

SW-EL'04 : semantic web for E-learning : applications of semantic web technologies for E-learning, ISWC'04 workshop, 8th November 2004, Hiroshima

Citation for published version (APA):

Aroyo, L. M., & Dicheva, D. (Eds.) (2004). SW-EL'04 : semantic web for E-learning : applications of semantic web technologies for E-learning, ISWC'04 workshop, 8th November 2004, Hiroshima. s.n.

Document status and date: Published: 01/01/2004

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.

• The final author version and the galley proof are versions of the publication after peer review.

 The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
 You may not further distribute the material or use it for any profit-making activity or commercial gain
 You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

ISWC'04 WORKSHOP # 3

SW-EL'04: Semantic Web for E-Learning

Applications of Semantic Web Technologies for E-learning

SWEL Workshop co-chairs Lora Aroyo Darina Dicheva

SWEL ISWC'04 Session chairs Riichiro Mizoguchi Yukihiro Itoh

ISWC 2004: International Semantic Web Conference Hiroshima, Japan, November 7-11, 2004

Preface

The Semantic Web offers new technologies to the developers of Web-based applications aiming at providing more intelligent access to and management of the Web information and semantically richer modelling of the applications and their users. An important target for Web application developers nowadays is to provide means to unite, as much as possible, their efforts in creating information and knowledge components that are easily accessible and usable by third parties. Within the context of Semantic Web, there are several hot issues, which allow achieving this reusability, shareability and interoperability among Web applications. Conceptualizations (formal taxonomies), ontologies, and the available Web standards, such as XML, RDF, XTM, OWL, DAML-S, and RuleML, allow specification of components in a standard way. The notion of Web services offers a way to make such components mobile and accessible within the wide sea of Web information and applications.

The research on Web-based educational systems (WBES) traditionally combines research interests and efforts from various fields. Currently, the efforts in the field of ontologies and Semantic Web play an important role in the development of new methods and types of courseware, including ITS, learning management systems, adaptive educational hypermedia, and various other Web-based educational applications.

Standardization of educational content specification and annotation, course components sequencing paradigms, user modelling and other aspects of educational systems, also play an important role in achieving more flexible and interoperable courseware. There is a significant effort performed by educational standardization institutions (e.g. IEEE Learning Technology Standards Committee, CEN/ISSS, IMS, the US ADLnet, CETIS, ARIADNE) to propose standards for various aspects of educational systems, such as IEEE Learning Object Metadata, IMS Learner Information Package Specification, etc. The research effort in both communities comes together in the attempt to achieve improved interoperability, personalization, adaptation and flexibility for single and group users of WBES (e.g. instructors, courseware authors and learners).

The goal of this workshop is to outline the state-of-the-art in the application of Semantic Web technologies and standards for e-Learning. We aim at exploring, among others, the relationships between the main components of educational systems presented by the existing reference models, i.e. resource representation, domain model, sequencing representation, user profiles, and adaptation and tutoring strategies, and the existing Semantic Web and educational standards. The workshop topics include:

- Using Semantic Web technologies to improve:
 - personalization, adaptation, knowledge and user modelling
 - information retrieval
 - authoring of WBES.
- Web standards and metadata specifications for WBES:
 - information exchange protocols between WBES
 - consistency in standards evolution
 - mappings between existing Semantic Web and educational standards
 - educational metadata specification languages.
- Semantic Web-based architectures for WBES
- Educational Web services
- Real-world systems, case studies and empirical research for Semantic Web-based WBES.

This workshop follows the successful workshop on Concepts and Ontologies in Web-based Educational Systems, held in conjunctions with ICCE'2002 in Auckland, New Zealand. It is part of the Second International Workshop on Applications of Semantic Web Technologies for E-Learning (SWEL'04), which is organized in three sessions held at three different conferences. The aim is to discuss the current problems in e-Learning from different perspectives, including those of Web-based ITS and adaptive educational hypermedia, and the implications of applying Semantic Web and educational standards and technologies for solving them:

- SW-EL'04 Session at AH, 23rd August, 2004 Session co-chairs: Peter Dolog and Martin Wolpers
- SW-EL'04 Session at ITS, 30th September, 2004 Session co-chairs: Vladan Devedzic and Tanja Mitrovic
- SW-EL'04 Session at ISWC, 8th November, 2004 Session co-chairs: Riichiro Mizoguchi and Yukihiro Itoh

We hope that this workshop will provide some new insights and serve as a catalyst to encourage others to investigate the potential of the emerging Semantic Web technologies for the Web-based educational systems and contribute to the realization of the vision of the Educational Semantic Web.

November, 2004

Lora Aroyo Darina Dicheva Riichiro Mizoguchi Yukihiro Itoh

Workshop Committees

General co-chairs:

Lora Aroyo, Eindhoven University of Technology Darina Dicheva, Winston-Salem State University

ISWC'04 session co-chairs:

Riichiro Mizoguchi, Osaka University Yukihiro Itoh, Shizuoka University, (SIG-IES, JSAI)

Program Committee:

Peter Brusilovsky (University of Pittsburgh) Paloma Diaz (Universidad Carlos III de Madrid) Vanja Dimitrova (University of Leeds) Erik Duval (Katholieke Universiteit Leuven) Jim Greer (University of Saskatchewan) Tsukasa Hirashima, (Hiroshima Univeristy) Ulrich Hoppe (University of Duisburg) Geert-Jan Houben (Technische Universiteit Eindhoven) Mitsuru Ikeda(JAIST) Judy Kay (University of Sydney) Kinshuk (Massey University) Erica Melis (Universität des Saarlandes, DFKI) Tanja Mitrovic (University of Canterbury) Ambjörn Naeve (Royal Institute of Technology) Ossi Nykänen (Tampere University of Technology) Gilbert Paquette (LICEF) Simos Retalis (University of Cyprus) Demetrios Sampson (Center for Research and Technology, Hellas - CERTH) Katherine Sinitsa (Kiev Univeristy) Amy Soller (ITC-IRST) Steffen Staab (AIFB, University of Karlsruhe) Julita Vassileva (University of Saskatchewan) Felisa Verdejo (Ciudad Universitaria) Gerd Wagner (Eindhoven University)

Additional Reviewers:

Vadim Chepegin (Eindhoven University of Technology) Christo Dichev (Winston-Salem State University) Renata Guizzardi (University of Twente) Carsten Ullrich (Universität des Saarlandes, DFKI)

Table of Content

Session I: Educational Services & Interoperability Issues

A Community Oriented Approach to Delivering Learning Services Enrico Motta & Arthur Stuff	1
Specifying Learning Object-Based Goals and Capabilities with WSMO Jose Manuel Lopez-Cobo, Miguel-Angel Sicilia, & Sinuhe Arroyo	5
Integration of an Ontology Manager to Organize the Sharing of Learning Objects in a P2P Network <i>Gautier Bastide</i>	11
Description of an Instructional Ontology and its Application in Web Services for Education <i>Carsten Ullrich</i>	17
A Pragmatic Approach to Support Concept-based Educational Information Systems Communication Darina Dicheva & Lora Aroyo	25

Session II: Knowledge & User Modelling

Using Semantic Web Methods for Distributed Learner Modelling Mike Winter, Christopher Brooks, Gord Mccalla, Jim Greer & Peter O'donovan	33
Identifying Relevant Fragments of Learner Profile on the Semantic Web <i>Peter Dolog</i>	37
Intellect Disclosure Support Based on Organizational Intellect Model Mitsuru Ikeda, Yusuke Hayashi, Youhei Tanaka, Masataka Takeuchi & Riichiro Mizoguchi	43
E-portfolios for Meaningful Learning and Automated Positioning Tommy W. Nordeng, Stian Lavik & Jarle R. Meløy	49

Session II: Ontologies for Instruction & Authoring

Building of an Ontology of the Goals of IT Education and its Applications Toshinobu Kasai, Haruhisa Yamaguchi, Kazuo Nagano & Riichiro Mizoguchi	55
From Annotated Learner Corpora to Error Ontology: A Knowledge-based Approach to Foreign Language Pedagogy Md Maruf Hasan, Kazuhiro Takeuchi, Virach Sornlertlamvanich & Hitoshi Isahara	67
The Use of Ontologies in Web-Based Learning Marvin Tan & Angela Goh	75
Integration of Hyperbooks into The Semantic Web Jean-Claude Ziswiler, Gilles Falquet & Luka Nerima	81
Ontologies for Authoring of Intelligent Educational Systems Lora Aroyo & Riichiro Mizoguchi	89

Posters

Adaptive Support to e-Learning Using Formal Specifications of Applications, Tasks, Content and User Models	95
Aude Dufresne & Mohamed Rouatbi	
A Metadata Editor of Exercise Problems for Intelligent e-Learning	00
Tsukasa Hirashima	99
Automatic Generation of Courseware for Economics Mathematics	102
Yukari Shirota	105
A Rule Editing Tool with Support for Non-Programmers in an Ontology-based Intelligent Tutoring	105
System	107
Eric Wang, Sung Ah Kim & Yong Se Kim	

A Community Oriented Approach to Delivering Learning Services

Enrico Motta and Arthur Stutt

Knowledge Media Institute The Open University Milton Keynes, MK7 6AA, UK e.motta@open.ac.uk

Abstract

In the past few years we have seen a growing interest in applying semantic web technologies to education. Much of this work has concentrated on the use of Learning Objects. This research is flawed in three ways: firstly, the notion of learning Objects (LOs) is deeply problematic, secondly, learners need more than ready access to pre-packaged fragments of knowledge, and, thirdly, a focus on LOs constrains the deployment of semantic web terminologies to the needs of providers and consumers of LOs. We propose an alternative approach which is not tied to LOs and which extends the use of semantic technologies as a means of providing learning services which are owned and created by a knowledge community.

1 Introduction

There is currently much interest in applying Semantic Web technologies to learning. Much has been written about how the Semantic Web can be used to provide easier access to Learning Objects, including many interesting proposals for infrastructures for sharing LOs and for interoperation among different repositories. However this work fails on three counts: firstly, LOs are deeply problematic, secondly, it pays little attention to the needs of learners, and, thirdly, it fails to make use of the possibilities inherent in semantic web technologies.

We propose an alternative model of the use of semantic technologies in learning which focuses on the delivery of certain high level learning services (such as sense-making, structure-visualization, support for argumentation, novel forms of content customization, novel mechanisms for aggregating learning material, and so on) but which:

- does not tie semantic technologies to LOs
- makes full use of the potential of semantic technologies

- is more appropriate for learners
- provides a means of contextualization and interpretation
- · makes navigation through web resources easier

By learning services we mean the use of semantically augmented web services as a means of implementing support for certain educationally important activities such as sense-making. Our approach to the provision of learning services has three main components: Knowledge Charts, Knowledge Navigation and Knowledge Neighbourhoods. In the example below we illustrate this using a learning service which provides contextualization via structures representing scientific and other controversies. A Semantic Browser is used to (a) provide access to this service and (b) to navigate through the controversy structures as an aid to sensemaking. Both of these represent forms of Knowledge Navigation. The learning service and the controversy structure (we call it a Knowledge Chart) are both owned and constructed by knowledge communities focused on particular disciplines, topics or interests inhabiting bits of the Semantic Web which we call Knowledge Neighbourhoods.

2 The approach in more detail

We are deeply sceptical of the value of LOs and of the social implications of their use. This does not mean that we eschew them altogether. LOs are one kind of web resource which we can use. At the same time we are concerned to get away from a model of learning which views it as the consumption of learning objects. We take a more constructivist, community-centred approach. For us much depends on the formulation of a prospective/story/narrative about a particular topic which captures a community perspective and which can be used in a variety of learning contexts.

However, our view of learning is broader than that taken by many educationalists since we see the semantic web as a location where learners of all types can acquire knowledge in a variety of formal and informal learning contexts. We therefore do not address issues such as different learning styles, the details of pedagogic strategies or the possibility of pedagogy-specific metadata models.

We have the following problems with LOs:

- LOs are immature with many competing metadata schemes.
- These may not be flexible enough to capture relevant characteristics.
- There are costs as well as benefits of annotation.
- There is little likelihood of automatic aggregation.
- While reusability is claimed as a good it might act as a means of monopolizing a market.
- LOs can be seen as tied to a pedagogy which sees people as simple information acquirers.
- LOs are often generic, reproducible, standardized products largely for passive consumption by individuals.
- LOs are mainly for individual consumption.

2.2 Learning services

While LOs are annotated with metadata which is principally intended to facilitate discovery, for us, true learning requires the ability to situate a thought in its context within or across disciplines as part of a narrative, a scientific controversy, or an analogical argument. We therefore recognize a variety of possible learning services such as sense-making, structurevisualization, support for argumentation, novel forms of content customization, novel mechanisms for aggregating learning material, and so on, we will discuss a service which contextualizes via representations of controversies. This service provides sense-making support and, since it represents controversies, a means of visualizing an argument structure which could, in principle be extended by the learner as part of the knowledge community.

To provide learning services, we need: a set of ontologies; a set of representations of knowledge which use some of these (Knowledge Charts); and, a set of tools (including tools for Knowledge Navigation and supporting Knowledge Neighbourhoods).

2.3 Ontologies

Learning services require three main types of ontologies: for topics or domains, community-oriented knowledge structures (Knowledge Charts) and learning communities.

Ontologies for domains abound. However, we are particularly interested in the concepts which are regarded by community members as the most salient and foundational for a domain. In physics this would be notions such as relativity. In Semantic Web studies this would be notions such as ontology. The ontologies for Knowledge Charts are less plentiful. Knowledge Charts include a range of high level representations of the most important or most controversial knowledge in a domain. For instance, we could have representations of the processes which underlie fossilization in palaeontology or the current controversy about global warming in climate science. We have identified three main sorts of Knowledge Chart: debates or controversies; narratives; and analogies. Each of these requires both a structural and a domain ontology. Debates require concepts such as *claim*, *ground*, *evidence* and *theoretical backing* while stories need *characters*, *events* and *motives*. Particular knowledge structures will also make use of ontological primitives such as *atmosphere*, *gas*, *pressure* and *energy* for the climate domain.

Finally, we need detailed ontologies for types of community and for the different roles individuals play in them. Thus there are communities with hobbyist interest in archaeology or the music of Mozart or professionals concerned to learn about the latest surgical techniques for hernia repair. Within these communities, individuals have different *statuses* (some are centrally concerned with the body of knowledge which the community creates and preserves while others have a more casual interest) and *roles* (some are teachers as well as contributors to the body of knowledge while others are solely consumers of knowledge).

2.4 Knowledge Charts

A Knowledge Chart is a partial, ontology-based representation of a story or a controversy about a topic for the purpose of supporting understanding. A knowledge chart normally makes use of one or more domain models providing the domain-level knowledge required for its formulation.

In Laurillard's work [Laurillard, 2002] on the use of learning technologies she identifies two important characteristics of learning: that it requires a means of seeing structure as well as the relations among structures. We thus need to provide a typology of knowledge charts (as indicated above) as well as a means of linking among them. This latter will be provided firstly by creating static and dynamic links from structure to structure (based on ontology mappings) and by using tools (see below) to carry out the linking.

As we have seen already we consider three types of knowledge chart to be crucial: debates/controversies, narratives and analogies.

Debates/controversies (most relevant for the SciControversy Learning service) are structured exchanges of positions, factual statements, rebuttals, attacks and so on. Controversies may be seen as a special sort of debate in which the exchanges are aimed at testing the validity of particular theoretical positions. Scientific controversies are a means to test and explore theoretical positions which are not widely accepted. For instance, Wegener's theory of continental drift was the topic of a scientific controversy in the last century. It is typical of controversies that they reach some sort of closure. No one now doubts that tectonic plates exist (al-

though some still argue about notions such as Darwinian evolution).

Narratives as we view them are the high level stories or meta-narratives which a discipline tells itself. For example, archaeology sees itself as currently a highly specialized, professional disciple concerned typically with access the 'archaeological record' rather than the discovery of buried treasures. It became this as a result of pioneering endeavours by individuals such as Worsaae in Denmark who moved the profession away from poorly thought out and experimentally inadequate excavations.

Analogies may be taken as a form of argument in which a discipline proceeds by mapping its state onto results or theories in some other discipline and using these to derive new results itself. For example, social sciences such as psychology often work by employing analogies from other, *harder* disciplines. For instance, cognitive psychology is a discipline which in large part derives its models from a central analogy between the working of the brain and the mechanisms of computers.

2.5 Tools

A whole range of ontologically-informed tools will be needed for the creation of learning services and knowledge structures as well as for the nurturing and support of knowledge communities. Central to these is the means of navigating knowledge structures — the Semantic Browser and the means of creating and updating the knowledge structures the Knowledge Chart (KC) constructor.

Unlike most of the tools and representations mentioned in this position paper, the Semantic Browser already exists. The Magpie Semantic Browser [Dzbor, 2004] has been developed in the Knowledge Media Institute as a means of accessing complementary material initially for students interested in learning about climate change and prediction. Essentially, it works as follows. The community or resource designer provides an ontology for a domain (for example, for climate science). At the same time a set of learning services is created and made available to the Magpie tool. These services can be as simple as glossary-lookups or as complex as simulators for some aspect of climate science. The designer provides mapping between concepts and services. The user of Magpie can decide which parts of the ontology to concentrate on. Magpie finds textual elements in the current web document which match the concepts and highlights these. When the user selects one and right clicks, a menu provides access to the range of associated services.

While the tool is already available, it will need to be modified to some extent to enable it to navigate through Knowledge Charts since these include graphical as well as textual elements. However, the main points to be made are:

- Magpie uses ontologies to construct links dynamically.
- Therefore it does not require pre-annotated resources including LOs.

• It can be extended with arbitrary learning services.

The KC constructor does not exist as yet. However it is our intention that it should be modelled on the wiki interactive web page creation so that communities can browse and create Knowledge Charts using the same combined Browser and Constructor.

Support for Knowledge Neighbourhoods is also an active research topic. While we already have a range of tools for supporting discussion, in our view, what is needed to nurture and support communities are, firstly, a range of environments which 'understand' community dynamics, i.e., which are underpinned by community ontologies, and, secondly, a range of tools which are part of these environments and which allow the collaborative construction of knowledge, knowledge structures such as Knowledge Charts, and learning services.

2.6 The SciControversy Learning service

We can imagine that our learner is reading a web page/document/learning object on climate change as part of some course on environmental studies. While some mention is made of alternative and competing viewpoints, this is not dealt with fully in the text. As she reads, our semantic browser indicates portions of the text with which it has associated services. In this case it can offer a service which displays an interactive view of the scientific controversy about global warming.



Figure 1 A schematic Knowledge Chart for the global warming controversy

Figure 1 represents this controversy in its barest essentials: a real Knowledge Chart for controversy would be much more complex. The figure shows two levels of Knowledge Chart. Level 1 shows the structure of an argument linking CO2 rise to climate change. Level 2 shows part of the ongoing scientific controversy about this linkage. If the learner clicks on the *Lomborg Sceptical Environmentalist* node, this will open up to provide a more detailed version of Lomborg's argument. Since Lomborg's argument about models is based on a view of what statistical models can do, the learner can now opt to follow a link to either a description of statistical models or a deeper view of Lomborg's argument here.

And so on. At each point in the debate model, the learner can access the original web resources of which the model is a summary. Of course, any new document or Chart could have further Knowledge Charts associated with it, which the learner can pursue in turn.



Figure 2 Our approach to delivering learning services

Figure 2 shows how the main components of our approach relate to each other to form the basis for delivering learning services.

2.9 Conclusion

We are currently working on realizing the framework outlined above. Our main aim is to have the SciControversy Learning service implemented within the next few months. At the same time we are trying to define a typology of possible learning services which would be educationally relevant and show the potential of the Semantic Web.

When it is operational, our approach will avoid the reductionism inherent in learning objects and related approaches and support users in making connections, in engaging in critical analysis, in locating the right knowledge and in making sense of pedagogic narratives. It thus stands a better chance of producing the sort of critical thinker able to deal with the complexity of the material available in any future knowledge based society.

References

- [Laurillard, 2002] Diane Laurillard. Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies. RoutledgeFarmer, London, 2002.
- [Dzbor, 2004] Martin Dzbor, Enrico Motta and John Domingue, Opening Up Magpie via Semantic Services, In Proc. of the 3rd Intl. Semantic Web Conference, November 2004, Japan.

Specifying Learning Object-Based Goals and Capabilities with WSMO

José-Manuel López Cobo

ATOS Origin Albasanz, 12 28037, Madrid, Spain jose.lopez@atosorigin.com

Miguel-Ángel Sicilia

Computer Science Department. Polytechnic School University of Alcalá, 28871 – Alcalá de Henares, Madrid, Spain msicilia@uah.es

Sinuhé Arroyo DERI-Innsbruck Technikerstrasse, 13 A-6020 Innsbruck, Austria sinuhe.arroyo@deri.org

Abstract

The Web Service Modelling Ontology (WSMO) allows the definition of Semantic Web Services in terms of capabilities, interfaces using mediators to link them to the goals that a user might have when consulting a service. The concept of pre- and post-conditions embodied in that capabilities have a correlate in the technique of learning object *Design by Contract*, pointing out to the adequacy of using WSMO to model flexible learning object search and discovery processes. This paper describes how WSMO can be used to specify goals and capabilities for learning object selection, using a subset of LOM as a case study.

1 Introduction and Motivation

Web Services are self-contained, self-describing software applications that can be published, discovered, located and invoked across the Web, and using standard Web protocols. Nowadays, the basic support for Web Services is provided through the SOAP, WSDL and UDDI specifications, which address message format, service description and service publishing and lookup, respectively. Nonetheless, this basic support still provides limited help in automating configuration and combination, which has fostered proposals like DAML-S and WSMF that employ Semantic Web technologies for service description and related aspects. The *Web Service Modeling Framework* (WSMF) [Fensel and Bussler, 2002] provides the conceptual model and related tools required to describe complex Web Services in a context of decoupled and scalable components [Arroyo *et al*, 2004].

The *Web Service Modeling Ontology* (WSMO)¹ refines the framework of WSMF providing a formal ontology and formal language. WSMO allows the definition of ontologies comprised by four building blocks: concepts, relations, axioms and instances. Then, these ontologies can be used to describe:

- Goals. A goal specifies the objectives that a client may have when he consults a Web Service.
- Web Services, described in terms of mediators, capabilities and interfaces.

Goals include the definition of post-conditions, that describe the state of the information space that is desired in terms of axioms, and Web Services include *capabilities*, which are in turn described by axioms representing preconditions, postconditions, assumptions and effects. Preconditions define requirements on the input, and post-conditions describe the outputs of the system (in case that preconditions was met in the invocation). Assumptions are previous conditions not limited to the inputs, but to the state of the world

¹ <u>http://www.wsmo.org</u>

in a broad sense, and effects describe such state of the world after the call. All these information items support the description of requirements and outcomes in a logical form. Concretely, WSMO has adopted the *F-Logic* [Kifer *et al.*, 1995] for their representation.

The recent approach of learning object *Design by Contract* (DBC) [Sicilia, 2003] has highlighted that a contractual approach using assertions in the form of pre- and postconditions leads to clearer semantics oriented to automation of the targeting and delivery of learning objects, taking into account the characteristics of the learner, of his/her interaction means and of the overall learning context.

The similarity of the WSMO and LO-DBC has leaded us to exploring the possibility of providing richer semantics to learning object contracts by using WSMO facilities. Concretely, the WSMO approach seems adequate for expressing learning object selection services. Such a service could be implemented as a WSMO Web Service, using as ontology some adapted learning technology schemas. The overall architectural issues for Web service-based e-learning systems has been addressed elsewhere [Xu and El Saddik, 2003 and Blackmon and Rehak, 2003], so that here we focus on ontology-based descriptions that provide a higher level of definitional fexibility to existing architectures.

Two scenarios have been identified in the direction described. First, the post-conditions section of the learning object contracts could be used as goals used in requesting appropriate learning objects. Second, "Instructional Services" (IS) could be considered as a specific kind of learning object implemented as a WSMF Web Service (WS), enabling scenarios of constraint-based search with the following pattern:

- A requester or client *R* (e.g. a Learning Management System or an agent) is looking for learning objects (being those simple contents or complex learning activities) for a given need.
- *R* is able of expressing the characteristics of the need in terms of: the characteristics of the user *U*, the characteristics of the interaction devices and software *D*, and the learning outcome desired *O*, e.g. in terms of expected acquired competencies.
- *R* uses a service *P* to find appropriate learning objects by sending the service *U*, *D* and *O*. The context of learning *C* can also be considered.
- P may check as preconditions that the devices and overall type of learner matches its current capabilities.
- *P* begins the execution of the service.

It should be noted that the non-synchronous nature of the WSMF as described in [Fensel and Bussler, 2002] allows for diverse implementations of such service, including those in which not all the input data is transferred at the beginning of the execution of the service, but concurrent with it. This enables smart approaches to selection without the need of completely transferring U, D and O at a time. In addition, being able to check some constraints as preconditions has the advantage of preventing execution in case that the ser-

vice is not capable of providing the service due to an *a priori* restriction.

The rest of this paper is structured as follows. Section 2 describes the basic elements of an ontology for describing simple learning object metadata and goals based on it. Section 3 sketches how Web Services mapping such goals can be described through an example. Finally, Section 4 provides some conclusions and future research directions.

2 Describing Learning needs as WSMO Goals

A basic ontology describing some essential metadata items in LOM could be described through the following definitions in WSML, which correspond to the elements depicted in Figure 1. It basically addresses language (which should be matched to that of the user described by U), cost (dependant on the context C of the scenario), basic technical requirements (part of D) and classifications (a basic form to express intended outcomes O). Some instances are defined for illustration purposes.

```
ontology http://www.uah.es/.../lom4flogic.wsml
namespace
  default=http://www.uah.es/..lom4flogic#,
 dc=http://purl.org/dc/elements/1.1#,
  wsml=http://www.wsmo.org/.../v0.2/20040418#
non-functional-properties
 dc:title "Learning Object Model for fLogic
Ontology"
  . . .
concept currency
 non-functional-properties
    dc:description "Represents the currency in
  which is expressed the cost of a Learning
  Object"
  currencyName oftype xsd:string
  currencyCode oftype xsd:string
concept cost
  non-functional-properties
    dc:description "Represents the cost of a
    Learning Object, expressed as an amount
and the currency in which is referred"
  hasCurrency oftype currency
  amount oftype xsd:float
concept humanLanguage
  non-functional-properties
    dc:description "a humanLanguage is a lan-
    guage in which can be expressed a Learning
    Object"
  name oftype xsd:string
  ISOCode oftype xsd:string
```





concept taxon

non-functional-properties

dc:description "a taxon is a concrete value for a Taxonomy in which a Learning Object can be classified"

idTaxon oftype xsd:string

valueTaxon oftype xsd:string

concept typesOfRequirement

non-functional-properties

dc:description "Defines a set of values which can be taken for a type of requirement"

idTypeOfRequirement **oftype** xsd:integer

labelTypeOfRequirement oftype xsd:string

concept technicalRequirement

non-functional-properties

dc:description "Technical requirements could be any constraint which has to be solved for the correct use of the Learning Object"

typeOfRequirement **oftype** typesOfRequirement

nameOfRequirement oftype xsd:string

minimumVersionOfRequirement oftype

xsd:integer

concept purposes

non-functional-properties

dc:description "Defines a set of values which can be taken for purposes of a LO"

idPurpose **oftype** xsd:integer

```
labelPurpose oftype xsd:string
```

concept classification

non-functional-properties

dc:description "Means of classify a learning object into one specific taxonomy"

taxonPath **oftype** taxon

purpose **oftype** purposes

concept learningObject

non-functional-properties

dc:description "Any digital entity that may be used for learning, education or training"

identifier oftype xsd:string

aggregationLevel oftype xsd:integer

locationURI oftype xsd:string

isClassifiedInto **oftype set** classification

hasCost **oftype** cost

hasTechnicalRequirements of type set technicalRequirement

languages **oftype set** humanLanguage

isPartOf oftype set learningObject

hasPartOf oftype set learningObject

variable LOone, LOtwo memberOf learningObject

variable X,Y memberOf xsd:integer

axiom inverseFunctionBelong

non-functional-properties

dc:description "Integrity Constraint: A learning Object is composed by Learning Objects"

logical-expression

"LOone[hasPartOf hasvalue LOtwo] <-LOtwo[isPartOf hasvalue LOone]."

axiom minimumLevelOfAggregation

non-functional-properties

dc:description "A Learning Object that is not composed by other learning objects has an aggregation level of 1"

logical-expression

"LOone[aggregationLevel hasvalue 1] <-

not LOone[hasPartOf hasvalue LOtwo]."

axiom levelOfAggregation

non-functional-properties

dc:description "A Learning Object that is composed by other learning objects has an aggregation level greater than the aggregation level of the learning objects which composed it"

logical-expression

"LOone[aggregationLevel hasvalue X] <-LOone[hasPartOf hasvalue LOtwo] and LOtwo[aggregationLevel has value Y] and X = Y + 1 ."

comment: instanceDefinitions

instance browser memberOf typesOfRequirement

idTypeOfRequirement **hasvalue** 1

labelTypeOfRequirement **hasvalue** "Browser"

instance operatingSystem memberOf typesOfRequirement

idTypeOfRequirement hasvalue 2

labelTypeOfRequirement hasvalue "Operating
System"

```
instance device memberOf typesOfRequirement
  idTypeOfRequirement hasvalue 3
 labelTypeOfRequirement hasvalue "Device"
instance educationalObjective memberOf pur-
poses
 idPurpose hasvalue 1
 purpose hasvalue "Educational Objective"
instance competency memberOf purposes
 idPurpose hasvalue 2
 purpose hasvalue "Competency"
instance spanish memberOf humanLanguages
  name hasvalue "spanish"
  ISOCode hasvalue "ES"
instance englishUK memberOf humanLanguages
  name hasvalue "englishUK"
  ISOCode hasvalue "en-UK"
instance englishUSA memberOf humanLanguages
 name hasvalue "englishUSA"
 ISOCode hasvalue "en-US"
instance portuguese memberOf humanLanguages
  name hasvalue "portuguese"
  ISOCode hasvalue "PT"
instance euro memberOf currency
  currencyName hasvalue "Euro"
  currencyCode hasvalue "EUR"
```

Goals in these conceptual models are defined by the postconditions required on the learning objects selected.

The overall goal find free learning objects in spanish for WAP devices that tell how to reach Guadalajara" can be expressed as follows:

goal http://www.uah.es/.../goalLO.wsml namespace

```
default=http://www.uah.es/.../goalLO#,
lom=http://www.uah.es/.../lom4flogic#,
dc=http://purl.org/dc/elements/1.1#,
wsml=http://www.wsmo.org/.../20040418#
```

non-functional-properties

dc:title "Obtaining a Learning Object" . . .

used-mediators

comment: At this moment of the work any mediator is needed.

postcondition

axiom findALearningObject

```
non-functional-properties
```

dc:description "The goal postcondition is represented as a fact. It represents that we want to find a free Learning Object in spanish for WAP devices for learn how to get to Guadalajara"

```
logical-expression
"someLearningObject memberOf
lom:learningObject
  lom:isClassifiedInto hasvalues # memberOf
 lom:classification
    lom:taxonPath hasvalue _# memberOf
    lom:taxon
      valueTaxon hasvalue "ddc:/History & Ge-
  ography/Geography & travel/Geography and travel in Europe/Guide to Guadalajara"],
      lom:purpose hasvalue lom:competency ],
    lom:hasTechnicalRequirements hasvalues _#
    memberOf lom:technicalRequirements
    ſ
      lom:typeOfRequirement hasvalue
 lom:device,
      lom:nameOfRequirement hasvalue "WAP"],
    lom:languages hasvalue lom:spanish,
    lom:hasCost hasvalue # memberOf lom:cost
      lom:amount hasvalue 0]]."
```

The outcomes of the goal (once linked to a service execution) are instances of learningObject with a number of constraints. Zero cost and language are represented as direct constraints, while the technical requirement is expressed by matching the collection implicit in Figure 1. The classification is used to match only elements linked to a given element in a taxonomy. It should be noted that using classifications this way results in "hardcoded" requirements and it should be replaced by a more flexible approach that uses concepts instead of strings, enabling eventual inference about subtypes.

3 Describing WSMO Capabilities for Learning Object Repositories

Capabilities offered by Web Services can be described within the same ontological framework. The following simple definition specifies a capability that could eventually fulfill the need expressed in the goal of the previous section.

```
webservice http://www.uah.es/.../ws.wsml
namespace
 default=http://www.uah.es/.../ws#,
 dc=http://purl.org/dc/elements/1.1#,
 wsml=http://www.wsmo.org/2004/d2/#,
 lom=http://www.uah.es/.../lom4flogic#,
  targetnamespace=http://www.uah.es/.../ws#
non-functional-properties
```

dc:title "Online Guide to Guadalajara for WAP devices"

used-mediator

comment: At this moment of the work any mediator is needed.

variable

?LObject memberOf lom:learningObject

?Language memberOf lom:humanLanguage

?Device memberOf lom:technicalRequirement
?Taxon memberOf lom:taxon

?Classification memberOf lom:classification

capability

postcondition

non-functional-properties

dc:description "the output of the service is a learn ing Object created for WAP devices, in spanish and about travel in Europe"

logical-expression

?LObject memberOf lom:learningObject
[

lom:languages hasvalue ?Language,

lom:technicalRequirement hasvalue ?Device,

lom:classification **hasvalue** ?Classifica-

[

```
lom:taxonPath hasvalue ?Taxon]] and
(?Language.ISOCode = "ES") and
```

(?Device.typeOfRequirement = lom.device and

?Device.nameOfRequirement = "WAP")

and

(?Taxon.valueTaxon = "ddc:/History & Geography/Geography & travel/Geography and travel in Europe/").

Learning object types [Sicilia, 2004b] could be integrated in goal and capability definitions directly, simply putting restrictions on type where necessary for filtering out some kinds of objects, since subsumption guarantees that specialized learning objects are directly considered.

4 Conclusions and Future Research Directions

The use of WSMO to describe contract-based learning object selection services has been sketched. A simple ontology and examples have been been provided.

Our current work is on the description of LOM in its entirety, and designing prototype selectors using the ongoing work on $WSMX^2$. Future work should also cover other e-learning specifications (e.g. SCORM, IMS) and also more precise semantics on scenarios involving automation of learning object selection [Sicilia, 2004a].

Ontology mediators could be used as a convenient approach to map different ontologies related to learning outcomes O, different kinds of user models U, or ontologies of devices for technical requirements C, e.g. the FIPA³ one.

References

- [Arroyo et al., 2004] Arroyo, S., Lara, R., Gmez, J. M., Berka, D., Ding, Y., Fensel, D.: Semantic aspects of Web Services, in Practical Handbook of Internet Computing. Munindar P. Singh, editor. Chapman Hall and CRC Press, Baton Rouge, 2004.
- [Blackmon and Rehak, 2003] William H. Blackmon, Daniel R. Rehak. Customized Learning: A Web Services Approach in Proceedings Ed-Media 2003, June 2003.
- [Fensel and Bussler, 2002] Fensel, D. and Bussler, C. The Web Service Modeling Framework (WSMF). *Electronic Commerce: Research and Applications*, 1 (2002) 113{137.
- [Kifer et al., 1995] M. Kifer, G. Lausen, and James Wu: Logical foundations of object oriented and frame-based languages. *Journal of the ACM*, 42(4):741-843, 1995.
- [Sicilia, 2003] Sicilia, M.A., S¶anchez, S. 2003. On the concept of learning object "Design by Contract". WSEAS Transactions on Systems, Vol. 2, Issue 3, pp. 612-617.
- [Sicilia, 2004a] Sicilia, M. A., Pagés, C., García, E. and Sánchez, S. 2004. Specifying semantic conformance profiles in reusable learning object metadata. In Proceedings of ITHET 2004 - 5th International Conference on Information Technology Based Higher Education and Training. Estambul.
- [Sicilia, 2004b] Sicilia, M.A., García, E., Sánchez, S. and Rodríguez, E. 2004. Describing learning object types in ontological structures: towards specialized pedagogical selection. In Proceedings of ED-MEDIA 2004 - World conference on educational multimedia, hypermedia and telecommunications, pp. 2093-2097.
- [Xu and El Saddik, 2003] Zhengfang Xu and Abdulmotaleb El Saddik. "A Web Services Oriented Framework for Wired and Wireless E-Learning Systems". In Proceedings of the Canadian Conference on Electrical and Computer Engineering (CCECE 2003) Montreal, Quebec, Canada, May 4-7, 2003.

² <u>http://www.wsmo.org/wsmx</u>

³ <u>http://www.fipa.org/specs/fipa00091/PC00091A.html</u>

Integration of an Ontology Manager to Organize the Sharing of Learning Objects in a Peer-to-Peer Network

Gautier Bastide, Pierre Pompidor, Danièle Hérin, Michel Sala

L.I.R.M.M.

Computer Science, Robotics and Microelectronics Laboratory 161 Rue Ada, 34392 Montpellier {bastide,pompidor,dh,sala}@lirmm.fr

Abstract

This paper is dedicated to the ontologies management, and more particularly to a tool called Kaon. One of its objectives is to propose the integration of a such application in a peer-to-peer platform. Indeed, the tools which are provided by Kaon can be used for the management of the distributed resources sharing. The integration of Kaon to model and to organize the knowledge in the Edutella network, for example, would allow the realization of more effective search engines. The Kaon platform is already used for the annotations management in Edutella project. It would be judicious to spread it to all the resources that are shared within the Internet network. The use of Kaon would allow to increase the efficiency of searching services thanks to the modeling of the semantic links between the various learning objects which are shared through the Edutella network

1 Introduction

The various search engines which are proposed and implemented within the framework of the Edutella project [Edutella; Nejdl et al., 2002] provide the user with a disorganized list of results. These work in the following way: the user enters a series of keywords or a request written in a language like Datalog or in RDF-QEL-i. Then the application passes on the requests, returns the results and presents them to the user shape of a list of resources which it is possible to consult by select on. Every answer is described by metadata. The problem lies in the fact that the data shown in the screen are not organized.

Furthermore, the relations between the various solutions that are proposed to the user are not valorized within the framework of their visualization. This inconvenience meets itself in the major part of the traditional search engines. It would be more judicious to advance the existing links between the various resources found by the search engine so as to facilitate the reading and the interpretation of the results. This proposition requires the use of an ontology management tool such as Kaon [Kaon; Volz et al., 2003; Bozak et al.,2002] to modelize the links between the resources. A Kaon server must be present in each super-peer of the Edutella network.

A first proposition would consist in showing the results using an ontologies graph which is similar to that presented in one of the interfaces of the Kaon API i.e. OI-Modeler [OI-Modeler, 2002]. The graph presented to the user would be simplified but would allow him to navigate through the various concepts present in the answer. So the use of Kaon to model the links between the resources allows us to increase the efficiency of the search engine and facilitate the legibility of the answers.

As regards the interface, the user must have three possibilities of action at the level of the search for information in the Edutella network:

- · Enter a series of keywords judiciously chosen,
- Load a request written in a language such as Datalog or RDF-QEL-i,
- Load an ontology which can be described in XML or in RDF.

Furthermore, the user must have the possibility of selecting a particular supplier of resources. If any Edutella supplier isn't selected, then the request will be spread to the whole network. The fact of giving the possibility to the user to select one or several Edutella suppliers of resources allows us to restrict the searches. Besides, if he wishes, the user has the possibility of querying a particular supplier. He can also spread his searches to a series of suppliers that he has selected beforehand according to his preferences (if he wants to omit some suppliers).

2 Management of Learning Objects Metadata

The use of metadata is needed for the efficiency of the system of search for information. All the learning objects intended for the sharing must be described by metadata that respect predefined schemas such as those which are proposed by IEEE LOM, IMS or SCORM. In this part, we consider only objectives annotations. There are two possibilities to annotate a document (i.e. To field corresponding to the metadata allowing us to describe this resource). Either it is the creator of the learning objects, which fills fields allowing us to describe it, or this stage is automatically made through tools of automatic annotations. The key points of this part are the maintenance of these metadata and the search for semantic links between the various resources.

The lack of annotations on a learning object could be resolved by analyzing the contents of documents to complete the missing data. This method of distribution of metadata by the analysis of the bibliographical references can be applied to one or several documents selected beforehand. It acts recursively. The point of departure for the application of this method is a document. Then, the method pursues its search in the resources quoted as references of the analyzed document. To avoid problems, it is thus necessary to introduce the notion of life cycle (Creation of an indication TTL). It allows one to stop the process beyond a certain number of documents which are gone through and so to avoid the infinite loops. Indeed, that this method works correctly, it is indispensable to have a core of resources which are manually annotated by their owner (Figure 1).



Figure 1. Creation of a core of annotated resources

At first, it is necessary to analyze the semantic contents of the document in order to establish a hierarchy of this component. In fact, the objective is to extract the ontology of the document. The metadata which are stored in the nodes allow to describe the various parts of the resource which is analyzed. A semantic link that must be created between two resources which are referenced in the same chapter of a document.

The completion of annotations contents by this method of metadata propagation brings the creation of two main cases:

• If the resource is not annotated.

In that case, it is necessary to look for the references supplied by the authors. They are generally situated at the end of the document in the bibliography. Then, it is enough to retrieve the corresponding information. Thus, the annotations of a resource will be made thanks to the metadata describing the documents, which are referenced. This operation necessarily arouses the creation of a semantic link between both resources. In that case, the reach of the method is global because the completion of metadata is made at the level of the document. A part of the information so harvested serve for describing the whole resource. It is necessary to retrieve the part of the ontology of the resource 1 on which is going to be put the metadata which correspond to the resource 2.



Figure 2. Annotation of learning objects: case n°1

• If the resource is already annotated.

In that case, it is enough to parse the document and to look for the references of the learning objects which are not annotated. When a resource is found, it is possible to fill in the information contained in its metadata with the corresponding data in the chapter being analyzed. A semantic link is then created between these two learning objects. In that case, the reach of the method is local because only a part of the learning object can be described by the metadata so collected. Only the information which corresponds to the chapter of the annotated resource can be used to describe the new document.

