Semantically aware agents

Luciano Serafini
ITC-IRST
Trento, Italia

Thanks to:

Paolo Bouquet
(DIT-Uni. Trento)

Stefano Zanobini
(DIT-Uni. Trento)

Simone Sceffer
(ITC-IRST)

Bernardo Magnini
(ITC-IRST)

Antonia Dona'
(Distributed Thinking)

Laura Olivetti
(Distributed Thinking)

Manuela Speranza
(ITC-IRST)

Christian Girardi
(ITC-IRST)

and many other who actvlly collaborated in the project:

edamok

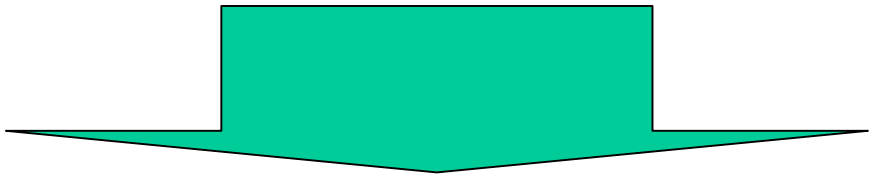
enabling distributed and autonomous **management of knowledge**

The dream

- A semantically aware agent is capable to interact with other semantically aware agents, via meaningful communication, without a predefined strong agreement on the underlying communication language.

Pre-defined agreement

- Offline agreement on the syntax of the agent communication language
- Offline agreement on the semantics of the agent communication language



Common shared (set of) ontologies

- Agent communication is limited only by the shared ontology

Hidden Semantics

- (semi) structured data contains a implicit semantics which allows one to “understand” data, i.e., to provide a semantics to data.

Student_ID	**Name**	**Course**	**Mark**
A23	Paolo Bouquet	Database 1	30
B34	Stefano Zanobini	Knowledge representation	27
C37	Simone Sceffer	Telecomunication	33

To interpret data, hidden semantics has to be made explicit

The unique identifier for a student

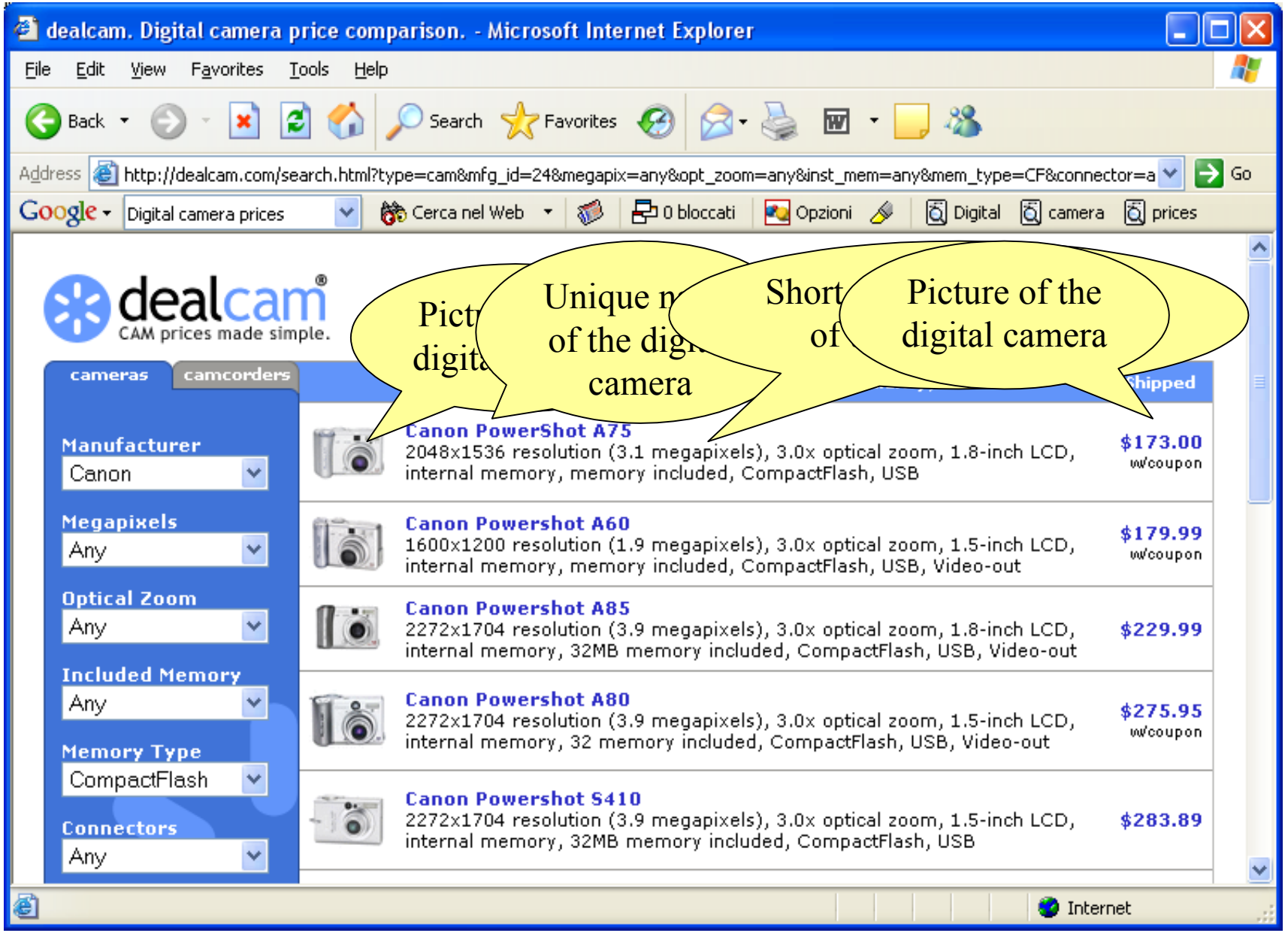
His/her name (first and last name)

