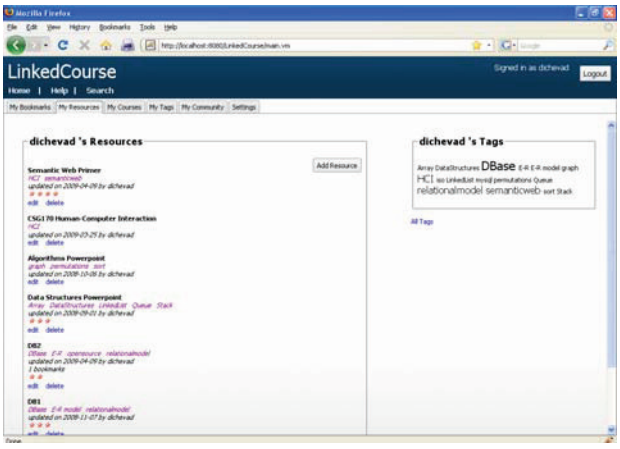


Figure 1. Screenshot of the *LinkedCourse* interface: a user's private space.

My Community connects a user to other *LinkedCource* users. Each user can add registered users to his/her community for easy access to their learning resources/bookmarks/tags, as well as for more convenient contact to them through

special services.

My News is the space where a user can receive information for newly added or modified resources of their interest (through RSS feeds) or such exported to them by members of *their community*.

Users can subscribe to courses, tags, and people in order to receive information about new resources submitted to a specific course, by a specific author, or tagged with a specific tag.

4. Conclusion

In this paper we propose a framework for rapid collaborative development and sharing of learning resources for emerging disciplines, which is built on a set of intuitions shared by a wide range of academics: that knowledge should be open to use and reuse; that collaboration should be easier; that people should get credit and kudos for contributing to education research; that there should be a way for instructors to publicly acknowledge reuse of open content; and that the ability of authors and instructors to readily and dynamically access and update learning material is especially important in rapidly changing fields.

A professional community will succeed if the participating members perceive some value in their participation. In this case, the value is in the content that no single instructor is normally able to develop on their own. A pool of up-to-date teaching materials made available to community members through sharing and collaboration provides value and motivation for sustainability. Providing an audience and means for expressing the self is another value factor for contributors seeking reassurance. We believe that an appropriate infrastructure can turn a learning repository into a space where content attracts people and people bring others who use and further evolve it.

References

[1] What makes people to participate? Online marketing strategies and techniques from Asia Pacific, (2007), available at: http://www.rockyfp.com/blog/what-motivates-people-to-participate/.

[2] Creative Commons http://creativecommons.org/

[3] H. Al-Khalifa, H. Davis: FolksAnnotation: A Semantic Metadata Tool for Annotating Learning Resources Using Folksonomies and Domain Ontologies. In: Proceedings of the Second International IEEE Conference on Innovations in Information Technology, Dubai, UAE (2006).

[4] S. Bateman, C. Brooks, G. McCalla: Collaborative tagging approaches for ontological metadata in adaptive e-learning systems, In Proceedings of 4th Int'l Workshop on Semantic Web for E-Learning - SWEL'06 (2006).

[5] J. Casey, J. Proven, D. Dripps: Geronimo's Cadillac: Lessons for Learning Object Repositories. 7 (2005).

[6] Report on the Workshop of Learning Object Repositories as Digital Libraries, D-Lib Magazine, October 2006, available at: http://www.dlib.org/dlib/october06/artacho/10artacho.html

[7] R. Lachica, D. Karabeg: Metadata creation in socio-semantic tagging systems: Towards holistic knowledge creation and interchange. In L.M. Garshol and L. Maicher (Ed.): Scaling Topic Maps, Topic Maps Research and Applications (TMRA 2007), LNCS, Springer Verlag (2007).

[8] L. Specia, E. Motta.: Integrating Folksonomies with the Semantic Web, Proceedings of the 4th European Conference on the Semantic Web (ESWC 2007), Innsbruck, Austria, June 3-7, 2007, LNSC Springer 4519 (2007), 624-639.

[9] P. Mika: Ontologies are us: A unified model of social networks and semantics, J. Web Semantics 5 (1) (2007), 5–15.

[10] D. Millen, J. Feinberg, and B. Kerr: Dogear: Social Bookmaking in the Enterprise, In Proceedings of CHI 2006, Montreal, Canada, April 2006 (2006).

Toward a Learning/Instruction Process Model for Facilitating the Instructional Design Cycle

Yusuke HAYASHI[1], Seiji ISOTANI[1], Jacqueline BOURDEAU[2] and Riichiro MIZOGUCHI[1]
[1] ISIR, Osaka University, Japan, {hayashi, isotani, miz}@ei.sanken.osaka-u.ac.jp
[2] LICEF Research Center, TÉLUQ-UQAM, Canada, jacqueline.bourdeau@licef.ca

Abstract. Many instructional design models have been proposed and their benefits are evident. However, there is lack of a common and formal notation to describe the product of the design. This causes difficulty in evaluating the product (the course) in the development. To eliminate the difficulty, we need a formal framework which has enough semantics for keeping the consistency of the product. Thus, this work aims at proposing a unified modeling framework for learning and instruction based on ontologies that has the potential to support some phases of instructional design. Furthermore, we give an example of how one-to-one instruction and collaborative learning are modeled on the proposed framework.

Keywords: ontological engineering, instructional design, collaborative learning

Introduction

A considerable number of instructional design (ID) models have been proposed. The main contribution of them is to provide systematic and reflective *processes* for developing learning/instructional courses. All of these process models share most of the same basic components: analysis, design, development, implementation, and evaluation [18]. Each component has a discipline for an assessment of the course in bringing about learning and a mechanism to improve the course if learning fails to occur as expected. Therefore the final product of ID (learning/instructional process description as a course) can be modified until it reaches the desired quality level [5].
In order to go through the whole ID process, it is necessary to ensure the consistency of the *product* of each phase across the overall process. However, there is (still) no real tradition in education of making formal notations of course designs. Such lack of common and formal notations makes the course development very local which hampers broader sharing between ID phases or stakeholders and impedes a better evaluation of design products [16]

To establish a common and formal notation, development and use of EMLs [21] and scripts [7], [10] have been moderately adopted by the community. Currently EMLs are integrated into IMS learning design (LD) specification as a standard [11] providing a sufficiently flexible framework that can be used to describe formally the design of almost any teaching-learning process [16]. Although such approach is much better than free handwriting notations, it neither helps users to keep the consistency/validity of the

course throughout the ID process nor allows for the development of intelligent tools that can support users during the design process.

Thus, the final goal of this study is to establish a comprehensive model for describing formally the design of variety forms of learning/instruction[1] (e.g. those summarized in [22]) through ontological engineering approach [3], [4], [19]. Especially, in this paper, we discuss a unification of one-to-one instruction, such as tutoring or individual e-learning course, and collaborative learning, in which learners teach and learn from each other. Although the attention to blended learning has been growing, most of the studies have been made on either type of them. Such a unified model will contribute to expansion of the range of instructional design and to share the design rationale of a course through the overall ID phases. Ontological engineering is expected to provide guidelines to find out the key concepts for such a unified model. In addition, while it cannot be discussed in this paper in detail, such a model is also expected to make contributions to modeling instructional design knowledge, which provides a valid composition of a model.

This paper is organized as follows: In section 2, instructional design processes are summarized and the requirements for comprehensive learning/instructional design process management are discussed together with its overview. In section 3, we describe ontologies we have proposed as the basis for a comprehensive model that support various forms of learning and instruction. The fourth section presents an example of modeling collaborative learning based on the Peer Tutoring theory. Finally, we conclude this paper with future directions of this study.

1. Towards a Comprehensive ID Process Management

This section gives an overview of the main phases of the available ID process models and discusses the requirement for comprehensive learning/instructional design process management. As mentioned in section 1, all ID process models share most of the same basic phases: The ***analysis*** phase involves analyzing a specific educational problem. The product of this phase is the terminal objective of the course. Usually, a list of questions is used to conduct analysis and the results are described narratively or in informal diagrams. In the ***design*** phase, learning/ instructional strategies to achieve the terminal objective are identified. The main product of this phase is a flow of learning/instruction which works as the mold for a particular learning/instruction. In the ***development*** phase, specific learning/ instructional materials used in the execution are assigned to the product of the design phase. In the ***implementation*** phase the course is delivered to learners and learning is conducted by it. The output of this phase is actual data of learning conducted by the course. Finally, in the ***evaluation*** phase, data collected in the implementation phase are compared with the design of the course. The gap between them is the point to be improved in the current course. Based on this result, the ID process returns to any other phase for improvement.

Through these phases, a course is produced as the final product that reaches the desired quality level. The problem pointed out here is that most of the products of each process are managed with narrative or simple, non-formal diagrams and tables [23].

[1] The term “instruction” is used in the wider sense in this paper therefore this means not only what a person does to instruct others but also what one does to support or facilitate learning of others [[2]]. The term “instructor” is also used in the sense.

Although IMS LD provides a formal framework to describe the products, this is just a format and does not have enough semantics for keeping their consistency or for assessing their validity [1].

This study proposes a framework to model the product (course) to manage the input and the output of each phase in the ID process comprehensively. If the framework has the potential to describe any learning/instruction process from the learning objective of a course to the learning materials employed in the course, the product can be maintained across the ID process consistently.

Figure1 ID process and ontology

We take the ontological engineering approach to tackle this issue through defining concepts related learning and instruction and organizing them as an ontology based on philosophical considerations. Figure 1 draws a rough sketch of a learning/instructional process model for facilitating the ID cycle based on such an ontology. The center of the figure denotes an ontology that defines concepts for modeling learning and instruction process as the product. The cycle around the ontology is the instructional design process composed of the typical basic phases. The ontology will be a foundation for maintenance of the product throughout the ID process. Currently the focus of this study is mainly on the input and output of the design phase (and a part of the development phase).

2. Ontologies for Modeling Learning and Instruction

We have proposed two ontologies for modeling learning: OMNIBUS [19], [20] and the Collaborative Learning (CL) ontology [12], [15]. Although the target of the former is one-to-one/more instruction and the latter is collaborative learning, both of them are based on the same working hypothesis and aim at providing a conceptual framework to model learning and instruction as well as structuring learning/instructional theories as guidelines to compose good learning and instructional scenarios. The core idea of these ontologies is that "learning" can be modeled as state change of learners. This is based on our working hypothesis that a sharable "engineering approximation" of the concept "learning" can be found in terms of the changes that are taking place in the state of the learners [8].

This core idea is conceptualized as *I_L event* and shared by the two ontologies. This concept, in which "I_L" stands for the relationship between Instruction and Learning, describes a learner state is achieved by the learner's action affected by the other's action, which can be considered to have any instructional effect. Under the concept of I_L event, the relationships among the actions and the learner's state change are conceptualized as one. This makes it possible to describe the relationships among various learning/instructional actions and state changes.

The following sub-sections describe, briefly, how individual learning and collaborative learning is modeled with I_L event in the two ontologies as the basis for a comprehensive modeling framework for the instructional design process.

2.1. OMNIBUS

One of the characteristics of OMNIBUS is to model learning/instructional process at various levels of granularity. At each level of granularity, learning/instruction process is modeled as a sequence of I_L event and the levels are multi-layered. In the layers, each I_L event at the upper level is related to I_L events at the lower one. This relation offers both top-down and bottom-down interpretations; the lower state changes of learner achieve the upper one and the upper action is realized by the lower ones, respectively. In OMNIBUS this is conceptualized as "WAY" In short, I_L events describe *what* to achieve and WAYs describe *how* to achieve it. Fig. 2 (a) shows an example of WAY. In the Fig. 2 (a), the oval nodes represent I_L events, and black squares linking the macro and the micro I_L events represent WAYs. Here, the macro I_L event has two WAYs; WAY1 and WAY2, and there is an "OR" relation between them. This indicates that there are two alternatives to achieve the macro I_L event.

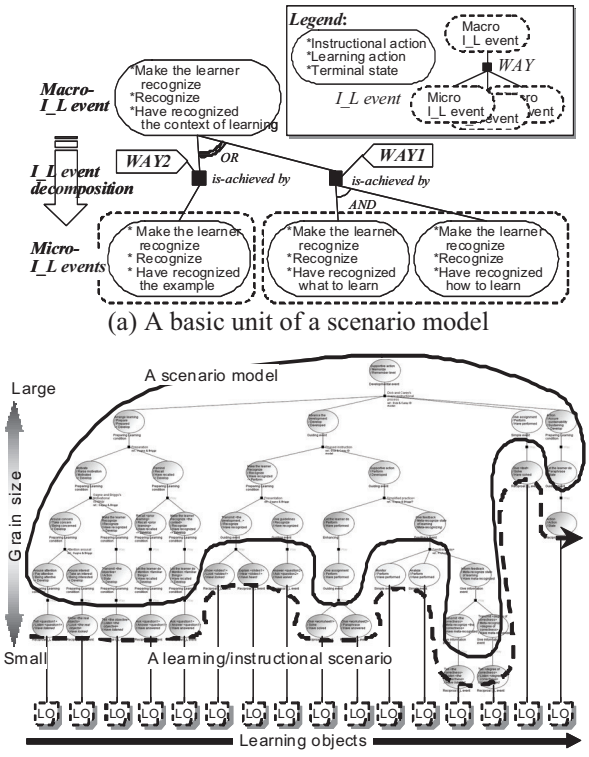


Figure 2 Scenario modeling based on OMNIBUS

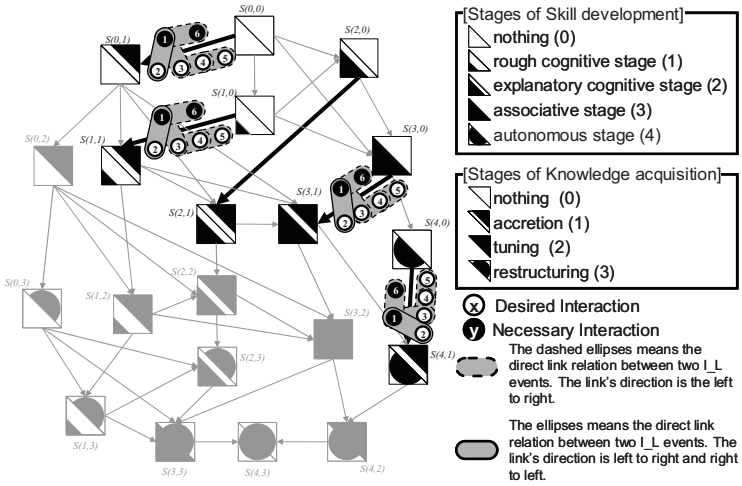
Based on OMNIBUS, a learning/instructional scenario is modeled described as a tree structure of I_L events decomposed by WAYs as shown in Fig. 2 (b). The leaf layer is a description of a learning/instructional scenario executed by instructors and learners, and is linked with LOs used in the execution. The tree structure excepting the leaf level explains the design rationale of the scenario and it works as the specifications of the LOs to be attached.

These concepts of I_L event and WAY also give a conceptual scheme to model strategies from learning/instructional theories. We have extracted 99 strategies from 11 theories and defined them as WAYs in OMNIBUS [9]. Such WAYs based on learning/instructional design knowledge, which includes learning/instructional theories, patterns and best practices, are called "WAY-knowledge" in OMNIBUS. This WAY-knowledge works as the guidelines for designing scenarios and as a justification to demonstrate their validity.

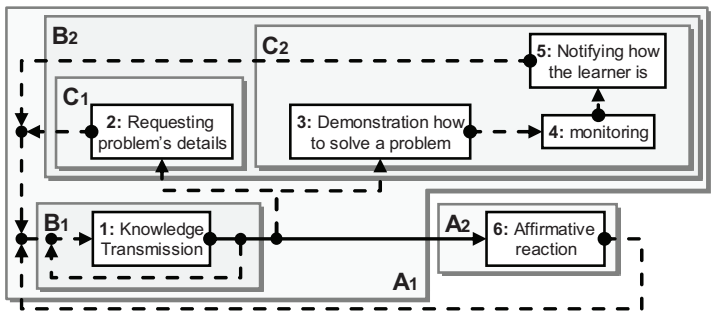
2.2. Collaborative Learning (CL) Ontology

The focal points of the CL ontology are also state changes, which are "learning", of each participant in collaborative learning and interactions between them. These are modeled as Growth Model Improved by Interaction Patterns (GMIP) [15] employing I_L event. Figure 3 shows an example of GMIP. GMIP has two components: one is Learner Growth Model (LGM) [13] and the other is Interaction pattern (IP) [14]. As shown in Figure 3(a), LGM represents, in a simplified way, possible transitions of states in the learner's knowledge acquisition process and skill development process as links in the graph. IP represents the flow of interaction between learners as shown in Figure 3(b), in which a node denotes an interaction modeled as I_L event. Through the connection of LGM and IP in GMIP each transition between states is connected with interactions between participants.

In collaborative learning, each participant is a learner with his/her own learning objective and sometimes his/her action helps or facilitates learning of others, which is refered to as instructional action in the conceptualization of I_L event. For example, in the theory of "Peer tutoring" [6], two types of role are defined: *PeerTutor-role* and *PeerTutee-role*. Participants assigned to a *PeerTutee-role* (PeerTutees) learn through being taght by the others assigned to a *PeerTutee-role* (PeerTutors). And the



(A) The LGM for Peer tutoring: Learning by being taught (Peer tutee)



(B) The Interaction pattern for Peer tutoring

Figure 3 Growth Model Improved by Interaction Patterns (GMIP)

PeerTutors also learn through teaching the PeerTutees. The important point here is that from the point of view of CL the PeerTutor does not act as a real instructor, who only teaches, because he/she is also a learner through learning by teaching. Such a dual-nature of a participant can be modeled by I_L events. Focusing on learning in *PeerTutee-role*, when a PeerTutee learns, a PeerTutor support the PeerTutee by teaching. On the other hand, focusing on learning in *PeerTutor-role*, when a PeerTutor learns, a PeerTutee support the learning by being taught. These are described in two different I_L events. Thus, in an I_L event, the PeerTutor teaches the PeerTutee, and, in another I_L event, he/she learns through teaching the PeerTutee.

GMIP defines one IP and one or more LGMs corresponding to each role. Thus, although Fig. 3 has only one LGM for PeerTutee-role, actually there is another LGM for PeerTutor-role. GMIP helps to explicitly show how learners in the group should interact with each other and the benefits for learners playing different roles. Thus, it becomes a powerful tool in helping designers to select appropriate interactions and roles to achieve desired learning goals.

3. An Integrated Model of Learning and Instruction

Based on the ontologies described in the previous section, we aim at modeling various forms of learning/instruction (eg. those summarized in [22]), which is the product of the ID process. As discussed previously, employing I_L event as the basis, GMIP allows to model roles of participants in collaborative learning and interaction among them to achieve the learning goals. Thus, each interaction between two roles/participants is modeled as I_L events, defining which participant learns or supports the learning in a given interaction.

Although GMIP currently aims at describing CL, it can be used to model other forms of learning. Consider the case shown in Fig. 4 where three roles are defined. PeerTutor ($Role_1$) teaches PeerTutee ($Role_2$) and, from the behavior of PeerTutor, Observer ($Role_3$) learns how to teach others. As stated above, the basic unit of GMIP is a set of LGMs and an Interaction pattern. In the $interaction_{12}$ each of PeerTutor and PeerTutee has its role's learning goal described as LGM_1 and LGM_2, respectively. On the other hand, in the $interaction_{13}$, only Observer has the learning goal because PeerTutor is just observed and does not always need to be conscious of the Observer. The interaction pattern is an aggregation of the interactions between these roles. The I_L event decomposition tree ($DT_{1\sim3}$ in Fig. 4) discussed in Section 3.1 fulfills a role to explain how each of the goals relates to the interaction pattern. In addition, an interaction pattern and some LGMs connected with the I_L event decomposition trees work as a generic model for learning and instruction. Even if the number of roles and interactions are increased, it can be modeled with additional LGMs and decomposition trees. On the other hand, in the case of one-to-one instruction, only an LGM and a decomposition tree are related with the interaction pattern because the learning goal of the instructor can be ignored, as in the example of the interaction between PeerTutor and Observer in Fig. 4.

Using this idea, we will show how to model CL as a formal product of the ID process with our proposed modeling framework through an example based on the theory "Peer tutoring" [6]. Figure 5 shows an example of collaborative learning model based on Peer tutoring. As mentioned above, in Peer tutoring, learners play two types of collaboration roles: the *peer tutor* role and the *peer tutee* role. The learning objective

for each role can be described in the LGMs shown in Fig 5 (x). Although there are some active paths in the LGMs (emphasized arrows in Fig. 5 ($x_{1,2}$)), the essence is that the objective of peer tutor is *Tuning* and the one of peer tutee is *Accretion* as shown in Fig 5 (x').

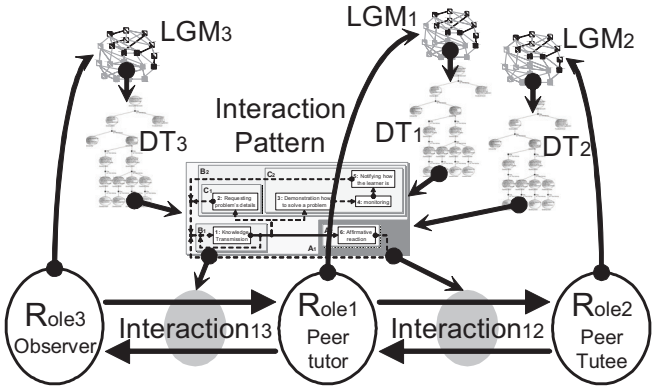


Figure 4. An overview of the integrated model

These objectives are achieved by the activities of participants assigned to the roles, which are informing the topic to the peer tutee by the peer tutor, practice by the peer tutee, and guiding the practice by the peer tutor. These activities are defined as an interaction pattern shown as Fig. 5 (z), which is the one redrawn from Fig. 3 (b) in order to establish it to the I_L event decomposition trees (Fig. 5 (y)). The I_L event decomposition tree supplies links between the objective and the interaction pattern, and explains the design rationale of the link.

I_L event decomposition trees are constructed along the decomposition of learning objectives. Here the root of each decomposition tree is set as the objective defined by the LGM. This state (change) is decomposed into smaller-grain-sized ones with learning and instructional actions. Fig. 5 (y) illustrates a path of decomposition to a leaf I_L event in each I_L event decomposition tree. Each I_L event is decomposed into some I_L events or embodied in a much more concrete I_L event until the objectives are achieved by actions.

The essential of the relation between an LGM and an I_L event decomposition tree (DT) is that the LGM represents what to achieve and the DT represents how to achieve. Of course, as stated above, the DT is also an accumulation of smaller-grain-sized what and how to achieve. The relation is the top level distinction of what and how to achieve as it were. This also means that the combination of a LGM and a DT is flexible. That is to say, an LGM can be achieved with some different DTs. In the case of Fig. 5, although the LGMs for Peer tutor and Peer tutee (Fig. 5(x_1) and (x_2)) are achieved by DTs based on Peer tutoring theory (Fig. 5(y_1) and (y_2)), these can be achieved by other DTs based on other theories, best practices or ideas of designers. If a designer wants to adopt a different way to achieve the LGM, he/she can set a different DT without changing the LGM.

The interaction pattern is the same as the sequence of the leaves of decomposition, which is interaction between the participants as a cycle of activities shown in Fig. 5 (z). This is expressive enough to give the flow of activities while the corresponding DT gives its rationale. For the participants of CL, interaction patterns are useful because they give concrete guidelines that tell them the required activities to carry out the CL effectively. On the other hand, for designers as well, they are important to identify the characteristics of interaction among the participants. Interaction among them is made explicit mainly in an interaction pattern, by which the DTs are correlated each other. A cluster of the components in the interaction pattern corresponds to intermediate I_L event in the tree. For example, the cluster A_1 in Fig 5 (z) corresponds to I_L event A_1 in

Fig 5 (y), and the cluster A_1 is composed of B_1 and B_2 in Fig 5 (z) because I_L event A_1 is decomposed into I_L event B_1 and B_2 in Fig 5 (y).

