On the use of Ontologies for Information Systems Interoperability

joint work with

Flavio Corradini, Rosario Culmone e Emanuela Merelli

Dipartimento di Matematica e Informatica
Università di Camerino
Italia

Sirolo 7-8 October 2005
Information Systems

• The information system is the component of an organization that manages (gets, processes, stores, communicates) the information of interest [Atzeni et al. 1999]
  • each organization has its own information system (protocols, terminology, processes, activities, resources, …),
  • usually, the information system operates in support to other components of the organization

• The very notion of information system is partly independent of its computerization
• The Organizational System dealing with rules and roles
e-Health scenario 1/3

Reference Architecture of openEHR (Electronic Health Record)
e-Health scenario 2/3

The Information System is responsible of

• information flows (processes)

• information resources (data)

[Ceri, 1999]
The EHR must contain the historical updated clinical information of an individual

1. The information must be complete as much as possible
   - Every clinical event, that happens in any place of the world, must be recorded
   - How to precisely describe any event?

2. The information must be accessible from each health structures placed everywhere
   - Every structure must be able to exchange clinical data
   - How to share information description?

Information systems interoperability
Information Systems Interoperability by integrating …

- Ontology
  - Explicit specification of a conceptualization. An abstract simplified view of the world represented for some purpose [Gruber '93]

- Activity-based Application (Workflow)
  - The computerized facilitation or automation of a business process, in whole or part. (from Workflow Management Coalition-Reference Model)
  - Agents are software components that:
    - embed a "complete" behaviour in their control,
    - are autonomous and react to external stimuli and environment
    - can be mobile (in the sense that they can move from a site another)

- Agent-based Middleware for mobile computing
From Data to Knowledge and vice versa

- Data + wrapper agents + common ontology + mapping tools
- Information + agent-based coordination + communication ontology
- Domain ontologies (human concepts) + workflow
- IS Interoperability through Integration
  - Meta-data
  - XML + RDF + ACL
  - Data format + schema
Hierarchy Enterprise Systems

Enterprise System
Organizational System
Information System
Computer-based system

Rules
Resources
Protocols
Roles and Responsibilities
Processes
Data
Applications
Middleware for mobile computation: motivations

- Middleware sits between the operating system and the application

  - Facilitates the development of distributed applications

  - Provides developers with abstractions, hiding details of distribution, enabling rapid, dependable development

  - Includes typical features as communication primitives, replication, concurrency management, etc.
What is HERMES?

- Hermes is a software tool (80 kb of Java code)

- Hermes supports the execution of software applications. The applications it supports can be:
  - Activity-based
  - Distributed

- Hermes provides service agents to
  - Ontology management
  - Web-services orchestration
  - Data repository wrapping
  - …
Main Aim for Developing Hermes

For the application domains we are interested in, applications must be as much as easy and intuitive possible to specify.

The programmer of applications has to concentrate on the involved activities and can ignore implementation strategies or methodological development issues.

Hence, we also demand to the (automatic) tool the bridging of the gap between an application specification and the application itself.
Hermes’ software architecture

User Application Workflow

Workflow Management

Application Agents

Application Agents Management

Service Agents

Core Level

User Layer

System Layer

Run-Time Layer
Hermes: The Agent-based Middleware
Activity-based applications for BioMedicine (O2I Project)

Use context: Stanford-Lab -- verification of proteins mutation experiment by in-silico reproducing

Goal: Retrieve abstracts from a molecular biology literature db for identifying the best cell line for reproducing a human TP53 mutation experiment linked to a particular tumour-habits-sex combination

Activities: by using Bioinformatics Services available in the net

1. Retrieve all mutations (IDs) observed in the 7th exon in men who are ex-smokers and drinkers
   • by searching p53 mutations database SRS (IST, Genova)

2. Retrieve all mutations (IDs) observed by using B9 cell line as original resource
   • by searching p53 mutations database SRS (IST, Genova)

3. Retrieve all abstracts of the correlated bibliographic references, of a specific mutation ID
   • by searching Medline (NCBI)
The Geographical Distribution

1st Activity
Retrieve all mutations (IDs) observed in the 7th exon in men who are ex-smokers and drinkers

2nd Activity
Find all mutations observed by using B9 cell line

3rd Activity
Retrieve all abstracts of the correlated bibliographic references, of a specific mutation ID

IST-Genova

NCBI

Hermes-Camerino

Research Institute MIS-Stanford
An Example of Workflow at User Level

1st Activity
Retrieve all mutations (IDs) observed in the 7th exon in men who are ex-smokers and drinkers.

