Building Blocks for a Smart Space for Learning™

Bernd Simon[*#)], Stefan Sobernig[*)], Fridolin Wild[*)], Sandra Aguirre[+)], Stefan Brantner[$)], Peter Dolog[†)], Gustaf Neumann[*)], Gernot Huber[*)] Tomaz Klobucar[&)], Sascha Markus[±)], Zoltán Miklós[*)], Wolfgang Nejdl[†)], Daniel Olmedilla[†)], Joaquin Salvachua[+)], Michael Sintek[§)], Thomas Zillinger[$)]

*) Vienna University of Economics and Business Administration, Vienna, Austria
#) Knowledge Markets Consulting GmbH, Vienna, Austria
+) Universidad Politécnica de Madrid, Madrid, Spain
$) BearingPoint Infonova GmbH, Graz, Austria
†) L3S Research Center and Hanover University, Hannover, Germany
±) Information Multimedia Communication (IMC), Saarbrücken, Germany
&) Institute "Jozef Stefan", Ljubljana, Slovenija
§) DFKI, Kaiserslautern, Germany

E-mail: elena-sqi@agws.dit.upm.es

Abstract

This case study summarizes the demonstration of a semantic network of interoperable educational systems referred to as Smart Space for Learning™. We started connecting several educational nodes in projects such as Elena, Prolearn, and Icamp. Integration was achieved by using the interaction standard SQI, common schemas for querying and results presentation, and query exchange language, e.g. QEL.

The paper particularly focuses on how heterogeneous nodes can be made interoperable by reusing generalizations of mediating components –building blocks for a Smart Space for Learning™.

1. Introduction

Carrying out a search constitutes a typical process for everyone involved in teaching and learning. For example, pupils complement things they have learned in school by looking for learning resources when working on a home assignment. Teachers aim at enriching their courses with external resources in order to reduce their own workload through re-use while increasing the effectiveness of their course at the same time. Employees responsible for personnel development try to find a cost-efficient but effective learning arrangement that offers the possibility of filling a particular competency gap.

The Web puts a huge number of resources within reach of anyone with Internet access. However, many valuable resources are difficult to find, because they are hidden in the closed and proprietary worlds of brokerage platforms, learning (content) management systems, streaming media servers and online collaboration tools. These systems - commonly referred to as learning object repositories - hold information on learning resources such as courses, online tutorials, lecture notes, electronic textbooks, tutoring sessions, quizzes, etc.

Learning object repositories lacking interoperability create three major drawbacks for its users [14]:

(1) The number of learning resources accessible is limited to those held in the local repository. Teachers, learners, and learning managers might miss the latest critical developments due to the fact that the existence of an external learning object is hidden from them.

(2) The way users can search for learning resources is restricted by proprietary semantics or by document-centered interfaces offered in today's search engines. While proprietary semantics can lead to effective search on the local repository, the lack of semantic alignment with the outside world reduces usability. On the other hand general search engines hardly provide means for effective searches. For example, it is difficult to satisfy a search such as 'List me all courses that improve my negotiation skills and are offered in Brussels between 15th of October and 30th of November and that are frequently visited by employees of the pharmaceutical industry' using state-of-the-art technology.

(3) Context information on the learning resource is limited to the local environment herby reducing the possibilities for deploying effective recommendation mechanisms. The evolution of Web search engines has largely benefited from re-using context information [18], e.g. by investigating the hyperlink structure of Web documents (‘who links to who?’). This kind of information is not accessible for closed educational systems, which is especially a drawback for making a highly complex decision such as selecting the right learning resource. The context of learning resources – especially information on its usage (‘who has successfully used which learning resource?’) – provides highly valuable information that is hidden in the closed and proprietary worlds of today’s repositories.

Beyond the individual level, institutions are increasingly demanding interoperability in the educational domain in order to realize scenarios, such as exchanging learner performance records, integrating course descriptions and evaluations from heterogeneous sources, and delivering learning using a multi-application learning environment.

The remainder of this paper first explores various interoperability-related aspects of specifying and implementing a Smart Space for Learning™. We demonstrate the use of the Simple Query Interface [14] as one part of the solution for making educational systems interoperable and exploring the Hidden (Educational) Web [11]. The subsequent section discusses our achievements in terms of identifying building blocks for interoperability. The paper concludes with a brief discussion on future work in the area.

2. Smart Spaces for Learning™

Our interoperability case study is realized in the context of personnel development. In order to achieve interoperability among heterogeneous systems for corporate personnel development, we adopted a novel search and retrieval infrastructure for learning resources referred to as Smart Space for Learning™. This infrastructure allows for the integration of heterogeneous educational nodes in a semantic network and provides ‘smart’ access technology for it. Integrated into a workflow support for goal-driven learning, personalized search, and feedback tools, the educational semantic network (the ‘space’) plays a crucial role for supporting corporate personnel development.

