Integrating Schema-specific Native XML Repositories into a RDF-based E-Learning P2P Network

Changtao Qu, Wolfgang Nejdl, and Holger Schinzel

Learning Lab Lower Saxony
University of Hannover, Germany

The 2nd International Conference on Dublin Core and Metadata Applications (DC 2002)
Florence, Italy, Oct. 13-17, 2002

Overview

- Edutella: a RDF-based E-Learning P2P (Peer-to-Peer) network
  - General architecture
- Integrating native XML repositories into Edutella: technical challenges
  - Incomparability between XML data model and Edutella Common Data Model (ECDM)
  - Incomparability between XPath (XML query language) and Datalog (ECDM's query language)
  - Query result processing
- Integration approach for schema-specific native XML repositories

  (DCMES RDF/XML binding, LOM/IMS XML binding, SCORM XML binding)
  - Task 1/3: Generate ECDM-based common data view of the native XML repositories
  - Task 2/3: Translate Datalog query into XPath query
  - Task 3/3: Transform local query results into Edutella common result exchange format
- Conclusions

Edutella: General Introduction

- Open source project (http://edutella.jxta.org)
  - Accommodate heterogeneous learning resource metadata repositories in a P2P manner
  - Facilitate the exchange of learning resource metadata between these repositories based on RDF
- Principal contributors
  - *Germany*: U. of Hannover; U. of Braunschweig; U. of Karlsruhe; IWF Göttingen; Tech. U. of Darmstadt; Fraunhofer ISST
  - *Sweden:* Royal Inst. of Technology, Stockholm; U. of Uppsala
  - *Austria:* Business U. of Vienna
  - *USA:* Stanford U.
- Voluntary contributors
  - About 30 software engineers coming from all over the world (until Oct. 1, 2002)

New Edutella contributors are welcome!

