Semantic Web technologies in Unit-net IEDI

Vadim Ermolayev
Zaporozhye State University, Ukraine
UnIT-Net: IT in University Management Network
TEMPUS/TACIS MP-JEP-2010-2003

http://eva.zsu.zp.ua/
http://www.zsu.edu.ua/
http://www.unit-net.org.ua/

UkrPROG’04, Kiev, 02-03.06.2004
Outline:

- What is the Semantic Web – just to remind ...
- UnIT-Net: the motivation, the domain, the project
- The State of the Art: the advances and the pitfalls
- Semantic Web technologies in UnIT-Net Infrastructure for Electronic Data Interchange
The Semantic Web*

- **W3C Initiative**
- **Aim**: to provide a comprehensible framework for identifying, representing and processing the SEMANTICS of Web resources
- The ultimate **vision**:
  - Worldwide distributed device for computation
  - Inhabited with artificial service providing agents

A Walkthrough Example

Return the list of the 1-st year CS students who:

- had received maximal grade in Mathematics at the entrance examinations
- and have failed to pass the 1-st Term examination in any basic course in Mathematics

Why?
Return the list of the 1-st year CS students who:

- had received maximal grade in Mathematics at the entrance examinations
- and have failed to pass the 1-st Term examination in any basic course in Mathematics
A Walkthrough Example

Return the list of the 1-st year CS students who:

- had received maximal grade in Mathematics at the entrance examinations

...and have failed to pass the 1-st Term examination in any basic course in Mathematics

Mathematics:
- Math Analysis
- Linear Algebra
- Analytical Geometry
...
To achieve and sustain dynamic improvement service-oriented organizations, like Universities, need an IT infrastructure that underpins:

- Flexible and robust management of their activities through Intelligent Distributed Information Retrieval
- Decision making support
Objective(s):
- Creation of the National “Network of Excellence”
- Dissemination of the best practices – IT in University Management
- Elaboration of the Specifications recommending the reasonable ways of using IT in University Management
- Design and implementation of the Research Prototype of the National Infrastructure for Electronic Data Interchange (motivation)

Participants:
- Kherson State University (project coordinator)
- Ministry of Education and Science of Ukraine
- Kharkiv national University
- Zaporozhye State University
- University of Nice – Sofia Antipolis, France
- Glasgow Caledonian University, UK

http://www.unit-net.org.ua/
The State of the Art ...

- **Not really a breakthrough** in the Domain!!!
- Lots of related work, e.g., TSIMMIS, MOMIS, BUSTER, DOME, InfoSleuth, KRAFT, OBSERVER, Ontobroker, PICSEL, SIMS, … (proves the importance)
- **Novelties:**
  - **Ontologies** specified in W3C emerging de facto standard language (**OWL DL**)
  - **Ontology-driven** Semantic Query Formulation, Transformation, … (ZSU RACING Project)
  - IR (RDB-structured) semantics is formalized by means of a semi-structured Ontology Specification Language (**OWL DL**)
  - Conceptually – one more layer (**Ontology**) of Semantic Specification on top of the IR schema
  - **Semantic Web Service** technology for Uniform IR Wrapping
- All these is in the **Mainstream of Semantic Web Activities**
Complications: Natural Distribution and Heterogeneity

- **Organizations** involved in the Educational framework are rightfully independent
- They own and maintain their data and knowledge sources autonomously
- Serious complications for their integration:
  - IR-s may be opened or closed to external access
  - IR-s may be provided by different hardware and software using various notations and protocols
  - IR-s may be disparately structured
  - IR-s may have different data models behind them
  - IR-s are semantically heterogeneous
The Principles of IEDI Architecture