All the metadata can't be propagated. Indeed, only those which correspond to the contents of the resource as the subject, the description, the keywords, the references can be duplicated. Other metadata (author, date of creation or publication, language) are not used within the framework of this method of propagation. The quantity of keywords obtained by this method can be quickly very important. Furthermore, some of these words can have no link with the resource for which we try to annotate. The problems so met in this method of propagation of the metadata are owed to the evolution and to the revision of the ontologies. To manage all these problems, it is indispensable to set up certain number of operations intended to manipulate the ontologies relative to the description of the documents for which we try to annotate.



Figure 3. Annotation of learning objects: case n°2

A first solution of this problem would consist in making the intersection of keywords relative to the resources pointed by this one. This method would allow to reduce the quantity of keywords being of use as description to the resource and would get a better result in their links with the document. This strategy must not be only applied to keywords but also to references. This method using the intersection of keywords does not allow to resolve all the problems relative to the fill of the metadata.

The methods of generalization on the local ontologies or external can also be a solution of the metadata management problems.

Another solution of the evolution and the revision of ontologies would consist in using logic languages which allow to create inferences. The aim is to manage the data inconsistency.

Another problem which arises in the annotations management concerns the validity of the information contained in the metadata allowing us to describe the learning objects. Indeed, an annotation can be false (that is inappropriate in the resource or inaccurate). The problem can be due to incorrect data entry made by the resource's owner or to an error at the level of the automatic creation of annotations. It is possible to solve this problem by the use of a system of level-headednesses attributed to every resource's annotation. The users of the Edutella network have to have the possibility of acting on the value relative to the validity of the information contained in the metadata describing a resource. These operations can be made through a check by the users at every consultation of a resource. If an annotation is considered incorrect, either the owner of the resource is warned then he can operate modifications, or the mechanism of automatic creation of annotations computes again the information contained in the metadata of the resource.

3 Management of the Ontologies at the Level of Super-node

The Kaon platform is used to manage the ontologies contained in the Edutella network. A Kaon server must be present on every super-node of the network. It is intended to manage the resources of its cluster and is necessary to know the big subjects of the other groups of peers.

The learning objects which are stored in the peers of the Edutella network are described by metadata and are grouped together in respect to ontologies. Indeed, they are considered as instances of concepts or sub-concepts. Every time a peer gives a resource, it must be declared to the Kaon server situated in the super-node to which it is connected. A new instance which square with the new resource will be created. The Kaon API is then going to be in charge of integrating the resource into the existing model by taking into account semantic relations.

First of all, the Edutella supplier peer creates a description of its capacities as well as resources which it suggests sharing in the form of an ontology. Indeed, the resources which the peer suggests supplying must be described by metadata which respect standards of type IEEE LOM or SCORM and must be organized in the form of ontologies containing concepts, sub-concepts and instances. These correspond to the resources. Every instance possesses properties described by metadata and have relations with the other entities of the model. Every Kaon server has to manage at least two ontologies (figure 1): one allowing us to store the characteristics of the peers which are contained in the cluster as well as super-node of the Edutella network and another designed to store the metadata being of use as description to the learning objects and allowing the modeling of the semantic links between these. The information concerning the capacities of every Edutella supplier is stored in the Kaon server.



Figure 4. Example of an Ontology Model for Super-peers' Management

The management of the knowledge is made on three levels which are interconnected:

• The first level: Learning objects.

It is the lowest level of the data model. It concerns the storage of the learning objects without metadata.

• The second level: metadata

This level contains the descriptions of the learning objects. The metadata generally follow a schema which is defined by standards such as IEEE LOM or SCORM.

• The third level: ontologies

This level contains the representation of the concepts, the sub-concepts and the links. This part allows one to organize and to manage components contained in the previous two levels. The instances of the ontology model contain the metadata (Level 2) which are used to describe the learning objects (Level 1). The learning objects (Level 1) are described by metedata (Level 2) and regrouped by ontologies (Level 3).

The main relations which arise in ontologies of learning objects are the following ones:

• Be_a_part_of

The relation $Be_a_part_of(x, y, i)$ means that x is a part of y. Thus, it is necessary to know the resource x if we want to study the resource y.

The value i represents the validity index of the relation (i.e. Reliable indication of the relation). In fact, it is a weight. This value has the same signification in the three following relations.

• Be explained by

The relation $Be_explained_by(x,y,i)$ means that the resource x can be explained by the resource y.

• Be required

The relation $Be_required(x,y,i)$ means that the resource x needs the resource y as pre-required.

• Be_suggested

The relation $Be_suggested(x,y,i)$ means that it is better to know the resource y before making the learning of the resource x. If you are interested in the resource x you can use it independently of the resource y. You don't have to know both resources.

The references supplied by the authors must be used to create semantic links between two resources. If a link doesn't exist between these two resources, a relation of type "Be_suggested" will be created. In this case, it is indispensable to create a relation which is the most flexible possible when we don't know the exact kind of link between two resources. Moreover, the kind of relation must be able to be modifiable by the authors of the resources. The goal of this operation is to improve the semantics of the model.

One of the most important points in the management of ontologies lies in their maintenance and in their evolution. Indeed, the model must be able to evolve every time that a new supplier connects to the network or every time that a peer share a new resource. The ontologies have to remain viable at all times. For that purpose, the Kaon API possesses numerous features which allow us to resolve this kind of problem. The updates must be made automatically while verifying the integrity of the model. It is indispensable to be able to introduce new concepts or even new relations into the existing model. For that purpose, it is necessary to be able to discover that the new resources really introduce new concepts and to determine their positions within the model. These operations must be automatically realized through syntactical analysers.



Figure 5. Example of an ontology model for mathematics [WebLearn, 2004]

4 Scenario of Connection of a Peer Supplier

The peer which tries to join the Edutella network looks for the super-node which appears to be the most suited to the fact that it suggests supplying as learning objects (Figure 6). For that purpose, the supplier peer is going to send a message to any super-nodes (which is used as a point of entry to the Edutella network). It will transmit the messages that contain the information about the new learning objects which are proposed to other super-peer of the Edutella network. When the super-node comes up to the expectations of the supplier peer, this one is going to be bound there. Then it will be sent to it all the information which it has collected beforehand. Once these operation made, the Kaon server which is situated on the super-node selected is going to have to update its model and so integrate the ontology proposed by the new peer. The lexical data must be stored in the Kaon server. The fusion of these ontologies is realized by the Kaon API and is made thanks to the tools that it possesses.



Figure 6. Connection of a New Edutella Supplier Peer Scenario

5 Scenario of Search for Information in the Network

Here is an example of scenario of search for information in a Edutella network using the Kaon platform to describe the learning object and model the links between the entities of the model.

The user enters a series of keywords, either loads a request written in languages such as Datalog or RDF-QEL-i, or even an ontology which can be described in XML or RDF. If the user loads an ontology, the search engine takes care to complete it by adding instances, sub-concepts as well as relations between the various entities of the model.

The peer translates the request into the basic language of a Edutella network i.e. RDF-QEL-i. Then it passes on it in its super-node. This is then going to query the Kaon server through this language of request. Thanks to the data stored by the Kaon server, it is possible to determine which peer of the cluster or which other super-node contain the information that we are looking for. Thanks to the data so harvested and to the dynamics tables of routing, the request can be passed on to the peers who may supply a correct answer.

If no good result is obtained, the request is passed on in the other nodes of the network.

Otherwise, it is going to send back the result of the request to the peer of origin in the form of a graph of RDF data. In this way, when a super-peer receives the request, it requests to the Kaon server which manages its resources and to send back if possible a RDF graph in reply.

When the peer receives the answers, it collects and reorganizes the graphs obtained so as to be able to present to the user an unique graph allowing one to show the links between the resources and the various concepts included in the result of his request.



Figure 7. Scenario of Kaon Integration in an Edutella Network

The scenario follow this model (Figure 4):

Step 0: Stage of initialization (cf. 4. Scenario of connection of a new supplier peer). The peers have to communicate with the Kaon server of the super-node to be able to model the semantics links between the resources which are shared at the level of the cluster. The resources must be described as instances of concepts or sub-concepts.

Step 1: The user creates or loads a request which is going to be spread in the Edutella network. Several methods are possible: either the user supplies a list of keywords judiciously chosen, or he directly loads a request which he has beforehand written in a language such as Datalog either RDF-QEL-i with i lower than 5, or he defines an ontology (Composed of concepts, by sub-concepts, by properties and by instances) written in a language as RDFS or OWL. Another possibility would be to propose to the user certain number of subjects and sub-subjects that would allow him to obtain more general information. Furthermore, the user has to have the possibility of selecting the language in which he wishes to obtain the answers (The lexical information must be stored in the user's machine).

Step 2: The request so formed is translated into a language (RDF-QEL-i) which is understandable by all the components of the Edutella network (Peers and super-nodes). The data are passed on in the super-node which is under the responsibility of the peer of the original request. This is going to be retrieved by the administrator of request of the super-node which then undertake to do its treatment.

Step 3: The request retrieved by the administrator is then going to be modified to be able to request to the Kaon server containing all the information necessary for the management of the ontologies modeling the cluster. Kaon can resolve requests in the RDF-QEL-i language.

Step 4: The Kaon server returns the result of the request to the super-node's administrator to know the list of peers of the cluster which may resolve the request. Two cases are then possible:

- If no result is found by the server, then the administrator of requests consults the dynamic tables of routing (*Step 7b*) to obtain the address of the super-node being able to solve the request. This request is going to be passed on in the super-peer so found (*Step 8*).
- Otherwise, the request is passed on to the peers which are so found (*Step 5a*, *Step 5b*). The peer are going to resolve locally the request and to send back the result under the shape of a RDF graph to the super-node (*Step 6a*, *Step 6b*). The data thus retrieved is analysed by the administrator. Two cases appear:

- Either, the results obtained are satisfactory (According to a certain number of pre-defined criteria),

In that case, the data which are collected by the administrator is sent to the original peer (*Step 7a*).

- or the results do not allow us to propose a correct answer.

In that case, the request is passed on to the other super-node of the Edutella network thanks to the dynamic tables of routing (*Step 8*).

Step 9: The results of the request are then passed on to the broadcasting peer of the request to be collected. The transmission of data answers is made through super-peers of the Edutella network. The results can be presented to the user. They are organized by ontology. The semantics links between the resources are shown in such a way that the user can find quickly what he is looking for.

The integration of the Kaon platform inside the Edutella super-peer allow us to improve the searching service. This tool is presented as a complement to the dynamic routing tables. The learning objects are described by metadata and are grouped by ontologies. The goal is to retrieve it and locate it easily and efficiently.

6 Conclusion

This article proposes the use of an ontology administrator like Kaon to modelize the links between resources which are shared within a peer-to-peer network like Edutella. The use of such a tool allows to improve the service of search for information of the Edutella peer-to-peer network. Indeed, the knowledge concerning the resources contained in the super-nodes of the network are grouped together by ontologies so as to be able to localize them more easily and more quickly. The scenario proposed in this paper describes the various phases relative to the implementation of an ontology management tool within super-peers. The integration of this type of tool to manage the resources is an undeniable contribution in the field of the information search within a totally decentralized environment. It allows one to model better knowledge contained in a network as well as their relations. This tool is presented as a complement to the dynamic tables of routing in the optics to more easily and quickly track down and localize one or several points of information which the user is looking for.

References

[Edutella] Project Edutella. http://www.edutella.jxta.org/

- [Nejdl et al., 2002a] W. Nejdl, B. Wolf, C Qu, S. Decker, M Sintek, A Naeve, M. Nilsson, M. Palmér, T. Risch. EDUTELLA: A P2P networking infrastructure based on RDF. 11th International World Wide Web Conference, Hawaii, USA, May, 2002.
- [Nejdl et al., 2002b] W. Neijdl, B. Wolf, S. Staab, J. Tane. EDUTELLA: Searching an Annotating Ressources within an RDF-based P2P Network. Semantic Web Workshop 2002 Honolulu, Hawaii, May 7, 2002.
- [Nejdl et al., 2003] W. Nejdl, W.Siberski, M. Sintek. Desing Issues and Challenges for RDF- and Shema-Based Peer-to-Peer System. ACM SIGMOD Record, Volume 32, Issue 3, September 2003.

[Kaon] Project Kaon. http://www.kaon.semanticweb.org/

- [Volz et al., 2003] R. Volz, S. Staab, D. Oberle, B. Motik. KAON SERVER - A Semantic Web Management System. Twielth Internationnal World Wide Web conferance, WWW 2003, Budapest, Hungary, May 20-24, 2003. ACM, 2003.
- [Bozsak et al., 2002] E. Bozsak, M. Ehrig, S. Handschuh, A. Hotho, A. Maedche, B. Motik, D. Oberle, C. Schmitz, S. Staab, L. Stojanovic, N. Stojanovic, R. Studer, G. Stumme, Y. Sure, J. Tane, R. Volz, V. Zacharias. KAON Towards a large scale Semantic Web. E-Commerce and Web technologies, Third Internationnal Conference, EC-Web 2002, Aix-en-Provence, France, September 2-6, 2002, Proceeding, volume 2455 of Lecture Notes in Computer Science, pages 304-313, Springer, 2002.
- [Kaon, 2004] KAON: The Karlsruhe Ontology and Semantic Web Framework. Developer's Guide for KAON 1.2.7, January 2004. <u>http://www.aifb.unikarlsruhe.de/WBS</u>
- [OI-Modeler, 2002] OI-Modeler User's Guide. November 27th, 2002. <u>http://www.aifb.uni-karlsruhe.de/WBS</u>
- [WebLearn, 2004] Workshop organized by the action « Semantic Web and E-Learning » of Kaleidoscope NoE, Paris, May 3-4, 2004.

Description of an Instructional Ontology and its Application in Web Services for Education*

Carsten Ullrich

German Research Center for Artificial Intelligence Saarbrücken, Germany cullrich@dfki.de

Abstract

In the last years, important steps have been undertaken to bring the e-learning web to its full potential. In this paper, I describe an ontology that can serve as a further step in this direction. The ontology captures the instructional function of a learning resource, in other words, its "essence" from a teaching/learning perspective, an aspect not vet covered by learning object metadata standards. It offers the well-known advantages of ontologies: it can provide humans with a shared vocabulary and can serve as the basis for the semantic interoperability for machines. The article motivates the need for such an ontology and describes several educational Web services that can benefit from it. To exemplify the generality of the ontology, the article describes how the ontology can be mapped onto several knowledge representations currently used in e-learning systems.

1 Introduction

1.1 Motivation

Imagine Eva, a teacher, preparing a lesson. Yesterday, in class, she introduced the concept of *gravity*. The learning progress analyzer of her pupils noticed that some kids were not able to apply the new knowledge. Therefore, Eva orders her authoring tool to search the web for examples and interactive exercises that specifically train the application of gravity. On the web-page of Anton, a fellow teacher, the tool finds the necessary resources, and, in addition, an abstract description of an instructional strategy which is based on a real-world-problem teaching approach, especially appropriate for learning physical concepts. The tool shows its findings to Eva and offers to feed them into a course generator. Eva accepts, and the course generator assembles a curriculum that follows the instructional strategy and is adapted to the knowledge of her class. The next day, her pupils work with the new learning materials. Depending on their personal interests, a browsing service adds links to learning resources that provide

real-world application of gravity. For Clara, it adds a link to a NASA site that describes the relation between gravity and space ships; Bert is offered a page describing airplanes. After the lesson a data mining service analyzes the paths of the pupils and makes suggestions to Eva what content to include permanently in the course.

In the last years, important steps have been undertaken to achieve such a scenario. The development of sophisticated web-based e-learning systems with a wide range of learner support on the one hand and integrating architectures on the other hand could sum up to a critical mass that brings the Web to its full e-learning potential. In this paper, I describe an ontology that can act as a binding glue between different systems and services and serve as a basis for interoperability with respect to instructional matters.

In the remainder of the introduction, I will briefly summarize the need and scope of the here proposed instructional ontology and describe the shortcomings of today's e-learning standards. The subsequent section portrays a number of potential educational Web services that can profit from such an ontology. Section 3, the main part of the paper, describes the ontology in detail. It is followed by a proof of concept illustrating how the ontology can be mapped onto three frequently used knowledge representations. The paper concludes with a description of related work.

1.2 Benefits of Using an Ontology

An ontology expresses a common understanding of a domain that serves as a basis of communication between people or systems. The need for ontologies has been widely recognized (for a recent discussion see [World Wide Web Consortium, 2004]) therefore I will only summarize some expected benefits. In education, widespread appliance of such a shared instructional vocabulary offers advantages for teachers and learners. A more accurate search for learning resources, made possible by the explicit instructional function, leads to better reuse and less duplication, hence faster authoring of curriculums. By seeking instructionally appropriate learning material, learners can bridge knowledge gaps more efficiently.

The pedagogically relevant information of the ontology also brings forth better Web services. It can increase the accurateness of a service because at design time, a Web service developer can foresee different functionality depending on the type of the resource. For most educational services,

^{*}This publication was generated in the LeActiveMath project, funded by the European Community (IST-2003-507826). The author is solely responsible for its content.

the information whether a resource contains a definition or an example will be of use. Similarly, service composition is enhanced. For instance, a requester service can require different actions from a provider depending on the instructional type of a resource. Furthermore, interoperability is eased. Then, in theory, each system can provide its own specialized service and make use of the services offered by others.

1.3 Scope of the Ontology

The ontology described in this paper provides a vocabulary that captures the "instructional semantics" of a virtual or text-book learning resource. In general, each learning object serves a particular pedagogical role. These roles are reflected in the classes of the ontology.

The ontology of instructional objects covers instructional theories only partially, namely those parts that describe the learning materials independently of a specific learning context. Hence, it does not encompass learning goals. Learning goals are one primary cause why in a specific context an instructional object is selected, but as instructional objects can serve to attain various goals and one goal of the ontology is re-use, I excluded learning goals and other context-specific information.

A concrete example illustrates best the entities described by the ontology. Figure 1 presents several learning resources (taken from [Bartle and Sherbert, 1982]), clearly divided into several distinct paragraphs. Each paragraph serves a particular *instructional role*. The first two introduce two concepts (a *definition* and a *theorem*), the third provides *examples of applications* of the concept, and the last one offers to the learner *activities to apply* the concept. These reseource can be assembled by an author or a Web service to compose a page in a course (as it was done here).

The ontology provides a standardized vocabulary of the instructional function of a resource. Additionally, the ontology can be used to partially describe the instructional strategy that underlies the composition of a curriculum. A (simplified) strategy for the example in Figure 1 is the following: To introduce a new concept x, present learning resources in the order *concept x, examples for x, exercises for x.*

Seminal work regarding ontologies and instructional design was done by Mizoguchi. [Mizoguchi and Bourdeau, 2000] lay out how ontologies can help to overcome problems in the domain of artificial intelligence in education. The work presented in this article was designed to be a step towards this goal.

1.4 Shortcomings of Today's E-Learning Standards

Today's standards prove the (commercial) importance of reuse and interoperability of learning material. For this article, particularly relevant standards are IEEE Learning Object Metadata (LOM, [IEEE Learning Technology Standards Committee, 2002]), and IMS Learning Design (LD, [IMS Global Learning Consortium, 2003]). LOM's educational category allows a description of resources from an instructional perspective. Possible types of learning resources are, among others, *Diagram, Figure, Table, Exercise, Narrative* **2.3.7 Definition** Let $a \in \mathbf{R}$ and v > 0. Then the *s*-neighborhood of *a* is the set $V_{\varepsilon}(a) := \{x \in \mathbf{R} : |x-a| \le \varepsilon\}$.

For $a \in R,$ the statement that x belongs to $V_e(a)$ is equivalent to either of the statements

$$-\varepsilon < x - a < \varepsilon \quad \Longleftrightarrow \quad a - \varepsilon < x < a + \varepsilon$$

(See Figure 2.3.2.)

FIGURE 2.3.2 An e-neighborhood of a.

a-0

2+8

2.3.8 Theorem Let $a \in R$. If x belongs to the neighborhood $V_e(a)$ for every $\varepsilon > 0$, then x = a.

Proof. If a particular x satisfies $|x - a| < \varepsilon$ for every $\varepsilon > 0$, then it follows from 2.2.9 that |x - a| = 0, and hence x = a. Q.E.D.

2.3.9 Examples (a) Let $U := \{x: 0 \le x \le 1\}$. If $a \in U$, then let ϵ be the smaller of the two numbers a and 1 = a. Then $V_i(a)$ is contained in U. Thus each element of U has some ϵ -neighborhood of it contained in U.

(b) If $I := \{x: 0 \le x \le 1\}$, then for any $\varepsilon > 0$, the ε -neighborhood $V_{\varepsilon}(0)$ of 0 contains points not in I, and so $V_{\varepsilon}(0)$ is not contained in I. For example, the number $x_{\varepsilon} := -\varepsilon/2$ is in $V_{\varepsilon}(0)$ but not in I.

(c) If $|x - a| < \varepsilon$ and |y - b| < v, then the Triangle Inequality implies that

$$|(x + y) - (a + b)| = |(x - a) + (y - b)|$$

 $\leq |x - a| + |y - b| < 2e$

Thus if x, y belong to the ε -neighborhoods of a, b, respectively, then x + y belongs to the 2 ε -neighborhood of a + b (but not necessarily to the ε -neighborhood of a + b).

Exercises for Section 2.3

```
1. Let a \in \mathbb{R}. Show that we have:

[a]_{a}(a) |a| - \sqrt{a^{2}}, (b) |a^{2}| = a^{2}.

2. If a, b \in \mathbb{R} and b \neq 0, show that |a/b| = |a|/|b|.

3. If a, b \in \mathbb{R}, show that |a + b| = |a| + |b| if and only if ab \ge 0.
```

Figure 1: A page from a mathematics textbook that contains several types of instructional objects.

Text, Exam. In LOM, these values are provided as a list, without taking into account the inherent structure which a representation as an ontology as envisaged in this article would provide. More critical, the LOM types mix instructional and technical information. The first three values of the above example describe the format of a resource, whereas the last three cover the instructional type. They represent different dimensions, hence should be separated. Furthermore, several instructional objects are not covered by LOM (e.g., *definition, example*). IMS LD describes ordered activities in learning and the roles of the involved parties. It does not represent single learning resources and their instructional functions.

To summarize, today's standards do not cover the instructional function of a learning resource, and, in addition, were not designed for the Semantic Web. However, the full elearning potential of the Web will only be reached if Semantic Web techniques are supported. The following section describes several educational Web services and their usage of an ontology of instructional objects.

2 Web Services Using an Instructional Ontology

This section provides several examples of Web services and their possible benefit from an ontology of instructional objects.

Course Generator. A course generator (e.g., [Ullrich, 2003]) assembles learning resources to a curriculum that takes into account the knowledge state of the leaner, his preferences, learning goals, and capabilities. If (third-party) resources are annotated by their instructional function, a course generator can include them in a curriculum. Work in this direction was done in Open Corpus Hypermedia. For instance, [Henze and Nejdl, 2001] propose an approach based on a domain ontology. The additional use of an instructional ontology can lead to a more accurate selection of the resource to be included.

Learner Modeling. A learner model stores personal preferences and information about the learner's mastery of domain concepts. The information is regularly updated according to the learner's interactions with the content. A user model server such as Personis ([Kay *et al.*, 2002]) can use the information about the instructional function of a learning resource for more precise updating. For instance, reading an example should trigger a different updating of the mastery of a concept than solving an exercise.

Browsing services. Services that support the user in browsing through content benefit if the instructional function of a learning resource is made explicit. They can better select and classify the presented objects. Systems that adaptively add links to content ([Brusilovsky *et al.*, 1998]), can decide what link to add and how to classify them more appropriately. Similarly, tools that generate concept maps can better adapt the maps to the intended purpose, both with respect to the selection and the graphical appearance of the elements. A dictionary that provides a view on the dependencies of the domain elements can sort the element with respect to their instructional type.

Authoring support. An ontology of instructional objects assists authors by allowing for better search facilities and by describing an conceptual model of the content structure. It offers teachers with a set of concepts at the adequate abstractness level to talk about instructional strategies. They describe their teaching strategy at a level abstract from the concrete learning resources. Hence, instructional scenarios can be exchanged and re-used. An ontology of instructional objects can additionally support the author by providing an operational model in the sense of [Aroyo and Mizoguchi, 2003] that provides hints to the author, e.g., what instructional objects are missing in his course.

Additional service that can profit from the ontology are, for instance, data mining, interactive exercises, and intelligent assistants.

3 Description of the Ontology of Instructional Items

The goal of this work is to provide an ontology that describes a learning resource from an instructional perspective. The ontology does not describe the content taught by the learning material, e.g., concepts in physic and their structure. Instead, each class of the ontology stands for a particular instructional role a learning resource, for instance a paragraph in a textbook, can play. For some objects, determining the role is straightforward. In most text-books, exercises are distinctly marked. For other objects, it may be less obvious.

In order to provide an ontology that can be applied in a large variety of contexts, it was necessary to analyze a significant amount of sources. Here, sources ranged from text classification ([Mann and Thompson, 1988]), over instructional design theories (e.g., [Reigeluth, 1999]) to instructional oriented knowledge representations which were implemented in e-learning systems (e.g., [van Marcke, 1998; Specht *et al.*, 2001; Pawlowski, 2002; Lucke *et al.*, 2003; Cisco Systems, Inc, 2003]).

In addition to theoretical applicability, an ontology should be easily understandable for authors. Therefore, one design goal was to come up with a limited set of classes which still encompasses all necessary instructional objects. Two teachers and two instructional experts reviewed the ontology, and, besides minor suggestions which were integrated, rated it very positively.

In the following, I will describe the classes and properties of the ontology of instructional objects. Figure 2 shows the class hierarchy. The ontology was implemented using Protégé ([Gennari *et al.*, 2003]).

Instructional Object. "Instructional object" is the root class of the ontology. Several properties are defined at this level: a unique identifier; "learning context", which describes the educational context of the typical target audience; and "field", which describes the field of the target audience. The field of an instructional object can differ from the domain the resource describes. For instance, a mathematical concept can be illustrated by an example from economics or from medicine. An additional slot includes Dublin Core Metadata, e.g., information about creator of the resource. The property "analogous" indicates that an instructional object shares some aspects with another instructional object.

Concept. The class "concept" subsumes instructional objects that describe the central pieces of knowledge, the main pieces of information being taught in a course. Pure concepts are seldom found in learning materials. Most of the time, they come in the form of one of their specializations "fact", "definition", "law", and "process". Albeit concepts are not necessarily instruction-specific because they cover types of knowledge in general, they are included in the ontology because they are necessary for instruction. Learning objects often have the instructional function of presenting a concept.

Concepts rarely stand alone, more often than not they depend on another concept. This is represented by the *dependson* property which has its range the class "concept".

Fact. An instructional object that is a "fact" provides information based on real occurrences; it describes an event or something that holds without being a general rule. An example is "Euclid lived from about 365 to 300 BC". In mathematics, the line of distinction between facts and examples is fuzzy as most facts can be considered as examples, for instance " $\sqrt{2}$ is irrational.".



Figure 2: Class hierarchy of instructional objects.

Definition. A "definition" is an instructional object that states the meaning of a word, phrase, or symbol. Often, it describes a set of conditions or circumstances that an entity must fulfill in order to count as an instance of a class. Examples for definitions are "A group is a mathematical structure consisting of..." and "The middle ages describes the period of time that...".

Law. An instructional object that is a "law" describes a general principle between phenomena or expressions that has been proven to hold, or is based on consistent experience.

Law of Nature. A "law of nature" is a scientific generalization based on observation. Typical examples are Kepler's first law of planetary motion: "The orbit of a planet about a star is an ellipse with the star at one focus.", or Einstein's equivalence of mass and energy: " $E = mc^2$ ". Similar laws of nature exist in biology, chemistry, etc.

Theorem. A "theorem" is an instructional object which describes an idea that has been demonstrated as true. In math-

ematics, it describes a statement which can be proven true on the basis of explicit assumptions. Examples are "The intersection of submonoids is a submonoid", or Gödel's incompleteness theorem.

Process. "Process" and its subclasses describe a sequence of events. The deeper in the class hierarchy, the more formal and specialized they become. A process provides information on a flow of events that describes how something works and can involve several actors. Typical examples are "the process of digestion", and "how is someone hired in a company".

Policy. A "policy" describes a fixed or predetermined policy or mode of action. One principal actor can employ it as an informal direction for tasks or a guideline. Curve sketching in mathematics, for instance, provides a general guideline of how to determine the essential parts of a function. Similar guidelines exist for analyzing a work of literature.

Procedure. A "procedure" consists of a specified sequence of steps or formal instructions to achieve an end. It can be as formal as an algorithm. Typical examples are Euclid's algorithm, or instructions to operate a machine.

Satellite. "Satellite" elements (the name was adopted from [Mann and Thompson, 1988]) subsume instructional objects which are not the main building blocks of the domain to be learned, but elements that provide additional information about the concepts. In principle, concepts provide all the information necessary to describe a domain. However, from an instructional point of view, the satellite objects contain crucial information. They motivate the learner, and offer engaging and challenging learning opportunities. Every satellite object offers information about one or several concepts. The identifiers of these concepts are enumerated in a "for" property.

Interactivity. An instructional objects that is an "interactivity" offers some kind of interactive aspect. It corresponds to the "active" type of interactivity in LOM's educational category. An interactivity is more general than an exercise as it does not necessarily have a defined goal that the learner has to achieve. It is designed to develop or train a skill or ability related to a concept. The difficulty of an interactivity is represented in the property of the same name.

The subclasses of "interactivity" do not capture technical aspects. In general, the way how an interactivity is realized, for instance as a multiple choice question or an erroneous example, is independent of its instructional function. A welldesigned multiple choice question can target knowledge as well as application of a concept.

Exploration. "Exploration" is an instructional object in which the user can freely explore aspects of a concept without a specified goal, or with a goal but no predefined solution path. Cognitive tools ([Lajoie and Derry, 1993],) or simulations are typical examples of an exploration object.

Real World Problem. "Real world problems" are frequently used in instructional design, especially in constructivist theories, e.g., [Jonassen, 1999]. They describe a situation from the learner's daily private or professional life that involves open questions or problems. Solving the problems requires knowledge about a set of concepts. Authentic real world problem are an excellent way of motivating the learner as they can directly experience the relevance of a concept. **Invitation.** An "invitation" is a request to the learner to perform a specific activity. For instance, it can consist of a call for discussion with other students. Meta-cognitive hints often have the form of an invitation, e.g., "Reflect on what you have learned".

Knowledge, Comprehension, Application, Analysis, Synthesis, Evaluation Exercise. Instructional objects from these classes correspond to typical exercises found in learning materials. The classes were adopted from [Bloom, 1956] and differ in the educational objective they aim the student to achieve, e.g., whether a learner can recall or apply a concept. Recently, new classifications have emerged, for instance PISA's *literacies* [Schleicher, 1999]. It may be necessary to include them in the future, but currently Bloom's taxonomy is dominantly used.

Example. An "example" serves to illustrate a concept. Similar to interactivities, it has a "difficulty" slot.

Knowledge, Comprehension, Application, Analysis, Synthesis, Evaluation Example. These subclasses of "example" are similar to those of exercises. They illustrate concepts with different educational objectives.

Non-Example. A "non-example" is an instructional object that is not an example of a concept but is often mistakingly thought of as one. It includes "counter-examples".

Evidence. An "evidence" provides supporting claims, for instance observations or proofs, made for a law or one of its subclasses. Therefore, the "for"-property of an evidence has a range the class "law".

Proof. A "proof" is a more strict evidence. It can consist of a test or a formal derivation of a concept.

Demonstration. A "demonstration" consists of a situation in which is shown that a specific law holds. Experiments in physics or chemistry are typical examples of demonstrations. Note that the demonstration of a procedure, e.g., by showing how a curve is sketched is not a demonstration in the heredescribed sense, but is an application example.

Explanation. An "explanation" provides additional information about a concept. It elaborates on some aspect, points out important properties.

Introduction. An "introduction" contains information that leads the way to the concepts.

Conclusion. A "conclusion" sums up the main points of a concept.

Remark. A "remark" provides additional, not obligatory information about an aspect of a concept. It can contain interesting side information, or details on how the concepts is related to other concepts.

4 Mapping Knowledge Representations onto the Ontology

An ontology fulfills its purpose if it is used by a large number of parties. As the ontology of instructional objects described in this article is a new development, only its potential usefulness can be shown. Section 2 outlined educational Web services and their benefit from the ontology. This section describes three knowledge representations, two of them used in e-learning systems, and shows that the ontology can be used to describe the representation.

4.1 DocBook

DocBook [Walsh and Muellner, 1999] serves a standard for writing structured documents using SGML or XML. Doc-Book elements describe the complete structure of a document down to basic entities, e.g., the parameters of functions. The elements in-between are the most interesting in the scope of this article. DocBook offers several elements that describe content at paragraph level (called "block" elements).

DocBook	Instructional Object	
Example	Example	
Procedure	Procedure	
CmdSynopsis/FuncSynopsis	Definition	
Highlights	Summary	
Para,Figure	depends on content	

Table 1: Mapping of a selection of DocBook elements and instructional objects

Table 1 contains a mapping between DocBook elements and instructional objects. "CmdSynopsis" and "FuncSynopsis" describe the parameters and options of a command; "Highlights" summarizes main points. As one can see, although DocBook was not designed for educational purposes, several elements such as "example" and "procedure" can be directly described by the ontology of instructional objects. However, such a table functions as a very rough guideline only. To infer the exact instructional purpose of a block element by its tag alone is not possible in general. Especially with regard to abstract elements such as "para" (paragraph) or "figure", the content has to be taken into account in order to determine the correct instructional function.

Although DocBook is not directly related to e-learning purposes, it is of interest here because its way of structuring content in rather unspecified paragraphs is similar to systems such as [Henze and Nejdl, 2001; Weber and Brusilovsky, 2001; Bra *et al.*, 2002].

4.2 WINDS

WINDS ([Specht *et al.*, 2001]), a Web-based Intelligent Design and Tutoring System, uses the adaptive learning environment ALE to provide several adaptive hypermedia features, e.g., adaptive link annotation. Its knowledge representation is based on Cisco's learning objects and provides the following types: Introduction, Issue, Fact, Definition, Example, Nonexample, Simulation, Process, Procedure, Guidelines, Criteria, Analogy, Instruction, Summary, Tests.

Most of the types can be directly matched onto the ontology, unsurprisingly, as Cisco's learning objects served as one source for the here described ontology, too. However, in the ontology, "analogous" is introduced as a property between two instructional objects, and not as a stand-alone element. The reason is that every object can serve as an analogy for another object, regardless of its type.

4.3 <ML $>^3$

The "Multidimensional Learning Objects and Modular Lectures Markup Language", $\langle ML \rangle^3$ ([Lucke *et al.*, 2003]), is used by 12 German universities that encoded 150 content modules. It is of particular interest in the scope of this article because its design was explicitly influenced by pedagogical considerations. $\langle ML \rangle^3$ represents learning materials in "content blocks". These blocks are of the type "definition", "example", "remark", "quotation", "algorithm", "theorem", "proof", "description", "task", or "tip". Because of its pedagogical background, most $\langle ML \rangle^3$ elements directly correspond to an instructional object. Some element, such as "quotation" and "description" can not be mapped directly. Again, it is necessary to asses the instructional purpose of the element. For instance, does the quotation serve as a bibliographical reference? Then it is no instructional object in the true sense. Or does the quotation introduce a concept, e.g., by citing a famous scientist? Then it would be classified as an "introduction".

5 Related Work

Seminal work regarding ontologies and e-learning was done by Mizoguchi. [Mizoguchi and Bourdeau, 2000] lay out how ontologies can help to overcome problems in artificial intelligence in education. [Aroyo and Mizoguchi, 2003] describe how an assistant layer uses an ontology to support the complete authoring, for instance by providing hints on the course structure. The work in this paper has a different focus but fits well in their approach.

[Dolog *et al.*, 2004] propose a architecture for personalized e-Learning based on Web services. In the architecture, a personal learning assistant integrates personalization services such as a recommendation and a link generation service and thereby provides a personalized access to learning resources. They also describe an ontology for learning resources. However, their ontology does not represent the instructional function of a resource. Both ontologies complement each other.

Using APeLS ([Conlan *et al.*, 2003]), an author can describe his courses on a level which abstracts from the concrete learning resources and focuses on learning goals. He writes a *narrative* which references to a group of candidate resources that each fulfill the learning goal. The concrete element chosen depends on the learner's individual context. In the current version of APeLS, learning goals are concepts of the domain being teached. Using the ontology of instructional objects allows for the integration of commonly shared instructional goals, thereby enabling for better re-use and adaptability.

An approach of integrating e-learning systems which is not based on web-services is described by [Brusilovsky, 2004]. In this architecture, called KnowledgeTree, a *portal* takes care of the management and offers an integrating access to the learning resources. The learning resources are not stored in a central repository but offered by *activity* and *value-adding servers*. An activity server presents the learning resources to the learner and handles her interactions with the resources. A value-adding server adds functionality to the resources, similar to the Web services described in Section 2. Hence, in the same way the Web services benefit from an ontology of instructional objects, KnowledgeTree servers can use this additional information to provide better learner support.

6 Conclusion

This article describes an ontology of instructional objects which captures the educational "essence" of a learning resource. This ontology is supposed to serve as a shared and common understanding that can be communicated between people and applications. A number of Web services were described to illustrate how they benefit from the ontology. Additionally, the connections between two document structuring standards and the ontology exemplified the applicability of the ontology.

An ontology is never completely stable and always the result of integrating different viewpoints. To stimulate discussion and to enhance the scope of the ontology, the author has set up a forum at his homepage (http://www.activemath.org/~cullrich/oio.html). It is the hope of the author that the ontology is one step forward to bring the Web to its full e-learning potential.

The author wishes to thank Kerstin Borau, the ACTIVEMATH-group, especially Erica Melis, Paul Libbrecht and George Goguadze, and the anonymous referees for their valuable suggestions.

References

- [Aroyo and Mizoguchi, 2003] L. Aroyo and R. Mizoguchi. Authoring support framework for intelligent educational systems. In U. Hoppe, F. Verdejo, and J. Kay, editors, *Proceedings of AI in Education, AIED-2003*, pages 362– 364. IOS Press, 2003.
- [Bartle and Sherbert, 1982] R.G. Bartle and D.R. Sherbert. *Introduction to Real Analysis*. John Wiley& Sons, New York, 1982.
- [Bloom, 1956] B.S. Bloom, editor. Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain. Longmans, Green, New York, Toronto, 1956.
- [Bra et al., 2002] P. De Bra, A. Aerts, D. Smits, and N. Stash. Aha! version 2.0 – more adaptation flexibility for authors. In Proceedings of the AACE ELearn'2002 conference, 2002.
- [Brusilovsky et al., 1998] P. Brusilovsky, J. Eklund, and E. Schwarz. Web-based education for all: A tool for developing adaptive courseware. *Computer Networks and ISDN* Systems, 30(1-7):291–300, 1998.
- [Brusilovsky, 2004] P. Brusilovsky. KnowledgeTree: A distributed architecture for adaptive e-learning. In Proc. of WWW2004 - The Thirteen International World Wide Web Conference, 2004.
- [Cisco Systems, Inc, 2003] Cisco Systems, Inc. Reusable learning object strategy: Designing and developing learning objects for multiple learning approaches, 2003.
- [Conlan et al., 2003] O. Conlan, D. Lewis, S. Higel, D. O'Sullivan, and V. Wade. Applying adaptive hypermedia techniques to semantic web service composition. In Proceedings of AH2003: Workshop on Adaptive Hypermedia and Adaptive Web-Based Systems, pages 53–62, 2003.

- [Dolog et al., 2004] P. Dolog, N. Henze, W. Nejdl, and M. Sintek. Personalization in distributed elearning environments. In Proc. of WWW2004 - The Thirteen International World Wide Web Conference, 2004.
- [Gennari et al., 2003] J. H. Gennari, M. A. Musen, R. W. Fergerson, W. E. Grosso, M. Crubézy, H. Eriksson, N. F. Noy, and S. W. Tu. The evolution of Protégé: an environment for knowledge-based systems development. *International Journal of Human-Computer Studies*, 58(1):89– 123, 2003.
- [Henze and Nejdl, 2001] N. Henze and W. Nejdl. Adaptation in open corpus hypermedia. *International Journal of Artificial Intelligence in Education*, 12, 2001.
- [IEEE Learning Technology Standards Committee, 2002] IEEE Learning Technology Standards Committee. 1484.12.1-2002 IEEE standard for Learning Object Metadata, 2002.
- [IMS Global Learning Consortium, 2003] IMS Global Learning Consortium. IMS learning design specification, February 2003.
- [Jonassen, 1999] D. Jonassen. Designing constructivist learning environments. In C. M. Reigeluth, editor, *Instructional Design Theories and Models: A New Paradigm of Instructional Theory*, volume 2, pages 215–239. Lawrence Erlbaum Associates, 1999.
- [Kay et al., 2002] J. Kay, R. J. Kummerfeld, and P Lauder. Personis: a server for user models. In P. De Bra, P. Brusilovsky, and R Conejo, editors, *Proceedings of* AH'2002, Adaptive Hypertext 2002, pages 203–212. Springer, 2002.
- [Lajoie and Derry, 1993] S. Lajoie and S. Derry, editors. Computers as Cognitive Tools. Erlbaum, Hillsdale, NJ, 1993.
- [Lucke et al., 2003] U. Lucke, D. Tavangarian, and D. Voigt. Multidimensional educational multimedia with <ML>³. In Proceedings of E-Learn 2003, World Conference on E-Learning in Corporate, Government, Healtcare, & Higher Education, pages 101–104, 2003.
- [Mann and Thompson, 1988] W. C. Mann and S. A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3):243–281, 1988.
- [Mizoguchi and Bourdeau, 2000] R. Mizoguchi and J. Bourdeau. Using ontological engineering to overcome AI-ED problems. *International Journal of Artificial Intelligence in Education*, 11(2):107–121, 2000.
- [Pawlowski, 2002] J. M. Pawlowski. Modellierung didaktischer Konzepte mit dem Essener-Lern-Modell. In Proceedings of Workshop Neue Medien in der Bildung: Standardisierung im e-Learning, 2002.
- [Reigeluth, 1999] C. M. Reigeluth, editor. Instructional Design Theories and Models: A New Paradigm of Instructional Theory, volume 2. Lawrence Erlbaum Associates, Mahwah, NJ, 1999.

