

Introduzione all'ontology engineering



Ernesto Damiani

*Dipartimento di Tecnologie
dell'Informazione –
Università di Milano
Polo di Crema*



Sommario

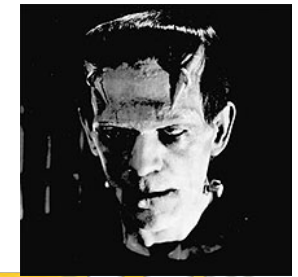


- ❑ Cos'è il Web of Knowledge ?
- ❑ I metadati
- ❑ I formati storici
- ❑ Nuovi formati: da XML a RDF
- ❑ RDF Schema e il Semantic Web
- ❑ Ontology Engineering

Web of Knowledge



- ❓ L'educazione moderna incoraggia la specializzazione. Lo storico e il fisico “generico” sono stati sostituiti da figure con competenze più focalizzate (e.g. lo storico del Rinascimento o l'astrofisico)
- ❓ **Però:** il progresso della conoscenza nasce dalle interconnessioni “lateral” tra discipline quanto e più che dalla ricerca “verticale”.
- ❓ **Conoscenza=informazione organizzata**



Un esempio: Frankenstein search (James Burke)

Search for: frankenstein

Filter search by place

Region

Africa
Antarctica
Asia
Australia

OR

Specific Place

England

[Choose from list](#)
[Choose from map](#)



Filter search by time

Period

1900s AD
1800s AD
1700s AD
1600s AD

OR

Specific Begin Date

1800

AD ☒ BCE ☐

Specific End Date

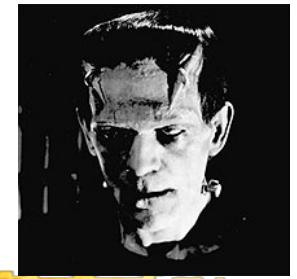
1900

AD ☒ BCE ☐

[Choose from timeline](#)

Search K-Web

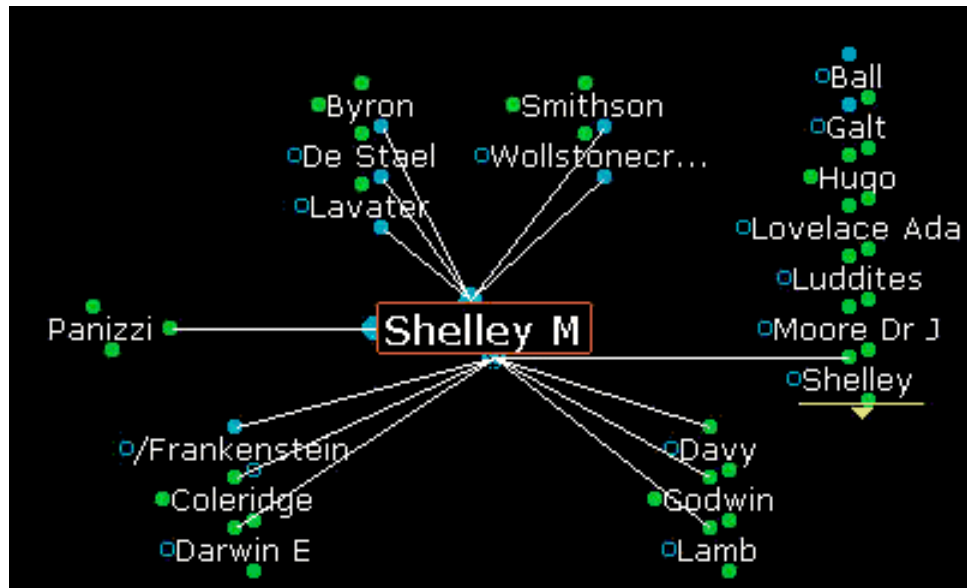
Frankenstein search: struttura d'accesso



Name/Description (what's this?)	Node Type (what's this?)	Nav. View (what's this?)	Detail View (what's this?)
 Frankenstein The original story was written in 1818 by Mary Shelley	 gateway		
other node Description goes here			
other node Description goes here			
other node Description goes here			

 Person |  Place |  Thing |  Event

Frankenstein search: i dati



b: 30 August 1797, London, England
m: to Percy Bysshe Shelley
d: 1 February 1851, London, England. Aged 54
children: Percy Florence Shelley
Parents: William Godwin, Mary Wollstonecraft
Siblings: none
Other relations:
Education: Privately at home



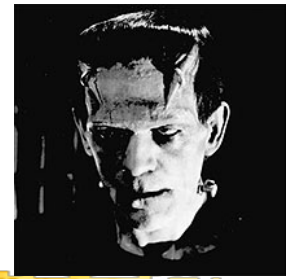
Assessment

Mary Shelley was an intellectual and radical, remembered as Shelley's wife and as the author of "Frankenstein."