As discussed in this section, through the line from a LGM to an Interaction pattern through a DT, the design rationale of collaborative learning scenario can be revealed and maintained across the phases of instructional design. In addition, I_L event decomposition tree is helpful to assess the consistency between the learning objectives and the interactions. For example, there are other ways to achieve making the peer tutee meta-recognize his/her own understanding (Fig. 5 (y_2)-B_2) than informing the peer

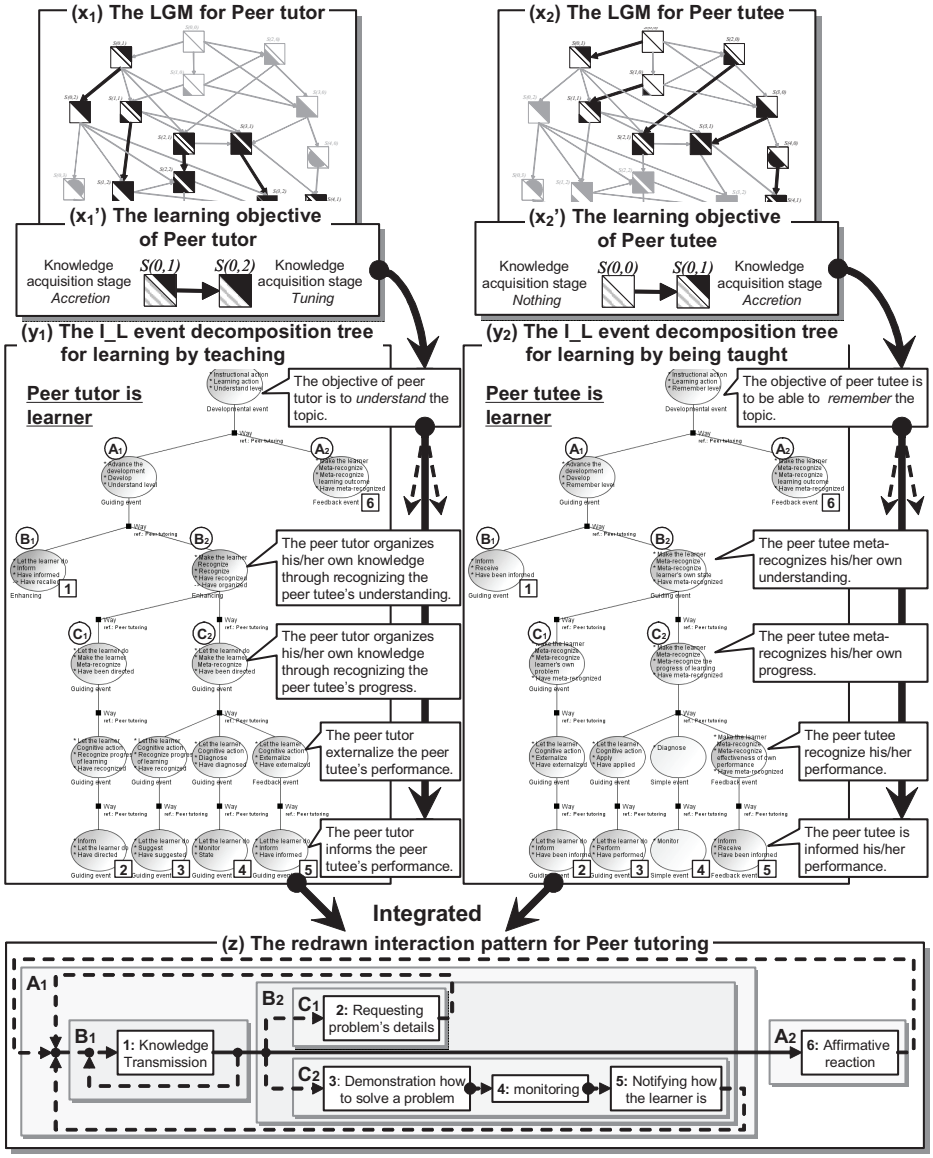


Figure 5. An example of the integrated model

tutee's performance (Fig. 5 (y_2)-5). An example is that the peer tutor demonstrates how the tutee solved the problem. In this case, although it is more difficult for the peer tutee to achieve, he/she can be trained in monitoring his/her own performance additionally. However, if the way is adopted, a problem occurs in learning of peer tutor. In this scenario the peer tutor learns through diagnosing the peer tutee's performance and informing the result. The peer tutor cannot learn by just demonstrating again. Like this, in our proposed modeling framework, such inconsistency between collaboration roles can be identified easier than other modeling such as IMS LD. However, in order to keep consistency among DTs, some sort of software module, which is aware of the ontologies, is required. OMNIBUS and CL ontology define constraints that provide possible relation between concepts for describing DTs. In addition to that, it is necessary to organize possible combinations of I_L events between roles, and to develop a mechanism outside of the modeling framework. This is a future issue.

If learning of PeerTutor is not intended in a learning session, this model can be considered to be the same as one-to-one instruction, in which PeerTutee learns through being taught by PeerTutor, neglecting the GMIP and the DT of PeerTutor. A set of a GMIP, a DT and an Interaction pattern is a basic unit. Depending on the form of learning and on the number of roles that have intended learning objectives in the learning session.

In conclusion, the presented framework allows for formally describing the product of the ID process for different forms of learning and, therefore, it helps to ensure the consistency of the *product* across the overall ID process and to manage the input/output of each phase of the ID process comprehensively.

4. Conclusion

The ID process is a complex task composed of many phases (analysis, design, development, implementation, and evaluation). To keep the consistency and the validity of the product (the course) in each phase, it is necessary to have a formal and semantically rich framework that allows for a better model of the product. Therefore, this paper discussed previous achievements on modeling individual and collaborative learning/instruction using ontologies, and how the accumulation of these past results together with a shared key concept to represent "learning" (I_L event) allow for the development of a framework that can describe formally learning and instructional scenarios. Such a description facilitates the sharing of the product of each ID phase and enables the systematic design of the course.

To show the potential use of our framework, section 4 presented an example that covers the design phase of the ID process showing the creation of collaborative learning activities based on the Peer Tutoring theory. Due to space limitation, we could neither discuss the usability of our model in other ID phases nor present more details about the framework. However its potential benefits to support the ID process has been demonstrated.

The future direction of this study will expand the proposed modeling framework to tackle many other difficulties found in other phases of the ID process. For example, the analysis and development phases need much more detailed attributes in the context of learning and the implementation and evaluation phases require a mechanism for data collection and comparison of it to the design of course.

Furthermore, in order to realize our proposed idea, it is necessary to build an authoring system based on the modeling framework. We have developed two authoring systems, which are SMARTIES [20] based on OMNIBUS and CHOCOLATO [16] based on the CL ontology. Based on these systems we would like to go on to integration of them to develop a unified system which supports the overall ID phases.

References

[1] Amorim, R., Lama, M., Sánchez, E., Riera, A., and Vila, X.: A Learning Design Ontology based on the IMS Specification. *Journal of Educational Technology & Society*, 9(1), 38-57, (2006).

[2] Carr-Chellman, A.A. and Hoadley, C.M.: Looking back and looking forward, Educational technology, 44(3), 57-59, (2004).

[3] Devedzic, V.: *Semantic Web and Education*, Springer ScienceBusiness Media, 2006.

[4] Dicheva, D.: O4E Wiki, http://o4e.iiscs.wssu.edu/xwiki/bin/view/Blog/Articles, (2008).

[5] Dick, W., Carey, L., & Carey, J. O.: *The systematic design of instruction*, Fifth edition, Addison-Wesley Educational Publisher Inc., (2001).

[6] Endlsey, W. R.: *Peer tutorial instruction*, Englewood Cliffs, NJ: Educational Technology, (1980).

[7] Harrer, A.: "An Approach to Organize Re-usability of Learning Designs and Collaboration Scripts of Various Granularities". *Proc. of ICALT 2006*, pp. 164-168, (2006).

[8] Hayashi, Y., Bourdeau, J. & Mizoguchi, R.: Ontological Support for a Theory-Eclectic Approach to Instructional & Learning Design, *Proc. of EC-TEL2006*, 155-169, (2006).

[9] Hayashi, Y., Mizoguchi, R., and Bourdeau, J.: Structurization of Learning/Instructional Design Knowledge for Theory-aware Authoring systems, *Proc. of ITS2008*, 573-582, (2008).

[10] Hernandez, D., Villasclaras, E.D., Asensio, J.I., Dimitriadis, Y.A., Retalis, S.: CSCL Scripting Patterns: Hierarchical Relationships and Applicability, *Proc. of ICALT 2006*, 388-392, (2006).

[11] IMS Global Learning Consortium, Inc.: IMS Learning Design. Version 1.0 Final Specification, (2003).

[12] Inaba, A., Supnithi, T., Ikeda, M., Mizoguchi, R., Toyoda, J.: How Can We Form Effective Collaborative Learning Groups?, Proc. of ITS2000, 282-291, (2000).

[13] Inaba, A., Ikeda, M., & Mizoguchi, R., What Learning Patterns are Effective for a Learner's Growth?, Proc. of AIED2003, 219-226, (2003a).

[14] Inaba, A., Ohkubo, R., Ikeda, M., & Mizoguchi, R.: Models and Vocabulary to Represent Learner-to-Learner Interaction Process in Collaborative Learning, Proc. of ICCE2003, 1088-1096, (2003b).

[15] Isotani, S. and Mizoguchi, R.: An Integrated Framework for Fine-Grained Analysis and Design of Group Learning Activities, Proc. of ICCE2006, 193-200, (2006).

[16] Isotani, S. and Mizoguchi, R.: Adventure in the Boundary between Domain-Independent Ontologies and Domain-Specific Content for CSCL, Proc. of the 12th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES'08), pp. 523-532, 2008.

[17] Koper, R.: An introduction to learning design, Koper, R. and Tattersall, C. (Eds.), Learning design A handbook on modeling and delivering networked education and training, 3-20, Springer-Verlag Berlin Heidelberg, (2005).

[18] Leshin, C. B., Pollock, J., & Reigeluth, C. M.: *Instructional Design Strategies & Tactics*, Englewood Cliffs, NJ: Educational Technology Publications, (1992).

[19] Mizoguchi, R. and Bourdeau, J.: Using Ontological Engineering to Overcome Common AI-ED Problems, *International Journal of Artificial Intelligence in Education*, 11(2), 107-121, (2000).

[20] Mizoguchi, R., Hayashi, Y., & Bourdeau, J.: Inside Theory-Aware & Standards-Compliant Authoring System, *Proc. of SWEL'07*, 1-18, (2007).

[21] Rawlings A., Rosmalen van P., Koper R., Artacho M., and Lefrere P., "Survey of Educational Modelling Languages (EMLs)," CEN/ISSS WS/LT (2002). (Available via the Internet. http://www.cenorm.be/cenorm/businessdomains/businessdomains/isss/activity/emlsurveyv1.pdf. Accessed 30 April 2009)

[22] Reigeluth, C.M.: What is instructional-design theory and how is it changing, Reigeluth, C.M. (Eds.), Instructional-design theories and models A new paradigm of instructional theory, Lawrence Erlbaum Associates, Mahwah, N.J., 5-29, (1999).

[23] Sloep, P., Hummel, H., and Manderveld, J.: Basic design procedures for E-learning, Koper, R. and Tattersall, C. (Eds.), Learning design A handbook on modeling and delivering networked education and training, 140-160, Springer-Verlag Berlin Heidelberg, (2005).

Open-corpus Personalization Based on Automatic Ontology Mapping

Sergey SOSNOVSKY
University of Pittsburgh, School of Information Sciences
135, North Bellefiled ave, Pittsburgh, PA, 15260, USA
sosnovsky@gmail.com

Abstract. Open-corpus personalization is an important problem for modern Web-based adaptive systems. Its solution can greatly advance dissemination of adaptive and intelligent technologies, especially in education. The existing closed-corpus adaptive learning systems operate with a limited amount of content in a limited number of learning domains. A teacher willing to provide students with personalized access to open-corpus online resources of her/his choice needs an easy way of organizing them into an adaptive tool. This work proposes an approaches for open-corpus personalization in the context of e-Learning that relies on automatic extraction of coarse-grained content models and translation of these models into a finer-grained domain ontology for student knowledge assessment. The approach is implemented in the Ontology-based Open-corpus Personalization Service (OOPS) that adaptively recommends online reading resources to students.

Keywords. Open-corpus personalization, ontology mapping, adaptive recommendation

1. Introduction

Web-based access has become a de-facto standard for most types of adaptive systems. WWW as an infrastructure for information and service delivery provides unique opportunities in terms of user access and availability, inter-component connectivity, and the volume of information resources. However, it also brings new challenges, such as privacy-enhanced personalization, reliable user identification, user model interoperability, etc. One of the adaptation problems that are very important in Web settings is open-corpus personalization (Peter Brusilovsky & Henze, 2007; Henze & Nejdl, 2001). In the pre-Web era, the volume of learning resources available for the user of an adaptive educational system was restricted to the pre-authored corpus of learning problems, pages of reading material, quiz questions, etc. The system "owned" its content and did not have access to any documents outside. On the Web, which is ultimately a global network of information items, every document is potentially available for any adaptive system. The challenge is how to maintain adaptive access to new documents.

This paper describes an approach to open-corpus personalization in the context of e-Learning. It augments existing online quiz system with adaptive recommendation of open-corpus reading material. A formal domain ontology is used as a basis for modeling student knowledge and as a reference model for coarse-grained structures of reading content. The open-corpus content targeted in this research includes various kinds of semi-structured online learning resources available on the Web, such as online tutorials, textbooks, digital libraries etc. Coarse-grained topic-based models of such resources are harvested from their hyper-linked structures and the HTML markup of the Web pages. The extracted models are translated into the central domain ontology with the help of ontology mapping. As a result, student knowledge modeled in terms of fine-grained ontology concepts is mediated into the topics of open-corpus resources. This allows adaptive recommendation of pages associated with appropriate topics.

The paper is structured as follows. Section 2 introduces the problem of open-corpus personalization in more details. Section 3 overviews the relevant work in the area of Semantic Web technologies for open corpus content modeling. Section 4 describes the details of the proposed approach, and the developed system. Finally, Section 5 discusses the future plans of this project.

2. The Problem of Open-corpus Personalization

One of the major challenges for a content-based adaptive adaptation is the efficient modeling of the adapted content in terms of domain elements. The interactions of a user with an adaptive system's content may be collected from the system's logs; however, the task of the adaptive system is to interpret this data and utilize it for adaptive support of future user interactions. Traditional content-based adaptation relies on composite domain models and associations between content items (tutorial pages, problems, etc.) and domain concepts. It allows adaptive systems to reason in terms of domain semantics instead of content and to generalize users' characteristics based on the history of their interaction with the system. A closed-corpus applications operating on a restricted set of content items require a domain expert to manually associate content items with domain elements at the authoring time. Figure 1 presents the general design of a closed-corpus adaptive system. The adaptive content is embedded in the system along with all other components. Once the system is developed, no new documents can be added to it. Authoring tools and shells (Murray, 1999) partially help to remedy this restriction, as they allow for reuse of adaptation components and reduce the authoring time. Every new system, though, would still require development of a content model.

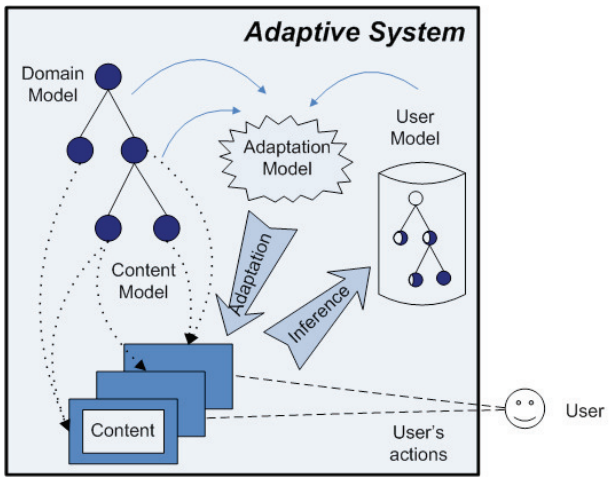


Figure 1. **Closed-corpus adaptive system**

In the Web settings, this approach no longer allows efficient user support. There is a large volume of relevant Web-content that has not been processed by the system developers. The system cannot trace user's interaction with this content; neither can it include it in the adaptive inference to maintain the user model adequately.

Web-based open-corpus personalization is not reduced to the modeling of open-corpus content. The relevant information needs to be discovered. New documents should be linked to the existing ones and users should be given an effective way to access them. The problems of dynamic content and dead links need to be addressed, as well. The content quality control is another issue, which is extremely important for e-learning. While recognizing all these challenges, this paper is mainly focused on modeling of adaptive content as a central part of open-corpus personalization process.

Figure 2 demonstrates a closed-corpus adaptive system facing the problem of open-corpus personalization in the context of e-Learning. The WWW can be viewed as a very large collection of educational material. For many subjects one can find online tutorials, textbooks, examples, problems, lectures slides, etc. Nowadays, teachers do not have to create a lot of content themselves; they can utilize the best of what is available.

For example, a teacher developing a course on Java programming might decide to use the Sun Java Tutorial (http://java.sun.com/docs/books/tutorial/), the electronic version of the chosen course book, the PowerPoint presentations of the course lectures, a quiz system and online code examples. Students should find these resources useful, however, they might get lost in this amount of content without additional guidance. Organizing adaptive access to the course resources would help solve the problem; appropriate resources will be recommended to the students based on their progress. However, even if one of these tools is implemented as an adaptive system (e.g. the quiz system), incorporating the rest of them into this system

is not possible unless it supports open-corpus personalization. An adaptive quiz system will have no knowledge about the rest of the content available to the students and can neither recommend it nor take students' interaction with this content into account to better model their knowledge.

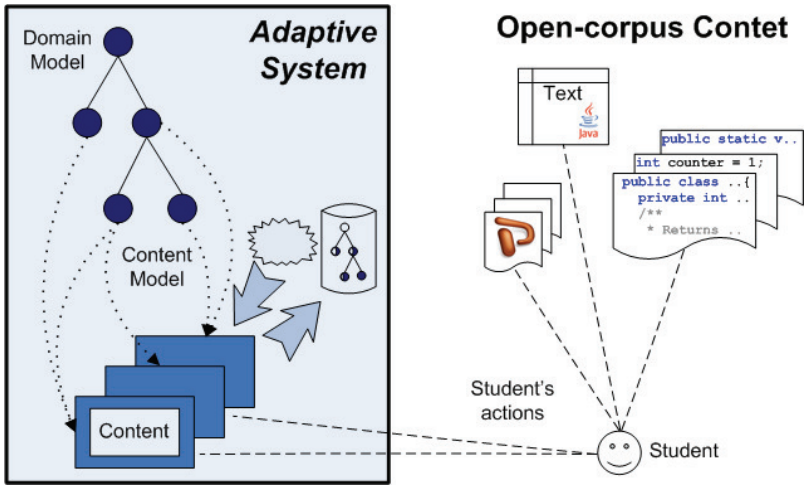


Figure 2. **Open-corpus personalization in e-Learning**

There are two approaches to deal with the open-corpus content problem that dominate the area of Web-systems: keyword-based and collaborative/social approaches. Keyword-based content modeling relies on automatic extraction of keyword vectors from documents' content. Many adaptive Web systems model user information interests or needs as vectors of keywords extracted from the documents that the user has browsed or requested (sometimes such vectors were enriched with tf*idf values), e.g. Letizia (Lieberman, 1995), WebMate (Chen & Sycara, 1998), NewsDude (Billsus & Pazzani, 1999), etc. However, the keywords support only shallow content models and does not ensure high level of modeling accuracy, which is an important drawback in the context of e-Learning, where the precision is required to be near perfect. Collaborative filtering (Konstan, et al., 1997) and social navigation (P. Brusilovsky, et al., 2004) are two other approaches that became popular for open-corpus personalization on the Web. Instead of content models and metadata they are based on the users themselves, the similarities between users and the patterns of social behavior they follow. Collaborative filtering looks for the like-minded users expressing similar interests/needs/preferences (purchasing the same products, rating items similarly, etc.) to generate recommendations of items that the user might like. Social navigation exploits the natural behavioral regularity of human users to follow the crowd by displaying the amount of attention users give to resources. Both these approaches provide the basis for full-scale open-corpus personalization; however, their application in e-Learning systems is limited. Collaborative systems provide an individual user with recommendations based on the evidence collected from other users, although learning is a very individual process. The fact that a certain student accessed the learning resources in a particular order or answered a problem correctly only after two incorrect attempts does not imply that other students will follow the same pattern. Social navigation, while being a very successful for content quality control, has no mechanism of guiding a student to the "right" resource at the "right" moment.

Most of the adaptive and intelligent educational systems rely on adaptation technologies that require accurate and meaningful content-based models, therefore the use of keyword-based and collaborative/social approaches for open-corpus personalization in the context of e-Learning is rather situational. They can supplement other solutions; however, hardly can serve as a basis for the broad class of adaptive educational systems.

3. Ontology-based Techniques for Open-corpus Personalization

Automatic and semiautomatic generation of content-based models and metadata, which is a central problem for open-corpus personalization, has been recently an important topic of research for Semantic Web community, as well. The Semantic Web platform requires effective means for solving so-called "knowledge acquisition bottleneck problem", which refers to the lack of ontologies and semantic metadata

necessary for dissemination of Semantic Web ideas. Several techniques relying on the use of Semantic Web ontologies for knowledge representation have been proposed for the open-corpus personalization problem. This section provides a short overview of this work.

One of the first attempts to apply ontologies for building an open-corpus adaptive system was made in KBS Hyperbook project (Henze & Nejdl, 2001, 2002). The authors recognized the inability of adaptive systems to personalize the external content and proposed a solution based on the strict separation of the adaptive functionality from the domain model and the content it describes. The system relied on representation of domain models as ontologies, which allowed authors to employ a standard set of inference rules for building the adaptive component of the system. This ensured the compatibility of the system's logic with any potential domain model as long as its representation followed the same ontological requirements. For any new content added to the system, a new content model could be created. If an author wanted to update an existing document space, s/he could simply modify its content model by adding semantic markup for the new documents. The authors implemented this approach for developing an educational hypermedia system for learning Java programming.

The main contribution of KBS Hyperbook is the ability to extend an existing adaptive system with novel content. However, it did not provide any functionality for automatic authoring support. The human expert still had to associate the new pages with the ontology concepts and update the content model manually. In our scenario, a teacher would have to manually link the external resources to the concept of the domain model. A typical index of a learning resource consists of many concepts (sometimes the number exceeds several dozens). Although the KBS Hyperbook approach opens the possibility of building open-corpus systems, it does not scale well given the need for manual indexing.

Several recent projects address the open-corpus content modeling problem by exploiting the ontology-based annotation techniques for automatic indexing of textual Web-resources with semantic metadata. One of them is *COHSE* (Conceptual Open Hypermedia Service), the joint project between Sun Microsystems and University of Manchester (Bechhofer, et al., 2003; Yesilada, et al., 2007). *COHSE* is based on the original idea of distributed link services (Carr, et al., 1995) working as intermediaries between Web clients and Web servers and augmenting Web documents with dynamic links. Architecturally *COHSE* works as a proxy. It intercepts a client's HTTP request, retrieves the original HTML document, enriches its content with new links and delivers the modified document to the user's browser. The extra links added by *COHSE* are based on the contents of both the original source document and the target document. They are placed in the way to connect the key pieces of text in the original document with relevant HTML context elsewhere. Such knowledge-driven linkage is performed on the basis of an ontology serving as a source of consensual domain semantics. The *COHSE* components apply ontology-based annotation technologies to associate automatically the pieces of documents with ontology concepts. When a document is requested, its annotations act as links to the concepts and then to other documents annotated with these (or related) concepts.