- [Schleicher, 1999] A. Schleicher, editor. *Measuring Student Knowledge and Skills – A New Framework for Assessment*. OECD, 1999.
- [Specht et al., 2001] M. Specht, M. Kravcik, L. Pesin, and R. Klemke. Authoring adaptive educational hypermedia in WINDS. In N. Henze, editor, *Proc. of the ABIS 2001 Workshop*, 2001.
- [Ullrich, 2003] C. Ullrich. An instructional component for dynamic course generation and delivery. In R. Tolksdorf and R. Eckstein, editors, *Proceedings of Berliner XML Tage 2003*, pages 467–473, 2003.
- [van Marcke, 1998] K. van Marcke. GTE: An epistemological approach to instructional modelling. *Instructional Science*, 26:147–191, 1998.
- [Walsh and Muellner, 1999] N. Walsh and L. Muellner. Doc-Book: The Definitive Guide. O'Reilly, 1999.
- [Weber and Brusilovsky, 2001] G. Weber and P. Brusilovsky. ELM-ART: An adaptive versatile system for web-based instruction. *International Journal of Artificial Intelligence in Education*, 12(4):351–384, 2001.
- [World Wide Web Consortium, 2004] World Wide Web Consortium. OWL web ontology language, use cases and requirements, February 2004.

A Pragmatic Approach to Support Concept-based Educational Information Systems Communication

Darina Dicheva

Winston-Salem State University Department of Computer Science 601 Martin Luther King, Jr. Drive, Winston Salem, NC 27110 dichevad@wssu.edu

Lora Aroyo

Eindhoven University of Technology Faculty of Mathematics and Computer Science P.O.Box 315, 5600 MD Eindhoven, The Netherlands 1.m.aroyo@tue.nl

Abstract

Significant efforts are currently focused on defining powerful frameworks and architectures to support interoperability and integration of Web-based Educational Information Systems (EIS). We approach this integration problem from a rather practical perspective and propose a pragmatic framework for supporting communication between existing concept-based EIS aimed at utilizing systems' resources. The framework allows two independent systems to share and interchange information solely through ontology-based communication without sharing data stores. As a basis of the framework, we define a communication ontology and propose an interaction protocol, CB-EIS IP, built over a SAAJ-enabled SOAP transport layer.

1 Introduction

Web-based learning support systems, including Educational Information Systems (EIS) that are aimed at providing resources and services for various educational goals and tasks attract a growing interest. Representatives of such systems are adaptive textbooks constructed with AHA! [De Bra et al, 2003], InterBook [Brusilovsky et al, 1998] and NetCoach [Weber et al, 2001], or adaptive courses prepared within ELM-ART [Brusilovsky et al, 1996], PAT Online [Ritter, 1997], AIMS [Aroyo et al 2001], etc. Most of these specialized educational systems and content providers support only a single task/function within the educational process. In order to support a richer set of educational functions and increase their effectiveness, such systems need to interoperate, collaborate and exchange content or re-use functionality. Consequently, considerable efforts are currently focused on defining powerful frameworks and architectures to tackle issues of integration and interoperability of such systems. These frameworks prove useful for developing future effective large-scale web-based educational systems. In this paper we try to approach the integration problem of present systems from a rather practical perspective and propose a pragmatic framework for supporting communication between existing concept-based EIS aimed at utilizing systems' resources.

The main goal of web-based EIS is to provide the learners with immediate, on-line access to a broad range of structured information. They also support more efficient task performance by offering learners a domain-related help in the context of their work. There are a number of conceptbased EIS already developed [Brusilovsky et al, 1998; Weber et al, 2001; Brusilovsky et al, 1996; Aroyo et al 2001; Dolog et al, 2004; Dicheva et al, 2004b] which typically include:

- concept-based (ontology-driven) subject domain,
- repository of learning resources,
- course (learning task) presentation,
- adaptation & personalization.

The fundamental feature of these systems is the *subject domain conceptualization*. It supports not only efficient implementation of their required functionality but also standardization: the concept structure can be built to represent a domain ontology that provides a broadly agreed vocabulary for domain knowledge representation. If the attached learning resources have also a standards-based representation as opposed to a system-specific internal representation, this will insure that the application's content is reusable, interchangeable, and interoperable. Good examples of such systems are AIMS (Adaptive Information Management System) [Aroyo et al, 2001] and TM4L (Topic Maps for Learning) [Dicheva et al 2004b], which we use as examples in our discussion. Though quite similar, these systems can be seen as complementary in the way they support learning tasks. While AIMS includes course representation and sequencing, TM4L is a digital library, which does not include direct course representation.

Integration and interoperability are very important for the EIS systems. If interoperable, two systems can benefit of additional functionality (supplied by the other system) and especially of sharing resources and common components, e.g. user models. In our example of AIMS and TM4L, TM4L can use AIMS course sequencing model, and resource metadata, while AIMS can use TM4L external and internal resources, domain and resource merging capability, text search, and external search. Our approach to the concept-based EIS integration problem is rather practical and based on sharing information between systems solely through communication without sharing data stores (e.g. providing data from one system on a request from another without allowing a general access to the private data store of the first system). The main questions related to the implementation of such communication concern the level of granularity of communicated information, the syntax and semantics of communication messages, and possible modes of use of user models (communicated or shared by systems). We have tried to answer these questions at two levels - a general one and a pragmatic one, which provided guidelines to the design of two corresponding frameworks for supporting communication between concept-based EIS. While the general framework fits well in the ambitious effort to define conceptually the shared and interoperable Educational Semantic Web by providing a powerful service-oriented architecture to support efficient communication between component-based EIS, the pragmatic one presents an efficient currently realistic solution by providing a constrained architecture for supporting shareability and exchangeability of existing systems' resources. Our implementation efforts as well as the focus of this paper are directed to the constrained architecture, since we believe that it will help to fill a gap between the current situation and the promises of the Educational Semantic Web of the future.

The paper is organized as follows. After a brief description of the general framework for supporting interoperability of various concept-based EIS in Section 2 (for details see [Dicheva et al, 2004a]), we propose a pragmatic approach for implementing the communication between two existing concept-based EIS (Section 3) by defining a communication ontology (Section 4) and proposing an interaction protocol, CB-EIS IP, built over a SAAJ-enabled SOAP transport layer (Section 5). We conclude with a short discussion.

2 General Architecture for Component-based EIS Interoperability

The proposed in [Dicheva et al, 2004a] general architecture for supporting component-based EIS include (see Figure 1):

- Stand-alone, component-based *independent EIS* using their private subject domain ontologies.
- *Information brokerage bureau* where all applications are registered.
- Services to support systems communication, e.g. for ontology mapping.
- *Communication bridges* between the systems supporting standardized transport mechanisms and a common interaction protocol.

The main purpose of the architecture is to support sharing and exchanging information between EIS initially designed to be standalone. This is achieved through communication between the systems (or their components) via services included in the framework to facilitate systems' communication. The services, including ontology-related services, are intended to support different specific aspects of the communication.

A communication is an interaction between two software systems (agents) guided by an interaction protocol. The communication between the systems requires not only standardized transport mechanisms and communication languages, but also *common content languages* and *semantics*. We have chosen XML as an 'information' content language to represent the content embedded in the messages in our architecture, as opposed to the commonly used 'logic' languages for representing the content, embedded in ACL (Agent Communication Languages) messages, such as KIF (Knowledge Interchange Format) [KIF, 1998], SL (Semantic Language proposed by FIPA) [FIPA-SL, 2004], and Prolog.



Figure 1. General architecture for component-based EIS interoperability.

As to the communication semantics, in order for the applications to understand each other we propose using a communication ontology that defines the vocabulary of terms used in the messages at both message and content layers (see Section 4). To interpret the requests and answers standardized domain ontologies, User Model ontologies, as well as upper-level ontologies such as, WordNet, etc. can be used.

Our next step is to *constrain* the proposed general architecture by considering *only two* communicating systems that "know" each other and "trust" each other. We consider that this is a common case and the goal is to find a configuration that will support such communication and allow sharing of systems' knowledge and resources.

3 Constrained Architecture for Conceptbased EIS Communication

Since our present goal is to support communication between already developed concept-based EIS, each system is assumed to be a standalone application and is not required to have a particular architecture or to "adapt" to the other system in the framework. We make two important presumptions and use them as a basis for our design of a constrained architecture:

- The two systems know and are committed to communicating with each other. This implies that the systems will communicate directly and there is no need of Information brokerage bureau for registering the applications. Note that one system can communicate with more than one other system in such a direct mode.
- · Concerning the services related to ontology mapping, an important presumption for our simplified framework is that the domain ontologies are created and used within one community, not different communities. This eases a lot the task since we may assume that in one community there exists an agreed upon understanding that favors the sharing of knowledge. Indeed, the goal of the EIS systems that we consider is to support learning in a specific course (discipline), for example a Database course. The community of users includes potential instructors (authors) and learners. Since the authors are knowledgeable in the specific subject domain, e.g. databases, it can be assumed that in defining the domain ontology they will use terms (concepts) that are accepted and agreed upon in that domain. This will remove the necessity of alignment and translation of domain ontologies. That is why ontology-related services are not included in the constrained architecture. Note that it will be useful though to include a service for merging ontologies. Currently, we delegate this task directly to the applications.

Table 1 compares the components in the general and constrained architectures.

Components	General Architecture	Constrained Architecture
Stand-alone EIS based on domain ontologies	Yes	Yes
Information Brokerage Bureau	Yes	No
Communication-supporting services	Yes	No
Communication Bridges	Yes	Yes

Table 1. Components required in both architectures

As shown in Table 1, the proposed constrained architecture includes only the two concrete communicating systems (e.g. AIMS and TM4L) and a communication bridge between them (see Fig. 2). Since in this architecture there are only two "committed" communicating systems, there is no real need of agent communication management, represented by Information brokerage Bureau and Communicationsupporting Services.

We propose a common interaction protocol for the communication between Concept-Based EIS (CB-EIS IP) built over a SAAJ-enabled SOAP transport layer. As a content language we use XML, which is designed to support data exchange interoperability between applications. In the next sections we first propose a communication ontology for the pragmatic framework and then discuss implementation details of the proposed constrained architecture, including the proposed interaction protocol.

In order for the communicating applications to understand

each other we need an ontology to provide a basis for shar-

ing a precise meaning of symbols exchanged during communication. An ontology denotes a representation vocabulary of a specific domain, and more precisely the conceptu-

Communication Ontology

4

alizations that the terms in the vocabulary are intended to capture [Chandrasekaran et al, 1999]. In our case we define a communication ontology which conceptualizes the domain of the communication between two concept-based EIS. We distinguish two parts, corresponding to both layers of an interaction between two communicating systems, the message layer and the content layer. Consequently, we propose the Communication Ontology (CO) to consist of communication content ontology and interaction protocol ontology (see Fig. 3):

- Communication content ontology (CCO) describes the content (knowledge) that can be exchanged by the systems (corresponds to the content layer).
- Interaction protocol ontology (IPO) specifies interaction communicative act types (corresponds to the message layer).

4.1 Content Ontology

The Communication Content Ontology defines the terms (concepts) needed to exchange messages, i.e. gives the meaning of the symbols included in the content expression.



Figure 2. Pragmatic integration framework for AIMS and TM4L

When two concept-based EIS exchange data, the message content will typically include two types of terms (concepts): terms belonging to the domain ontology of the sender (the application sending the message) and terms categorizing domain term(s). The latter belongs to the general information model of concept-based EIS. For example, the sender can send a request for information of the kind "Send me all relationships in which you believe concept 'ER-model' is involved". In this message 'ER-model' is a term (concept) from the subject domain of the requesting application, while 'relationship' and 'concept' are terms belonging to the information model of concept-based EIS.

Thus in our framework, the content ontology consists of two parts: the application domain ontologies (DO) of the involved EIS and an application domain-independent ontology defining the concept-based information model of EIS (EISO). The latter includes basic terms describing the information model of concept-based EIS, such as concept, concept name, relationship type, relationship role, etc. Figure 3 presents an excerpt from this ontology. In the proposed framework, each application uses the common EISO ontology and its own domain ontology. For this reason we have depicted application domain ontologies separately from the EIS ontology in Fig. 1.

4.2 Interaction Protocol Ontology

The Interaction Protocol Ontology (IPO) defines terms related to message types, reasons, and preconditions. While the communication content ontology is generally independent of the framework's functionality, the IP ontology has to reflect its functionality (e.g. whether it supports agent communication).



Figure 3. An excerpt from the communication ontology


Figure 4. An excerpt from the IPO ontology

Messages represent communicative acts denoting the actions related to communication. In general, communicative acts (performatives) include (1) queries, (2) responses, (3) informational, (4) capability definition, (5) generative, and (6) networking (see KQML [Finin et al, 1994]). Since the two applications in our constrained architecture are "committed" to collaborate, the communication between them is very simple and does not require the typical variety of message types, for example, types such as agree, accept, cancel, propagate, and refuse, as well as defining message preconditions and reasons. Thus, in our case we choose the IPO ontology (Fig. 4) to include the following message types:

Status

- Failure: Informing that an action was attempted but the attempt failed.
- Not understood: message or Domain Ontology term.
- **Success:** Informing that an action was attempted and the attempt succeeded.

Query

- Query-know: Asking whether the receiver knows about an object corresponding to an EIS Ontology term/category (e.g. specific concept, relationship, etc).
- Query-confirm: Asking whether a proposition is true.
- Query-object: Asking for an object or all objects of specific category in the EIS Ontology (e.g. concept, relation, etc).
- Query-if-object: Asking for objects as in 'queryobject' but in case a specified proposition is true.

Response

- **Response-know:** Informing the receiver whether or not the sender knows about the specified object.
- **Response-confirm:** Confirming to the receiver that the specified in the query proposition is true or not.

- **Response-object:** Sending to the receiver the objects specified in the request.
- Response-if-object: Sending to the receiver the objects specified in the request only if the specified proposition is true.

5 Communication Bridge

As a basis of the transportation mechanism in our framework we have chosen SOAP (Simple Object Access Protocol) [SOAP], which is a standard lightweight protocol for exchanging information in a decentralized, distributed environment. It complies with the WS-I Basic Profile 1.0 specifications and therefore supports interoperability across platforms, operating systems, and programming languages. It actually permits an exchange of messages in XML format between physically distributed machines. More specifically, the communication bridge is based on using the SOAP with Attachments API for Java (SAAJ). The SAAJ API, allows creating XML messages that conform to the SOAP 1.1 and WS-I Basic Profile 1.0 specifications. A SAAJ client is a standalone client. It sends point-to-point messages, i.e. a message goes from the sender directly to its destination. Messages sent using the SAAJ API are request-response messages. They are sent over a SOAP connection, which sends a message (request) and then blocks until it receives the reply (response). A SOAP message is an XML document. It always has a required SOAP part, and it may also have one or more attachment parts (that can contain any kind of content). The SOAP part must always have an envelop, which contains a SOAP body.

To realize the communication between two concept-based EIS, we propose an interaction protocol, CB-EIS IP (Concept-Based EIS Interaction Protocol), which provides the real semantics of the communication between them. Since the message content language in the framework is XML, we have defined a DTD for XML files representing the content of interaction messages that conform to this protocol.

```
<!ELEMENT message (queryMessage | responseMessage) | statusMessage>
<!ELEMENT queryMessage (query-know | query-confirm | query-object | query-if-object)>
<!ELEMENT query-know (commOntoTerm, dmOntoTerm)>
<!ELEMENT query-confirm (proposition)>
<!ELEMENT query-if-object (proposition, query-object)>
<!ELEMENT query-object (objectSpec, categorySpec)>
<!ELEMENT proposition (relation, dmOntoTerm, dmOntoTerm)>
<!ELEMENT categorySpec (commOntoTerm)>
<!ATTLIST categorySpec type (category | ALL )>
<!ELEMENT objectSpec (relOperator, commOntoTerm, dmOntoTerm)>
<!ATTLIST objectSpec type (object | ALL )>
<!ELEMENT relation (#PCDATA)>
                                 <!-- term from DO -->
<!ELEMENT dmOntoTerm (#PCDATA)>
                                 <!-- term from DO -->
<!ELEMENT commOntoTerm (#PCDATA)> <!-- term from EISO -->
```

Figure 5. An excerpt of the DTD definition of the CB-EIS IP

The DTD definition is based on the developed Communication Ontology (CO). An excerpt of the DTD document is given in Figure 5. This DTD allows sending messages like the following:

• A request asking whether the recipient "knows" the concept 'relational model':



• A message, containing a "yes" response to the previous request:

```
<message>
<response-know type = known/>
</message>
```

• A message, requesting the relationships in which concept 'ER-model' is involved:



The CO interface modules in our architecture (see Fig. 2) are responsible for translating the messages (requests and responses) from the native language of EIS (e.g. TM4L or AIMS) into the language of the universal CB-EIS IP and vice versa. We plan to develop an API for Java (EISIPAJ), to be used by the CO interface for creating and interpreting XML files (representing the content of interaction messages) that conform to the CB-EIS IP. The CO interface is built on top of a SAAJ module and uses it to realize the CB-EIS IP with SOAP messages (the CB-EIS commands are embedded within the SOAP body).

Thus, in the proposed pragmatic framework, two independent systems can share and interchange information solely through ontology-based communication without sharing data stores. This removes any constraints on the systems architecture as well as the necessity of developing a 'wrapper' system, i.e. an environment that host the communicating systems. The only requirement for the systems is to be furnished with a plug-in realizing a CO interface that enables sending and receiving messages conforming to the proposed CB-EIS IP (through a SAAJ client and a SAAJ servlet).

6 Conclusion

We believe that the time for implementing large-scale educational web-service frameworks hasn't come yet. Thus our efforts are focused on increasing the use and efficiency of present, i.e., already developed or currently being developed systems, more specifically concept-based educational information systems. We propose to complement their functionality by supporting them to ask external 'known' peersystems for information, possibly involving informationproviding processing.

We approach the problems related to systems integration and communication at two levels: a general level, proposing a powerful service-oriented framework to support efficient communication between component-based EIS, and a pragmatic one, illustrating an efficient proof of concept for supporting shareability and exchangeability of system resources, applicable in the context of the current educational computing advancement. We believe that the proposed constrained architecture will contribute to filling the gap between the current realistic situation and the desired future educational semantic web. As part of the framework we have defined a communication ontology consisting of communication content ontology and interaction protocol ontology and have embedded the latter within the CB-EIS IP. We have illustrated the concrete realization of the interaction protocol ontology within the constrained architecture. This way, we show how two independent systems can share and interchange information solely through ontology-based communication without sharing data stores.

The proposed framework for supporting communication between applications will eliminate in many cases the need for exporting the entire application domain model or other application model to another application. Thus, this will be an alternative to interchanging and merging domain models. The advantage is in eliminating duplication of stored information, which is unlikely to be often used. In addition, if an application has a specific concept-based application model with no corresponding model in the other system, import will not work and the proposed communication is the only way for the second system to use information from the first one. This will also solve problems related to shareability and reusability for already developed applications that don't use standards-based information but rather their own internal representations.

Acknowledgements

We would like to thank Christian Tsolov at the Univerity of Twente, The Netherlands and the paper reviewers for many valuable comments and suggestions on how to improve the paper.

References

- [Aroyo et al, 2001] L. Aroyo and D. Dicheva. AIMS: Learning and Teaching Support for WWW-based Education. *International Journal for Continuing Engineering Education and Life-long Learning*, 11(1/2): 152-164.
- [Brusilovsky et al., 1998] P. Brusilovsky, J. Eklund and E. Schwarz. Web-based education for all: A tool for developing adaptive courseware. Computer Networks and ISDN Systems. In *Proceedings of the 7th International WWW Conference*, pages 291-300. Springer.

- [Brusilovsky et al., 1996] P. Brusilovsky, E. Schwartz, G. Weber. ELM-ART: An Intelligent Tutoring System on the World Wide Web. In Proceedings of the 3rd International Conference on Intelligent Tutoring Systems, pages 261-269, Springer.
- [Chandrasekaran et al., 1999] B. Chandrasekaran, J. Josephson and R. Benjamins. What are Ontologies and Why Do We Need Them? J. *IEEE Intelligent Systems*, January-February, pages 20-26, 1999.
- [De Bra et al., 2003] P. De Bra, A. Aerts, B. Berden, B. de Lange, B. Rousseau, T. Santic, D. Smits, and N. Stash. AHA! The Adaptive Hypermedia Architecture. In *Proceedings of the 14th ACM Conference on Hypertext and Hypermedia*, pages 81-85, 2003. ACM Press.
- [Dicheva et al., 2004a] D. Dicheva and L. Aroyo. General Architecture Supporting Component-based EIS Interoperability. In *Proceedings of the Workshop on Semantic Web Technologies for E-Learning*, ITS'04, September 2004, Maceió, Brazil, 21-28.
- [Dicheva et al., 2004b] D. Dicheva, C. Dichev, Y. Sun and S. Nao. Authoring Topic Maps-based Digital Course Libraries. In *Proceedings of the Workshop on Applications* of Semantic Web Technologies for Adaptive Educational Hypermedia, AH 2004, August 2004, Eindhoven, The Netherlands, pages 331-337.
- [Dolog et al., 2004] P. Dolog, N. Henze, W. Nejdl and M. Sintek. Personalization in Distributed eLearning Environments. In Proceedings of the 13th International WWW Conference.
- [Finin et al., 1994] T. Finin, R. Fritzson, D. McKay and R. McEntire. KQML as an agent communication language. In Proceedings of the 3rd International Conference on Information and Knowledge Management, pages 456-463.
- [FIPA-SL, 2004]. FIPA SL Content Language Specification. Available at:

http://www.fipa.org/specs/fipa00008/SC00008I.html

- [KIF, 1998]. Knowledge Interchange Format. Available at: http://logic.stanford.edu/kif/dpans.html
- [Ritter, 1997] S. Ritter. PAT Online: A Model-Tracing Tutor on the World-Wide Web. In Proceedings of the Workshop on Intelligent Educational Systems on the 6th WWW conference, pages 11-17.
- [SOAP, 2004]. Available at: http://www.w3.org/TR/soap/
- [Weber et al., 2001] G. Weber, H. C. Kuhl and S. Weibelzahl. Developing adaptive Internet-based courses with the authoring system NetCoach. In *Proceedings of the 3rd International Workshop on Adaptive Hypermedia*.

Using Semantic Web Methods for Distributed Learner Modelling

Mike Winter¹, Christopher Brooks¹, Gord McCalla¹, Jim Greer¹ and Peter O'Donovan²

University of Saskatchewan

ARIES Laboratory

Saskatoon, SK., Canada S7N 5A9 ¹{mfw127, cab938, mccalla, greer}@cs.usask.ca ²peo499@mail.usask.ca

Abstract

Developing a learner model containing an accurate representation of a learner's knowledge is made more difficult in distributed learning environments where the learner uses multiple applications and resources to accomplish learning tasks. To help reduce this difficulty we describe a semantic web approach to representing student models based on distributed student data. We also present a proposal for revising those student models based on arbitrary, web-based learner actions.

1 Introduction

Current online learning is described as often taking place in an 'adaptive learning community' [Gaudioso and Boticatio, 2003] in which online learners use a wide variety of resources to help them perform their problem-solving tasks. These resources include a wide variety of web-pages, instant messaging, online discussion and peer-help tools. In this paper, we present an integrated learner modelling architecture using RDF, RDFS and SOAP that effectively stores and transmits learner information from multiple sources.

The outline of this paper is as follows: Section 2 describes the development of a RDF/RDFS based learner model for a first-year computer science class and the use of the Massive User Modelling System (MUMS) [Brooks *et al.*, 2004] which allows the collection of learner modelling information from diverse application sources for use by our learner models. Section 3 describes how the integrated system is currently deployed for hundreds of students while Section 4 discusses our future goals including the use of information retrieval techniques with MUMS to update our learner models.

2 Granular Learner Models with RDF and RDFS

The first challenge for a learner modelling system in a distributed learning environment is to effectively attach meaning to the learner data it is receiving. Using RDF to model learners has many advantages for this task. First, RDF is a well-specified semantic data model that can be easily serialized between systems, allowing easy sharing of learner models and learner information between interested components. Second, popular RDF packages such as Jena¹ allow for the easy manipulation of RDF graphs, including reasoning capabilities that allow a modelling component to make inferences regarding learners over multiple ontologies. Finally, RDF is able to refer to an arbitrary number of ontologies within a single graph. This allows a student modelling component to accurately model many different aspects of a learner by combining statements that use different ontologies in the same graph. The learner modelling component that we have developed uses multiple RDF schema (RDFS) ontologies to define the classes and relationships contained in RDF graphs that act as student models. The two main ontologies we have encoded in RDFS to express learner model information are listed below.

1. Granularity Hierarchies. To define concept maps for the domains being studied by the learners in our system, we use the granularity hierarchy formalism which is an extended semantic network that defines both specialization and aggregation relationships between topics [McCalla *et al.*, 1992]. In the granularity hierarchy formalism, a K-Cluster represents a particular semantic aggregation of topics while an L-Cluster represents a particular semantic specialization of a topic. A topic can have more than one K-Cluster and/or L-Cluster relationship. The major advantage granularity hierarchies provide in terms of domain modelling is the ability to represent a domain at multiple levels of detail simultaneously. Currently, a domain map has been

¹ http://jena.sourceforge.net

developed using this method that completely models the topics within a first-year Computer Science course at the University of Saskatchewan (Figure 1). This domain map contains over five-hundred topic nodes and thousands of granularity hierarchy relationships between them.



Figure 1. Section of a Granular Learner Model

2. Ontology of Learning Outcomes. For purposes of learner modelling, a concept map is not enough; the knowledge of particular learners must be added to instantiations of the map. Student knowledge of a topic can be represented as an increasing degree of proficiency as detailed by Bloom's taxonomy [Bloom, 1956]. We have developed an RDFS version of the taxonomy encompassing its eight levels of knowledge ranging from Knowledge (basic recall) to Evaluation (assess and contrast). To build learner models of real learners, quizzes have been developed for an online version of the first-year Computer Science course modeled above. Answers to questions are annotated with the topic(s) that they cover and the Bloom's level of knowledge that will be demonstrated if they are answered correctly.

2.1 MUMS: Collecting Distributed Learner Modelling Data

MUMS is an event system designed to collect and distribute notifications of user actions from the applications where they happen to interested third parties [Brooks *et al.*, 2003]. Applications that generate events are called **Producers**, applications that receive events are called **Consumers** or **Modellers**, and the application that routes messages from Producers to Consumers is called a **Broker**. **Filters** act as intermediaries between the Broker and Consumers, and provide miscellaneous security, routing and reasoning services. Consumers send event subscription requests (or queries for archival user information) to the Broker and then receive user events as they happen in real-time.

There are three design principles underlying the MUMS system: interoperability, extensibility and scalability. Interoperability is necessary because of the diversity of applications that are involved in generating and consuming user events. Interoperability is ensured in MUMS in two ways. The first is by an implementation in of the Web-Services Events (WS-E) [Catania et al., 1985] specification in WSDL, with a SOAP binding. This specification details the type and format of messages that should be sent between components in a web-service based event system. The use of web-service standards enables a language and systemneutral transport protocol for event messages. The second way interoperability is ensured in the MUMS system is by the use of RDF statements as the payload of each notification. RDFS or OWL schemas provide the ontologies for the RDF payloads and allow the various applications on the MUMS network to have a shared semantic understanding of user events. Extensibility is an important feature of the MUMS network because Modellers and Producers are volatile and may be online or offline at any given moment. Extensibility is provided for by the subscription mechanism of the MUMS system. Consumers send subscription requests in RDQL to the broker regarding arbitrary events (usually involving certain users or groups of users) and receive events generated by applications across the MUMS network that match the subscription request. Another mechanism ensuring extensibility is the lightweight API that easily allows components to talk to the MUMS network. Scalability is ensured by the distributed nature of the network and the clustering capabilities inherent in the router design.

3 The Integrated Modelling Network

The learner modelling component we have developed is a consumer on the MUMS network, with a separate instantiation of the domain model for each learner. As the MUMS events are encoded in RDF, they are simple for the modelling component to either add directly on to the existing RDF learner model or to use them as the input for inference. Currently, the learner modelling system is deployed for a firstyear Computer Science course at the University of Saskatchewan that has around three-hundred students enrolled, thirty-five of them through an online version of the course. The learner modelling system takes as inputs the answers to quizzes in the online course, as mentioned previously, as well as the learners' other actions in the online course, such as a reading a lesson or working with an interactive program. In addition, all of the students' activities on an online class discussion board are sent over the MUMS network and received by the learner modelling component. A MUMSenabled web proxy is also available for use in research studies. While we are just starting to build the learner models for the first time, the combination of our RDF-based user messaging system and our RDF learner models has been effective in combining distributed sources of learner information into coherent and accessible learner models.

4 Next Steps: Using Information Retrieval Techniques for Student Model Updating

Once the events about a student's behaviour have been transmitted to a student modelling system by the MUMS network there still exists the difficult problem of determining what relevance those events have in relation to its understanding of the learner's knowledge and plans. One already-implemented approach to translating events was discussed in the last section where the answers to quiz questions have pre-determined mappings to learner knowledge assessments. However, the MUMS network is able to transmit information from any arbitrary application, including ones where learner actions are not pre-analyzed. The remainder of this section will detail a proposed general approach to translating events involving a learner's interaction with text-based resources, such as web pages and message board postings, to appropriate learner model revisions.

Assuming a learner model like that discussed in Section 2, a learner's reading of a web page will have to be translated into an update of the model's understanding of the learner's domain knowledge. The way in which a textual resource view has to be interpreted in terms of the learner's knowledge gain can be further decomposed into two separate problems: determining the topic(s) of the textual resource and determining the amount of knowledge the learner has gained from the resource. One way in which the topic of the web page can be determined is by associating a representative piece of text with each knowledge node in the domain model and then using an appropriate information retrieval technique such as vector scoring to determine which topic the web page is most likely to be about. Determining the knowledge gain of the learner resulting from his/her viewing the text resource is trickier because the gain would vary based on the attention the learner paid to the resource, the quality of the resource, and other factors which would not generally be known to the student modelling system. To gather relevant data for this task, a study is being designed that will use MUMS events from the online course and the associated online discussion board that are hooked into the MUMS systems as producers, generating events each time a learner interacts with them. In addition, any other web-based resources that the users of the study access will generate events on the MUMS network through the MUMS-enabled web proxy. The tests in the online course are divided into pre and post-lesson components, and the resulting change in the learner's knowledge as discovered by the tests can then be correlated with the web resources they have viewed.

Acknowledgments

This work was partially supported by the Government of Canada's NSERC LORNET NCE grant.

References

- [Gaudioso and Boticatio, 2003] Gaudioso, Elena and Boticatio Jesus, G. Towards Web-based Adaptive Learning Communities. Proceedings of the 2003 International Conference on Artificial Intelligence in Education. Sydney, Aus., IOS Press.
- [Brooks et al., 2004] Brooks, Christopher, Winter, Mike, Greer, Jim and McCalla, Gord. The Massive User Modelling System (MUMS). *Proceedings of Intelligent Tutoring Systems 2004*. 2004, To appear.
- [Catania et al., 1985] Catania, Nicholas, Kumar, Pankaj, Murray, Bryan, Pourhedari, Homayoun, Vambenepe, William and Wurster, Klaus. Web Services Events (WS-Events) Specification v. 2.0. Hewlett-Packard Company.

http://devresource.hp.com/drc/specifications/wsmf/WS-Events.pdf

- [McCalla et al., 1992] McCalla, G. I., Greer, J. E., Barrie, B., and Pospisil, P. Granularity Hierarchies, International Journal of Computers and Mathematics with Applications (Special Issue on Semantic Networks), Vol. 23, pp. 363-375, 1992.
- [Bloom, 1956] Bloom, B.S. (Ed.) Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain. New York; Toronto: Longmans, Green, 1956.

Identifying Relevant Fragments of Learner Profile on the Semantic Web*

Peter Dolog

L3S Research Center University of Hannover Expo Plaza 1, 30539 Hannover, Germany, dolog@l3s.de

Abstract

The aim of this paper is to discuss how to identify distributed learner profile fragments on the semantic web. The learner profile fragments are modelled employing vocabulary suggested by several standards for learner profile. The learner profile fragments are maintained as standalone semantic networks of objects in RDF. The objects are instances of concepts labeled by terms from the standards. The identification of the profile fragments needed for example by adaptation services is performed as unification of identification records maintained on different sites. Queries sent to the edutella P2P network provide virtual views which connect those stand alone object networks. The queries can be constructed according to specific needs of personalization techniques, which can be provided as personalization services in a P2P learning network.

Keywords: Distributed User Modelling, RDF/RDFS, Learner Profile Fragment, Learner Profile Standards

1 Introduction

Recent advances in technologies for web-based education provide learners with a broad variety of learning content available. Learner may choose between different lecture providers and learning management systems to access the learning content. On the other hand, the increasing variety of the learning material influences effort needed to select a course or training package which will effectively build skills required for changed business situation. Adaptive support based on learner needs, background and other characteristics can help in selecting appropriate learning and during learning.

Information about a learner is crucial for enabling such adaptation. As the learner may take courses and training in different learning management systems, fragments of his profile are maintained on different sites. The systems should be able to collect those fragments to enable adaptation. This situation raises a question how to identify the relevant fragments of a learner profile distributed over the systems. In this paper we discuss an approach how to identify the distributed learner profile fragments in P2P environment. The fragments are maintained in RDF according to a vocabulary prescribed by standards for learner profiles.

The rest of the paper is structured as follows. Section 2 provides a sample scenario which drives the descriptions in the paper. Section 3 provides a discussion on simplified user conceptual model typically used in adaptive systems based on terminology taken from several learner profile standards. Section 4 discusses our approach to identification of learner profile fragments based on local identification schemes. Related work is discussed in the Sec. 5. Paper conludes with sumary and remarks on ongoing research (Sec. 6).

2 Sample Scenario

To motivate our approach we refer to a sample scenario. Alice is trying to improve her skills in programming of accounting software. She has a degree in computer science and experience in programming of a text editor. She is looking for a training package where she will experience common problems and approaches in programming the accounting software. Alice has an application to access and search a network of learning providers. Her profile about her learning performance at the university is available from the university provider. Her portfolio is available directly from her application.

As the situation shows, the Alice profile fragments have to be retrieved from several places. Those places usually use different identification mechanisms. For example, university identifies a learner by his matriculation number. The company has its own identification scheme for identifying its employees. Alice uses application which employs different identification scheme as well.

Figure 1 depicts the architectural outline for the Alice scenario. Alice accesses the provided courses through her personal learning assistant (PLA). The PLA uses the Edutella consumer to query connected systems. The PLA maintains the identification entries used at the previously accessed systems (the University and CompuTraining provider in our case). The university provider maintains Alice performance during her university studies. The training provider followed Alice performance in the course on programming accounting software and stores it in its metadata stores.

^{*}This work is partially supported by EU/IST ELENA project IST-2001-37264 (http://www.elena-project.org).



Figure 1: An architecture showing Alice scenario

Both learning providers could use additional external services which followed Alice performance. There are two possibilities to handle this situation. In the first case, the services maintain Alice performance records identified by their identification schemes. The learning provider provides a routing and mapping between its scheme and the service identification schemes. In the second case, the Alice performance from the services is stored at the learning provider. Both situations are possible, thus an algorithm for collecting learner profile fragments has to support both situations.

In addition, the providers and services could use different data models. Data integration problem has to be studied in this context, to be able to exchange learner profile fragments between learning services.

In the following, we will address issues related to learner identification on different distributed systems while the data model for learner profile fragments stays uniform.

3 Modelling Learner Features

Semantic web description formats allow us to express information as a network of associated objects described by a certain type. Therefore, each system, which Alice used to access her training or course, maintains a small network of objects describing Alice in each relevant node of the learning provision network.

The main concepts identified in scenario are performance, portfolio and learner as such. Current versions of learner profile standards provide vocabularies to describe such concepts as discussed for example in [Dolog and Nejdl, 2003]. The use of standards allows us to reduce variability in data models used to maintain learner profile records. For example, the IEEE PAPI describes learner performance as a learning experience measured by achieved competency value and portfolio as anything created during the learning experience or anything which supported the learning experience. Both concepts are described by its properties.

The performance and portfolio objects have to be associated to an object which represents Alice (instance of the Learner class). Such objects have several performance and portfolio records and possibly their real name. To enable multiple identifications (pseudonyms), the Learner class points to several identification records which belong to different systems (providers). This allows us to route requests to particular providers and to use object identifiers used at those providers. As the identification might be time limited, "valid to" and "valid from" dates can be associated to the identification records.

Figure 2 depicts a conceptual model needed for the Alice scenario discussed above. The conceptual model is an excerpt of the conceptual model used in Elena project. Further concepts have been considered, such as learner goal, preferences, competencies, and certificates.

4 Identification of Relevant Distributed Fragments of Learner Profile

According to above proposed conceptual model, any system can choose its own identification mechanism. The system can assign locally unique identifiers to distinguish learners. However, it is required to provide the identifiers according to the conceptual model described above. Learner accesses train-



Figure 2: Conceptual model for learner profile from Alice scenario

ing and courses through his personal learning assistant (PLA). However, the PLA uses learning provider services with own identification mechanism. The providers can expose learner identifier used to identify learner a performance record belongs to. The identifier can be then store together with the provider identifier at the learner's PLA. Similarly, if a learning provider accessed further external services, the learner identifiers at those services have to be provided.

Figure 3 depicts an excerpt of instances from Alice' performance and identifications. Alice is identified as Al at the university provider and as Li at the training provider. The model also contains an instance about her learning experience at the university in a course on programming. The learning experience at the CompuTraining in the course on programming accounting software is also depicted.

The parts of the figure which are overlapping are maintained independently at the providers. The unification of the identifications for particular systems is performed when an adaptive system/service searches for learner profile fragments needed for adaptation. The systems can maintain the learner profile fragments by learner API designed according to the learner profile fragments schemas. The API, schemas, and a system prototype for browsing such learner profiles can be found at http://www.l3s.de/~dolog/ learnerrdfbindings/.

Algorithm. Following algorithm applies when system searches for relevant fragment of a learner performance:

- Retrieve all instances of the Identification concept for current user;
- Search instances of the Learner concept on systems referenced in each identification entry;
- If there are further systems referenced in the identification records at the remote systems, reapply this algorithm with the records;
- Retrieve all objects as instances of concepts needed for adaptation (e.g. learner performance);

To illustrate the algorithm, let us refer back to the Alice scenario. We use the Edutella [Nejdl *et al.*, 2002] to submit queries to the P2P network. The Edutella P2P infrastructure

allows us to connect peers which provides metadata about resources described in RDF. Edutella also provides us with a powerful Datalog based query language, RDF-QEL. The query can be formulated in RDF format as well, and it can reference several schemas.

In the following we will use the QEL selection syntax where three parameters (subject, predicate, object) are used to retrieve instances of RDF classes. The syntax of such selection in QEL is s(subject, predicate, object). The selection will retrieve all the resources which contain assertions with the subject, predicate, and object. Any of those parameters can be used as variables.

As we assume a uniform data model suggested above, the query can be formulated in terms of the data model.

```
s(Alice, learner:identification, Ident),
    s(Alice, rdf:type, learner:Learner),
    s(Alice, learner:learner_id, LID),
    s(Ident, rdf:type,
        learner:Identification),
    s(Ident, learner:provider, PID),
    s(PID, rdf:type, learner:Provider)
    s(Ident, learner:ID, LPID).
s(Alice, learner:performance, Perf),
    s(Perf, rdf:type,
        learner:Performance),
    s(Perf,
        learner:learning_experience, LEX).
```

First, all identification records of Alice are retrieved together with local learning performance. The selection query for learner identifiers is constructed based on the Identification concept (the learner: prefix is an abbreviated namespace of the learner schema). The remote learner identification is maintained as a pair of provider and learner identifiers (PID, LPID) maintained in the provider and ID attributes. It is allowed to have one learner identifier valid for several providers. In that case, multiple pairs are retrieved. According to the Alice scenario, the program finds the identifications of the university provider and the CompuTraining provider.

The join selection for performance is constructed based on the Performance concept. The performance maintains a learning_experience attribute where an identifier of a resource taken during the study is stored.



Figure 3: An excerpt of instances of Alice's performance at different systems under different identification

As the external providers can have similar identification records for third systems, a query has to be constructed for each tuple (PID, LPID) to find the identifications at the third systems. A query to retrieve also performance values from the external systems is constructed similarly to the previous example.

```
s(Lremote, learner:identification,
  RemoteLID),
    s(Lremote, rdf:type, learner:Learner),
    s(Lremote, learner:learner_id,
      LIDremote),
    s(RemoteLID, rdf:type,
      learner: Identification)
    s(RemoteLID, learner:provider, PID),
    s(RemoteLID, learner:ID,
      LPID)
s(Lremote, learner:identification,
 RemoteLIDI)
    s(RemoteLIDI, rdf:type,
      learner: Identification),
    s(RemoteLIDI, learner:provider,
      PIDExternal),
    s(RemoteLID, learner:IDExternal,
      LPID).
```

If there is a non empty result set of identification entries for the third systems, the query construction is reapplied until there are no more systems to contact.

Discussion. The approach to distributed learner modelling presented in this paper is currently under development in EU/IST Elena project. The exchange model for learner profiles described in this paper has been implemented for example in the PLA [Dolog *et al.*, 2004a] and is currently under development in the Personal Reader [Dolog *et al.*, 2004b]. The

advantage of this approach is that it relies on standards for learner profiling which allows to construct uniform queries. The identification mechanism suggested here allows to use local learner identifiers and the mapping between them is performed according to the records which maintain learner identifiers at the neighboring providers and/or services. The records also provide us with routing information for queries, i.e. which providers to contact to retrieve additional information about learner.

There are some open issues which still have to be resolved. There is a very likely situation that the internal data model for learner profiles is different from the one based on standards. The providers have to support query rewriting functions to rewrite received queries into their internal data model. Another solution would be to provide mapping services between schemas employed as discussed for example in [Dolog *et al.*, 2004a; Simon *et al.*, 2004]. Another important problem is how to address different attribute value ontologies for example for concepts learned or competencies acquired. The ontology mapping has to be employed also in that case.

Another problem which is currently discussed is where to put the reasoning about the query construction. The queries for the algorithm proposed in this paper can be constructed at a mediator (e.g. the PLA). Another approach would be that each provider will be able to construct additional queries if there are external systems to be contacted according to the identification records. This would mean that each node in the network will construct and submit queries just to its neighbors.