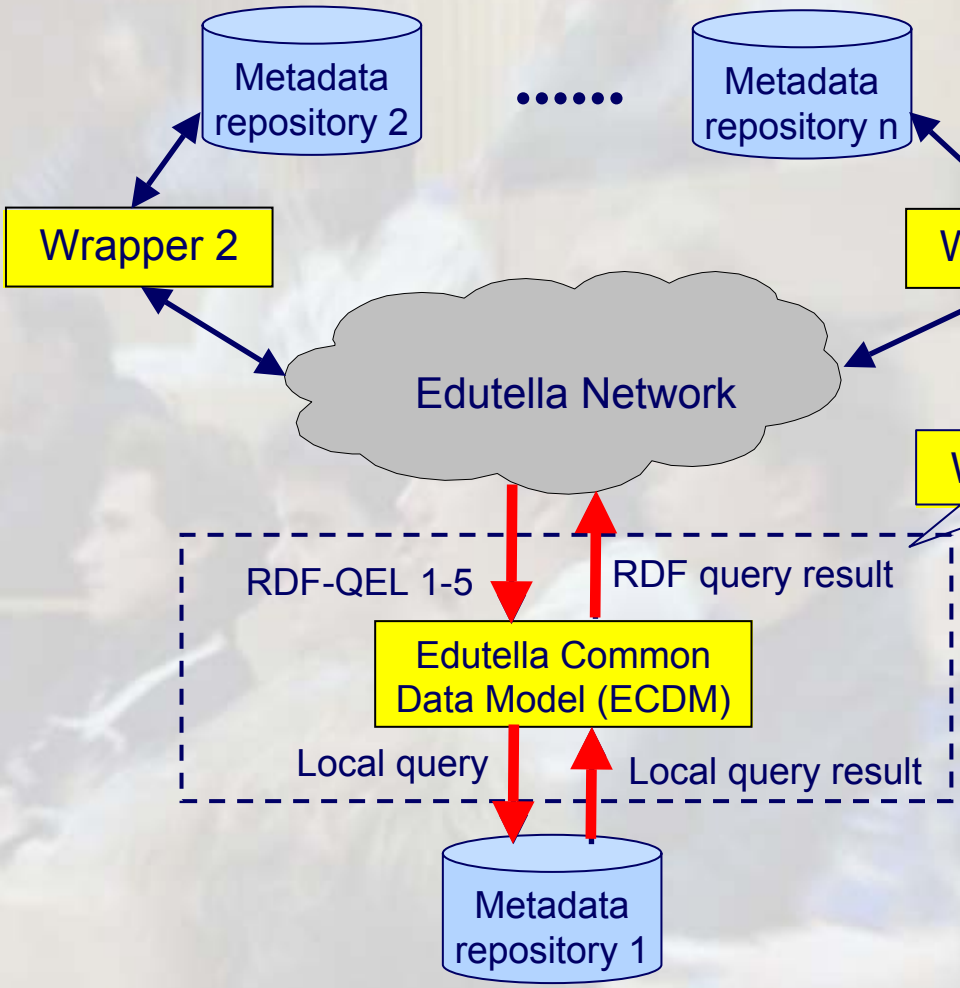
Edutella: Basic Assumptions

- Basic assumption 1/2:
  - lots of learning resource repositories, which typically employ various back-ends, various metadata schemas, and various architectures, etc., have already existed in many institutions. (troubles: isolate information islands, lack of interoperability between each other)
- Basic assumption 2/2:
  - Many institutions are reluctant to give up their control over learning resources developed or accumulated by themselves, which is currently troubling many central-server based approaches to learning resource sharing, e.g., eLearning "portals". (troubles: "portals" are costly but unprofitable)
- Troubleshooters: 
  - P2P: enable institutions to actively participate in a global sharing network without losing the control over their learning resources.
  - RDF: across heterogeneous learning resource metadata repositories to provide basic interoperability.

Edutella: The First Step

- Focus: accommodate learning resource metadata repositories
  - DCMES; IEEE LOM/IMS Learning Resource Metadata Spec.; ADL SCORM
- Current application context and testbed
  - Project: PADLR (Personalized Access to Distributed Learning Repositories) (http://www.learninglab.de/workspace/padlr/index.html)
  - EU/IST project: Elena - Creating a Smart Space for Learning (www.elena-project.org)
  - EU/IST project: UNIVERSAL (http://www.ist-universal.org)
  - Project: ULI (Universitärer Lehrverbund Informatik) (http://www.uli-campus.de/index.html)
  - Project: Teachware on Demand (http://www.teachware-on-demand.de/index.php)
  - and some others ……
- The first prototype has been running for about half a year

Edutella: General Architecture



4 types of wrappers are developed:

- RDB Oracle 8i (SQL)
- OODB ConceptBase (O-Telos)
- Apache Xindice (XPath)
- Flat RDF file (Jena RDQL)

Responsibilities of the wrapper program:

- 1/3: Generate the ECDM-based common data view of individual repositories
- 2/3: Translate common query language (RDF-QEL) into local query language
- 3/3: Transform local query results into Edutella common result exchange format

Edutella: ECDM, RDF-QEL, & Query Result Exchange Format

- Common Data Model: Edutella Common Data Model (ECDM)
  - Binary relational data model with a full compliance with RDF data model
  - Internal query language: Datalog
  - External query language: RDF-QEL
- Common Query Language: Edutella Query Exchange Language (RDF-QEL)
  - RDF QEL1: conjunctive query
  - RDF QEL2: RDF QEL1 + disjunctive query
  - RDF QEL3: RDF QEL2 + query negation (SQL92)
  - RDF QEL4: RDF QEL3 + linear recursive query (SQL3)
  - RDF QEL5: RDF QEL4 + arbitrary recursive query
- Common Result Representation: Edutella Query Result Exchange Format
  - Fully in RDF syntax

Why Do We Need XML-based Repositories in Edutella?