A similar approach is implemented in *Magpie* (Domingue, et al., 2004; Dzbor & Motta, 2007). Unlike *COHSE*, *Magpie* works on the client side as a browser plug-in. It analyses the content of the HTML document being browsed on-the-fly and automatically annotates it based on a set of categories from an ontology. The resulting semantic mark-up connects document terms to ontology-based information and navigates the user to the content describing these terms. For example, if *Magpie* recognizes that a particular phrase is a title of a project, it populates a menu with links to the projects details, research area, publications, members etc.

The main drawback of this approach is its reliance on a number of natural language patterns and heuristics that are often situational. Modern ontology-based text annotation techniques are capable of producing good results in recognizing named entities, such as places (countries, cities), people, organizations etc. They also implement procedures for recognizing pieces of content with very specific formatting, like dates, times, coordinates, citations etc. Annotation of general content in terms of an arbitrary ontology may create a challenge for these tools. In the context of e-Learning, such limitations can satisfy a very small number of domains and types of content. Nevertheless, ontology-based annotation can provide the functionality necessary for open-corpus personalization. Some work in this field has been done recently for the task of adaptive recommendation. Systems like *Quickstep* (Middleton, et al., 2001), *Foxtrot* (Middleton, et al., 2003) and *MyPlanet* (Kalfoglou, et al., 2001) apply ontology-based annotation to automatically index dynamic textual content and recommend it adaptively.

One of the promising directions to advance the automatic knowledge acquisition technologies, which should also greatly contribute to open-corpus personalization, lies on the border of Semantic Web and

Social Web technologies. Several research teams seek to bridge socially authored folksonomies and formal ontologies in order to enrich huge pools of social tag-based metadata with semantics and structure (Angeletou, et al., 2007; Van der Sluijs & Houben, 2009). Online services like twine.com, realtravel.com, and freebase.com employ sets of existing ontologies to automatically annotate web-content generated or discovered by their users.

4. Ontology-based Open-corpus Personalization Service

4.1 The approach

Much of today's Semantic Web research is carried on under the slogan "A little semantics goes a long way" (Hendler, 2007). Several small pieces of RDF-based metadata linked with roughly specified relations cannot maintain powerful ontological inference or ensure best possible formalization of a domain semantics, however implemented on a WWW-scale such network of shallow knowledge can support many interesting applications. Systems like Revju.com (Heath & Motta, 2008) and Headup (SemantiNet, 2008) are among multiple recent examples of this paradigm implementation.

This project follows a similar pragmatic approach in attempt to create an effective application by exploiting coarse-grained semantics of semi-structured textual Web-resource. The main idea behind the project is to benefit from a large base of high-quality online reading material that has been authored by domain experts for those who need to learn about the domain. Information resources like tutorials, books, and digital libraries are usually created by knowledgeable authors with a certain domain structures in mind, which can be regarded as domain models. Such models are reflected in the form of tables of content, hyperlinks connecting the pages and HTML markup. They are always coarse-grained (with a topic corresponding to a page or a section of a page) and often subjective (different authors can structure the same domain differently). However, they allow breaking the reading material into logical pieces corresponding to meaningful domain categories. Human readers usually do not have problems with navigating through these categories (or topics) and understanding the material they structure. A minimal set of requirements to a topic as a domain modeling unit allows automatic harvesting of topic-based models from semi-structural online resources. Once the topic-based model of the Web-resource is created, it can be used by an adaptive system as the basis for adaptive presentation of this resource. For example, once the AES understands that a student needs help on a certain topic, the textual information associated with this topic can be retrieved and presented to the user.

A potential criticism of this adaptation approach is that it makes adaptive decisions based on coarse-grained domain models. How effective the recommendation of online reading material will be if a student receives a large piece of text rather than an extract that is relevant to the concept(s) s/he has been working on? Our experience with coarse-grained adaptation implemented in the system QuizGuide has shown that large topics can serve as the basis for successful adaptation approaches that lead both to the increase in students' performance (P. Brusilovsky, Sosnovsky, Yudelson, et al., 2005) and motivation (P. Brusilovsky, et al., 2006). Although topic-based adaptation demonstrates a number of positive features, topic-based user modeling suffers from low accuracy and predictive validity. A topic turns out to be an effective adaptation unit; however, it is too large for precise identification of student knowledge. A single topic represents too much domain knowledge; therefore, when a student makes different mistakes in the problem belonging to the same topic, there is no way for a topic-based user model to distinguish these mistakes. At the same time, an adaptive reaction attributed to the same topic does not create much of a problem; a student makes the mistakes within a single topic and receives an appropriate topic-wise intervention from the system (S. Sosnovsky & Brusilovsky, 2005).

Another problem of an open-corpus adaptive system extracting the topic-based domain models from several collections of documents is the necessity to aggregate users' activity with the documents from different collections. Otherwise, the system can model users' characteristics only separately for all topic-based models and adapt the resources from different collections only based on their own models. Essentially, such a system will act as a cluster of several independent systems not capable of delivering adequate and consistent adaptation. For instance, if a student accessed a document about `for`-loops in Java from one tutorial, the system will not be able to use this information to understand that s/he probably does not need to see a similar page from another tutorial.

To remedy both problems, this paper proposes a solution based on the automatic mapping of a central fine-grained ontology for modeling students' knowledge and automatic mapping of the concepts from this ontology to the topics from the extracted models of online document collections. A fine-grained ontology ensures a high-quality overlay knowledge modeling. At the same time, it is used as a referential domain model for all harvested topic-based models to deliver coherent adaptation across multiple collections of documents. The central ontology and the topic-based models are aligned using an ontology mapping technique, where topic-based model is treated as a second ontology. This approach has been implemented in the ontology-based open-corpus personalization service (OOPS), which is described in the next two subsections.

4.2 The architecture

OOPS is developed as a value-adding service that augments existing e-Learning infrastructure with adaptive recommendation of open-corpus reading content. OOPS does not perform knowledge modeling itself, neither it has capabilities for knowledge tracing. The only type of educational content available to OOPS is pages of reading material that cannot provide reliable evidence for student knowledge update. Instead, OOPS rely on a third-party application that is capable to elicit current levels of students' knowledge and can benefit from timely recommendation of supporting readings. In its current implementation OOPS interacts with the user modeling server CUMULATE (P. Brusilovsky, Sosnovsky, & Shcherbinina, 2005), to receive students' knowledge levels and submit students' reactions to recommendations. The students' knowledge are assessed based on their work with QuizJET system (Hsiao, et al., 2008) that delivers online self-assessment quizzes and evaluates students' answers. QuizJET questions are indexed with the concepts of domain ontology, thus providing the means for overlay knowledge modeling.

Essentially, QuizJET and CUMULATE form the distributed version of the traditional close-corpus e-Learning system, where every learning resource is described with the domain metadata at the design time, so that tracking student activity with these resource allows modeling of students knowledge and maintaining adaptive access to this and other pre-authored resources. Figure 3 presents this architecture with central components and OOPS components painted differently. Upon retrieving student knowledge from the central user modeling component OOPS uses the mapping between the central ontology and the harvested topic-based model to translate student's knowledge into this coarse-grained model. As a result OOPS computes a list of five recommendations that are the most important for a student based on his knowledge and the question the student is currently working on.

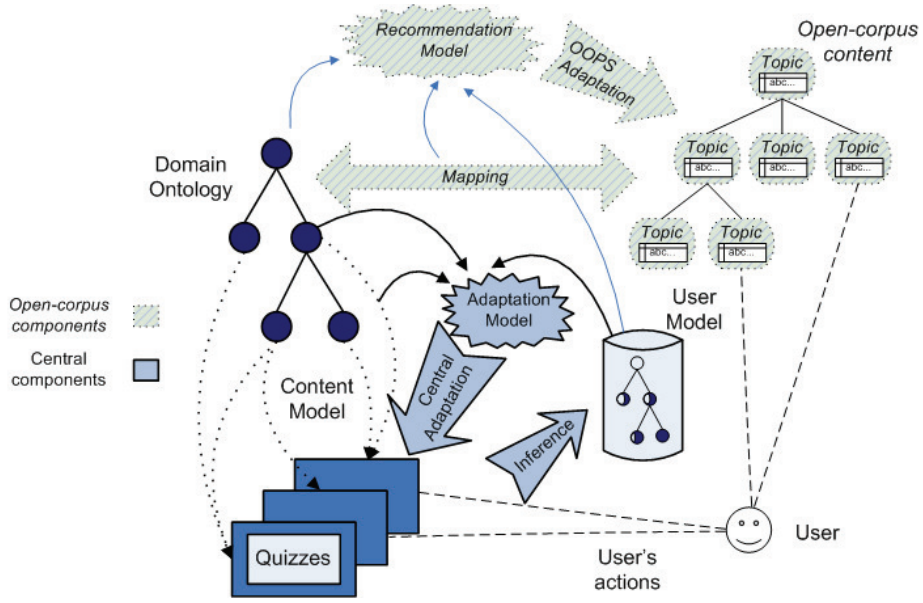


Figure 3. **OOPS Architecture**

Currently OOPS has been implemented only for one domain – Java programming language. Although, from the architecture point of view, all its components are domain independent. To start working OOPS requires an existing domain ontology and a URL of a collection of relevant reading resources. The ontology used for Java implementation of OOPS has been designed in the University of Pittsburgh, primarily, to support the development of adaptive educational content for Java and facilitate integration of multiple educational systems in this domain, while ensuring the objective representation of Java semantics. The level of granularity of the terminal concepts has been chosen to maintain the adequate modeling of students' knowledge with the necessary details. At the same time, the goal was not the comprehensive representation of all aspects of Java programming technology, therefore certain parts of the domain stay out of the scope if this ontology. The ontology is implemented as a light-weight OWL-Lite ontology. It specifies about 350 classes connected to each other with one of the three relations: standard *rfs:subClassOf* (hyponymy relation), a transitive relation pair *java:isPartOf – java:hasPart*, (meronymy relation), and *java:relatedTo*, which have been introduced to model a semantic connection between two classes, where neither of two previous relation types can been used. Figure 4 presents an extract of the Java ontology. The ontology can be accessed at http://www.sis.pitt.edu/~paws/ont/java.owl

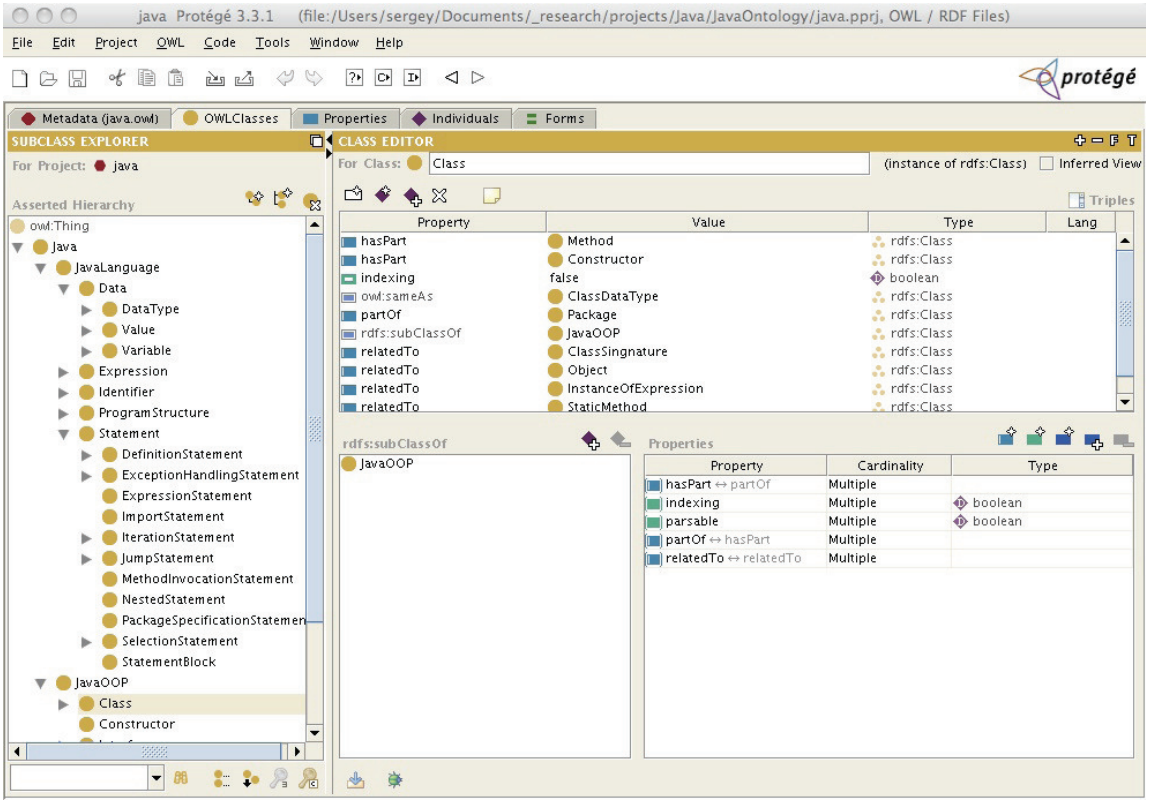


Figure 4. **Java Ontology**

The topic-based model has been harvested from a part of electronic version of introductory Java programming textbook (Horstmann, 2007). The number of topics in this model is comparable to the number of ontology concepts, however, the granularity of concepts is higher. Every topic corresponds to a single section in the book. However, every such section can be indexed by a set of ontology concepts. Figure 5 presents the extract from this model.

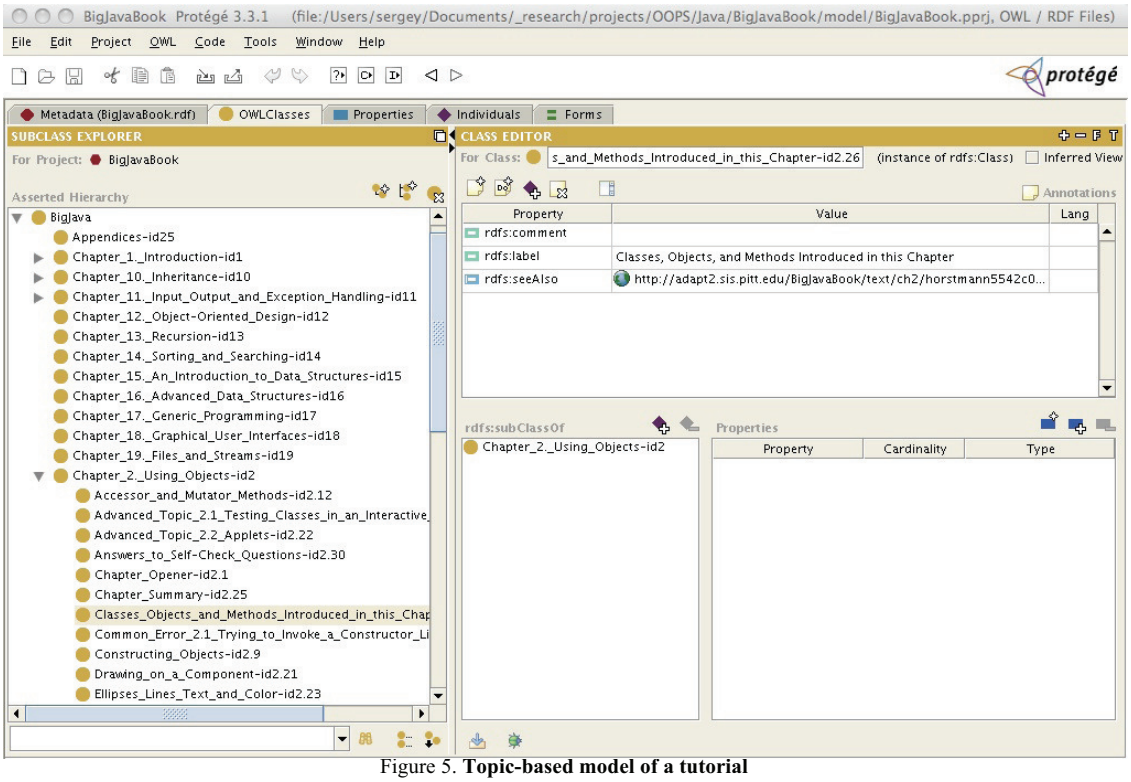


Figure 5. **Topic-based model of a tutorial**

4.3 The interface

OOPS is implemented as a value-adding service. It acts as a wrapper for the host content by enriching it with adaptive recommendations. Figure 6 present an example of a list of recommendations generated by OOPS for a QuizJET question.

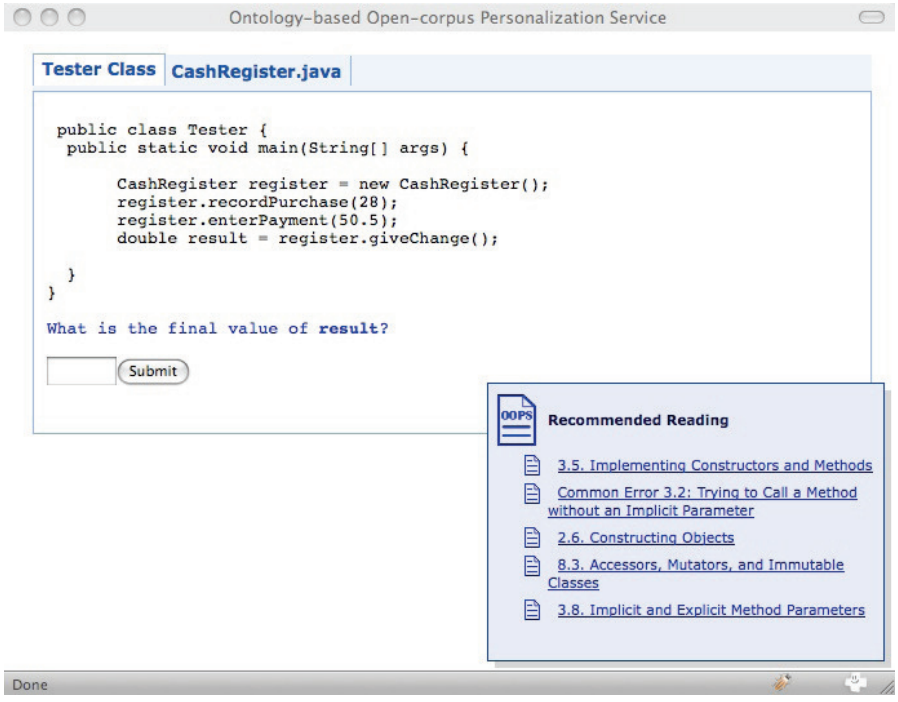


Figure 6. **OOPS: list of recommendations**

If a student decides that a particular reading can help her to solve the question, she can click on the link and open the window with this reading. Figure 7 demonstrates the result of a click on the first link form the previous list of recommended pages. Once the student has finished reading the page she can collapse its window by clicking one of two closing buttons ("This was USELESS" and "This was USEFUL"). The two exit buttons allow students to report their opinion about the quality of recommendations and allow the system to collect these reports for future data processing. After closing the reading window, OOPS shows the list of recommendations again.

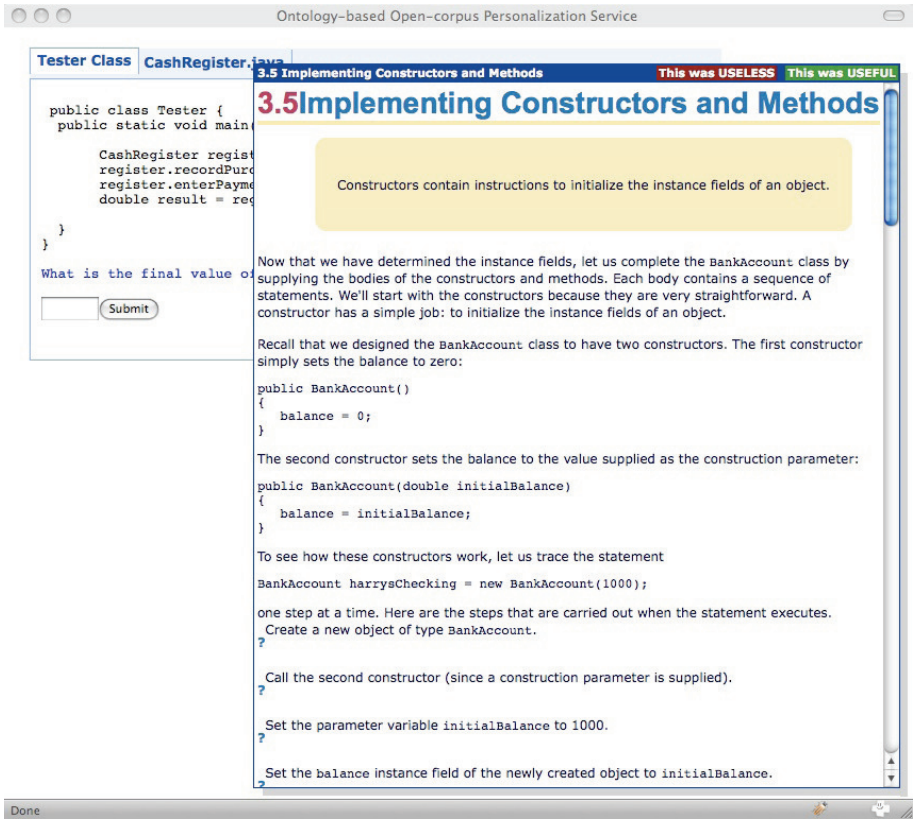


Figure 7. **OOPS: recommended page**

5. Discussion

This project addresses the problem of open-corpus content-based personalization in the context of e-Learning and propose a solution based on ontology mapping technologies and minimalist topic-based approach to adaptation. This work is built upon the results of our previous studies that investigated the pros and cons of coarse-grained adaptation and user modeling (P. Brusilovsky, et al., 2006; S. Sosnovsky & Brusilovsky, 2005) and another work evaluating the applicability of automatic ontology mapping for meditation between overlay user models (S. Sosnovsky, et al., 2007). The results of these experiments allow us to expect that the approach we propose can support open-corpus personalization.

We plan to evaluate this approach with students studding introductory Java programming. Another version of the system is being developed for SQL domain, as well. Finally, we would like to incorporate social navigation techniques into the OOPS service, which should help to filter out low-quality content and improve the results of pure-mapping-based recommendation.

References

Angeletou, S., Sabou, M., Specia, L., & Motta, E. (2007). Bridging the Gap Between Folksonomies and the Semantic Web: An Experience Report. In *Proceedings of Workshop on Bridging the Gap between Semantic Web and Web 2.0 at ESWC'2007*.

Bechhofer, S., Goble, C., Carr, L., Kampa, S., Hall, W., & De Roure, D. (2003). COHSE: Conceptual Open Hypermedia Service. In S. Handschuh & S. Staab (Eds.), *Annotation for the Semantic Web* (pp. 193-211): IOS Press.

Billsus, D., & Pazzani, M. J. (Year). A hybrid user model for news story classification. In *Proceedings of 7th International Conference on User Modeling (UM'1999), Banff, Canada* (pp. 99-108).

Brusilovsky, P., Chavan, G., & Farzan, R. (Year). Social adaptive navigation support for open corpus electronic textbooks. In P. D. B. a. W. Nejdl (ed.), *Proceedings of Third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2004), Eindhoven, the Netherlands, August 23-26* (pp. 24-33).