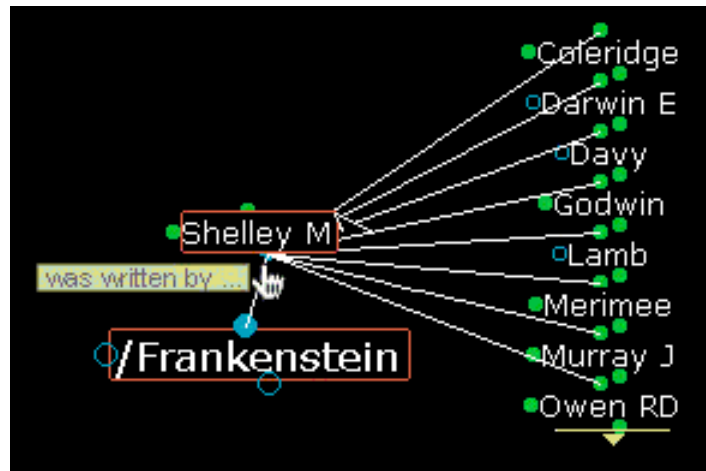
Bibliography

Spark Muriel, *Mary Shelley*, Constable & Co., Ltd., London, 1988.
Sunstein, Emily, *Mary Shelley, Romance and Reality*, John Hopkins University Press, Baltimore, 1989.

Frankenstein search: i metadati



relazioni



vincoli

Show only nodes that are in:



Show only nodes that are:

- ☒ People
 - ☒ Familial
 - ☒ Business
- ☐ Places
- ☒ Things
- ☐ Events

tipi

Metadati: una visione 'ingenua'



? Dati che descrivono altri dati

? e.g. il catalogo di una biblioteca

? Spesso piu' facili da trattare dei dati stessi

? Il formato dei metadati è una decisione del catalogatore, quello dei dati no..

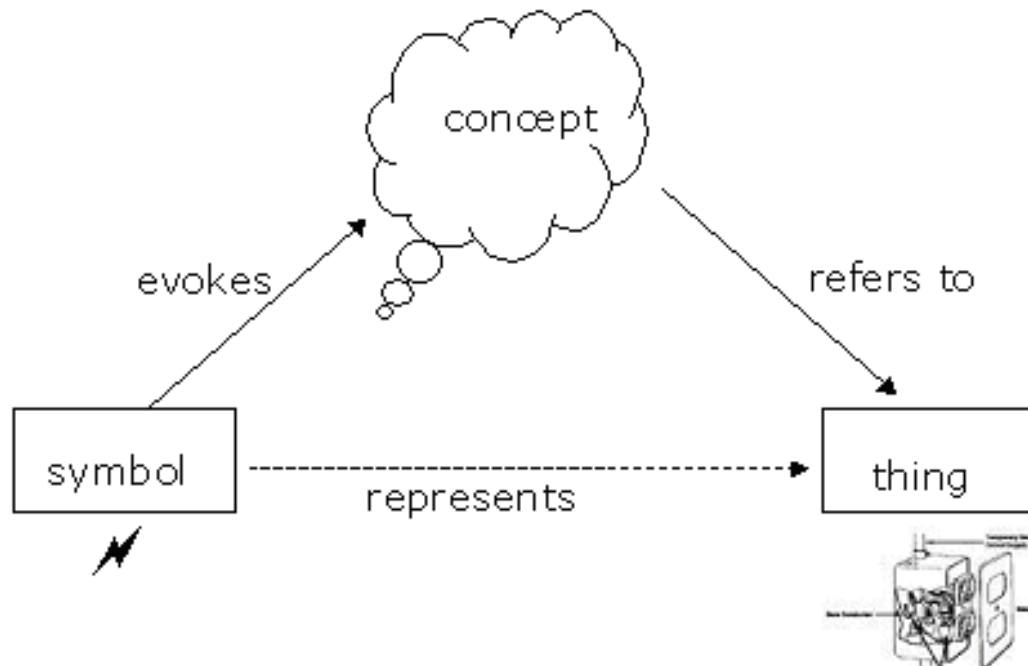
Quindi:



- ❑ Metadati = 'etichette descrittive'
esprimono:
 - ❑ Struttura e semantica dei datima anche altro:
 - ❑ Autore
 - ❑ Copyright
 - ❑ Informazioni di autenticazione
 - ❑ Protezioni e permessi d'accesso
 - ❑ Prezzo

- ❑ Usi dei metadati
 - ❑ Catalogazione dei dati
 - ❑ Elaborazione automatica dei dati da parte di agenti software
 - ❑ Controllo degli accessi
 - ❑ Filtraggio dei flussi informativi

Metadati: una visione “colta”



Il “meaning triangle” (Ogden and Richards, 1923)

Esempi 'storici'



 PICS

 Dublin core

PICS



- ❑ Platform for Internet Content Selection
- ❑ www.w3.org/PICS/
- ❑ Recommendation dal 1996
- ❑ Condivisione e interscambio di valutazioni su siti e pagine Web
 - ❑ etichette PICS trasportate negli header HTTP or come marcatori META
 - ❑ from document server
 - ❑ or an independent *label bureau*
 - ❑ firmati elettronicamente e associati con un *digest*

Esempio



```
<html>
  <head>
    <META http-equiv="PICS-Label" content='
(PICS-1.1 "http://www.gcf.org/v2.5"
  labels on "1999.11.05T08:15-0500"
    until "2001.12.31T23:59-0000"
      for "http://w3.org/PICS/Overview.html"
        ratings (suds 0.5 density 0 color/hue 1))
    '>
  </head>
  <body>
    <p>qui va il documento</p>
  </body>
</html>
```

Applicazioni di PICS



- ❑ Protezione dei minori
 - ❑ CyberPatrol, EvalWeb, NetNanny, ...
- ❑ Utente genera o ottiene un *profilo*
- ❑ L'applicazione confronta rating e profilo e compie una delle seguenti azioni:
 - ❑ suggest
 - ❑ search
 - ❑ monitor
 - ❑ warn
 - ❑ block