One of the course (s)he is/has been enrolled

The mark (s)he obtained

Student_ID	Name	Course	Mark
A23	Paolo Bouquet	Database 1	30
B34	Stefano Zanobini	Knowledge representation	27
C37	Simone Sceffer	Telecomunication	

A more complex example



Humans do a lot of semantic explicitation

- Pictures and graphical elements
- Position and structure of the data
- Context
- Natural language understanding
- Common sense knowledge
- Domain knowledge,
- …

THIS IS TOO MUCH FOR AN AUTONOMOUS AGENT

Agents interact via (semi)structured data and schema

```
<book ISBN="1234">
   <author>
        <first name> Stefano </first name>
        <last name> Benni </last name>
   </author>
   <title> Bar Sport </title>
   <price currency="Euro">20</price>
</book>
```

Making explicit the semantic hidden in (semi)-structural data

- Natural language (The labels of XML schema are meaningful expressions of the natural language + some new syntax (e.g., space is replaced by capital letters)
- The schema structure (e.g., the tag <A> nested in the tab <B> describe an attribute of an element of type <B>)
- Commonsense and domain specific knowledge (e.g., books must have an ISBN and usually have at least one author)
- Context (e.g., if an xml record is returned by a e-commerce web service, then “price” stands for the selling price)

Natural Language

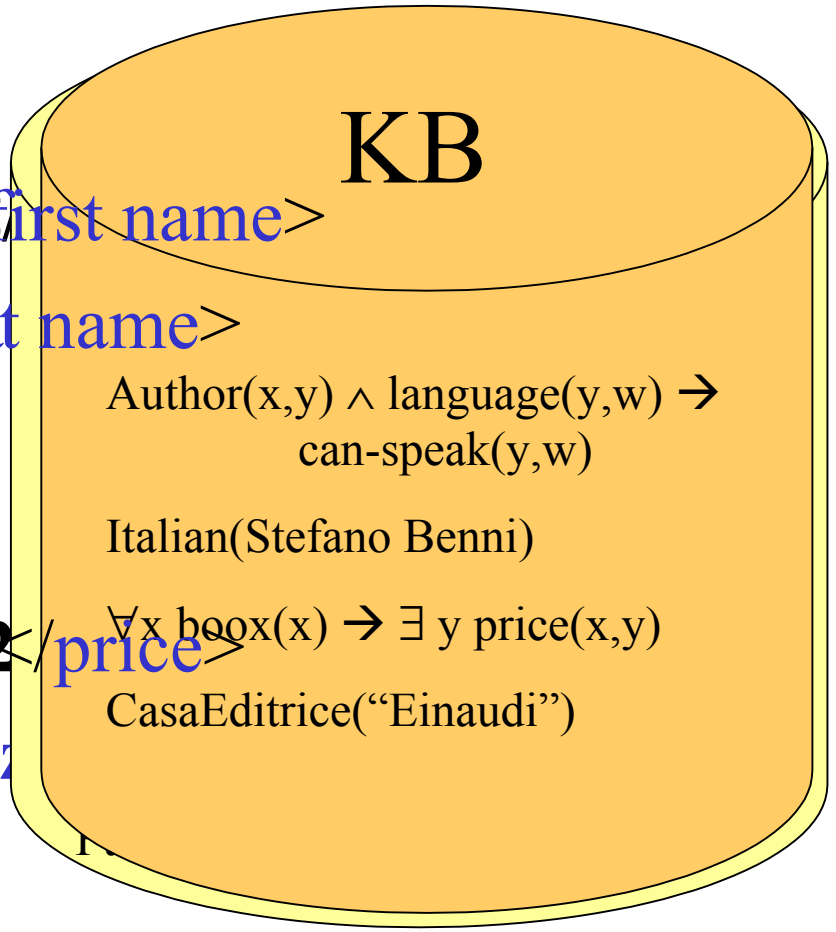
```
<book ISBN="12345">
  <author>
    <first name>Stefano</first name>
    <last name>Bernini</last name>
  </author>
  <title> Bar Sport </title>
  <price currency="Euro">20</price>
</book>
```

Structure

```
<book ISBN="1234">
    <author>
        <first name> Stefano </first name>
        <last name> Benni </last name>
    </author>
    <book ISBN="1234">
        <title> Bar Sport </title>
        <price currency="Euro"> 20 </price>
</book>
```

Domain knowledge

```
<book ISBN="12234">
   <author>
        <first name>Stefano</first name>
        <last name>Benni</last name>
   </author>
   <title> Bar Sport</title>
   <price currency="Euro">202</price>
</sgruntz>Einaudi</sgruntz
</book>
```



Context

sra.itc.it/people/serafini/mybooks.xml

```
<book ISBN="1234">
    <author>
        <first name> Stefano </first name>
        <last name> Benni </last name>
    </author>
    <title> Bar Sport </title>
    <price currency="Euro">20</price>
</book>
```

Problem

- How to represent the semantics made explicit by using linguistic, structural, contextual and domain knowledge?
- This is a good task for logic
  - Its enough expressive to encode explicited semantics
  - Its meaning (its semantics) is commonly accepted
  - Support fast reasoning procedure

Which logic?