In the case of inter-organization network, privacy and security issues has to be addressed to protect sensitive data. The identification mechanism has to be combined with distributed policies and credentials evaluation. Both, the identification records and learning related learner features, has to be protected and disclosed just to trusted parties. An important question in this context is how to protect information which was already disclosed to a system which is asked by third external system to provide the information.

5 Related Work

Work on integration of distributed user model fragments which are needed for specific task was presented in [Niu *et al.*, 2003; Vassileva *et al.*, 2002]. Their work similarly as our is based on an idea that just particular fragments in specific combination are needed for different computation purposes. In our work we applied the standard vocabulary to reduce negotiation overhead needed when the heterogeneous fragments and schemas are employed.

The identification through pseudonymity was applied also in [Kobsa and Schreck, 2003]. The pseudonymity was treated as kind of protection mechanism. Here we apply different pseudonyms in learner identification on different systems to compute relevant fragments of learner profile.

As the learner data are sensitive, the trust and security issues have to be further investigated. We have already proposed extensions towards standard based vocabulary for privacy purposes in learner modelling. Those extrensions and work reported in several other projects like [Kobsa and Schreck, 2003; Bohrer and Holland, 2000; Heckmann, 2003; Maler *et al.*, 2003] should be further investigated in the context of our approach.

Trust negotiation is another interesting related work in the context of open environment. First steps towards trust negotiation in open p2p network was presented in [Nejdl *et al.*, 2003]. Guarded distributed logic programs are used to encode policies and enforce them during resource attribute exchange and negotiation. We are currently investigating whether this approach is suitable also for learner attribute exchange. Our metamodel for policies to protect learner information is flexible and allow us to use any language when there is appropriate interpreter available. Appropriate subclass of the policy class is then used to identify which iterpreter to run.

6 Conclusions and Further Work

This paper reports on recent work for the development of learner profile for the ELENA project. We showed that we can definitely benefit from learner profiles standards because they provide a vocabulary which was agreed in a broader context.

The RDF and RDFS allow us to use different schemas and query languages such as QEL. The QEL allows us to integrate reasoning capabilities over personal profile in P2P network. This is step towards P2P RDF based environment where personalization techniques can be implemented as services.

Many issues still have to be resolved. The technical infrastructure for this approach to personalization has to be investigated in more detail and mechanisms for provision, searching, and using such personalization services have to be introduced. Mapping or mediating between different schemas should be investigated as well when we want to provide communication between different peers. Different identification schemes have to be investigated more deeply to support better exchange of learner profile fragments between distributed nodes. Experiments with analyzed privacy technologies and dynamically switching between them have to be investigated to support flexibility in open environment also in the context of security.

Acknowledgements. We would like to thank Wolfgang Nejdl and Tomaž Klobučar for extensive discussions which helped to improve this work. We would also like to thank anonymous reviewers for comments which helped to improve this paper.

References

- [Bohrer and Holland, 2000] Kathy Bohrer and Bobby Holland. Customer profile exchange (cpexchange) specification. http://www.cpexchange.org/, October 2000.
- [Dolog and Nejdl, 2003] Peter Dolog and Wolfgang Nejdl. Challenges and benefits of the semantic web for user modelling. In Proc. of AH2003 — Workshop on Adaptive Hypermedia and Adaptive Web-Based Systems, WWW2003 Conference, Hypertext'03 Conference, User Modelling Conference 2003, Budapest, Hungary, Nottinngham, UK, Johnstown, PA, USA, 2003.
- [Dolog et al., 2004a] Peter Dolog, Nicola Henze, Wolfgang Nejdl, and Michael Sintek. Personalization in distributed e-learning environments. In Proc. of WWW2004 — The Thirteen International World Wide Web Conference, New Yourk, May 2004. ACM Press.
- [Dolog et al., 2004b] Peter Dolog, Nicola Henze, Wolfgang Nejdl, and Michael Sintek. The personal reader: Personalizing and enriching learning resource using semantic web technologies. In Wolfgang Nejdl and Paul De Bra, editors, Proc. of AH2004 — International Conference on Adaptive Hypermedia, volume 3137 of LNCS, Einghoven, The Netherlands, August 2004. Springer.
- [Heckmann, 2003] Dominik Heckmann. Integrating privacy aspects into ubiquitous computing: A basic user interface for personalization. In *Proceedings of the AIMS 2003, Artificial Intelligence in Mobile System*, 2003.
- [Kobsa and Schreck, 2003] Alfred Kobsa and Jörg Schreck. Privacy enhancing technologies — approaches and development. ACM Transactions on Internet Technology, 3(2):149–183, May 2003.
- [Maler *et al.*, 2003] Eve Maler, Prateek Mishra, and Robert Philpott. Bindings and profiles for the oasis security assertion markup language (saml) v1.1. OASIS Standard, September 2003.
- [Nejdl et al., 2002] W. Nejdl, B. Wolf, C. Qu, S. Decker, M. Sintek, A. Naeve, M. Nilsson, M. Palmér, and T. Risch. EDUTELLA: a P2P Networking Infrastructure based on RDF. In *In Proc. of 11th World Wide Web Conference*, Hawaii, USA, May 2002.
- [Nejdl et al., 2003] Wolfgang Nejdl, Daniel Olmedilla, and Marianne Winslett. Peertrust: Automated trust negotiation

for peers on the semantic web. Technical Report, Learning Lab Lower Saxony, University of Hannover, November 2003.

- [Niu et al., 2003] Xiaolin Niu, Gordon McCalla, and Julita Vassileva. Purpose-based user modelling in a multi-agent portfolio management system. In Peter Brusilovsky, Albert T. Corbett, and Fiorella de Rosis, editors, *In Proc. of User Modeling 2003, 9th International Conference, UM 2003*, Johnstown, PA, June 2003. Springer, LNAI 2702.
- [Simon et al., 2004] Bernd Simon, Peter Dolog, Zoltán Miklós, Daniel Olmedilla, and Michael Sintek. Conceptualising smart spaces for learning. *Journal of Interactive Media in Education. Special Issue on Educational Semantic Web*, (9):221–239, May 2004.
- [Vassileva et al., 2002] Julita Vassileva, Gordon McCalla, and Jim Greer. Multi-agent multi-user modelling in I-Help. User Modeling and User-Adapted Interaction, 2002.

Intellect Disclosure Support Based on Organizational Intellect Model

Mitsuru IKEDA, Yusuke HAYASHI

School of Knowledge Science, Japan Advanced Institute of Science and Technology 1-1 Asahiidai, Tatsunokuchi, Ishikawa, 9231292 Japan {ikeda, yusuke}@jaist.ac.jp

Youhei TANAKA, Masataka TAKEUCHI, Riichiro MIZOGUCHI

The Institute of Scientific and Industrial Research (ISIR), Osaka University 8-1 Mihogaoka, Ibaraki, Osaka, 5670047 Japan {youhei, takeuchi, miz}@ei.sanken.osaka-u.ac.jp

Abstract

To establish the intellectual identity of the organization, it is important for every organization to revitalize creative activity inside and attract intellectual interest from outside the organization. This paper proposes a framework to support attraction of outside interest by disclosing the organizational intellect. We have developed models of organizational intellect and a support environment for creation and inheritance of organizational intellect based on the model. This paper proposes concepts to design attractive information for the outside in terms of intellectual activity, and a support system to disclose organizational intellect based on those concepts.

1 Introduction

The variety and growth of intellects in an organization are major sources of high competitive power for an organization [Nonaka 1995]. Regarding the growth of organizational intellect, it is important for each organization to exchange intellects not only internally, but also externally [Wenger 2002]. Lundkvist provides an example of importance of external exchange [Lundkvist 2004]. He analyzed the relation between a software company and the user group of the company's product. And then he reported that the users have played important role as innovators.

This study is intended to develop information systems to support both internal and external exchange. This paper specifically addresses the latter: it proposes a framework to support attracting intellectual interest from the outside by disclosing the organizational intellect effectively.

The next section will try to clarify our conception of the term "intellect" and then propose a model based on which computers support the creation and inheritance of organizational intellect. Section 3 presents an overview of intellect exchange support based on the model. Section 4 will discuss organizational intellect disclosure support further. In section 5 describes metadata for organizational intellect disclosure. Finally, section 6 present concluding remarks.

2 An Organizational Intellect Model

2.1 Intellect

The terms 'knowledge,' 'intellect,' and so on are used with various meanings, so there appear to be no definite meanings for them [Liebowitz, 1999]. Though it is difficult to define them strictly in a consistent manner, to show subjects of this study, we will take some exemplary definitions from the literature.

Brown and Duguid [Brown and Duguid, 2000] argue convincingly that knowledge is more than just information because it

- usually entails a 'knower',
- appears harder to detach than information, and
- is something what we digest rather than merely hold.

Tobin draws distinctions between data, information, knowledge, and wisdom [Tobin, 1996].

- 1. *Data*:
- 2. *Information*: = *Data*+ relevance + purpose
- 3. *Knowledge*: = *Information* + application
- 4. *Wisdom*: = *Knowledge* + intuition + experience

In this research, the term 'intellect' is used to express our idea similar to Brown and Duguid's argument about 'knowledge' and Tobin's 'wisdom'. Having an intellect means not only merely knowing something, but also digesting it through creation or practical use. It also means that the intellect cannot be separated from a person because it includes skill and competency. This is the major reason why we introduce the term "intellect." We aim to support creation and inheritance of organizational intellect by managing information concerned with intellect.

2.2 Dual loop model

Our goal is to present a framework of information systems that supports all the activities from the practical ones in an organization to ones oriented to knowledge creation and skill/competency development.

In this research, based on Dynamic Theory of Organizational Knowledge Creation [Nonaka 95], some activities related to the formation of organizational intellect are explained from both viewpoints of the 'individual' as the substantial actor in an organization and the 'organization' as the aggregation of the individuals.

The two viewpoints are modeled as two separated loops of activities with explicit interactions between them. The whole model called "Dual Loop Model [Hayashi et al., 2001]" is illustrated in figure 1. It works as the reference model for designing an intellect exchange support environment.

The dual loop model is constructed from formative process of an individual's intellect (figure 1 (A), personal loop) and formative process of organizational intellect (figure 1 (B), Organizational loop), and it represents the flow of intellects between them. Intellect creation activities in this dual loop model are explained in the following.

2.2.1 Personal loop

The personal loop is a loop of individual activities related to intellect. As shown in figure 1(A), it consists of four processes: *internalization*, *amplification*, *externalization* and *combination*.

Events in an organizational loop

This loop is considering two types of modes of individual activity; a learning mode, in which an individual acquires intellect from his/her surroundings, and a creative mode, in which he/she creates original intellect. A typical activity in the learning mode is one in which the members acquire intellect of which the significance is approved in an organization. Systems supporting the leaning and the creative modes can be considered the learning support and creative thinking support systems, respectively. A possible common basic requirement for supporting these two modes is:

 Easy access to useful intellect for intellect acquisition and creation activities.

This is closely equivalent to the considerations in the study of Ogata et al's knowledge awareness support [Ogata et al., 2000], kMedia [Takeda et al., 2000], and L-EVIDII [Ohira et al., 2001]. These researches aim to support individual activities in a community.

In our research, in addition, we aim to support making harmony between the individual activities and organizational activities which give direction to the individual activities based on a vision and strategy of the organization. We will describe the organizational activities in 2.2.2. We develop this idea in a framework that promotes the 'appropriate creation/distribution' of intellect in an organization based on knowledge management theory.

Basic requirements for each mode of the personal loop are:

- for the learning mode, preparing and implementing a rational learning process for an organization, and
- for the creative mode, supporting communication of intellect, e.g. acquiring knowledge and imparting it to others, as the basis of individual amplifying process.



Figure 1. Dual loop model (partly simplified)

In figure 1, nodes from 1 to 7 represent the events of the individual activities. Typical starting events for the learning and creative modes assumed in the dual loop model are nodes 5 and 7, respectively in figure 1. Node 5 represents an event in which 'significant intellect in an organization' should be acquired, and node 7 is an externally-triggered event that represents a start of the creation of original intellect. These are defined in connection with a user's activity conditions and an organization's loop events.

2.2.2 Organizational loop

An Organizational loop is an abstract model, reflecting members' activities in personal loops in an organization as intellect inheriting and creating activities from an organizational viewpoint. The typical activities include acquisition and creation of intellect inside and outside an organization.

The loop consists of four processes: *internalization, so-cialization, externalization* and *combination*. In figure 1, nodes 8, 11 and 12 represent the events that trigger off individual activities performed in the personal loop process. For example, node 8 represents such an event as 'intellect distributed by individuals', node 4 represents 'obtaining intellect from others'. The arrow from node 8 to node 4 shows a causal link between the two events.

Furthermore, this dual loop model can explain conditions of creation and inheritance of the organizational intellect. For example, an organization that frequently has events in the socialization process (at the top left) and rarely has events in the combination process (at the bottom right) mean that even though an individual actively carries out intellect acquisition and creating activities, they are not likely to be recognized as 'organizational intellect'. Lack of relation between activities of individuals and ones of the organization can be identified as the causes. Further, when an organization has events only in the internalization process in the Organizational loop (at the bottom left), it can be seen that a tendency of the organization leans to knowledge acquisition activity in practice. Thus, the dual loop model is also useful as a reference for analyzing the proper process of intellect acquisition, passing down and creation in an organization.

3 Overview of Intellect Exchange Support

This study explores the following important issues that support information systems:

(A) Revitalization of activities for creation and inheritance of organizational intellect

- Supplying guidelines to direct organization members to the desired process of creation and inheritance of organizational intellect.
- Encouraging organization members to become aware of the relationships among people, intellects, and vehicles. Through that awareness, they can derive answers to questions such as: *Who knows the intellect well? Who should collaborate? Which medium is useful to obtain the intellect?*

(B) Disclosing organizational intellect to the outside

- Clarifying the intention of disclosure based on a deep understanding of the organizational intellect.
- Producing a presentation with the most suitable style for showing the intellect.

Figure 2 shows an overview of this project, focusing on (B). The dual loop model (DLM) and intellectual genealogy graph (IGG) form a foundation to provide awareness information on the organizational intellect for both organizational members and outsiders. That awareness will involve not only the meaning of the intellect itself but also its formative process.

As mentioned in the previous section, DLM represents a formative process of intellect in an organization, both from the viewpoint of the individual and the organization. This



Figure 2. An overview of intellect exchange support

model serves the members as a guideline for organizational activity and the organization as a reference for analyzing its condition of creation and inheritance of the organizational intellect.

IGG represents chronological correlation among persons, activities, and intellect in an organization as an interpretation of the activities of organization members based on DLM [Hayashi et al., 2002]. IGG is generated from activities with vehicles. A vehicle is a representation of intellect that mediates intellect among people: text, figures, voice, and so on.

On the similar lines of research, ScnolOnto Project [Buckingham et al., 2000] aims to model the formative process of ideas in the academic community, paying attention to the *claims* described in vehicles (research literature). They propose a model for authors or readers to describe their interpretation of the *claims* on the vehicle and relationship among them. On the other hand, our study is interested in semi-automatic extraction of IGG by DLM-based interpretation of the observed activities, such as creating/revising/referring the vehicles.

A site-map is a model describing the structure of intellects to disclose. The model consists of a content level model and a presentation level one. The content level model is a subset of an IGG. That level model is extracted with the intention of disclosing the organizational intellect. The content level model is transformed into the presentation level model to allow its display on a WEB browser.

Based on these models, this project is intended to develop information systems to support both (A) and (B), as mentioned above. To support (A), it is crucial to prompt the members' spontaneous activity by providing organizational intellect awareness information based on IGG, as well as to direct their activity by presenting guideline on the activity along to DLM. On the other hand, to support (B), it is crucial to prompt the organization to grasp a comprehensive view of its own organizational intellect by also presenting IGG to enable the organization to prepare its best materials for disclosure. Moreover, it is important to prepare a mechanism for conversion from the content of disclosure to its presentation.

This brief paper is insufficient to allow comprehensive discussion on all the aspects presented in the previous section. This paper specifically addresses features of the framework focusing on Intellect disclosure support function. In the following, we will see the model and support functions for organizational intellect disclosure. Detailed explanation about DLM and IGG can be found in [Hayashi et al., 2001, 2002].

4 Organizational intellect disclosure support

The outline of tasks that designers carry out in the site-map design process is the following. During the content level model design process, designers select an organizational intellect that will be disclosed to the outside. And then, in the presentation level model design process, designers decide how to display the organizational intellect to the outside. This is similar to common tasks undertaken during website design.

This section presents concepts and support functions related to the design processes.

4.1 Concepts for organizational intellect disclosure

Site-map model consists of the presentation level model and the content level one. Table 1 summarizes concepts to describe the content level model and the presentation level one. The content level model describes meaning and intention of the disclosure information. The presentation level model is embodied as web pages displayed to the outsiders.

Most important of all, the relation between the presentation level model and the content level one describes the contextual information of intellect disclosure, that is, the relation between meaning and intention of the disclosure information and the embodiment of it as web pages. The base unit of mapping between the content level model and the presentation level model is a *Description* and a *Page*. The content level model plays a role of metadata for the corresponding presentation level model.

4.2 Support functions for organizational intellect disclosure

This study aims to design and develop an organizational intellect disclosure support environment. Here we will see necessary functions of the environment.

• Lead designers into coordinating content and intention of disclosure: The concepts mentioned in table 1 are provided as a basis of Site-map design for designers through the environment. Those concepts facilitate designers' recognition of the importance of coordi-

Table 1. Site-map model concepts

Level	Concept	Explanation
Content level	Description	Description of a person, an intellect, a vehicles and an activity in IGG
	Attractive frame	A network of <i>descriptions</i> to be disclosed to the outside This is extracted from IGG with the organization's
		intention.
	Subject	A description of a person, an intellect, a vehicle or an activity that is presented as a subject of an Attractive frame
	Related items	Descriptions presented together with the Subject
	Theme	Description of intention of an Attractive frame.
	Subject	A person, an intellect, a vehicle or an activity that is a noteworthy item in the Attractive frame. It corresponds to
		the <i>subject</i> of the <i>attractive frame</i> .
	Purpose	Expectant effects of the attractive frame on the outside.
	Perspective	Necessary relations to display the <i>Subject</i> attractively according to the <i>purpose</i> .
	Site pattern	Pattern of extraction of an attractive frame from IGG.
Presentation level	Page	A web page that expresses a <i>description</i> .
	Cluster	A network of <i>pages</i> that corresponds to an <i>attractive frame</i>
	Cluster top page	A page that corresponds to the <i>subject</i> of an <i>attractive frame</i> .



Figure 3. The generation of a site-map model

nating disclosure content and intention.

- *Provide the lines of thought in Site-map design by Site pattern*: The site pattern describes noteworthy relations in IGG according to the intention of the disclosure. Based on the description, the environment provides for designers with the candidates for *Attractive frames* as reference information.
- Convert the content level model to the presentation level model: The environment converts the content level model, which is represented by RDF, to the presentation level model, which includes web pages represented in HTML.

Figure 3 shows a site-map model generation image. Fig. 3 (C) shows an image of one of the web pages resulting from the generation. The web pages are included in the presentation level of the site-map model. Fig. 3 (C) displays a new paper just submitted to an international conference and the hyperlinks to those people, intellects, vehicles, and activities that are related to the intellect of the paper. The hyperlinks in this web page are set based on relations in the IGG (A).

Figure 3(A) shows an IGG, which has all nodes and links that are retrospectively accessible from the subject paper. Arrows indicate direct links among people, intellects, vehicles, and activities. A typical directed link means, for example, that a destination intellect is derived from a source one. The relations reflected in the hyperlinks are selected by the designer according to the intention of the disclosure. In this case, the intention specifically addresses the organization members' contribution to the subject paper. Tracing the links retrospectively from the subject in IGG, the designer prunes away confidential and irrelevant nodes to secure the disclosure information and render it to be easily understandable by outside entities. Figure 3(B) shows an attractive frame extracted from IGG(A). The attractive frame has broken lines, which indicate that the links have been pruned away. The remaining nodes are important activities or intellects in the formative process of the subject. This extraction can reveal relations that are not described clearly in the vehicle. Finally, this model is converted to a web page as a presentation level model.

5 Metadata for Organizational Intellect Disclosure

This paper defined the framework to describe contextual information of the organizational intellect. Contextual information includes people and vehicles that relate to the intellect, and the intellect's role. That contextual information is extracted from IGG. Metadata describing the contextual information are shown in Fig. 4.

These metadata show that a person made medium#1, named ontological engineering, with intellect#1 through intellectlevelactivity#1. The meta data elements are defined in DLM ontology. A part of the ontology described with RDF schema is shown in Fig. 5.

6 Concluding Remarks

This paper discusses organizational intellect disclosure support. That support is intended to activate intellect exchange and growth of mutual understanding among organizations. This study will also accumulate site patterns and develop a support environment using semantic web technologies.







Figure 5. DLM ontology using RDF Schema

References

- [Brown and Duguid, 2000] Brown J. S. and Duguid P. *The Social Life of Information*, Harvard Business School Press, Boston, 2000.
- [Buckingham et al., 2000] Buckingham Shum, S., Motta E. and Domingue J. ScholOnto: An Ontology-Based Digital Library Server for Research Documents and Discourse, *International Journal on Digital Libraries*, 3 (3), pp.237-248, 2000.
- [Hayashi et al., 2001] Hayashi, Y., Tsumoto, H., Ikeda, M., and Mizoguchi, R. Toward an Ontology-aware Support for Learning-Oriented Knowledge Management, *Proc. of ICCE*'2001, pp.1149-1152, 2001.
- [Hayashi et al., 2002] Hayashi, Y., Tsumoto, H., Ikeda, M., and Mizoguchi, R. An Intellectual Genealogy Graph -Affording a Fine Prospect of Organizational Learning-. *Proc. of ITS 2002*, pp.10-20, 2002.
- [Liebowitz, 1999] Liebowitz J. (Eds.): Knowledge Management Hand Book, CRC Press, 1999.
- [Lundkvist 2004] Lundkvist, A. User Networks as Sources of Innovation, *Knowledge Networks: Innovation Through Communities of Practice*, pp. 96-105, 2004.

- [Nonaka and Takeuchi, 1995] Nonaka, I., and Takeuchi, H. The Knowledge-Creating company: How Japanese Companies Create the Dynamics of Innovation, Oxford University Press, 1995.
- [Ogata et al, 2000] Ogata, H, Matsuura, K., Yano, Y Active Knowledge Awareness Map: Visualizing Learners Activities in a web Based CSCL Environment, *Proc. of NTCL2000*, pp.89-97, 2000.
- [Ohira et al, 2001] Ohira, M., Nakakoji, K., Kurakawa, K., Yamamoto, Y.: L-EVIDII: an Environment for Supporting Small-Group Project-based Collaborative Learning, *Transaction of Japanese Society for Information and Systems in Education*, Vol. 18, No, 3,4, pp. 328-339, 2001 (in Japanese).
- [Tobin, 1996] Tobin D. Transformational Learning: Renewing Your Company thought Knowledge and Skills, John Wiley and Sons, 1996.
- [Takeda et al, 2000] Takeda, H., Matsuzuka, T. and Taniguchi, Y.: Discov-ery of Shared Topics Networks among People - A Simple Approach to Find Community Knowledge from WWW Bookmarks -, *Proc. of PRICAI* 2000, pp. 668-678, 2000.
- [Wenger 2002] Wenger, E. et al. *Cultivating Communities of Practice: A Guide to Managing Knowledge*, Client Distribution Services, 2002.

E-portfolios for Meaningful Learning and Automated Positioning

Tommy W. Nordeng, Stian Lavik and Jarle R. Meløy

Cerpus AS, N-8432 Alsvåg, Norway {tommy, stian, jarle}@cerpus.com

Abstract

When it comes to lifelong learning and e-learning, we are steadily moving towards distributed and self-organized networks where multiple content providers offer parts and pieces, not complete vertical systems. This spurs development of new methods and techniques to position learners in these networks. Positioning requires that characteristics of the learner are mapped onto characteristics of learning materials and curricula. In this paper we describe BrainBank Learning, a web application for construction of individual topic maps as a mean for learning, and discuss the potential of such knowledge maps for automated computersupported positioning. We also present current work on developing, evaluating and utilizing topic maps-based applications to support meaningful learning and deeper understanding.

1 Introduction

Web-based learning has in general become more popular in education and business training. A lot of computer-aided learning software exist to aid learning, web applications as well as offline systems. The tools vary from customized learning applications to edutainment and simple communication systems. However, abundant digital resources and tools do not necessarily solve any problems if they by the end of the day contribute to increase the chaotic pressure of information on the learners. The main problems related to using educational hypertext for learners are cognitive overload, disorientation and distraction, poor narrative flow, and poor conceptual flow [Jacobson et al 1996].

Educational practices are changing from being predominantly teacher-led to largely student-centered. But how can the students themselves be able to assess their position relative to a future learning environments consisting of a diverse set of learning activities from which learners somehow may take their pick? The learner's history and goals define an entry position relative to the learning activities. A different entry position is likely to result in a different partition of the set of available activities in activities to skip and to complete. Different entry points will thus result in different paths through the set of relevant learning activities. Computer supported positioning in learning networks could contribute to the formidable set of hurdles that arises in such a scenario. In fact, it assumes answers on a substantial number of the research questions that were recently proposed for intelligent information systems [Cherniavsky et al 2002]. In this article we focus on how the learner's history can be recorded and stored in electronic portfolios.

Electronic portfolios have traditionally been defined as an organized collection of digital and/or analog artifacts and reflective statements that demonstrate growth over time [Barett 2001]. In a broader perspective e-portfolio has been defined as a tool that can provide sophisticated control of one's virtual identity [Treuer et al 2003]. A fundamental characteristics of an e-portfolio in this perspective is that the virtual identity is stored using a common set of functional and organizational standards. Wilbert Kraan puts it this way: "Without the means to output e-portfolio data in a standard format, it's next to useless" [Kraan 2003].

Topic Maps [Park 2002] is a hypertext navigation metalayer above electronic information sources supporting topical finding of various kinds of resources, e.g., documents, graphics, images, database records, audio/video clips. As a result of a special characteristic of the topic maps model is a clear separation between the description of the information structure and the physical information resources. The navigation meta-layer is independent of the format of the actual resources and enables the creation of an external index that makes the information findable. The main topic maps components are topics, associations, and occurrences [Biezunski et al 1999]. Using those elements, one can create maps in document repositories.

The topics represent the subjects, that is, the things that are in the application domain, and make them machine process-able. They can have zero or more topic types and also have names (a base name and variants for use in specific processing contexts). A topic association represents a relationship between topics. Associations can have types (e.g., illustrated by, example of, written in, etc.) and define roles of the participating topics (e.g., example—concept description; prerequisite— result; document—language). Occurrences instantiate topics to one or more relevant information resources. An occurrence can be anything; most often it is a URI or a document (article, picture, video, etc.). Scope defines the extent of validity of a topic characteristic assignment: the context in which a name or an occurrence is assigned to a given topic, and the context in which topics are related through associations. One useful and potentially very powerful application of scope is to permit capture of different viewpoints of the subject. Another important concept in TM is identity. Two topics are the same if both have the same name in the same scope or both refer to the same subject indicator. The topics and all their characteristics could be merged if this condition holds. Topic maps provide a language to represent the conceptual knowledge with which a student can distinguish learning resources semantically. Apart from their major purpose of indexing information resources, topic maps embody knowledge. A semantically rich topic map would enhance the value of a teaching unit. Moreover, topic maps are very suitable for representing the course unit ontological structure.

In a recent paper [Dichev et al 2004] discuss the advantages of Topic Maps in education from 3 perspectives; the learners', authors' and courseware developers' perspectives. Authors will benefit from knowledge externalization support, effective management and maintenance of knowledge and information, augmented learning space beyond teaching space, rapid and efficient courseware development, collaborative authoring and personalized courseware presentations. For courseware developers Topic Maps supports building ontology-aware applications, open ended learning environment, adaptive educational applications and courseware templates. It offers enhanced navigational retrieval tools, reuse, sharing and interoperability, and TM tools and API's are available from several, including open, sources. Moreover, for the learners Topic Maps offers efficient contextbased retrieval of relevant online information, acquisition of new topical knowledge, deeper understanding of the domain conceptual relations, better information comprehension true visualization, domain navigation and browsing awareness, and finally customized views, adaptive guidance and context based feedback

For some reason, Topic Maps has so far not been utilized extensively for education and learning purposes. In fact, we are only aware of a few such commonly available applications. [Dicheva et al 2004] has recently developed TM4L (Topic Maps for Learning), a framework supporting the development of ontology-aware repositories of learning materials. Our contribution, BrainBank Learning, has been focused on learning, rather than education. The main goal of the work presented in this paper was to build a bottom-up application where the learners can construct their own learning ontology and curriculum during a course or a complete study.

2 Results and Discussion

2.1 BrainBank Learning

BrainBank Learning (BBL) [Lavik et al 2004a] was developed as a web application (http://brainbank.no) for learning of concepts (their content) and associations (how concepts relate). It works with standard Internet browsers, which means that educational institutions are not dependent on any other installation to use the application. Users enter the application through individual accounts. Topics (concepts) that the learner meets during education activities are entered and described. The topics can then be connected by linking phrases to form propositions or associations: The learner creates his own associated network of topics that represents his knowledge. This way of documenting in the learning process is good for the learner's understanding of the area of study (placing knowledge in a context), as well as navigating and overview of the acquired knowledge later on. To further describe topics and associations, digital resources such as documents, pictures, movie clips and sound clips can be attached to the topics. These resources can be either linked to or uploaded to and stored in BrainBank. BBL is based on the Topic Maps standard, including the XML format supporting the Topic Maps ISO standard (XTM) [Topic Maps 2001] and it was implemented using the Ontopia Knowledge Suite (http://ontopia.net). As the Topic Map standard defines an effective way of representing information, through topics and associations etc. [Biezunski et al 1999], BBL now uses this Topic Maps technology to represent the data in the application.

A case study has been done to evaluate practical use of BBL and to find out if it helps improve learning to become more effective. The project has been a cooperative effort between Krsitin Bjørndal at PLP (Program for learning and practical pedagogy at the University of Tromsø), Cerpus AS and Alsvåg primary and secondary school. The project has been reported [Bjorndal 2003] and thoroughly discussed [Lavik et al 2004b] elsewhere and the main focus here is rather to catch on with unleashed potential and prospects for improvement. Based on the replies from the pupils (in interviews), three separate aspects were identifiable: BBL as an e-porfolio, as a learning strategy and as a medium and method for assessment.

2.2 E-portfolio

The pupils expressed that they would prefer to structure and store their knowledge in BrainBank rather than in paper notebooks. Pupils often think of repetition of learnt material as boring, but it is widely acknowledged that repetition is one of the best ways of storing knowledge. Seven out of the group of sixteen pupils stated that BBL helped in remembering what they had learned. According to these pupils, BBL mainly helped because they could easily go back and take a look at what they did earlier, what they had written down of keywords and associations (e.g.: "We can save things, so we won't forget it. It's simply to enter BrainBank, and there we have it. It's easy to save and easy to retrieve. We learn more and more through the years.") The same pupils said that they regularly used BrainBank to repeat for themselves what they had learned (e.g.: "You kind of get a repetition of what is learned when typing it into BrainBank. When I'm in 9th grade, I can look back on what I learned in 8th grade."). The pupils also expressed that they were motivated to document

their knowledge thoroughly by the fact that it is properly stored: "I'm so proud when I see how many keywords I've got in BrainBank!" one pupil said.

Some criticism has been raised by both pupils and teachers on the way hierarchical structures are built in BBL. Although BBL is related to the central ideas of concept mapping (a pedagogic method) as defined by Joseph Novak [Novak 1990; Novak 1991; Novak et al 1991], it differs by the fact that it does not demand knowledge to be expressed hierarchical. On the contrary, with BBL a user can build and browse complex multi-directional associative structures, across context and disciplines. It is however quite possible to build hierarchical structures using the Topic Maps standard. Even within the standard itself, there is support for typing in a hierarchical manner. In the learning process, hierarchy and (not least) typing can be quite useful to understand structures and trees of concepts. For example, it is valuable knowledge in itself to know that 'cat' is a mammal. And as long as it is also known that mammal is an animal, 'cat' will have to be an animal, which represents even more knowledge. We will include ways to build hierarchic structures in upcoming versions of BBL by implementing topic types. Topic typing is an efficient way to express simple propositions like 'cat is a (type of) mammal'.

In concept mapping, the idea of focus is essential [Novak 1991]. There is always some kind of focus point where the mapping starts. Defining a context will always increase the value of information and knowledge. By somehow telling that a particular view of a piece of knowledge belongs to a particular context, it is easier to relate new chunks of knowledge to pre-acquired knowledge, and it is also easier to see the purpose of the knowledge in its current location. In addition, especially with young learners, it is important to be able to divide the knowledge into manageable chunks during the learning process. We will include a feature in BBL that makes it possible to create themes, and to use a theme to build a small knowledge map within the boundaries of the perspective. BBL is all the way centered on topics, and a theme can consist of one or more topics. As the learner acquires new knowledge and relates it to the theme, it makes sense right away, at least where it is put. Later on, as deeper understanding develops, the knowledge map belonging to the theme can be merged into the learner's main (complete) knowledge map. However, the theme is still kept as an identity to allow focused navigation, searching, etc.

Some pupils did complain that BBL is suffering from the lack of a powerful visualization of concepts and their relations. Numerous reports have documented the power of the concept mapping. Implementation of graphical edition of concept map-like structures in BBL would thus substantially increase the value of the tool as a pedagogical method for meaningful learning. BBL has now implemented Ontopia's VizigatorTM, a generic Topic Maps visualization tool developed by Ontopia [Ontopia], and we also intend to enable editing of such graphically visualized maps. Ontopia's Vizigator™ based TouchGraph's is on (http://touchgraph.sourceforge.net/) technology for visualizing map structures using Java Swing components. Another important strategy is to support, direct and/or indirect import/export from front end software for mind mapping and concept mapping. Interestingly, CmapTools [Canas et al 2004] already supports XTM 1.0, and we believe that the concept mapping community should strive to decide on a common standard, preferentially XTM, for digital concept maps.

Successful learning often takes place within a sosiocultural context where an interaction between humans is essential [Vygotsky 1978]. Interaction between the learner and the teacher is supported in BBL. However, cooperation between peers is widely accepted as a useful way of learning and some pupils did ask for such features. The ability to work in projects, where peers have equal access to all project resources is one attractive way of doing this. The project members should be able to share resources from their personal brainbanks with the project, as well as accept the project resources and import them into their brainbanks. Moreover, learners should be able to share knowledge maps and resources with the world by publishing them on public searchable web pages, free for anyone to browse. It is expected that this may help exchange of knowledge between peers. It is also important for the interaction between learner and teacher to have the ability to share resources. The teacher needs to be able to transmit resources to the learners. This could be possible in several ways, but as a principle, the learner should have to actively accept new resources. This is to ensure that the learner always is actively aware that he has received something new that can be used in his own knowledge structure. Furthermore, teachers should be able to share learning resources, from complete ontologies to simple learning objects, so that developed learning resources could be reused not only by the developer, but also colleagues and other teachers.

Research data indicates that the learners need curriculum and ontological support to responsibly record and manage their e-portfolios [Treuer et al 2003]. In an ongoing project, Dichev et al [2004] aim to develop ontology-based courseware that supports learners in their reflection on knowledge, and that students can use to navigate and search course related material by broadly understood categories. With Topic Maps-based digital course libraries coming up [Dicheva et al 2004], it will be very interesting to study how successful students can construct individual knowledge maps with predefined ontologies as a knowledge backbone. However, as helpful as it is to have good tools for individual learning, the world of information we live in is more and more based on networking and interaction with many instances and sources. The Internet is no doubt an important source for information, but the amount of information out there is so vast and overwhelming that new and better methods are needed to search and navigate. A useful point when trying to retrieve information from digital sources would be: What exactly is the learner's current knowledge in the area in question? Could we in some way analyze the learner's already acquired knowledge to help him locate new information that is relevant to him in his current position?

If representing the knowledge using a map like structure one could try to build some sort of mechanism that could analyze the documented knowledge and search the digital sources with the outcome of that analysis to determine what information that really is relevant. Leake et al [2004] have developed a model for this using concept maps. They used the locations and relations of the nodes in the concept map to automatically create queries for the Google internet search robot. As the hierarchical structure of a concept map supplies means to weigh concepts used in a semantic web search the new Theme feature in BBL can do a similar job. It gives a main topic (the perspective) that gives the boundaries for the scope. It is possible in the topic maps structure in BrainBank to start with one main topic and then count the radius: The distance from the main topic to other topics. This can also indicate a topic's level of relevance and it can help balancing the search and make it more accurate. However, because Google is not able to analyze what any retrieved page is actually about it is likely that such queries still would result in a lot of false positives. We are currently building a search function that automatically uses associations in BBL to focus the queries, and linguistic characterization and indexing to match the retrieved document content to the queries. Hopefully, this function will allow the users to spend less time on browsing and more time on learning.

2.3 Learning Strategies

BBL was designed to be a tool for meaningful learning [Ausubel et al 1978] within a constructivist learning environment [Wilson 1996; Jonassen et al 1999]. The tool was inspired by the ideas of knowledge building developed by Joseph Novak and colleagues [Novak 1977]: It stimulates the learning process as the learner continuously reflects through and updates his own knowledge and stores it in BrainBank. This is because he has to discriminate received information to extract the essence of the information to document it in BBL, and also by relating new information to already existing knowledge by associating new topics to existing ones and describing the relation between them [Novak 1990; Novak 1991]. Some of the pupils said that they now pay more attention to how they are learning and made explicit statements that indicates that they have started a process of reflecting on their own learning process as such (e.g.: "You become more aware of what you read when writing keywords: You pay more attention. When I do my homework more in-depth, because I'm going to find keywords."). One of the main conclusion drawn by [Bjørndal 2003] is that BBL is a good learning strategy. Assessment

Another important issue that came up during the project is how detailed and how often the teachers should evaluate the pupils. BBL opens for both formative and summative methods for evaluating students, which makes it a promising instrument for modern forms for education. Teachers and supervisors can at any time take a look at what their students and pupils has documented in BrainBank. This way, they both evaluate progress and the knowledge documented. By examining the associations the students have made between topics, the teacher gets an impression on how much the students really understand of the area of study as well. However, for the teacher this kind of detailed evaluation of many pupils is time consuming. Even if this challenge is not related directly to BBL (a teacher can simply choose not to use it for evaluation) the new options of assessment bring this issue out into the light. A possible answer to this could be to automate the analysis of the end product (summative assessment) by using techniques like latent semantic analysis [Landauer et al 1998] or by comparing topic maps: Several tools for comparing concept maps have been described [Chang et al 2001; Biswas 2001], but such systems are often restricted to particular subject domains, vocabularities and even map building environments. The Reasonable Fallible Analyzer [Conlon et al 2004] strives to be flexible in this respect: When comparing a map with any other map (for instance an expert map) it is honest and says it is likely to be wrong. The point is that the learner becomes aware of similarities and differences between different maps, and by arguing with the program, deeper understanding will be achieved. Results from a practical case [Conclon 2004] suggest that the Reasonable Fallible Analyzer is a promising tool for formative self-assessment, and at least with respect to time consumption a good alternative to diagnostic assessment done by the teachers with shortage of time.

3 Conclusion

As e-learning still strives to honor it's promises it is getting increasingly complex, partly because it deals with one of the most intricate disciplines in modern research: human cognition. Development and improvement of methods and techniques for handling different levels of granularity and use of networking needs to coincide with development of and commitment to standard ways of handling the increased complexity. These methods and techniques should be focused on the learners, rather than merely teacher-led. BrainBank Learning unleashes powerful support for learning, and it does so using a technological standard that is inherently fit for the purpose. There is a huge potential for improvements on several areas, such as peer cooperation, assessment and positioning. We will continue our mission and aim to develop these and others areas in close relation with learners and teachers and pedagogical and technical researchers.

References

- [Jacobson et al. 1996] M. Jacobson, C. Maouri, P. Mishra, & C. Kola. Learning with hypertext learning environments, *Journal of Educational Multimedia & Hypermedia*: 5, 239-281.
- [Cherniavsky et al. 2002] J. C. Cherniavsky & E. Soloway. A survey of research questions for intelligent informa-

tion systems in education. Journal of Intelligent Information Systems: 18, 5-14.

- [Barett 2001] H.C. Barrett. Electronic Portfolios in Educational Technology Encyclopedia: http://transition.alaska.edu/www/portfolios/encyclopedia entry.htm
- [Treuer et al 2003] P. Treuer & J. D. Jenson. Electronic Portfolios Need Standards to Thrive, *Educause Quarterly*. 2, 34-42.
- [Kraan 2003] W. Kraan. The Open Source Portfolio Initiative releases 1.0 in CETIS: http://www.cetis.ac.uk/content/20030820153519
- [Park et al 2002] J. Park & S. Hunting. XML topic maps: Creating and using topic maps for the webmation retrieval, and application of AI methods and tools, Addison-Wesley, Moston, MA.
- [Biezunski et al 1999] M. Biezunski, M. Bryan & S. Newcomb. ISO/IEC 13250: 2000 topic maps: Information technology -- Document Description and Markup in http://www.y12.doe.gov/sgml/sc34/document/0129.pdf
- [Dichev et al 2004] C. Dichev, D. Dicheva & L. Aroyo. Using Topic Maps for Web-based Education, *Advanced Technology for Learning*: 1, 1-7.
- [Dicheva et al 2004] D. Dicheva, C. Dichev, Y. Sun, & S. Nao. Authoring Topic Maps-based Digital Course Libraries. In Proceedings of SWEL'04 @ AH'04 workshop.
- [Lavik et al 2004a] S. Lavik & T. W. Nordeng. BrainBank Learning – Building Topic Maps-Based E-portfolios. In the Proceedings of the 1st International Conference on Concept Mapping, Pamplona, Spain.
- [The Topic Maps.org Authoring Group 2001] The Topic Maps.org Authoring Group. XML Topic Maps (XTM) 1.0. The Topic Maps Consortium: http://www.topicmaps.org/xtm/index.html
- [Biezunski et al 1999] M. Biezunski, R. Bryan & S. Newcomb. ISO/IEC 13250:2000 Topic Maps: Information Technology -- Document Description and Markup
- [Bjørndal 2003] K. E. W. Bjørndal. Sluttrapport fra prosjektet BrainBank ved Alsvåg barne- og ungdomsskole i Øksnes kommune (Report from the BrainBank project at Alsvåg primary and secondary school)
- [Lavik et al 2004b] S. Lavik & T.W. Nordeng. BrainBank Learning - building personal topic maps as a strategy for learning in Idealliance, http://idealliance.org/papers/dx_xmle04/papers/04-02-03/04-02-03.html.
- [Novak 1990] D. Novak. Concept maps and Vee diagrams: Two metacognitive tools for science and mathematics education. *Instructional Science*. 19, 29-52.
- [Novak 1991] J. D. Novak. Clarify with concept maps. *The Science Teacher*. 58, 45-49.