2nd Activity
Find all mutations observed by using B9 cell line.

3rd Activity
Retrieve all abstracts of the correlated bibliographic references, of a specific mutation ID.

4th Activity
[Intersezione_Mutazioni = 0]

[Intersezione_Mutazioni > 0]
Activities Database for O2I

Activities in Cell Line domain

A1: Find information about the cell line named x
A2: Find all cell lines derived from a specific tumour or pathology
A3: Find all cell lines producing a specific protein
A4: Given a specific cell line, find all related bibliographic references
A5: Given a specific cell line, find all information about produced proteins

Activities in Mutation domain

B1: Find all mutations observed in a specific intron/exon in subjects with specific sex and life habits (i.e. smokers / drinkers)
B2: Find all mutations in subjects affected by a given pathology
B3: Find all subjects affected by a tumoural pathology and with a given protein mutation
B4: Find all mutations observed by using a given cell line
B5: Given a specific mutation, find all abstracts of the correlated bibliographic references

Activities in Bibliographic domain

C1: Select all abstracts of bibliographic references, whose text includes a given term
C2: ........................................

Activities in Hermes

H1: Merge in AIXO-WA
H2: ........................................
Domain Ontology
Hermes Middleware Architecture for O2I

User Layer
- User Interface
- Application Workflow
- Workflow Mng
  Editor + Compiler

System Layer
- Mobile User Agents
- Workflow Mng
- Editor + Compiler
- Wrapper Service Agent
- Matchmaker Service Agent
- WebServices Service Agent
- Ontology Mng Service Agent

Run-Time Layer
- Retrieval Service Agent
- Wrapper Service Agent
- Matchmaker Service Agent
- WebServices Service Agent
- Ontology Mng Service Agent
- Temporary Data Repository
- EMRL
- FASTA
- ASN.1
- GenBank
- RDB
- HTML
- XML
- TXT
- Local Repositories

emanuela.merelli@unicam.it
Workflow (Context-Aware) Compiler

The compiler generates a pool of cooperating agents form the workflow specification.

Two steps compilation:
• Step 1: from user level Wf to agent level Wf
  • User Level Activity Database (ULAD)
    • contains the mapping from each user activity to an agent level Wf specification

• Step 2: agent synthesis
  • Database of Skeletons (DoS)
    • contains the “empty” implementation of an agent role (skeleton)
  • Agent-Level Activity Implementation Database (ALAID)
    • contains the implementation of each agent level activity

Context-awareness
The compiler takes advantage of information about the state of the global environment during the agent synthesis.
Compiler Step 1: from User to Agent Level Workflow

Agent1 VARIABLES
- sequence: String[]
- PDBs: PDBstruct[]
- Cstructures1: Cstructare[]

select similar sequences

Agent1

select PDB structures from sequences

Agent1

Agent2 VARIABLES
- sequence: String[]
- articlesFile[]
- Cstructures2: Cstructare[]

select articles from sequences

Agent2

ULAD

emanuela.merelli@unicam.it

23
Compiler Step 2: executable application agents
Ontology Based Wrapper Agent

Mapping function
Fig. 1. An Ontological Graph for the lexicon $L_1$

An ontological graph is formalized by the following definition:

**Definition 1 (Ontological Graph).** An ontological graph $O=(N,A)$ is a directed, node and arc labelled over the lexicon $L$, graph. Where $N$ is the finite set of ontology concepts and $A$ is the finite set of relations among concepts. The node labelling function, $\lambda : N \rightarrow C$ uniquely associates a node to a concept in the lexicon. The arc labelling function $\delta : A \rightarrow R$ uniquely associates an arc to a relation in the lexicon.

3.1 Algebraic operators

- projection $\pi : O \times C \rightarrow O$

- similarity $\sigma : 2^N \times 2^N \rightarrow [0,1]$

- enriching $\sqcup : N \times N \rightarrow O$
Semantic Similarity Functions

1. **Similar by relation**: Given a description of two concepts we measure the semantic similarity of two concepts by recursively comparing the existing relations among all the concepts involved in the description.