A driving force for the design of our Smart Space for Learning™ was an extensive requirements analysis emphasizing the need for integrating divergent resource providers into one common access tool. Hereby, potential learners are not depending on offers by a particular provider or are restricted to a specific learning format, e.g., a costly classroom-based course. On the contrary, they can expand their search to several types of learning formats and providers, e.g., books available from the Amazon Media Store.

Another driving force for setting up a Smart Space for Learning™ is provider’s interest in disseminating their educational services and products via electronic distribution channels. We identified a trend that these providers increasingly offer electronic distribution channels in machine-processable formats such as XML, while technology leaders even offer Web Services that allow for querying.

The infrastructure underlying a Smart Space for Learning™ builds on three corner stones:

The *Simple Query Interface* (SQI, see [14]) as interaction standard for both providing and consuming learning resources’ metadata within SQI-compliant networks. We opted for a service-oriented architecture and therefore make use of the SOAP Web Service Bindings for SQI. Figure 1 illustrates an exchange process, where Learning Repository A (the source) submits a query to Learning Repository B (the target). Metadata are stored using different means, such as file-based repositories, (possibly distributed) relational databases, XML databases, or RDF repositories. It is assumed that both systems have agreed upon a common query language beforehand. The concepts used in the query statement are part of a common (query) schema. At Repository B, the interface component might need to transfer the query from the common query language to the local one. Also some mappings from the common to the proprietary schema might be required before submitting the search. This task is performed by a wrapper component. Once the search has yielded results, the results set is forwarded to the source, formatted according to a common results format.

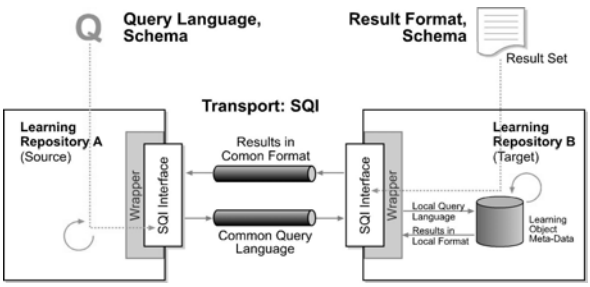

Figure 1. Communication between two Repositories

Queries are issued against a *common metadata schema*, which also informs results presentation. The Elena Common Schema, for example, is realized as an application profile that enriches the element battery of the IEEE Learning Object Metadata (LOM) Standard by mixing in other metadata standards [12]. Elena employs RDF bindings for the common schema. Users' queries against this schema are expressed in the Query Exchange Language (QEL, see [9]), initially devised as a building block of Edutella [8], and mappings are provided to bridge common and local schemas [10, 7].

As third corner stone, we provide *re-usable software components* for integrating existing systems with minimum effort. They are applied for aligning the common and various local representation models (representation management and matching) and for mediating between query languages involved (query transformation).

3. Building Blocks

So far, we have connected a number of providers to our network that can all be accessed from the EducaNext Portal for Exchanging Learning Resources (see http://www.educanext.org/networksearch/) and from the personnel development portal Human Capital Development (HCD) Online.

We have identified the following integration cases which seem to form recurring patterns for developing an integrated service-oriented architecture in the educational domain: (Web) Service Gateway, RDBMS-based Integration, RDF and XML File Harvesting, XML Database Integration, HTTP Proxy. The different types of integration cases are exemplified below.

Gateway: Amazon.com (http://www.amazon.com/), the leading online book-store, offers a proprietary application programming interface (API) that was wrapped by an SQI gateway. Amazon's E-Commerce Service (AECS) is built upon Web Service technology. The search format is determined in the API. Incoming QEL queries are translated into the Amazon search format and results are transformed back to a QEL response format. A caching mechanism accelerates the search for recurring search requests.

Metadata is not directly expressed in known learning related metadata standards. The search interface Amazon offers is not a query interface for their database, but provides similar options as the advanced search on their web site. Searchers may use a rich combination of the logical operators (AND, OR, NOT), however, these have Amazon specific semantics. The gateway makes use of the SQI libraries originally stemming from the Edutella wrapper. Metadata quality is generally high. A direct, hard-coded schema mapping is applied.

RDBMS-based Integration: Clix is a leading commercial learning management system developed by the IM-C AG. The SQI implementation and realization of metadata integration resulted in a wrapper-like architecture tightly integrated into the Java component hierarchy of Clix, but running as stand-alone application. Clix uses a relational database to store its metadata. Metadata schemata are freely configurable within Clix and are not restricted to specific standards. Therefore, the metadata schema may vary across different customized Clix instances. Usually, a basic metadata stock exists. However, only the title is mandatory. The mappings are developed for the particular metadata schema and have to be adopted for different local metadata schemes. The heavy use of free-text information makes the development of the mappings difficult. In Clix, different rights can be assigned to users. Users with appropriate rights may update or enrich metadata. Clix offers a versioning system and process management support, thus enabling metadata quality management. The relational databases uses SQL as local query language, the SQI target uses the QEL-to-SQL translation component created for the Edutella integration case. The wrapper development incorporates experiences gained from interfaces to other systems (Peoplesoft/SAP) via JMS and from import facilities for SCORM and AICC resources.