- [Novak et al 1991] J. D. Novak & J. Wandersee. Special Issue on Concept Mapping in *Journal of Research in Science Teaching*.
- [Cañas 2004] A. J. Cañas, Hill, G., Carff, R., Suri, N., Lott, J., Eskridge, T., Gómez, G., Arroyo, M. & Carvajal, R. CmapTools: A Knowledge Modeling and Sharing Environment. In Proceedings of the 1st International Conference on Concept Mapping, Pamplona, Spain.
- [Vygotsky 1978] L.S. Vygotsky. Mind in Society, Harvard University Press, Cambridge, MA.
- [Leake 2004] D. Leake, A. Maguitman, T. Reichherzer, A. Cañas, J. Carvalho, Marco, Arguedas, Marco & Tom Eskridge. "Googling" from a Concept Map: Towards Automatic Concept-Map-Based Query Formation. In Proceedings of the 1st International Conference on Concept Mapping, Pamplona, Spain.
- [Ausubel et al 1978] D. P. Ausubel, Novak, J.D. & Haneisan, H. (1978) Educational Psychology: A Cognitive View, 2 edn, Holt, Rinehart and Winston.
- [Wilson 1996] B. Wilson. Constructivist learning environments: Case studies in instructional design, Educational Technology Publications, New Jersey.
- [Jonassen et al 1999] D.H Jonassen & Rohrer-Murphy, L. Activity theory as a framework for desinging constructivist learning environments, Educational Technology Research and Development. 47, 61-79.
- [Novak 1977] J. D. Novak. A Theory of Education, Cornell University Press, Ithaca, New York.
- [Landauer et al 1998] T. Landauer, Foltz, P & Laham, D. An introduction to latent semantic analysis, Discourse Processes. 25, 259-284.
- [Chang et al 2001] K. Chang, Sung, Y. & Chen, S. Learning through computer-based concept mapping with scaffolding aid, *Journal of Computer Assisted Learning*. 17, 21-33.
- [Biswas et al 2001] G. Biswas, Schwatz, D. & Bransford, J. Technology Support for Complex Problem Solving in Smart Machines in Education (Forbus, K. & Feltovich, P., eds), MIT Press, Boston, MA.
- [Conlon et al 2004] T. Conlon & Bird, D. 'Please Argue, I Could be Wrong': a Reasonable Fallible Analyzer for Student Concept Maps. In Proceedings of the World Conference on Eductional Multimedia, Hypermedia and Telecommunications, Lugano, Switzerland.
- [Conlon 2004] T. Conlon. 'But is My Concept Map Any Good?': Classroom Expereience With the Reasonable Fallible Analyzer. In Proceedings of the 1st International Conference on Concept Mapping, Pamplona, Spain.

Building of an Ontology of the Goal of IT Education and its Applications

Toshinobu Kasai, Haruhisa Yamaguchi

¹Faculty of Education, Okayama University 3-1-1 Tsushima-naka, Okayama, 700-8530 Japan {kasai, hyamagut}@cc.okayama-u.ac.jp

Kazuo Nagano

University of the Sacred Heart 4-3-1 Hiroo Shibuya-ku, Tokyo, 150-8938 Japan nagano@fushigi.net

Riichiro Mizoguchi

Institute of Scientific and Industrial Research, Osaka University 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047 Japan miz@ei.sanken.osaka-u.ac.jp

Abstract

In Japan, interest in IT education has continued to grow. Most goals of IT education involve meta-ability, which cannot be fully learned by traditional Japanese instructional methods. It is difficult to design effective IT education materials, and at present there are few specialists in IT education. For this reason, it is necessary and important to provide IT instructors with a powerful help system that can locate and provide access to a variety of useful information resources. To that end, we are building a system that reconstructs the resources according to the various viewpoints based on an ontology of IT education we built. Further, we propose a framework to make use of the results of another ontology by alignment of these ontologies based on Semantic Web technology.

1 Introduction

As a result of widespread use of the Internet and the development of numerous large information systems, the necessity and importance of information technology (IT) education have increased. In Japan, high school instruction in IT began in April 2003. However, most of the IT teachers are incumbent teachers in general subjects; as of April 2004, there were very few specialist teachers of IT. As a result, it is likely that most of the teachers of IT courses lack the specific skills for teaching this topic. Further, it is difficult for them to gain the necessary knowledge and skills, because the educational goals and techniques for IT instruction are not yet clearly defined. For example, most of the teachers who are not specialists mistakenly believe that use of the technology itself is the main goal of IT education, though the ability to use information systems is a more complex and indispensable aspect of IT education.

Many instructors and researchers have published their opinions on various concepts of IT education and the relationships between these concepts [The Meeting of Tuesday, 2001; Ministry of Education, 2000; Ohiwa et al., 2001]. Most of them take into account factors that are useful during the usual instructional design process, such as situations and areas of content, in order to meet the educational goal. But it is also necessary, given that the main goal of IT education is to enhance the meta-ability to make use of information in various situations, to add educational goals that are related to the main goal of an instructional "unit", keeping in mind the content and situations addressed in that particular unit. This problem can be solved by teachers who have technical knowledge as a result of their prior learning and experiences. For teachers who are not specialists in information technology, it is difficult to comprehend this problem. Consequently, a framework that reconstructs these useful resources from various viewpoints and in response to teachers' requests is necessary.

Many organizations provide web pages that show teachers of IT education various useful resources--e.g., digital content, lesson plans, and Q&A [The Meeting of Tuesday, 2003; Okayama Prefectural Information Education Center, 2004; NICER, 2003]. However, it is very difficult to collect the necessary resources for teachers because relevant web pages are too numerous, and their formats and viewpoints are not unified even when the resources have the same purpose.

One cause of these problems is that various concepts of IT education are not yet clearly defined. Because most of the guidelines and commentaries about the subject present the concepts in a disorganized fashion, we believe that these concepts are not conveyed to the teachers effectively. To solve this problem, it is necessary to clarify and to articulate the fundamental concepts of IT education. We consider that ontological engineering can assist in meeting this goal. The ontology provides a common vocabulary and set of concepts about IT education and can promote the reuse and sharing of these concepts among teachers. However, because the ontology is too abstract for teachers to understand, we think that it is not effective to directly provide teachers with the ontology. So, in this study, we use the ontology as a basis and introduce IT education goals, which are more familiar with teachers than ontology, to define other support information resources does not allow authoring of metadata. We propose to make use of the results of this research by identifying relations between this ontology and our ontology. This framework is compliant with the *openness* of the Semantic Web in that it allows alignment of separate ontologies. In this framework, because a system can reconstruct information resources annotated using another ontology, many information resources on IT education can be used more effectively.

The outline of our approach is described in the next section, after which the characteristics and benefits of our ontology are provided in detail. In addition, we show a prototype system based on Semantic Web technology that provides teachers with various information resources.

2 An Outline of Our Approach that is Compliant with the *Openness* of the Semantic Web

In this chapter, we describe the framework for realizing a system that provides teachers of IT education with useful resources in accordance with the various viewpoints that they



Figure 1. The outline of our approach that is compliant with the openness of the Semantic Web

for IT education. If useful web resources for IT education are tagged on the basis of ontology, they can be accessed according to the various viewpoints they represent. This framework is realized based on Semantic Web technology.

There is one previous report [The Meeting of Tuesday, 2001] that classifies the goal of IT education and gives meaning to resources about IT education, though from an ontological engineering viewpoint the classification may need modification and the method of giving meaning to

might have. This framework is an example of the Semantic Web application system that is open to the decentralized world. An outline of this framework is shown in Fig.1.

This framework includes two instances of Semantic Web components: one is based on our ontology, which is described later in detail. We authored metadata of various resources about IT education in RDF using the ontology of IT education as the tag; the other is based on the Goal List of IT education [The Meeting of Tuesday, 2001], which was taken from other research.

The purpose of this Goal List is to provide teachers with viewpoints from which to evaluate the learner's activity during instruction in IT education. Because this Goal List was not generated based on the ontology theory, its quality is not as high as that of an ontology (we explain this problem in detail later.). However, this Goal List already has been so widely used with the same purpose as an ontology that many information resources that support teachers for IT education in Japan are annotated using it. Therefore, in this paper, we regard this Goal List as an ontology.

This Goal List was created by one of the authors of this paper. Although this fact seems to suggest that these two ontologies were built by the same organization and that this framework is closed or centralized, that is not the case; in this paper, we consider the Goal List as a result of other organization's research, because this classification was established before the author began to contribute to the current joint research.

In this study, we realize semantic integration between the metadata based on separate ontologies by describing relations between our ontology and the Goal List clearly. For example, in this framework, the system can reconstruct lesson plans tagged based on the Goal List from the viewpoint of the ontology and provide with them. In addition, the system can integrate lesson plans based on the Goal List with digital contents based on the ontology which are able to be used in each step in them. With this framework, it becomes possible for teachers to use many useful resources on IT education more effectively for a wider range of purposes.

3 An Ontology of the Goal of IT Education

In this chapter, we describe the ontology of the goal of IT education. First, we show a hierarchy in which the classification is based on *is-a* relation, and we show another hierarchy in which it is based on *part-of* relation. Next, we explain why both *is-a* relation and *part-of* relation are necessary, and describe the difference between these two relations. Further, we show the benefits of our ontology in detail by comparing it with other classifications.

3.1 The is-a Hierarchy of the Goal of IT Education

In this section, we show the *is-a* hierarchy of the goal of IT education. A part of the *is-a* hierarchy as the ontology of the goal of IT education is shown in Fig. 2. This ontology was built on the editor "Hozo" [Kozaki *et al.*, 2000], which is an environment for building ontologies.

The ontology of the goal of IT education consists solely of



Figure 2. A part of the ontology of the goal of IT education (*is-a* hierarchy)

concepts of the goal of IT education. Stratification based on *is-a* relation is the essential property of these concepts, and ensures that no confusion of various concepts occurs; such confusion can obstruct teachers' understanding of concepts of IT education. This stratification is one of the characteristics of the ontology, and is one of the important reasons that we applied the ontology theory.

In this study, for this ontology, we extracted three concepts that can be the goal of IT education. These are "Knowledge about information", "Practical ability to act in the information society", and "Independent attitude toward participating in the information society"¹. This classification is compliant with Bloom's taxonomy of instructional objectives [Bloom *et al.*, 1971]. Furthermore, we classified these three concepts.

Bloom's taxonomy of instructional objectives classifies the whole of the goal to be attained in instruction into three domains. The first domain is a cognitive domain, which consists of various goals regarding comprehension of various goals regarding formation of interest, attitude, and value, and is related to enhancement of ability to judge properly and to adapt. The last domain is a psychomotor domain, which consists of various goals regarding acquirement of skills of manipulation and execution. It is clear that the three concepts extracted as goal of IT education correspond to Bloom's taxonomy of instructional objectives as follows: "Knowledge about information" and the cognitive domain, "Independent attitude toward participation in the information society" and the affective domain, and "Practical ability to act in the information society" and the psychomotor domain.

"Practical ability to act in the information society" can be specialized in "Practical ability to utilize information" and "Ability to act based on the information ethics". The former is further specialized in "Meta ability to utilize information", "Ability to process information" and "Ability to utilize an information system". This classification is based on Bloom's



Figure 3. The layered structure of the practical ability to utilize information

knowledge and enhancement of intellectual ability. The second domain is an affective domain, which consists of

taxonomy of instructional objectives as shown in the upper part of Fig.3. The Motor Layer corresponds to the psychomotor domain which is based also on Bloom's taxonomy. In this layer, the ability to manipulate the information system is the concept of goal of IT education. The Intellectual Layer corresponds to the cognitive domain, and the ability to process information is comprised of concepts in this layer.

¹ In Fig.2, a root concept of *is-a* hierarchy (Goal of IT education) and these three concepts are linked by an *is-a* relation. Strictly, these are not *is-a* relations but relation that these concepts can get value of a goal of IT education as a roll concept.

Further, we can classify this concept under an ability to process information by human ("Ability to process information by human") and an ability to process information by computer ("Ability to process information by computer"). The Affective Layer corresponds to the affective domain, which is based on Bloom's taxonomy, and the ability to analyze a project and to practice utilization of information from the meta-level ("Ability to analyze a project" and "Ability to analyze a way to utilize information") are concepts represented in this layer.

As stated above, "Ability to process information" is specialized into "Ability to process information by human" and "Ability to process information by computer". This classification is based on a viewpoint that is whether man is conscious of information processing by computer or not². Furthermore, these two concepts are classified into five concepts as shown in the lower portion of Fig.3. We can specialize ("Ability to design model"), design of structures of data and steps to process information ("Ability to design algorithms"), and description of steps for information processing by using concrete programming languages ("Ability to develop programs"). Note here that the abstraction of information without taking computer processing into account belongs to "Ability to edit information". This distinction is usually difficult to make. In this study, we decide that one aspect of abstraction is whether it's done for processing by computer. Even if a modeling process about computer processing is performed unconsciously, such an ability is classified with "Ability to process information by computer" if it is a procedure, formulation, or theory on computer processing.

"Ability to act based on the information ethics", which is another specialization concept of "Practical ability to act in the information society", is classified as shown in Fig.4. The ability to act based on the information ethics for taking part in





Figure 4. A structure of the ability to act based on the information ethics

"Ability to process information by human" into input or output of information, into or from human as a medium ("Ability to input or output information") and processing and creation of information, which have been input into human as a medium ("Ability to edit information"). "Ability to input or output information" is specialized in extraction and collection of information from the real world ("Ability to collect information") and reporting and sending of information to the real world ("Ability to report information"). "Ability to edit information" is specialized into "Ability to analyze information" and "Ability to create information".

We can divide "Ability to process information by computer" into the following three subabilities: abstraction of information with consciousness of processing by computer the information society can be classified into an ability to act by a subject person so as not to become an assailant ("Ability to act actively based on the information ethics") and an ability to act so as to avoid being a victim ("Ability to act passively based on the information ethics"). Further, the ability to act based on the information ethics can be classified in a two-dimensional space spanned by an axis that represents the owner of information that is a behavioral object into self or others and an axis that represents a target to take into consideration ethically into owner's right or duty. The ability to act based on the information ethics can thus be classified into four areas as shown in Fig.4.

When "Ability to act actively based on the information ethics" and "Ability to act passively based on the information ethics" are interpreted in this point, the former is in areas of the second and the fourth quadrants, the latter in areas of the first and the third quadrants. "Ability to act actively based on the information ethics" is specialized in "Ability to act actively for others' rights" and "Ability to act actively for self duties", and "Ability to act passively based on the information ethics" is specialized in "Ability to act passively for self rights" and "Ability to act passively for others' duties". We

² "A learner processes information by using computer" is not contained in "Ability to process information by human" or "Ability to process information by computer" because the ontology of the goal of IT education discusses only concepts of goal of IT education, not other concepts such as learning activity. However, goal of IT education contained in this activity belongs to "Ability to process by human".

can divide these concepts further into specialized concepts related to intellectual property rights such as information protection, information communication morals, information expression morals, and information reliability.

Finally, we briefly describe the *is-a* hierarchy under the concept of "Knowledge about information". First, "Knowledge about information science" and "Knowledge for taking part in the information society". "Knowledge about information science" is specialized in "Knowledge about an information system", "Knowledge about the expression of information", "Knowledge about the design of information processing", and "Knowledge about the communication of information" from contents of knowledge. "Knowledge for taking part in the information society" is specialized in "Knowledge about the communication of information" from contents of knowledge. "Knowledge for taking part in the information society" and "Knowledge about the information society" and "Knowledg

"Knowledge about an information system" is specialized in "Knowledge about hardware" and "Knowledge about software". "Knowledge about the design of information the information society" is specialized in "Knowledge about the influence of information", "Knowledge about the value of information", and "Knowledge about the value of information". "Knowledge about the information ethics" is specialized in "Knowledge about intellectual property rights", "Knowledge about the protection of information", "Knowledge about morals regarding the expression of information", "Knowledge about the morals regarding the communication of information", and "Knowledge about the reliability of information".

3.2 A *part-of* Hierarchy of the Goal of IT Education

In this study, we also describe a *part-of* hierarchy for the goals of IT education. A part of the *part-of* hierarchy is shown in Fig.5.

A concept that shows the whole of the goal of IT education is "Ability to utilize information" provided by the Ministry of Education in Japan. And the Ministry of Education prepared three viewpoints of this concept. These are "Practical ability of using information", "Scientific understanding of informa-



Figure 5. A part of the part-of hierarchy of the goal of IT education

processing" is specialized in "Knowledge about designing models", "Knowledge about designing algorithms", and "Knowledge about developing programs". "Knowledge about the communication of information" is specialized in "Knowledge about networks" and "Knowledge about the information communication technology". "Knowledge about

tion", and "Awareness toward participation in the information society"; this relation can be interpreted with *part-of* relation. In this study, we classified these three viewpoints in more detail.

The structure of the *part-of* hierarchy is almost the same as that of the *is-a* hierarchy for these concepts. For example, the

lower hierarchy of "Practical ability of using information" is almost the same as the lower hierarchy of "Practical ability to utilize information" in the *is-a* hierarchy, and the lower hierarchy of "Scientific understanding of information" has almost the same structure as the lower hierarchy of "Knowledge about the information science" in the *is-a* hierarchy. This correlative relationship is typical between *part-of* relations and *is-a* relations. Here, we discuss a common problem that is caused by *is-a* relation and *part-of* relation.

It is known that there are some kinds of *part-of* relations that use different semantics [Mizoguchi *et al.*, 1999]. The classification into three viewpoints prepared by the Ministry of Education is a *part-of* relation called *Function-part-of* [Mizoguchi *et al.*, 1999]. This relation is one of *part-of* relations that each makes a "functional" contribution to the whole. In this case, the structure of the *part-of* hierarchy does not become the same as that of the *is-a* hierarchy. However, where concepts such as knowledge and ability, which are included in our *part-of* hierarchy, are concerned, such a problem occurs, because *part-of* relation we need is *Operation-part-of*, which requires careful attention for differentiation between the *is-a* and *part-of* relations.

One of the examples of the *Operation-part-of* is found in the case of plant operation. Consider a normal operation (operation in a normal situation) and a restoration operation (one in a situation where recovery is implemented from above). Both operations are apparently subclasses of the operation class. At the same time, however, if the operation is regarded as a working process, this relation can be interpreted as also a *part-of* relation, given that the whole operation process is composed of a normal operations and restoration operations. The nature of this relation is that a class to which a whole concept belongs is a super class of a class to which all partial concepts belong. When we consider these from a viewpoint of time or space, we can interpret these as *part-of* relations.

For example, when we regard knowledge as a field of study, we find that *is-a* relation is suitable, but when we regard knowledge as what a learner should learn and use, we find that *part-of* relation is more suitable. In other words, it means that in order to learn knowledge it is necessary to learn all more detailed knowledge. In the same manner, when we regard an ability as simply an ability, we find that *is-a* relation is suitable, but when we regard ability as what a learner possesses and performs, we find that *part-of* relation is more suitable.

For example, when we regard "Ability to process information by human" as an ability, each of the concepts ("Ability to collect information", "Ability to analyze information", "Ability to create information", and "Ability to report information"), which are subordinate concepts is the ability to process information by human and we can find that those relations are *is-a* relations. However, when we regard "Ability to process information by human" from the viewpoint of a process of processing, the ability to process information by human is realized by all concepts ("Ability to collect information", "Ability to analyze information", "Ability to create information", and "Ability to report information"), and we can find that those relations are *part-of* relations. In the case of knowledge and capability, as in this example, the structure of the *part-of* hierarchy and the structure of the *is-a* hierarchy are almost the same, though their meanings are quite different. The other classification of the goal of IT education mentioned above does not make a clear distinction between *is-a* relation and *part-of* relation. Such confusion is obstructs the understanding of teachers of the goal of IT education. One of the advantages of our research is that it separates the hierarchy of *is-a* relation and that of *part-of* relation completely.

The necessity of having a *part-of* hierarchy that has almost the same structure as an *is-a* hierarchy occurs when an instance of a class of a middle concept³ is made. When an instance of a class of a middle concept in an *is-a* hierarchy is made, we can obtain an instance of a goal of IT education that belongs to one of the sub-classes of its middle concepts. In other words, we cannot obtain an instance of a goal of IT education from each of those sub-classes that belongs to the class of the middle concept in an *is-a* hierarchy. In a *part-of* hierarchy, on the other hand, we can obtain an instance of a goal of IT education from each of those sub-classes that belongs to the class of the middle concepts. Therefore, we described not only an *is-a* hierarchy but also a *part-of* hierarchy of goal concepts in our research.

In the target world in which the structure of an *is-a* hierarchy is almost the same as that of a *part-of* hierarchy, the confusion of *is-a* relation with *part-of* relation can obstruct a user's understanding of the distinction. In the usual classification of the goal of IT education, *is-a* relation and *part-of* relation were also confused. This is also one of potential obstructions to teachers' understanding of IT education. Thus, it is necessary to distinguish hierarchies based on *is-a* relation and *part-of* relation clearly, we have done here. The concrete obstruction of the confusion of *is-a* relation and *part-of* relation is described in detail in 3.3.2.

3.3 A Benefit of the Ontology of the Goal of IT Education

In this section, we show the advantages of the ontology of the goal of IT education over other classifications from two viewpoints. First, we describe how confusion among goal concepts is obstructive, referring to the Goal List [The Meeting of Tuesday, 2001] introduced in Section 2. Next, we explain another ontology, which a student in our laboratory built, then we describe the obstructive confusion of *is-a* relation and *part-of* relation by comparison of our ontology with this ontology.

There is No Confusion with Other Concepts

In this paragraph, as an example of the current standard classification of goal of IT education, we take up the Goal List [The Meeting of Tuesday, 2001] which is well known in

³ Here, "middle concept" means a node which is not a leaf node in the tree structure. In other words, it is a node which has more than one child node.

Japan. Though this Goal List was not generated based on the ontology theory, this Goal List already has been so widely used with the same purpose as an ontology in Japan. So, it is necessary for this Goal List to have the same natures as the ontology. For this reason, we compare this Goal List and our ontology from the viewpoint of the ontology theory. The classification of the Goal List of IT education is shown in Fig.6.

The Goal List has three top-level categories, "Practical Ability of using information", "Scientific understanding of information", and "Awareness toward participating in the information society", which the Ministry of Education pre-

A : Practical Ability of using information

- (1) Expression and communicating information
 - a: Expression of information
 - b: Communication by a media
- (2) Using information in problem solving
 - c: Discovery and planning of a problem
 - d: Collection of information
 - e: Classification, analysis and judgment about information
 - f: Reporting and sending of information
- (3) Selecting the means of information (information media, computer, network)
- g: Selecting the means of information
- B: Scientific understanding of information
 - (4) Understanding various means of information and its advantages and disadvantages
 (5) Basic theories and methods of information processing, information technology,
 - and human's information recognition
- C : Awareness toward participating in the information society
 - (6) Awareness toward information
 - m: Awareness toward information
 - (7) Information ethics and responsibility in sending out information
 - n: Information ethics

Figure 6. The classification of the Goal List of IT education

by superficial differences among learning activities.

pares in more detail in the same way as our *part-of* hierarchy. For this purpose, examples of more concrete learning activities that are easy for teachers to understand are provided with a level that shows when learners should attain this goal, though these levels are not shown in this figure. We think that it is more suitable for teachers' understanding to provide them with information on activities related to concepts of learning. We think that it is easier for teachers to grasp each description when concepts of learning activities are included in the information provided. Further, it is difficult to set a level of difficulty for a goal of IT education without presenting concepts of learning activities. Consequently, the Goal List has many advantages as information that is provided to teachers directly.

However, the Goal List has some faults from the viewpoint of classification of the goal of IT education. Although, essentially, the classification of educational goals should be performed by extracting the intrinsic goals that should be attained in education and systematizing them, in many of the current classifications, we find concepts other than goals; for example, learning activity and learning environment related to goals are incorrectly mixed up. For example, in the case of the Goal List, the concept of the goal "Selecting the means of information" contains not only the goal of IT education but also that of a learning environment in which learning occurs. Given these considerations, to the best of our knowledge, there is no goal classification that properly captures the intrinsic educational goals of IT education without any confusion regarding learning activity, standard of evaluation for education, etc. It is difficult to separate various concepts related to IT education, because most goals of IT education are meta-abilities that are attained in the process of problem solving. Considering the fact that the purpose of the classification of the goal of IT education is to give teachers a clear understanding of the educational goals, our goal ontology is more suitable, based on the fact that it reveals the inherent conceptual structure of educational goals and thereby facilitates a teacher's understanding of those goals.

No Confusion between is-a relation and part-of relation

An ontology of the goal of IT education that was built by a student in our laboratory whose understanding about the ontology theory was insufficient is shown in Fig.7. In this paragraph, we show obstruction caused by confusion between the *is-a* and *part-of* relations through the comparison of this ontology and our ontology.

The student described all relations as *is-a* relations without considering the meaning of classification. As a result, the viewpoint of the classification is not unified, and a distinct *part-of* relation exists. The first obstruction related to this

This is influenced by the purpose of the Goal List, which is to provide teachers with viewpoints of evaluation of learners of IT education. Moreover, systematization like that in this example, in

which other concepts are mixed, sometimes causes another problem: the extracted concepts are not completely independent of each other. For example, in the case of the Goal List, both "Expression of information" and "Reporting and sending of information" are subordinate concepts of "Practical ability of using information" and have the same goal as "Ability to report information". When teachers use such a classification, their awareness to this goal may be obstructed confusion is that the inheritance of attributes, which is one of the biggest advantages of the *is-a* hierarchy, is not realized in this hierarchy. Additionally, it can confuse teachers by classifying a concept based on an *is-a* relation in spite of having classified it originally based on *part-of* relation.

For example, the topmost classification in Fig.6 is originally based on *part-of* relation, but it is classified based on *is-a* relation. This can promote misunderstanding by teachers that the educational goal of three subordinate concepts can be contradiction as shown by "1" in Fig.6. This concept is inappropriate because both concepts of knowledge and ability of utilization are confused in it.

These examples illustrate how the mixture of *is-a* and *part-of* relations can confuse both users and authors. Therefore, our ontology of the goal of IT education, which incorporates the distinction between *is-a* relation and *part-of* relation and the exclusion of other concepts, is meaningful.



Figure 7. An ontology of the goal of IT education built by a student in our laboratory whose understanding of the ontology theory was insufficient

attained independently to attain the educational goal of the super ordinate concept, when, because this hierarchy is part-of relation (function-part-of), the goal of the superordinate concept cannot be actually attained unless all goals of subordinate concepts are attained with each role in the whole. This can also obstruct teachers' understanding of each specific goal of IT education. For example, in the part-of hierarchy, teachers can easily understand that "Ability to analyze a project" is an ability to analyze a project that can be solved by processing of information, which is done later in the whole process for solving it, because the definition of *part-of* relation shows that it is a part of the whole process of processing information directly. However, in an is-a hierarchy, teachers can easily misunderstand that this concept is an ability to consider a project with no relation to processing of information because subordinate concepts are independent in this hierarchy.

Further, the author of this ontology created concepts such as "Knowledge to make use of information" to get rid of these

4 Prototype based on Semantic Integration

In this chapter, we describe a prototype system for supporting teachers based on the above framework. Resources used by this system are simple lesson plans on the Web (called Digital Recipes) provided by Okayama Prefecture Information Education Center. These Digital Recipes are open to the public as resources related to concepts of the Goal List. However, they were not described as metadata; we authored the metadata of these resources from the viewpoint of the Goal List.

The layered structure of the prototype system we built is shown in Fig.8. This system is constructed in four layers. The bottom layer is the ontology layer. In this layer, we define all of the concepts related to the above the ontology of the goal of IT education and the Goal List of IT education. The second layer is the RDF-Schema Layer. In this layer, the vocabularies of classes and properties used in the third layer, the RDF Model Layer, are defined. There are four schemata in this layer. As two schemata in them, the vocabularies of classes and properties related to the ontology and the Goal List defined in the bottom layer are defined. The third schema defines the vocabularies related to the fundamental academic ability of IT education. The purpose of describing this ability is to show more clearly the essence of the concepts of the goal of IT education by identifying its differences from the academic ability, which is attained in other subjects. For example, there is an ability to express something, which is the fundamental academic ability of the a resource that shows the relations between the Goal List, the ontology of the goal of IT education and the fundamental academic ability of IT education. The Goal List provides teachers with some examples of more concrete learning activities with a level that shows when learners should attain each goal classified. For this resource, we authored metadata of these learning activities, which belong to the respective concepts of the Goal List, by using the vocabularies defined in the RDF-Schema Layer related to the ontology of the goal of IT education and the fundamental academic ability of IT education. The other resource is the description of the metadata of the Digital Recipes. We described the same contents as the resources that are open to the public in RDF.



Figure 8. The layered structure of the prototype system based on semantic integration

ability to report information. The difference between them is "Use of various ways to express information". We do not describe this fundamental academic ability in this paper in detail. The fourth schema defines the vocabularies of classes and properties for description of resources of the Digital Recipes this prototype system processes.

The third layer is the RDF Model Layer. In this layer, we can author metadata of various resources by using the vocabularies defined in the RDF Schema Layer. For this prototype system, we authored metadata of two resources. One is Thus, this metadata is described based on the Goal List.

The topmost layer is the Web Layer. In this layer, the system analyzes the metadata described in RDF and provides teachers with web pages that are reconstructed as HTML files. For example, the screen shot on the right in Fig.8 shows a web page that the system analyzes the metadata of a Digital Recipe, which a teacher requests, and provides him/her with it. These web pages are almost the same as the contents of the resources provided by Okayama Prefecture Information Education Center.

The prototype system explained in detail in this section is a system that converts the contents of the screen shot on the right in Fig.8 to the contents of the screen shot on the left. The system analyzes the metadata of a Digital Recipe and extracts concepts of the Goal List tagged in this resource, then the system extracts the concepts of the ontology of the goal of IT education related to those concepts of the Goal List from the other resource (Description of the relations between the ontology and the Goal List) in the RDF Model Layer. The system integrates the original resources with extracted concepts of the ontology and outputs it as an HTML file. The example of integrated information that is output by the system is the screen shot on the left in Fig.8.

This prototype system can provide teachers with the integrated benefits of both ontologies. In this example, for each step in a flow of the instruction, the viewpoints of evaluation are provided with expression that is easy for teachers to understand these meaning as the benefit of the Goal List, and the goals of IT education, which is easy to be hidden in the shadow of it as the benefit of the ontology, are provided.

And, in this framework, the system can extract Digital Recipes which contains a particular goal of IT education by using the relations between the Goal List and our ontology. Furthermore, the system can also extract Digital Recipes that teachers will be able to develop easily into the instruction which contains a particular goal of IT education by using the description of the relations between our ontology and the fundamental academic ability. We have already finished implementation of these functions.

This example of processing by the prototype system is based on a very simple mechanism, though this system realizes the fundamental function of semantic integration based on two ontologies. In this framework, the system will be able to integrate more complicatedly resources based on two ontologies according to the features and their manner of use.

5 Summary

In this paper, we described the ontology of the goal of IT education in detail. And, we proposed the framework to make use of the results of another ontology by alignment of these ontologies based on Semantic Web technology.

The ontology of the goal of IT education consists solely of concepts of the goal of IT education as the *is-a* hierarchy without other concepts. And, we considered the difference between *is-a* relation and *part-of* relation and classified separately the *part-of* hierarchy. Then, we showed the advantages of our ontology over other classification from the above two viewpoints.

We described the prototype system which realizes semantic integration by the alignment between our ontology and the Goal List.

In future work, we intend to build a more effective system based on this framework according to these features and manner of use. And, we intend to build another system to support dynamically to teachers when they design instruction. For realization of this system, we think to use not only the framework shown in Fig.8 but also the several kinds of "part-of relations" between goals of IT education.

Acknowledgments

This work is supported in part by Grant-in-Aid for Scientific Research (A) No. 14208029 and Grant-in-Aid for Young Scientists (B) No. 15700133 from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

- [The Meeting of Tuesday, 2001] The Meeting of Tuesday, The Goal List of Information Education, Mail-Magazine of the Meeting of Tuesday, http://kayoo.org/home /project/list.html, 2001.
- [Ministry of Education, 2000] Ministry of Education, Commentary of the Course of Study for the Subject "Information" in the High School, Kairyudo Publishing, 2000.
- [The Meeting of Tuesday, 2003] The Meeting of Tuesday, Digital Contents related to the Information Systems and the Structure of the Information Society, http://kayoo.org/home/mext/joho-kiki/, 2003.
- [Okayama Prefectural Information Education Center, 2004] Okayama Prefectural Information Education Center, Okayama Prefectural Information Education Center HomePage, http://www.jyose.pref.okayama.jp, 2004.
- [NICER, 2003] NICER, National Information Center for Educational Resources HomePage, http://www.nicer.go.jp, 2003.
- [Kozaki *et al.*, 2000] Kouji Kozaki, Yoshinobu Kitamura, Mitsuru Ikeda, Riichiro Mizoguchi, Development of an Environment for Building Ontologies which is based on a Fundamental Consideration of "Relationship" and "Role", The Sixth Pacific Knowledge Acquisition Workshop (PKAW2000), pp.205-221, 2000.
- [Bloom et al., 1971] Benjamin Samuel Bloom, J Thomas Hastings, George F Madaus, Handbook on formative and summative evaluation of student learning, McGraw-Hill, 1971.
- [Ohiwa *et al.*, 2001] Hajime Ohiwa, Takahiro Tachibana, Tooru Handa, Yasushi Kuno, Takeo Tatsumi, How to Instruct in the Information Course, Ohmsha, 2001.
- [Mizoguchi et al., 1999] Riichiro Mizoguchi, Mitsuru Ikeda, Yoshinobu Kitamura, Foundation of Ontological Engineering -An Ontological Theory of Semantic Links, Classes, Relations and Roles-, J. of Jpn. Soc. for Artificial Intelligence, Vol.14, pp.1019-1032, 1999.
- [Okayama Prefectural Information Center, 2002] Okayama Prefectural Information Education Center, Collection of Web Recipe and Web Worksheet, http://www2.jyose.pref.okayama.jp/cec/webresipi/resipi/ resipi.htm, 2002.
From Annotated Learner Corpora to Error Ontology: A Knowledge-based Approach to Foreign Language Pedagogy

Md Maruf Hasan¹, Kazuhiro Takeuchi, Virach Sornlertlamvanich, Hitoshi Isahara

Thai Computational Linguistics Laboratory (TCL)²

National Institute of Information and Communications Technology (NICT), Japan NECTEC #224-5, 112 Pahonyothin Road, Klong Luang, Pathumthani 12120, Thailand mmhasan@acm.org, kazuh@nict.go.jp, virach@tcllab.org, isahara@nict.go.jp

Abstract

The mainstream corpus-based linguistics research focused on collecting and annotating well-formed language usage (i.e., correct instances). In recent years, Annotated Learner Corpora, with ill-formed instances and respective error annotations, are growing rapidly. In general, Annotated Learner Corpora consist of tuples with two essential attributes and one or more annotation attributes: <I: incorrect instance, C: respective corrected instance, A: annotations>. Initiatives in developing Annotated Learner Corpus are novel attempts to capture and codify error knowledge in machine understandable forms - often using markup languages such as, XML. However, given the fact that linguistic errors are complex and multifaceted phenomena: spoken vs. written language errors, L2/L1 language acquisition errors (related to different native language and age groups); competence/tacit vs. performance/explicit errors, etc., a generalized and flexible representational framework is desirable. We used Semantic Web inspired tools and technologies, and ontology-based modeling and representation of errors to facilitate integration, sharing and reuse of heterogeneous annotated learner corpora and respective error knowledge. We also analyze how such Error-ontology Driven Knowledgebase (EKB) can be used in developing sophisticated applications for foreign language teaching and learning.

1 Introduction

Although the mainstream corpus-based linguistics research focused on collecting and annotating well-formed language usage (i.e., correct instances), *Annotated Learner Corpora*, with *ill-formed* instances and respective error annotations, are growing rapidly in recent years. In general, Annotated Learner Corpora consist of tuples with two essential attributes and one or more annotation attributes: <I: *incorrect instance,* **C:** *respective corrected instance,* **A**_{*i*}*: annotations*>.

For a particular error, the first two attributes, *incorrect instance* (I) and *respective corrected instance* (C), identified by different annotators may vary (e.g., *Noun-Verb Agreement Error*: either N or V can be corrected to achieve conformity). For the optional annotation attributes (A_i s), with the *What-scheme*, the annotations are usually done by tagging the error with an *error-type* from a predefined error-hierarchy. With a *Why-scheme* such annotations are *narra-tive* - based on natural language descriptions of the error. Each scheme has its obvious pros and cons in terms of machine and human readability. Both *What-* and *Why-* annotation schemes are being widely used. Nonetheless, both schemes can also be adopted simultaneously.

Sharing and reusing learner corpora annotated with different set of error-tags (*What-scheme*) are difficult. Moreover, with narrative annotation (*Why*-scheme), the conversion of human-understandable descriptions in natural language into machine-readable forms remains to be a major challenge.

Annotated Learner Corpora are novel attempts to capture and codify error-knowledge. However, given the fact that linguistic errors are complex and multifaceted phenomenon: spoken vs. written language errors; L2/L1 language acquisition errors (by different native-language and age groups); competence/tacit vs. performance/explicit errors, etc., a generalized and flexible representational framework is desirable. If carefully adopted, Semantic Web inspired tools and technologies such as, ontology-based modeling and representation of linguistic errors can facilitate integration, sharing and reuse of heterogeneous annotated learner corpora and respective error knowledge efficiently.

In this paper, we investigate the initial development of an *Error Ontology* from a Learner Corpus (namely, the **NICT JLE Corpus** which consists of Japanese Learners' Spoken English with errors annotated). We will explain how such an

¹ Currently a *Lecturer* at Shinawatra University, Thailand ² The TCL Lab is one of the NICT's overseas laboratories located at Thailand Science Park.

Error-driven Knowledge-base (**EKB**) can be used to develop (1) *Smart Pedagogical Dictionaries* which includes error instances as well as respective corrections, annotations and associated error knowledge through logical inferences, (2) *Robust Natural Language Processing Tools* (e.g., Morphological Analyzers and Parsers) that are capable of detecting linguistic errors, and (3) *Higher-Order Error Typology* which may capture higher level error phenomena (e.g., Discourse, Pragmatic and Interlanguage errors [Selinker, 1972].

We also argue that by integrating the EKB with Learning Objects [Wiley 2002], we can potentially obtain an ideal *Personalized Language Learning Environment*. Similar to *Data-Driven Learning* (**DDL**) of foreign language [Hadley 1997], which is based on *positive instances* (well-formed corpora) in target language, the EKB may help foreign language (FL) teachers and learners to customize text-materials and to develop personalized lesson-plans based on errors (*negative instances*). Such an Error-Driven Learning (**EDL**) may be more effective in pedagogical contexts.

2 Corpus Linguistics and Learner Corpora

Corpus Linguistics is not a particular linguistic paradigm but a methodology as defined by Leech: a methodological basis for pursuing linguistic research [Leech, 1992]. Such a claim can be validated by surveying the various types of corpora available and the types of analyses (research) conducted on those corpora. Corpora (collections of externalized utterances/instances of tacit linguistic competence) and computing technologies have been (arguably!) playing an important role in analyzing human language, behavior and cognition.

The first computer corpus ever created is the Brown Corpus is a balanced corpus – one million words of edited written American English taken from 2,000-words samples representing different genres of written English (and made available in electronic format through ICAME; ICAME also distributes a variety of other corpora [ICAME]. Such balanced corpora are valuable resources for Linguists who want to use corpora primarily for the purpose of linguistic description and analysis on different genres. On the other hand, for the Penn Treebank Corpus [Treebank], the focus is not on balancing the collection of text but on the size (4.9 million words) so that statistically significant parameters can be computed. Computationally oriented Natural Language Processing (NLP) researchers have done significant work in developing morphological analyzers (part-of-speech taggers), syntactical analyzers (parsers) using the large collection of well-formed (correct) instances of English using machine learning and other empirical techniques.

The *CHILDES Corpus* contains transcriptions of children speaking in various communicative situations. Such a corpus has been widely used by psycholinguists interested in child language acquisition [MacWhinney, 1996]. The proliferation of Learner Corpora is partly due to the efforts in developing effective text-materials to teach foreign languages and standardized test criteria to evaluate language proficiency of second language (L2) learners. *The Cambridge Learner Corpus* [CLC] consists of more than 10million words of student essay exams collected and annotated from over 35,000 students from 150 countries with 75 different first languages [L1]. Such a corpus and its subsets are valuable resources for research in foreign language teaching and learning. The *NICT JLE Corpus* (formerly known as Standard Speaking Test Corpus or SST Corpus) is a joint effort between *ALC Press* and the *National Institute of Information and Communications Technology* (NICT), Japan [Tono *et al.*, 2001; Izumi *et al.*, 2003]. This corpus consists of 1-million words of Japanese Learners' (spoken) English.

The corpora cited above can roughly be categorized into two types – corpora with well-formed (correct) instances and corpora with ill-formed (incorrect) instances. It is obvious that incorrect instances of language serve as a good source of knowledge in language pedagogy, as well as, first and second language acquisition (FLA/SLA) research [Corder, 1983]. However, incorrect instances of language are difficult to gather and annotate to build a corpus of reasonable size and quality [Becker *et al.*, 2003]. Notwithstanding though, *Learner Corpora* for English as well as those for other languages, such as Japanese Learner Corpus [Nagoyadai] are steadily becoming available.

