UNIVERSITÀ DEGLI STUDI DI TRENTO

Managing diversity in Knowledge

Fausto Giunchiglia

ECAI 2006, Riva del Garda, Trento

To be cited as: Fausto Giunchiglia, "Managing Diversity in Knowledge", Invited talk, ECAI 2006. Online presentation, reachable from http://www.disi.unitn.it/~fausto/knowdive.ppt

Outline

- The problem: the complexity of knowledge
- The solution: managing diversity
- Some early work
- Three core issues

Managing knowledge (and data)

The "standard" Approach:

- Take into account, at ***design time***, the future dynamics.
- Design a ***"general enough"*** representation model, able to incorporate the future knowledge variations.
- Most commonly: design a ***global*** representation schema and codify into it the ***diverse*** knowledge components.

Examples: Relational and distributed databases, federated databases, ontologies, knowledge bases, data bases in the Web (information integration), …

Why the current approach?

- It is conceptually "simple"
- It has been successfully and extensively used in the past
- There is a lot of know-how
- It works well also in "controlled" (not too) open applications
- It satisfies the companies' desire to be in control of their data
- It is reassuring: it is "easy" to establish *right* … and *wrong*
- It is deeply rooted in our logical and philosophical tradition

… it should be used as much as possible!

However…
Ex. 1: business catalogs (~ 10^4 nodes)

UNSPSC

- Top
  - Industrial Manufacturing and Processing Machinery and Accessories
    - Lapidary machinery and equipment
    - Leatherworking repairing machinery and equipment
    - Industrial process machinery and equipment and supplies
      - Separation machinery and equipment
      - Cutting tools
        - Drills
        - Reamer cutting tool
        - Form tools or toolbits
        - Taps or dies
        - Broach cutting tool
        - Gear cutting tools
        - Rotary burrs
        - Regrind or reclaim or coating services for cutting tools
        - Countesink tool or counterbore tool
        - Machinery cutting knives or knife assemblies
      - Assembly machines
      - Paint systems
    - Foundry machines and equipment and supplies
    - Workshop machinery and equipment and supplies

eCl@ss

- Top
  - Machine, apparatus
    - Heat exchanger
    - Boiler, furnace
    - Sterilizer
    - Cleaning installation
    - Sound damper, pulsation damper
    - Cutting machine
      - Plasma cutting machine
      - Cutting machine (other)
        - shears (manufacturing of glass)
        - melt machine (manufacturing of glass)
      - Cutting machine (parts)
      - Cutting mach. (maint., serv.)
      - Cutting mach. (repair)
    - Textile machine
    - Pressure machine

The problem: the complexity of knowledge

- ***Size***: **the sheer numbers – a huge increase in the number of *knowledge producers* and *users*, and in their *production/use capabilities***
- ***Pervasiveness***: **knowledge, producers, users pervasive in space and time**
- ***Time unboundedness*** **- two aspects:**
  - **knowledge continuously produced, with no foreseeable upper bound.**
  - ***Eternal Knowledge***: **produced to be used indefinitely in time (e.g. my own family records, cultural heritage)**
- ***Distribution***: **knowledge, producers and users very sparse in distribution, with a spatial and a temporal distribution**

The core issue: knowledge diversity

- ***Diversity:*** **unavoidable … in knowledge, producers and users**

- ***Dynamics (of diversity)*****: new and old knowledge, often referenced by other knowledge, will (dis)appear virtually at any moment in time and location in space.**

- **Unpredictability (of the dynamics of diversity): the future dynamics of knowledge unknown at design *and* run time.**

Semantic heterogeneity

- Two (data, content or knowledge) items are ***semantically heterogeneous*** when they are diverse, still being a representation of the same phenomenon (example: 1Euro, 1.25$)
- The ***semantic heterogeneity problem*** is an instance of the problem of diversity

Semantic heterogeneity and diversity: business catalogs

UNSPSC

- Top
  - Industrial Manufacturing and Processing Machinery and Accessories
    - Lapidary machinery and equipment
    - Leatherworking repairing machinery and equipment
    - Industrial process machinery and equipment and supplies
      - Separation machinery and equipment
      - Cutting tools
        - Drills
        - Reamer cutting tool
        - Form tools or toolbits
        - Taps or dies
        - Broach cutting tool
        - Gear cutting tools
        - Rotary burrs
        - Regrind or reclaim or coating services for cutting tools
        - Countesink tool or counterbore tool
        - Machinery cutting knives or knife assemblies
      - Assembly machines
      - Paint systems
    - Foundry machines and equipment and supplies
    - Workshop machinery and equipment and supplies

eCl@ss

- Top
  - Machine, apparatus
    - Heat exchanger
    - Boiler, furnace
    - Sterilizer
    - Cleaning installation
    - Sound damper, pulsation damper
    - Cutting machine
      - Plasma cutting machine
      - Cutting machine (other)
        - shears (manufacturing of glass)
        - melt machine (manufacturing of glass)
      - Cutting machine (parts)
      - Cutting mach. (maint., serv.)
      - Cutting mach. (repair)
    - Textile machine
    - Pressure machine

Outline

- The problem: the complexity of knowledge
- **The solution: managing diversity**
- Some early work
- Three core issues

A paradigm shift: Managing diversity in knowledge

Consider *diversity* as a *feature* which must be maintained and exploited (at run-time) and not as a *defect* that must be absorbed (at design time).

