Information and Knowledge Interoperability

IASI-CNR

Laboratory for Enterprise Knowledge and Systems (LEKS)

Michele Missikoff, Francesco Taglino, Fabrizio Smith
Outline

• Interoperability
  – Semantic interoperability
• Introduction to semantic web
  – Introduction to Ontologies
• Macro-architecture semantic services
• Knowledge interoperability
• Information Interoperability
IASI-CNR, LEKS: Profile

• Spin-off from the Research Area: Knowledge Bases and Information Systems at IASI-CNR
• Goal: experiment scientific results, test software tools, participate to applied projects
• Staff: ca 20, including students from Engineering and Informatics, Laurea Breve/ Specialistica/PhD
• Funding: totally from projects and private grants, approx 600 K€ per year
• More than 60 publications in the last 5 years
• Founded: November 1999
Interoperability reference model
(FP6 EU Athena project)

<table>
<thead>
<tr>
<th>Provided</th>
<th>Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enterprise (Knowledge)</td>
<td>Enterprise (Knowledge)</td>
</tr>
<tr>
<td>(Business) Processes</td>
<td>(Business) Processes</td>
</tr>
<tr>
<td>Services</td>
<td>Services</td>
</tr>
<tr>
<td>Information/Data</td>
<td>Information/Data</td>
</tr>
</tbody>
</table>

- Collaborative Enterprise Modelling
- Cross-Organisational Business Processes
- Flexible Execution and Composition of Services
- Information/Data Interoperability
- Semantic Mediation Interoperability
What is Semantic Web?

A vision of possibilities

“The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”

Current situation of the web

The web contains a **tremendous amount** of data/information/knowledge that is:

- freely available
- easily reachable
- covering all possible areas of human activities

But ...

- Contents are **poorly organised**, mainly in textual form
- Content organisation, when present, greatly varies
- Automatic filtering/extraction of content is very difficult (Google...); the output remains in the (original) text form
Traditional Web

Factual Resources (FR): Data, music, pictures, …

(HTML, MP3, jpeg, mpeg, …)

Computer: rendering without “understanding”
Semantic Web

Knowledge Resources (KR):
Concepts, semantic nets, ontologies, …

Knowledge Network
- RDF, OWL, Rules
- Semantics (Ontologies)

Network (HTML)
An Ontology is ...

“... a theory about the nature of beings” (Philosophical view)
“... a formal, explicit specification of a shared conceptualisation.” (AI view)*

'Formal' refers to the fact that the ontology should be machine understandable.

'Explicit' means that the type of concepts used and the constraints on their use are explicitly defined.

'Shared' reflects the notion that ontology captures consensual knowledge, that is, it is not restricted to some individual, but accepted by a group.

A 'conceptualisation' refers to an abstract model of some phenomena in the world, which identifies the relevant concepts of that phenomena.

Computational Ontologies

• Ontologies represent traditionally a philosophical subject, but we are interested in their computational versions

• An Ontology is a socio-cultural phenomenon, but we need to describe the concepts in a formal and unambiguous way, processable by a computer

An ontology contains:

– a set of concepts (e.g., entities, attributes, processes) relevant in a given domain

– their definitions and inter-relationships (conceptualization)

• To be used by computers, ontologies must

  – have precise definitions, with a mathematical semantics (Tarski)
  – be able to grow and adapt to current usage by both human and computer users, reflecting the evolving of modeled reality
Conceptualization & Instatiation

A doggish entity

Concept Descr.

myDog

Instance Descr.

concept

individual

conceptualization

Classification
From Terms to Knowledge

Knowledge Base

Ontology

Thesaurus (Taxonomy)

Glossary

Terminology

Directories

Databases

Doc Repositories

White Pages

Yellow Pages

KB = Concepts + rules + Instances + documents
Unary modeling notions

• Domain neutral
  – Class, entity, object, relation, type, ...

• Possible domain categorization (OPAL)
  – Business Object
  – Process
  – Actor
  Other categories
  – Decision, event, message, goal, ...
Binary modeling notions

• **Vertical**
  - Refinement (Gen/Spec)
  - Decomposition (PartOf/HasPart)
  - Membership (ElemOf/HasElem)
  - Predication* (hasData/ObjProp)
  - Instantiation

• **Horizontal**
  - Similarity
    - Similarity degree
  - Association+ (dataProperty, objectProp)
  - Relatedness
    - Named/unnamed relationships
  - Implication
    - New knowledge
    - Actions