Annotated learner corpora attempts to capture (or codify) significant amount of error knowledge in explicit and implicit forms based on the annotation schemes. Learner corpora may be annotated with either or both *What* and *Why* attributes, as well as other annotation attributes (A_i s) depending on target applications. However, reuse and sharing of annotated learner corpora become difficult when heterogeneous error typologies and annotation schemes are used in annotating such corpora. In order to capture and codify error knowledge from heterogeneous learner corpora in an integrated manner, an ontology-based modeling and representation of linguistic error knowledge remains to be ideal as we have justified through the rest of this paper.

2.1 The NICT JLE Corpus and Applications of Learner Corpora

Unlike the *Cambridge Learner Corpus* which consists of written essays, the *NICT JLE Corpus* consists of spoken language instances (transcriptions of oral proficiency interviews), and it is in public domain. One of the unique features of *NICT JLE Corpus* is that each data-set also includes the interviewee's (anonymous) proficiency profile and other information based on the ACTFL-ALC **Oral Proficiency Interview** criteria (aka, OPI or SST criteria).



Figure 1: Concordancer output showing article-related errors from NICT JLE Corpus Source: NICT Technical Report

The NICT-JLE Corpus consists of 300 hours of interview data (each interview lasts about 15 minutes), which is about 1-million words. Each data-set is annotated with one of the nine proficiency levels based on SST evaluation criteria. The annotation is XML based with a predefined set of errortags (error-typology). It should be noted that certain error knowledge such as, those related to pronunciations and fluencies are not available in the NICT JLE Corpus due to the simple XML-based annotation scheme used with a predefined set of error-tags inherently biased on lexical and syntactic errors. For a thorough analysis of linguistics error phenomena, it is desirable that error features other than the lexical and syntactic errors are also annotated with equal importance. In ontology-based approach, due to the expressive power of ontology, it is possible to capture such error knowledge into an Error Ontology using appropriate errorobjects and attributes. In an ideal scenario, we may also be able to integrate errors spoken language errors with those of written language. For a complete description of NICT JLE error-tags and tagging guidelines, please refer to the SSTrelated Web site maintained by Yukio Tono [SST]. Our ontology-based representation is discussed briefly in Section 3. It should be noted that a Complete Error Typology essentially includes subclasses for orthographic, morphological, morpho-syntactic, syntactic, semantic, pragmatic and other errors [Becker *et al.*, 2003; Crysmann, 1997]. The NICT JLE error-typology, however, is a subset thereof (representing only spoken language errors for Japanese learners' spoken English with syntactic and lexical errors annotated). Our target *Error Ontology* aims at capturing the *Complete Error Typology* of linguistic errors in and across languages.

Like any other Learner Corpus, the *NICT JLE Corpus* also includes tools such as, a *Tag Editor* to facilitate error annotation, and a *Concordancer* to facilitate error analysis. Figure 1 shows an example of errors related to the use of article/determiners using Concordancer. It should be noted that Japanese language has no article-system and therefore, Japanese speakers often tend to misuse articles in English.

Similar analysis tools are also used for other Learner Corpora. For example, the *Cambridge Learner Corpora* includes Concordancer and a set of statistical analysis tools as shown in Figure 2.

It should be noted that such annotation and analysis tools are not interoperable. By converting Learner Corpora annotations into an Error Ontology, it is possible to develop interoperable tools as pointed out in the subsequent Sections.

younger generation can <#U><	#MT> for #MT	longer s	share the wisdom of	the older	genera	tions share th	
iteressted interested #S <	#MT> <u>in</u> #MT <	#FV>to a	attend attending #</td <td>FV> one of</td> <td>your s</td> <td>eminars about</td>	FV> one of	your s	eminars about	
ir students to participate <	:#MT> in #NT [Statistic	5	1			
<#W>Only <	:#MT> on #MT 🛅					th	
And <	:#MT> on #NT	View sta	tistics by: Word	- Positio	n: 0	Close	
some of them are in danger <	#MT> of #NT	TICH Ltd	word	Loudo		le	
<#BT> On #NT Print							
the phone, here is a list <	:#MT> of #NT		Value	# Cites	2 -		
<	#MT> The In th	1	to	197	19.7	Ver Charley	
school wants to provide us <	#HT> with #HT</td <td>2</td> <td>for</td> <td>147</td> <td>14.7</td> <td>View Littes</td>	2	for	147	14.7	View Littes	
Please write <	:#MT> to #MT	3	of	135	13.5	······	
1 in order to let you know <	#HT> about #H</td <td>4</td> <td>in</td> <td>133</td> <td>13.3</td> <td>User Tag Ich</td>	4	in	133	13.3	User Tag Ich	
<	#MT> First At	5	on	68	6.8		
regards #S occurs #RV <	#MT> to #MT	6	about	62	6.2	I	
/orry #TV , even thinking <	#HT> about #H</td <td></td> <td>with</td> <td>52</td> <td>- <u>5.2</u></td> <td>.5</td>		with	52	- <u>5.2</u>	.5	
10 hours later, she learnt <	#MT> of #MT	9	at ha	32	3.2	Helo	
a were going to provide us <	#HT> with #HT</td <td>10</td> <td>from</td> <td>28</td> <td>2.8 -</td> <td><u>щеф</u> #Т</td>	10	from	28	2.8 -	<u>щеф</u> #Т	
prry about #RJ this, but <	#MT> over #NT</td <td></td> <td></td> <td></td> <td></td> <td>en</td>					en	
it twice, 'Let's find out <	#HT> about #H</td <td></td> <td></td> <td></td> <td></td> <td>ıd</td>					ıd	
I am writing <	#MT> to #NT y	ou to an	nounce <#UA>you #</td <td>(UA> that<#</td> <td>¢UP>.<!--#</td--><td>UP></td></td>	(UA> that<#	¢UP>. #</td <td>UP></td>	UP>	
theatre's restaurant, and <	<pre> to<!--#NT--> my surprise <#U>was when<!--#U--> I <#RV>realized discovered<!--/pre--></pre>						
'#RT> what they have asked <	1 <#NT> for #NT through their protest.						
(RT> main European cities, <	ties, <#MT> for #MT example<#MP> , #NP London, Paris, Amsterdam, etc.						
I vould like to suggest </td <td>NT> to<!--#NT--> yo</td> <td>ou that</td> <td>it <#TV>would be i</td> <td>s<!--∦TV--> ex</td> <td>tremely</td> <td>important to</td>	NT> to #NT yo	ou that	it <#TV>would be i	s ∦TV ex	tremely	important to	
In contrast <#NT> to #NT these advantages <#R>the canteen they offer does not seem							
1 can <#RV>see watch #RV <	<pre><#NT> for<!--#NT--> one or two hours a day<#MP> ,<!--#NP--> but the more you <#RV</pre>						
(FV>to ask asking #FV me <#HT> about #HT my <#FN>taste tastes #FN preferences #FN and <							
include are included #TV <#MT> in #NT the fee exactly<#RP>. ? #RP							
Scerned, they didn't focus <#NT> on #NT these subjects in an interesting way.							
(ND> intimate relationship <#NT> with #NT because the gap between the Honeychurches and the Emerso							
, #MP: but she explained <</td <td>#NT> to<!--#NT--> as</td> <td>e <∄HC> </td> <td>that<!--#NC--> she was</td> <td>in Londor</td> <td>for he</td> <td>r job because</td>	#NT> to #NT as	e <∄HC>	that #NC she was	in Londor	for he	r job because	
ok Fahrenheit 451 relates <	#MT> to #MT .						
(RP> they also have tennis <#MT> as #MT <#MD> a #MD <#DJ>sport sports #DJ <#AGN>activities ac							
Once we got <#HT> to #NT the hotel I wasn't surprised about the service.							

Figure 2: Concordancer output showing errors related to *prepositions*. The small window further shows the statistics for each preposition sorted by ascending order. *Source: CLC Web Site*

In the original *NICT JLE Corpus*, (in most cases) both the interviewers and interviewees are native Japanese speakers. The developer of *NICT JLE Corpus* also intends to add native speakers' version of the interviews, as well as, the back-translated version. Back-translation is a well-known method used to evaluate machine translation systems. Features extracted from the back-translated version and respective native speaker's version may be useful to describe *interlanguage* of particular L2 groups using corpus-linguistics methodologies. The readers should note that such additional features can also be easily represented in the error ontology in terms of additional slots.

Several applications of *NICT JLE Corpus* have already been reported. Such applications include (1) the analysis of a particular type of errors - for example, the article usage pattern in Novice, Intermediate and Advanced level Japanese speakers of English, and (2) automatic level checking – the assignment of a proficiency level by analyzing features extracted from the interview data using machine learning algorithms, and so on as reported in [Izumi *et al.*, 2003]. The first application is based on quantitative analysis of error types and directly relevant to errors and foreign language pedagogy. The second application, however, heavily relies on features other than errors (e.g., vocabulary usage, fluency, etc.). In fact, almost all the 17 features used with machine learning algorithms are not directly relevant to errors. The result of automatic level checking obtained in this way maybe justified for the oral proficiency test however, it may not be appropriate for written language or for a comprehensive proficiency test. As we pointed out earlier, linguistic errors are complex and multi-faceted phenomena and results based on simplified assumptions may not be appropriate in the context of language acquisition and FL pedagogy.

3 The Error Ontology

Annotation of Learner Corpora with respective error-related information is a novel attempt to capture and codify error knowledge. However, as explained above, the information included in the data-set varies with the information annotated explicitly in the learner corpora.



Figure 3: A screenshot of the Error Ontology showing some of the objects and attributes in the Error Ontology

Linguistic errors are not as straightforward as one may apparently think. As explained earlier, for a particular error, two annotators may suggest 2 different $\langle I, C \rangle$ tuples. Even when the $\langle I, C \rangle$ attributes are unique, the corresponding A_{iS} may not be unique depending on the error-typology (*Whatscheme*) and depending on the annotators' styles (*Whyscheme*), etc. Moreover, for a particular language, errors found in spoken languages differ from those in the written forms. Needless to say that each language has some unique errors which are not usually occur in other languages (e.g., errors inherent to Japanese in terms of *polite expressions* and *case markers* etc.). It is also well observed fact that certain error phenomena in L2 acquisition are more common

for particular L1 speakers (e.g., *article*-related errors for Japanese speakers learning English).

The *NICT JLE Corpus* attempts to codify many useful attributes for each interviewee, which include the interviewee's age, sex, written language proficiency (ToEFL and ToEIC scores, if available), duration of overseas-stay, etc. Although some spoken language features (e.g., Fillers) are explicitly annotated in the NICT JLE Corpus, certain other features (e.g., the *number* of vocabulary uttered during the interview and the *level of sophistication* of those vocabulary according to SST criteria; the rate of utterance and fluency, etc.). This is probably due to the limitations of flat XMLbased representational scheme used in corpus annotation. It should be noted that in some cases, these implicit attributes can be calculated on-the-fly to develop applications. However, it is desirable that these attributes are explicitly included with the preferred representational framework.

With the earlier explanation of various linguistic error phenomena and error-related attributes as well as different Learner Corpora initiatives, it is desirable that we investigate linguistic error phenomena in an integrated fashion. Historically, the methodologies in corpus linguistics have been highly influenced by the development of new technologies and algorithms in computing discipline. The Semantic Web tools and technologies which attempt to facilitate sharing, reuse and manipulations of heterogeneous (unstructured) information into a sharable and reusable repository [Corcho et al., 2003] have their own potentials in modeling and representing linguistic errors phenomena in a uniform (machine-understandable) manner. By adopting an ontology-based modeling and representation we can make use of sophisticated tools and technologies developed by the Semantic Web community world-wide over the years.

In order to do so, we first identified the objects (classes and subclasses) and attributes (slots) and their relationships prevailing in the NICT JLE Corpus (and data-sets). The NICT-JLE annotation scheme uses a *What-scheme* to annotate errors with error-types from a predefined hierarchical error-typology. It was fairly straightforward to convert the error-type hierarchy into an ontological class hierarchy. We also model the learners, annotators, proficiency levels as classes with respective subclasses and attributes in the error ontology. Figure 3 shows a screenshot of the Error Ontology developed using Protégé [Protégé; Knublauch, 2003].

We use semi-automatic approach to transfer XML-based error annotations of the *NICT JLE Corpus*, other learner corpora as well as expert (language teachers) knowledge to populate the Error Ontology. Error-instances are directly related to respective error-objects as well as other objects (such as, SST proficiency level) through inheritance relationships. Protégé's multiple-inheritance and inference capability [Duineveld *et al.*, 1999] also allow us to relate spoken error instances with written language proficiency, and if available, with written language error-instances - for example through SST Proficiency Level. Protégé also provides an easy-to use GUI which is suitable to be used by annotators and FL teachers and learners to add or update data in the Error Ontology. We plan to develop a Web-based collaborative interface for FL teachers and learners.

Our Error Ontology is still in prototype stage, and we are refining the ontology to make it capable of capturing error knowledge from heterogeneous Learner Corpora – error instances from written and spoken languages of L1 and L2 learners of various age and language groups. Upon completion, the Error Ontology will be the basis of the Error Knowledge Base (EKB) for research in FLA/SLA and FL pedagogy research. Further details about the Error Ontology will be made available on the web [EO-EKB] in HTML, RDF and other formats.

With the current Error Ontology, it is possible to make simple queries such as, categorizing errors by SST Proficiency Levels. However, that is not the ultimate goal of this research. Our focus is to build a generalized Error Ontology and Error-driven Knowledge-base which will eventually become the basis of several sophisticated applications (cf. Section 4 and Appendix A). We will subsequently develop intelligent applications to demonstrate the usefulness of such an EKB.

4 Potential Applications of Error-driven Knowledge-Base (EKB)

In this section, we will outline some of the potential applications which can be built with the help of the error-driven Knowledge-base (EKB).

Smart Pedagogical Dictionary: As shown in Figure 1 and 2, annotated learner corpora and accompanied tools are useful in locating error instances, respective corrections and other vital statistics about errors. Such tools have been helping language researchers, FL teachers and learners in various ways. However, with the EKB it will be possible to build even smarter Pedagogical Dictionaries. For instance, such dictionaries may make use of sophisticated algorithms such as, Collaborative Filtering, to explore error phenomena in more details because our EKB includes not only error-instances but also learner's profiles and other attributes that facilitate logical inferences in the error-space. FL teachers and learners will be benefited by using Smart Dictionaries since this is not only based on direct pattern-matching and frequency-counts, but also based on intelligent inferences.

Robust NLP Tools: The most widely used NLP tools such as, Part-of-Speech Taggers and Parsers are built with features and parameters extracted from general corpora with positive instances. Those parameters do not directly reflect error-related knowledge. With any POS Tagger or Parser, one may successfully parse an erroneous sentence such as, "Does you play football?", and notice no indications of errors in the output. Robust NLP tools need to incorporate features extracted from incorrect instances to highlight linguistic errors. In order to do so, robust NLP applications need to make use of error-related features which is readily available in the EKB. For instance, the existence of bigrams features, such as "does you" in the EKB is potential evidence (negative feature) for a Robust Tagger or a Parser.

Higher Order Error-Typology: Most annotated learner corpora capture and include adequate annotations for lexical and syntactic errors. However, due to the limitation of representational power in the annotation method, it becomes difficult for those learner corpora to capture and annotate higher-level errors such as, semantic, pragmatic and interlanguage error phenomena. In an ontology-based representation, it is comparatively easier to capture such error knowledge and therefore, higher-level error analysis may become easier with the ontology driven KB. For instance, for the

SST corpus, when a back-translated version of an error instance is available, such information can be aligned with each <**I**, **C**> pairs for further analysis of *interlanguage* at the application level.

Integration of Error Ontology with Learning Object through LO Metadata:

The Learning Objects approach [2] is inspired by the *object-oriented* technology. Learning Objects are independent and reusable units of instructional materials with specific learning goals. LOs are typically designed through analysis and decomposition of traditional instructional materials and annotated with learning object metadata. By aligning or mapping error-types with learning object metadata may provide an ideal personalized FL Learning Environment. Unlike Data-Driven Learning (DDL) [Hadley, 1997] Error-driven Learning (EDL) suits better in the context of FL pedagogy.

Collaborative Error Annotation: Collecting incorrect instances is more painstaking task than collecting correct instances. Annotation of incorrect instances is another overhead in building a learner corpus. The EKB Error Typology may be used to help FL teachers annotate learner's errors preferably over the WWW in a collaborative fashion.

Readers who regularly mark student assignments of a big class know the pain of handwriting similar comments on similar mistakes on each student's assignment. With the help of the EKB, FL teacher may be able to fairly automatically annotate (comment on) student's error.

5 Conclusions

The criticism of corpus-linguistics came from Noam Chomsky, where he argued: a corpus is by its very nature a collection of externalized utterances - it is performance data and is therefore, a poor guide to modeling linguistic competence. However, statistical natural language processing intensively used well-formed instances of language usage (from annotated corpora) and developed sophisticated applications over the years. Ambitious AI approaches such as, the OpenCYC initiative has been suffering significant delay to prove its potentials in full swing due to the fact that common sense modeling in a universal domain is still a far-reaching a goal given the state of the art technology and the resources mobilized for the task. Comparatively speaking, ours is a novel approach to utilize mature Semantic Web tools and technologies to model a smaller domain (linguistic errors) with explicit error instances (performance data/corpora) and competency knowledge (i.e., features not available explicitly in annotated learner corpora: error-contexts, learner's profiles, L1/L2 transfer, etc.) in an error ontology. Such a modeling and representation has an added advantage of making logical inferences and investigating error phenomena in both empirical and rational point of views.

We agree that much of the potentials of such an ontologybased modeling and the ontology-driven KB have yet to be justified without demonstrating them in real-life applications as outlined in Section 4. Nonetheless, the novel approach portrayed in the entire paper essentially shed lights to a new paradigm in corpus-based SLA research and FL pedagogy using Semantic Web inspired technologies. A simplified example is also included in Appendix A. Given the innate nature of language and tacit nature of knowledge, the error-driven modeling approach presented in this paper may enhance our know-how in analyzing human language, behavior and cognition.

Acknowledgments

The first author likes to thank the National Institute of Information and Communications Technology (NICT), Japan for its kind invitation as a Visiting Researcher during the period of 03/2003-09/2004. Most of the work discussed in this paper was undertaken at NICT's overseas lab located at Thailand Science Park. We are indebted to many foreign language teachers for their valuable comments and feedback. We specially like to thank Midori Tanimura, Akiko Harada and Rika Nohata for their helps with this project.

A A Simplified Example of EKB Based Application in FL pedagogy

The following example shows how the Error Ontology and the EKB may play an important role in FL pedagogy.

A common mistake made by beginner's level English learners is the incorrect use of preposition. This example illustrates how to check the correct use of preposition with the verb, *agree*. Taking the same verb, *agree* as an example, we also show how to generate typical correct usage examples with the help of EKB, NLP Tools and Learner corpora. Natural language processing tools plays important role in tokenization, parsing, etc. EKB is used for the logical inferences.

a. Input, Source of Error knowledge

Expert Knowledge in natural language: The object of agree with is a living thing The object of agree to is a non-living thing

Learner Corpora Annotation:

Erroneous instances with relevant corrections

Error Scope:

object-of "agree with" and "agree to"

Error Ontology:

Error Ontology combines error knowledge from experts and error instances from the Learner Copora

b. Error Checking

object-of(agree-with) = X; living-thing(X)=true -> correct

object-of(agree_with) = X; living-thing(X)=false -> incorrect

c. Generation of Typical Usage Example

Forall ?living-thing()

if living-thing() = proper-name
gen-sentence(subject-of(agreewith)+'agree with'+ living-thing());

if living-thing() = pronoun
gen-sentence(subject-of(agreewith)+'agree with'+objective-caseof(living-thing))

References

- [Becker et al., 2003] M. Becker, A. Bredenkamp, B. Crysmann and J. Klein. Annotation of Error Types for German Newsgroup Corpus. In Abeille A. eds, *Building and Editing Parallel Corpora*, Kluwer, pp. 89-99. Also at http://treebank.linguist.jussieu.fr/pdf/toc.pdf
- [CLC] The Cambridge Learner Corpus, http://uk.cambridge.org/elt/corpus/clc.htm
- [Corcho, 2003] Oscar Corcho, Mariano Fernández-López and Asunción Gómez-Pérez. Methodologies, tools and languages for building ontologies: where is their meeting point? *Data & Knowledge Engineering*, Volume 46(1) pp. 41 - 64
- [Corder, 1983] S.P. Corder. A role for the mother tongue. In S. Gass & L. Selinker ed. *Language Transfer in Language Learning*. Rowley, MA: Newbury House
- [Crysmann, 1997] B. Crysmann. Fehlerannotation, Technical Report, DFKI GmbH.
- [Duineveld et al., 1999] A. Duineveld, R. Studer, M. Weiden, B. Kenepa and R. Benjamis. WonderTools? A compara-tive study of ontological engineering tools. In Proceedings of the 12th Knowledge Acquisition Workshop (KAW99), Banff, 1999

- [EO-EKB] Error Ontology and EKB Web site, http://tcllab.org/mmhasan/eo-ekb/
- [Hadley, 1997] G. Hadley. Sensing the Winds of Change: An Introduction to Data-Driven Learning, http://web.bham.ac.uk/johnstf/winds.htm Hadley/DDL
- [ICAME] The ICAME Corpus Collections on CDROM, International Computer Archive of Modern and Medieval English, http://nora.hd.uib.no/icame/newcd.htm
- [Izumi et al., 2003] E. Izumi, T. Saiga, T. Supnithi, K. Uchimoto, H. Isahara. The Development of the Spoken Corpus of Japanese Learner English and the Applications in Collaboration with NLP technology, In Proceedings of Corpus Linguistics 2003, UK. Also available in NICT Technical Report Collection 2003.
- [Knublauch, 2003] H. Knublauch. An AI tool for the real world: Knowledge modeling with Protégé. Also at http://www.javaworld.com/javaworld/jw-06-2003/jw-0620-protege_p.html
- [Leech, 1992] G. Leech, Corpora and Theories of Linguistic Performance, In Svartnik ed. *Directions in Corpus Linguistics*, Berlin, Mouton De Gruyter, pp. 105-122
- [MacWhinney 1996] B. MacWhinney. The CHILDES System, American Journal of speech Language Pathology, Vol. 5, pp. 5-14
- [Nagoyadai] Learner Corpora for Japanese, Nagoya University, Linguistics Department, Japan, http://cookie.lang.nagoya-u.ac.jp/pub/data/
- [Protégé] The Protege Web Site, http://protege.stanford.edu/
- [Selinker, 1972] L. Selinker. "Interlanguage", International Review of Applied Linguistics. Vol. 10, pp. 209-230
- [SST] The NICT JLE Corpus (formerly, SST Corpus) Web site, http://leo.meikai.ac.jp/~tono/sst/
- [Tono *et al.*, 2001] Y. Tono, T. Kaneko, H. Isahara, T. Saiga, E. Izumi, M. Narita and E. Kaneko. The Standard Speaking Test (SST) Corpus: A 1 million-word spoken corpus of Japanese Learners of English and its implications for L2 lexicography. http://leo.meikai.ac.jp/~tono/sst/asialex2001.pdf
- [Treebank] The Penn Treebank Project, http://www.cis.upenn.edu/~treebank/home.html
- [Wiley, 2002] D. Wiley ed. *The Instructional Use of Learning Objects*, http://www.reusability.org/read/

The Use of Ontologies in Web-based Learning

Marvin Tan and Angela Goh

School of Computer Engineering, Nanyang Technological University, Blk N4, Nanyang Avenue, Singapore 639798 {ps7726426d, asesgoh}@ntu.edu.sg

Abstract

This paper proposes the use of domainspecific ontologies to provide semantics-based classification, navigation and query for learning repositories. Ontologies are used with metadata to classify resources into specific domains. Concepts from the ontologies are associated with these resources, thereby facilitating navigation of the e-learning material via a conceptual map. We also propose semantics support for query processing. By selecting a specific domain, the corresponding ontology will enable users to formulate conceptual-based multi-criteria queries, leading to more relevant and precise results.

1 Introduction

The advent of the Internet has changed computer-based learning tremendously. Learners have increasingly turned to web-based learning in order to access learning resources across the World Wide Web. In the past, learning repositories provided storage for large volumes of information, with index-and-search capabilities to facilitate self-paced learning. The resources available from these libraries contain links to other web resources holding the actual content. They are often catalogued using proprietary taxonomies or classification schemes and the accompanying metadata are usually proprietary. Each Learning Repository (LR) may have its own classification taxonomy. As a result, different digital libraries were unable to exchange resources with one another efficiently, reducing the effective reach of these libraries. Users also had to cope with the differences between metadata formats and terminologies employed, making the resource identification process more laborious. The introduction of metadata standards in the learning domain, such as the IEEE/IMS LOM (Learning Object Metadata) standard [IMS Global., 2001], has corrected much of these shortcomings. The metadata standards are meant to support the exchange of learning resources between repositories. Theoretically, users are able to identify these resources via a common set of description terminologies. In practice, however, although the IMS Metadata standard provides elements for classification of resources, these elements are not mandatory and the elements basically accept any kind of text input. This can be a problem, as users may find it difficult to relate learning resources from different repositories. This is because the 'meta-data content' created are neither represented in machine-readable semantics nor created using standardized semantics. Ontologies may help to achieve semantic interoperability. An ontology is defined as "an explicit specification of a conceptualization and provides an agreement about a shared conceptualization of a given domain of interest" [Gruber, 1993]. With proper semantic support in the form of standardized ontologies, the sharing of learning resources will be more effective. This paper proposes a framework that harnesses the benefits of ontologies and machine-readable semantics. The objectives are summarized as follows:

- To provide semantics-based classification of learning resources based on metadata
- To provide visualization of conceptual maps for domain ontology and the lists of associated resources.
- To provide more semantically precise query capabilities for identification and retrieval of resources using domain-specific ontologies.

2 Overview of Proposed Framework

Figure 1 shows a general overview of an Ontology-based Framework and its interaction with both human users and different data sources. Details of each component will be elaborated on in the following sections.

2.1 Ontology-Based Classifier (OBC)

The first component is the **Ontology-Based Classifier (OBC)**. Learning resources are classified into different domains with the use of an ontology. For the purpose of testing, the Upper Cyc Ontology [Cyc, 2002] was used. This can be replaced by other suitable ontologies such as SUMO (Suggested Upper Merged Ontology) [SUMO, 2004]. By using the metadata available from the resource, the OBC

will match the key terms with concepts within the ontology to determine the best match domain as well as the other possible domains the resource may belong to.



Fig. 1. Overview of Proposed Ontology-Based Framework

We will assume that the metadata taken from the resource is adequate and semantically-correct. While this assumption may not be valid for some resources, most of the resources encountered should have basic metadata that will describe the content of the resources adequately. We do not expect any system to be able to operate properly with inaccurate or incorrect metadata. An overview of the OBC is shown in figure 2.

The classification process is as follows:

- 1. The elements, 'title', 'keyword', 'description' and 'catalog' are retrieved from the XML metadata document that accompanies the learning object.
- NLP pre-processing techniques such as "stopping" and "stemming" are performed on the text blocks (e.g. description). For each remaining term, the frequency is obtained and a histogram-like data structure is obtained.
- 3. Each word obtained is sent to an interface to discover if the concept represented by the word is found in any of

the available ontologies. If found, the concept and the location of the concept is stored with the original word.

4. The concepts are then associated to the particular learning resource object and are used to classify the resource. As the concepts may be of different levels, top-level concepts will form the basic categories of this resource. Currently, we assume that every concept carries the same weight, as the ontologies we are employing do not allow us to impose weights directly. In addition, ontologies are represented in graph-like forms that do not show any hierarchy. Hence, it will be difficult to introduce weights without affecting the ontologies.

Using an IMS record taken from [Li, 2003], an example of this classification process is follows: Some of the critical terms obtained at the end of the preprocessing include microbes, organism, biosphere, entities, fossil and cycle. The terms, microbes, organism, biosphere, fossil are found in the Biology domain, while entities and cycle are found in other general ontologies. Hence, the resource is clearly in the Biology domain.

Learning resources can also be classified across domains. Concepts may also belong to several domains depending on the context. It has been mentioned that weights are not imposed at concept level but terms that belong solely to a particular domain are more indicative of the domain. Similarly, terms belonging to many domains are less authoritative. Hence, we will assign weights to concepts according to the number of domains they belong to. The significance of the concept in the classification process is inversely related to the number of domains the concept can be found under. In addition, we will consolidate the frequency of the terms or concepts under multiple domains. The domain that appears most frequently could be the common domain that these concepts belong to. In our framework, we maintain a list of resources associated with each concept within the domain-specific ontology and each list will be updated during the classification process.



Fig. 2. Overview of Ontology-Based Classifier

2.2 Ontology-Based Navigator (OBN)

The second component is the **Ontology-Based Navigator (OBN)**. Currently most LRs allow the browsing of resources through the navigation of the classification hierarchies. We propose to navigate resources using a conceptual map view. By displaying the concepts from the ontology in a graphical form, users may navigate the concepts and view the list of resources associated with each concept. Figure 3 shows how the different elements interact with the OBN.

Navigation Interfaces for LRs are extremely important as they provide the first level of support for users looking for specific information. Many LRs provide two basic methods of accessing available learning resources. The first method is to provide a search interface that allows users to look for resources based on certain criteria. The query processor will then present a list of possible results. Details of the resource querying will be given in the next section. The second method is to provide browsing based on topics and keywords. Usually, the user is presented with a list of main topics or category. The user selects a particular category and the system presents another list of sub-topics under the category selected. In this way, the user is able to 'specialize' a request for a certain class of materials.

Keyword-based browsing refers to the presentation of a list of indexed keywords to the user. When the user selects a particular keyword, he or she is presented with the list of resources containing the selected keyword. Both forms of browsing have their merits in providing a more structured and organized approach to filter learning material. However, they narrow the learning experience as users tend to zoom into an specific area of interest without getting to learn about related concepts and ideas. There is little room for adaptive learning.

The user will be better rewarded if he or she is able to see the related concepts of a category or concept selected. We propose the following solution, which caters to different learning experiences. The classical system of hierarchical taxonomies is provided for users who are regular users of the LR and are only interested in particular resources. For general users, a 'concept map and index' interface is provided. On the same interface, an index list of all the top-level concepts are presented to the user. Selection of any concept results in the visualization of all directly related concepts and cursor support is provided to allow the users to navigate to indirectly-related areas of the conceptual graph. In this way, users can process related knowledge before zooming into the specific resources. To prevent information overload, the 'concept map' will be designed in a similar way as 'online street maps', where basic operations such as zoom and directional navigation are provided. Only limited information will be available to the user at any point of time, improving the assimilation of the information and any decision-making process. A typical navigational model of LRs is shown in figure 4(a) while our enhancement is presented in figure 4(b).

2.3 Ontology-Based Query Manager (OBQM)

The third component is the **Ontology-Based Query Builder-and-Processor (OBQM)**, which allows users to build ad-hoc queries based on the concepts provided within a domain-specific ontology. The criteria for querying may be formulated using the concepts from the ontology. Search engines today are built on top of huge networks of indexed keywords with neither semantics nor contextual information.



Fig. 3. Overview of Ontology-Based Navigator



Fig.4. (a) Existing Navigational Model of LRs and (b) our Proposed Enhancement

An example search was performed at www.google.com for the keyword, "arm". We had expected articles and information on a part of human anatomy such as the bones and ligaments making up the arm. Instead, the first few results were on Microprocessors and the remaining results included brand names, club names, firearms and an acronym for 'Atmospheric Radiation Management'. Within the top 10 results, there was not a single link to an article on the 'arm' anatomy. When we refined our search to human arm, the results were better, returning articles related to the anatomy, animation or news containing the words human arm. However, the number of results returned was 1.5 million articles, which makes it impossible for a user to filter through individually. If the user was only strictly interested in human arm anatomy, then the top 10 results contained only 4 totally relevant results (2 of which were from the same site).

The problem of irrelevance remains in current search engine technologies that rely simply on keywords. In the context of Learning Repositories, this problem could be better managed. As each individual resource can be classified within specific domains, we could use the domains as primary filters. Users will be able to choose the domain ontology for the search to be performed on.

In addition, we also propose the use of domain-specific concepts to refine these queries. For example, when a user selects the Biology Ontology, a list of related concepts will be made available. Logical operators such as 'AND', 'OR' or 'NOT' will also be made available for the user to create

combinations of these concepts. In this way, we will improve the quality of the results returned during the query process, using ad-hoc query builders of different complexities for different types of users. Novice users may adopt the keyword-domain search while more experienced users may use the available concepts to formulate more complex queries for better precision. The results will be presented in a ranked list. However, we propose the inclusion of the option to invoke the OBN. This allows users to obtain an overview of the concepts related to the results. Figure 5 presents an overview of the OBQM.

3 Domain Registry

As seen in figures 2, 3 and 5, a Domain Registry (DR) is present to keep additional information, which will enable our framework to function independently of the existing repositories. This Domain Registry is a novel idea of placing critical index information of learning resources from multiple learning repositories. To have the DR work successfully, we make an assumption that most learning repositories have a script or interface that exposes the record of a learning resource to Internet users. For web-based LRs, such interfaces are usually web services or dynamic web pages that can display learning resource details. The Domain registry will act as an independent information store, which allows users to access learning resources from different LRs through external interfaces.



Fig. 5. Overview of Ontology-Based Query Manager

Each LR is registered at the DR by administrators who are required to provide a script or an interface URL and the necessary key parameters to identify or access specific learning resources. At the same time, the DR should also provide single-sign-on for users wishing to access several repositories through a common framework.

Hence, the DR plays a key role in establishing transparent loosely-coupled integration between different learning repositories. The registration of these repositories must be made relatively simple to encourage use of this service and as part of an open-source initiative, free access to the resources should be made available. However, the creation of such registries may encounter problems. One of the most critical is scalability. One of the objectives of this project is to provide ontological support for learning resources across the World Wide Web. Hence, we require the registry to be accessible on the Internet. Over the years, different communities have recognized that centralized registries are usually not scalable and not feasible unless the registry is monitored regularly and information is replicated across servers. Hence, we would like to examine peer-to-peer frameworks [Neijdl et al., 2002] to implement distributed domain registries. The availability and reliability of such registries will be critical to the success of our proposed framework.

4 Related Work

In the past few years, there has been much interest in the roles of ontologies in the learning environment. One of the earliest research efforts on eLearning and the Semantic Web was discussed in [Stanjonovic *et al.*, 2001], where an eLearning scenario was presented. Ontology-based descriptions of content, context and structure of learning materials are employed to provide flexible and personalized access to the resources. The main difference between their scenario and our framework is the use of ontologies. The main ontology

they adopt is a 'course ontology' that describes the context and structure of the learning resources. They also make use of the Ontobroker [Decker *et al.*, 1999] inference engine to answer queries and infer new facts from the existing information. However, the ontology used in [Stanjonovic *et al.*, 2001] is restricted to course description and the links with the learning resources are more structural than semantic.

The generation of flexible taxonomies for learning resources is also discussed in [Papatheodorou et al., 2002]. Their methodologies are based on data mining techniques. The ontology discovery process consists of basic NLP (Natural Language Processing) and machine learning techniques such as pattern discovery. The main drawback in this process lies in the requirement of training data and its intractability. Since the level of accuracy cannot be guaranteed using machine learning techniques and large corpuses of training data are hard to come by, we prefer to employ heuristic techniques to achieve similar results. However, there was one interesting issue raised in the paper - the adaptive discovery of ontology-based taxonomies may result in the limitless depth of the hierarchical taxonomies and the display of the taxonomies may result in poor humancomputer interaction design and information overload for the user. Similarly, the challenge of building and maintaining metadata using an ontology based on an extended dictionary was addressed by [Apted et al., 2004].

Other related research in the support of learning using ontologies includes the 'Collaborative Courseware Authoring Support' [Dicheva *et al*, 2002] and 'Objectoriented collaborative course authoring environment'[Cristea and Okamoto, 2001]. [Dicheva *et al*, 2002] provides support for concept-based web courseware authoring with the use of a domain ontology. The ontology can be used by authors in queries or by the system to perform semi-automatic authoring activities. Courseware are then manually linked to the concepts by the authors. The system also provides domain engineering of the ontology. Hence, the ontology has to be built from scratch by the courseware authors. The two main problems are, firstly, the ontology is expected to grow exponentially and links to new concepts may have to be created continuously for earlier created courseware. Secondly, with the authors creating and maintaining the domain ontology, the resultant ontology becomes proprietary and unsynchronized with external ontologies.

"My English Teacher" [Cristea and Okamoto, 2001] focuses on an English upgrading course authoring environment. In this project, 'concept mapping' is used to support authoring where the concepts are built using keywords. 'Concept mapping' is the graphical technique of representing concepts and corresponding hierarchical relationships to present the domain model. It presents some ideas of resolving the display of concept maps or hierarchical views that are too wide or deep. At the same time, it also explores automatic concept mapping to content, which we will refer to in our project. However, the emphasis is on courseware authoring for a proprietary domain, whereas we are proposing a more generic framework for learning repositories

5 Conclusion

One of the key problems affecting the effectiveness of Learning Repositories is the lack of semantic support. Without this support, users are forced to spend more time to review irrelevant learning resources. At the same time, non-standard classification schemes also prevent efficient interoperability between these repositories.

Our proposed framework will remove most of these inadequacies by providing support in the form of commonly available domain-specific ontologies. By provided semantic support for the classification, navigation and query processing for Learning Repositories, users will benefit from a clearer conceptual view of resources as well as a more efficient learning experience through more precise retrieval of relevant information.

A key issue in our future work is to support Ontology growth. As research on Ontology engineering and management is still at an early stage, it is envisaged that considerable improvements to ontological support and tools will be seen in near future. In addition, the ontologies to be employed will constantly be updated. Hence, our framework must cater for the possible growth of the underlying ontologies. In our framework, local copies of the ontologies will be used. The local copies will be compared periodically with the master copies and the changes will be propagated to the rest of the system. In this way, we will be able to maintain the semantic accuracy of the repository and at the same time maintain conformance to the ontologies employed.

References

- [Apted *et al.*, 2004] Trent Apted, Judy Kay and Andrew Lum, Supporting metadata creation with an ontology built from an extensible dictionary. In *Proceedings of* 3rd Int, Conf on Adaptive Hypermedia & Adaptive Web-Based Systems August 2004. Springer-Verlag.
- [Cristea and Okamoto, 2001] Alexandra Cristea and T. Okamoto, Object-oriented collaborative course authoring environment supported by concept mapping in MyEnglishTeacher, Educational Technology & Society 4 (2), 104–115, 2001.
- [Cyc., 2002] Cyc, http://www.cyc.com/cycdoc/vocab/ upperont-diagram.html, OpenCyc Selected Vocabulary and Upper Ontology, 2002.
- [Decker et al, 1999] Stefan Decker, Michael Erdmann, Dieter Fensel and Rudi Studer. Ontobroker: Ontology based access to distributed and semi-structured information, In Database Semantics - Semantic Issues in Multimedia Systems, *IFIP TC2/WG2.6 Eighth Working Conference on Database Semantics (DS-8)*, Rotorua, New Zealand, pages 351–369, January 1999.
- [Dicheva et al, 2002] Darina Dicheva, Lora Aroyo and Alexandra Cristea. Collaborative Courseware Authoring Support, In *Computers and Advanced Technology in Education*, Cancun, Mexico, pages 52– 57, May 2002.
- [Gruber, 1993] Tom Gruber. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. In Formal Ontology in Conceptual Analysis and Knowledge Representation, N. Guarino, R. Poli (Eds.), Kluwer Academic Publishers, 1993.
- [IMS Global., 2001] IMS Global, http://www.imsglobal. org/metadata/index.html, IMS Learning Resource Meta-data Specification Version 1.2, 2001.
- [Li., 2003] Yaxin Li, Biological control: A guide to natural enemies in North America, sMETE.org, URL http://www.nysaes.cornell.edu/ent/biocontrol/, 2003.
- [Neijdl et al., 2002] Wolfgang Nejdl, Boris Wolf, Changtao Qu, Stefan Decker, Michael Sintek, A. Naeve, M. Nilsson, M. Palmer, T. Risch. Edutella. A P2P networking infrastructure based on rdf. 11th International World Wide Web Conference, Honolulu, Hawaii, USA, May 2002.
- [Papatheodorou et al, 2002] Christos Papatheodorou, Alexandra Vassiliou and Bernd Simon. Discovery of ontologies for learning resources using word-based clustering, ED-MEDIA 2002, Denver, USA, 2002.
- [Stanjonovic *et al*, 2001] Ljiljana Stojanovic, Steffen Staab and Rudi Studer. elearning based on the semantic web. WebNet 2001 - *World Conference on the WWW and Internet*, Florida, USA, pages 23-27, 2001.
- [SUMO, 2004] SUMO Suggested Upper Merged Ontology. http://ontology.teknowledge.com/, 2004.

Integration of Hyperbooks into the Semantic Web

Jean-Claude Ziswiler, Gilles Falquet and Luka Nerima Centre universitaire d'informatique CUI – University of Geneva

24, rue du Général Dufour – 1211 Geneva 4 – Switzerland {Jean-Claude.Ziswiler, Gilles.Falquet, Luka.Nerima} @cui.unige.ch

Abstract

A crucial aspect of the Semantic Web is the capacity to add formalized meanings to information to enable non-human actors to process it. This is usually accomplished by linking the information to an ontology that describes the domain's concepts. In the Web's context it does not seem realistic to represent this semantic layer on a central server, as this model would not reflect the characteristics of presently used, non-centralized networks like the current Web. Therefore we are confronted with a huge number of locally developed and stored ontologies, and we need some kind of integration techniques to connect ontologies developed for the Semantic Web. In this article, we describe our experience with ontology-based e-learning systems and we propose a mechanism to integrate such systems into a Semantic Web context. We concretely present our hyperbook model and show how hyperbooks can be integrated into digital libraries by an ontology mapping procedure.

1 Introduction

Most of the digital libraries available on the Web are collections of documents stored on a document server where users can for instance use full-text search engines. Results are then presented in title lists with annotated, small informational fragments of the (perhaps) most relevant documents. By clicking on these references, users can see the whole text or information, which is stored in the form of a PDF-file or in a similar format. Such systems are simple to build, but have their drawbacks. From the semantic point of view, they are in fact static hypertext systems. There is no semantic representation of the content, and different systems and servers are not interrelated.