Etichette Recreational Software Advisory Council (RSAC): Violenza



 (category (transmit-as "v") (name "Violence")
(label (name "**Conflict**") (description "Harmless
conflict; some damage to objects") (value 0))
(label (name "**Fighting**") (description "Creatures
injured or killed; damage to objects; fighting")
(value 1)) (label (name "**Killing**") (description
"Humans injured or killed with small amount of
blood") (value 2)) (label (name "**Blood and Gore**")
(description "Humans injured or killed; blood and
gore") (value 3)) (label (name "**Wanton
Violence**") (description "Wanton and gratuitous
violence; torture; rape") (value 4)))

Problemi



- ❓ Chi sceglie le etichette?
- ❓ Chi ne definisce la sintassi?
- ❓ Come determinarne il significato?
- ❓ **A quale dato sono associate?**

Dublin Core



- 15 etichette
- Metadati bibliografici
- Campi tipati
 - vocabolario controllato
 - ISO 8601 (date)
 - tipi di risorse
 - URLs, ISBNs
 - RC 1766 (lingue)

Elementi Dublin Core (1)



? Title

? Creator (persona, organizzazione, servizio)

? Subject (lista di parole chiave dal vocabolario controllato)

? Publisher

? Contributor

? Date

? Type (e.g. collezione, evento, immagine, oggetto fisico, software, suono, ...)

? Format

Elementi Dublin Core (2)



? Identifier

? Source

? Language (lingua)

? Relation

(collegamento a una risorsa correlata)

? Coverage (es. luogo, tempo, a giurisdizione)

? Rights (proprietà intellettuale, copyright)

? Description

(riassunto, indice, o testo libero)

Metadati basati su XML

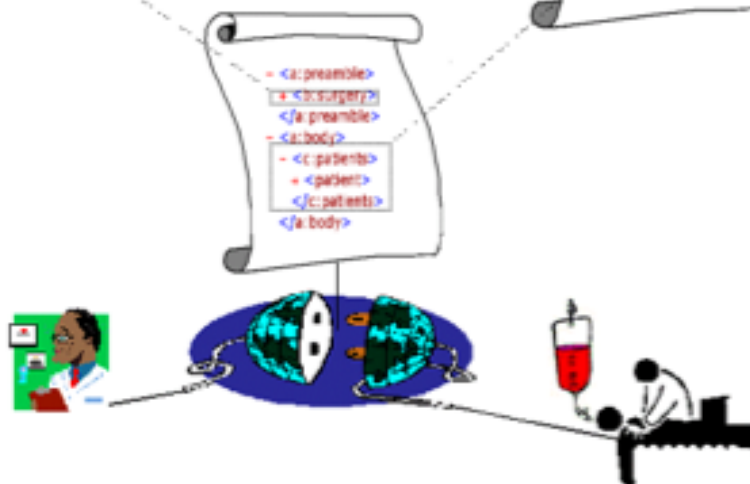


- ❑ XML Schema (perchè no?..)
- ❑ RDF (Resource Description Framework)
- ❑ RDF Schema

Uso di schemi XML come rappresentazione di metadati



- Problema principale: il *potere espressivo*



- **Ambiguità nel significato dell'inclusione di elementi**

RDF



- ❑ Resource Description Framework
- ❑ W3C Recommendation, dal Febbraio 1999
 - ❑ si basa su XML e spazi di denominazione
- ❑ Indipendente dalla piattaforma
- ❑ Eredita'
 - ❑ PICS
 - ❑ biblioteche digitali, SGML e rappresentazione della conoscenza
- ❑ Scopo principale
 - ❑ rendere le etichette elaborabili da agenti software oltre che leggibili dall'uomo

Primi passi con RDF



- ❑ I dati sono descritti da *asserzioni*
- ❑ Un'asserzione e' composta da
 - ❑ Una *risorsa*
 - ❑ Pagina Web o parte di essa, intero sito Web...
 - ❑ ...un documento XML ...un libro ... qualunque cosa dotata di URI
 - ❑ una *proprietà*
 - ❑ e.g. Titolo, Data, Tipo
 - ❑ un valore
 - ❑ es. stringa, numero, frammento XML, un'altra risorsa

Risorse



Dalla RFC 2396:

-  "A resource can be anything that has identity. Familiar examples include an electronic document, an image, a service (e.g., "today's weather report for Los Angeles"), and a collection of other resources. Not all resources are network "retrievable"; e.g., human beings, corporations, and bound books in a library can also be considered resources. The resource is the conceptual mapping to an entity or set of entities, not necessarily the entity which corresponds to that mapping at any particular instance in time. Thus, a resource can remain constant even when its content---the entities to which it currently corresponds---changes over time, provided that the conceptual mapping is not changed in the process".