A paradigm shift

- ***FROM:*** knowledge assembled by the *design-time combination* of basic building blocks. Knowledge produced *ab initio*
- ***TO:*** knowledge obtained by the *design* and *run-time adaptation* of existing building blocks. Knowledge no longer produced *ab initio*

New methodologies **for knowledge representation and management**

- design of (*self*-) *adaptive* knowledge systems
- develop methods and tools for the management, control and use of *emergent* knowledge properties

Handling diversity - Step 1: design knowledge to be "local"

- ***FACT 1:*** **Acknowledge that complexity and unpredictable dynamics are such that we can *only* build *local knowledge,* satisfying some set of local goals (though as broad as possible). This knowledge defines a viewpoint, a partial theory of the world**
- ***GOAL:*** **Design local knowledge which is *optimal* for the goals it is meant to achieve [[Diversity is a *feature*! *…* the WWW is not an "*implementational mistake"*]]**
- ***ACTION:*** **Implement local knowledge as a suitable *local theory*.**

A toy example – 2

Two local theories …

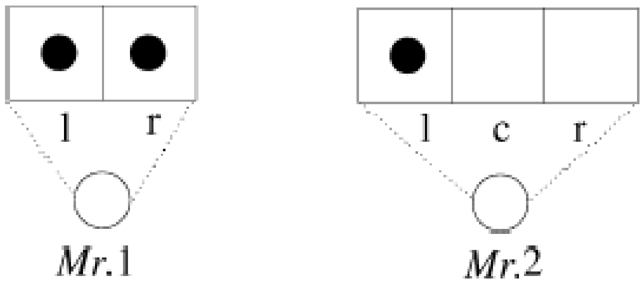


Figure 2: *Mr*.1 and *Mr*.2's contexts.

… and the world

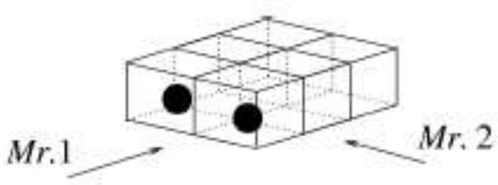


Figure 1: The magic box

A real world example: Business catalogs (contexts)

UNSPSC

- Top
  - Industrial Manufacturing and Processing Machinery and Accessories
    - Lapidary machinery and equipment
    - Leatherworking repairing machinery and equipment
    - Industrial process machinery and equipment and supplies
      - Separation machinery and equipment
      - Cutting tools
        - Drills
        - Reamer cutting tool
        - Form tools or toolbits
        - Taps or dies
        - Broach cutting tool
        - Gear cutting tools
        - Rotary burrs
        - Regrind or reclaim or coating services for cutting tools
        - Countesink tool or counterbore tool
        - Machinery cutting knives or knife assemblies
      - Assembly machines
      - Paint systems
    - Foundry machines and equipment and supplies
    - Workshop machinery and equipment and supplies

eCl@ss

- Top
  - Machine, apparatus
    - Heat exchanger
    - Boiler, furnace
    - Sterilizer
    - Cleaning installation
    - Sound damper, pulsation damper
    - Cutting machine
      - Plasma cutting machine
      - Cutting machine (other)
        - shears (manufacturing of glass)
        - melt machine (manufacturing of glass)
      - Cutting machine (parts)
      - Cutting mach. (maint., serv.)
      - Cutting mach. (repair)
    - Textile machine
    - Pressure machine

Which world? How much of it?

Handling diversity – Step 2: knowledge sharing via interoperabilty

- ***FACT:*** **Acknowledge that we are bound to have *multiple diverse* theories of the world (and also of the same world phenomena)**
- ***GOAL:*** **Make the local theories *semantically interoperable* and exploit them to build solutions to "global" problems (e.g. eBusiness, knowledge sharing)**
- ***ACTION:*** **Implement semantic interoperability via *semantic mappings (context mappings)* between local theories.**

A real world example - more: Partial agreement between catalogs

- Top
  - Industrial Manufacturing and Processing Machinery and Accessories
    - Lapidary machinery and equipment
    - Leatherworking repairing machinery and equipment
    - Industrial process machinery and equipment and supplies
      - Separation machinery and equipment
      - Cutting tools
        - Drills
        - Reamer cutting tool
        - Form tools or toolbits
        - Taps or dies
        - Broach cutting tool
        - Gear cutting tools
        - Rotary burrs
        - Regrind or reclaim or coating services for cutting tools
        - Countesink tool or counterbore tool
        - Machinery cutting knives or knife assemblies
      - Assembly machines
      - Paint systems
    - Foundry machines and equipment and supplies
    - Workshop machinery and equipment and supplies

- Top
  - Machine, apparatus
    - Heat exchanger
    - Boiler, furnace
    - Sterilizer
    - Cleaning installation
    - Sound damper, pulsation damper
    - Cutting machine
      - Plasma cutting machine
      - Cutting machine (other)
        - shears (manufacturing of glass)
        - melt machine (manufacturing of glass)
      - Cutting machine (parts)
      - Cutting mach. (maint., serv.)
      - Cutting mach. (repair)
    - Textile machine
    - Pressure machine

Ex.: <Id, Drills, Cutting machine (other), subsumes>

Handling diversity – Step 3: knowledge sharing via adaptivity

- ***FACT:*** **Acknowledge that in most cases straight interoperability *will not work* due the different goals and requirements**

- ***GOAL:*** **Make the local theories and context mappings *adaptive* and adapt them as needed at any new use**

- ***ACTION:*** **Implement (partial) adaptivity as a set of (meta)-data: *implicit assumptions***

A real world example - more: The two catalogs' implicit assumptions

- Top
  - Industrial Manufacturing and Processing Machinery and Accessories
    - Lapidary machinery and equipment
    - Leatherworking repairing machinery and equipment
    - Industrial process machinery and equipment and supplies
      - Separation machinery and equipment
      - Cutting tools
        - Drills
        - Reamer cutting tool
        - Form tools or toolbits
        - Taps or dies
        - Broach cutting tool
        - Gear cutting tools
        - Rotary burrs
        - Regrind or reclaim or coating services for cutting tools
        - Countesink tool or counterbore tool
        - Machinery cutting knives or knife assemblies
      - Assembly machines
      - Paint systems
    - Foundry machines and equipment and supplies
    - Workshop machinery and equipment and supplies

- Top
  - Machine, apparatus
    - Heat exchanger
    - Boiler, furnace
    - Sterilizer
    - Cleaning installation
    - Sound damper, pulsation damper
    - Cutting machine
      - Plasma cutting machine
      - Cutting machine (other)
        - shears (manufacturing of glass)
        - melt machine (manufacturing of glass)
      - Cutting machine (parts)
      - Cutting mach. (maint., serv.)
      - Cutting mach. (repair)
    - Textile machine
    - Pressure machine

Implicit assumptions:

<Focus = Tools and process>

<Area = Mechanical Eng.> ...