Brusilovsky, P., & Henze, N. (2007). Open Corpus Adaptive Educational Hypermedia. In P. Brusilovsky, A. Kobsa & W. Nejdl (Eds.), *The Adaptive Web: Methods and Strategies of Web Personalization* (pp. 671-696). Heidelberg, Germany: Springer Verlag.

Brusilovsky, P., Sosnovsky, S., & Shcherbinina, O. (2005). User Modeling in a Distributed E-Learning Architecture. In L. Ardissono, P. Brna & M. A. (eds.), *Proceedings of 10th International Conference on User Modeling (UM'2001), Edinburgh, UK* (pp. 387-391).

Brusilovsky, P., Sosnovsky, S., & Yudelson, M. (2006). Addictive links: The motivational value of adaptive link annotation in educational hypermedia. In V. Wade, H. Ashman & B. Smyth (eds.), *Proceedings of 4th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2006), Dublin, Ireland* (pp. 51-60).

Brusilovsky, P., Sosnovsky, S., Yudelson, M., & Chavan, G. (2005). Interactive Authoring Support for Adaptive Educational Systems. In C.-K. Looi, G. McCalla, B. B. & B. J. (eds.), *Proceedings of 12th International Conference on Artificial Intelligence in Education, AIED'2005, Amsterdam, Netherlands* (pp. 96-103).

Carr, L., DeRoure, D., Hall, W., & Hill, G. (1995). The distributed link service: A tool for publishers, authors and readers. *World Wide Web Journal 1*(1), 647-656.

Chen, L., & Sycara, K. (Year). A Personal Agent for Browsing and Searching. In *Proceedings of 2nd International Conference on Autonomous Agents, Minneapolis/St. Paul, May 9-13* (pp. 132-139).

Domingue, J., Dzbor, M., & Motta, E. (Year). Magpie: supporting browsing and navigation on the semantic web. In C. Rich & N. J. Nunes (eds.), *Proceedings of 9th International Conference on Intelligent User Interfaces (IUI '04), Funchal, Madeira, Portugal, January 13-16* (pp. 191-197).

Dzbor, M., & Motta, E. (Year). Semantic Web Technologies to Support Learning about the Semantic Web. In R. Luckin, K. R. Koedinger & J. Greer (eds.), *Proceedings of 13th International Conference on Artificial Intelligence in Education (AIED'2007), Marina Del Ray, CA, USA, July 9-13* (pp. 25-32).

Heath, T., & Motta, E. (2008). Revyu : Linking reviews and ratings into the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web 6*(4), 266-273.

Hendler, J. (2007). The dark side of the semantic web. *IEEE Intel ligent Systems 22*(1), 2-4.

Henze, N., & Nejdl, W. (2001). Adaptation in open corpus hypermedia. *International Journal of Artificial Intelligence in Education 12*(4), 325-350.

Henze, N., & Nejdl, W. (Year). Knowledge Modeling for Open Adaptive Hypermedia. In P. De Bra, P. Brusilovsky & R. Conejo (eds.), *Proceedings of 2nd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2002), Malaga, Spain, May, 2002* (pp. 174-183).

Horstmann, C. (2007). *Big Java* (3d ed.): John Wiley and Sons, Inc.

Hsiao, S., Brusilovsky, P., & Sosnovsky, S. (2008). Web-based Parameterized Questions for Object-Oriented Programming. In *Proceedings of E-Learn 2008, Las Vegas, NV, USA* (pp. submitted).

Kalfoglou, Y., Domingue, J., Motta, E., Vargas-Vera, M., & Shum, S. B. (2001). MyPlanet: an ontology-driven Web-based personalised news service. In *Proceedings of Workshop on Ontologies and Information Sharing at IJCAI'01, Seatlle, WA, USA*(pp. 140-148) (Available at: http://www.informatik.uni-bremen.de/agki/www/buster/IJCAIwp/Finals/kalfoglou.pdf).

Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). GroupLens: applying collaborative filtering to Usenet news. *Communications of the ACM 40*(3), 77-87.

Lieberman, H. (Year). Letizia: An agent that assists web browsing. In *Proceedings of International Joint Conference on Artificial Intelligence, Montreal, Canada, August* (pp. 924–929).

Middleton, S. E., De Roure, D. C., & Shadbolt, N. R. (Year). Capturing knowledge of user preferences: ontologies in recommender systems. In *Proceedings of 1st International Conference on Knowledge Capture (K-CAP '01), Victoria, British Columbia, Canada* (pp. 100-107).

Middleton, S. E., Shadbolt, N. R., & De Roure, D. C. (Year). Capturing interest through inference and visualization: ontological user profiling in recommender systems. In *Proceedings of 2nd International Conference on Knowledge Capture (K-CAP '03), Sanibel Island, FL, USA* (pp. 62-69).

Murray, T. (1999). Authoring Intelligent Tutoring Systems: An Analysis of the State of the Art. *International Journal of Artificial Intelligence in Education 10*, 98-129.

Sosnovsky, S., & Brusilovsky, P. (2005). Layered Evaluation of Topic-Based Adaptation to Student Knowledge. In *Proceedings of Workshop on Evaluation of Adaptive Systems at UM'2005, Eduburgh, UK*(pp. 47-56) (Available at: http://www.easy-hub.org/hub/workshops/um2005/doc/Sosnovsky,%20and%20Brusilovsky.pdf).

Sosnovsky, S., Dolog, P., Henze, N., Brusilovsky, P., & Nejdl, W. (2007). Translation of Overlay Models of Student Knowledge for Relative Domains Based on Domain Ontology Mapping. In R. Luckin, K. R. Koedinger & J. Greer (eds.), *Proceedings of 13th International Conference on Artificial Intelligence in Education (AIED'2007), Marina Del Ray, CA, USA, July 9-13* (pp. 289-296).

Van der Sluijs, K., & Houben, G.-J. (2009). Community Based Disclosure of Multimedia Datasets Using the Semantic Web. In S. Berkovsky, F. Carmagnola, D. Heckmann & T. Kuflik (Eds.), *Ubiquitous User Modeling* (pp. to appear): Springer.

Yesilada, Y., Bechhofer, S., & Horan, B. (2007). *COHSE: Dynamic Linking of Web Resources* (No. TR-2007-167): Sun Microsystems.

Ontology-based Support for Designing Inquiry Learning Scenarios

Stefan WEINBRENNER[a,1], Jan ENGLER[a], Lars BOLLEN[b] and Ulrich HOPPE[a]

[a]*COLLIDE Research Group, University of Duisburg-Essen*
47047 Duisburg, Germany

[b]*University of Twente, PO Box 217,*
7500 AE Enschede, The Netherlands

Abstract. Ontology support has been incorporated into a design environment for inquiry learning scenarios in the context of an ongoing European project. The underlying ontology is being elaborated as an interdisciplinary agreement on basic concepts, their relationships and attributes between designers and developers. The ontology encapsulates concepts such as "pedagogical plan", "scenario", "learning activity", or "tool" together with compatibility relations. This forms the semantic basis of a cooperative graphical editing environment. Similarity measures addressing structural aspects (graph matching) as well as semantic similarities allow for generating social recommendations in the SCY[2] design community.

Keywords. Learning Design, Pedagogical Scenarios, Ontologies.

1. Introduction

Although often disregarded, there is a common agreement in the TEL community that ontologies can enhance and enrich learning environments, e.g. in the form of compentence ontologies used for learner profiling [1] or for building a framework for instructional and learning theories [2]. Our approach uses semantic representations (in the form of an ontology) as an internal resource in a specific learning environment that is currently being developed in the new European research project SCY. In this sense, our approach differs from a general "semantic web" approach (cf. [3]), which would extend to a larger and open collection of web resources. From an educational perspective, an obvious added value of using an ontology is to define a set of terms and their relations and then use these as a shared vocabulary. This shared vocabulary can then serve as a common verbal denominator for a development and design group to synchronise and co-align their wording and conceptual perspectives. By doing that an ontology can help to avoid misunderstandings and to improve the precision of communication in general. In addition to the provision of a shared vocabulary, an

[1] Corresponding author.

[2] SCY – "Science created by You" is an EU project of the 7th Framework Programme. For more information, see http://www.scy-net.eu (last visited in April 2009).

ontology may help inferring semantic relations between objects and thus may indicate recommendations or possible flaws in a certain domain.

In this paper we elaborate on the communication and conceptualisation perspective by introducing specific system support through providing ontology-aware tools for educational designers. In this sense, our target situation is very similar to the one of the SMARTIES scenario editing environment based on the pedagogical ontology OMNIBUS [4]. However, in our approach there is no given equivalent of OMNIBUS as a general basis of pedagogical concepts and patterns, which are used as generic or foundational concepts to underpin specific designs. Instead, the SCY ontology with different sections addressing domain knowledge, pedagogical and task concepts is being built up according to the needs and specific choices (e.g., example domains) of the project, specifically oriented towards inquiry style learning. Also, we do not want to encompass a large variety of pedagogical approaches to be modelled in detail but concentrate on SCY specific methods of teaching and learning. Rather than defining these in very flexible, generic form (as in OMNIBUS), we aim at practically oriented blueprints without capturing "pedagogical causality".

In general, ontologies can capture semantics (concepts) from domains of interest as well as methodological and task knowledge and support reflection and self-description in the context of the system environment. As for our SCY project, we see the following specific roles of ontologies:

- Represent terminological/semantic basis for human-human interdisciplinary exchange;
- Serve as meta-level description for storage structures (repositories);
- Standardise and interlink metadata vocabulary (e.g., for the learning objects created by the students, domain and student models);
- Provide basic concepts for the definition of agent behaviour on a semantic level;
- Facilitate high level inter-operability between tools and services.

It has turned out turns that building up an ontology for the SCY project both enforced and established an agreement on the basic terminology in terms of concepts, their relationships, and attributes between designers and developers from different academic disciplines. The ontology encapsulates the agreements on basic terms such as "mission", "pedagogical plan", "scenario", "learning activity", "tool", or "scaffold" together with lexical enumerations of certain elements (activities, tools, types of learning objects) and their compatibility relations. This forms the semantic basis of the editing environment. Similarity measures addressing structural aspects (graph matching) as well as semantic similarities allow for generating social recommendations in the SCY design community.

2. Educational Design for CSCL Scenarios

Context and requirements for our developments originate from the SCY project, which aims at delivering a flexible, open-ended learning environment for science education and scientific discovery learning. SCY integrates a number of different tools, background material, collaborative activities, etc. to form "missions", which are then accomplished by groups of learners in an environment called "SCY-Lab". SCY-Lab provides adaptive support and pedagogical scaffolds to help learners on their missions. One central aspect in SCY are "emerging learning objects" (ELOs) [5], which denote

re-usable, sharable products of learning activities, created by students themselves, rather than traditional learning objects, which are typically used in the sense of pre-fabricated learning material.

At the very beginning of SCY, it became obvious that the expected, complex interplay of tools, activities, scaffolds and ELOs needs to have an external representation for discussion, authoring, exchange, and potentially to be machine-readable and machine-interpretable. A number of languages and tools for the design and storage of learning processes are available, like IMS-LD[3], LAMS [6] and COML [7], each with particular strengths and weaknesses. For example, IMS-LD strongly builds on XML representations, which are hardly comprehensible for non-experts. LAMS appears as an combined authoring-player suite that makes it hard to integrate third party tools and COML focuses on the use and integration of hardware like mobile devices or shared displays. However, one strong deficit of all these approaches with respect to the SCY requirements is the lack of support for emerging learning objects.

These shortcomings called for the development of a comprehensible, flexible and pedagogical neutral learning design language. As a result, the concept and graphical representation of so-called Learning Activity Spaces ("LAS") emerged, which is described in detail in [8]. A LAS is defined as a coherent and intuitive set of activities supported with specific tools and scaffolds. The input and output of a LAS are described in terms of a set of artefacts usually created by students (i.e. the ELOs).

Based on these premises, a graph-based modelling tool has been developed to support the creation and exchange of LASs. This tool, called "SCY Scenario Editor" (or "SCY-SE"), builds on the graph-based modelling tool FreeStyler and enhances it through the use of ontologies. To be flexible and to provide a high degree of representational freedom, but still be able to interpret and intelligently support a designer, the SCY-SE tool relies on an ontology in the backend. The ontology contains a representation of available tools, activities, types of ELOs, scaffolds and the relations between those entities and is able to provide information to identify possible LAS elements, to give recommendations and to discover discrepancies in the LAS design.

In the SCY project, the same ontology is planned to be used for other purposes, too, like query extension when searching for ELOs, identifying applicable tools for an ongoing learning activity or for deciding on an appropriate scaffold for the current situation. However, this paper deals with the usage of ontologies in a graph-based learning design editor. The following chapters will elaborate on technical requirements and use cases of ontology-based models for pedagogical scenarios.

3. Implementation Platforms

3.1. Blackboard Architecture based on Tuple Spaces

Our approach uses a blackboard architecture for the backend, both as a container for the evolving ontologies as well as a communication medium for additional agents. Such a blackboard system comprises a central server with several clients that interact with each other only indirectly by exchanging information through the blackboard [9]. Following a distributed problem-solving paradigm, the blackboard approach is very

[3] For the full specification and more information in IMS-LD, see http://www.imsglobal.org/learningdesign (last visited in April 2000).

appropriate in situations with multiple clients or agents that have quite different functionalities and differing implementation styles. In these situations, in which a problem solving strategy is divided into several steps with each step covered by one agent, there is often an implicit order of processing a query, but without having tightly connected components, i.e. without any explicit, external ordering. This corresponds to a "loose coupling" approach, which essentially means that all participating components were designed using minimal assumptions on the knowledge and awareness of each other. These minimal assumptions lead to a high robustness and failure tolerance and make it easy for designers to flexibly adapt the architecture.

TupleSpaces, as introduced by [10] together with the Linda language for distributed processing, are a well known means to implement a blackboard architecture. Here, a central server acts as a relay station for exchanging messages. These messages are stored in the form of tuples, i.e. in ordered sequences of typed literals. The most prominent recent implementations are JavaSpaces (Sun Microsystems) and TSpaces by IBM [11]. Our ontology backend, however, is built on top our own implementation called SQLSpaces [9]. Since the SQLSpaces were developed on top of a relational database, they have an inherent support for persistent and efficient data storage and provide transactions. Additionally to typical convenience features of other TupleSpaces implementations like asynchronous notifications (callbacks), automatic expiration of tuples, blocking and non-blocking commands, bounded queries, a graphical web-interface and a user management, the SQLSpaces have other, new features like multi-language support (currently supported languages are Java, C#, Ruby, Prolog and PHP), wildcard fields, space-based rights management and versioning.

3.2. SWAT

The SQLSpaces are the technical foundation for the actual ontology management facility called SWAT (Semantic Web Application Toolkit). It has been developed by the Collide Research Group in the Ontoverse project [12] that aimed at developing a web-based platform for collaborative ontology creation in the Life Sciences. SWAT is a framework that uses the SQLSpaces as a platform for a multi-agent system that encapsulates all the functionality in several agents coordinated via the SQLSpaces platform. Therefore, it uses several spaces inside the SQLSpaces server as disjoint communication channels.

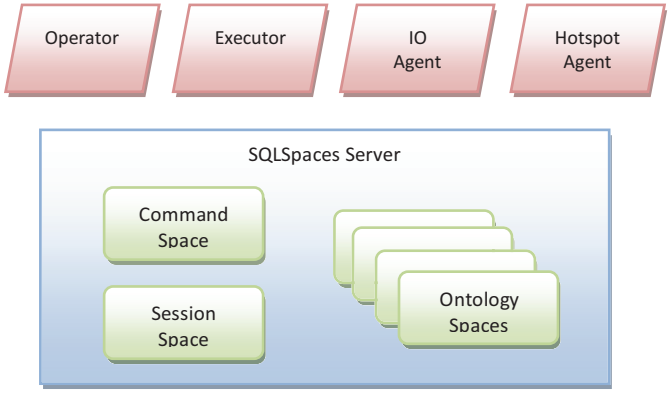


Figure 1. SWAT components

A simplified overview of the SWAT architecture is shown in 1. At the top of the picture some of the agents of SWAT are shown, namely the operator (for supervising other agents), the executor (for starting and restarting other agents), the IO agent (for importing and exporting OWL files) and the Hotspot agent (for calculating modification statistics during the evolution of an ontology). As described in section 5, this paper will explain the design and implementation of another functional agent that is capable of calculating similarity of certain entities on the base of an ontology. At the bottom of Figure 1, the SQLSpaces server is shown that holds all the spaces for SWAT. The command space is the central coordination space for all agents. The session space acts as a kind of logging and coordination space for a user. Finally, for each ontology there is one space in which the whole ontology is stored in the form of RDF triples.

3.3. FreeStyler

FreeStyler [13] is a collaborative modelling environment based on the idea of shared workspaces offering freestyle-sketching as well as different visual languages. FreeStyler can be used in collaborative as well as in single-user scenarios . The support for different representations is very flexible, as it based on a plugin-framework for graph-based visual languages that can be implemented as plugins through a common interface or API. This API was also used to implement the user frontend of the scenario-editor. Figure 2 shows an example scenario design, which consists of several ELOs (green diamonds) and LASs with substructures (blue boxes) thus illustrating the two levels of representation (macro level: senarios / micro level: LASs).

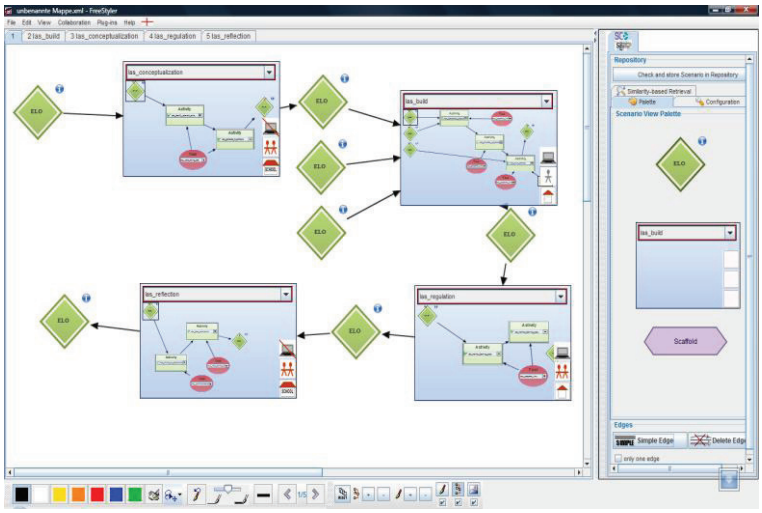


Figure 2. SCY-SE as FreeStyler plugin.

4. Usage Scenarios

The ontology serves as a basis to implement four groups of user support features that all serve different purposes and that all affect different phases of the creation process.

In addition, the way of using the ontology differs from group to group. This will be explained in the following.

4.1. Possible Values

As described in section 2, the graphical modelling language uses the notion of a LAS, which forms the low level structure consisting of atomic building blocks like tools and activities. Furthermore, there are concepts like LAS types, activity types, concrete tools and ELO types. For each of these concepts there is a limited number of possible values to be chosen from a designer. However, it would be quite inflexible to have these lists of possible values hard coded to the application. Therefore, these lists are fetched from the ontology, where they are represented as instances of the corresponding concepts. Using the ontology as a knowledge store for all possible values of certain assignments makes it easy to extend the lists without changing the actual application code and it gives the possibility of changing the values at runtime.

4.2. Recommendations

Another way of using the ontology is to make specific recommendations of values according to the context. In the SCY project it has been clearly defined what activities are useful and adequate in which kinds of LASs. However, these relations between LASs and activities are not mandatory, so it is also possible to create LASs comprising not recommended activities. The recommendations are shown in SCY-SE as green “checks” (meaning “everything is fine”) or yellow warning symbols (meaning “this is not recommended, though might be ok”) in the activities drop-down box and are of course context-sensitive depending on the LAS this activity is located in. Similar recommendations can relate to the formats a tool supports (not yet implemented), so that the application itself would warn a user who tries to consume or produce an ELO with a tool that is not capable of processing this type of data.

4.3. Warnings

SCY-SE can check for several types of constraints and constraint violations in given designs, some of which are ontology based. The simpler, obvious ones apply to the way nodes are connected. Considering the notion of input and output ELOs, it is obvious that these nodes have a natural direction that edges to them should follow. Therefore, an output ELO needs some activity that produces it and an input ELO needs some activity that consumes it. In addition, it does not make any sense for an activity to produce an input ELO, because this ELO is the output ELO of some other LAS, where it has already been produced etc. Moreover, ELOs can have more or less precisely defined types. These types also have a relation to the tools (that support activities), because the tools need to be compatible with this type. Constraints like the directions of connections between ELOs and activities are quite simple and do not need any background domain knowledge in an ontology, whereas the compatibility of formats and tools are obviously well stored in ontological structures.

4.4. Similarity

It is of course possible to save learning designs that were created with SCY-SE as XML files. However, it might also be an advantage to have a central server, where all users can store and exchange their results. Having such a central repository has mainly two purposes: On the one hand, it might support the designer in finding similar scenarios. Access to similar scenarios may help to improve the currently developped design by comparison and reuse. On the other hand, the list of similar scenarios can be seen as a mediating object between several designers. This retrieval mechanism uses some basic mathematical principles such as distance measure and grpah matching techniques as well as relations from the ontology (for determining what is more related and similar) in the calculation of similarity.

5. Architecture and Implementation

SCY-SE has been implemented as a plugin for FreeStyler and provides the means to create pedagogical scenarios based on a specific graphical language developed within the SCY project (see section 2). To implement the functionalities described in the section 4 this plugin utilises the SWAT framework (see section 3.2). The ensuing overall architecture is outlined in Figure 3.

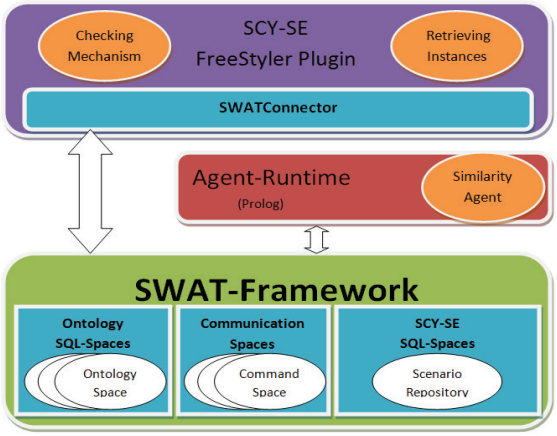


Figure 3. A simplified overview of the SCY-SE architecture.

The FreeStyler plugin for SCY-SE on the top is connected to the SWAT framework through a facility called SWAT Connector. This component encapsulates the communication to the underlying backend and allows easy extension for later work. In the lower part of the figure the SWAT framework with all its spaces is shown. Besides the SWAT-typical spaces already mentioned, there is an additional space for SCY-SE. As the name implies the scenario repository space represents a global storage for all scenario designers to save and share their results. In the center of the figure, the agent runtime component is shown. We used SWI Prolog [4] to implement this

[4] For more information, see http://www.swi-prolog.org

component whose main purpose is currently to provide the similarity calculation. Following the strategy of loose coupling, the ensuing messages between this component and SCY-SE are all passed via the command space of SWAT.

The exact use cases within the scenario design process, where the ontology access is being used, according to the general ideas described in section 4, are the following:

- *Startup*: fetch instances of LASs, activities, tools, ELO types
- *LAS editing*: compatibility between LAS and activity
- *Save*: checking based on possible relations between the ontology entities
- *Retrieval*: similarity-based search for other scenarios

Technically speaking the first point is implemted as a query to the *instance-of* relation of the corresponding ontology concept, whereas the second and the third point use certain object properties of the concepts "LAS" and "Tool". Since the fourth point relies on the functionality of the Prolog agent, which is in this use case the component that accesses the ontology store, the communication scheme and the similarity calculation algorithm is described first.