Even if most Web sites have a dynamic part, it is normally limited to information which is semantically poor and which can be easily stored in a common database management system. For instance, most e-commerce shops have such functionalities. A virtual bookstore has information about titles, authors, page-numbers and prices, but the content of the books is often not formally represented. Systems including formal concept representation give for instance the possibility of a non-linear reading or the presentation of narratives in multiple forms. Therefore, we propose an approach that allows fragmenting text sources and linking these fragments to domain ontologies. This is the main idea of what we call a hyperbook.

Such a dynamic hyperbook model results in a local, noncentralized and heterogeneous system. Heterogeneity is one of the caracteristics of the Web and we anticipate it to be a permanent one. We even believe that one of the major keys to the success of the Internet is the heterogeneity of the systems where all users may create the pages in their own way. When building semantic web applications, we should not restrict authors when they build the content. We furthermore should try to find solutions that involve a great flexibility. But to avoid the risk that users create isolated systems, we propose to assemble many hyperbooks into a digital library. The main task of this process is the integration of different domain ontologies.

The ontology integration problem is one of the most challenging tasks in the Semantic Web. The integration procedure for digital libraries we present uses specific properties of hyperbooks or similar e-learning systems like fragments, documents and personalization elements. But we can imagine that the approach could be extended to other domains, to the extent that there is a good information base or a document repository available.

The paper is organized as follows: Section 2 describes the existing research in the domain of hyperbooks and presents our model. Section 3 includes the step from hyperbooks to a digital library, and section 4 concludes the article with an outlook about future research questions.

2 Hyperbooks

There is presently no consensus on a common virtual document or hyperbook model. Nevertheless, most of the proposed models are comprised of (at least) a domain ontology and a fragment base. These models generally differ on the user interface part, i.e. how to specify the production of user-readable documents (with declarative languages, through specific ontologies, with inference rules, etc.). Crampes and Ranwez [Crampes and Ranwez, 2000] propose two models of virtual documents. Both of them use domain ontologies for indexing informational fragments (the resources). In the first case, a "conceptual backward chaining" strategy can construct reading paths corresponding to the user objectives (described in terms of conceptual graphs). In the second case, a pedagogical ontology defines teaching rules, which guide the assembling of fragments to produce documents with respect to a predefined pedagogical approach. These rules, in particular, force the order of appearance of information in the documents. An inference engine generates documents that satisfy these rules.

The InterBook adaptative hyperbook project [Brusilovsky et al., 1998a] is based on two models: The domain model and the student model. The domain model is represented by a network whose nodes correspond to domain concepts and links to their relationships. In fact, the domain model corresponds to what is named ontology in many research paper. The student model describes the student knowledge as well as the student learning goals, both expressed in terms of the concepts of the domain model. A glossary is used to describe the navigational paths in an InterBook hyperbook. Several books can be integrated in a bookshelf. They are connected by sharing the same set of domain concepts, thus avoiding the ontology integration problem.

Iksal and Garlatti [Iksal et al., 2001] propose a comprehensive and detailed model of virtual documents. It is based on four ontologies for modeling the domain, the metadata, the applications and the user. These ontologies allow a fully declarative approach to document composition.

Another approach is the Scholarly Ontologies (ScholOnto) project [Buckingham Shum et al., 2000]. It defines a digital library server that supports multiple and possible conflicting interpretations of a research document. Essential parts are the formalized encoding of interpretations of a scholarly document and the compilation to create semantic hypertexts. Objects (concepts or data) are connected by multi-typed links defining the properties of the objects. Reading a document is defined as going through a set of claims. Claims are the basics of formalized interpretations, which are considered as two connected objects and a (optional) typed link. The result is a set of semantic annotations about the document's contributions like citations to other key literature. As we will see, this approach is similar to our hyperbook model in the way in which links are typed.

[Dicheva et al., 2004] present a framework for standards-based, ontology-aware course libraries and an environment for building, maintaining, and using such libraries. The aim is to provide a system built of (existing) learning resources to assist students. Dicheva et al. propose a topicmap based system that allows authors to describe the metadata of their learning material according to Dublin Core (DCMI) or IEEE LOM. The model is close to our approach in the way relationships were represented. Each relationship has a type and authors can use a pool of predefined relationship types. Bocconi [Bocconi, 2003] describes a hypertext generation system to automatically select and compose scholarly hypermedia. The presented content is generated through a domain ontology containing the concepts and their relations and a discourse ontology containing different roles and narrative units describing different genres. The discourse ontology holds a very detailed and highly formalized description of the points of interest that a user can have about a domain. By integrating hyperbooks into digital libraries, we should also consider the question how to integrate elements stored apart of the domain ontology like essentials for hypertext personalization purposes. For this task, the abovementioned model seems to be very interesting and might be a good base.

In general, the hypertext personalization problem [Brusilovsky, 1998b] seems essential for developing hyperbook systems in the domain of e-learning. In [Wu et al., 2001], the authors propose a model of adaptive hypertext which includes a domain model, a user model and adaptation rules. The domain model is a semantic network consisting of domain concepts and relations between concepts. This model serves essentially to define adaptation rules, depending, for instance, on the concepts known or appropriated by the user. We have included in a similar way some adaptive mechanisms, such as points of view, into our system [Falquet et al., 2004].

A new research field has emerged in the last years that concentrates on the concept of personalizable virtual documents [Crampes, 1999; Garlatti and Iksal, 2001]. Personalizable virtual documents are defined as sets of elements (often called fragments) associated with filtering, organization and assembling mechanisms. According to a user profile or user intensions, these mechanisms will produce different documents adapted to the user needs. The model we will present is based on this approach.

Our hyperbook model is build upon a *fragment repository*, a *domain ontology*, and an *interface specification* [Falquet and Ziswiler, 2003]. The fragments and the ontology including their interconnecting links form the structural part of the hyperbook (Figure 1).

The basic informational contents of the hyperbook are made of reusable fragments. These fragments can be small texts, or even illustrations or programming code, but we want to avoid that authors write large fragments. They have to divide the content sources, like documents, and to place the created fragments around concepts. Fragments are connected by structural links, for instance from fragments to sub-fragments. These typed links indicate the roles they play in a group of fragments (compound fragments). For instance an exercise could be made up of a question fragment, one or more answer fragments, and a discussion.

The semantic structure is described by a domain ontology. It is intended to hold a formal representation of the domain's concepts and used for indexing or qualifying the fragments.



Figure 1. The hyperbook structure

By establishing typed links between fragments and concepts, the information content stored in the fragments is referenced by the concepts of the domain ontology. This shows relations between the fragments and the concepts, but also puts the fragments into a context or a semantic environment. Typical link types are *Definition*, *Property*, *Example*, *Illustration*, *Exercise*, *Instance*, or *Reference*. Link types should be predefined, so that authors of a hyperbook can choose from a limited number of unambiguous defined link types.

The fragment-ontology structure allows representing different elements of an e-learning system. Domain knowledge can be represented in different, sometimes contradictory ways according to pedagogical or narrative purposes. Besides, we can annotate concepts with examples, illustrations, exercises and solutions. Even a discussion is possible (*topic* and *message* fragments connected through *about* and *replyto* links), and authors can give their opinion about a subject (arguments, opinions).

This hyperbook structure enables to generate different hypertext views. As the number of concepts in the ontology is generally much smaller than the number of information fragments, the user can browse the ontology and then go down to the connected fragments. Fig. 2 shows an extract of a hyperbook that we have used in a computer science course this year. Around a concept, we found links to other concepts (some of them are even automatically generated by link inference techniques) and annotated fragments like comments and examples.

This model seems more adequate to e-learning purposes than approaches that are based on large ontology repositories, sometimes also on several types of ontologies (toplevel ontology, application ontology). Our experience has shown that people are more used to building small text fragments and annotating them with the most important concepts of the domain than creating complex ontology structures. The result will be small, but expressive domain ontologies. Of course, a more formal representation of the content might have advantages for the ontology integration process. In the next section, we show how we integrate such small ontologies by using the fragment repositories.

[New Fragment] [Enter] [Concepts] [Fragments] [My Fragments] [Upload] Finate-state Machine ofsi1 Concepts A Finite-state Machine is composed of <- reference Regular Language <- reference Accepting a string <- [Equivalence] -> FSM non-deterministic 1. An alphabet of symbols A = {s1, s2, ..., sn} A set of states $Q = \{q1, q2, ..., qn\}$ <- [Comparison] -> Turing Machine 3 A initial state q1 included in Q A set of final states included in O 4. 5. A transition function d of Q x A in Q [Modify] [Delete] [Create] Fragments Comment: Definition FSM - general remarque A finite-state Machine is deterministc, intuitively if there is only one way of resolution for a given Example word: being given a initial state and a character, Tennis and FSM there is not more than one final state.

Figure 2. An extract of a virtual hyperbook

3 From Hyperbook to Digital Library by Ontology Alignment

As all the hyperbook models presented above are based on ontologies, their integration plays a major role in the domain of virtual books and *a fortiori* in the domain of virtual libraries. If we suppose that each virtual book has its own domain ontology, we need an integration to create a semantically coherent virtual library. It is important to note here that it is not very realistic to suppose that all the books will be linked to the same (global) ontology, because either such an ontology does not currently exist or even if it exists, it contains only stable and well-established concepts and will not have the desired level of specialization or diversity. Thus, it will not be convenient for books on new and advanced topics.

Our ontology integration approach consists in preserving the concepts and the links of the initial ontologies and establishing relations of equivalence between elements of the origin ontologies. This approach seems to be much more flexible than a complete fusion of the hyperbook's ontologies. In the domain of e-learning with non-centralized, semantic information systems, we have to consider that content is under a permanent evolution process. New parts of documents will continuously be added to the Digital Library.

We are focusing on work that is based on *alignment* techniques. This means bringing two ontologies into mutual agreement by extending the ontology with drawing links between concepts [Klein, 2001]. As subcategory, *mapping* ontologies means relating similar concepts or relations from different ontologies to each other by an equivalent relation. In both cases, the existing ontologies will persist.

Some heuristics for semi-automatic ontology integration can be found in different approaches. They are based either on identifying structural or naming similarities as can be found in the SMART algorithm [Noy and Musen, 1999] (now integrated in the interactive ontology merging tool PROMPT) or on machine learning techniques, used for instance in the GLUE system [Doan et al., 2002]. Other approaches like OBSERVER [Mena et al., 2000] use semantic interrelations (limited to the analysis of taxonomic relations) to specify mappings between (not synonymous) concepts of two ontologies. It defines lower and upper bounds for the precision and recall of ontology-crossing queries that are based on manually defined subsumption relations.

As a conclusion to this introduction to ontology integration, we will remember the following points:

- Ontology integration is essential for building Digital Libraries or semantic web applications, because it does not seem realistic to assume that there is one common ontology that covers all the domains and subjects. Using local domain ontologies allow users to express concepts in a higher explicitness.
- Automatic ontology mapping for building Digital Libraries is a *multi-level* task. We should take advantage

of the larger information base of the hyperbooks (taxonomic relations between concepts, but also other *attributes* of the concepts and especially *instances* and annotated *fragments* or *documents*). Including all this information will allow automatic procedure, contrary to other applications in the Semantic Web where automatic ontology mapping seems more difficult.

3.1 Ontology Mapping by measuring the semantic similarity

For mapping ontologies, we propose to calculate semantic similarities between the concepts. We have noticed that in the existing literature, ontology mapping is a crucial assumption for measuring similarity. We will not consider these approaches, mainly because most of them carry out ontology mapping manually and just focus on how to represent them. This shows that developing algorithms for ontology integration is a crucial, but also a critical task, and might serve not only to research in the domain of semantic similarity measurement, but also to the whole Semantic Web community.

[Weinstein and Birmingham, 1999] present an overview about semantic similarity measures. They have identified three groups:

The first category are *filter functions* and are presented as inexpensive, universally applicable, and appropriate for identifying initial sets of candidate recommendations. A first subcategory concerns measures based on path distance, but they are labelled fragile due to their sensitivity to the degree of detail in the ontological structure. When classifying new concepts, the measures can change although the compared concepts have not changed. Other filter measures leverage the assumption of measures based on path distance, that local concepts in differentiated ontologies inherit. They use inheritance links to identify concepts and the definitions of both the source and target concepts.

The *matching-based functions* build and evaluate (maximal) one-to-one correspondence between elements of concept definitions represented as graphs. They enable analysis of similarities and differences between the concepts.

The *probabilistic functions* require domain-specific knowledge of the joint distribution of primitives. To model the joint distribution, Bayesian Networks are necessary.

The identified groups cross a spectrum with respect to the *degree of required knowledge specification*. Filter functions do not involve role semantics. The matching-based functions use knowledge of roles considered independently, and the probabilistic functions exploit knowledge of the interactions among roles.

3.2 Similarity measure in Digital Libraries

For the integration of hyperbooks with their domain ontologies into a Digital Library, we focus on techniques that try to establish one-to-one correspondences between ontologies of two hyperbooks. We first carry out the computation of the semantic similarity to decide if two concepts belong to the same semantic field or not. Using string similarity measurement approaches might be the first and simplest task for this computation. However, considering only the syntax of the concepts in question is too basic. In the domain of Digital Libraries, we can first assume that there are not many words with the same spelling in the different ontologies, and second that there exists terms with the same spelling, but which differ semantically (polysemy). Figure 3 shows extracts from four sample-ontologies, each speaking about Football. As we can see, the word "Football" does not always stand for the same meaning.

Using the WordNet ontology or another available toplevel ontology for word disambiguation and resolving the meanings of polysemous words might be a solution. But in these kinds of ontologies, we can't usually find specific concepts, as only more general terms are represented. Looking at the ontologies in Figure 3, we can assume that there are at least some words in the upper-class level that can be found in the WordNet ontology. This would allow introducing some links between the different ontologies. But as our experience with e-learning systems has shown, authors do not normally model the upper level part of their ontologies. Another question is also whether the time needed to calculate word disambiguation with WordNet is reasonable.

The domain ontology of a Digital Library is not just a terminological ontology where the collection of concepts are organized by a partial order. It is an axiomatized ontology whose concepts are distinguished by different kinds of relations to other concepts. We have mentioned above the importance of link typing in a hyperbook structure, and we can now take advantage of this formalized representation for the ontology integration process. Thus, we need a similarity measure approach that considers more elements than just terminological relations. For this, we propose to use an extension of the technique of [Rodríguez and Egenhofer, 2003]. The approach is based on entity classes that are groups of equivalent or very similar words. The similarity between two entity classes is the weighted sum of three measurements: similarity of the terms (set of synonyms), similarity of the semantic neighbourhood (set of concepts close to the entity class in the graph) and similarity of the attributes (characteristics and properties, so called distinguished features like set of values). The similarity function takes into account the depth of the entity classes relative to their respective ontologies. The approach considers also cognitive properties of similarity by introducing asymmetric measurement of semantic similarity. The authors argue that according to people's judgement, the base and the target concept can have different roles. For instance, the perceived similarity from a concept to its super-concept is greater than the perceived similarity from the super-concept to the concept. Finally, comparing not only sets of synonyms, but involving also distinguished features into the similarity measures means introducing grades of similarity. The result is no longer a simple binary expression (same or different word), it allows to detect also similar words.

As we have mentioned, most of the known techniques for semantic similarity measurement need a primarily integrated ontology. As we want to invert the process (integrating ontologies by first determining the semantic similarity measurement), we profit of the above-described approach because it can process the measures in not yet related ontologies. This is done by simply establishing a relation to an imaginary and more general entity class "anything" from the two root-concepts of the ontologies in question. This allows to calculate of the distance from two entity classes to the immediate super-class even if there is no "natural" common super-class.



 Established by syntactical similarity measure, but rejected when calculated in conjonction with WordNet disambiguation

Figure 3. String similarity measurement with word disambiguation by WordNet

3.3 Discussion

A first problem of the algorithm of Rodríguez and Egenhofer that is to mention is the fact that the assigned weights of each specification component depend on the characteristics of the ontologies. For instance, the authors mention that, when many polysemous terms occur within the ontologies, the synonym set component might not be a good indication of similarity. In consequence, the relative weight assigned to this component should be reduced. But it seems critical when we have to analyze the occurrence of polysemy first. If we want to decide how the weights have to be assigned, we have to look at the three components in detail.

First, comparing different synonym sets by a word matching process is a very basic level of similarity measure. When checking the number of common and different words in the synonym sets, there is a gradual similarity measure, but polysemy will not be detected.

Second, the semantic neighbourhood is determined by analyzing related super-concepts of an entity class. The links are labelled with types like "is-a" (hyponymy) or "part-whole" (meronymy). The authors use path distance (with a preliminary defined upper-bound) to define the semantic neighbourhood of an entity class. In a hyperbook, we can assume that the semantic neighbourhood is a priori of "good quality" for similarity measures. The problem is how we should determine the boundary of this distance.

But the most important problems can be found in the third component, introducing distinguished feature. Although when they provide a more formal description of a domain, Rodríguez and Egenhofer have mentioned in their evaluation section that feature matching alone is insufficient for detecting the most similar entity classes as many entity classes share common features or have a common superclass from which they inherit common features. Even in combination with the other two components, the *recall* is lower for finding the *most equivalent* concepts than without using distinguished features.

The approach detects equivalences between concepts of two ontologies, but by using distinguished features, it also returns similar concepts. This is an important fact. In contrary to the ontologies that Rodríguez and Egenhofer have used for the evaluation of their approach (top-level ontologies like WordNet), we are primarily confronted with small domain ontologies. This means that we can't be sure that there are many equivalent concepts. To get a stronger mapping between the two ontologies, we need an approach that can at least detect some similar concepts.

At this moment, it seems important to remark that this approach is adapted to hyperbooks and Digital Libraries and most probably can't be used in other Semantic Web applications as such. A high degree of domain formalization and a strong explicitness of concept representation are important assumptions.

Before we explain how we use an extended concept of distinguished features to improve semantic similarity, we want to mention another unresolved problem of this approach. In fact, there probably is no single root in each of the ontology to compare. But our model presented in section 2 explicitly allows a flat hierarchical ontology structure, where a single root might not appear very often. Such a hyperbook model might be much closer to e-learning purposes. So, the question remains how to build a common virtual root if relevant domain ontology is empty. We could try applying the approach of Rodriguez iteratively, but applying the process to an even smaller part of the ontology might not represent the best solution.

The above-mentioned approach of [Doan et al., 2002] is particularly interesting because they deal with taxonomical relations and instances to establish links between two ontologies. Contrary to Rodriguez, they don't establish an artificial top-level root concept to measure the semantic similarity, they primarily match the instances with machinelearning technique to be able to define the semantic similarity of two concepts. Then, they define distributions over the concepts to determine the final semantic similarity by a user-supported approach. Finally, they determine which of the derived relations will be considered into the ontology mapping according to given *domain constraints* and *relaxation labelling* (taking into account the neighbourhood of a concept).

3.4 Improvements

In our case of virtual documents, we make use of additional information to evaluate the similarity between concepts. We also consider the annotated fragments and we take advantage of the fact that links between fragments and concepts are typed. As we have seen above, the fragments were annotated according to a predefined list of link types. If two concepts A and B are bound by links of the same type t to sets of fragments t(A) and t(B) respectively, the "documentary" similarity between t(A) and t(B) will be taken into account in the definition of the similarity between A and B. To define the similarity between t(A) and t(B), we will use a traditional technique of information retrieval (for instance, the cosine between the *tf-idf* vectors representing the documents in the space of terms [Salton, 1989]). Then, we define the similarity between t(A) and t(B) based on the similarities between documents (for example by taking the maximum similarity found between all the fragments of t(A) and t(B)). The similarities obtained for all types of links will then be added up to the similarity measure computed at the conceptual level.

It is important to remark that link typing is crucial here. Indeed, the comparison makes sense only if the compared fragments play the same role with respect to a concept. If, for instance, fragment a is an *example* of concept A whereas b is a *counterexample* of B, a strong similarity between aand b does not imply a strong similarity between A and B, on the contrary.

4 Future Research Questions and Conclusion

In this article, we have presented our hyperbook model, which is based on domain ontologies and fragment repositories and an integration process into a Digital Library, which is an extended version of the approach of Rodríguez and Egenhofer. We have applied their algorithm to Domain ontologies by involved typed relations between fragments and the domain ontology. In future research, we will work on the integration process, especially on the validation of matched links between concepts of the domain ontology, but also on the integration of hypertext personalization elements like points of view. This also seems an interesting approach for validating the established relations because points of view help to describe the concept of an ontology more precisely.

References

- [Bocconi, 2003] S. Bocconi. Automatic Presentation Generation for Scholarly Hypermedia. In Proceedings of the 1st International Workshop on Scholarly Hypertext at the Hypertext 2003 conference, Nottingham U.K., 2003.
- [Brusilovsky et al., 1998a] P. Brusilovsky, J. Eklund, E. Schwarz. Web-based Education for All: A Tool for Development Adaptative Courseware. In Computer Networks and ISDN Systems (Proceedings of the Seventh International World Wide Web Conference), April 1998.
- [Brusilovsky, 1998b] P. Brusilovsky. Methods and techniques of Adaptive Hypermedia. User Modeling and User-Adapted Interaction. Adaptive Hypertext and Hypermedia, Kluwer, 1998, 1-43.
- [Buckingham Shum et al., 2000] S. Buckingham Shum, E. Motta, J. Domingue. ScholOnto: An Ontology-Based Digital Library Server for Research Documents and Discourse, In International Journal on Digital Libraries, vol. 3 (3), Springer-Verlag, Berlin, 2000.
- [Crampes, 1999] M. Crampes. User Controlled Adaptivity versus System Controlled Adaptivity in Intelligent Tutoring Systems. In Proceedings of the Artificial Intelligence in Education Conference, Le Mans, France, 1999
- [Crampes and Ranwez, 2000] M. Crampes, S. Ranwez. Ontology-Supported and Ontology-Driven Conceptual Navigation on the World-Wide Web. In Proceedings of the ACM Hypertext 2000 Conf., San Antonio, 2000.
- [Dicheva et al., 2004] D. Dicheva, C. Dichev, Y. Sun, S. Nao. Authoring Topic Maps-based Digital Course Libraries. In Proceedings of the Workshop on Applications of Semantic Web Technologies for Adaptive Educational Hypermedia (SW-EL@AH'04), in conjunction with AH 2004, Eindhoven, The Netherlands, August 23-26, 2004.
- [Doan et al., 2002] A. Doan, J. Madhavan, P. Domingos, A. Halevy. Learning to Map between Ontologies on the Semantic Web. In Proceedings of the ACM World Wide Web Conference, Honolulu, Hawaii, May 2002.

- [Falquet and Ziswiler, 2003] G. Falquet, J.-C. Ziswiler. A Virtual Hyperbooks Model to Support Collaborative Learning. In AIED 2003 Supplemental Proceedings, Sydney, Australia, July 2003.
- [Falquet et al., 2004] G. Falquet, L. Nerima, J.-C. Ziswiler. Adaptive Mechanisms in a Virtual Hyperbook. In Proceedings of the 4th IEEE International Conference on Advanced Learning Technologies (ICALT2004), Joensuu, Finland, August 30 - September 1, 2004.
- [Garlatti and Iksal, 2001] S. Garlatti, S. Iksal. Revisiting and Versioning in Virtual Special Reports. In Proceedings of the Hypertext 2001 conference, Arhus, Denmark, 2001.
- [Iksal et al., 2001] S. Iksal, S. Garlatti, P. Tanguy, F. Ganier. Semantic Composition of Special Report on the Web: A Cognitive Approach. In Proceedings of the H2PTM'01 conference, Valenciennes, France, 2001.
- [Klein, 2001] M. Klein. *Combining and relating ontologies: an analysis of problems and solutions*. IJCAI2001 workshop on Ontologies and Information Sharing, Seattle, 2001.
- [Mena et al., 2000] A. Mena, A. Illarramendi, V. Kashyap, A. Sheth. OBSERVER: An Approach for Query Processing in Global Information Systems Based on Interoperation across Pre-existing Ontologies. In Distributed and Parallel Databases, Vol. 8 (2), 2000. pp. 223-271.
- [Noy and Musen, 1999] N. F. Noy, M. A. Musen. An Algorithm for Merging and Aligning Ontologies: Automation and Tool Support. In Proceedings of the Workshop on Ontology Management at the Sixteenth National Conference on Artificial Intelligence (AAAI-99), Orlando, Florida, AAAI Press, 1999.
- [Rodríguez and Egenhofer, 2003] M. A. Rodríguez, M. J. Egenhofer. Determining Semantic Similarity among Entity Class from Different Ontologies. In IEEE Transactions on Knowledge and Data Engineering, Vol. 15, No. 2, March/April 2003.
- [Salton, 1989] G. Salton. Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Addison-Wesley, 1989.
- [Weinstein and Birmingham, 1999] P. C. Weinstein, W. P. Birmingham. Comparing Concepts in Differentiated Ontologies. 12th Workshop on Knowledge Acquisition, Modelling and Management, Banff, Canada, 1999.
- [Wu et al., 2001] H. Wu, E. de Kort, P. De Bra. Design Issues for General-Propose Adaptive Hypermedia Systems. In Proceedings of the ACM Hypertext 2001 Conference, Aarhus, Denmark, 2001, pp. 141-150.

Ontologies for Authoring of Intelligent Educational Systems

Lora Aroyo

Eindhoven University of Technology Faculty of Mathematics and Computer Science P.O.Box 315, 5600 MD Eindhoven, The Netherlands l.m.aroyo@tue.nl

Riichiro Mizoguchi ISIR, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047 Japan miz@ei.sanken.osaka-u.ac.jp

Abstract

The goal of this paper is to illustrate the beneficial role of ontologies in achieving efficient authoring support for Intelligent Educational Systems (IES). We present our ideas within an ontology-driven framework EASE offering power with respect to the functionality, generic approach for its support of instructional strategies and user-friendliness in its interaction with the author. A central function in it has an authoring task ontology that at a metalevel defines and controls the configuration and tuning of an authoring tool for a specific authoring process. In this way we achieve more control over the evolution of the intelligence in IES and reach a computational formalization of IES engineering.

1 Introduction

For many years now, various types of Intelligent Educational Systems (IES) have proven to be well accepted and have gained a prominent place in the field of courseware [Murray, 2003b]. IES also have proven [Brusilovsky, 2003; Murray, 2003a] that they are rather difficult to build and maintain, which became, and still is, a prime obstacle for their wide spread popularization. The dynamic user demands in many aspects of software production are influencing research in the field of intelligent educational software as well [Ainsworth et al, 2003]. Problems are related to keeping up with the constant requirements for flexibility and adaptability of content and for reusability and sharing of learning objects [Devedzic et al, 2000].

Thus, the IES engineering is a complex process, which could benefit from a systematic approach, based on a com-

mon models and a specification framework. This will offer a common framework, to identify general design and development phases, to modularize the system components, to separate the modeling of various types of knowledge, to define interoperability points with other applications, to reuse subject domains, tutoring and application independent knowledge structures, and finally to achieve more flexibility and consistency within the entire authoring process. Beyond the point of creation of IES, such a common engineering framework will allow for structured analysis and comparison of IES and their easy maintainability.

Currently, a lot of effort is focused on improving of IES authoring tools to simplify the process and allow timeefficient creation of IES [Murray, 2003a; Redfield, 1997; Vassileva, 1995]. Despite this massive effort, there is still no complete integrated methodology that allows to distinguish between the various stages of IES design, and also to (semi-)automate the modeling and engineering of IES components, as well as providing structured guidance and feedback to the author. There are efforts to decrease the level of complexity of ITS building by narrowing down the focus to a set of programming tasks and tools to support them [Anderson et al, 1995], and by limiting the view to only correct or incorrect 'solutions to a set of tasks' [Ritter et al, 2003]. As a way to overcome the complexity without decreasing the level of 'intelligence' in IES, [Ritter et al, 2003] proposes an approach for separation of authoring components, and [Murray, 2003a] offers a KBT-MM a reference model for authoring system of a knowledge-based tutor, which is storing the domain and tutoring knowledge in "modular components that can be combined, visualized and edited in the process of tutor creation".

A considerable amount of the research on knowledgebased and intelligent systems moves towards concepts and ontologies [Mizoguchi et al, 2000] and focuses on knowledge sharing and reusability [Chen et al, 1998; Ikeda et al, 1997]. Ontologies allow the definition of an infrastructure for integrating IES at the knowledge level, independent of particular implementations, thus enabling knowledge sharing [Breuker et al, 1999]. Ontologies can be used as a basis for development of libraries of shareable and reusable knowledge modules [Aroyo et al, 2002b] and help IES authoring tools to move towards semantics-aware environments.

In compliance with the principles given by [Murray, 2003a] we present an integrated framework that allows for a structured approach to IES authoring, as well as for automation of authoring activities. Characteristic aspect of our approach is the definition of different ontology-based IES intelligence components and the definition of their interaction. We finally aim in obtaining an evolutional (self-evolving) authoring system, which will be able to reason over its own behavior and subsequently change it if is necessary. In Section 2 we illustrate the aspects of the IES authoring process. In Section 3 we shortly present the main points of the authoring task ontology, and in Section 4 the architectural considerations of the EASE framework are sketched. Finally, we present our conclusions and intentions for future work.

2 Authoring Support Approach

The approach we take follows up on the efforts to elicit requirements for IES authoring, define a reference model and modularize the architecture of IES authoring tools [Aroyo et al, 2004]. We describe a model-driven design and specification framework that provides functionality to bridge the gap between the author and the authoring system by managing the increased intelligence. It accentuates the separation of concerns between subject domain, user aspects, application and the final presentation of the educational content. It allows to overcome inconsistencies and to automate the authoring tasks. We show how the scheme from [Murray, 2003a] can be filled with the 'entire intelligence of IES', split into collaborative knowledge components.

First, we look at the increased intelligence. Authoring of IES is a process with an exponentially growing complexity and it requires many different types of knowledge and considering various constraints, requirements and educational strategies [Nkambou et al, 1996]. Aiming at (semi)-automated IES authoring we need to have explicit representations of the strategic knowledge (rules, requirements, constraints) in order to be able to reason within different authoring contexts and situations. Managing of the increased intelligence is therefore a key issue in authoring support.

Second, we consider the conceptual distance between the user and the system. According to [Mizoguchi et al, 2000; Redfield, 1997] the authoring tools are neither intelligent nor user-friendly. Special-purpose systems provide extensive guidance, but the disadvantage is that changing such systems is not easy, and the knowledge and content can hardly be reused for their educational purposes [Murray, 2003b]. Thus, structured guidance is needed in this complex authoring process.

Our ultimate aim is to attain seemingly conflicting goals: to define authoring support in a powerful, generic and easy to use way. The power comes from the use of ontologybased approach. The generality is achieved with the help of a meta-authoring tool, instantiated with the concrete learning context to achieve also the power of a domain specific tool. The ease of use comes from the combination of the previous two. A characteristic aspect of our approach is the use of Authoring Task Ontology (ATO) [Aroyo et al, 2002a; Aroyo et al, 2003] as part of the authoring environment, which enables us to build a meta-authoring tool [Aroyo et al, 2004] and to tailor the general architecture to the needs of each individual system.

2.1 IES Authoring

Characteristically, ITS, maintain and work with knowledge of the expert, learner, and tutoring strategies, to capture the student's understanding of the domain and to tailor instructional strategies to the concrete student's needs. Thus, the provision of user-oriented (adapted) instruction and adequate guidance in IES depends on:

- maintaining a model of the domain, describing the structure of the information content within IES (based on concepts and their relationships);
- maintaining a personalized portal to a large collection of well organized and structured learning/teaching material resources.
- maintaining a model of the user to reflect the user's preferences, knowledge, goals, and other relevant instructional aspects;
- maintaining instructional design, assessment, adaptation and sequencing models;

In line with this we structure the complexity of the entire authoring process by grouping various authoring activities to:

- model the subject domain / domain knowledge;
- · maintain and modify learning objects;
- define the learning goals / learning activities;
- apply instructional strategies for individual and group learning;
- apply assessment strategies for individual and group learning;
- specify a learner model / characteristics;
- specify learning sequence(s) out of learning and assessment activities.

To support these authoring tasks we employ knowledge models and capture all the processes related to those tasks in corresponding authoring modules. Our final goal is to realize an evolutional (self-evolving) authoring system, which will be able to reason over its own behavior.

3 Authoring Task Ontology

The authoring task ontology (ATO) serves as a shared vocabulary to describe problem-solving structures of all existing tasks domain-independently [Jin et al, 1997]. It is a meta-level ontology of upper level concepts of the specific IES authoring ontologies. Its role in an authoring environment is to support the verification of the authoring activities and to allow the authoring system to be reusable. The main parts of ATO (described in details previously in [Aroyo et al, 2002a; Aroyo et al, 2003; Aroyo et al, 2004] are:

- basic ATO concepts
- primitive activities
- · authoring tasks

The basic ATO concepts are used in the formulation of the authoring tasks. We build upon the authoring concepts introduced by [Mizoguchi et al, 1997]: (1) generic nouns reflecting the roles of the objects in the authoring process, (2) generic verbs representing authoring activities over the objects, (3) generic adjectives representing the modifications of the objects and (4) other authoring task specific concepts. We extend this set and make it IES domainspecialized.

The primitive activities [Aroyo et al, 2002] in ATO are defined as atomic methods over objects (e.g. domain and course concepts, topics, learning objects, user model and

user profile attributes, cognitive characteristics, learning goal) within a specific structure in the authoring system, such as domain model, user model, user profile, course sequence/structure, or learning goal representation hierarchy. Those primitive activities constitute a basic functional formalism that expresses how the object changes the structure, or the structure is manipulated.

Finally, we define authoring tasks, as a hierarchy of higher-level (composite) functions to represent conceptual categories of relationships (interdependence) between primitive functions. These relationships present certain aggregation criteria (including causal and other relations among components) that are used for grouping primitive tasks into higher-level classes of authoring and system tasks. This way we can construct/identify functional groups of authoring tasks. The higher-level tasks represent a role of one base function for another base function. They are concerned not with the actual change in the objects, but with their actual function in the process of authoring IES. We define those tasks with conditions for their primitive parameters in order to be able to achieve specific authoring goals.

4 EASE Architectural Issues

We take a goal-centered approach to achieve separation of the data (content), the application (educational strategy), the instructional goals and the assessment activities.



Figure 1. EASE Reference Architecture

We follow explicitly the principles supported also by KBT-MM [Murray, 2003a] to separate 'what to teach' into modular units independent of 'how to teach' and to present learning goals separately from the instructional content. Thus, we have a clear distinction between the content and the computational knowledge, where the learning goal plays a connecting role in order to bring them together within the specific context of each IES (Fig. 1). In this way we also allow reusability of general knowledge on instructional design and strategies.

In other words, the Collaborative Learning Strategy (CLS) module provides the author with the appropriate group learning strategies and requirements for them via the Sequence Strategies Authoring (SS) module. To generate explanations and guidance about the recommended strategies CLS uses Collaborative Learning Ontology which is a system of concepts to represent collaborative learning sessions and Collaborative Learning Models inspired by learning theories [Inaba et al, 2000; Supnithi et al, 1999].



Figure 2. Assessment Module Interactions

The core of the intelligence in the EASE architecture comes from the communication or interactions between the components. There are two "central" components here, the Sequencing Strategies Authoring (SS) and the Authoring Interface (AI). The AI is the access point for the author to interact with the underlying concepts, models and content. The SS interacts with the other components in order to achieve the most appropriate learning sequence for the targeted learner (Fig. 2).

At a conceptual level the IES author interacts with the Learning Resources (LR) and with the Domain Model (DM) authoring modules, for example to handle the learning objects. While the author is working with DM, an interaction is required between DM and LR to determine available resources to link to domain concepts. At the user (learner) level the author interacts with the Simulated User Model (SUM) component in order to determine the use of UM (update rules) within the IES application. At the application level the author interacts with the A and SS modules.

The authoring rules in the Assessment knowledge base trigger interaction in order to realize various aspects of the test generation process. An authoring support rule in the CLS's knowledge base, on the other hand, produces recommendations and can be triggered by either the author or the system.

5 Conclusion

The goal of this research is to specify a general authoring framework for content and knowledge engineering for Intelligent Educational Systems (IES). The main added value of this approach is that on the one hand the ontologies in it make the authoring knowledge explicit, which improves the basis for sharing and reusing. On the other hand, it is configurable through an evolutional approach. Finally, this knowledge is implementable, since all higher-level (meta-level) constructs are expressed with a limited class of generic primitives out of lower-level constructs.

Within the EASE framework we have identified the main intelligence components and have illustrated their interaction. Characteristic for EASE is the use of ontologies to provide common vocabulary and common understanding of the entire IES authoring processes. This allows for interoperation between different applications and authors.

Acknowledgements

The work was made possible with the kind support by the Mizoguchi Lab, Osaka University. We would like also to thank Darina Dicheva for her active involvement and the numerous valuable discussions with respect to ATO. We are also grateful to Akiko Inaba and Larisa Soldatova for their essential contributions for the definition of the EASE

architecture, particularly with respect to the collaborative learning and testing aspects of the authoring process.

References

- [Ainsworth et al, 2003] Ainsworth, S., Major, N., Grimshaw, S., Hayes, M., Underwood, J., Williams, B., & Wood, D., REDEEM: Simple Intelligent Tutoring Systems from Usable Tools, In Murray, Ainsworth, & Blessing (eds.), Authoring Tools for Adv. Tech. Learning Env., 205-232.
- [Aroyo et al, 2002a] Aroyo, L. & Dicheva, D., Courseware Authoring Tasks Ontology. International Conference on Computers in Education (ICCE'02), December 2002, Auckland, New Zealand.
- [Aroyo et al, 2002b] Aroyo, L., Dicheva, D., & Cristea, A. (2002). Ontological Support for Web Courseware Authoring. In Proceedings of ITS 2002 Conference, 270-280.
- [Aroyo et al, 2003] Aroyo, L, & Mizoguchi, R., Authoring Support Framework for Intelligent Educational Systems. In Proceedings of AIED 2003 Conference.
- [Aroyo et al, 2004] Aroyo, L. & Mizoguchi, R., Towards Evolutional Authoring Support. *Journal for Interactive Learning Research*. (in print)
- [Anderson et al, 1995] Anderson, J., Corbett, A. Koedinger, K., & Pelletier, R., Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, 4(2), 167-207.
- [Breuker et al, 1999] Breuker, J., Bredeweg, B., Ontological Modelling for Designing Educational Systems, In Proceedings of the Workshop on Ontologies for Intelligent Educational Systems at AIED'99 Conference.
- [Brusilovsky, 2003] Brusilovsky, P., Developing Adaptive Educational Hypermedia Systems. In Murray, Ainsworth, & Blessing (eds.), Authoring Tools for Advanced Technology Learning Environments, Kluwer Academic Publishers, 377-409.
- [Chen et al, 1998] Chen, W. Hayashi, Y., Kin, L. Ikeda, M. and Mizoguchi, R., Ontological Issues in an Intelligent Authoring Tool, In Proceedings of ICCE 1998 Conference, (1), 41-50.
- [Devedzic et al, 2000] Devedzic, V., Jerinic, L., Radovic, D., The GET-BITS Model of Intelligent Tutoring Systems. *Journal of Interactive Learning Research*, 11(3), 411-434.
- [Ikeda et al, 1997] Ikeda, M., Seta, K., Mizoguchi, R., Task Ontology Makes It Easier To Use Authoring Tools. In Proceedings of IJCAI 1997 Conference.
- [Inaba et al, 2000] Inaba, A., Supnithi, T., Ikeda, M., Mizoguchi, R., & Toyoda, J., How Can We Form Effective Collaborative LearnningGroups - Theoretical justification of Opportunistic Group Formation with

ontological engineering. In Proceedings of ITS 2000 Conference, 282-291.

- [Jin et al, 1997] Jin, L., Ikeda, M., Mizoguchi, R. & Kitakaoka, Y., Role Explication of Simulation in Intelligent Training Systems by Training Task Ontology, In Proc. of AI-ED-97 Workshop on Architectures for intelligent simulation-based learning environments, p. 41-48.
- [Mizoguchi et al, 1997] Mizoguchi, R., Ikeda, M., & Sinitsa, K., Roles of Shared Ontology in AI-ED Research: Intelligence, Conceptualization, Standardization, and Reusability. In Proc. of AI-ED 97, pp.537-544, Kobe, Japan
- [Mizoguchi et al, 2000] Mizoguchi, R., & Bourdeau, J., Using Ontological Engineering to Overcome Common AI-ED Problems, *International Journal of AIED*, 11(2), 107-121.
- [Murray, 2003a] Murray, T., Principles for pedagogyoriented Knowledge-based Tutor Authoring Systems. In Murray, Ainsworth, & Blessing (eds.), Authoring Tools for Advanced Technology Learning Environments, Kluwer Academic Publishers, 439–466.
- [Murray, 2003b] Murray, T., An Overview of ITS authoring tools. In Murray, Ainsworth, & Blessing (eds.),

Authoring Tools for Advanced Technology Learning Environments, Kluwer Publishers, 491–544.

- [Nkambou et al, 1996] Nkambou, R., Gauthier, G., Frasson, C., CREAM-Tools: An Authoring Environment for Curriculum and Course Building in an ITS. *Computer Aided Learning and Instruction in Science and Engineering*, 186-194.
- [Redfield, 1997] Redfield, C. L., An ITS Authoring Tools: Experimental Advanced Instructional Design Advisor, AAAI Fall Symposium, 72-82.
- [Ritter et al, 2003] Ritter, S., Blessing, S., & Wheeler, L., Authoring tools for Component-based Learning Environments. In Murray, Ainsworth, & Blessing (eds.), Authoring Tools for Advanced Technology Learning Environments, Kluwer Academic Publishers, 467-489.
- [Supnithi et al, 1999] Supnithi, T., Inaba, A., Ikeda, M., Toyoda, J., & Mizoguchi, R., Learning Goal Ontology Supported by LearningTheories for Opportunistic Group Formation, Proc. of AIED'99, Le Mans France, 67-74.
- [Vassileva, 1995] Vassileva, J., Dynamic Courseware Generation: at the Cross Point of CAL, ITS and Authoring. In Proceedings of ICCE 1995 Conference, 290-297.