RDF and XML File Harvesting: ULI (Universitärer Lehrverbund Informatik) is a cooperative project of universities funded within the Programme for Investment in the Future by the German Federal Ministry of Education and Research. ULI tries to establish an exchange of course material, courses and certificates in the area of computer science. Eleven German universities with eighteen different professors have agreed to exchange their courses and to allow students from one university to attend courses at another university, using advanced e-learning technologies.

The data from the providers within ULI are merged into a single, centralized RDF-file. The SQI target is a file-based provider with static content, re-using the Java Edutella classes. To update the storage, the RDF-file needs to be exchanged, thus avoiding complex replication mechanisms.

ULI uses specific metadata subsets of Dublin Core (DC) and LOM, dc:subject classifications are provided with the hierarchical sub- and super-topics of the ACM Computing Classification System. Overall the quality of metadata is high. Queries are accepted in QEL and are then transformed to the local query language RDQL (RDF Data Query Language (RDQL) [13]. Results are returned as QEL result sets encoded in RDF/XML, whereby conformity with the common schema is ensured. More details on the applied mapping technique can be found in [10].

XML Database Integration: KnowledgeBay is a German portal, offering several hundred recorded audio and video lectures. In the case of KnowledgeBay an SQI target for an XML database is provided. The XML database serves as a cache for data from the underlying relational databases. Queries are answered based on the data stored in this XML database. Data from the legacy systems is replicated as complete copy, thus avoiding synchronization. The regular and automatic update of the database content is not fully realized at the moment: the relevant software is developed; however, organizational issues still have to be resolved. The Open Source native XML data base eXist [2] is used. The XQuark Bridge [3] is facilitated to transform the data stored in the relational database to XML. The QEL parser from the Edutella implementation was re-used. Furthermore, a re-usable generic QEL-2-XQuery mapping component was developed from this case [7].

HTTP Proxy: Seminarshop.com is a large content pool of courses, federating about 650 training providers with approximately 20.000 resources in Austria, Switzerland, and Germany. Seminarshop.com primarily addresses company owners, entrepreneurs, managers, and executives. The SQI implementation in Java is a wrapper providing a protocol gateway from SQI to plain HTTP, yielding XML results compliant to the common schema, i.e. a simple transformation into an RDF/XML encoded QEL result-set is applied. Incoming queries are issued live to the Seminarshop's local MySQL database via a restricted HTTP interface requiring authentication. The conventional SQI libraries were re-used. A generic, pattern-based QEL parsing framework was developed for this case which deconstructs QEL queries into generic composite object structures that can be converted easily.

Seminarshop uses a proprietary, non-standard-compliant metadata schema of adequate quality regarding completeness. Nevertheless, the metadata storage is managed decentrally by each of Seminarshop's clients through metadata feeds, so the quality of the entries is heterogeneous (ambiguity of entries, diverging interpretation of metadata attributes, duplicate values). Incoming queries are translated to parameterized HTTP requests, allowing phrase search and field-specific value restrictions. The wrapper itself realizes simple conjunctions and disjunctions based on query plans and duplicate elimination. Seminarshop returns results in an XML format, which is converted with a simple transformation into an RDF/XML encoded QEL result set.

4. Discussion & Outlook

This paper proposes SQI as one component of a universal interoperability layer for building a Smart Space for Learning™. Beyond communication, a representation pillar needs to be realized by providing a common schema and a corresponding common query language. From these experiences, we identified the following issues that require further investigation.

4.1 User requirements for search

We have learnt from a preliminary service quality evaluation that users appreciate most the possibility of being able to specify a search on a very detailed level. Attributes of learning activities like location, time period, topic, price or particular industries are of outstanding importance when providing a high quality search tool.

4.2 (Meta-)Data quality

The higher the complexity of a common schema, the higher the demands for management of data quality becomes. Smart Spaces for Learning™ currently face a trade-off situation between rolling out a highly functional, but complex common schema and opting for a simple common schema which is easier to populate but hardly provides any added value to standard search technology such as Google. Therefore, we consider both user interaction for validating plus enhancing metadata and transparent metadata (repository) quality measurements as crucial.

4.3 Automated Metadata Mining

We opted for a (semi-) automated approach for creating additional metadata [5] by introducing a special service called Semantic Annotator. The goal of the Semantic Annotator is the (semi-) automatic annotation of learning resources by analysing their description with the help of topic ontologies. We consider the inclusion of semantic annotators as important for enriching existing metadata. However, further research is needed to enhance the techniques deployed so far.