The overall process of calculating the similarity is shown in Figure 4. In the first step SCY-SE constructs lists of the relevant entities of the scenarios to be compared. These relevant entities consist of the three list types ELOs, tools and activities, which are grouped by the corresponding LASs of the two scenarios. These three lists again can be divided into two categories: Activities and tools are atomic values, so that lists of them can be compared directly through the "weighted symmetric distance" of two sets. This distance is based on the cardinality of the symmetric difference, which is then divided by the cardinality of the union of the two sets A and B:

$$SD_{A,B} = \frac{\#((A \cup B) - (A \cap B))}{\#(A \cup B)}$$

However, in our case this distance calculation is not a plain binary comparison, but it has been improved by weights based on the ontological proximity between the elements. The resulting similarity value ranges between 0 and 1.

ELOs, on the other hand, are not suited for a plain comparison as mentioned above. In fact, they are built up from four properties and therefore have an intrinsic complexity that requires a different approach for calculating the similarity. At first glance, it is possible to calculate the similarity of two ELOs A and B by comparing the four properties, so if two ELOs share three of their four properties, they would have a similarity of 0.75. Nevertheless, the comparison of A and C might result in an even higher similarity. So, each ELO of the first list needs to be compared with each of the second one. The outcome of this step is a matrix of similarity values. To condense this matrix to one single similarity value that expresses the similarity of these two lists, we decided to use the so-called "Hungarian method" for graph matching [14] to solve this assignment problem (of which ELO is identified with which other ELO).

As shown in Figure 4, the calculation inside the Prolog agent is done in several steps. After the lists have been written to the SQLSpaces server, a "start comparing" tuple is written to inform the agent about a new query. Then the agent fetches the lists and calculates the weighted symmetric distance for each of the lists. At that point the matrix for the ELOs is already calculated by using the "Hungarian method". The next step uses the merged matrix of these three matrices. The merging process is simply done by adding and normalising the three single matrices. The resulting matrix represents the similarity of the LASs of the two scenarios. The Hungarian method is

then used again to determine the best mapping between the LASs of the two scenario with respect to the similarity encoded in the matrix. The Hungarian method was chosen here since it is quite standardised can guarantee the optimality of the result. The method may maytake several iterations, but will eventually terminate and return the optimal solution. From the resulting mapping, a single value can be easily calculated by normalising the selected matrix entries. This value is then returned to the already waiting SCY-SE instance that presents the similarity in a user-oriented way.

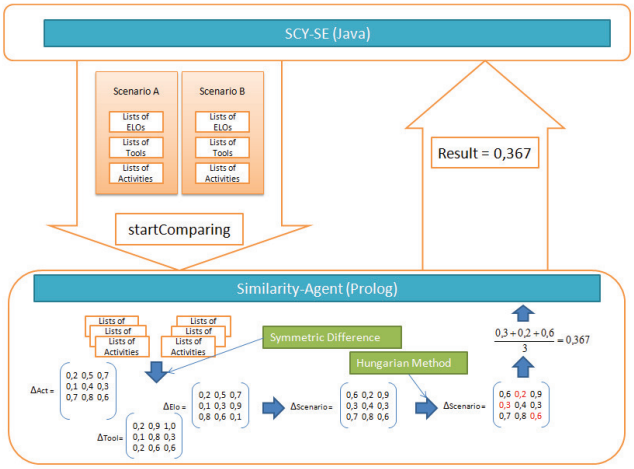


Figure 4. The sequence of actions in the similarity agent.

Using ontologies in pedagogical learning design in the use cases mentioned above, allows us to increase flexibility and correctness of the learning scenario. At the same time, this approach provides a higher precision according to the similarity measurement.

6. Next Steps

In the work presented so far the modelling tool FreeStyler has been extended with a plugin to describe abstract learning scenarios in the context of the SCY project. The possibility to describe these scenarios leads to a next step: Instantiating those abstract "pedagogical plans" towards concrete task-related "missions". The designed scenarios can be a basis for these missions, which can be enriched by domain specific content. By using the ontology as a common repository for the elements of the scenarios and by the ontology-based checking mechanism, it is guaranteed that the designed scenarios are consistent with the core concepts of the SCY project.

The similarity measurement is currently based on the comparison of activities, tools and ELOs as described above. To improve this functionality, it will be useful to include resources and scaffolds, which are also elements of a scenario (already available in the SCY-SE FreeStyler palette).

Currently, according to the definition of the scenario modelling language, ELOs can only appear as inut or output objects to LASs. In addition to this we have considered the provision of "intermediate ELOs". These special ELOs would be used to connect activities inside a LAS. A potential implementation could make use of the

possibility to create custom edges in FreeStyler. These edges could then be enriched with operational semantics, so that the intermediate ELOs could serve as anchor points for further features. An interesting support for the designers of scenarios may be a repository of predefined LASs. These "standard LASs" would be LASs that contain a typical set of activities and tools, so that designers are able to start designing a scenario from these generic LASs.

In the present implementation, it is not possible to create custom LAS types or other elements and save them to the ontology. Yet, the plugin only reads from the ontology, but one can think of an "ontology modification mode" through which the set of elements stored in the ontology can be extended. To support this the SWAT framework already provides some easy-to-use functionalities to manipulate to underlying ontology.

References

[1] Kalz, M., J.v. Bruggen, E. Rusman, B. Giesbers, and R. Koper, Positioning of Learners in Learning Networks with Content-Analysis, Metadata and Ontologies. *Interactive Learning Environments*, 15.2007.

[2] Hayashi, Y., J. Bourdeau, and R. Mizoguchi. Ontological Support for a Theory-Eclectic Approach to Instructional and Learning Design. in *Proceedings of the European Conference on Technology Enhanced Learning (EC-TEL 2006)*. 2006. Crete, Greece.

[3] Devedžić, V., *Semantic Web and Education (Integrated Series in Information Systems)*. 2006: Springer-Verlag New York, Inc.

[4] Hayashi, Y., J. Bourdeau, and R. Mizoguchi. Toward Establishing an Ontological Structure for the Accumulation of Learning/Instructional Design Knowledge. in *Proc. of 6th International Workshop on Ontologies and Semantic Web for E-Learning (SWEL '08)*. 2008. Montreal, Canada.

[5] Hoppe, H.U., N. Pinkwart, M. Oelinger, S. Zeini, F. Verdejo, B. Barros, and J.L. Mayorga. Building Bridges within Learning Communities through Ontologies and Thematic Objects. in *Proc. of the International Conference on Computer Supported Collaborative Learning (CSCL 2005)*. 2005. Taipei, Taiwan.

[6] Dalziel, J.R. Implementing Learning Design: The Learning Activity Management System (LAMS). in *Interact, Integrate, Impact: Proceedings of the 20th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*. 2003. Adelaide, Australia.

[7] Niramitranon, J., M. Sharples, and C. Greenhalgh. COML (Classroom Orchestration Modelling Language) and Scenarios Designer: Toolsets to Facilitate Collaborative Learning in a One-to-One Technology Classroom. in *Proceedings of the International Conference on Computer in Education (ICCE2007)*. 2006. Hiroshima, Japan.

[8] Lejeune, A., M. Ney, A. Weinberger, M. Pedaste, L. Bollen, T. Hovardas, H.U. Hoppe, and T.d. Jong. Learning Activity Spaces: Towards flexibility in learning design? in *Proc. of the 9th IEEE International Conference on Advanced Learning Technologies (ICALT 2009)*. 2009. Riga, Latvia.

[9] Weinbrenner, S., A. Giemza, and H.U. Hoppe. Engineering heterogenous distributed learning environments using TupleSpaces as an architectural platform. in *Proc. of the 7th IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*. 2007. Los Alamitos (USA): IEEE Computer Society.

[10] Gelernter, D., Generative Communication in Linda. *Acm Transactions on Programming Languages and Systems*, 7(1).1985. p. 80-112.

[11] Wyckoff, P., S.W. McLaughry, T.J. Lehman, and D.A. Ford, T spaces. *IBM Syst. J.*, 37(3).1998. p. 454-474.

[12] Malzahn, N., S. Weinbrenner, P. Hüsken, J. Ziegler, and H.U. Hoppe. Collaborative Ontology Development - Distributed Architecture and Visualization. in *Proceedings of the German E-Science Conference*. 2007. Baden-Baden, Germany: Max Planck Digital Library.

[13] Hoppe, U. and K. Gassner. Integrating collaborative concept mapping tools with group memory and retrieval functions. in *Proceedings of the International Conference on Computer Supported Collaborative Learning (CSCL2002)*. 2002. Hillsdale (USA): Lawrence Erlbaum.

[14] Kuhn, H.W., The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2.1955. p. 83-97.

Model Driven Architecture: How to Re-model an E-learning Web-based System to Be Ready for the Semantic Web?

Leyla ZHUHADAR and Olfa NASRAOUI
Knowledge Discovery and Web Mining Lab
Department of Computer Engineering and Computer Science
University of Louisville, KY 40292, USA

Robert WYATT and Elizabeth ROMERO
Office of Distance Learning
Division of Extended Learning and Outreach
Western Kentucky University, KY 42101, USA

Abstract: HyperManyMedia is a search and browse information retrieval system that provides access to online learning materials. This system offers metadata/ontology-based searching mechanism that allows learners to enter a simple search term and then recommends them with metadata/semantically related terms (concepts/subconcepts). HyperManyMedia integrates several technological components, such as machine learning, clustering (text analysis), knowledge-based modeling (ontology), and graphical data mapping. The first part of this paper proposes an architecture for adding metadata: (1) Domain-knowledge extraction; (2) Parsing learning objects (Lectures) and adding the metadata; (3) Reconfiguring the E-learning platform; and (4) Encapsulating the metadata within the "HyperManyMedia" platform. The second part of this paper proposes an architecture for adding the semantic: (1) Semantic Representation (knowledge representation); (2) Algorithms, which are the core engine of this study; and (3) Personalization to deal with information filtering.

Keywords. knowledge engineering, ontology, semantic web, information retrieval, E-learning

Introduction

The association between machines and thinking has been investigated for sometime, mainly in the field of Artificial Intelligence, this association started when Charles Babbage built his Analytical Engine, followed by Konrad Zuse with his creative work on the general-purpose computer, and Alan Turing's with a *Machine Intelligence* [9]. In 1999, Tim Berners-Lee, with his invention of the Web,

lightened this dream again, with a prophecy of software agents that can act on behalf of human which can reason about data, a place where the Web being a machine-readable information whose meaning is well-defined by standards [3]. In 2002, a new infrastructure of the Web was defined to help this dream to come true. A new way of data representation, such as RDF [1], was adopted to provide a common format for expressing information about information (metadata). However, in 2006, Tim Berners-Lee stated that "despite the progress of his dream (the Semantic Web) a lot of challenges still exist: the reuse of information is limited, how to effectively query an unbounded Web of linked information repositories? how to align and map different data models? how to visualize and navigate the huge connected graph of information? [4]" This paper is not going to answer these questions, but it will put some spotlights on the importance of the "Model Driven Architecture;" and how we can re-model our system to be ready for the Semantic Web, how we can add Metadata to our resources to make them reachable, and how we can extract the "Ontology" of a specific domain to make it understandable from the semantic prospective.

Of course, we were facing many challenges in the efforts to accomplish a full vision of the semantics of our domain, but we obtained quite impressive results by using the standards (RDF and OWL), framework (Jena)[2] and platform (HyperManyMedia[3]).

1. Background and Related Work

The main research question guiding this work is whether it is feasible and beneficial to add the metadata and semantic representation to learning resources, while still being able to retrieve personalized learning resources that are satisfactory and effective for the learner. The main goals of "HyperManyMedia[4]" system are to –

- **Provide the learner with a metadata search engine** to overcome the limitations of searching for learning objects
- **Deliver a personalized semantic information retrieval system** to learners using ontologies
- **Be an E-learning open source repository**

This research has been implemented on a real platform called "HyperManyMedia" at Western Kentucky University. It takes into consideration the personalization aspect, and enabling learners to retrieve personalized information. Over all, the system, as shows Figure 1., provides learners with a multi-model interface:(1) Generic search, (2) Metadata-driven approach, and (3) Semantically enriched approach. The main idea of multi-model metadata/semantic driven approach is to

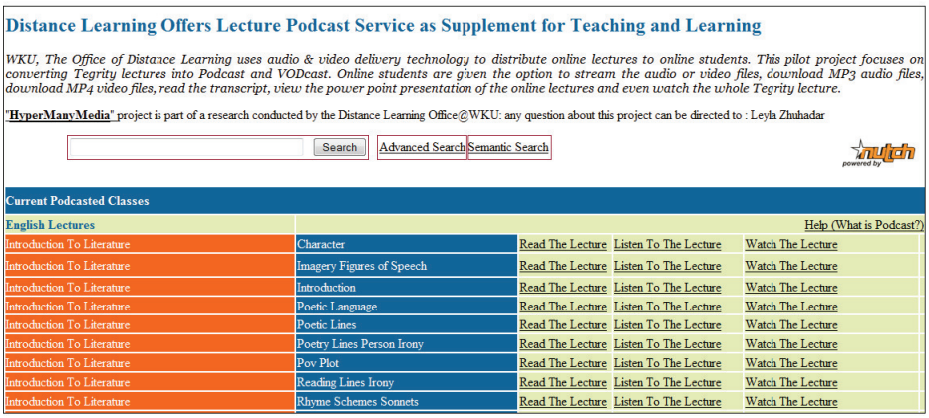
[1] Resource Description Framework

[2] http://www.w3.org/2001/sw/

[3] HyperManyMedia (http://hypermanymedia.wku.edu): We proposed this term to refer to any educational material on the web (hyper) in a format that could be a multimedia format (image, audio, video, podcast, vodcast) or a text format (webpage, powerpoint).

[4] http://hypermanymedia.wku.edu

Figure 1. HyperManyMedia Platform



provide the learners with possibilities that fit their searching style. A generic search engine model enables learners to search for learning materials based on keyword search. This model is similar to most generic search engines such as Google or Yahoo. Learners who are more adapted to this technique have the capability to use it. However, this approach has the following drawbacks:

- Searching for a specific college, course name, topic, media format is time consuming
- Searching for combinations of results is impossible (e.g., finding all video lectures in the college of business related to accounting)
- Low precision
- Inefficient ranking

Metadata enabled content repositories can provide learners with fast and accurate information retrieval. The Metadata model provides learners with high precision and effective ranking to learning materials. However, this approach has the following drawbacks:

- Limited in quantity, complexity and semantics
- No room for revision and validation
- Ambiguity
- One-way flow (on server-side)
- Can not be used for generating a learner profile

Potentially, the most significant model between the three is the personalized semantically enriched model. The advantages of this model can be summarized as follows:

- Learners were recommended with resources that had similar semantic meaning to the learning objects that they were looking for
- Learners were provided with resources similar to their profile
- Learners were provided with additional learning objects as recommendation based on similar concepts/subconcepts
- The Learners' profiles were updated based on their accumulated activities
- High precision and recall

The research field of Semantic E-learning covers a wide range of research problems. In Table 1, we cover some of the Semantic Web efforts in the context of E-learning.

Table 1. Semantic E-learning Research

An ontology-driven authoring tools' framework that the Educational platform can benefit from.	[2]
An ontology supported learning process enhances the activities between faculty and students in Web-based learning environments, and surveyed the relationship between Artificial Intelligence in Education (AIED) and Web Intelligence (WI).	[5]
A framework for ontology enabled annotation and knowledge management in collaborative learning environments. It provides semantic content retrieval with the personalization aspect expressed using ontologies.	[13]
An ontology-based document-driven memory in E-learning, that used two ontologies: a generic ontology related to the domain of training in general and a domain-specific ontology to deal with the application at hand.	[1]
An E-Learning web services architecture that can provide students with the following: registration, authentication, tutoring, question-answer queries services and an annotated feedback.	[10]
A framework for personalized E-learning in the Semantic Web where the hypertext structures were automatically composed using the semantic web services. The framework utilized reasoning rules to provide personalized hypertext relations between the domain and the learner.	[7]
An automated semantic annotation for multimedia learning objects in Educational platform.	[11]
The Unfolding of Learning Theories: Its Application to Effective Design of Collaborative Learning: It uses Ontologies and Semantic Web Services for Intelligent Distributed Educational System.	[8]
Using semi-automatic ontology extraction to create draft topic maps. It uses Topic maps to encode knowledge through the design and implementation of plug-in in TM4L editor.	[12]
This research exploits the concept of subject identity in learning content authoring, It uses Wikipedia articles as a source for (1) consensual naming, (2) and subject identifiers. This research is implemented in the Topic Maps for e-Learning tool (TM4L).	[6]

In our previous work we investigated different aspects of semantic web, some of them related to the personalization using clustering algorithms, also we studied user profiles based on situations where the domain and the user profile evolve, for more details refer to [17,14,15,16,18].

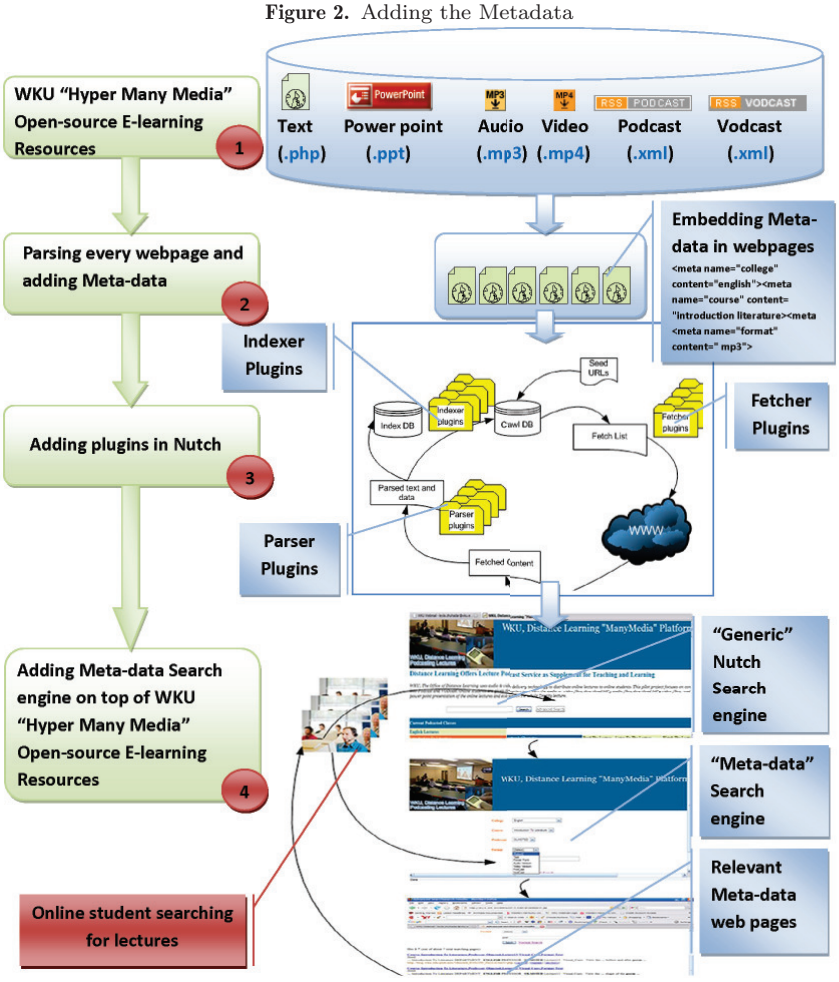
2. "HyperManyMedia" Architecture

The system architecture is divided into two phases. Phase 1: Adding the metadata, and Phase 2: Designing the E-learning domain ontology and adding the semantic.

2.1. Implementation of Phase 1: Adding the Metadata to each E-learning Resource

1. **Domain-knowledge Extraction**
   Each learning resource (lecture) is delivered in six different media formats:

Figure 2. Adding the Metadata



text, powerpoint, audio, video, podcast, and vodcast. Each lecture is a learning object used in online courses and taught by different professors. Information about each lecture was extracted and saved.

2. **Parsing learning objects (Lectures) and adding the metadata**
   All the resources (lectures) are located on the server were parsed using a java application. This application parses each webpage to find the specific location for the metadata. The following metadata information has been added: college name, course name, professor name, lecture name, media format type.
3. **Re-configuring the E-learning Platform**
   "HyperManyMedia" uses several refinements on VSM by extending the Boolean vector model and adding weights associated with terms and fields. The scoring algorithm is influenced by the sum of the score for each term of a query where each field is the product of the following factors: Its "tf", "idf", and index-time boosting. This score is computed as following:

$$Score(q,d) = coord(q,d) \times queryNorm(q) \times \sum(tf(tind) \times idf(t)^2 \times t.getBoost() \times norm(t,d))$$

To add a boosting measure to metadata, "HyperManyMedia" boosting algorithm was changed as following:

$$ModifiedBoost = \alpha\,(url) + \beta\,(anchor) + \gamma\,(content) + \delta\,(title) + \epsilon\,(meta-data)$$

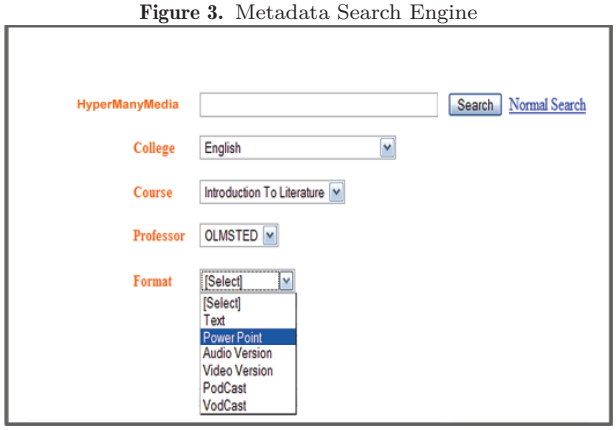
Table 2 shows the modified boosting weights after adding the metadata.

Table 2. Metadata Boosting Weights

Field	URL	Anchor	Content	Title	Metadata
Boost	4.0	2.0	1.0	1.0	5.0

4. **Encapsulating the Metadata within the "HyperManyMedia" Platform**

Figure 3. Metadata Search Engine

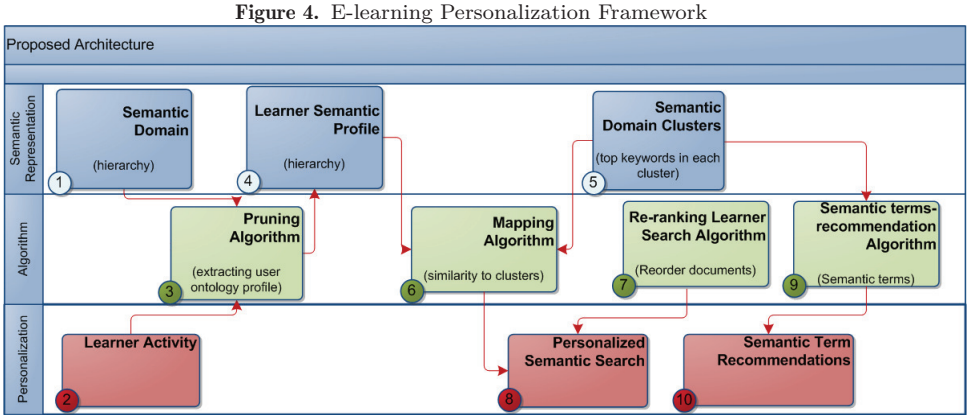


Since all lectures in "HyperManyMedia" platform were parsed and metadata were embedded, any search for a metadata will bring the related topic, see Figure. 3.