- Suggestions:
  - **Propositional logic**, for simple cases
  - **Description Logics** for more comples static and taxonomic knowledge (e.g., db-schema, xml-data schema)
  - **Temporal/dynamic logic**, for knowledge about dataflow and actions (e.g., descriptions of web services)

Logic is not enough

A formula

$$\forall x\,(P(x) \rightarrow \exists y(R(x,y) \wedge R(y))$$

does not carry much semantics. (P, Q, and R, do not make any sense to me, … and perhaps to any agent too)

We need to provide a set of meaningful (non logical) primitive symbols (constant, function and predicate)

WORDNET

- Wordnet is a DB of senses which are anchored to words
- For every (english/italian …) word <w>, Wordnet provides a set of senses

  <word>#1, word#2,…,<w>#N
- one for each sense in which <w> can be used
- Plus for each sense
  - a gloss (which explains in NL the use of this sens)
  - a isa, part-of hierarchical organization of the senses.
  - Domain information
  - And some other generic relations between sense
- WORDNET is a widely shared and commonly acc epted tool, in the NL community

Logic + WORDNET

$\forall x\ (book\#1(x) \rightarrow \exists\ y\ (author\#2(x,y) \wedge person\#1(y))$

(FOL)

$book\#1 \subseteq \exists author\#2.person\#1$

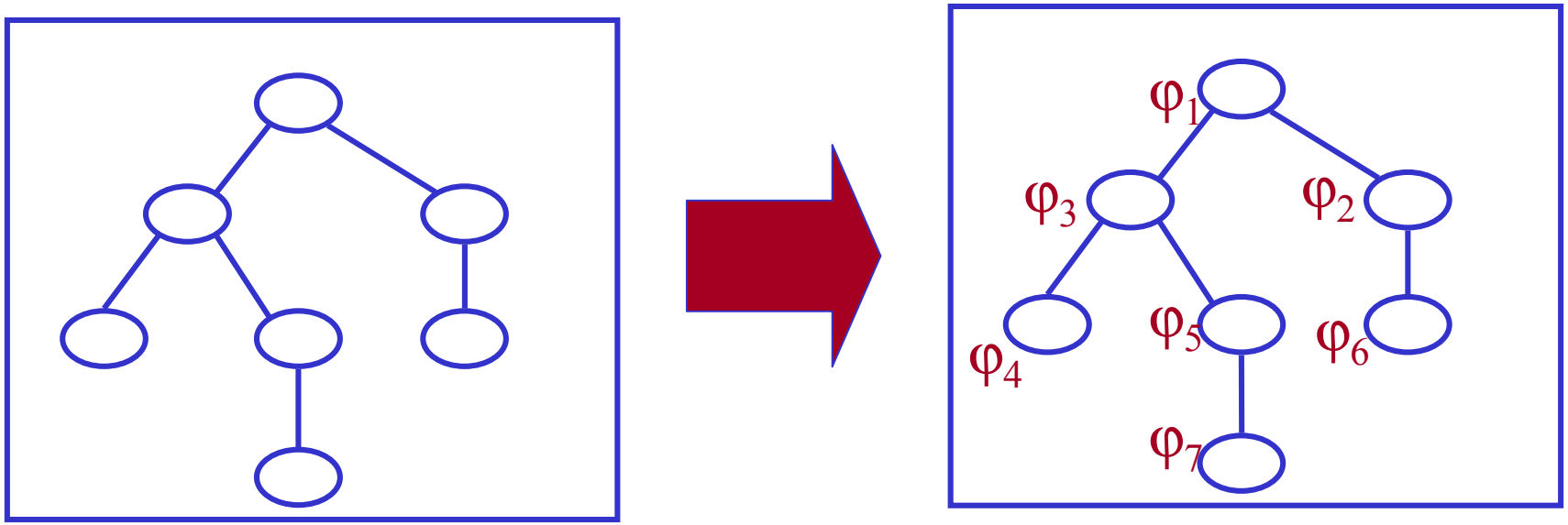
(DL)

Is the concrete representation of the semantics:

Every book is written by somebody who is a person

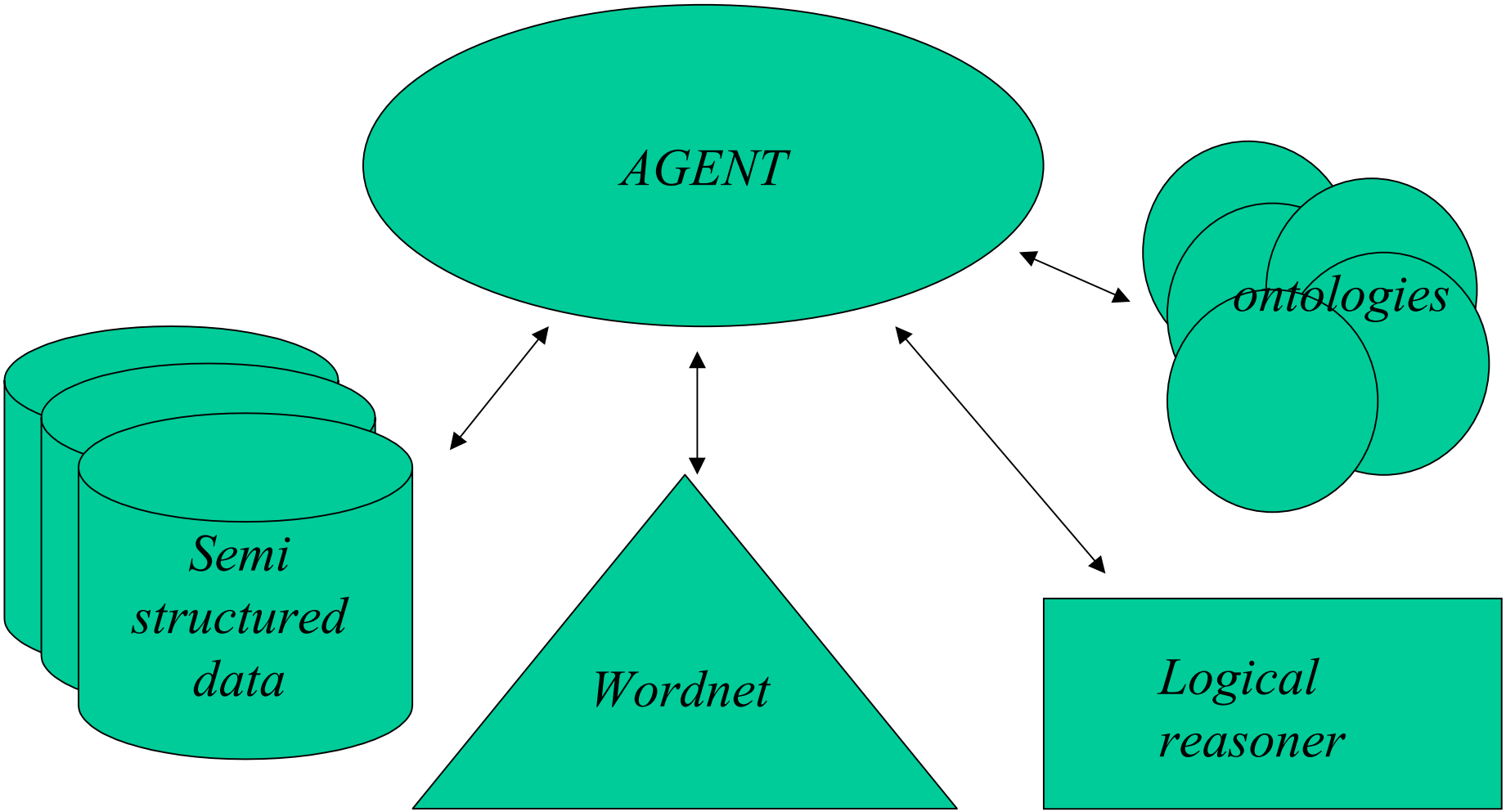
Recalling (and making more precise) the main objective