<Focus= tools>

<Area= Engineering> ...

Implicit assumptions

- Data and knowledge depend on many, unstated, ***implicit assumptions*** (goals, local state of affairs, time, location, …)
- Implicit assumptions are ***indefinitely many***, but finite in any moment in time
- ***Only some*** implicit assumptions can be memorized and/ or reconstructed
- ***Adaptivity*** is (partially) obtained by providing the means to represent implicit assumptions, to reason about them (add, modify, learn, …), and to use them to adapt local knowledge

A knowledge system

A knowledge system (component) is a 4- tuple:

< id, Th, M, IA >

Where:

- **Id: *unique* identifier**
- **Th: Theory – it codifies, in a proper *local* representation formalism, the *local* knowledge of the world**
- **M: a set of mappings – they codify the *semantic relation* existing between (elements of) local theories.**
- **IA: a finite but unbound set of assertions, written in some local metalanguage – they allow for the representation of *implicit assumptions***

Outline

The problem: the complexity of knowledge

- The solution: managing diversity
- **Some early work:** reusing, sharing, adapting language (ontologies) in the Web
  - C-OWL: Representing semantic mappings [Bouquet, Giunchiglia et al., ISWC'03, book in Spring 2007]
  - Semantic Matching: Discovering semantic mappings
  - Open Knowledge: Exploiting local theories and semantic mappings
- Three core issues

C-OWL: Contextual Ontologies

Contextual ontology = Ontology + Context mappings

Key idea:

1. **Share as much as possible (extended OWL import construct)**
2. **Keep it local whenever sharing does not work (C-OWL context mappings)**

Note: Using context allows for incremental, piece-wise construction of the Semantic Web (bottom up vs. top down approach).

C-OWL (1): multiple indexed ontologies

(Indexed Ontologies): Each ontology O_i and its language are associated a **unique identifier** i (e.g., $i{:}C$, $j{:}E$, $i{:}\exists r.C$)

(OWL space): A OWL space is a family of ontologies $\{\langle i, O_i \rangle\}$

(Local language): A local concept (role, individual), C_i (R_i, O_i) which appears in O_i with index i.

C-OWL (2): local Interpretations and domains

Consider the OWL space *{<i, O_i>}.* Associate to each ontology O_i *a* OWL interpretation I_i

(Local Interpretations): A C-OWL interpretation I is a family $I = \{I_i\}$, *of* interpretations I_i called the ***local interpretations*** of O_i.

Note: each ontology is associated with a local Interpretation

(Local domains): each local interpretation is associated with a ***local domain*** and a local interpretation function, namely

$$I_i = \langle \Delta^{Ii}, (.)^{Ii} \rangle,$$

Note: Local domains may overlap (two ontologies may refer to the same object)

C-OWL (3): context mappings

(Context mappings): A context mapping from ontology O_i to ontology Oj has one of the four following forms,

$$i:x \xrightarrow{\sqsubseteq} j:y, \quad i:x \xrightarrow{\sqsupseteq} j:y, \quad i:x \xrightarrow{\equiv} j:y, \quad i:x \xrightarrow{\perp} j:y, \quad i:x \xrightarrow{*} j:y,$$

with *x, y* concepts (individuals, roles) of the languages L_i and L_j

(Domain relations): Given a set of local interpretations

$$I_i = <\Delta^{Ii}, (.)^{Ii}>$$

with local domains Δ^{Ii} , a domain relation r_{ij} is a subset of Δ^{Ii} *x* Δ^{Ii} (a mapping between Δ^{Ii} and Δ^{Ii})

C-OWL: two examples

Example 1: Sale:Car and FIAT:car describe the *same* set of cars from two different viewpoints (sales and maintenance), and therefore with different attributes. We cannot have equivalence, however we have the following contextual mappings:

$$\mathsf{Sale}:\mathsf{Car} \xrightarrow{\equiv} \mathsf{FIAT}:\mathsf{Car}$$

Domain relation satisfies:

$$r_{ij}(Car^{ISale}) = Car^{IFIAT}$$

Example 2: Ferrari sells two cars which use petrol. Mappings:

$$\mathsf{WCM}:\mathsf{Petrol} \xrightarrow{\sqsupseteq} \mathsf{Ferrari}:\mathsf{F23}$$

$$\mathsf{WCM}:\mathsf{Petrol} \xrightarrow{\sqsupseteq} \mathsf{Ferrari}:\mathsf{F34i}$$

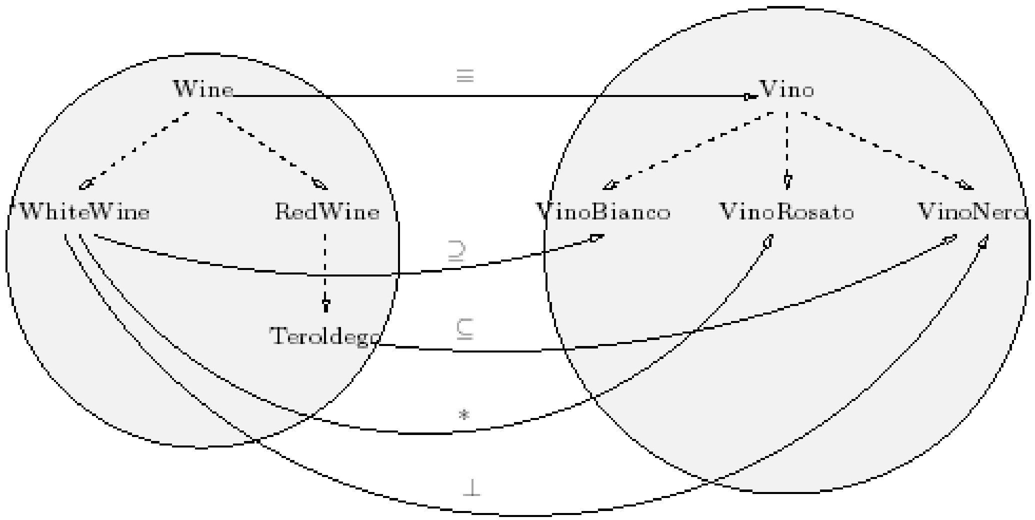
Domain relation satisfies:

$$r_{WCM,\ Ferrari}(Petrol)^{IWCM} \supseteq \{F23^{IFerrari}, F34i^{IFerrari}\}$$

C-OWL: the vision

A **contextual ontology** is a pair:
- OWL ontology
- a set of context mappings

A **context mapping** is a 4-tuple:
- A mapping identifier
- A source context
- A target context
- A domain relation



NOTES:
- a C-OWL space is a set of contextual ontologies
- mappings are objects (!!)