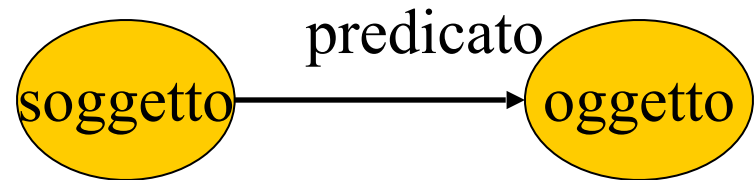
Risorse e URI



- ❑ Le risorse sono identificate da URI
- ❑ La risorsa identificata da una URI
puo' essere *astratta* = ***non
recuperabile in Rete***
- ❑ Sintassi: [absoluteURI | relativeURI]
["#" fragment-id] /

Altre visioni di RDF

- ? Sintassi XML standard per rappresentare *grafi* etichettati diretti
- ? Sintassi XML standard per rappresentare *fatti* in logica del primo ordine senza variabili



Asserzione :=
(predicato,soggetto,oggetto)
Predicato = URI
Soggetto = URI
Oggetto = URI o costante
Oggetto =
Predicato(Soggetto)

Esempio di modello RDF

<http://www.crema.unimi.it/Home/Damiani.htm>

Autore

“La persona id=240, il cui nome e’ Ernesto Damiani e l’e-mail edamiani@crema.unimi.it e’ l’autore di <http://www.crema.unimi.it/Home/Damiani.htm>”