- Edutella is a fully RDF-based P2P network, which determines that RDF is the most naturally applicable metadata representation in Edutella, and thus RDF-based repositories are the most natural form of Edutella content providers. However.....
- Currently XML-based learning resource repositories still occupy a dominant place in E-Learning in comparison to RDF-based ones.
  - XML has a much flatter learning curve and more straightforward binding strategy to the popular learning resource specifications.
  - XML has a longer history to be applied for binding learning resource specifications (e.g., LOM/IMS XML binding available in Aug. 1999, whereas its RDF binding only in June 2001)
  - Native XML databases provide a straightforward way for constructing XML-based learning resource metadata repositories

Native XML Repository: Features & Available Products

Features:

- 1/5: All XML metadata profiles can be directly stored and managed in their original hierarchical form without the need of any pre-processing
- 2/5: Database schema is virtually identical to the XML data schema defined by XML DTD (Schema)
- 3/5: XML metadata profiles with the identical data schema can be stored in a collection and thus be queried as a whole
- 4/5: Easy update through direct manipulation on XML fragments instead of on the whole XML profile
- 5/5: W3C XPath and W3C XQuery support

Available products

- Open source: Apache Xindice (dbXML), etc. xindice (Zeen-dee-chay)
- Commercial: Tamino, Ipedo, etc. Tamino XML Server IPEDO

Integrating Native XML Repositories: Technical Challenges

- Challenge 1/3: incomparability between XML Data model and ECDM
  - Tree-like hierarchical data model (XML) vs. binary relational data model (ECDM)
  - Outcome: the mapping from schema-agnostic XML data model to ECDM is found to be unrealistic
- Challenge 2/3: incomparability between XPath and Datalog
  - Relationally incomplete query language (XPath) vs. relationally complete query language (Datalog)
  - Outcome: only RDF QEL1-3 (conjunctive, disjunctive, query negation) can be translated into XPath
- Challenge 3/3: query result processing
  - Query results returned by native XML repositories (XML fragments) vs. Edutella Common Result Exchange Format (RDF model)
  - Outcome: additional processing is needed to transform local query results into Edutella RDF-based common result exchange format

Integration Approach: Overview

- Resolution to technical challenge 1/3
  - Focused on 3 schema-specific native XML repositories only (DCMES RDF/XML binding, LOM/IMS XML binding, SCORM XML binding)
  - DCMES RDF/XML binding as the basis schema (LOM/IMS, SCORM XML binding are downwards mapped to DCMES RDF/XML binding, and then use the same integration approach)
- Resolution to technical challenge 2/3
  - Focused on RDF QEL1-3 only (conjunctive query, disjunctive query, and query negation)
- Resolution to technical challenge 3/3
  - RDF parser (Jena) is adopted to process local query results (transform XML fragments into RDF model)