- **Mediator architecture** with the centralized mediator
- **Hybrid** approach to **knowledge representation**
  - Centralized **Mediator Domain Ontology** (MDO)
  - De-centralized **Information Resource Ontologies** (IRO)
- Use of **IR Registration** to allow the resource become available for querying
- Does **not** provide **full automation** for ontologies’ mapping and alignment
- **Rewriting** technique with **mappings** and **late binding** to produce, process, and perform queries
IEDI Architecture in a Nutshell

a User having an arbitrary query

IEDI

IR Providers which own disparate resources

CS Student IR

Univ. Entrant IR
IEDI Architecture in a Nutshell

User Layer

Mediator Layer

IR Wrapper Layer

IR Layer

Request to formulate a query

Q-ry Results in terms of MDO

 Pose queries

an IR Provider

Register IR-s

Maintain IR changes

Request to formulate a query

Q-ry Results in terms of MDO

 Pose queries

an IR Provider

Register IR-s

Maintain IR changes

*IEDI Mediator*

- Query Formulation Server
- Sub-Query Extraction Server
- Sub-Query Execution Server
- Results Mark-Up Translation Server

*IR Layer*

- IR

*IR Wrapper Layer*

- IR Wrapper Web Service
- IR Wrapper

*IRKB*

- AUPO
- MDO
- IRDMO

User

MKB

User

MKB
The Tasks for UNIT-NET IEDI

- **To Query Distributed Semantically Heterogeneous Information Resources**
- **To Register Information Resources**
- **To Maintain Coherent Semantic Descriptions**

What **IEDI** is **NOT** Supposed to Do:
- IR updates
- Results Fusion
IEDI: User Categories and Roles

- **An Authorized USER (AU):**
  - Poses queries in the terms of University Management Domain (a TOOL, a LANGUAGE)

- **A MEDIATOR ONTOLOGIES ENGINEER (MOE):**
  - Maintains Domain Ontology KB (a TOOL)
  - Interacts with RESOURCE ONTOLOGY ENGINEERS for:
    - Registering their Resources (semi-automatic, a TOOL)
    - Aligning Domain and Resource Ontologies (Semi-automatic, a TOOL)

- **An IR ONTOLOGY ENGINEER (IROE):** ...

- **An IR PROVIDER (IRP):**...
IEDI: Functionalities

- **Query** *(automatic)*
  Distributed
  Semantically
  Heterogeneous
  Information
  Resources

- **Register** *(semi-auto)*
  Information
  Resources
  *(ontology merge)*

- **Maintain** *(semi-auto)*
  Coherent
  Semantic
  Descriptions
  *(ontology alignment)*

- **!!!** Semi-automatic, authorized, secure ...
Formulate Query
Get Query Results

Formulate Query

Query O’k?

Yes

No

Yes

No

Critical?

No

Report: Ontology Problem

Yes

Extracted?

No

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?

Yes

Extracted?
Semantic Web components in IEDI

- Ontology Language: **OWL** (W3C recommendation)
- Ontologies at Mediator and IR layers
- Mediator Query Language: **RDQL** (W3C recommendation)
- Mark-up Language: **XML** (W3C standard recommendation)
- Ontology processing tools
- **Semantically reinforced Web Services** for IR wrapping
**Ontologies** are developed to provide a machine-processable semantics of IR-s that can be communicated between different software and humans.

- **An ontology is a formal, explicit specification of a shared conceptualization**
  - **Conceptualization** - a simplified abstract model of some object or phenomenon in the world which identifies the relevant concepts of that object or phenomenon.
  - **Formal** ...
  - **Explicit** ...
  - **Shared** ...

IEDI: Hierarchy of Ontologies
## IEDI: Usage of Ontologies

<table>
<thead>
<tr>
<th>Processes</th>
<th>ULO</th>
<th>MDO Core</th>
<th>MDO</th>
<th>IRDMO</th>
<th>UPRO</th>
<th>IRO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query distributed autonomous semantically heterogeneous information resources</td>
<td>--</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R/U</td>
<td>R</td>
</tr>
<tr>
<td>Register new information resource</td>
<td>R</td>
<td>R</td>
<td>R/U</td>
<td>R/U</td>
<td>--</td>
<td>R</td>
</tr>
<tr>
<td>Maintain coherent semantic descriptions</td>
<td>R</td>
<td>R/U</td>
<td>R/U</td>
<td>R/U</td>
<td>R/U</td>
<td>R/U</td>
</tr>
</tbody>
</table>