- enabling agents to compute the semantics (i.e., the logical form) of each component of a given schema




ISBN#2 ∧ ∃book-identifier.book#2
book#2
∃author#1-1.book#2
<book ISBN="1234">
<author>
<first name> Stefano </first name>
<last name> Benni </
∃firstname#1-1.∃author#1-
</author>
<title> Bar Sport </title>
<price currency="Euro">20</price>
</book>

Semantically aware agent architecture



The problem of semantic coordinatio

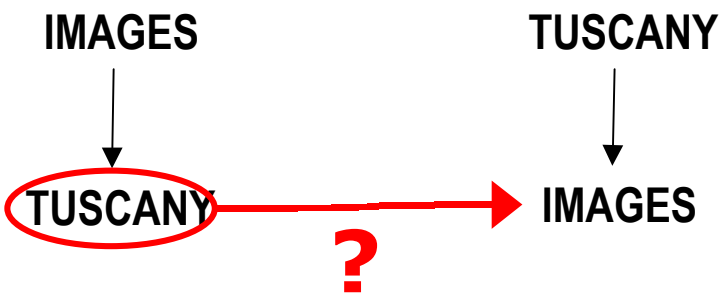
A *semantic coordination problem* is defined as a situation in which:

- two (or more) parties have an interest in finding an agreement on the **semantic relations** between their local models,

BUT

- there are multiple possible semantic relations (i.e., mappings across local models) among which they need to select the right (or a good enough) one.

- We consider the case of semantic coordination that allows agents to share hierarchical classifications of documents



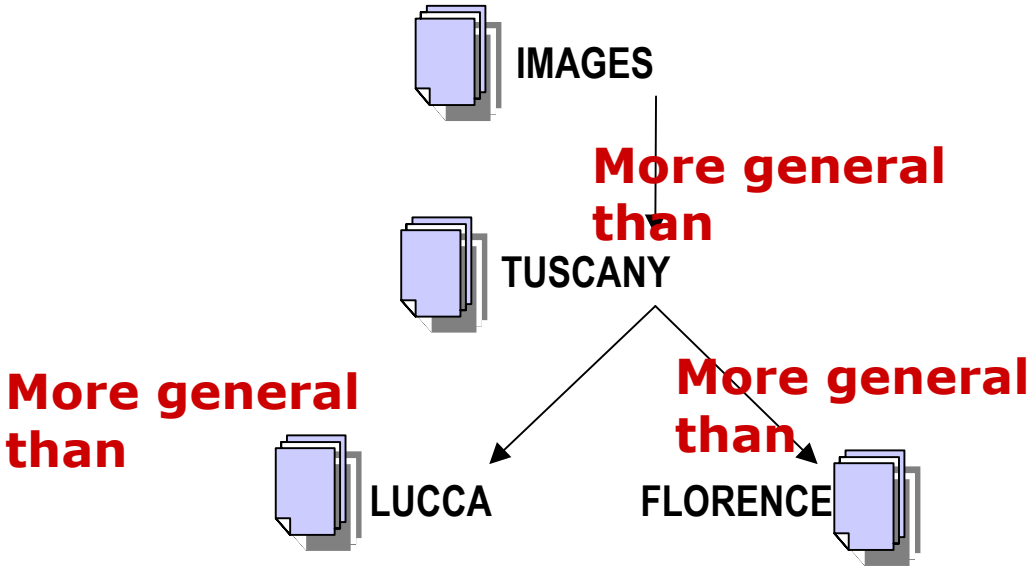
EQUIVALENT?

MORE GENERAL?

LESS GENERAL?

DISJOINT?