Integration Approach: DCMES RDF/XML Binding

- Feature: XML binding is also in valid RDF syntax

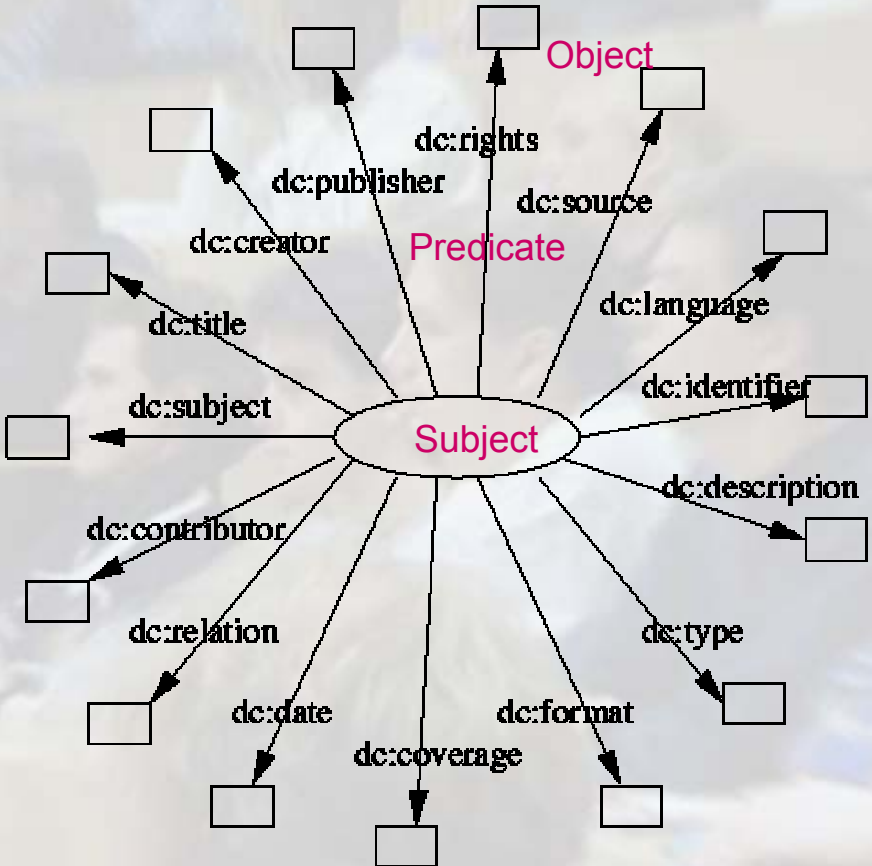
```
<!ENTITY rdfns 'http://www.w3.org/1999/02/22-rdf-syntax-ns#' >
<!ENTITY dcns 'http://purl.org/dc/elements/1.1/' >
<!ENTITY % rdfnsdecl 'xmlns:rdf CDATA #FIXED "&rdfns;"' >
<!ENTITY % dcnsdecl 'xmlns:dc CDATA #FIXED "&dcns;"' >
<!ELEMENT rdf:RDF (rdf:Description)* >
<!ATTLIST rdf:RDF %rdfnsdecl; %dcnsdecl; >
 <!ENTITY % dcmes "dc:title | dc:creator | dc:subject | dc:description|
dc:publisher | dc:contributor | dc:date | dc:type | dc:format |
dc:identifier | dc:source | dc:language | dc:relation | dc:coverage |
dc:rights" >
<!ELEMENT rdf:Description (%dcmes;)* >
<!ATTLIST rdf:Description rdf:about CDATA #IMPLIED>
<!ELEMENT dc:title (#PCDATA)>
<!ATTLIST dc:title xml:lang CDATA #IMPLIED>
<!ATTLIST dc:title rdf:resource CDATA #IMPLIED>
<!ELEMENT dc:creator (#PCDATA)>
<!ATTLIST dc:creator xml:lang CDATA #IMPLIED>
<!ATTLIST dc:creator rdf:resource CDATA #IMPLIED>
.....
```

XML DTD of DCMES RDF/XML binding (in part)

Source: Beckett, D., E. Miller, and D. Brockley, Expressing Simple Dublin Core in RDF/XML

Integration Task 1/3: Generate Common Data View

- The hedgehog model of DCMES RDF/XML binding



Object: *u2[@rdf:resource]* or *u2[text()]*

Predicate: *u1/** as *u2*

Subject: *//*[@rdf:about]* as *u1*

The hedgehog model of DCMES RDF/XML binding

Source: Kokkelink, S., and R. Schwänzl, Expressing Qualified Dublin Core in RDF/XML

Integration Task 2/3: Translate Datalog into XPath (Part 1)

- Typical RDF QEL3 query represented in Datalog

> *H(X) :- P1(X,U), NOT P2(X, V)*
>
> *H(X) :- P3 (X,W)*
>
> *H* is head; *P1*, *P2*, *P3* are predicates; *X* is variable; *U*, *V*, *W* are constants.