*has binary range, e.g., age
An example: Semantic Net

- CAR
  - myCar
  - fleet
  - garage
  - PartOf
  - RelTo
  - Inst
  - Pred
  - Spec
  - Gen
  - Ssim
  - Assoc
  - HasPart
  - driver
  - pickup
  - wheel
  - color
  - limo
Macro-architecture of Semantics-based Infrastructures developed by IASI

- **SOBE + OPAL**: Social Ontology Building and Evolution
- **Document Repository**: Stores all documents.
- **Reference Ontology**: Provides a structured view of the domain.
- **Sem Annot**: Semantic Annotation for Structured Docs.
- **SA4SD OFV**: Semantic Annot for Structured Docs Ontology-based Feature Vectors.
- **SRRG**: Semantic Reconciliation Rules Generation.
- **SIRE**: Semantic Interop Runtime Engine.
- **Semantic Reconciliation of biz docs**: Processes documents for semantic reconciliation.
- **Semantic BP**: Semantic Business Processes.
- **Query/Reasoning**: Tools for querying and reasoning on semantic data.
- **BPAL**: Business Processes Annotation Language.
- **SemSim**: Semantic Similarity Search and matchmaking.
- **Semantic Cluster Mgmt**: Management of semantic clusters.
Knowledge Interoperability services
Objectives

Achieving balanced and harmonised Collaborative Network

• Modelling Enterprise Semantic Profile (ESP) focussing on Competencies and Skills (CS)

• Semantically enriching CS of each enterprise in the CN to:
  – Understand the overall CS of the CN
  – Identify CS gaps and complementarities
  – Define the “ideal” CS asset for the CN and identify the missing CS (e.g., to be acquired with new partners)

To this end, we need to develop:

• Semantic Supporting Services: SSS, necessary for KIS
• Knowledge Interop Services: KIS, to support the harmonisation of CS for better enterprise cooperation in the CN
We consider: CS of partners, ideal CS of CN, CS Missing Gap
Beyond the State of the Art (1)

- **Objective:** achieve a unified method to model Enterprise capabilities (i.e., CS)
- **SotA:** Very complex and heterogeneous Enterprise Modelling (EM) frameworks (e.g., POP*, CIMOSA, GERAM, Zachman)
- **Innovation:** Describe CS by using ontology-based Enterprise Semantic Profiles (ESP):
  - Extract CS from the human-oriented knowledge (filtering the technical documentation)
  - Map the filtered Enterprise Knowledge to the CSOneto, creating the ontology-based ESP
Beyond the State of the Art (2)

- **Objective:** ontology-based analysis of overlapping and complementary CS through *Semantic Similarity Reasoning*
- **SotA:** A plenty of semantic similarity methods (e.g., Lin, Resnik, ...) for comparing pairs of concepts, not semantic profiles
- **Innovation:** define an enterprise semantic matchmaking method for
  - Computing similarity b/w ESP
  - Identify gaps/overlapping among ESP
Knowledge EI Services

Semantic Supporting Services (SSS)
- Mass Ontology Building (from doc & Web 2.0)
- Enterprise Semantic Profiling
- Enterprise Semantic Matchmaking

Knowledge Interoperability Services (KIS)
- Semantic analysis of new partners entering CN
- Semantic analysis of enterprises’ CS in CN, to identify CS gap & overlapping
- Semantic enhancement of CN CS, by transferring the Knowledge acquired during VO activities / loss of competencies
**Ontology Building**

- Dictionaries
- Thesauri...
- Domain related documents

**Semantic Annotation**

- Enterpr Related docs
- Social Ontology Building and Evolution srv

**Semantic Matchmaking**

- ESP X
- ESP Y
- Enterprise Semantic Matchmaking srv

**Semantic Supporting Services**

- CSOnto
- Experts Community
- Semantic effectiveness of CS in a CN
- Semantic coverage of a company entering the CN
- Competences transfer from a VO to the CN

**Knowledge Interoperability Services**

- Enterprise Semantic Profiles
- Semantic Similarity
- Semantic Coverage

---

**Definitions**

- **CS**: Competences & Skills
- **CN**: Collaborative Network
- **VO**: Virtual Organization
Social Ontology Building and Evolution (SOBE)
• Step-wise approach through five incremental steps (Milestones)
  – **Lexicon (M1)**: plain list of terms
  – **Glossary (M2)**: terms + natural language definition
  – **Concept Categorization (M3)**: in accordance with the OPAL (e.g., Object, Process, Actor)
  – **Taxonomy (M4)**: definition of ISA hierarchy
  – **Ontology enrichment (M5)**: additional relationships (e.g., predication, relatedness)
The SOBE process