- **R** – usage for reference purposes only
- **R/U** – used as a reference and is updated
- **--** – not used
Ontologies: Walkthrough Example

ULO: Artifact

IRO: CS Students, Univ X

MDO:
- Mathematics
  - Math. Analysis
  - Analytical Geometry
  - General Algebra
  - Subject of Mathematics

IRO: CS Students, Univ Z

ULO: Immaterial

IRO: Subject in Mathematics

MDO:
- Algebra
  - Linear Algebra
  - Geometry
  - Math. Statistics

IRO: Algebra

Why?
Semantically Reinforced Components

- Query Formulation Server
- Sub-query Extraction Server
- Results Mark-up Translation Server
- IR Wrapper
Query formulation (in the terms which are convenient and understandable for the specific AU) – manual, by the Tool

Query Transformation: reformulating the query in the terms of MDO (preserving the recall) – automatic

The GOAL: transform the Initial Query (IQ) to the Resulting Query (RQ) in the terms of the MEDIATOR DOMAIN ONTOLOGY

The procedure:
- Form the Query Plan (QP) by parsing the IQ
- Use the User’s Profile to map the key words of QP to the concepts of the DOMAIN ONTOLOGY
- Use semantic relationships between the concepts of the DOMAIN ONTOLOGY to add more semantics to RQ

The Basic Principle: – IQ preservation
IQ Preservation

- IQ preservation principle:
  - Strategic goal is to gain more Recall and more Precision
  - I.e., relevant RQ results $r(RQ)$ should be the sub-set of all IQ results $t(IQ)$ and, ideally, the difference $t(IQ) \setminus r(RQ)$ should contain only irrelevant results
  - Consequently, RQ should have the same or the broader meaning than IQ

- Transformation mappings are produced in the way providing that the recall of the RQ is at least the same than the recall of the IQ
IEDI: Implemented Components

- User Profile Ontology Editor
- Tool for IQ Plan editing and approval
  - Contribution of the RACING project
    http://racing.zsu.zp.ua
- Parts of IEDI Mediator Query Formulation Tool
IEDI Mediator: Sub-Query Extraction*

1. Preliminary grouping
2. Finding Determining Concepts
3. Concept mapping
4. Slot mapping
5. Ensuring that sub-query results will be correct RDF graphs
6. Forming RDQL SELECT sections
7. Forming RDQL AND sections

Walkthrough Example: IQ (RDQL)