Outline

- The problem: the complexity of knowledge
- The solution: managing diversity
- Some early work
  - C-OWL: Representing semantic mappings
  - Semantic Matching: Discovering semantic mappings [Giunchiglia et al, ISWC**, ESWC**, ECAI'06]
  - Open Knowledge: Exploiting local theories and semantic mappings
- Three core issues

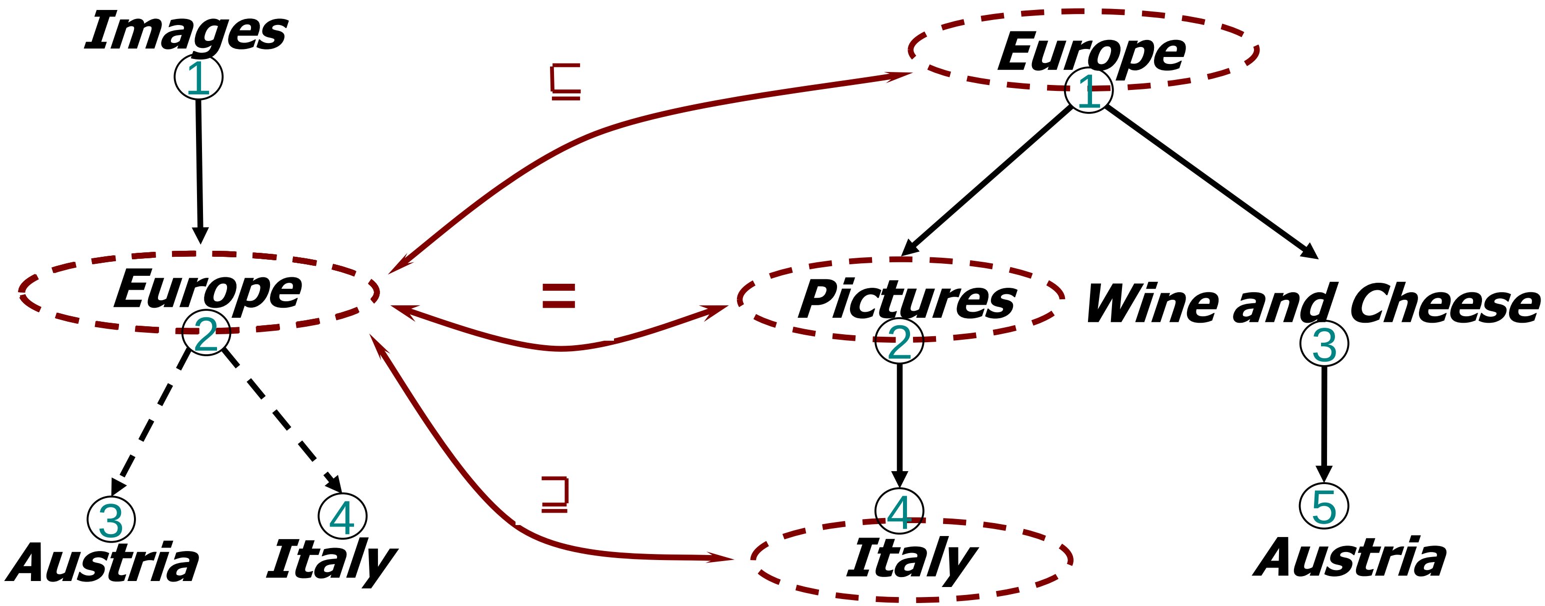
An example: Matching catalogs for eBusiness

- Top
  - Industrial Manufacturing and Processing Machinery and Accessories
    - Lapidary machinery and equipment
    - Leatherworking repairing machinery and equipment
    - Industrial process machinery and equipment and supplies
      - Separation machinery and equipment
      - Cutting tools
        - Drills
        - Reamer cutting tool
        - Form tools or toolbits
        - Taps or dies
        - Broach cutting tool
        - Gear cutting tools
        - Rotary burrs
        - Regrind or reclaim or coating services for cutting tools
        - Countesink tool or counterbore tool
        - Machinery cutting knives or knife assemblies
      - Assembly machines
      - Paint systems
    - Foundry machines and equipment and supplies
    - Workshop machinery and equipment and supplies

- Top
  - Machine, apparatus
    - Heat exchanger
    - Boiler, furnace
    - Sterilizer
    - Cleaning installation
    - Sound damper, pulsation damper
    - Cutting machine
      - Plasma cutting machine
      - Cutting machine (other)
        - shears (manufacturing of glass)
        - melt machine (manufacturing of glass)
      - Cutting machine (parts)
      - Cutting mach. (maint., serv.)
      - Cutting mach. (repair)
    - Textile machine
    - Pressure machine

Ex.: <Id, Drills, Cutting machine (other), subsumes>

Toy example: a small Web directory



$< ID_{22}, 2, 2, = >$

$< ID_{21}, 2, 1, \sqsubseteq >$ $\Rightarrow$ $< ID_{22}, 2, 2, = >$

$< ID_{24}, 2, 4, \sqsupseteq >$

Algo
Step 4

The two key problems

1. **Ontologies (Web directories? Classifications?) - Vast majority (including catalogs) are ambiguously and partially defined:**
   1. Meaning of labels is ambiguous (labels are in Natural Language)
   2. Labels are (somewhat) complex sentences
   3. Meaning of links is ambiguous (no labels or ambiguous labels)
   4. A lot of background knowledge is left implicit

2. **Matching - The notion of matching is not well defined:** many, somewhat similar, notions and corresponding implementations can be found in the literature...