Adaptive Support to Elearning, Using Formal Specifications of Applications, Tasks, Content and User Models^{*}

Aude Dufresne¹, Mohamed Rouatbi²

¹ Department of Communication and ²Informatique et Recherche Opérationnelle, ² University of Montreal, C.P. 6128 Succ Centre-Ville, Montreal, H3C 3J7, Canada aude.dufresne@umontreal.ca

Abstract

Some of the first applications of ontologies to support elearning were to generalize adaptive interfaces. Thus ontologies were used to search and suggest among conceptually structured learning objects. More recently ontologies have been extended to describe pedagogical structures in order to base selection on ID principles, and to support course designers in their tasks. But to generalize further, another level of ontology is needed describing applications and the interaction between the user and the system, the model of the interface. This paper presents our work to generalize the ExploraGraph system, in order to offer generic adaptive support. The extraction and specification of formal ontologies for different models makes it possible to describe axioms and boolean reasoning linking task, user, ID and interactions models, in general terms and for different simultaneous applications.

1 Using Ontologies to Support Elearning

Adaptive hypermedia systems were the first types of systems to incorporate annotations to support navigation in elearning environment. This metadata annotation was part of the first ontological information added to content to organize and support learning. As suggested by authors [Brusilovsky, 1996; Dufresne, 1997] the structure of information can used to hide or highlight elements to suggest a progression or to give feedback. In the elearning environment ExploraGraph, the structure of navigation is organized around the structure of concepts, tasks or documents using composition and precedence relations. It can then be dynamically annotated by the learner himself (to change is user's model) or the system to adapt and give feedback - open or highlight elements, propagate user models using the structures of relations. Authors have tried to generalize the principles of adaptive hypermedia by describing the meta-level ontologies of the adaptation in elearning environments. Various <u>architectures</u> were described using those formal classifications and adding different functions to support the learner and also the designers of elearning environments.

Other systems have tried to use ontological relations among metadata associated to learning objects to search and propose elearning content to learners [Mendes and Sacks, 2003]. Whether using manual entered metadata or textual clustering analysis, the system used a learner model and a model of the domain to order dynamically relevance of elements to be proposed to learners.

The same type of system where then proposed to designers of elearning environments. The ontology of elearning was also extended to include models of instructional design theories [Bourdeau, et al., 2004]. Formal specifications of elements of instructional design made it possible to define axioms and reasoning to support further designers of elearning content. Architecture for adaptive elearning environments were defined following this paradigm, as well as toolkits to facilitate course editing with adaptive functions [Ceri, et al., 2004; Cristea and Aroyo, 2002].

Our research stems from an ITS perspective trying to incorporate in elearning environments more interactive support to individual learners. As such, the idea of elearning as a mere collection of elearning elements appears limited for more active or complex learning like those in science. In particular in the context of distant education, activities needed to be supported further [Brusilovsky, 2001; Dufresne, 2000; Dufresne, et al., 2003; Dufresne and Hudon, 2002]. It appears important to support learners along various dimensions, which have little to do with the structure of content, but that could profit from an ontological formalization, for example those related to:

- preferences and cognitive styles in the interaction with elearning environments [Dufresne and Turcotte, 1997] [Bull and McCalla, 2002];
- Competence in using the elearning application;
- Agenda support for orientation, deadlines, synchronous meeting;
- Motivational incentive and feedback; etc.

^{*} This research was supported by NSERC, FCI and VRQ grants.

Also to support scientific activities specialized environments, like virtual labs or conceptual tools, were used for which it was necessary to extend support and feedback, using the ontology of the domain, For example in a virtual labs, the learner must specify an hypothesis with dependent, independent and control variables, etc. Specific axioms can be defined on how this should be done and how it can be supported. Although the ExploraGraph environment [Dufresne, 2000, 2002; Dufresne and Hudon, 2002] made it possible to add dynamic adaptation, support and feedback following learner progression, these adaptations were not defined as an ontology in general terms and independent of the application. Also we had no possibility to extend the support to an external application with its own ontology.



Figure 1. Architecture of the generic and multi-applications advisor

2 Architecture of the Generic Advisor

To make the advisor more generic and independent of applications we are now using XML Schema Definition (XSD), were structure of elearning elements, content, tasks, applications and user models are described in generic terms, compatible with various metadata standards.Other XSD structures were also defined to represent external applications components (forms, control elements, objects, attributes, events, functions). For example a virtual experimentation environment (EXAO) was described, in terms of application, but also functions to be applied for diagnosis. Finally user and group models were also generalized in a XSD, so they could be easily augmented and adapted depending on the content or theoretical models. XSD made it possible to represent models of support at a generic level that could be instantiated to specific applications and tasks [Dufresne, et al. 2003].

Figure 2 gives an example of an XSD structure that represent the task of an experiment in the EXAO environment and of parameters of the user models to be considered.

Once the generic schemas are described, **specific instances** can be created or generated following those XSD. In the case of the EXAO environment, instances are extracted directly from the Visual Basic code (Instances for VarValues). In the case of ExploraGraph, task and concept XML structures are exported from the SQL database were they are kept (structures of concepts or tasks in a course). In both cases instances are translated as XML structures in concordance with the XSD structure.

The generic rule editor can then be used to open those XML descriptions of structures and XML instances to create decision rules to support elearning interactions. Those rules can be made at different levels of generality using boolean modifiers. Rules can be defined using representations from multiple applications thus creating inter-application interactions and reasoning mechanisms. For example, information in the task structure, information in the user's model and actions in the EXAO virtual lab environment can be combined to define support interventions, to adapt the task environment or to modify the user model. Conditions in rules are defined as internal (user model) or external conditions (events) and are associated to sequences of internal (update user models) or external actions (messages, feedback to user, control of navigation in the conceptual structures inside Explora-Graph.

User Task	User model		
<xs:schema targetnamespace<xs:complextype<="" td=""><td colspan="3"><xs:complextype name="UserModel"></xs:complextype></td></xs:schema>	<xs:complextype name="UserModel"></xs:complextype>		
name="ExperienceStruct">	<xs:sequence></xs:sequence>		
<xs:sequence></xs:sequence>	<xs:element name="HelpLevel" type="xs:int"></xs:element>		
<xs:element name="DependantVar" type="VarValues"></xs:element>	<xs:element name="EffortInTask" type="xs:int"></xs:element>		
<xs:element name="IndependentVar" type="VarValues"></xs:element>	<xs:element name="IndepVarIdentif" type="xs:int"></xs:element>		
<xs:element name="Hypothese" type="xs:string"></xs:element>			
<td></td>			
XML Rule produced by the generic rule editor and used by <rule appid<br="" ruleid="CerifyVariable" type="Action Rule"><condition> <object_variable appid="EXAO" objectid="Form.FormHypoth
ObjectAttID=" text"=""></object_variable> <operation name="Diff"></operation> <object_variable appid="ExploraGraph" objectid="Task.ExperienceH
ObjectAttID=" text"=""></object_variable> </condition> <action_group> <external_action appinter="" id="Agent-Hacker-Animation" stat<="" state="" th=""><th>the generic advisor ="ExploraGraph" Multiplicity="False"> hese.TextField.IndepVarTxt" ObjectTypeID="TextField" Iypo.ExperienceSpecific" OjectTypeID="DependantVar" ppID="Agent-Hacker-Animation"> ETER> orrectly identified the dependant variable were UserSpecificModel" ObjectTymeID="UserModel"</th></external_action></action_group></rule>	the generic advisor ="ExploraGraph" Multiplicity="False"> hese.TextField.IndepVarTxt" ObjectTypeID="TextField" Iypo.ExperienceSpecific" OjectTypeID="DependantVar" ppID="Agent-Hacker-Animation"> ETER> orrectly identified the dependant variable were UserSpecificModel" ObjectTymeID="UserModel"		
<pre><objec1_variable add"="" appid="ExploraGe
<FUNCTION Name=" indepvaridentif"="" objectid="Modele.ModeleUsag
ObjectAttID="></objec1_variable><terme t<br="" value="-10"> </terme></pre>	er.UserSpecificModel" ObjectTypeID="UserModel" raph"/> `ype="integer"/>		
<td></td>			

Figure 2. Examples of parts of XSD structures for task and user model and of a rule produced and used to adapt support in the interaction

The generic advisor is an independent application that runs on the client-side (similar to [Ceri, et al., 2004]) for performance reason. It receives from the different applications the XML description of the structure of their objects, and the decision rules created with the generic editor or from ExploraGraph The advisor checks for the validity of the elements he receives and then open channels to listen to activities in the various applications. The advisor reacts dynamically to the user activity and to changes in his user model. His reactions can be external actions in the one of the applications or internal operations on the user model. Thus the combination of the different ontologies and rules makes it possible to define interactions among different applications, support messages, feedback and adaptive functions in the various environments as well as reasoning on the user model.

3 Discussion and Conclusions

Our work with the XSD generic modeling, the generic rule editor and the generic advisor, is similar to [Ceri, et al., 2004] client-based UML-Guide system: it runs on the client-side and use tasks and conceptual structures to diagnosis and express adaptive guiding to a specific learner. As proposed by them it can integrate other personalization parameters and adaptive actions, like feedback, incentives and internal representation updating. The use of generic structures and objects and Boolean modifiers, makes it possible to describe guiding at different level of generality, independent of the specific objects.

In the process of extracting a more formal and generic model of adaptive support, and as we tried to make it independent of our authoring tool, we encountered different tasks which are now in progress:

- Describe ontologies at a more generic level (independent of instances)
- Define mechanisms to extract or export instances from applications following those generic models
- Develop communication interfaces between our elearning authoring tool for the description of conceptual maps of tasks, concepts and support rules, and the generic rule editor in order to add components associated to external applications.
- Develop communication interfaces between applications and the advisor
- Develop the generic repository for XML structures describing the different models, their instances and the rules.
- Define or use a generic repository for the User Models in communication with the different applications (like the MUMS system [Brooks, et al., 2004].
- Define templates of generic adaptive rules which could be specialized for a specific context.

ExploraGraph first aim was to offer a user-friendly application for the description of elearning environments. Trying to integrate adaptive support and especially to develop a more generic and inter-applications system have lead us away from this simplicity. We still have to experiment how easily teachers can use the system to specify support rules. More research might be needed to improve the links between the rule editor and ExploraGraph. Finally now that some models of XSD for tasks, user models and applications have been developed it will be interesting to see how it can incorporate more theory aware models.

References

- [Bourdeau, et al., 2004] Jacqueline Bourdeau, Riichiro Mizoguchi, Valéry Psyché, and Roger Nkambou, Selecting Theories in an Ontology-Based ITS Authoring Environment, presented at ITS'2004, 2004.
- [Brooks, et al., 2004] Christopher Brooks, Mike Winter, Jim Greer, and Gordon McCalla, The Massive User Modelling System (MUMS), presented at ITS'2004, 2004.
- [Brusilovsky, 1996] Peter Brusilovsky, Methods and Techniques of Adaptive Hypermedia, *User Modeling and User-Adapted Interaction*, pages 687-129, 1996.

- [Brusilovsky, 2001] Peter Brusilovsky, Adaptive Hypermedia, User Modeling and User Adapted Interaction, 11:87-101, 2001.
- [Bull and McCalla, 2002] Susan Bull and Gord McCalla, Modelling Cognitive Style in a Peer Help Network, *Instructional Science*, 30(6):497-528, 2002.
- [Ceri, et al., 2004] Stefano Ceri, Peter Dolog, Maristella Matera, and Wolfgang Nejdl, Model-Driven Design of Web Applications with Client-Side Adaptation, presented at ICWE'04, Munich, Germany, 2004.
- [Cristea and Aroyo, 2002] Alexandra Cristea and Lora Aroyo, Adaptive Authoring of Adaptive Educational Hypermedia, in *Adaptive Hypermedia and Adaptive Web-Based Systems*, vol. LNCS 23:47: 122-132 Springer, 2002.
- [Dufresne, 1997] Aude Dufresne, From adaptable to adaptive interface for distance education, presented at Workshop on "Intelligent Educational Systems on the World Wide Web", Artificial Intelligence in Education: Knowledge and Media in Learning Systems, pages 94-98, Kobe, Japan, 1997.
- [Dufresne, 2000] Aude Dufresne, Model of an Adaptive Support Interface for Distance Learning, presented at ITS'2000, pages 334-343, Montréal, 2000.
- [Dufresne, 2002] Aude Dufresne, The ExploraGraph advising system : an ergonomical evaluation of the editor., in *TICE'2002*, pages 299-306, Lyon, France, 2002.
- [Dufresne, et al., 2003] Aude Dufresne, Josianne Basque, Gilbert Paquette, Michel Leonard, Karen Lundgren, and Sandrine Prom Tep, Vers un modèle générique d'assistance aux acteurs du téléapprentissage, *Sciences et Techniques Éducatives*, 10:57-88, 2003.
- [Dufresne and Hudon, 2002] Aude Dufresne and Martin Hudon, Modeling the learner preferences for embodied agents : experimenting with the control of Humor, presented at ITS 2002 Workshop on individual and group modelling methods that help learners understand themselves, pages 43-51, San Sebastian, Spain, 2002.
- [Dufresne and Turcotte, 1997] Aude Dufresne and Sylvie Turcotte, Cognitive style and its implications for navigation strategies, *AIED'97: Knowledge and Media in Learning Systems*, pages 287-293, Kobe, Japan, 1997.
- [Mendes and Sacks, 2003] M. E. S. Mendes and L. Sacks, Dynamic Knowledge Representation for e-Learning Applications, in *In: , Enhancing the Power* of the Internet - Studies in Fuzziness and Soft Computing, L. A. Z. In: M. Nikravesh, B. Azvin and R. Yager, Ed., pages 255-278, Springer, 2003.

A Metadata Editor of Exercise Problems for Intelligent e-Learning

Tsukasa HIRASHIMA

Hiroshima University, 1-4-1 Kagamiyama Higashi-Hiroshima, 739-8527 JAPAN tsukasa@isl.hiroshima-u.ac.jp

Abstract

In this paper, a metadata-authoring method of exercise problems is proposed and a metadata editor with the method is described. In the metadataauthoring method, a new problem is characterized using differences from the basic problem. In the metadata editor, a user can make a new problem by changing the basic problem. The changes represent the differences between the new problem and the basic problem. The new problem is then characterized by the differences from the basic problem. We call the metadata-authoring method "Differential Indexing".

1 Introduction

Our research target of intelligent e-learning is the exercises for letting students master the use of solution methods in mathematics, arithmetic, physics, and so on. Although students are usually taught a solution method using a basic problem, the students who can solve the basic problem cannot always solve exercise problems that can be also solved by the same solution method. The origin of the difficulty of the exercise problems is the differences between the basic problem and the exercise ones. Then the students learn how to deal with the differences through the exercises. Therefore, in such exercises, differences from the basic problem are the most important characteristics of the problems. Based on this consideration, in order to realize adaptive control of the difficulty of problems in e-Learning, we propose a metadata model called "differential indexing" [Hirashima et al, 2002]. Then, we have developed a prototype of a webbased metadata editor with the differential indexing. In the editor, authors make the metadata by re-placing the elements composing the basic problem with the elements composing the exercise problems. Therefore, the authors don't have to know the format of the meta-data. Moreover, the editor can diagnose the metadata semantically. Furthermore, it can acquire knowledge that is used to the diagnosis in the metadata authoring. In addition, "web-based" means that it is possible for a lot of people to use the editor and to share the problem data. These features of the editor will be useful in order to use it in real world.

2 Differential Indexing

The differential indexing is proposed based on MIPS that is a model of problem process of the problem structure [Hirashima et al, 1992]. When the problem structure drawn from a problem is transformed to the problem structure corresponding to the one of the base problem, the solution method applied to the base problem can be also applied to the problem. In the transformation process, the differences are detected. In other words, the differences express the problem solving process of the exercise problems. Therefore, the metadata made by the differential indexing include enough information to solve the exercise problems. By comparing the two metadata, similarities and differences between the two exercise problems are derived. Then, the similarities and differences represent semantic relation between the problems.

In the differential indexing, the following two differences from a base problem can be categorized: (1) an instance difference, and (2) a structure difference. The structure difference is divided as (2a) a structure difference that can be complemented with fact knowledge, and (2b) a structure difference that can be complemented with operational knowledge. In the following section, these differences are described concretely.

3 Metadata Editor

Figure 1 is the interface of the metadata editor. (Currently, only a Japanese version is implemented. The words in the figures were translated into English). The interface is composed of (a) Problem sentences field, (b) Problem-indexing field, and (c) Problem confirmation field. When a user selects a solution method, the base problem of the solution method is given in Problem-indexing field (in this section, a

user means a problem author). In Problem-indexing field, each line corresponds to a basic relation composed of an object, an attribute, and a numerical value. The left-hand side column is the statement of the way to give the value. The "given value" is the value that is given in the problem directly. "Not given (fact)" is the value that is not given in the problem, but can be complemented by fact knowledge. "Not given (operational)" is the value that is not given in the problem, but can be derived by operational knowledge. These basic relations correspond to the problem structure of the base problem. By changing the concepts or the statements, the problem structure of the new problem is generated. The metadata of the new problem are described as the changes.

In Figure 1, the base problem of the crane-turtle method is given in Problem-indexing field. In Problem sentence field, a user then writes sentences of a new problem that can be solved by the same solution method. The sentences are not interpreted by the system but saved as the body of the problem data. The sentences are used as they are when the problem is given to a learner.

📲 problem-making ed	itor			<u>- 0 ×</u>			
main information							
There are 40 pupils in the class.The average score of the test of the pupils is 69.The average score of the test of a boy is 65.The average score of the test of a girl is 75.How many boys are there?How many girls are there?							
	object	attribute		vaľue			
, answer	cranes	number		1			
answer	turtles	number		Problem			
given	a crane	number of legs	2	indexing			
given	a turtle	number of legs	4	field			
given	cranes and turtles	number of legs					
given	cranes and turtles	number of legs					
Problem confirmation field 警告 : ヌプレット ウィンドウ							

Figure 1. The Interface of Metadata Editor

After that, the user replaces concepts in Problem-indexing field with the concepts used in the new problem. Throughout this process, the problem structure that has different instances but has the same structure with the base problem is generated. Before making a problem including structure differences, a user has to make a problem structure that has only the instance differences. Then the statements of the way to give values are changed to make the problem structure that includes structure differences. The editor then requests the user to input the way to complement the value. In the case of "not given (fact)", the editor requests the user to input fact knowledge that has the same form with a basic relation. In the case of "not given (operational)", the editor requests the user to input operational knowledge and additional basic relations. The operational knowledge is the form of the operational relation between the basic relations. The additional basic relations should be included in the problem sentences.

For example, Problem-1 shown in Problem sentence field can be solved by the same solution method with the base problem shown in Problem indexing field. To make the problem structure of the Problem-1, the user has to first make the problem structure that includes only the instance difference. Therefore, a user replaces "the total number of legs of cranes and turtles" with "the total score of pupils" in problem-indexing field. Then, the statement of the way to give the value should be changed. In the problem, there is no total score of a test of pupils, but it can be derived with the total number of pupils and their average score. Therefore, the user should state that the value is "not given (operational)". Then, the user should input an additional basic relation (Object: pupils, Attribute: average score, Value: 69) and an operational knowledge ("average score of pupils" multiplied by "the number of pupils" is "the total score of pupils").

After the user has finished changing the problem structure in the problem-indexing field, the editor provides a user with problem sentences that are generated from the problem structure to help them check the problem structure in the problem confirmation field. Because the editor can solve the problem by applying the solution method to the problem structure, it also provides the user with the explanation of the calculation used to derive the answer as the operation among the basic relations.

References

- [Hirashima et al, 2002] T. Hirashima and Akira Takeuchi. Differential Indexing: A problem-authoring method for computer-based problem practice. In Proceedings of the international e-Learning Conference, pp.1612-1615.
- [Hirashima et al, 1992] T. Hirashima, A. Kashihara and J. Toyoda. Providing Problem Explanation for ITS. In Proceedings of the International Conference on Intelligent Tutoring Systems, pp.76 83.
Automatic Generation of Courseware for Economics Mathematics

Yukari Shirota

Faculty of Economics, Gakushuin University 1-5-1 Mejiro, Toshima-ku, Tokyo 171-8588, Japan yukari.shirota@gakushuin.ac.jp

Abstract

Today, an increasing number of universities use distance learning systems that leverage the World Wide Web. However, teachers developing the corresponding learning materials face a cost problem in that the work requires much time, practice, and devotion on their part. To solve this issue, we have developed a system -- e-Math Interaction Agent -that automatically generates learning materials using Semantic Web technologies, such as XML and XSLT. Knowledge databases containing math formulas and basic economic knowledge form the core mechanism of the system. Given the necessary mathematical problem definition data, the system can automate the target courseware by using these knowledge bases.

1 Introduction

Today, an increasing number of universities use distance learning systems that leverage the World Wide Web. However, teachers developing the corresponding learning materials face a cost problem in that the work requires much time, practice, and devotion on their part. To solve this issue, we have developed a system -- e-Math Interaction Agent -- that automatically generates learning materials using Semantic Web technologies, such as XML and XSLT [Shirota, 2004A and B]. Knowledge databases containing mathematical formulas and basic economic knowledge form the core mechanism of the system. Given the necessary mathematical problem definition data of which size is small, the system can automate the target courseware by using these knowledge bases.

The system differs from existing courseware automation systems in that it features

- (1) interactive dialogues with a virtual character that are pre-programmed into the XSL stylesheets,
- (2) a solution plan and calculations that are automated from a knowledge base of mathematical formulas and economical rules, and

(3) mathematical software that generates the mathematical expressions in MathML format and image files.

My final goal is to formalize a teaching model for a wide range of mathematical problems that includes how to solve the problems and guide students. When teachers use our system, they will be released from tedious XML programming activities and thus able to devote their energies to more creative work.

This paper describes how the e-Math Interaction Agent dynamically automates Web-based materials to be presented interactively on Web browsers. In the next section, we will explain the design principles and a system model of our proposed courseware automation. In Section 3, the developed system architecture will be described. The prototype system that automates learning materials to teach optimization problems in mathematics will be shown. Discussions and conclusions are given in the last section.

2 Automatic Generation Process Model

First, we shall explain our proposed model of automatic generation processes for math problems (See Figure 1). The input data is the definition data of a math problem. Suppose that the mathematical problem is named 'Problem A.' We wish to automatically generate a solution plan specific to Problem A. Namely, the output of the automatic generation process is the learning material specific to Problem A.

The core parts of the process are the "general solution plan model" and "the general model of teacher interaction" that are represented by the two hexagons. The "general solution plan model" describes how to solve a problem of the same or a similar type. The "general model of teacher interaction" defines what a virtual teacher dialogues with a student and how the virtual teacher guides a student. It is also defined for the same or similar types of problems. The followings are typical type names:

- (1) Optimization problem of single variable functions.
- (2) National income determination modeling problem.
- (3) Optimization problem of multivariable functions.
- (4) Constrained optimization problem with Lagrange multipliers.



Figure 1. The automatic generation process of courseware using the general solution plan model.

For each problem type, these two models must be defined in advance by a "system supervisor" who is both a computer expert and a math teaching specialist well-versed in solving the problems and teaching them to students.

Based on the proposed model, we have developed our e-Math Interaction Agent system to automate learning objects. The above-mentioned problem types (1) and (2) have been already developed as our prototype systems. In the next section, automation of problem type (1) courseware will be shown as an example.

3 e-Math Interaction Agent System Architecture

In this section, we shall outline our e-Math Interaction agent system. The Interaction Agent is written using Perl script language and the agent invokes four sub-modules: (1) Inference Engine (Prolog interpreter), (2) Mathematical Software for symbolic mathematical processing, (3) Equation Server to generate a MathML file, and (4) Web-Page Generator, which infers the solution plan and generates the corresponding XML files. These, in turn, are displayed by the Interaction Agent using the XLST stylesheet. As the above-mentioned (2) mathematical software, we now use Maple which is widely used for math education.

We apply our proposed automation methods as a prototype system to solve mathematical optimization problems of single variable. Although many kinds of optimization problems exist, they share the same schema which includes the following steps:

 Determine the quantity to be maximized or minimized and write the equation for it—in words first, if necessary.

- (2) Use the constraints of the problem to write the equation in terms of only one independent variable, and simplify the equation.
- (3) Find the first derivative, set it equal to zero, and solve the equation.
- (4) Test to determine whether critical points are maxima or minima.
- (5) Check for inflection points.
- (6) Answer the question posed in the problem.

Therefore the metadata properties of the problems, such as the given and unknown data, can be easily defined. Our defined attributes in a meta-level description file are as follows: (1) data, (2) unknown, (3) given, (4) relationship, and (5) find. The data schemas in a meta-level description file are explained in [Shirota 2004A]. As these data schemas are available to define all mathematical problems except proof problems, our proposed methods of automating mathematical learning materials have high descriptive power. Using the meta-level description file as the input information, the Web-Page Generator can automatically generate the corresponding XML files.

Figure 2, 3, and 4 show sample screens of the generated Web pages. The target problem there is a maximizing profit problem. As shown in Figure 2, the total revenue function (TR) and the total cost function (TC) are given. The variable "Q" represents a quantity. The student has to set up the profit function using an economical relationship "profit = total revenue – total cost". The relationship is stored in a knowledge database as an economical rule and used to set up the equation.

Then, the system finds the first derivative, sets it equal to zero, and solves the equation. The generated learning materials are shown in Figure 3. The candidate critical points are Q=1 and Q=25. Test to determine which critical point is the maximum, the system takes the second derivative, evaluates the critical points, and checks the signs. As shown in Figure 4, the maximum point is '(25, 6750)'. In addition, the graph can be displayed which is drawn by the mathematical software Maple.

4 Discussion and Conclusions

In this paper, we have described the e-Math Interaction Agent that dynamically automates on-line Web-based materials. For such automation, Semantic Web techniques are effective. In our implementation of the e-Math Interaction Agent, we used XML and XSLT to automate learning objects.

The Interaction Agent is written using Perl script language and the agent invokes four sub-modules: (1) Inference Engine (Prolog interpreter), (2) Mathematical Software for symbolic mathematical processing, (3) Equation Server to generate a MathML file, and (4) Web-Page Generator, which infers the solution plan and generates the corresponding XML files. These, in turn, are displayed by the Interaction Agent using the XLST stylesheet.

We apply our proposed automation methods as a prototype



Figure 2. The generated learning materials that explain the problem of maximizing profits.



Figure 3. The generated learning materials in which the first derivative is calculated, set equal to zero, and solved.



Figure 4. The generated learning materials in which the answer is shown only after the student tests to determine whether the critical points are maxima or minima.

system to solve mathematical optimization problems. Although many kinds of optimization problems exist, they share the same schema. Therefore the metadata properties of the problems, such as the given and unknown data, can be easily defined. Using the meta-level description file as the input information, the Web Page Generator can automatically generate the corresponding XML files. Our defined attributes in a meta-level description file are as follows: (1) data, (2) unknown, (3) given, (4) relationship, and (5) find. As these data schemas are available to define all mathematical problems except proof problems, our proposed methods of automating mathematical learning materials have high descriptive power.

In general, the human resource cost for Web-based learning material development is quite high. It takes teachers much time, practice, and devotion to design and develop learning materials from scratch. The cost is particularly high when teachers try to create learning materials to help students interactively and naturally. However, our proposed automation methods enable any teacher to generate his/her own Web-based sophisticated learning materials. I believe that we can leverage knowledge bases and Semantic Web technologies for most of this work, and thereby largely relinquish it to computers. To justify this belief, we have developed a prototype system that can automate Web-based learning materials. Our proposed automation methods can release mathematics teachers from tedious XML programming, so that they may devote their energies to more creative work.

Acknowledgments

This research is supported in part by the Japanese Ministry of Education, Science, Sports, and Culture under Grant-in-Aid for Scientific Research (C) (2)15606014.

References

- [Shirota, 2004A] Yukari Shirota. Knowledge-Based Automation of Web-Based Learning Materials Using Semantic Web Technologies. In Proceedings of the Second International Conference on Creating, Connecting and Collaborating through Computing (C5), pages 26—33, Kyoto, Japan, January 2004. IEEE Computer Society.
- [Shirota, 2004B] Yukari Shirota. A Metadata Framework for Generating Web-Based Learning Materials. In Proceedings of the 2004 International Symposium on Applications and the Internet (SAINT 2004) Workshops, pages 249--254, Tokyo, Japan, January 2004. IEEE Computer Society.

A Rule Editing Tool with Support for Non-Programmers in an Ontology-Based Intelligent Tutoring System

Eric Wang, Sung Ah Kim, and Yong Se Kim CREative Design & Intelligent Tutoring Systems (CREDITS) Research Center Sungkyunkwan University Suwon 440-746 Korea wang@me.skku.ac.kr, sakim@skku.edu, yskim@me.skku.ac.kr

Abstract

We are developing an ontology-based intelligent tutoring system, in which domain, pedagogical, and tutoring knowledge is represented as ontologies. Inference rules are a key representation of pedagogical knowledge for automated evaluation. However, the use of existing rule-based languages requires programming skill. Rule editing can be made more widely accessible to non-programmers through the development of smarter tool support. Through analysis of inference rules used in an existing intelligent tutoring system, we identify common idioms within rules that are simple to express in natural language, but which vary widely in their complexity of implementation in the Jess rule language. We have developed a prototype rule editing tool in which these idioms are provided as keywords, with automatic translation to complete rules, which simplifies the rule editing process.

1 Introduction

We have previously developed a multimedia application called Visual Reasoning Tutor (VRT) [Hubbard *et al.*, 1996], which uses the *missing view problem* as a mechanism to develop the visual reasoning abilities of design and engineering students, as shown in Figure 1.



Figure 1. Visual Reasoning Tutor

From 2001–2003, we have developed a successor system called Intelligent VRT (IVRT), which embeds VRT within an intelligent tutoring system framework. IVRT uses

manually-encoded learning contents consisting of *skills*, *lessons*, *and problems*, and a simple learner model that records student skill scores and activity history.

2 Ontology-Based Intelligent Tutoring System

Ontologies support knowledge sharing and reuse, by both humans as well as computers [Gruber, 1993]. We have migrated to an ontology-based approach to intelligent tutoring systems, shown in Figure 2. Domain knowledge (learning contents), learning process knowledge, tutoring knowledge, and learner information are formalized as ontologies. Pedagogical knowledge is also represented as inference rules, which are executed at run-time by a separate inference engine. We use the Protégé ontology editor [Protégé, 2004] with OWL plugin, and Jess [Friedmann-Hill, 2003] as the rule inferencing engine, with XSLT conversion from OWL to Jess.



Figure 2. System architecture of ontology-based intelligent tutoring system

3 Pedagogical Rules in IVRT

IVRT's teaching strategy is to show a subset of lessons and problems based on the learner's current skill scores. Items that the learner has already mastered, and items for which the learner is not yet ready, are not shown. As the learner solves problems satisfactorily, the learner's skill scores increase, which causes new lessons and problems to become visible, and previous ones to be hidden.

This high-level strategy is implemented using inference rules. Terms used in these rules are defined in our learning contents ontology for the IVRT domain, as follows:

- A skill has an *activeness* property, which is true or false, and zero or more *required* skills, arranged in a hierarchical structure.
- A lesson or problem has one or more *associated* skills, and has a *visibility* property, which is either true (it is shown to the learner) or false (hidden).
- A global property of *skill satisfaction* is defined by a system predicate, which takes a skill and returns true or false. Each teacher can customize this test.

Selected rules are shown below. These rules are given in a quasi-formal manner using natural language, which was to ease discussion of the rules without requiring Jess expertise.

Rules to Activate Skills

- 1. For each skill: if it has no required skills, activate it.
- 2a. For each skill: if **any** required skill is not satisfied, deactivate this skill.
- 2b. For each skill: if **all** required skills are satisfied, activate this skill.

Rules to Show and Hide Lessons

- 3a (Default strategy) For each lesson: if **all** associated skills are active, show it.
- 3b (Default strategy) For each lesson: if **any** associated skill is inactive, hide it.
- 4 (Special strategy) For each lesson: if **any** associated skill is active, show it.

Rules to Show and Hide Problems

- 5a For each problem: if **all** associated skills are active, show it.
- 5b For each problem: if **any** associated skill is inactive, hide it.

3.1 Implementation 1: Rules in Jess

A full implementation of IVRT's strategy is straightforward in a standard rule language such as Jess, totaling about 20 Jess rules. However, this requires a Jess programmer's skill. This tends to exclude any users who are not skilled Jess programmers. We assume most teachers lack sufficient programming skill to rely entirely on this approach.

3.2 Implementation 2: SWRL Ontology using Protégé OWL Interface

We have modeled a subset of SWRL rule syntax [Horrocks *et al.*, 2004] as an OWL ontology in Protégé, by defining SWRL terms as classes, and SWRL grammar rules as properties of these classes. Then we were able to use Protégé OWL's user interface to construct instances of SWRL rules, i.e. as a rudimentary rule editor. We integrated this interface with Jess using XSLT conversion, so that SWRL rule instances edited in Protégé are immediately updated in the Jess run-time environment.

This approach was successful insofar as it gave us a rule editing capability. However, it faced two severe drawbacks:

- 1. *Verbosity.* A SWRL rule naturally has a hierarchical structure. To instantiate such a rule as an instance of an ontology required instantiating every element and subexpression separately, in a bottom-up manner. This was a tedious process, even for trivial rules.
- 2. *Programming skill*. To create rules that would work properly after conversion to Jess still required expertise in Jess. Hence, we judged this approach to be no simpler than programming in Jess directly.

4 A Rule Editing Tool with Support for Non-Programmers

We have identified as a desideratum within our ontologybased intelligent tutoring system environment to make rule editing *accessible to non-programmers*. That is, it should provide intelligent support to hide or reduce the complexity of programming.

4.1 Identification of Common Idioms

We considered the actual rules used in IVRT, and also plausible rules within a typical teaching strategy, i.e. which could reasonably be expected to be reused by many teachers across many domains. When these rules are written at a fairly abstract level, using natural language, certain idioms emerged. These idioms correspond to everyday concepts in natural language, on which most people can agree at a nontechnical level.

Two common idioms used throughout IVRT's rules are "if any" and "if all", as highlighted in bold font in Section 3. When implemented in Jess rules, they require substantially different techniques, due to Jess's own characteristics.

4.2 Mapping of Idioms to Rule Fragments

For each idiom, we define a mapping to a Jess rule fragment, which is a portion of a Jess rule. A rule fragment could be as simple as a single keyword¹ in the language. More generally, it consists of a block structure within a rule, and it may introduce variables, or even multiple rules.

We have identified the following idiom-to-fragment mappings. For each mapping, we show the idiom in quasiformal natural language in italics, followed by its Jess rule fragment. Unimportant details of Jess rule syntax are shown in gray text.

• for each s in a set S (defrule R1 (S ?s) => ...)

"for each": This idiom expresses a simple iteration over a set S of facts. As this is a fundamental operation in any rule language, Jess performs this iteration implicitly, without

¹ This correspondence reflects the fact that the design of a programming language itself involves a choice among a range of possible programming idioms, and the idioms chosen will thereafter be trivial to use within that language, by design.

requiring any language keyword. Hence, it suffices to just specify the set itself, using one pattern (S ?s), where ?s denotes a Jess variable.

• for each s in set S that satisfies property P (defrule R2 (S ?s) (P ?s) => ...)

To restrict this iteration to a subset of S that satisfies an additional property P, we simply add a second pattern for P.

```
• if any s in set S satisfies property Q
(defrule R3 (S ?s) (exists (Q ?s)) => ...)
```

"if any": This idiom differs from *"for each"* in that we don't need to visit each element that satisfies the property. Instead, we halt the iteration as soon as any one succeeds. This idiom maps to the Jess keyword **exists**.

```
if all s in set S satisfy property Q
(defrule R4
; Let all other rules go first before checking the 'not'
(declare (salience -1))
; Guard (to ensure volatility, and to exclude empty set)
(and (S ?s1) (Q ?s1))
; All
(not (and
  (S ?s2) (~Q ?s2) ; ~Q is the negation of property Q
))
=> ...)
```

"if all": This idiom is also an iteration over a set. However, by its nature, it *must* visit every element of the set. Jess does not implicitly handle this operation. We apply De Morgan's law to convert the *all* idiom to *not any*, which Jess does provide as primitives. Hence, this idiom expands to a block structure, using the Jess **not** and **and** keywords.

Jess's **not** keyword introduces three complications.

- 1. *Volatility.* The pattern immediately preceding a **not** must be volatile: the rule will be re-checked only when that pattern changes. We handle this by adding a guard clause before the *all* block.
- 2. *Empty set.* **not and** gives a false positive for an empty set. The guard clause also prevents this.
- 3. *Temporal dependency.* **not** assumes that all other rules have reached a quiescent state (i.e. are no longer changing any facts). This requires temporal ordering among rules, which maps to a Jess **salience** declaration.

4.3 Simplified Syntax with Idioms as Keywords

We have defined a simplified rule syntax, in which the common idioms appear as keywords. This supports a non-programmer who thinks at the level of the natural-language idioms, by hiding their implementation details. As this rule format is essentially text-based, any text editor would be sufficient in theory. For added convenience, we have developed a graphical front-end using Java Swing, shown in Figure 3, which allows selection of the idiom keywords from menus.



Figure 3. Rule editor interface with conversion to Jess

Rules written in this simplified syntax are mapped into Jess rule fragments, using a standard recursive descent parsing approach. The fragments are then merged into complete Jess rules, and are immediately evaluated, which updates Jess's run-time environment.

5 Pedagogical Knowledge Editing in ITS Development

We are integrating the rule editing capability into a distributed, persistent ITS development framework. In this framework, the ontologies serve as repositories for many learning domains, tutoring strategies, and bodies of pedagogical knowledge, accumulated over time and across many individual teachers and courses. Rule editing is then a subtask of the more general operation of *pedagogical knowledge editing*. Each individual teacher composes her own tutoring strategy model from the tutoring ontology and a library of previously-developed inference rules, which exploits knowledge reuse. A teacher can customize her model by defining her own pedagogical knowledge as new inference rules. We support local extensions to a particular model, as well as extension of the ontologies themselves by using an ontology editor.

The ontologies and tutoring strategy models are accessible over the web, using standard web services. For distributed rule editing, we are exploring the further enhancement of the rule editing tool as an embedded Java applet, or a separate web-enabled application.

6 Rule Editing for Contents Presentation Design

We are developing an intelligent learning environment targeting heritage education, using ontology-based learner modeling to customize and refine the learning interaction [Kim *et al.*, 2004]. To support this, we have developed a learner ontology based on [Chen & Mizoguchi, 1999]'s approach. A learner is modeled with profile information containing personal data, comprehensive assessment of learner's capabilities, dynamic assessment of learner's

current mood and knowledge, low-level activity records for every learner action and system activity, and processed data obtained from the activity records. In addition, our learner ontology incorporates multiple sets of learner preferences, including Myers-Briggs Type Indicators, Felder & Silverman's Index of Learning Styles [Felder, 2002], and Chen & Mizoguchi's learning preferences. Learner models are inferred from the learners' records of interaction with the system, using data mining.

Felder & Silverman's Index of Learning Styles classifies a learner along 4 axes, with two extremes per axis: (S)ensory–I(n)tuitive, (V)isual–(A)uditory, A(c)tive–(R)eflective, and Se(q)uential–(G)lobal. We abbreviate each value to 1 letter, denoted by the parentheses. Each of the 16 combinations defines a set of learning preferences, which determines the best way in which learning contents should be presented to those students.

For a teacher, this becomes a task of *contents presentation design*. Specific learning content objects are annotated with properties defining the learning styles in which they are to be used. Using the rule editor, the teacher then defines the contents presentation knowledge as inference rules, which take a student's Felder & Silverman learning style as input, and selectively enables and disables learning content objects, and also adjusts their sizes, colors, positions, etc. Examples of the customized contents presentation for two combinations NARG and SVCQ are shown in

Figure 4.





Figure 4. Contents Presentation Design using Felder & Silverman's Index of Learning Styles

A future extension of this work is to deduce these inference rules automatically from visual exemplars of the learning contents presentation, which further simplifies the teacher's task.

7 Conclusion

Inference rules are a significant form of knowledge representation for pedagogical knowledge within an ontology-based ITS, which accurately records a teacher's intent, while supporting automatic execution. We have identified a desideratum to make pedagogical rule editing accessible to non-technical users. To support this, we have identified common rule idioms at the natural language level, and developed a mapping technique from these idioms into complete Jess rules. This supports a simplified rule syntax in which the idioms appear as keywords, hiding the complexities of their implementation.

References

- [Chen & Mizoguchi, 1999] Chen, W., and Mizoguchi, R. Communication Content Ontology for Learner Model Agent in Multi-agent Architecture. *Proc. 7th Int'l. Conf. on Computers in Education*, pp. 95–102, Chiba, Japan.
- [Felder, 2002] Felder, R. Learning and teaching styles in engineering education. *Engineering Education*, **78**(7), pp. 674–681.
- [Friedmann-Hill, 2003] Friedmann-Hill, E. Jess in Action. Manning, Greenwich.
- [Gruber, 1993] Gruber, T. R. Toward principles for the design of ontologies used for knowledge sharing. In N. Guarino and R. Poli, (Eds.), *Int'l. Workshop on Formal Ontology*, Padova, Italy, 1993. Revised version also published in *Int'l. Journal of Human-Computer Studies*, 43(5-6), pp. 907–928, Nov./Dec. 1995.
- [Horrocks et al., 2004] Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosof, B., and Dean, M. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C Member Submission, <u>www.w3.org/</u> <u>Submission/2004/SUBM-SWRL-20040521/</u>, May 2004.
- [Hubbard et al., 1996] Hubbard, C., Mengshoel, O. J., Moon, C., and Kim, Y. S. Multimedia Instructional Software for Visual Reasoning: Visual Reasoning Tutor (VRT). Proc. Int'l. Conf. on Multimedia Computing and System, pp. 261–268, Hiroshima, Japan, Jun. 1996.
- [Kim et al., 2004] Kim, Y. S., Kim, S. A., Wang, E., Park, B. J., Jeon, K. J., and Cho, Y. J. An Intelligent Learning Environment for Heritage Alive. To appear in *Proc. Conf. of KSPE* (special session on intelligent HCI), Busan, Korea, Oct. 2004 [in Korean].

[Protégé, 2004] www-protege.stanford.edu.