A case study:
Coordinating hierarchical classificat

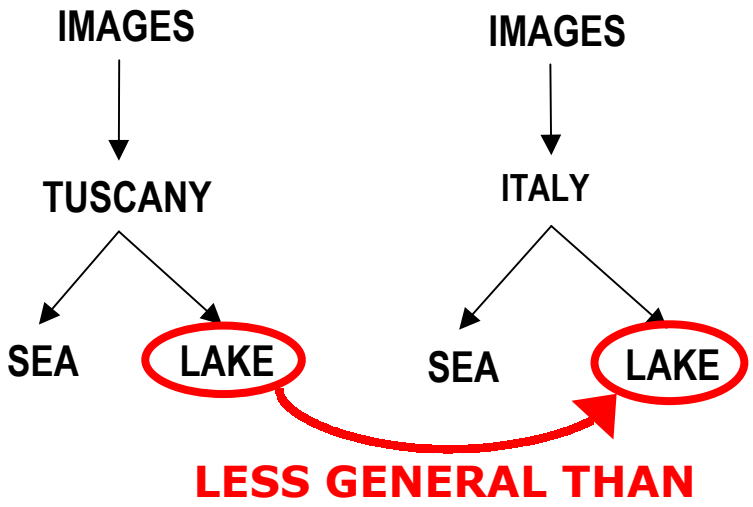


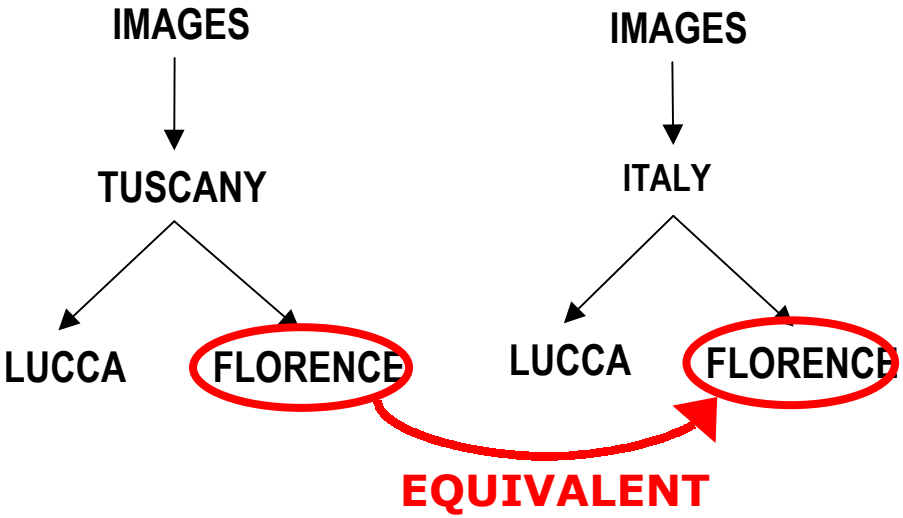
Instances of HCs:

- Web directories (Google, Yahoo, Looksmart, ….)
- File systems (folders structures)
- Content management structures
- Portal structures

edamok

Graphs is not enough – Part I





ITC irst

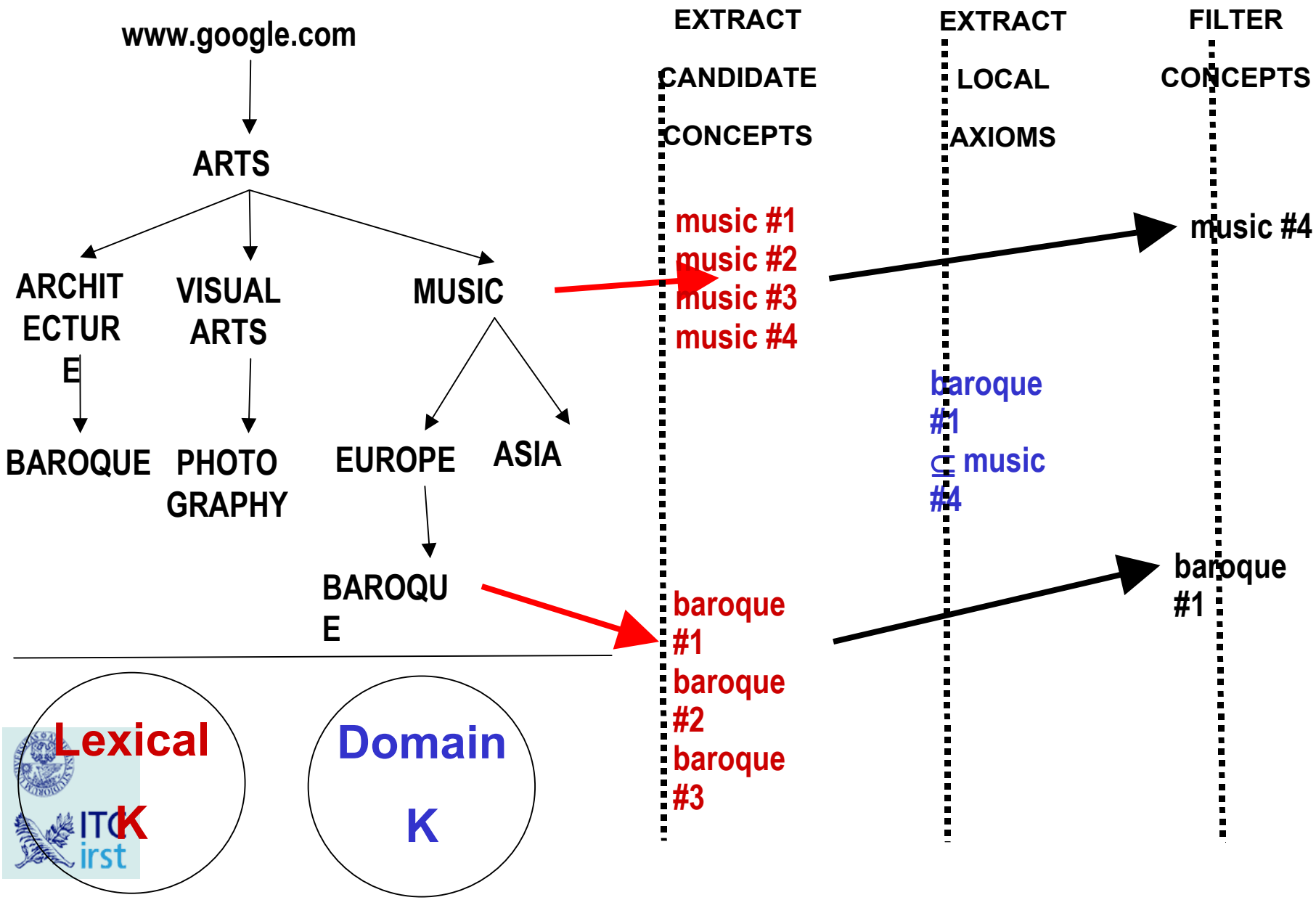
- ***SEMANTIC ESPLICITATION***

  First we compute the hidden semantics for each node of the hierarchical classifications