2.2. Implementation of Phase 2: Personalized Semantic Information Retrieval System

The architecture is divided into three layers as shown in Figure. 4 : (1) Semantic representation (knowledge representation), (2) Algorithms (core software), and (3) Personalization interface. The design of personalized semantic information retrieval system adds another dimension to the metadata search engine that can manage, and collect data that permits high levels of adaptability and relevance to learner's profiles. Within this context, the design framework consists of the following stages: (1) building the E-learning ontology using the "HyperManyMedia" colleges and courses as concepts and sub-concepts, (2) generating the semantic

Figure 4. E-learning Personalization Framework



learner's profiles (thus their *user models*) as an ontology from their navigation logs which record which lectures have been accessed, (3) re-ranking the learner's search results based on the matching between the learning content and the user profile, and (4) providing the learner with semantic recommendations during the search process.

2.2.1. Semantic Domain Structure

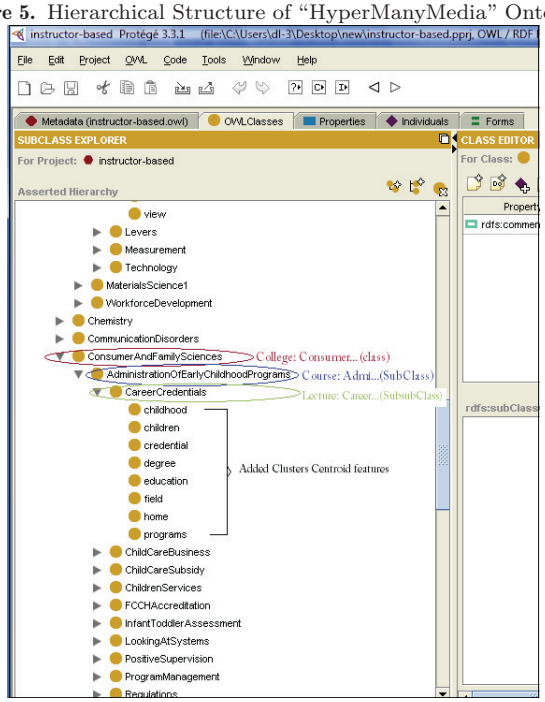
Let R, represents the root of the domain which is represented as a tree, and C_i represents a concept under R. In this case: R $= \cup_{i=1}^{n} C_i$, where n= Number of concepts in the domain, each concept C_i consists either of sub-concepts which can be children ($C_i = \cup_{j=1}^{m} SC_{ji}$) or of leaves which are the actual lecture documents ($\cup_{k=1}^{l} d_{ki}$). We encoded the above semantic information into a tree-structured domain ontology in OWL, based on the hierarchy of the E-learning resources. The root concepts are the colleges, while the subconcepts are the courses, and the leaves are the resources of the domain (lectures).

2.2.2. Adding the Ontology to "HyperManyMedia" Platform

Protégé[5] has been used as a framework application to design the "HyperManyMedia" ontology. Figure 5. shows part of "HyperManyMedia" ontology in Protégé. When a learner submit a query, the semantic search engine maps the query to the ontology file (maps to concept/subconcept that contains this query). In case that the concept/subconcept is found, the search engine retrieves the documents that contains that concept/subconcept.

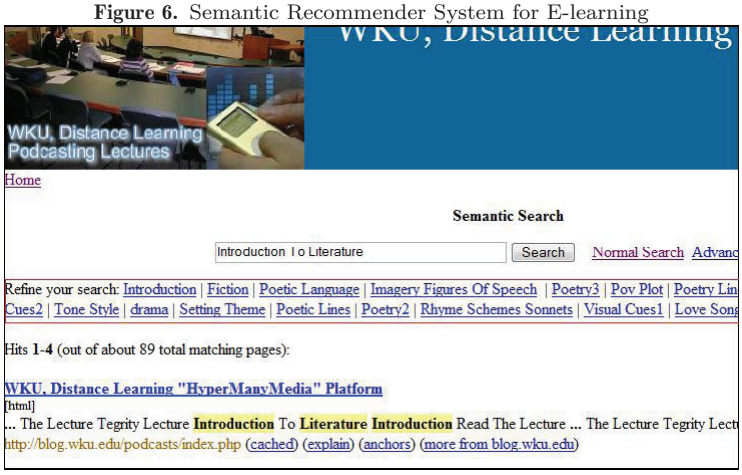
[5]http://protege.stanford.edu/

Figure 5. Hierarchical Structure of "HyperManyMedia" Ontology



2.2.3. Semantic Recommender System for E-learning

Figure 6. Semantic Recommender System for E-learning



For each query q submitted by a learner, a semantic mapping between the query and the learner's semantic profile brings all the concepts/subconcepts/ as recommendation. This framework allows the learner to navigate through the semantic structure of his/her query, as shown in Figure. 6, by possibly clicking on one of the recommended terms. The effect of this action is to add the selected term to

the query and repeat the search. Therefore the search is finally personalized via a query expansion using the recommended term that has been selected.

3. Conclusion

In this paper, we presented the design and implementation of a focused metadata/semantic information retrieval system that facilitates the search for learning objects in an E-learning domain. This research has been implemented on a real platform called "HyperManyMedia" at Western Kentucky University. It took into consideration the personalization aspect, and enabled learners with the retrieval of personalized information. The main purpose of multi-model metadata/semantic driven approach was to provide the learners with possibilities that fit their searching style. "HyperManyMedia" search engine has been ranked number 24 on "The Ultimate Guide to Using Open Courseware[6]," (between Cambridge University and Harvard Business). The next five are Princeton, Stanford, Yale, Johns Hopkins, and Boston College. Our future plan is to merge "HyperManyMedia" repository with as many external open source resources as we can accommodate, such as MIT OpenCourseWare [7] and Berkeley Webcast [8].

4. Acknowledgments

This work is partially supported by the National Science Foundation CAREER Award IIS-0133948 to Olfa Nasraoui.

References

[1] M.H. Abel, A. Benayache, D. Lenne, C. Moulin, C. Barry, and B. Chaput. Ontology-based Organizational Memory for e-learning. *Educational Technology & Society*, 7(4):98–111, 2004.

[2] L. Aroyo and D. Dicheva. The New Challenges for E-learning: The Educational Semantic Web. *Educational Technology & Society*, 7(4):59–69, 2004.

[3] T. Berners-Lee and M. Fischetti. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. Harper San Francisco, 1999.

[4] T. Berners-Lee, W. Hall, J. Hendler, N. Shadbolt, and D.J. Weitzner. COMPUTER SCIENCE: Enhanced: Creating a Science of the Web. *Science*, 313(5788):769–771, 2006.

[5] V. Devedžić. Web Intelligence and Artificial Intelligence in Education. *Educational Technology & Society*, 7(4):29–39, 2004.

[6] C. Dichev, D. Dicheva, and J. Fischer. Identity: How To Name It, How To Find It. In *Workshop on I3: Identity, Identifiers, Identification, 16th Intl Conference on World Wide Web, WWW 2007, May 8-12, 2007*, 2007.

[7] N. Henze, P. Dolog, and W. Nejdl. Reasoning and Ontologies for Personalized E-Learning in the Semantic Web. *Educational Technology & Society*, 7(4):82–97, 2004.

[8] S. Isotani and R. Mizoguchi. The Unfolding of Learning Theories: Its Application to Effective Design of Collaborative Learning. *SWELŠ07: Ontologies and Semantic Web Services for Intelligent Distributed Educational Systems*, page 44.

[6] http://www.collegedegree.com/library/college-life/the_ultimate_guide_to_using_open_courseware

[7] MIT OpenCourseWare: http://ocw.mit.edu/OcwWeb/web/home/home/index.htm

[8] Berkeley Webcast: http://webcast.berkeley.edu/

[9] P. McCorduck. *Machines Who Think: A Personal Inquiry Into the History and Prospects of Artificial Intelligence*. AK Peters, Ltd., 2004.

[10] E. Moreale and M. Vargas-Vera. Semantic Services in e-Learning: an Argumentation Case Study. *Educational Technology & Society*, 7(4):112–128, 2004.

[11] Stephan Repp, Serge Linckels, and Christoph Meinel. Towards to an automatic semantic annotation for multimedia learning objects. In *Emme '07: Proceedings of the international workshop on Educational multimedia and multimedia education*, pages 19–26, New York, NY, USA, 2007. ACM.

[12] S. Roberson and D. Dicheva. Semi-automatic ontology extraction to create draft topic maps. In *Proceedings of the 45th annual southeast regional conference*, pages 100–105. ACM New York, NY, USA, 2007.

[13] S.J.H. Yang, I.Y.L. Chen, and N.W.Y. Shao. Ontology Enabled Annotation and Knowledge Management for Collaborative Learning in Virtual Learning Community. *Educational Technology & Society*, 7(4):70–81, 2004.

[14] L. Zhuhadar and O. Nasraoui. Personalized cluster-based semantically enriched web search for e-learning. In *CIKM 08: ONISW The 2nd International workshop on Ontologies and Information Systems for the Semantic Web.*

[15] Leyla Zhuhadar and Olfa Nasraoui. Semantic information retrieval for personalized e-learning. *Tools with Artificial Intelligence, 2008. ICTAI '08. 20th IEEE International Conference on*, 1:364–368, Nov. 2008.

[16] Leyla Zhuhadar, Olfa Nasraoui, and Robert Wyatt. A comparsion study between generic and metadata search engines in an e-learning environment. In *IKE*, pages 500–505, 2008.

[17] Leyla Zhuhadar, Olfa Nasraoui, and Robert Wyatt. Metadata domain-knowledge driven search engine in "hypermanymedia" e-learning resources. In *CSTST '08: Proceedings of the 5th international conference on Soft computing as transdisciplinary science and technology*, pages 363–370, New York, NY, USA, 2008. ACM.

[18] Leyla Zhuhadar, Olfa Nasraoui, and Robert Wyatt. Dual representation of the semantic user profile for personalized web search in an evolving domain. In *Proceedings of the AAAI 2009 Spring Symposium on Social Semantic Web, Where Web 2.0 meets Web 3.0*, pages 84–89, 2009.

Using Ontologies for Modeling Educational Content

Vanessa Araujo BORGES and Ellen Francine BARBOSA
University of São Paulo (ICMC/USP) – São Carlos (SP), Brazil
{va.borges, francine}@icmc.usp.br

Abstract. Content modeling plays a fundamental role in the development process of educational modules.In spite of its relevance, there are few approaches for modeling educational content. Motivated by this scenario, in a previous work we proposed *IMA-CID* (Integrated Modeling Approach – Conceptual, Instructional, Didactic) – an integrated approach for modeling educational content.In this work we discuss the evolution of *IMA-CID* by exploring the use of ontologies at its conceptual level. The goal is to provide a better comprehension of the knowledge domain to be taught as well as to ease the knowledge sharing and reuse among authors. We illustrate our ideas by using an ontology of software testing for developing an educational module on this domain. The development of a supporting tool to help on the importation of ontologies and on the automated edition, interpretation and "execution" of the *IMA-CID* models is also discussed.

Keywords. Ontology, content modeling, educational modules, supporting tool.

1. Introduction

Several initiatives on using computing technologies have been investigated in order to facilitate the learning processes in general. The idea is to provide ways to establish quality educational products, capable of motivating the learners and effectively contribute to their knowledge construction processes in active learning environments.

Educational modules, which consist of concise units of study delivered to learners by using technologies and computational resources [1], can be explored in this perspective. Similar to software products, educational modules require the establishment of systematic development processes to produce reliable and quality products. In short, the development of such modules can involve developers from different domains, working on multi-disciplinary and heterogeneous teams, geographically dispersed or not. They should cooperate, sharing data and information regarding the project. Furthermore, there is a need for adaptability and reusability – educational modules should be seen as independent units of study, subject to be adaptable and reusable in different educational and training scenarios, according to parameters such as the learner's profile, instructor's preferences, learning goals, course length, among others.

Motivated by this scenario, in a previous work we proposed *IMA–CID* (Integrated Modeling Approach – Conceptual, Instructional and Didactic) [1] – an integrated approach for modeling educational content, composed by a set of models, each one considering specific aspects of the development of educational content.

In this work we intend to explore the use of ontologies [5] as a supporting mechanism for modeling the content of educational modules. The goal is to evolve *IMA–CID* by using ontologies at the conceptual level of the approach in order to provide a better

comprehension of the domain to be taught as well as to ease the knowledge sharing and reuse among authors/designers. We illustrate our ideas by using an ontology of software testing [3,2] for developing an educational module on this domain. The development of a supporting tool to help on the importation of ontologies and on the automated edition, interpretation and "execution" of the *IMA–CID* models is also discussed.

The remainder of this paper is organized as follows. In Section 2 we describe the main aspects of *IMA–CID*. In Section 3 we discuss how ontologies have been explored for evolving *IMA–CID*. In Section 4 we illustrate the application of our ideas into the development of an educational module for the software testing. In Section 5 we present an automated tool for modeling and generating educational content according to the new version of *IMA–CID*. Finally, conclusions and further work are presented in Section 6.

2. *IMA–CID*: An Integrated Approach for Modeling Educational Content

Content modeling plays a fundamental role in the development process of educational modules. It helps the author to determine the main concepts to be taught, providing a systematic way to structure the relevant parts of the domain [1]. Actually, how the content is structured impacts on the reusability, evolvability and adaptability of the module. Despite its relevance, there are few approaches for modeling educational content.

Motivated by this scenario, we proposed *IMA–CID* (Integrated Modeling Approach – Conceptual, Instructional and Didactic) [1] – an integrated approach for modeling educational content, composed by a set of models, each one considering specific aspects of the development of learning content. The *Conceptual Model* consists in a high-level description of the knowledge domain, representing its main concepts and the relationships among them. In order to construct the conceptual model, we focused on the conceptual mapping ideas [7]. The *Instructional Model* characterizes what kind of additional information (e.g., facts, principles, procedures, examples, and exercises) can be used to develop learning materials. The *Didactic Model* characterizes the prerequisites and sequences of presentation among conceptual and instructional elements.

We have also introduced the idea of *open specifications*, which provide support for the definition of dynamic contexts of learning. Depending on aspects such as audience, learning goals and course length, distinct ways for presenting and navigating through the same content can be required. An open specification allows to represent all sequences of presentation in the same didactic model. So, from a single model, several versions of the same content can be generated according to different pedagogical aspects.

3. Evolving *IMA–CID* Approach by Using Ontologies

An ontology is a formal explicit specification of a shared conceptualization [5]. That is, a simplified way of perceiving a piece of reality, often conceived as a set of relevant terms and their relationships, whose structure is constrained by some rules. Based on the principles and characteristics of ontologies, we are now interested in exploring them as a supporting mechanism for modeling educational content, as part of *IMA–CID*.

As a formal and declarative knowledge representation, an ontology includes [5,8]: (i) the vocabulary required for referring to the concepts in the domain; and (ii) the logical statements which describe what the concepts are and how they are related. Hence, it provides a vocabulary for representing and communicating knowledge about some topic as well as a set of relationships which hold among the concepts in that vocabulary.

Such definition matches with the goals of the conceptual modeling phase of IMA−CID. Based on this, we have extended IMA−CID to allow that both conceptual mapping and ontologies can be used for structuring and representing the knowledge domain. By using ontologies at the conceptual level of IMA−CID we intend: (1) to provide a better comprehension of the knowledge domain to be taught; (2) to ease the knowledge sharing among authors; (3) to provide a well-established structure for a knowledge repository; and (4) to provide support for interoperability, considering the relationship among different paradigms and languages. Notice that the use of ontologies (at the conceptual level) can also be explored together with the idea of open specifications (at the didactical level) aiming at providing knowledge reuse in different learning contexts.

We have also extended IMA−CID at the instructional level. In this case, we have adopted a specific ontology for establishing the media to be related to the information items and instructional elements. The *ALOCoM-Ontology* (*Abstract Learning Object Content Model - Ontology*) [9] establishes a formal representation for learning objects and their components. In short, it distinguishes three types of components [9]: content fragments, content objects and learning objects. To define the adequate media for the information itens and instructional elements we have explored the set of content fragments, which characterizes: (1) continuous elements (audio, video, simulatios and animations); and (2) discrete elements (texts, graphics, links and images).

It is worth to notice that the establishment of adequate media at the instructional level of IMA−CID, specially the continuous ones, is a relevant aspect for the development of interactive educational content, capable of motivating the learners and effectively contribute to their knowledge construction processes in active learning environments.

Moreover, the adopted representation is in agreement with the *ALOCoM framework*, which supports the use of XML schemas for importing and exporting the educational content for different models and specifications, such as *SCORM* (*Sharable Content Object Reference Model*)[1] and LOM (*Learning Object Metadata*) [6]. The standardization obtained from the use of *ALOCoM-Ontology* aims to guarantee interoperability, sharing and reuse to the educational content developed according to the IMA−CID approach.

4. An Educational Module for the Software Testing Domain

We have applied the IMA−CID approach into the development of an educational module for the software testing domain. As the conceptual model we have used *OntoTest* [3,2], an ontology of software testing, which aims to support acquisition, organization, reuse and sharing of knowledge on the testing domain. Due to the complexity of the testing domain, we have adopted a layered approach to the development of *OntoTest*. On the ontology level, the Main Software Testing Ontology addressed the main concepts and relations associated with testing. On the sub-ontology level, specific concepts were refined and treated into details – testing process, testing artifacts, testing steps, testing strategies and procedures, and testing resources . For the sake of illustration, Figure 1 shows one of the *OntoTest* sub-ontologies – Testing Strategy and Procedure.

Based on the concepts and relations represented into *OntoTest*, we have developed the instructional and the didactical models, according to the IMA−CID approach. For the sake of space, these models will not be illustrated here. In the end, the software testing educational module was composed by concepts, facts, principles, procedures, examples

[1] http://adlnet.org

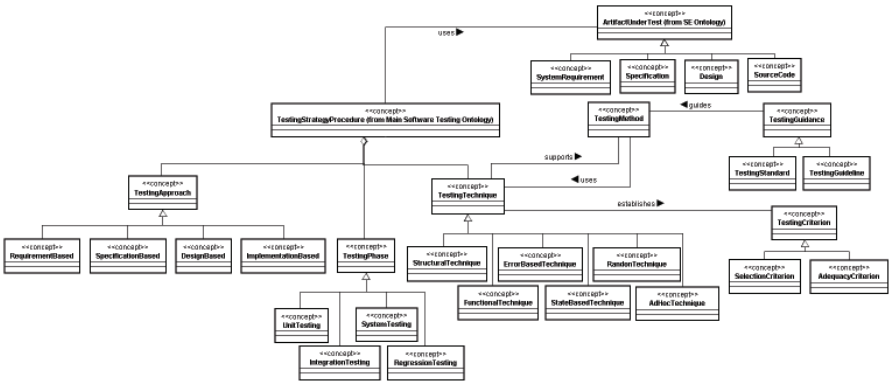


Figure 1. *OntoTest*: Testing Strategy and Procedure Sub-Ontology

and exercises, which were modeled and implemented as a set of slides, integrated to HTML pages, text documents, learning environments and testing tools.

To provide a preliminary evaluation on the effectiveness of the testing module, it was applied in: (1) a three-hour short-course; and (2) in two one-semester undergraduate courses at ICMC/USP [4]. The results obtained so far provide some evidences on the practical use of the $\mathcal{IMA}-\mathcal{CID}$ approach (and also the adoption of ontologies at its conceptual level) as a supporting mechanism to the development of effective educational modules. However, we highlight that applying it without an automated support is an error-prone activity. So, we are working on the development of a tool for helping the construction of the $\mathcal{IMA}-\mathcal{CID}$ models. An overview on the $\mathcal{IMA}$*Tool* is provided next.

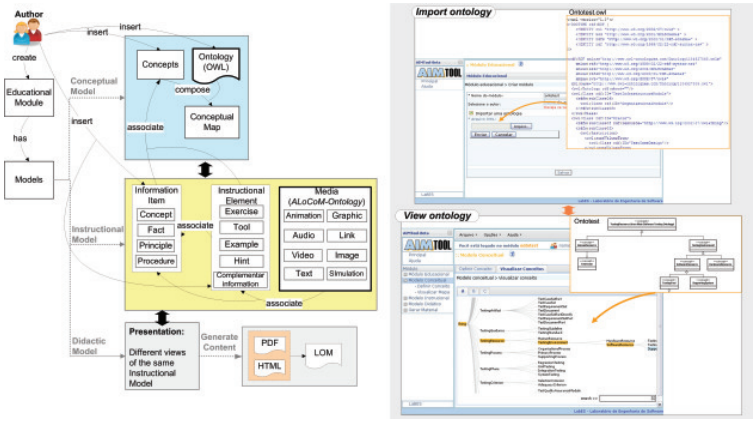
5. An Automated Tool for Modeling and Generating Educational Content

$\mathcal{AIM}$*Tool* aims at providing automated support for content modeling, focusing on the collaborative construction of the $\mathcal{IMA}-\mathcal{CID}$ models. We also intend to use its resulting specifications on the automatic content generation, which could be customized according to pedagogical interests, learner's profile, instructor's preference, course length, etc.

Figure 2(a) summarizes how $\mathcal{IMA}$*Tool* works, based on the $\mathcal{IMA}-\mathcal{CID}$ models. An ontology is imported as an OWL file, playing the role of the conceptual model of $\mathcal{IMA}-\mathcal{CID}$ to support the concepts definition. Information items and instructional elements are related to the concepts (ontology terms), establishing the instructional model. Notice that the media is classified according to the *ALOCoM-Ontology*. The didactic model is developed by defining the navigation sequence among the objects already modeled. Finally, from the didactic model, $\mathcal{AIM}$*Tool* can automatically generate and package the content according to the LOM specifications. Figure 2(b) illustrates *OntoTest* being imported and visualized. $\mathcal{AIM}$*Tool* has been developed in Java, as a Web application. We are now in the final phase of its development, working on the content generation module.

6. Conclusions and Further Work

In this paper we discussed some aspects of evolving $\mathcal{IMA}-\mathcal{CID}$, specially by using ontologies at its conceptual level. To illustrate our ideas, an ontology of software testing was used as the conceptual model of $\mathcal{IMA}-\mathcal{CID}$ for developing an educational module on this domain. As a further work, we intend to keep investigating the use of ontolo-



(a) General Structure (b) Ontology importation and visualization

Figure 2. *IMATool*: Overview

gies to support the development of the other *IMA–CTD* models, specially the didactic one. We are also motivated to keep evolving and evaluating the mechanisms we have proposed in different domains, for other areas and broader projects. At the very end, we are interested in establishing a culture for "open educational modules" so that the use and evolution of them by a broader community would be better motivated and become a reality. The adoption of ontologies should be explored in this perspective as well.

Acknowledgments

The authors would like to thank the Brazilian funding agencies (FAPESP, CAPES, CNPq) and to the QualiPSo Project (IST-FP6-IP-034763) for their support.

References

[1] E. F. Barbosa and J. C. Maldonado. An integrated content modeling approach for educational modules. In *IFIP 19th World Computer Congress – International Conference on Education for the 21st Century*, pages 17–26, Santiago, Chile, August 2006.

[2] E. F. Barbosa, E. Y. Nakagawa, and J. C. Maldonado. Towards the establishment of an ontology of software testing. In *18th International Conference on Software Engineering and Knowledge Engineering (SEKE 2006)*, pages 522–525, San Francisco, CA, July 2006. Short Paper.

[3] E. F. Barbosa, E. Y. Nakagawa, A. C. Riekstin, and J. C. Maldonado. Ontology-based development of testing related tools. In *20th Int. Conference on Software Engineering and Knowledge Engineering (SEKE 2008)*, pages 697–702, San Francisco, CA, July 2008.

[4] E. F. Barbosa, S. R. S Souza, and J. C. Maldonado. An experience on applying learning mechanisms for teaching inspection and software testing. In *21st Conference on Software Engineering Education and Training (CSEET 2008)*, pages 189–196, Charleston, SC, April 2008.

[5] T. R. Gruber. Towards principles for the design of ontologies used for knowledge sharing. *Int. Journal Human-Computer Studies*, 43(5/6), 1995.