4.4 Coupling result processing

OpenSearch primarily focuses on providing syndication of search results retrieved [1]. In the context of SQI, it is important to find means for leveraging ranking mechanisms to the configuration of the actual lookup. This allows for reducing network load, since metadata that is anticipated not to be of high relevance to the searcher can be omitted.

4.5 Refocusing interoperability

SQI helps reducing complexity with respect to interaction in a heterogeneous network of repository architectures. However, the underlying information integration architecture based on a common schema and various wrappers between the global and local sphere introduces another level of complexity: A network lookup currently involves a considerable overhead in terms of transformations of search requests. Initial requests that are formulated by the user and expanded automatically according to the users' profile are first transformed into equivalent expressions of a highly powerful formal query language, e.g. Elena's common query language QEL. This query is then distributed across the SQI-based network. Overhead and perceived complexity finally materialize when the global query is re-transformed into a local query or search expression, in most cases less complex and less powerful than the global one due to the representation models supported locally. The overhead in mediating between constraint expressive power at the local and extensive expressive capabilities at the global level can only partly be compensated by reusable service components. Affiliated initiatives that make use of SQI in their settings, such as Fire [6], opted for reduced global querying facilities based on keyword bags and basic Boolean connectives. This approach, however, risks re-introducing major drawbacks associated with heavily reduced expressive power, e.g. by satisfying aforementioned users' demand for being able to sufficiently restrict queries.

At the level of query languages, design requirements for more generic query languages for web-based search and retrieval were recently reintroduced into the discussion [4]. As a necessary complement, the development of query language derivates, that adapt to the problem domain of teaching and learning, is needed.

5. Acknowledgements

This work is supported by European Commission via the IST projects Elena, iCamp, and Prolearn.

6. References

[1] OpenSearch: http://opensearch.a9.com/, 2005.

[2] eXist. http://exist.sourceforge.net/, 2006.

[3] XQuark. http://xquare.objectweb.org/, 2006.

[4] F. Bry, C. Koch, T. Furche, S. Schaffert, L. Badea, and S. Berger, "Querying the Web Reconsidered: Design Principles for Versatile Web Query Languages.," International Journal on Web Information Systems vol. 1, pp. 1-21, 2005.

[5] K. Cardinaels, M. Meire, and E. Duval, "Automating Metadata Generation: the Simple Indexing Interface," in Proceedings of the 14th World Wide Web Conference. Chiba, Japan, 2005, pp. 548-556.

[6] D. Massart, How to light FIRE: http://fire.eun.org/howto.pdf, 2005.

[7] Z. Miklós and S. Sobernig, "Query Translation between RDF and XML - A Case Study in the Educational Domain," in WWW Workshop on Interoperability of Web-Based Educational Systems, D. Olmedilla, N. Saito, and B. Simon, Eds. Chiba, Japan: CEUR, 2005, pp. 27-35.

[8] W. Nejdl, B. Wolf, C. Qu, S. Decker, M. Sintek, N. Ambjörn, M. Nilsson, M. Palmer, and T. Risch, "EDUTELLA: A P2P Networking Infrastructure Based on RDF," in Proceedings of the 11th International World Wide Web Conference. Hawaii, USA, 2002.

[9] M. Nilsson and W. Siberski, RDF Query Exchange Language (QEL) - Concepts, Semantics and RDF Syntax: http://edutella.jxta.org/spec/qel.html, 2003.

[10] D. Olmedilla and M. Palmèr, "Interoperability for Peer-to-Peer Networks: Opening P2P to the rest of the World," in Workshop on Interoperability of Web-Based Educational Systems, D. Olmedilla, N. Saito, and B. Simon, Eds. Chiba, Japan, 2005.

[11] S. Raghavan and H. Garcia-Molina, "Crawling the Hidden Web," in Proceedings of the Twenty-seventh International Conference on Very Large Databases, 2001.

[12] G. Rivera, B. Simon, J. Quemada, and J. Salvachua, "Improving LOM-Based Interoperability of Learning Repositories," in Proceedings of the OTM 2004 Workshop on Ontologies, Semantics and E-learning, October, 2004, E. Duval and P. Spyns, Eds.: Springer LNCS (3292), 2004.

[13] A. Seaborne, RDQL - A Query Language for RDF: W3C Member Submission 9 January 2004, http://www.w3.org/Submission/RDQL/, 2003.

[14] F. van Assche, E. Duval, D. Massart, D. Olmedilla, B. Simon, S. Sobernig, S. Ternier, and F. Wild, "Spinning Interoperable Applications for Teaching & Learning using the Simple Query Interface," Educational Technology & Society (ISSN 1436-4522), vol. 9, 2006.