Retrieve the list of the 1-st year students who have received maximum grade (5) in Mathematics at the University entrance examinations and have failed to pass the 1-st semester examination in any basic course in Mathematics (got unsatisfactory grade - 2).

```
SELECT ?firstName, ?secondName, ?lastName, ?specialityName, ?sessionExTitle
WHERE
  (?x, stud:first_name, ?firstName), (?x, stud:second_name, ?secondName),
  (?x, stud:exams_passes, ?y),
  (?y, stud:exam_title, ?entrantExTitle), (?y, stud:exam_type, ?examType1),
  (?y, stud:entrant_grade, ?entrantGrade), (?y, stud:examOnDiscipline, ?r1),
  (?z, stud:exam_title, ?sessionExTitle), (?z, stud:exam_type, ?examType2),
  (?z, stud:session_grade, ?sessionGrade),
  (?a, stud:specialityName, ?specialityName),
  (?r1, stud:disciplineName, ?entrDiscName), (?r1, stud:includes, ?i1),
  (?r2, stud:disciplineName, ?sessionDiscName), (?r2, stud:includes, ?i2),
  (?i1, stud:disciplineName, ?discName1),
  (?i2, stud:disciplineName, ?discName2)
AND (?examType1 eq "Exam"), (?examType2 eq "Exam")
AND (?entrDiscName eq "Mathematics")
AND ((?sessionDiscName eq "Mathematics")
AND ((?entrantExTitle eq ?discName1) || (?sessionExTitle eq ?discName2)))
AND ((?sessionExTitle eq "Linear Algebra") ||
  (?sessionExTitle eq "Mathematical Analysis"))
AND (?entrantGrade eq "5")
AND (?sessionGrade eq "2")
AND (?semesterNum eq "1")
USING stud FOR <MDO-URL#>
```
Walkthrough Example: Results (RDQL)

```
SELECT ?firstName, ?secondName, ?lastName, ?specialityName, ?sessionExTitle
WHERE
(?x, abo:aboName, ?firstName),
(?x, abo:secondName, ?secondName),
(?x, abo:surname, ?lastName),
(?x, abo:passes, ?y),
(?x, abo:AboSpec, ?a),
(?y, abo:EntrantExamName, ?entrantExTitle),
(?y, abo:examType, ?examType1),
(?y, abo:grade, ?entrantGrade),
(?y, abo:examOnDiscipline, ?r1),
(?a, abo:specialityName, ?specialityName),
(?r1, abo:disciplineName, ?entrDiscName),
(?r1, abo:includes, ?i1),
(?i1, abo:disciplineName, ?discName1),
AND (?examType1 eq "Mathematics")
AND (?entrantExTitle eq ?discName1)
AND (?entrantGrade eq "5")
USING abo FOR <IRO-Entrant-URL#>
```

```
SELECT ?firstName, ?secondName, ?lastName, ?specialityName, ?sessionExTitle
WHERE
(?x, stud:name, ?firstName),
(?x, stud:secondName, ?secondName),
(?x, stud:surname, ?lastName),
(?x, stud:examPasses, ?z),
(?x, stud:onSpec, ?a),
(?z, stud:examName, ?sessionExTitle),
(?z, stud:examType, ?examType2),
(?z, stud:grade, ?sessionGrade),
(?z, stud:semesterNum, ?semesterNum),
(?a, stud:specialityName, ?specialityName)
AND (?examType2 eq "Exam")
AND ((?sessionExTitle eq "Linear Algebra") ||
    (?sessionExTitle eq "Mathematical Analysis"))
AND (?sessionGrade eq "2")
AND (?semesterNum eq "1")
USING stud FOR <IRO-Student URL#>
```
Retrieve the list of the 1-st year students who:
- have received maximum grade (5) in Mathematics at the University entrance examinations
- and have failed to pass the 1-st semester examination in any basic course in Mathematics (got unsatisfactory grade - 2).

Retrieve the list of the 1-st year students who have received maximum grade (5) in Mathematics at the University entrance examinations

Retrieve the list of the 1-st year students who have failed to pass the 1-st semester examination in any basic course in Mathematics (got unsatisfactory grade - 2).
IEDI: IR Wrapping

- **IR Wrapper** design is based on Web Service Technology

- IR Wrapper Web Service is Semantically Reinforced by:
  - Generic IR Wrapper
  - Specific IR Wrapper **binding**

- which use **IRO** for their operations

IEDI Generic Wrapper and Wrapper Bindings

Web Service Port

IR Wrapper Web Service

Translate Terminology

Translate Query Notation (RDQL-IRQL)

Perform IR Query

Mark-Up Query Result (in terms of IRO)

IR Wrapper Server

IRO

WKB

IR Layer

Query Results (Marked-up in terms of IRO)

Query (RDQL in terms of IRO)

IR specific (wrapper binding)

IR invariant (generic wrapper)

Wrapper Layer

Query Results (Plain Text)

IR Server
IEDI: Implemented Components

- Generic IR Wrapping Web Service
- Wrapper Testing Suite
- IR Wrapper for ZSU University Entrant IR
Unit-net IEDI: to Round up …

- That is what we have done in the project ... so far

- Semantic web technologies are used (and developed) for:
  - representing different aspects of knowledge
    - domain, resource, user profile, mapping, high-level
  - formulating, transforming, splitting down the queries to sub-queries
  - IR wrapping
  - Query results mark-up
That’s it …

Shall be happy to know the answers

Mentioned papers and these slides are available from:
http://eva.zsu.zp.ua/eva_personal/evapubs.htm