[6] IEEE Learning Technology Standards Committee. Learning Object Metadata (LOM), June 2002.

[7] J. D. Novak. Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching*, 27:937–949, 1990.

[8] M. Uschold and M. Grüninger. Ontologies: Principles, methods and applications. *Knowledge Engineering Review*, 11(2), June 1996.

[9] K. Verbert, J. Jovanovic, D. Gasevic, E. Duval, and M. Meire. Towards a Global Component Architecture for Learning Objects: A Slide Presentation Framework. In *Proc. Of the 17th ED-MEDIA*, 2004.

Ontology of the Learner's Annotation Objectives

Hakim MOKEDDEM [1], Faiçal AZOUAOU [1], Cyrille DESMOULINS [2]
[1] *National School of Computer Science (E.S.I)*
Oued Smar Algiers, Algeria
{h_mokeddem, f_azouaou}@esi.dz
[2] *CLIPS- IMAG, University Joseph Fourrier*
BP 53 38041 Grenoble cedex 9 France
{Cyrille.desmoulins}@imag.fr

Abstract. This article aims at defining semantic annotation ontology of the learner, in order to use it in a pedagogical annotation tool. All the annotations created by the learner with this tool constitute a pedagogical memory for him.. To identify and model the annotation semantics, we develop an ontology of annotation objectives using the approach proposed by [1], then we describe the ontology implementation in the EasyAnnotation semantic tool.

Keywords. e-learning, annotation, learner, ontology, active learning situation.

Introduction

Learner carries out various learning activities, during which, he handles different types of learning objects. The latter can be an exercise, a simulation, a text, a course, etc. The handling of these objects occurs in the context of active learning in which the learner becomes an actor responsible of his own learning.

When doing these activities, the learner needs to memorize his ideas directly on the learning objects he is using in order to reuse them later. Consequently, we consider the set of annotations created by the learner, as his pedagogical memory, which contains prints of his learning.

Each of the created annotation can be described using several properties (shape, anchor…), but the most important one is its semantics, which corresponds to the learner's implicit objective during the annotation creation. These objectives can be: to memorize an error, to memorize a question, etc. The lost of the annotation semantics makes it often useless.

So that these annotations can be handled by software agents, and shared with other learners, annotation semantics has to be formalized each time an annotation is created. Among the various possible models to represent this semantics, a solution is to choose an ontological representation, as it enables *a formal and explicit specification of a shared conceptualization* [2].

To design a quality ontology, in this work we follow the methodology proposed by Noy [1] and made popular by the protégé tool [3].

We start this article by developing the annotation objectives ontology following the Noy's methodology [1] and then, we describe the annotation tool, EasyAnnotation, in which it has been implemented.

1. Ontology of the Objectives of Annotation

In order to model the annotation semantics, we use the concept of ontology [5], *which offers a specification of a conceptualization of a domain knowledge* [6]. This domain of knowledge is in our case the learner's annotation objectives.

There are several methods for the development of ontologies. To design the ontology of the learner's annotation objectives, we follow the iterative method for the development of ontologies proposed in [7]. We describe below how we follow each stage of this method.

1.1. Domain and scope of ontology

We start the development of ontology by defining its domain and scope, this is by answering to the following questions**:**

- What is the field covered by the ontology?
  - The field of our ontology is the objective of the learner's annotations in an active learning situation. Our ontology includes the concepts which describe the objectives of the annotations produced by the learner in the realization of his learning activities.
- What are the ontology development's goals?
  - Ontology is designed with an aim of formalizing and clarifying the semantic (objective) of the annotations produced by the learner. This formalization enables us to implement the semantics of the annotations in an annotation tool dedicated to the learner..

1.2. Re-use of related work

The objective of this stage is to re-use the existing ontologies even if they have a different objective. We can re-use all or a part of these ontologies after having adapted them to our needs.

Mille proposes in [4] a formalization of ontology of the learner's annotation objectives which only contains generic concepts hence, it does not clarify the pedagogical semantics of the learner's activities.

However, in the literature, we can find specifications that clarify the pedagogical semantics like: LOM [8] that identifies the concepts of description of the learning resources, and IMS-LD [9] that identifies the necessary concepts to modelling the learning process. Nevertheless these two specifications are not specific to learner's annotation activity.

1.3. Identification and structuring of the ontology concepts

We present the resulting ontology in figure 1 (see below) that shows the learner's annotation objectives structured by the relation "is-a".

Figure1. Ontology of learner's annotations objectives

To identify the main concepts of our ontology, we focus on the three previous works in the following way:

We extract the generic objectives of annotation specified by Mille [4] and enrich them with the pedagogical semantics specified by LOM and IMS-LD.

In this design process we follow a TOP-DOWN approach, starting with the most generic concepts, then an iterative process enabling us to improve the concepts and the structure of the ontology.

We use the Protégé [3] editor to design our ontology. It graphically allows building the hierarchy of the classes and the edition of ontology in the desired language.

1.4. Classes properties and instances

Since the classes alone do not provide sufficient information to represent the objective of the learner's annotation, we must then describe the internal structure of each concept, and explain the instances of classes. In our case all the ontology concepts have two properties: its identifier and its name...

2. Implementation in EasyAnnotation

EasyAnnotation is a web based annotation tool. It is built on the basis of Mozilla Firefox. It enables the learner to annotate directly on web pages; it is composed of a rich interface allowing the learner to keep traces of the produced annotations. Figure 2 shows the toolbar interface of EasyAnnotation and some created annotations.

The first version of EasyAnnotation is a Firefox **add-on** developed in JavaScript using the Document Object Model (DOM) technology. It allows the storage of the annotations internally in the learner's computer. The following scenario explains the interaction between the user and the annotation tool at the creation of a new annotation.

- The user selects some passages of a text in the document
- Then he chooses the form of the annotation
- The browser pops up a new window (see Figure 3), where the user can type the text of his annotation, and choose the type of the annotation that represents our ontology.
- The user saves the annotation.

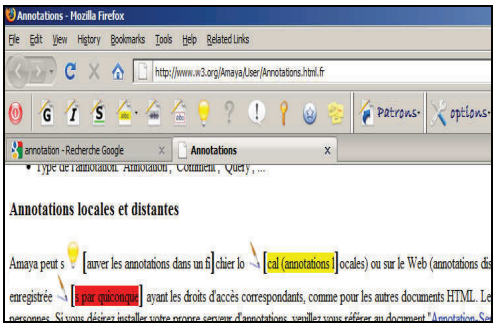


Figure2. EasyAnnotation toolbar

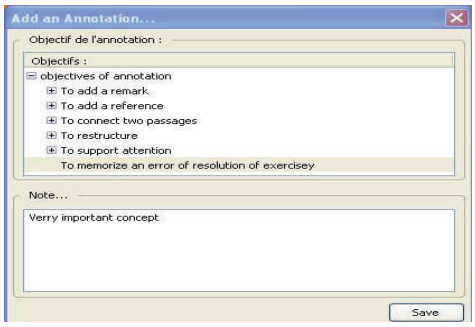


Figure3. The annotation window

3. Conclusion and Future Work

Our goal in this article is to clarify the semantic of the annotation produced by the learner by handling the various learning objects. The clarification of semantic learner's annotation will enable him to memorize his traces which represent his ideas, knowledge and remarks.

We developed for this, an ontology of the annotation's semantics that includes generic properties (to Add a remark, to Criticize, to Develop, to connect two passages, to support the attention…) and others rather specific which characterizes the learner and his activities (to criticize a course, to connect two questions, to develop a concept,…).

The purpose of the development of the ontology is to implement it in an annotation tool dedicated to the learner.. We implemented the first prototype version of EasyAnnotation that supports semantic learner's annotation of web pages content.

In our perspective, we will try to enrich the annotation's tool with a "domain ontology" which characterizes a particular domain of learning, and an ontology of the context which describes the context in which the learner annotates his document. In addition, to extend the annotation tool to support sharing annotation between learners.

We can also enrich the semantic annotation ontology by adding detailed classes that inherit the existing classes. This enrichment may be achieved through the monitoring of the annotations produced by the learners.

References

[1] Noy, N.F. and D.L. McGuinness. *Ontology Development 101: A Guide to Creating Your First Ontology* (2002).

[2] Gruber, T., R., *A Translation Approach to Portable Ontology Specifications*. Academic Press (1993).

[3] Noy, N.F., et al., *Protégé-2000: an open-source ontology-development and knowledge-acquisition environment*. AMIA Annu Symp Proc. (2003), p. 953.

[4] Mille, D., *Modèles et outils logiciels pour l'annotation sémantique de documents pédagogiques. Thèse.*, in *Département informatique*. Université Joseph-Fourier: Grenoble (2005), 173 pages.

[5] Guarino, N. *Understanding, Building, and Using Ontologies : A Commentary to 'Using Explicit Ontologies in KBS Development'.* in *IJHCS* (1997).

[6] Gruber, T.R., *Towards Principles for the Design of Ontologies Used for Knowledge Sharing.* Int'l J. Human-Computer Studies, 1995. **43**(5/6): p. 907–928.

[7] Noy, N. and D. McGuinness, *Ontology Development 101: A Guide to Creating Your First Ontology.* Knowledge Systems Laboratory (2001).

[8] IEEE, *IEEE Standard for Learning Object Metadata, Learning Technology Standards Committee.*

[9] Design, I.L. *IMS Learning Design* (2005) [cited; Available from: http://www.imsglobal.org/learningdesign/.

Comparing Three Approaches to Assess the Quality of Students' Solutions

Niels PINKWART and Frank LOLL
Clausthal University of Technology, Germany

Abstract. Collaborative filtering has potential for usage in Social Semantic Web e-learning applications: the quality of a student provided solution can be heuristically determined by peers who review the solution, thus effectively disburdening the workload of teachers and tutors. This paper compares three collaborative filtering algorithms which are based on different paradigms – one based on the assumption that a student who can classify the quality of given alternative solutions correctly is also able to provide a high-quality solution himself, another following a classical peer review paradigm, and a third being a mixture of both. An evaluation of the algorithms with data collected during a lab study showed that all algorithms can classify peer solutions correctly. Thus, these approaches have high potential as a support for classic academic teaching in larger classes.

Introduction

The term *Social Semantic Web* (SSW) describes an emerging design approach for building and using *Semantic Web* applications which employs *Social Software* and *Web 2.0* approaches. In SSW systems, groups of humans are collaboratively building domain knowledge, aided by socio-semantic systems [1]). The collaboration process of the users in SSW systems can have multiple purposes like group based creation of domain ontologies or the collaborative semantic classification of content (determination of properties of ontology elements). Both are potentially valuable in education. While the former can be a technique for collaborative knowledge building through jointly structuring an unknown knowledge domain, including the discussion of domain concepts and relations, the latter allows for jointly annotating or evaluating learning materials [2] and for heuristically determining the quality of task solutions through collaborative efforts.

This paper presents an example for the latter type of SSW systems in education. We compare three different approaches for solving the problem of automatically determining student solution's quality. Two of these approaches make use of peer reviews: the quality of a student's task solution is determined heuristically by assessments of other students. Typical points of critique concerning a peer review approach in education are related to the students' lack of knowledge and experience in assessing task solutions and to the risks of intentional manipulation [3, 4]. Yet, this form of using SSW approaches for education disburdens tutors and, at the same time, provides the possibility for students to train their critiquing skills. If there are tasks which allow for more than one correct solution, students have a chance to learn different acceptable ways to solve a problem. Also, students may empathize with other learners' problems easier and understand reasons for wrong task solutions sometimes better than experts, which can make their reviews sometimes more valuable than those of experts [5].

In spite of their potential however, peer review mechanisms have only been rarely used and empirically evaluated with respect to their effectiveness in the e-learning sector till now. Some of the few existing systems that make use of a peer-review approach to assess solutions quality are PeerGrader (PG) [6, 7], SWoRD [8], and LARGO [9]. While the systems' approaches are promising, they are limited in several ways: They are either specialized for a particular application area such as legal argumentation (LARGO) or writing skills training (SWoRD), or they involve a rather complicated and long-term review process (SWoRD, PG). In this paper, we present and compare 3 heuristics for estimating the quality of student solutions that are not constrained to a specified task area and that do not require time-consuming re-writing phases but only short quality assessments. While the first algorithm implements a "plain peer review" strategy, the second algorithm additionally relies on the hypothesis that a student who is able to correctly assess other student's solutions (i.e., classify them as poor or good) is likely to have provided a good solution himself – since a good judgment about solution quality can only be made when a task has been understood and solved. The third heuristics relies only on the latter hypothesis and does not include any peer reviews.

1. Peer Review Based Heuristics

The first heuristics consists of two components – an *evaluation rating* and a *quality rating*. The application scenario for this algorithm is then when students work on a task and provide a solution, they are asked to assess some alternative solutions afterwards.

The first component of the heuristics is the *evaluation rating*. Once a student has provided a solution, it is presented to other students to be assessed. All assessments get collected, averaged and weighted (assessments of better students get higher weights). An illustrating example: Assume a solution gets four assessments $w_1=0.9$, $w_2=0.2$, $w_3=0.4$ and $w_4=0.5$ from students whose own solutions have (system-internal) *quality ratings* of $q_1=0.8$, $q_2=0.1$, $q_3=0.3$ and $q_4=0.7$. The first assessment gets a higher weight than the others because the student who provided it has a higher *quality rating* as compared to the others. His opinion is thus considered as more important than the other students' opinions by the system heuristics. Then, the *evaluation rating* for the assessed solution is calculated by:

$$eval = \frac{1}{\sum_{i=1}^{j} q_i}\left(\sum_{i=1}^{j} w_i \cdot q_i\right) \Rightarrow \frac{1}{1.9}(0.9 \cdot 0.8 + 0.2 \cdot 0.1 + 0.4 \cdot 0.3 + 0.5 \cdot 0.7) \approx 0.63$$

The *quality ratings* are calculated based on the *evaluation rating* scores with an additional damping factor to assure that for solutions with very few peer reviews, no extremely low or high scores are possible. The *evaluation rating* gets weighted dependent on the number of received assessments p for a solution. Its impact thus increases with an increasing number of assessments. In the formula, *base* denotes a starting value (default: 0.5), and the constant c corresponds to the weight of this starting value relative to the weight of the peer reviews. In our study described in the next sections, we have chosen $c=3$. Thus, the *quality rating* q is calculated by:

$$q = \frac{c}{p+c} base + \frac{p}{p+c} eval$$

2. Base Rating Heuristics

The algorithm presented in the previous section assumes an equal start rating of *0.5* for all the student solutions (if no peer reviews are available, then *p=0* in the quality rating formula). Based on the assumption that a student who can classify the quality of given alternative solutions correctly is also able to provide a high-quality solution himself, the heuristics can be improved by replacing the static start value with a dynamically calculated *base rating* for a student's solution. The rationale here is simple: students who classify good solutions as good (and poor ones as poor) are likely to have understood the problem, and thus have probably provided a good solution themselves. Once a student provided n assessments $w_1, \ldots, w_n$ for n other student's solutions (which have *quality ratings* of $q_1, \ldots, q_n$ themselves), the *base rating* is calculated (see below). Independent of the quality of the solution that a student has to assess, this formula allows *base ratings* between *0.0* and *1.0*. To illustrate this: Assume there are solutions with *quality ratings* of q_1=0.35, q_2=0.6 and q_3=1.0. The worst ratings a user might make here, i.e. the ratings with the highest possible difference between q_i and w_i, are w_1*=1.0,* w_2*=0.0 and* w_3*=0.0*.

$$base = 1 - \frac{1}{n}\sum_{i=1}^{n}\frac{|w_i - q_i|}{\max(q_i, 1-q_i)} \Rightarrow 1 - \frac{1}{3}\left(\frac{|0.35-1.0|}{\max(0.35;0.65)} + \frac{|0.6-0.0|}{\max(0.6;0.4)} + \frac{|1.0-0.0|}{\max(1;0)}\right) = 0.0$$

In summary, we have so far proposed two algorithms to heuristically determine the quality of student solutions. While the first (called PR ONLY) makes use of a classic peer review approach with a fixed "base value" for those solutions that were not assessed yet, the second one (PR + BASE) replaces this base value with a base rating, calculated dynamically. A third variant (BASE ONLY) is to use *only* the base rating formula. We tested the three algorithms with data that we collected for a lab study described in [10]. Originally, the study was conducted to evaluate a slightly different algorithm; however parts of the log data – which essentially contains the student solutions and their peer reviews – can be used to test the algorithms described in this paper as well. We next briefly describe the software and the experimental procedure, and then present the results of our analysis, focusing on the question which (if any) of the algorithms works best for classifying student's solution quality.

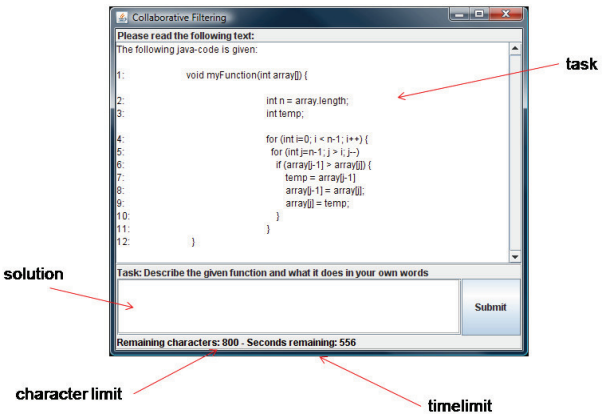


Figure 1. User interface

3. Implementation

After an initial login to the system, students go through the following phases as they use the system: (1) Work on task (Figure 1 shows a sample task on Java programming), (2) Assess 3 alternative solutions of different students on a scale from 0 to 10 for the just completed task (Figure 2), and (3) Repeat steps 1 & 2 as long as there are tasks to complete. The alternatives to be assessed in step 2 were chosen randomly. Solutions with less received reviews were preferred and a set of alternatives with similar quality was avoided.

4. Study Description

We conducted a controlled lab study in May 2008 at Clausthal University of Technology with 23 students, 11 female and 12 male. The participation was voluntary and paid. The students had to work on 12 tasks from various knowledge areas. The tasks were of the following types: (1) text summaries, (2) text interpretations, (3) knowledge tests without possibility to guess, and (4) knowledge tests with possibility to guess.

In the first task type (text summaries), the participants got articles dealing with different topics (e.g. a text about Second Life), which differed in their level of complexity and required, at least in parts, domain-specific knowledge to get the main points, which had to be summarized. The second task type (text interpretations) focused a fact-based news article about the take-over of DoubleClick by Google. The students were asked to mention and discuss possible concerns towards privacy of customers based on facts in the text. The third task type (knowledge tests without possibility to guess) consisted of 5 tasks where guessing was not possible (e.g. the calculation of a derivative of a function to calculate the slope at a given coordinate). The last task type (knowledge tests with possibility to guess) consisted of problems which could at least be approximated by logical deduction even without specific knowledge. An example here was the estimation of the population of Austria by means of a text about the size of the country.

The students had an overall time limit of 75 minutes. Furthermore, each task had a character limit as well as a time limit (see Figure 1), which served as orientation what kind of solution was expected. All participants were instructed to assess alternative solutions even if they did not know the correct solution for a task. To solve the cold-start problem [11] and offer alternative solutions also for the first participants who took part in the study, we provided 3 alternative solutions of different quality per task.

5. Results

To evaluate the results of the heuristics, all solutions were manually graded independently by two human experts (a professor of computer science and an advanced graduate student) on a scale from *0* to *10*. To check whether the human graders' assessments were similar (if human graders disagree, then a realistic baseline for the heuristics is hard to define), we first calculated inter-rater reliability based on Cronbach's Alpha [12]. We received excellent agreements with α-values between 0.834 and 0.982 in the different task groups. Therefore we averaged both human graders' scores and used the resulting "human grading" as a baseline for evaluating the algorithm results. Specifically, to analyze the quality of the three different heuristics,

we computed the correlation between the algorithm output and the human grading.For the "PR ONLY" heuristics, the result was a correlation of 0.53 – thus, a medium-to-large correlation between the algorithm's quality heuristics and the human grading. Apparently, the peer review paradigm worked quite well for the test scenario. For the "PR + BASE" algorithm, the correlation was not significantly better (0.54). We conclude from this that adding the (relatively complicated) base rating calculations did not change much in terms of the algorithm's predictive power. However, using only the base ratings (algorithm "BASE ONLY") also still produced a considerable correlation to the human grading of 0.46. For calculating the "BASE ONLY" values (cf. section 2: quality scores needed to calculate base ratings!), we used the human provided grades, so that this study confirms the hypothesis that a student who is able to recognize good (or poor) solutions is also likely to have provided a good solution himself.

6. Conclusion

Overall, the data analysis confirmed our expectations and even partially exceeded them. All three tested *SSW* methods for building quality related metadata about learning objects (here: student solutions) produced acceptable results, measured in terms of correlation to data provided by human experts. Particularly, it did not matter if "classical" peer review approaches were used, if a different strategy that relies on recognizing good students based on their correctly classification of others' solutions was followed, or if the two were combined (though the latter produced the best results).

References

[1] Morville, P. (2005). Ambient Findability. O'Reilly Media.
[2] Walker, A., Recker, M. M., Lawless, K., Wiley D. (2004). Collaborative Information Filtering: a review and an educational application. International Journal of AIED, Vol. 14, No. 1/2004 - pp. 3-28
[3] Dancer, W. T., Dancer, J. (1992). Peer Rating in Higher Education. Journal of Education for Business, 67, pp. 306–309.
[4] Mathews, B. (1994). Assessing Individual Contributions: Experience of Peer Evaluation in Major Group Projects. British Journal of Educational Technology, 25, pp. 19–28.
[5] Hinds, P. J. (1999). The Curse of Expertise: The Effects of Expertise and Debiasing Methods on Predictions of Novice Performance. Journal of Experimental Psychology: Applied, 5, 205–221.
[6] Gehringer, E. F. (2001). Electronic Peer-Review and Peer Grading in Computer-Science Courses. In Proceedings of the 32nd SIGCSE Technical Symposium on Computer Science Education, February 2001, Charlotte, North Carolina, United States, pp. 139 – 143.
[7] Lynch, C., Ashley, K., Aleven, V., & Pinkwart, N. (2006). Defining Ill-Defined Domains; A Literature Survey. In Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains at the 8th International Conference on Intelligent Tutoring Systems (pp. 1-10). Jhongli (Taiwan)
[8] Cho, K., Schunn, C. D. (2007). Scaffolded Writing and Rewriting in the Discipline: A Web-Based Reciprocal Peer-Review System. Computers & Education, Vol. 48 (3).
[9] Pinkwart, N., Aleven, V., Ashley, K., Lynch, C. (2007). Evaluating Legal Argument Instruction with Graphical Representations Using LARGO. In Proceedings of the 13th International Conference on Artificial Intelligence in Education (p. 101-108). IOS Press.
[10] Loll, F., Pinkwart, N. (2009). Using Collaborative Filtering Algorithms as eLearning Tools. In Proceedings of the 42nd Hawaii International Conference on System Sciences.
[11] Maltz, D., Ehrlich, E. (1995). Pointing the Way: Active Collaborative Filtering. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.
[12] Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. Psychometrika, 16(3).

The MATHESIS Ontology: Reusable Authoring Knowledge for Reusable Intelligent Tutors

Dimitrios SKLAVAKIS [a,1] and Ioannis REFANIDIS [a]
[a] *Department of Applied Informatics, University of Macedonia, Thessaloniki, Greece*

Abstract. This paper describes the MATHESIS Ontology, which is part of the MATHESIS project that aims at the development of an intelligent authoring environment for reusable model-tracing math tutors. The purpose of the ontology is to provide a semantic and therefore inspectable and reusable representation of the declarative and procedural authoring knowledge necessary for the development of a model-tracing tutor as well as of the declarative and procedural knowledge of the tutor under development. While the declarative knowledge is represented with the basic OWL components, i.e. classes, individuals and properties, the procedural knowledge is represented via the *process model* of the OWL-S web service description ontology. By using OWL-S, every authoring or tutoring task is represented as a composite process. Based on such an ontological representation, a suite of authoring tools will be developed at the final stage of the project.