<http://www.crema.unimi.it/docenti/240>

nome

Email

Ernesto Damiani

edamiani@crema.unimi.it

Sintassi

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
```

```
<rdf:description about="http://www.crema.unimi.it/Home/Damiani">
```

```
<s:Creator>Ernesto Damiani</s:Creator>
```

```
</rdf:description>
```

```
</rdf:RDF>
```

RISORSA

PROPRIETA' - VALORE

(in questo caso costante)

SPAZIO DI DENOMINAZIONE =
VOCABOLARIO DI DOMINIO =
ELENCO DI PROPRIETA'

Iper testi e RDF



❓ xlink semplici: niente semantica!

❓ xlink tipati

❓ xlink:arcrole := predicato

❓ xlink:href := oggetto

❓ xlink:role := tipo

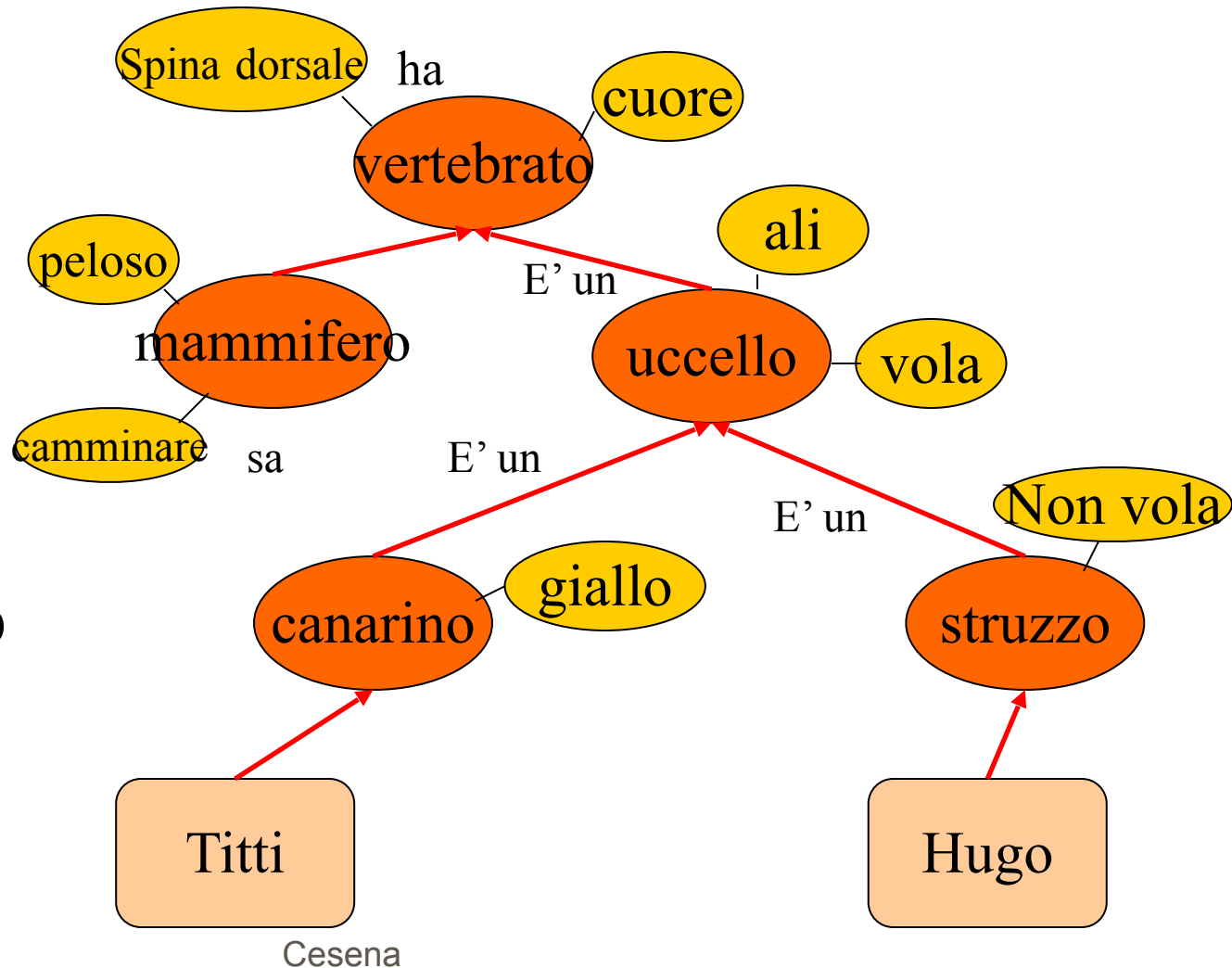
Schemi RDF



- ❑ Specificano le proprietà che possono comparire nelle asserzioni RDF
- ❑ Non sono solo elenchi di proprietà
 - ❑ Specificano le relazioni tra le proprietà
 - ❑ Dichiarano classi (e sottoclassi) di risorse
- ❑ Sviluppo consortile

Cosa sono gli schemi RDF ?

- ? Una vecchia idea:
le reti semantiche
 - ? formalismo IA,
circa 1970
- ? Organizzano
concetti in modo
elaborabile da
agenti software
- ? Primo passo verso
ontologie di
dominio



Interrogazioni ai metadati



❓ "Titti sa volare?"

❓ "Hugo ha le ali?"

❓ "Titti ha la spina dorsale?"

❓ "Quanti canarini vivono in gabbia?"

Cosa sono gli schemi RDF ? (2)



☐ Schemi RDF = Reti semantiche sul Web

☐ Nodi=risorse (URI)

☐ Archi

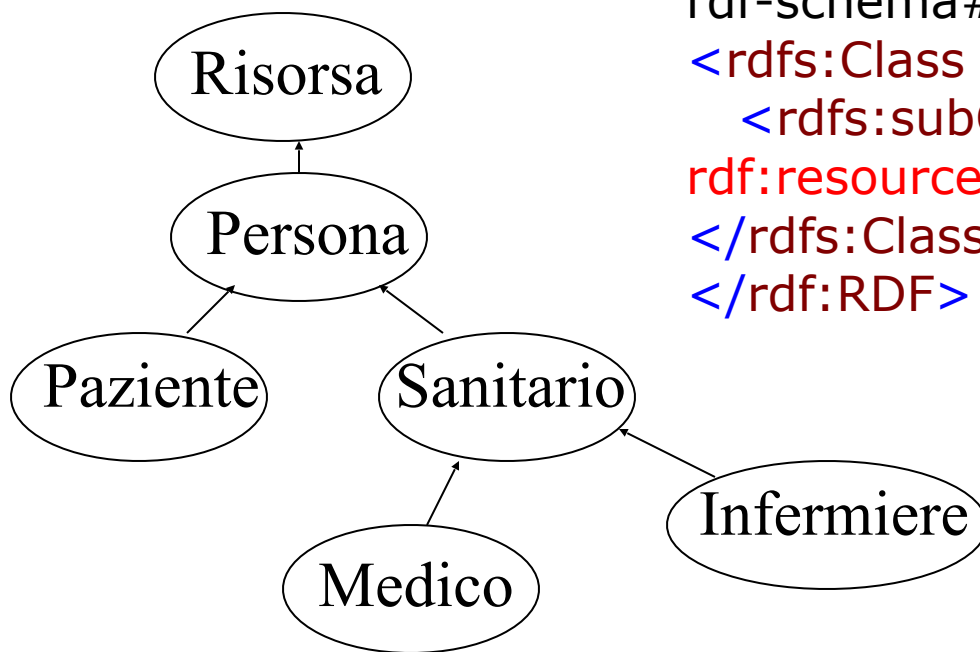
☐ rdfs:Class

☐ rdfs:Property

☐ rdfs:subClassOf