Problem 1: ontologies
Dealing with ambiguity and partiality

Translate classifications into (lightweight) ontologies according to the following (not necessarily sequential) phases

1. **Compute the background knowledge: extract it from existing resources (e.g., Wordnet, other ontologies, other peers, the Web, …)**

2. **For any label compute the *concept of the label:* translate the natural language label into a description logic formula (using NLP)**

3. **For all nodes compute the concepts at nodes: compose concepts of labels into a complex formula which captures the classification strategy**

Problem 2
Formalize Semantic Matching

Mapping element is a 4-tuple < ID_{ij}, $n1_i$, $n2_j$, R >, where

- ID_{ij} is a unique identifier of the given mapping element;
- $n1_i$ is the *i-th* node of the first graph;
- $n2_j$ is the *j-th* node of the second graph;
- R specifies a **semantic relation** between the concepts at the given nodes

Computed *R's*, listed in the decreasing binding strength order:

equivalence { = };
more general/specific { $\sqsupseteq$, $\sqsubseteq$ };
mismatch { $\perp$ };
overlapping { $\sqcap$ }
… I_dont_know.

Semantic Matching: Given two graphs *G1* and *G2*, given a node $n1_i \in G1$, find the mapping with the strongest semantic relation *R'* holding with node $n2_j \in G2$

Problem 2
Implement semantic matching

The idea: **reduce the matching problem to a validity problem**

Let

Wffrel (C1, C2)

be the relation to be proved between the two concepts *C1* and *C2*, where:

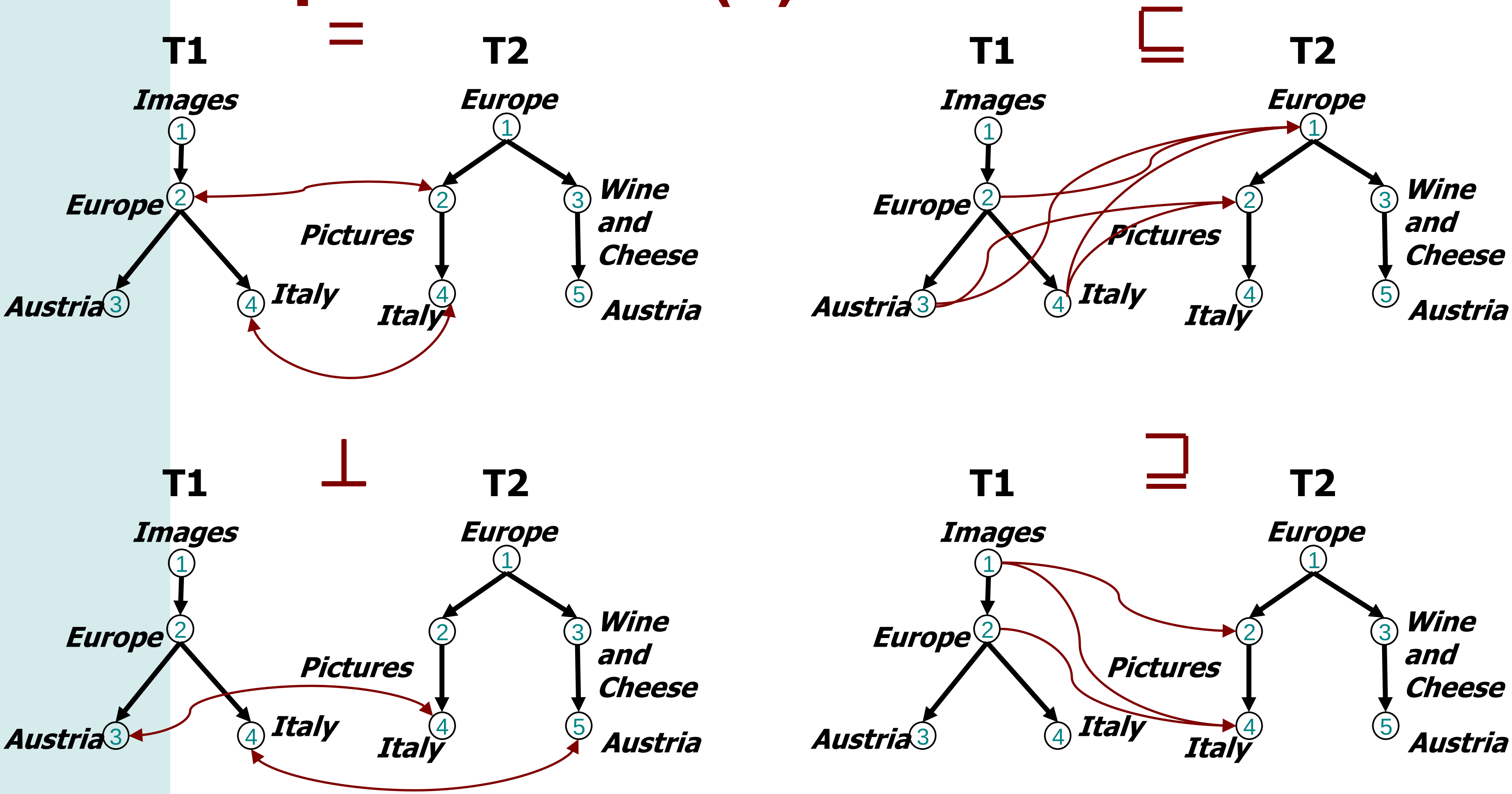
- ***C1 equiv C2*** **is translated into** $C1 \leftrightarrow C2$
- ***C1 subsumes C2*** **is translated into** $C1 \rightarrow C2$
- $C1 \perp C2$ **is translated into** $\neg(C1 \wedge C2)$

Then prove

"Background knowledge" $\rightarrow$ ***Wffrel*** $(C1_i, C2_j)$

… using SAT

Step 4: cont'd (2)



Does this really work?
Recall (incompleteness)!