Keywords. authoring systems, ontologies, semantic web, model-tracing tutors

Introduction

Intelligent tutoring systems and especially model-tracing tutors have been proven quite successful in the area of mathematics [1]. Despite their efficiency, these tutors are expensive to build both in time and human resources. The main goal of the ongoing MATHESIS project is to develop authoring tools for model-tracing tutors in mathematics, with knowledge re-use being the primary characteristic of the authored tutors as well as of the authoring knowledge used by the tools.

The MATHESIS ontology is an OWL ontology developed with the Protégé-OWL ontology editor. Its development is the second stage of the MATHESIS project. Aiming at the development of real-world, fully functional model-tracing math tutors, the project is being developed in a bottom-up approach. In the first stage the MATHESIS Algebra Tutor was developed in the domain of expanding and factoring algebraic expressions [2]. The tutor is web-based, using HTML and JavaScript. The authoring of the tutor as well as the code of the tutor were used to develop the MATHESIS ontology in a bottom-up way, as it will be described later.

The rest of the paper is structured as follows: Section 1 gives a brief presentation of the process model of OWL-S. Section 2 describes the part of the ontology that

[1] Corresponding Author: Dimitrios Sklavakis, Department of Applied Informatics, University of Macedonia, Egnatia 156, P.O. Box 1591, 540 06 Thessaloniki, Greece; E-mail: dsklavakis@uom.gr.

represents the model-tracing tutor(s), while Section 3 describes the representation of the authoring knowledge. Section 4 presents related work and, finally, Section 5 concludes with a discussion about the ontology and further work to complete the MATHESIS project.

1. The OWL-S process model

OWL-S is a web service description ontology designed to enable the tasks of (semi-) automatic discovery, invocation, composition and interoperation of Web services. It provides a language for describing service compositions. Every service is viewed as a Process. There are three subclasses of Process, namely the AtomicProcess, CompositeProcess and SimpleProcess.

Composite processes are decomposable into other composite or atomic processes. Their decomposition is specified by using control constructs such as Sequence and If-Then-Else. Any composite process can be considered as a tree whose non-terminal nodes are labeled with control constructs. The leaves of the tree are invocations of other processes, composite or atomic. These invocations are indicated as instances of the Perform control construct.

2. Tutor Representation in MATHESIS ontology

The MATHESIS project has as its ultimate goal the development of authoring tools that will guide the authoring of real-world, fully functional model-tracing math tutors. This means that during the authoring process and in the end, the result will be program code that implements the tutor, i.e. the ontology must be able to represent the program code. For this reason, in the first stage of the MATHESIS project the MATHESIS Algebra tutor was developed to be used as a prototype target tutor.

The MATHESIS Algebra tutor is a Web-based, model-tracing tutor that teaches expanding and factoring of algebraic operations: monomial and polynomial operations, identities, factoring. It is implemented as a simple HTML page with JavaScript controlling the interface interaction with the user and implementing the tutoring, domain and student models. Therefore, it is the representation of the HTML and JavaScript code that forms the low-level MATHESIS ontology of the tutor as described below.

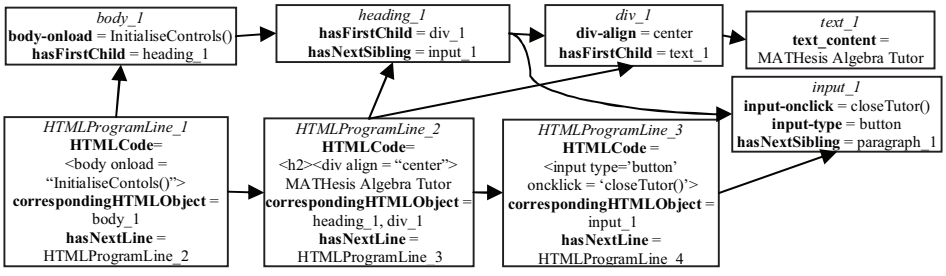


Figure 1. The HTML code and the corresponding DOM representation

2.1. Representation of the HTML Code of the Tutor

The representation of the HTML code and the corresponding Document Object Model (DOM) of the user interface are shown in Figure 1. Each line of the HTML code is represented as an instance of the HTMLProgramLine class having three properties: the HTMLCode, hasNextLine and correspondingHTMLObject. The last one points to the HTMLObject defined by the HTML code.

Each HTMLObject has the corresponding HTML properties as well as the hasFirstChild and hasNextSibling which implement the DOM tree. Therefore, there are two representations of the HTML code enabling a bottom-up creation of the ontology (from HTML code to DOM) and a top-down (from the DOM to HTML code).

2.2. Representation of the JavaScript Code of the Tutor

The representation of the JavaScript code is shown in Figure 2. Each line of the JavaScript code is represented as an instance of the JavaScript_ProgramLine class having three properties: the javascriptCode, hasNextLine and hasJavaScriptStatement. The last one points to a JavaScript_Statement instance which is an AtomicProcess of the OWL-S process model.

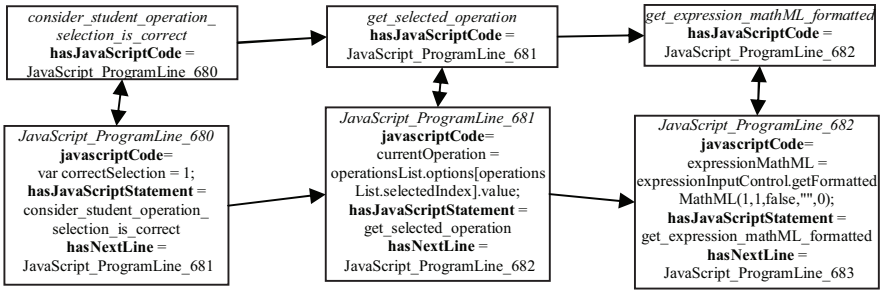


Figure 2. The JavaScript code and the corresponding JavaScript_Statement Atomic processes

Once again, there are two representations of the JavaScript code enabling a bottom-up creation of the ontology (from JavaScript code to JavaScript_Statement atomic processes) and a top-down (from the JavaScript_Statement atomic processes to JavaScript code).

2.3. Representation of the Tutoring Model

Being procedural knowledge, the model-tracing algorithm is represented as a composite process named Model_Tracing_Algorithm, shown in Figure 3. Each step of the algorithm is also a composite process. For example the Task_Execution_Expert_Process step can be described by an algorithm that performs other composite processes. These processes are instances of subclasses of the Task_Execution_Expert_Process class, shown in Figure 4. During the authoring of a specific tutor, the authoring tools will parse the tree of the Model_Tracing_Algorithm composite process and invoke for each tutoring task a corresponding authoring task represented also as a composite process in order to implement the tutoring task for the specific tutor (described in Section 3).

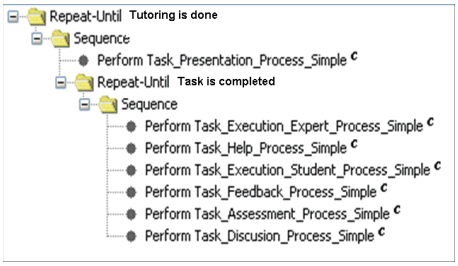


Figure 3. The Model_Tracing_Algorithm process

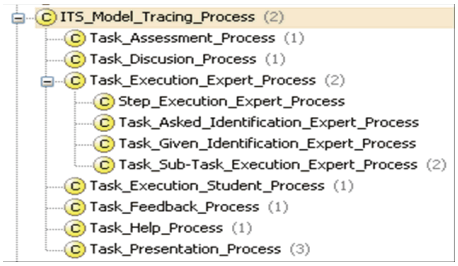


Figure 4. The model-tracing processes taxonomy

2.4. Representation of the Domain Expert Model

In a model-tracing tutor the Domain Expert Model executes the next step of the problem and produces the correct solution(s) to compare them with the student's proposed solution. If the solution step is simple, then it is represented as an instance of the atomic process JavaScript_Statement (see Section 2.2). If the step is complex, then it is represented as a composite process. This analysis ends when the produced composite processes contain only atomic processes, i.e. JavaScript_Statement instances.

3. Authoring Knowledge Representation in MATHESIS Ontology

As mentioned in Section 2.3, for each tutoring task of the model-tracing algorithm, there is a corresponding authoring task in the MATHESIS ontology, represented as a composite process. The authoring_task_execute_task_by_expert (Figure 5) for example corresponds to the Task_Execution_Expert_Process_Simple tutoring task (Figure 3). The define_data_structures_for_knowledge_components process, one of the composite processes that form this authoring task, is shown in Figure 6.

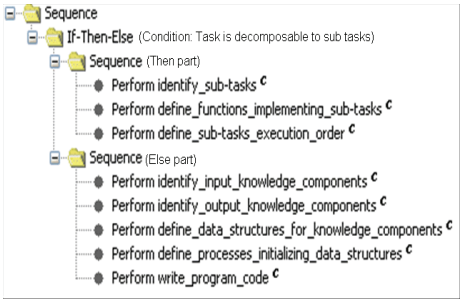


Figure 5. The authoring process for the Execute_Task_By_Expert tutoring task

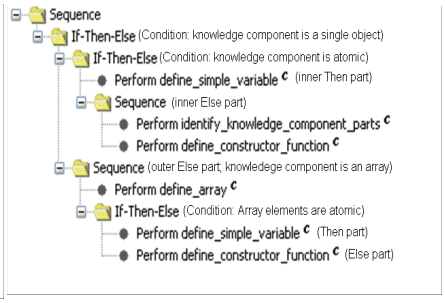


Figure 6. The authoring composite process define_data_structures_for_knowledge_componets

Based on all the above representations, the overall authoring process will have as follows: The tools will parse the model-tracing algorithm (Section 2.3). For each step of the algorithm, the corresponding authoring process will be called and traced. This authoring process will guide the author in creating recursively the various parts of the

tutor (Sections 2.1, 2.2, 2.4); as a consequence, any newly created tutor part becomes new knowledge in the ontology to be used later.

4. Related Work

The use of ontologies and semantic web services in the field of ITSs is relatively new. Ontological engineering is used to support authoring of instructional scenarios [3], provide educational feedback, or plan learning resources. However, there is a lack of semantic expressiveness and, more important, the difficult task of using and integrating low-level learning services to compose more complex ones, is not faced at all.

It is this difficult task that the MATHESIS project is trying to address using as low-level learning services the concept of problem-solving tasks and providing through the MATHESIS ontology a semantic description of these tasks and the way they can be combined to create more complex learning services (intelligent tutors).

5. Discussion and Further Work

In an overview of intelligent tutoring system authoring tools [4], it is suggested that authoring tools should support interoperability, re-usability, durability, scalability and accessibility. Even in this preliminary form, the MATHESIS ontology provides a proof-of-concept that it can serve as the basis for the development of authoring tools and implemented tutors that will match these criteria. The main reason for this claim is that the ontology provides an open, modular and multi-level representation (ranging from conceptual design to program code) of both authoring and tutoring knowledge.

Of course, it is expected that a lot of work has to be done: the ontology must be extended, refined and formalized. This will be done by representing the whole MATHESIS Algebra tutor into the ontology. Because of the tremendous workload this task entails, an initial set of authoring tools are being developed: parsers for HTML to/from MATHESIS interface model and for JavaScript to/from MATHESIS JavaScript_Statements/Program code translation; interpreters for the authoring and tutoring OWL-S processes; visualization tools for the authoring processes, the tutoring model (model-tracing algorithm) and the tutor parts being developed. Most of these tools will be implemented as Java plug-ins for the Protégé-OWL editor, accessing and updating the MATHESIS ontology.

References

[1] K.R. Koedinger, J.R. Anderson, W.H. Hadley, M.A. Mark, Intelligent Tutoring Goes to School in the Big City, *International Journal of Artificial Intelligence in Education* **8** (1997), 30-43.

[2] D. Sklavakis, I. Refanidis, An Individualized Web-Based Algebra Tutor Based on Dynamic Deep Model Tracing, *Proceedings of the 5th Hellenic Conference on Artificial Intelligence (SETN '08)*, LNAI **vol. 5138**, 389-394, Springer, Berlin/Heidelberg, 2008.

[3] R. Mizoguchi, Y. Hayashi, J. Bourdeau, Inside Theory-Aware and Standards-Compliant Authoring System, *Proceedings of the Fifth International Workshop on Ontologies and Semantic Web for E-Learning (SWEL '07)*, http://compsci.wssu.edu/iis/Papers/SWEL07-Proceedings.pdf.

[4] T.Murray, An Overview of Intelligent Tutoring System Authoring Tools: Updated Analysis of the State of the Art, *Authoring Tools for Advanced Technology Learning Environments*, 491-544, Kluwer Academic Publishers, Netherlands, 2003.

Linked Data as a Foundation for the Deployment of Semantic Applications in Higher Education

Thanassis TIROPANIS[1], Hugh C. DAVIS, Dave MILLARD, Mark WEAL and Su WHITE
University of Southampton, UK

Abstract. The value of semantic technologies in the context of learning and teaching has often been associated with the use of reasoning to support learning processes. This paper discusses the value of linked data in addressing data interoperability and integration across higher education institutions and repositories. This value is related to higher education challenges and a proposal on deploying linked data in higher education is presented and discussed.

Keywords. Higher education, semantic technologies, linked data, learning

Introduction

The relevance of semantic technologies to learning and teaching has been examined in a number of different contexts in recent years [2, 4], among which is Higher Education (HE). The development of semantic technologies for learning often required agreement on ontologies, annotation of available resources and reasoning to facilitate learning related processes. The requirement for agreed ontologies has often presented a hurdle in the deployment of semantic technologies for learning on a large scale, involving resources in different administrative domains; on the other hand, the use of expressive ontologies on a smaller scale featured advanced reasoning to match learners and resources. The linked data movement advocates a bottom-up approach to ontology agreement [3] by shifting the focus first to the exposure of data in machine processable formats like RDF before agreeing on ontologies for specific applications.

Semantic tools and services relevant to higher education have been prototyped and, as a recent survey [5] shows, they can address the needs of students, teachers and researchers. However, support for additional kinds of higher education users, such as assessors, admissions teams, programme administrators, do not seem to be supported.

The JISC funded project SemTech[2] (Semantic Technologies for Learning and Teaching) performed a survey of semantic technology adoption in the UK higher education sector to outline a roadmap for the future adoption of semantic technologies. A workshop organised by SemTech in January 2009 identified a number of higher education challenges that semantic technologies were expected to address.

[1] Corresponding Author.
[2] http://www.semtech.ecs.soton.ac.uk

This paper discusses these challenges and argues for the creation of a linked data infrastructure on which semantic technologies relevant to higher education can be deployed. Section 1 discusses the relevance of semantic technologies generally to HE challenges and the extent to which these challenges can be addressed by semantic technologies. Section 2 investigates the relevance of linked data specifically to these HE challenges and issues. Section 3 proposes institutional policies that would help to foster linked data deployment across HE.

1. The Value of Semantic Technologies for HE Support

The HE challenges that could be addressed by semantic technologies, as identified at the SemTech workshop, consisted of a number of institutional challenges as well as challenges related to learning and teaching processes. The institutional challenges are particularly related to UK HE but it is expected they are relevant to HE in other countries too. A brief summary of these challenges includes:

- Visibility of degree programmes and research output of HE institutions
- Curriculum design
- Recruitment and retention of students
- Efficiency of accreditation
- Collaboration across departments and institutions through workflows
- Integration of knowledge capital, cross-curricular initiatives
- Transparency of data held by educational institutions

Semantic technologies are expected to provide for more efficient discovery of degree programmes to match the background and objectives of prospective students; the research output of institutions could be more visible to potential funding bodies. Student retention could be supported by more efficient monitoring of student activity and assessment of their progress. Institutional data is often dispersed across databases and is often not interoperable; semantic technologies could provide for such integration and support workflows and collaboration across departments or even institutions. Requirements for transparency on HE processes could be supported by making institutional data available in open formats, which could also assist in obtaining accreditation for degree programmes by external bodies. Curriculum design could also be supported by establishing how different curricula across HE institutions compare to each other and identify potential gaps that new degree programmes could address.

Apart from the institutional challenges, some of the learning and teaching challenges that were identified include:

- Support for course creation and delivery workflows
- Group formation for learning and teaching activities
- Support for critical thinking and argumentation
- Efficient construction of personal and group knowledge
- Assessment, certification and addressing of plagiarism

Work that has been reported in the literature in recent years and prototypes do address these areas, however, the survey performed by SemTech showed that there are no widely adopted tools and services to address these needs.

1.1. Scale

The lack of widely adopted semantic technologies for learning and teaching and the rarity of semantic applications to address HE challenges raise the following questions:

- What is the value of semantic technologies in the higher education domain?
- Is there a sufficient amount of HE data to perform reasoning on?
- Can HE data be exposed in formats easily mapped to ontologies?

The results of the survey performed by SemTech[3] showed that the value of semantic technologies in HE is primarily in well-formed metadata, secondarily in data interoperability and integration and thirdly in data analysis and reasoning. The surveyed tools and services were found to be collaboration tools (e.g. *Compendium*[4]), searching and matching tools (e.g. *Arnetminer*[5]), repositories and VLEs (e.g. *SKUA*[6]) or infrastructural tools supporting semantic annotation, integration, metadata storage and queries (e.g. *D2R server*[7]). The *potential added value* of semantic technologies per category outlined as follows:

- Collaboration tools could benefit from data integration and reasoning for inline recommendation of resources on other repositories.
- Searching and matching tools benefit from data integration and reasoning on a larger scale.
- Repositories, VLEs and annotation tools could provide additional value by linking to other repositories and by exposing machine processable data.

To reach the *potential added value* of semantic applications in higher education, the scale of availability of higher education data is critical and so is interoperability of metadata across institutions and repositories. Most of the higher education challenges, as identified in the previous paragraphs, rely on data integration on a large scale. For example, when it comes to HE curriculum design or alignment, course information is currently available on the Web pages of HE institutions but often not in machine processable formats; exposing HE curricula as linked data in RDF using SPARQL endpoints would enable relevant searches over different HE institutions, comparisons and analysis for this end. In this example, the power of linked data is in the scale or available data sources rather than reasoning.

It would be important that data exposed in linked data formats that can easily mapped to potentially more expressive ontologies for the development of specific semantic applications that employ advanced reasoning.

1.2. Reasoning

Certain applications establish the value of semantic technologies in advanced data analysis and reasoning and do not require large-scale data interoperability from the start.

---

[3] http://semtech-survey.ecs.soton.ac.uk
[4] http://www.aktors.org/technologies/compendium/
[5] http://www.arnetminer.org/
[6] http://www.myskua.org/
[7] http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/

Such applications include *Debategraph*[8] and *Cicero*[9] and can support learning based on critical thinking and argumentation. Nevertheless, even these applications could feature added value given the availability of additional resources in semantic formats. For example, argumentation tools could enable the discovery of resources to second certain arguments or to link to other argumentation data on additional platforms. Similarly, tools that rely on deep linguistic analysis of resources with textual descriptions to perform reasoning (e.g. *COGITO®* by *ExpertSystem* [10]) could benefit from interoperability with additional resources in additional repositories.

2. Exposing HE Linked Data

A significant amount of information is already exposed by institutions on their public Web pages. This information could also be exposed in RDF as linked data and help address institutional challenges such as exposing institutions' expertise and making their curricula and syllabi available for semantic technology enabled matching to prospective student interests. Infrastructural tools like *D2R server*, *Talis*[11] or *Virtuoso*[12] could provide for exposing data in relational databases as RDF via SPARQL endpoints. The availability of large RDF repositories (e.g. *RKBExplorer*) could also host linked data from a number of institutions and provide for optimised storage and searches [1].

Learning and teaching resources currently available in VLEs and internal repositories could expose their metadata via plugins; such extensions are already featured by repositories for publications such as *EPrints*[13] or *DSpace*[14].

Agreement on common URIs for RDF across institutions and repositories would be desirable but not required. URIs could be HE institution specific, VLE specific, standard (e.g. *DublinCore*) or agreed community-wide. A high degree of reusing URIs will make the mapping of linked data to higher ontologies more efficient.

2.1. Issues

Exposing linked data in HE can provide significant value in addressing HE challenges and in supporting learning and teaching activities. At the same time, there are certain challenges that need to adequately discussed and addressed.

It seems that by exposing information already publicly available as Web pages can support applications with valuable features (e.g. exposing degree program information, research output, expertise, and accreditation related information). Technologies like GRDDL[15] could support transition from HTML to linked data. However, additional data to address student retention like course evaluation data would potentially have to be available to selected parties. Other data would have to be sufficiently anonymized before exposed to any third party in order to protect personal information.

[8] http://debategraph.org/
[9] http://cicero.uni-koblenz.de/wiki/index.php/Main_Page
[10] http://www.expertsystem.net/
[11] http://www.talis.com/platform
[12] http://virtuoso.openlinksw.com/
[13] http://www.eprints.org/
[14] http://www.dspace.org/
[15] http://www.w3.org/TR/grddl/

The performance of even lightweight reasoning over resources dispersed across repositories is another issue to be considered. RDF can be stored and queried via SPARQL endpoints at each institution, or could be stored in larger RDF repositories that support optimised queries. Certain information may be required to remain within institutions (e.g. information of a sensitive nature or information frequently updated) while other information could be stored in large RDF repositories.

The cost of exposing linked data is another issue for consideration. Despite the availability of even free or open source *RDFizers*[16] there are additional cost parameters to consider. The existence of additional barriers due to institutional or government policies deserves further investigation.

Potentially novel teaching and learning activities enabled by linked data need to be identified and properly documented to enhance our understanding on the pedagogical potential of semantic tools and services over linked data.

3. Ways Forward

The potential of a deployed linked data field across the higher education sector has been argued in the previous sections together with the challenges that need to discussed, understood and addressed. It seems that there is significant value to be obtained by exposing information currently publicly available as HTML and that there is additional value in exposing data currently available in internal databases.

Taking this forward requires institutional policies on exposing linked data in a way potentially similar to when policies were established for exposing institutional information in HTML. Case studies of applications that could address HE challenges need to be conducted to identify more precisely what information each institution should consider exposing. Successful cases of using semantic technologies in a pedagogically meaningful way need to be documented and become available across institutions.

The cost of exposing linked data and of maintaining triple stores and SPARQL endpoints needs to be investigated. At the same time, the deployment of education related triple stores that will host metadata from institutions that are not able to support their own RDF repositories could be discussed. Best practises for exposing institutional data securely and selectively can be documented and studied by HE institutions.

References

[1] Harris, S. and Gibbins, N., 3store: Efficient Bulk RDF Storage. *In Proceedings 1st International Workshop on Practical and Scalable Semantic Web Systems*, Sanibel Island, Florida, USA, 2003.

[2] Johnson, L., Levine, A., & Smith, R., *The 2009 Horizon Report*, Austin, Texas: The New Media Consortium, 2009.

[3] O'Hara, K. and Hall, W., Semantic Web, *In: Marcia J. Bates; Mary Niles Maack; Miriam Drake (eds), Encyclopedia of Library and Information Science Second Edition,* Taylor & Francis, 2009.

[4] Ohler, J., The Semantic Web in Education – What happens when the read-write web gets smart enough to help us organize and evaluate the information it provides?, *EDUCAUSE Quarterly Magazine*, 31 (4), 2008.

[5] Tiropanis, T., Davis, H., Millard, D., Weal, M., *Semantic Technologies for Learning and Teaching in the Web 2.0 era-A survey*, 1st Web Science Conference, Athens, 2009, http://journal.webscience.org/166/

[16] http://simile.mit.edu/wiki/RDFizers