- **Enterprise Docs**
  - Terms Extractor
    - Gloss Extractor
      - Gloss Validator
        - N-Lexicon
          - N-Lexicon
            - N-Glossary
              - Concept Categorization environment

- **Google define Wordnet**

- **Taxonomy Extractor**
  - Taxonomy Proposer & Validator
    - N-Taxonomy
      - Initial Ontology
        - Ontology Enrich.
          - N-Ontology
            - M1

- **Terms Validator**

- **E-Lexicon**

- **Gloss Validator**

- **E-Glossary**

- **Glossary Validator**

- **M1**

- **M2**

- **M3**

- **M4**

- **M5**
Driven by Web 2.0 and social communities philosophy

• **Voting**: accept/discard results of the automatic extraction (lexicon and glossary)

• **Proposing**: new terms and definitions to be validated by participants

• **Discussing**: for reaching an agreement on glossary definitions (dedicated forums)
A community of domain experts to support the ontology building process

• **Ontology master**: the responsible of the ontology building task

• **Participant**: who takes part to the validation and enrichment of extracted knowledge

• **Moderator**: who leads the forums in the glossary building discussion
Social participation functions: GUI

Term Validation

OM Term Validation

Participants validation

Ontology Master final validation

Validation Deadline
Social participation functions:

GUI

Gloss Validation

<table>
<thead>
<tr>
<th>3B Terms</th>
<th>Definitions</th>
<th>Accept</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>asynchronous execution (2 definitions found)</td>
<td>Asynchronous execution is the recommended mode of operation as this gives a client more control over the execution of the workflow as we describe shortly.</td>
<td>✔</td>
<td>edit</td>
</tr>
<tr>
<td>business process (2 definitions found)</td>
<td>Business process is an application thing, not just a layer across the enterprise.</td>
<td>✔</td>
<td>edit</td>
</tr>
</tbody>
</table>

OM Gloss Validation

<table>
<thead>
<tr>
<th>3B Terms</th>
<th>Definitions</th>
<th>Accept</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>asynchronous execution (2 definitions found)</td>
<td>Asynchronous execution is the recommended mode of operation as this gives a client more control over the execution of the workflow as we describe shortly.</td>
<td>✔</td>
<td>edit</td>
</tr>
<tr>
<td>business process (2 definitions found)</td>
<td>Business process is an application thing, not just a layer across the enterprise.</td>
<td>✔</td>
<td>edit</td>
</tr>
</tbody>
</table>

Participants validation

Validation Deadline

Forum discussion Deadline

Participants

Moderator
Term extractors (TE)

- 6 TE Analyzed. Best results from
  - TermExtractor from Univ La Sapienza” (precision=0.5, recall 0.4)
  - Alchemy (precision=0.3, recall 0.4)

- To integrate more than one TE

Gloss Extractor

- Current integration of the Gloss Extractor from Univ Sapienza (Rome)
  - Access multi sources (e.g., Google define, Wordnet)
Case study: IVSZ ontology on ICT

Docs from 8 enterprises of the IVSZ cluster (Hungarian Association of Information Technology Companies), involving a team of 4 people.

Lexicon
- 102 automatically extracted terms
- 21 accepted terms after the validation

Glossary
- definitions automatically extracted and validated within 2 sessions (voting and forum stages)

Next steps
- Building the taxonomy (ISA hierarchy)
- Use the ACM categorization and other dictionaries to enrich the ontology
**Weighted Reference Ontology**

- The ontology as a (WRO):
  - **Concept weight**: probability that a digital resource (DR) is characterised by the concept

- Different approaches for assigning weights
  
  E.g.:
  
  - **Uniform probabilistic distribution** (bottom up and independant on the annotated resources)
  
  - **Relative frequency** (top down and dependant on the annotated resources)
An example of WRO (uniform probabilistic distribution)
Enterprise profiling: Objectives

• To build a semantic profile for each of the companies in the Collaborative Network

• Automatic support through
  – Knowledge (chunks) extraction from enterprise related documents
  – Semantic filtering guided by the reference ontology

• Automatically proposed profile

• People from each company to validate the proposed profile
OFV-based Semantic Profile

• Semantic profile represents an ontology-based annotation of an enterprise
• Each enterprise can be annotated with one or more concepts
• We refer to this annotation as:

  OFV

  Ontology-based Feature Vector
Enterprise Semantic Profile Building

Documents describing company A

CSOnto

Web Resources

Chunks Extraction

Semantic Filtering

esp_A (automatically proposed)