- Basic constructs of Datalog
  - Atom: describe a ground assertion, e.g., P(arg1, arg2)
  - Rule (Head, Body)
- Datalog query
  - Conjunctive query: sets of atoms in a rule's body
  - Disjunctive query: sets of rules with the identical head
  - Query negation: Boolean operator "NOT"

Integration Task 2/3: Translate Datalog into XPath (Part 2)

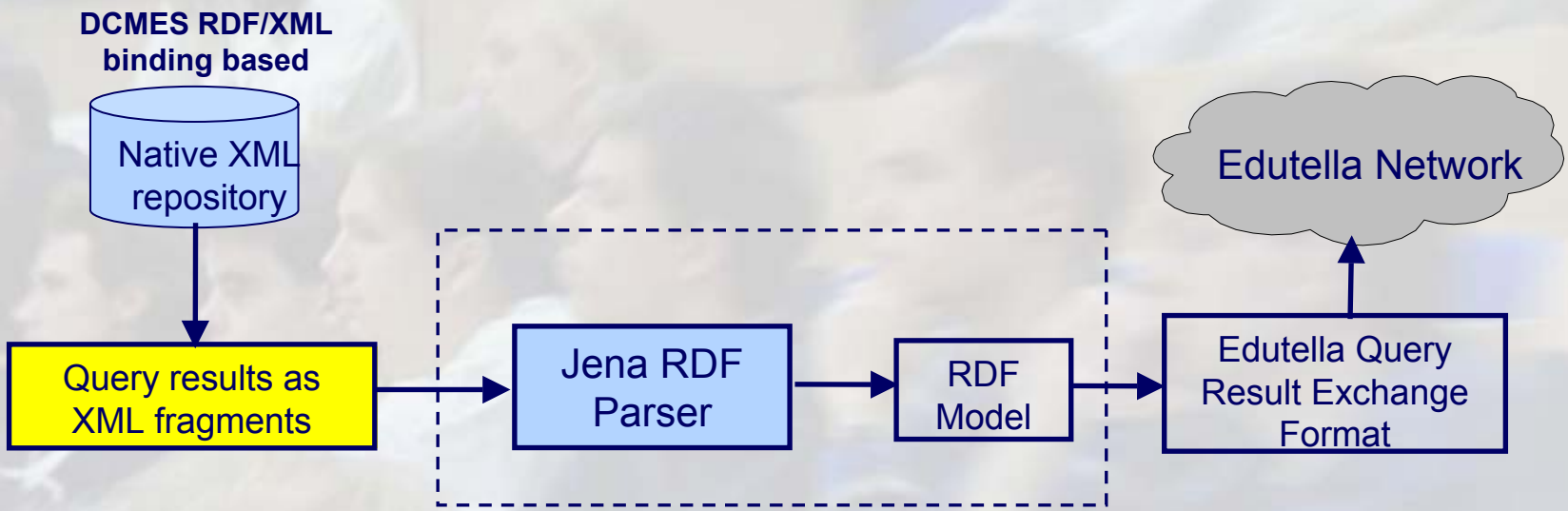
- Identical RDF QEL3 query represented in XPath

```
//*[ @rdf:about and
  (P1 [@rdf:resource=U] or P1 [text()=U])  and
 not (P2 [@rdf:resource=V] or P2 [text()=V])  ]
|
//*[ @rdf:about and
  (P3[@rdf:resource=W] or P3 [text()=W])  ]
```

- Basic constructs of XPath
  - Pattern; filter ([filter pattern])
- XPath query
  - Conjunctive query: sets of filter patterns combined with Boolean operator "and"
  - Disjunctive query: sets of patterns combined with union operator "|"
  - Query negation: Boolean operator "not"

Integration Task 3/3: Query Result Processing

- Jena RDF parser is used to transform XML fragments into RDF Model



Workflow of query result processing

Integration Approach: LOM/IMS & SCORM XML Binding

- Downwards mapped to DCMES RDF/XML binding (using W3C XSLT)