NLP techniques evaluation [Magnini et *al. 2004*]

- **Google vs. Yahoo:** Architecture (Arc.) and Medicine (Med.) parts
- Precision (Pr.), Recall (Re.), F-measure (F)
- CtxMatch (baseline)

		Pr.	Re.	F
Arc.	equiv.	.33 (.25)	.04 (.04)	.07 (.07)
	more g.	.92 (.93)	.42 (.44)	.58 (.60)
	less g.	.88 (.90)	.62 (.41)	.73 (.56)
Med.	equiv.	.27 (.25)	.07 (.05)	.11 (.08)
	more g.	.91 (.95)	.48 (.45)	.63 (.61)
	less g.	.83 (.86)	.61 (.54)	.70 (.66)

The background knowledge problem!

Outline

- The problem: the complexity of knowledge
- The solution: managing diversity
- Some early work
  - C-OWL: Representing semantic mappings
  - Semantic Matching: Discovering semantic mappings
  - Open Knowledge: Exploiting semantic mappings and local theories [FP6 EC project. Partners: Edinburgh, Trento, Amsterdam, Barcellona, Open University, Southampton]
- Three core issues

Open Knowledge: Semantic Webs through P2P interaction

Abstract: **We present a manifesto of kowledge sharing that is based not on direct sharing of "true" statements about the world but, instead, is based on sharing descriptions of interactions ...**

... [This] narrower notion of semantic committment ... Requires peers only to commit to meanings of terms for the purposes and duration of the interactions in which they appear.

... This lightweight semantics allows networks of interaction to be formed between peers using comparatively simple means of tackling the perennial issues of query routing , service composition and ontology matching.

Web Site: www.openk.org

Open Knowledge: Key ingredients

1. **Peer-to-peer (P2P) organization at the network and knowledge level (e.g. autonomy of the peers, no central ontology, diversity in the data, metadata and ontologies, ...)**
2. **Interactions specified using interaction models**
3. **P2P peer search mechanism**
4. **Semantic agreement via semantic mappings built dynamically as part of the interaction**
5. **Good enough answers: answers which serve the purpose given the amount of resources (no requirement of correctness or completeness)**
6. **Knowledge adaptation via approximation in order to get answers which are good enough**

Outline

- The problem: the complexity of knowledge
- The solution: managing diversity
- Some early work
- Three core issues

The need for common (shared) knowledge

- **FACT:** Common (shared) knowledge (e.g. shared ontologies) is ***easier*** to use
- **ISSUE:** How can we construct ***common*** knowledge components (e.g., from context mappings to OWL import), possibly mutually inconsistent, also understanding their applicability boundaries
- **SUGGESTED APPROACH:** Common knowledge should not be built a priori (in the general case). It should "emerge" as a result of a ***incremental process of convergence*** among views, goals, … of peers.

The lack of background knowledge

FACT1: There is evidence that a major bottleneck in the use of knowledge based systems is the lack of the background knowledge (Giunchiglia et al, ECAI 2006; Frank Van Harmelen et al, ECAI 2006 C&O wshop invited talk)

FACT 2: In certain high value areas large domain specific knowledge bases have been built in a systematic way (e.g., the medical domain). However this approach will not scale to commonsense knowledge

FACT 3: The commonsense knowledge of the world is essentially unbound. No knowledge base will ever be “complete”

ISSUE: What is the “right” background knowledge? How do we construct it?

The knowledge grounding problem

- **FACT 1: Two main approaches to data and knowledge management:**
  - **the *top down deductive* approach, e.g., the use of ontologies, classifications, knowledge bases, …**
  - **the *bottom up inductive* approach, e.g., data or text mining, information retrieval, ...**

- **FACT 2: Both approaches have their weakenesses:**
  - **The top down approach will always miss some of the necessary background knowledge**
  - **The bottom up approach uses oversimplified models of the world**

- **ISSUE: We need to fill the gap … composing strengths and minimizing weakenesses**

Conclusion

- Handling the upcoming complexity of knowledge requires the development of new paradigms.
- Our proposed solution: managing diversity
- Three steps: local theories + mappings + adaptation
- … Still at the beginning with many unsolved core issues, most noticeably: how to build common knowledge, how to build background knowledge and how to ground knowledge into “objects”

Acknowledgements

C-OWL: Paolo Bouquet, Frank Van Harmelen, Heiner Stuckenschmidt, Luciano Serafini

Semantic Matching: Pavel Shvaiko, Mikalai Yaskevich, Ilya Zaihrayeu

Open Knowledge: Dave Robertson, Frank Van Harmelen, Carles Sierra, Alan Bundy, Fiona, McNeill, Marco Schorlemmer, Nigel Shadbolt, Enrico Motta, …

… and many others

References (http://www.dit.unitn.it/~knowdive/)

F. Giunchiglia: Managing Diversity in Knowledge In preparation. Mail to: fausto@dit.unitn.it

F. Giunchiglia,M.Marchese, I. Zaihrayeu: Encoding Classifications into Lightweight Ontologies. ESWC'06.

M. Bonifacio, F. Giunchiglia, I. Zaihrayeu: Peer-to-Peer Knowledge Management . I-KNOW'05.

F. Giunchiglia, P.Shvaiko, M. Yatskevich: S-Match: an algorithm and an implementation of semantic matching. ESWS'04.

Bouquet, F. Giunchiglia, F. van Harmelen, L. Serafini, H. Stuckenschmidt: C-OWL: Contextualizing Ontologies . ISWC'03.

F. Giunchiglia, F. van Harmelen, L. Serafini, H. Stuckenschmidt: C-OWL . Fothcoming book.

F.Giunchiglia, I.Zaihrayeu: Making peer databases interact – a vision for an architecture supporting data coordination. CIA'02

P. Bernstein, F. Giunchiglia, A. Kementsietsidis, J. Mylopoulos, L. Serafini, and I. Zaihrayeu: Data Management for Peer-to-Peer Computing: A Vision , WebDB'02.

C. Ghidini, F. Giunchiglia: Local models semantics, or contextual reasoning = locality + compatibility. Artificial Intelligence Journal, 127(3), 2001.