ESP Validation

esp_A (validated)
Examples of ESP

\[ esp_A = [\text{InformationSystem}, \text{ErrorHandlingAndRecovery}, \text{ObjectOrientedDevelopment}] \]

\[ esp_B = [\text{WorkflowManagementSystem}, \text{RequirementsAnalysisAndSpecifications}, \text{ObjectOrientedDevelopment}] \]
Enterprises Semantic Matchmaking

Enterprise A Sem Profile

Enterprise B Sem Profile

ESM

Semantic Similarity
Semantic Collaborative Network

Enterpr Docs

Potential new CN partner

Enterpr Docs

CN partner

Enterpr Docs

CN partner

Enterpr Docs

CN partner

Enterpr Docs

CN partner

Potential new CN partner

CN Semantic Profile

CSOnto

Semantic CN
Semantic effectiveness of CS in a CN

**CN** = Collaborative Network
**CS** = Competences & Skills

- Partners’ CS (ESP)
- Ideal CN-CS (CSOnto)
- CN-CS Gap
- CN-CS on Target (Strengths and Weaknesses)
Partner entrance in the CN

CN = Collaborative Network
CS = Competences & Skills

Partners’ CS (ESP)
Ideal CN-CS (CSoOnto)
Semantic Coverage of the entering company

New CN partner

CN = Collaborative Network
CS = Competences & Skills
Information Interoperability
The Interoperability problem

Different terminology
Different data structure
Different data organization
The EAI state-of-the-Art solution

App_A  Adaptor App_A/App_B  App_D

App_C  Adaptor App_A/App_C  Adaptor App_C/App_B  App_B

- Geometric growth $O(n^2)$ of the adapters respect to the number of parties
- Maintenance issues
The Semantic Reconciliation approach

Set up phase
• Reference ontology
• Semantic annotation of the documents to be exchanged
• Transformation rules for each document type
  – Forward
  – Backward

Run-time phase
• Accepting the message (as issued by the sender)
• Transform the payload according to the Fw Tr rules
• Produce the “neutral” ontology representation
• Transform the payload according to the Bw Tr rules
• Deliver the message to the receiver, formatted according to the receiver specs
The Semantic Reconciliation approach

- **Mapping** = Semantic Annotation + Transf rules
- Mapping expressed in terms of a RO
- **Two sets of rules** for each SA (linear growth, O(n))
- Reconciliation through composition of rules

- **Composition of Transformation rules**
  - Forward (App\textsubscript{X}2RO)
  - Backward (RO2App\textsubscript{X})
Semantic Reconciliation Suite

- Athos: Reference Ontology
- SMD: Semantic Annotation Tool
- Annotation Repository
- SRGS: Transformation Rules Building Tool
- Transformation Rules Repository
- SIRE: Reconciliation Engine
- App_A
- Doc Instance_A
- App_B
- Doc Instance_B

Ontology Management System

Set up phase

Run-time phase
Sem Annot of Structured Documents

Schema Document (e.g., PurchaseOrder)

Reference Ontology (on e-Proc)

- rs:company
  - rs:works_for
  - rs:employee
    - rs:name::string

- ro:organization
  - ro:has_employee
    - ro:person
      - ro:has_name
        - ro:name
          - ro:firstname::string
          - ro:surname::string

Naming Mismatch
Attribute granularity Mismatch

Schema Document (e.g., PurchaseOrder)
Sem Annot of Structured Documents
A classification of Mismatches

**Lossless Mismatch**

- **Naming**: different labels for the same entities
- **Abstraction**: Level of specialization/refinement of the information. The same concepts are recognized, but they are defined at different levels of abstraction
- **Structuring**: The same set of concepts is modeled, but it differs the way these concepts are structured by means of relations
- **Subclasses-Attribute values**: an attribute with predefined value set is represented by a set of subclasses
- **Class-relation**: a concept is represented as a relation
- **Attribute Granularity**: The same information is decomposed into a different number of attributes (or sub-attributes)
- **Attribute Assignment**: Two conceptualization differ in the way they assign attributes to concepts
- **Complex Attribute**: A set of attributes is grouped and represented as a concept.
- **Encoding**: Different formats of data or units of measure.