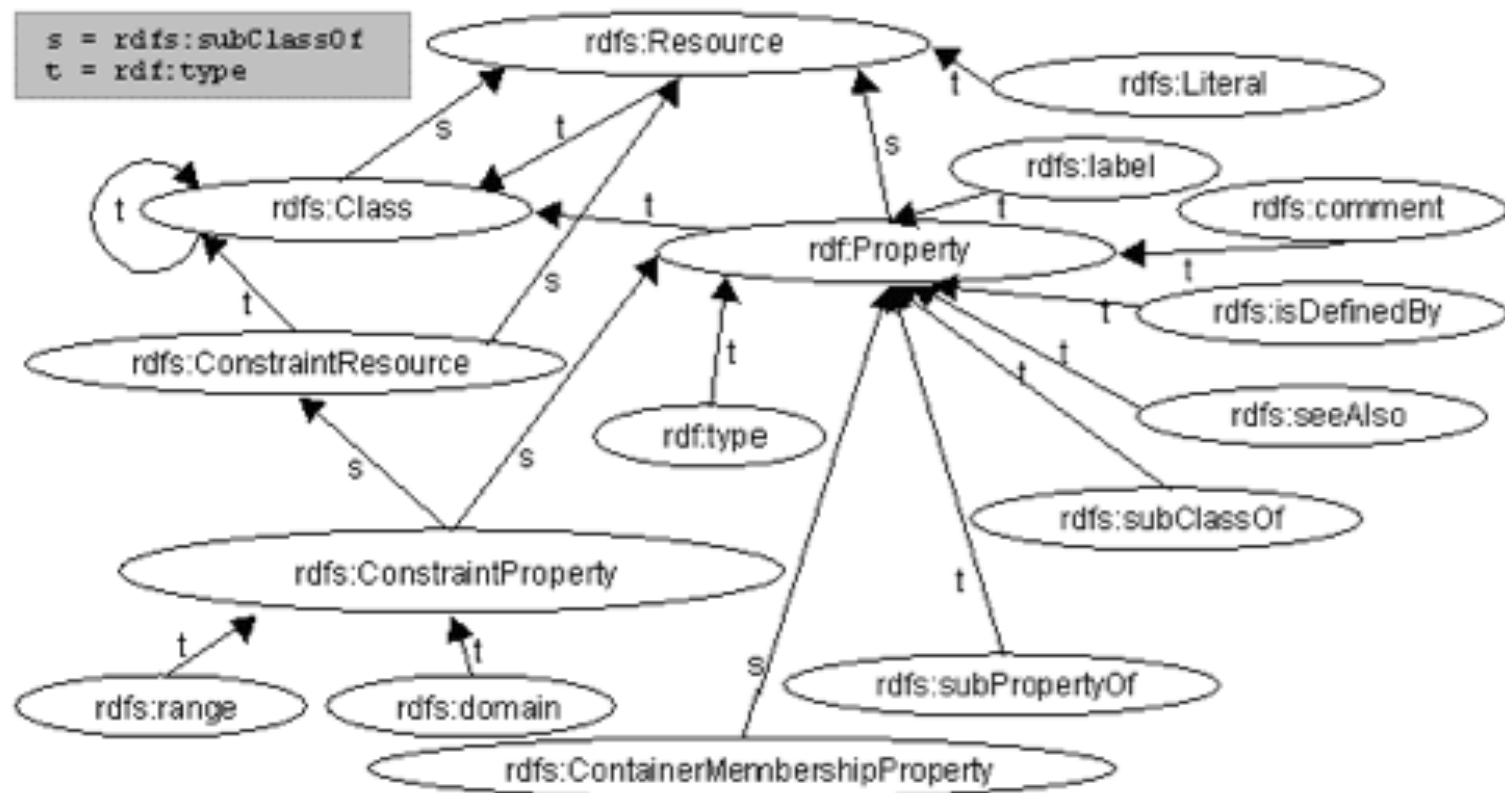
☐ rdf:type

Esempio



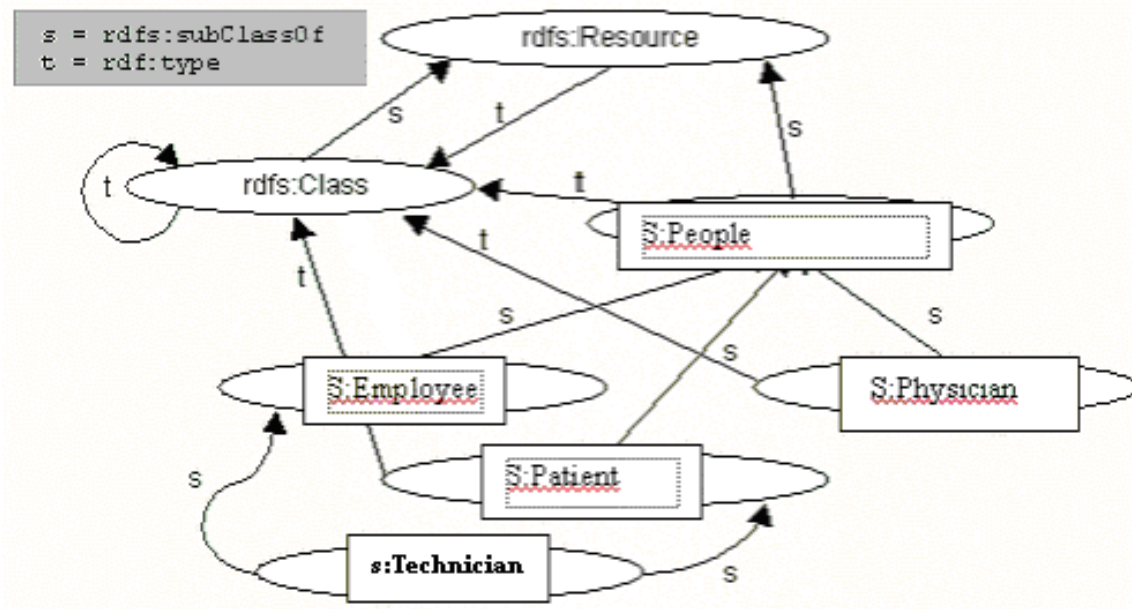
```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/
1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/
rdf-schema#">
  <rdfs:Class rdf:ID="Sanitario">
    <rdfs:subClassOf
rdf:resource="#Persona"/>
  </rdfs:Class>
</rdf:RDF>
```

RDFS: : il metaschema



a.k.a. "lo schema degli schemi"

Ontologie e RDFS



- Ontologie=schemi RDFS **scritti in RDFS**

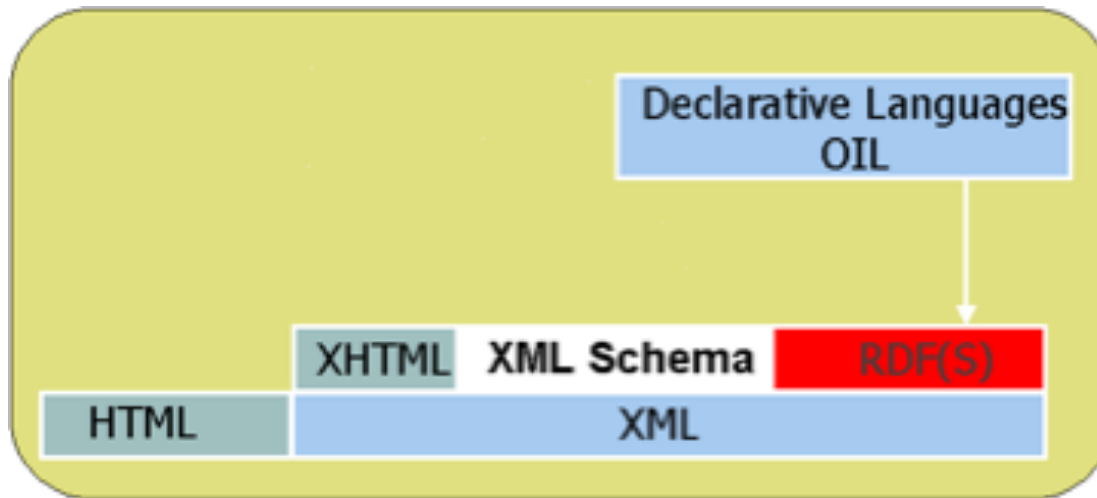
- Definiamo Class e Property con nome e cognome

- Possiamo **reificare**, cioè dire che una property ha una property o che una class appartiene a una class

Linguaggi avanzati

Schemi RDFS già pronti che definiscono property interessanti

Potere espressivo uguale a quello della logica descrittiva



DARPA Agent Markup Language (DAML)

Ontology Interchange Language (OIL)

OWL=DAML+OIL
Standard W3C

Resource Directory Description Language (RDDL)



- ❓ Proposto inizialmente (Gennaio 2001) per dare un significato ai riferimenti agli URI degli spazi di denominazione
- ❓ Identifica *classi di risorse* a cui si accede attraverso collegamenti ipertestuali dandone *scopo* e *natura* (spazi di denominazione: un caso particolare)
- ❓ formato leggibile sia dagli utenti umani sia dalle macchine
- ❓ Adottato da XML Schema, Schematron, RSS, Examplotron, XSLT Extension framework, SWAG
- ❓ Un semplice approccio al “Semantic Web”

RDDL: Esempio