2. **Max-Structural Match**: Given a description of two concepts we measure the semantic similarity of two concepts by recursively comparing the most similar structural matching.
Semantic Similarity

Similarity by relations

Structural similarity

emanuela.merelli@unicam.it
Ontology Functional Similarity

Given two ontological graphs $O' = (N_1, A_1)$, $O'' = (N_2, A_2)$ and two set of nodes $H \subset N_1$ and $K \subset N_2$, the similarity between $H$ and $K$ is measured by the following function

$$f(H, K) = \begin{cases} 
1 & \text{iff} \quad \tau(H) = \tau(K) = 0 \\
0 & \text{iff} \quad (H = \emptyset \text{ xor } K = \emptyset) \text{ or } (\tau(H) = 0 \text{ xor } \tau(K) = 0) \\
\sum_{r \in R_{H,K}} f(\text{prj}_r(H), \text{prj}_r(K)) & \text{otherwise} \\
\frac{\text{card}(R_{H,K})}{\text{card}(R_{H,K})} & \text{otherwise}
\end{cases}$$

Where $\tau(N)$ is the number of outgoing arcs from the set of nodes in $N$; $R_{H,K}$ is the set of relation associated to the arcs outgoing from nodes in $H$ and $K$; $\text{prj}_r(H)$ and $\text{prj}_r(K)$ is the set of nodes reachable from any node in $H$ or in $K$ respectively through the relation $r \in R$.  

emanuela.merelli@unicam.it
Main Publications


- F. Corradini, E. Merelli, M. Vita. A Multi-Agent System for Modelling Carbohydrate Oxidation in Cell. First International Workshop On Modelling Complex Systems, MCS’05, LNCS, 2005

- F. Corradini, R. Culmone, E. Merelli. Integrating mobile agent and ontologies in a distributed environment, working paper.

- F. Corradini, E. Merelli, M. Ruffino Ontology-drive Information System Architecture in the Knowledge Management Perspective, working paper
Hermes Software

- Downloadable from

  http://hermes.cs.unicam.it
  or
  http://www.bioagent.net

Thanks to the audience!
Domain Ontology Description: functional role
Communication Ontology: MM and Environment

<<Communication>>
AgMatriceMitocondriale - AgStatoAttuale1
Ontology: Gruppo Dati OPP
Language: XML
Protocol: FIPARequest

<<Communication>>
AgMatriceMitocondriale - AgStatoAttuale2
Ontology: Gruppo Dati Ciclo Krebs
Language: XML
Protocol: FIPARequest

<<Communication>>
AgMatriceMitocondriale - AgStatoAttuale3
Ontology: Gruppo Dati OPP
Language: XML
Protocol: FIPARequest

<<Communication>>
AgMatriceMitocondriale - AgStatoAttuale4
Ontology: Gruppo Dati Ciclo Krebs
Language: XML
Protocol: FIPARequest

<<Communication>>
AgStatoAttuale - AgMatriceMitocondriale1
Ontology: Gruppo Dati OPP
Language: XML
Protocol: FIPARequest

<<Communication>>
AgStatoAttuale - AgMatriceMitocondriale2
Ontology: Gruppo Dati Ciclo Krebs
Language: XML
Protocol: FIPARequest

<<Communication>>
AgStatoAttuale - AgMatriceMitocondriale3
Ontology: Gruppo Dati OPP
Language: XML
Protocol: FIPARequest

<<Communication>>
AgStatoAttuale - AgMatriceMitocondriale4
Ontology: Gruppo Dati Ciclo Krebs
Language: XML
Protocol: FIPARequest

<<Communication>>
AgStatoAttuale - AgMatriceMitocondriale
Ontology: Gruppo Dati OPP
Language: XML
Protocol: FIPARequest

<<Communication>>
AgStatoAttuale - AgMatriceMitocondriale2
Ontology: Gruppo Dati Ciclo Krebs
Language: XML
Protocol: FIPARequest

<<Communication>>
AgStatoAttuale - AgMatriceMitocondriale3
Ontology: Gruppo Dati OPP
Language: XML
Protocol: FIPARequest

<<Communication>>
AgStatoAttuale - AgMatriceMitocondriale4
Ontology: Gruppo Dati Ciclo Krebs
Language: XML
Protocol: FIPARequest

<<Agent>>
AgMatriceMitocondriale
Ossidazione Piruvato: Gruppo Dati OPP
Ciclo di Krebs: Gruppo Dati Ciclo Krebs
Fase: Fase

<<Agent>>
AgStatoAttuale
Dati Condivisi: Gruppo Dati Condivisi

emanuela.merelli@unicam.it