F. Giunchiglia, Contextual reasoning, Epistemologia - Special Issue on I Linguaggi e le Macchine, 1993.

F. Giunchiglia, P.Shvaiko: Discovering Missing Background Knowledge in Ontology Matching. ECAI 2006 .

Managing knowledge … in the Web

The novelty: Lots of *pre-existing* knowledge systems, developed independently, most of the time fully autonomous

The predominant approach (so far):

- Reduce to the "standard" approach,
- Integrate the pre-existing knowledge systems by building, at *design time*, a *"general enough"* representation model,
- Most commonly: design a *global* representation schema

Issues: knowledge merging, consistency, how to deal with granularity of representation, …

Example: Information integration (databases and ontologies). Integration via a design time defined global schema / ontology (a single virtual database/ ontology).

However…

Ex.2: web classifications (~ 10^3 nodes)

Google | **Looksmart**

However…
Ex.3: Intranet applications

Difficulties (failures) in knowledge integration attempts

- Multinational CV management and sharing
- Collaborative design
- Mailbox heterogeneity (... and attachments)
- ...

Why it will get worse

Over time, the complexity of knowledge **and its *interconnections* will grow to the point where we can no longer fully and effectively understand its global behaviour and evolution:**

- **We will build and interconnect systems on top of a landscape of existing highly interconnected systems**
- **Each system and its interconnections has/had its own producers and users but the whole will not**
- **Some existing systems and their interconnections will not be accessible or will not be changeable; they will be given to us as a an asset/ sunk cost**
- **Systems will increasingly need to be adapted at run-time;**

A toy example:
Mr.1 and Mr.2 viewpoints

The two local theories ...

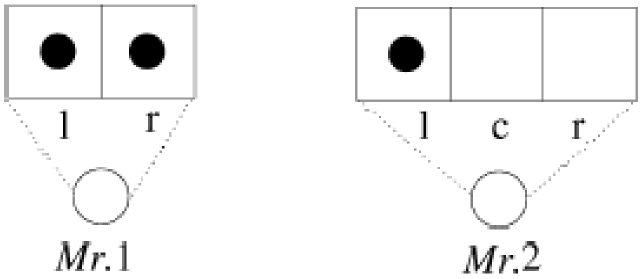


Figure 2: $Mr.1$ and $Mr.2$'s contexts.

Which world? How much of it?

A toy example – more: Partial agreement between Mr.1 and Mr.2

The two local theories agree to some extent …

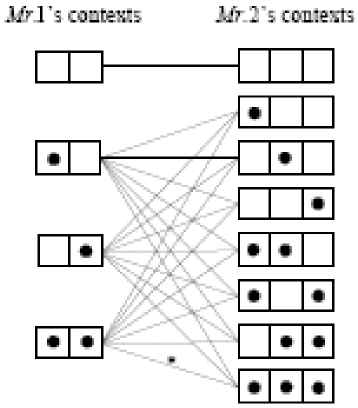


Figure 3: Compatible contexts of $Mr.1$ and $Mr.2$.

Example: if Mr.1 sees one ball then Mr.2 sees at least one ball (one, two, or three)

Outline

- The problem: the complexity of knowledge
- The solution: managing diversity
- Some early work
- Three core issues

The application area

Application area: reusing, sharing, adapting language in the Web

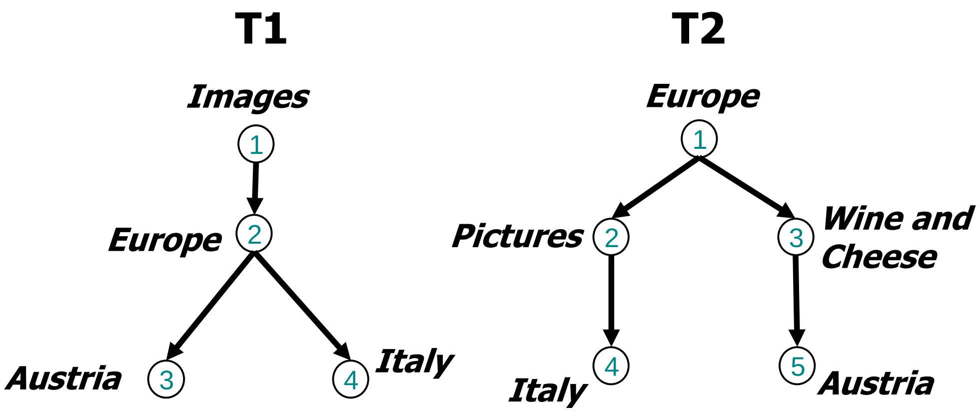
Local theories (languages): ontologies, taxonomies, classifications, …

Some early work:

- C-OWL: Representing semantic mappings
- Semantic Matching: Discovering semantic mappings
- Open Knowledge: Adapting and exploiting local theories and semantic mappings

Problem 1: ontologies
Phase 1: compute the background knowledge

The idea: Exploit pre-existing knowledge, (e.g., Wordnet, element level syntactic matchers, other ontologies, other peers, the Web …)



Results of step 3:

T1 \ T2	C_{Europe}	$C_{Pictures}$	C_{Wine}	C_{Cheese}	C_{Italy}	$C_{Austria}$
C_{Images}		$=$				
C_{Europe}	$=$				$\sqsupseteq$	$\sqsupseteq$
$C_{Austria}$	$\sqsubseteq$				$\perp$	$=$
C_{Italy}	$\sqsubseteq$				$=$	$\perp$

Problem 1: ontologies
Phase 2: compute concepts of labels

The idea: Use Natural language technology to translate natural language expressions into internal **formal** language expressions (**concepts of labels**)

Preprocessing:

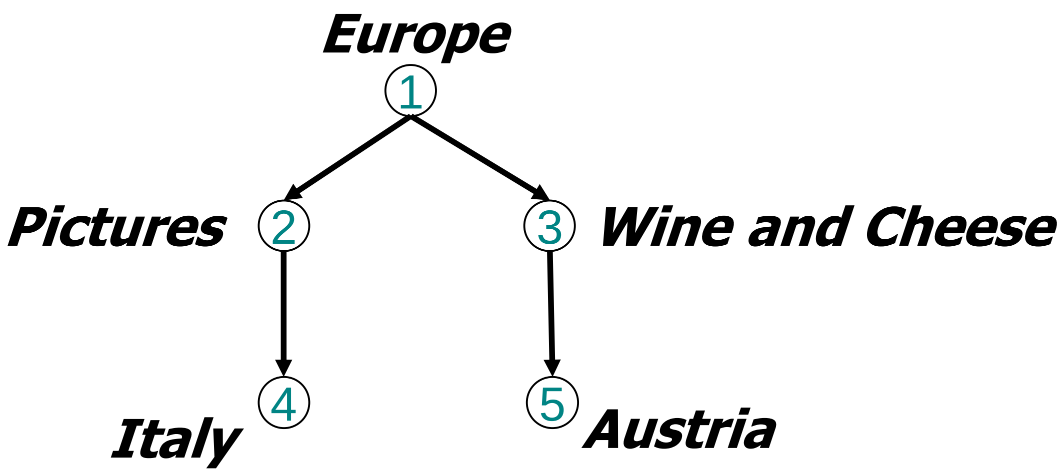
- **Tokenization.** Labels (according to punctuation, spaces, etc.) are parsed into tokens. E.g., Wine and Cheese → <Wine, and, Cheese>;
- **Lemmatization.** Tokens are morphologically analyzed in order to find all their possible basic forms. E.g., Images → Image;
- **Building atomic concepts.** An oracle (WordNet) is used to extract senses of lemmatized tokens. E.g., Image has 8 senses, 7 as a noun and 1 as a verb;
- **Building complex concepts.** Prepositions, conjunctions, etc. are translated into logical connectives and used to build complex concepts out of the atomic concepts

  E.g., $C_{\textit{Wine and Cheese}} = \langle Wine, U(WN_{Wine})\rangle \sqcup \langle Cheese, U(WN_{Cheese})\rangle$,

  where U is a union of the senses that WordNet attaches to lemmatized tokens

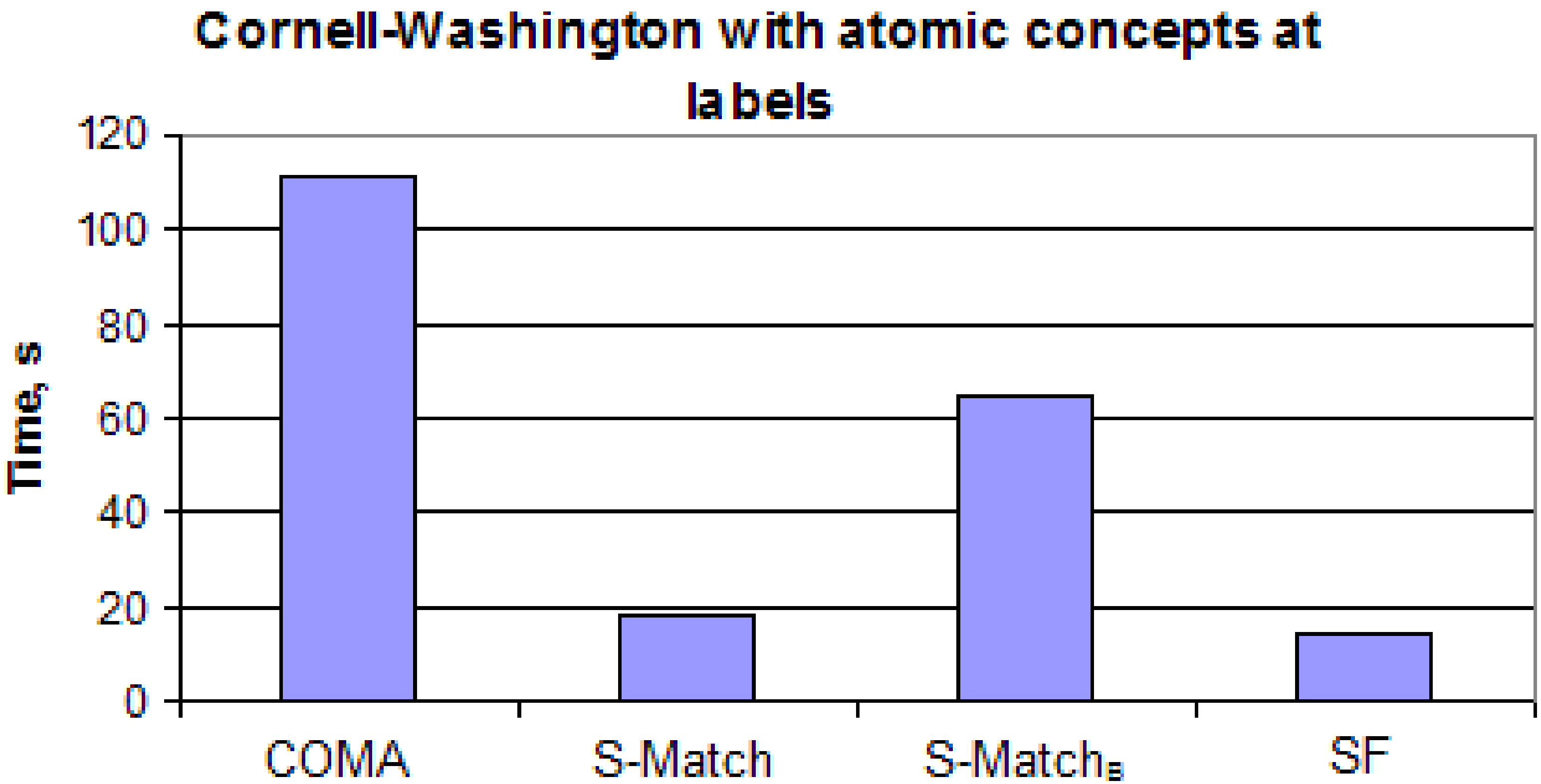
Problem 1: ontologies
Phase 3: compute concepts at nodes

- **The idea:** **extend concepts at labels by capturing the knowledge residing in a structure of a graph in order to define a context in which the given concept at a label occurs**
- **Computation (basic case):** **Concept at a node for some node *n* is computed as an intersection of concepts at labels located above the given node, including the node itself**



$$C_4 = C_{Europe} \sqcap C_{Pictures} \sqcap C_{Italy}$$

Does this really work? Efficiency?



Trees max. depth	# of nodes per tree	# of labels per tree	Average # of labels per node
10/8	253/220	253/220	1/1