```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rddl="http://www.rddl.org/">
  <rddl:resource rddl:id="www.questaURL.org/questodoc.xml"
xlink:arcrole="http://www.http://www.rddl.org/purposes#schema-
  validation"
xlink:role="http://www.w3.org/2001/XMLSchema"
xlink:title="uno schema XML"
xlink:href="http://www.crema.unimi.it/loschemadiquestodoc.xsd">
    <p>Testo descrittivo</p>
  </rddl:resource>
</rdf:RDF>
```

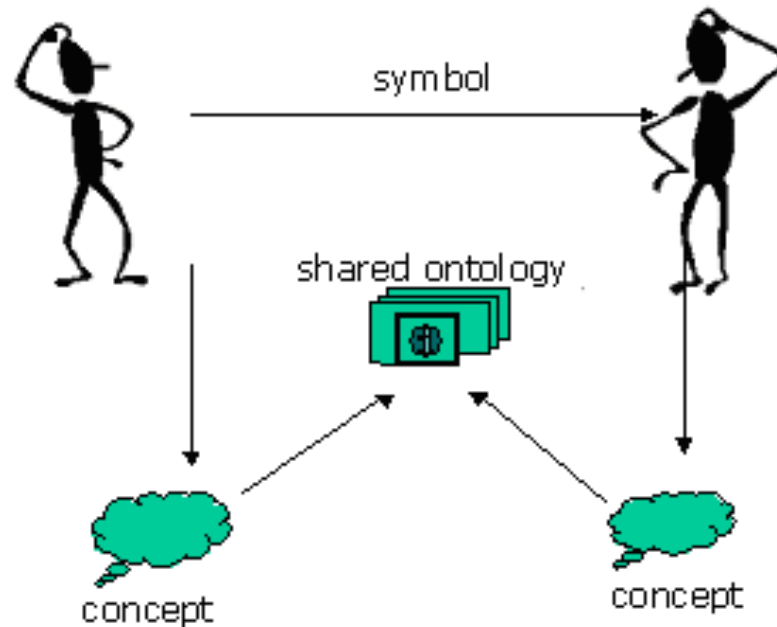
RDDL Esempio (3)

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rddl="http://www.rddl.org/">
  <rddl:resource
    xlink:role="http://www.crema.unimi.it/application/java-archive"
    xlink:arcrole="http://www.rddl.org/purposes/software#xslt-extension"
    xl:href="thisNS-xslt-extension.jar">
    <p>le estensioni XSLT definite in questo namespace sono in questo
      JAR</p>
  </rddl:resource>
</rdf:RDF>
```