**Lossy Mismatch**

- **Overlapping**: There is an intersection between the extensions of different concepts/attributes/roles.
- **Subsumption**: There is an inclusion between the extensions of different concepts/attributes/roles.
- **Categorization**: Two conceptualizations distinguish the same concept but divide it into different and incomparable sub-concepts.
- **Coverage**: Two conceptualizations do not model all the entities or information of a given domain.
Sem Annot of Structured Documents (Mapping Discovery)
Concept Mismatches Pattern (1)

Atomic Concepts (naming)

$$A \equiv B$$

$$\forall x. PO(x) \leftrightarrow PurchaseOrder(x)$$

Conjunctive Concepts (naming, abstraction)

$$A \equiv B \land C$$

$$\forall x. WorkingStudent(x) \leftrightarrow Employee(x) \land Student(x)$$

Disjunctive Concepts (naming, abstraction)

$$A \equiv B \lor C$$

$$\forall x. Person(x) \leftrightarrow Male(x) \lor Female(x)$$
Concept Mismatches Pattern (2)

- **Role Restrictions** *(naming, abstraction)*

\[ A \land \exists R.\text{Thing} \models \subseteq \triangleright B \]

\[ \forall x. \exists y \text{Person}(x) \land \text{WorksFor}(x, y) \leftrightarrow \text{Employee}(x) \]

\[ A \land \exists R.C \models \subseteq \triangleright B \]

\[ \forall x. (\exists y. \text{Person}(x) \land \text{WorksFor}(x, y) \land \text{University}(y) \leftrightarrow \text{Researcher}(x)) \]

- **Attribute Restrictions** *(naming, abstraction, subclass-attribute)*

\[ A \land \exists P.v \models \subseteq \triangleright B \]

\[ \forall x. \text{Person}(x) \land \text{hasSex}(x, "male") \leftrightarrow \text{Male}(x) \]
Role Mismatches Pattern

Atomic Roles (naming)

\[ R_1 \equiv \subseteq R_2 \]
\[ \forall x, y. \text{hasBuyer}(x, y) \leftrightarrow \text{Buyer}(x, y) \]

Inverse Roles (naming, abstraction, organization)

\[ R_1 \equiv \subseteq R_2^- \]
\[ \forall x, y. \text{worksFor}(x, y) \leftrightarrow \text{employed}(y, x) \]

Path Roles (naming, abstraction, organization)

\[ R_1 \cdot R_2 \equiv \subseteq R_3 \]
\[ \forall x, y. (\exists z. \text{hasHeader}(x, z) \land \text{hasBuyer}(z, y) \leftrightarrow \text{hasBuyer}(x, y)) \]

Constrained Path Roles (naming, abstraction, organization)

\[ A.R_1.B.R_2.C \equiv \subseteq R_3 \]
\[ \forall x, y. (\exists z. \text{hasHeader}(x, z) \land \text{hasBuyer}(z, y) \land \text{PO}(x) \land \text{Header}(z) \land \text{Buyer}(y) \leftrightarrow \text{hasBuyer}(x, y)) \]
Attribute Mismatches Pattern

Atomic Attribute (naming, attribute assignment)

\[ P_1 \equiv \sqsubseteq \sqsupseteq P_2 \]
\[ \forall x, y. hasName(x, y) \leftrightarrow Name(x, y) \]

Function Application (naming, encoding, attribute assignment)

\[ \forall x, y. bDay(x, y) \leftrightarrow Birthday(x, date2string(y)) \]

Attribute Composition (naming, granularity, attribute assignment)

\[ \forall x, s1, s2. firstName(x, s1) \land surname(x, s2) \leftrightarrow hasName(x, fn : concat(s1,s2)) \]

Constrained Attributes (naming, organization, attribute assignment)

\[ A.P_1 \equiv \sqsubseteq \sqsupseteq P_2 \]
\[ \forall x, y. hasName(x, y) \land Person(x) \leftrightarrow Name(x, y) \]
Semantic Reconciliation Rules Generation

∀x, s, s1, s2, s3.buyer_info(x), vat(x, s), address(x, s1), split(s1, s2, s3)
→
∃y.purchaser(x), address(y), hasAddress(x, y), vatNumber(x, s), streetName(y, s2), streetNumber(y, s3)

[ (?x rdf:type rs:buyer_info) (?x rs:vat ?s)
  (?x rs:address ?s1) regex(?s1, SPLIT-PATTERN, ?s2, ?s3) makeTemp(?y)
  ->
  (?x rdf:type ro:purchaser) (?y rdf:type ro:address) (?x ro:hasAddress ?y)
  (?x ro:vatNumber ?s) (?y ro:streetName ?s2) (?y ro:streetNumber ?s3)
]
Semantic Interop Runtime Engine

• Actual document reconciliation from A to B is performed applying the defined reconciliation rules

• Two transformations:
  – **Forward**, from the A format into the Ontology representation (FWD rules)
  – **Backward**, from the Ontology representation into the B format (BWD rules)
Conclusions

Semantic technologies for

• Knowledge interoperability for managing enterprise collaborative networks
• Information interoperability for business document exchange