- ***SEMANTIC COMPARISON***

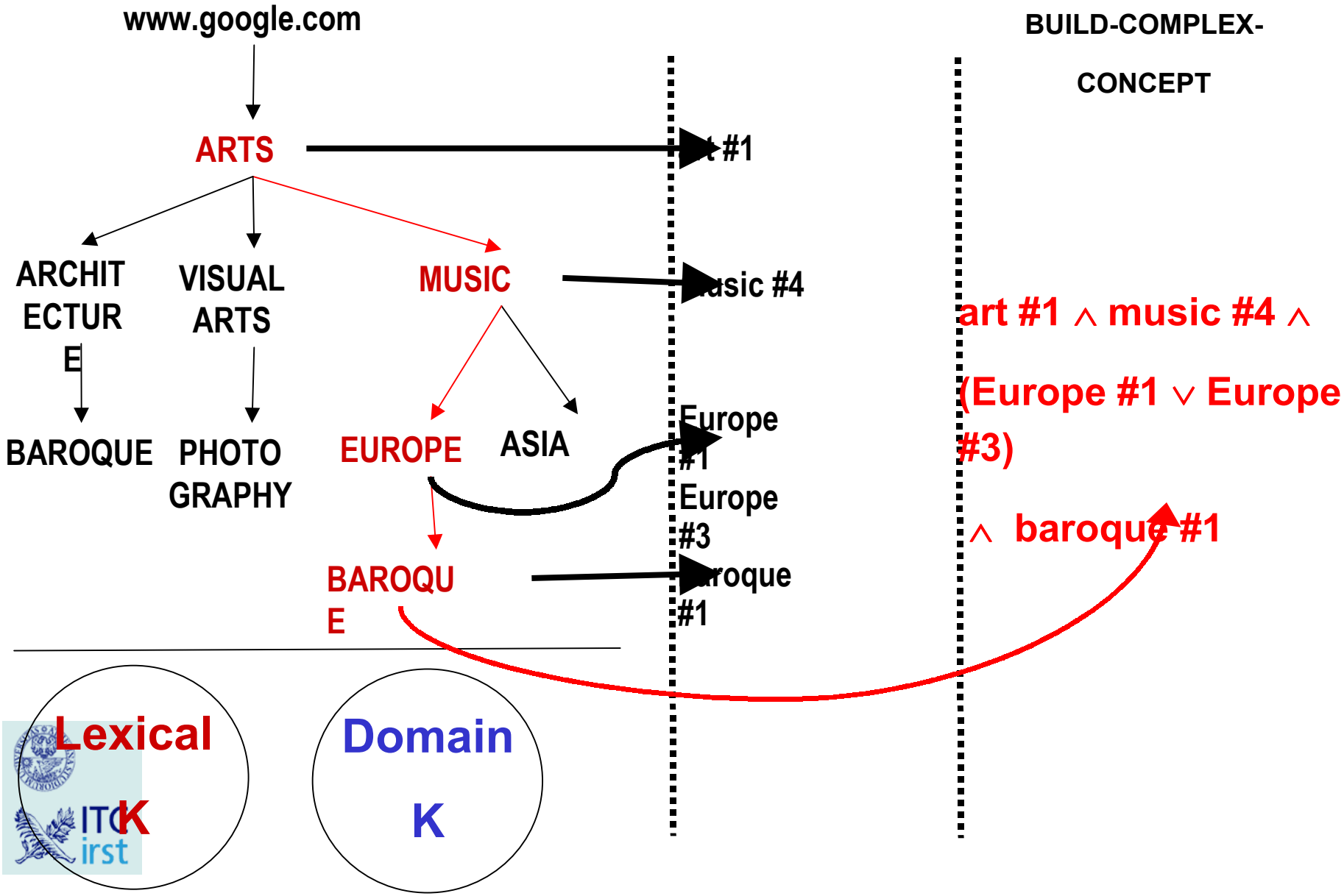
edamok

Semantic explicitation - I



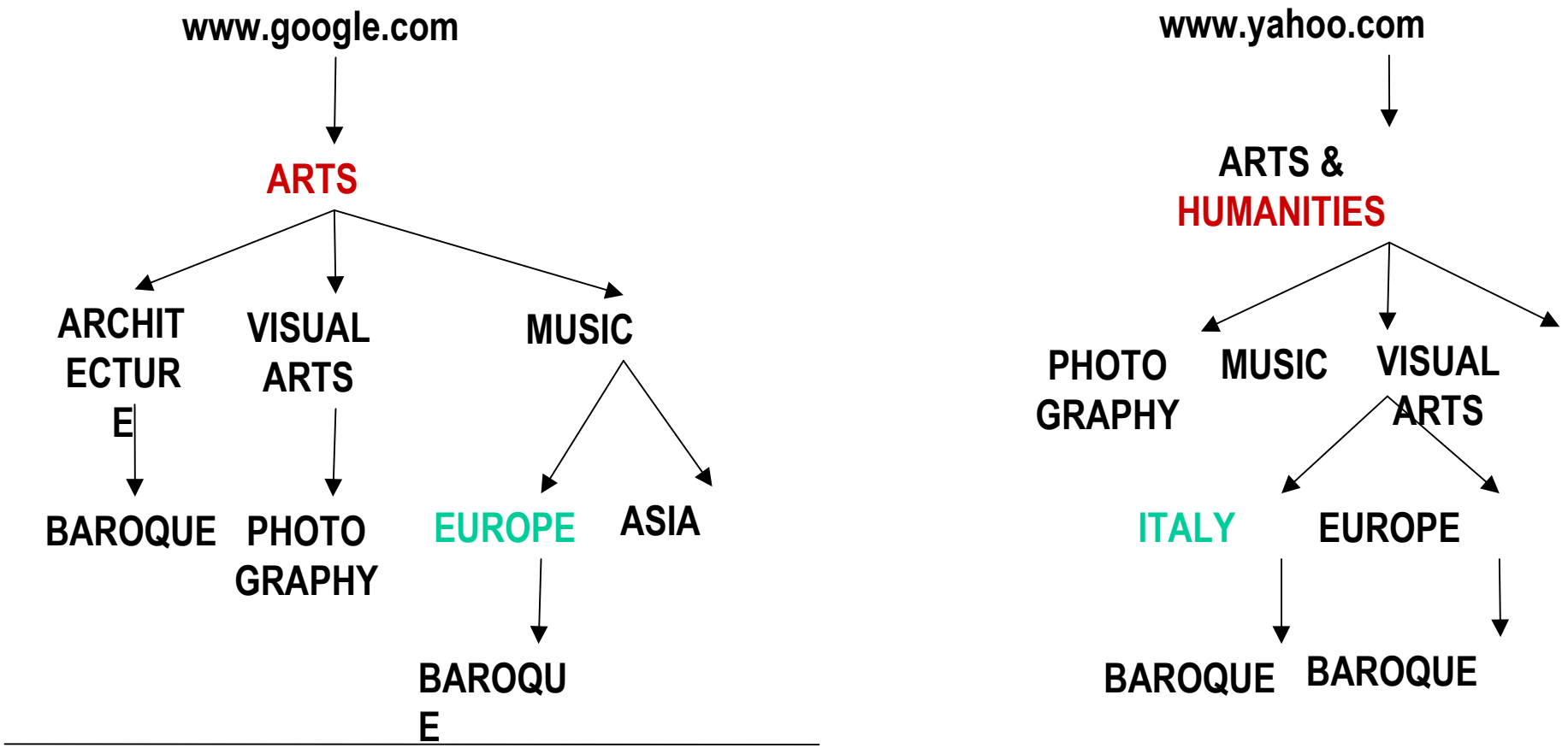
edamok

Semantic explicitation - II



edamok

Semantic comparison - I



ITC irst

Domain K

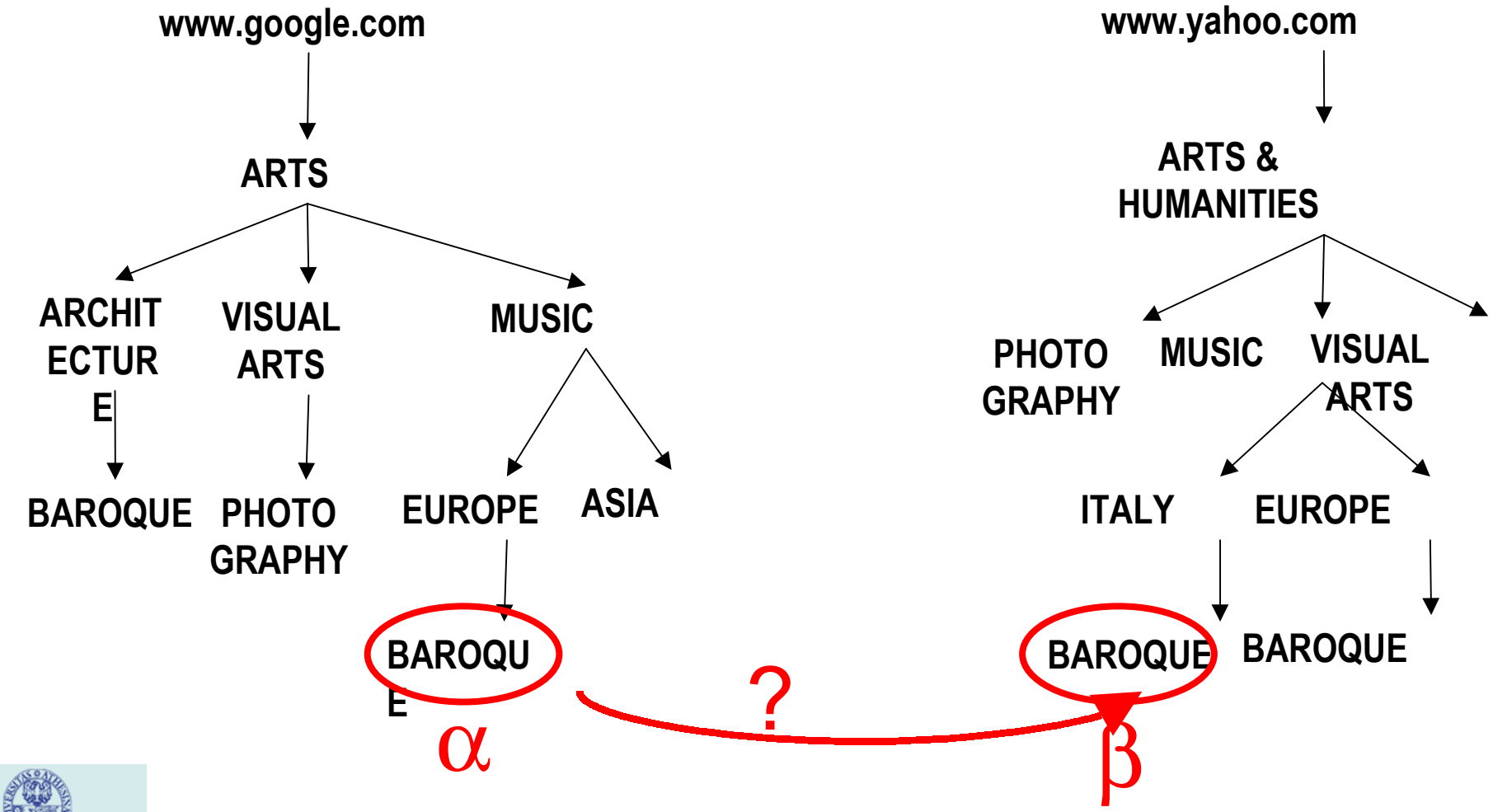
B =

music#4 → art#1

Italy#1 → Europe#3

edamok

Semantic comparison - II



ITC irst

edamok

Semantic comparison - III

DISJOINT? $B \models \neg (\beta \wedge \alpha)$? → no