LOM/IMS (SCORM)	DCMES
1.1.2:General.Identifier.Entry	DC.Identifier
1.2:General.Title	DC.Title
1.3:General.Language	DC.Language
1.4:General.Description	DC.Description
1.5:General.Keyword or 9:Classification with 9.1:Classification.Purpose equals "Discipline" or "Idea".	DC.Subject
1.6:General.Coverage	DC.Coverage
5.2:Educational.LearningResourceType	DC.Type
2.3.3:LifeCycle.Contribute.Date when 2.3.1:LifeCycle.Contribute.Role has a value of "Publisher".	DC.Date
2.3.2:LifeCycle.Contribute.Entity when 2.3.1:LifeCycle.Contribute.Role has a value of "Author".	DC.Creator
2.3.2:LifeCycle.Contribute.Entity with the type of contribution specified in 2.3.1:LifeCycle.Contribute.Role.	DC.OtherContributor
2.3.2:LifeCycle.Contribute.Entity when 2.3.1:LifeCycle.Contribute.Role has a value of "Publisher".	DC.Publisher
4.1:Technical.Format	DC.Format
6.3:Rights.Description	DC.Rights
7.2.2:Relation.Resource.Description	DC.Relation
7.2:Relation.Resource when the value of 7.1:Relation.Kind is "IsBasedOn".	DC.Source

The rules used to map LOM/IMS (SCORM) to DCMES

Source: IEEE LTSC, IEEE LOM working draft 6.4

Integration Approach: LOM/IMS & SCORM XML Binding

- Some discussions

  - 1/5: The mapping from LOM/IMS (SCORM) to DCMES will unavoidably lose some metadata info. However, the most of lost info. are only useful for detailed resource description, not for simple resource discovery.

  - 2/5: It is difficult to generate the ECDM-based common data view of LOM/IMS (SCORM) based metadata repositories using XPath because their XML data schemas are more complicated than DCMES.

  - 3/5: Using XPath to query some metadata entries of LOM/IMS (SCORM), e.g., lom:Classification, etc., is basically unusual and unpractical (DCMES's 15 metadata entries are sufficient for discovery).

  - 4/5: W3C XQuery might provide some new resolutions. However, using XQuery might be more expensive than directly constructing RDF-based repositories using LOM/IMS (SCORM) RDF binding.

  - 5/5: For some complicated data schema based repositories, using RDF binding metadata and high-performance back-end systems is the natural (sometimes unique) way to realize the integration.

Conclusions

- 1/4: Although XML is still popular today, RDF should have the future.
- 2/4: Edutella is focused on the future, but still gives XML a place in a RDF world.
- 3/4: Our integration approach, motivated by the current usage of native XML repositories in E-Learning, addresses the immediate integration need of Edutella
- 4/4: Although the integration approach is proved to be feasible, we still encourage the use of RDF-based repositories in the Edutella network.

Thank You for Your Attention

- Related work to this presentation:

[1] Nejdl, W., B. Wolf, C. Qu, S. Decker, M. Stintek, A. Naeve, M. Nilsson, M. Palmer, and T. Risch, Edutella: A P2P Networking Infrastructure Based on RDF, in Proc. of the 11th International World Wide Web Conference (WWW 2002), ACM Press, Honolulu, Hawaii, USA, May 2002.

[2] Qu, C., and W. Nejdl, "Towards Interoperability and Reusability of Learning Resource: a SCORM-conformant Courseware for Computer Science Education", in Proc. of the 2nd IEEE International Conference on Advanced Learning Technologies (IEEE ICALT 2002), IEEE Computer Society Press, Kazan, Tatarstan, Russia, Sept. 2002.

Any Questions ?

M. Eng. Changtao Qu
Email: qu@learninglab.de Homepage: http://www.kbs.uni-hannover.de/~changtao
Learning Lab Lower Saxony, University of Hannover, Germany (http://www.learninglab.de)