EQUIVALENT? $B \models \beta \equiv \alpha$? → no

LESS GENERAL? $B \models \alpha \rightarrow \beta$? → no

MORE GENERAL? $B \models \beta \rightarrow \alpha$? → yes

Where: β = ((art#1 $\vee$ humanities#3) $\wedge$ music#4 $\wedge$ Italy#1 $\wedge$ baroque#1)

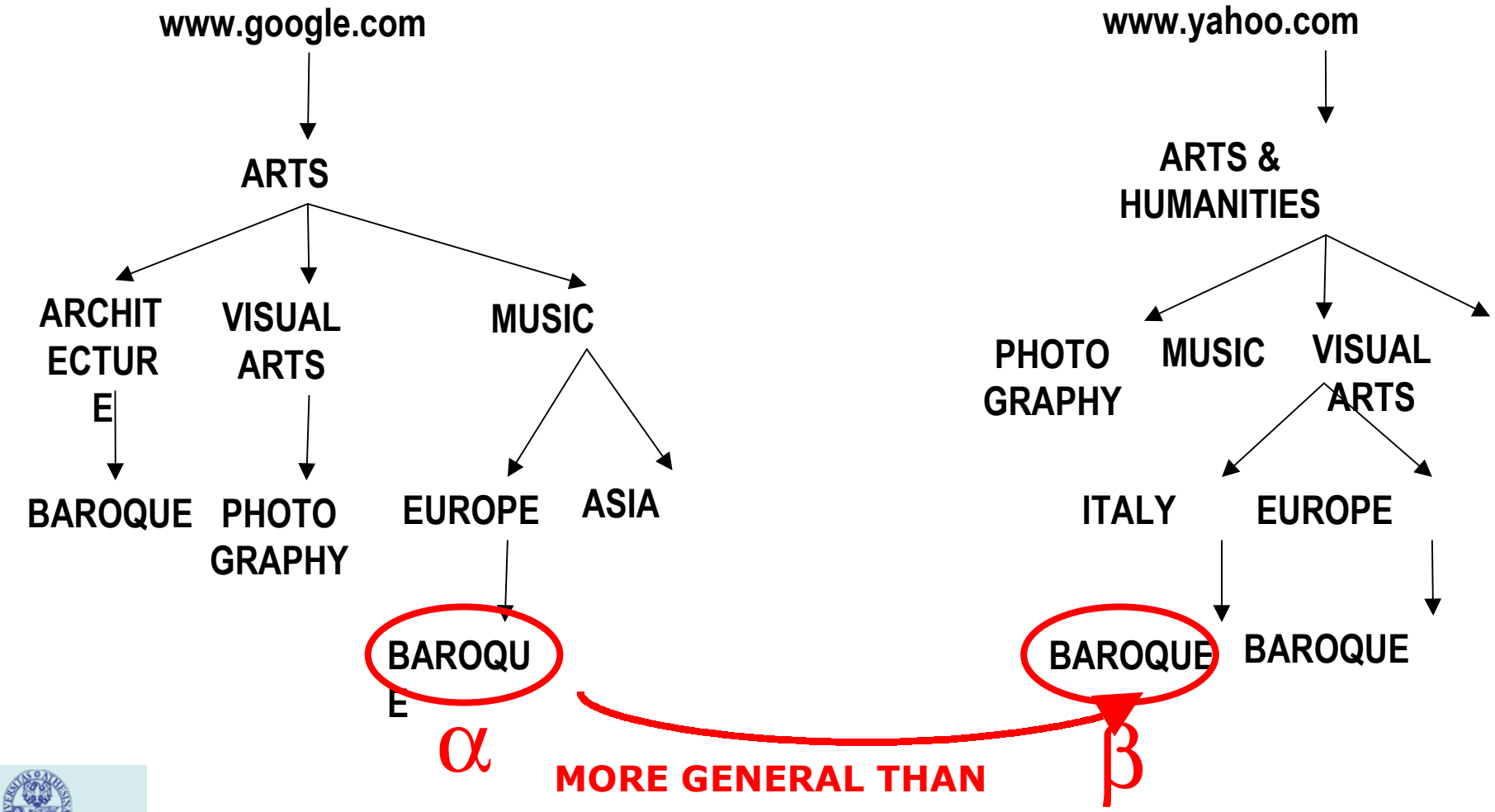
α = (art#1 $\wedge$ music#4 $\wedge$ (Europe#1 $\vee$ Europe#3) $\wedge$ baroque#1)

B = (Music#4 $\rightarrow$ Art#1) $\wedge$ (Italy#1 $\rightarrow$ Europe#3)

ITC irst

edamok

Outcome



ITC irst

DEMO

Sharing documents between Harry and Jerry

Two semantically aware agents

Conclusions

- *Interoperability can be reached also by improving the semantics awareness of the autonomous applications (agents)*
- *Natural language processing and ontological reasoning are combined in a nice and new way to compute hidden semantics*
- *The proposal is only partial, We don't consider data instance (Future work → integration with machine learning)*