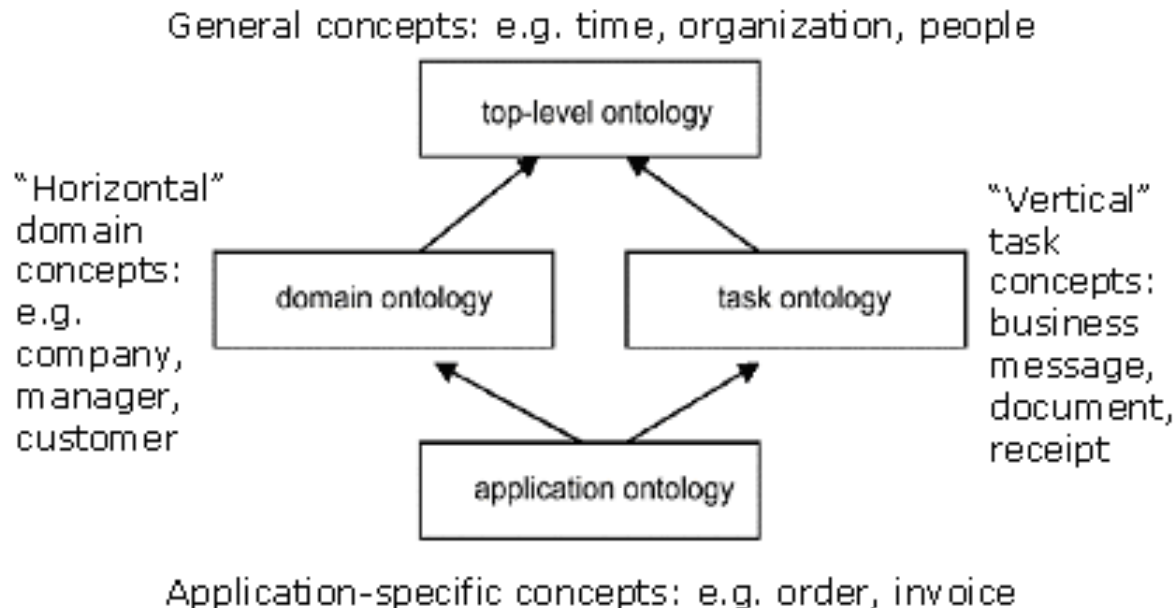
Ontology Engineering

Comunicazione intelligente usando ontologie



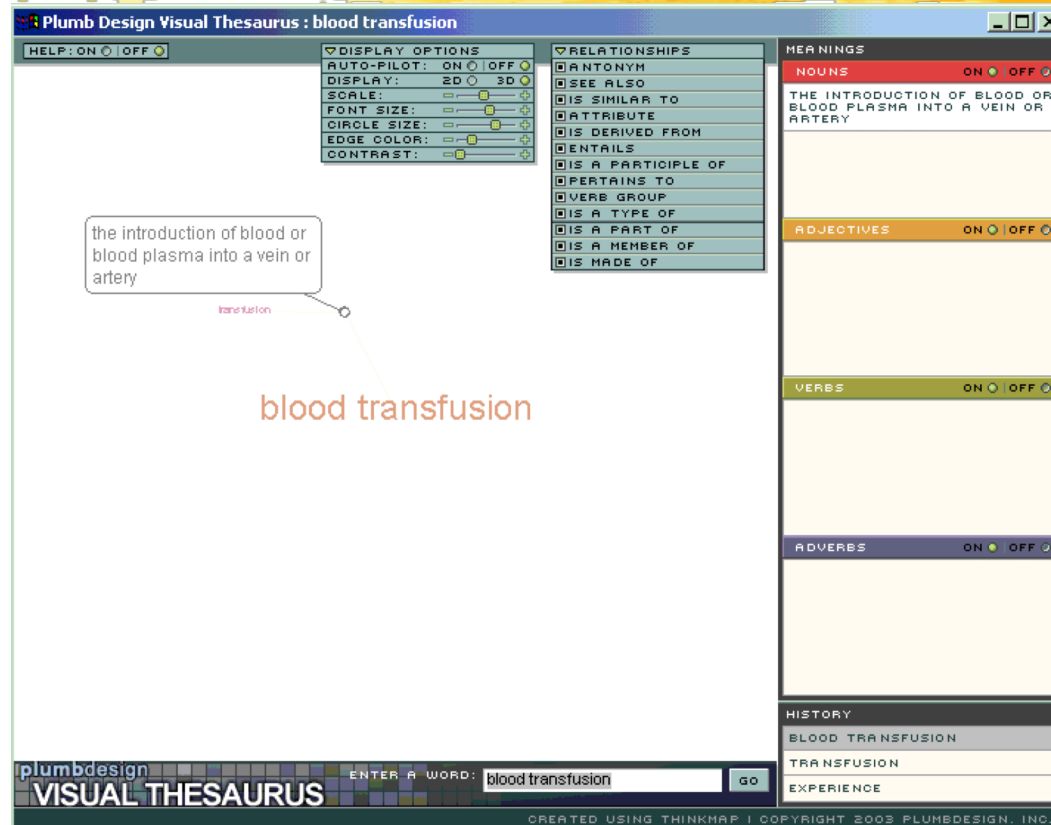
RDF Schema: linguaggio standard per rappresentare e scambiare ontologie

Tipi di ontologie



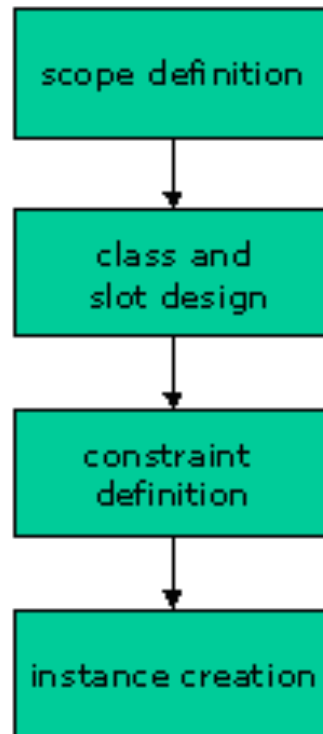
Necessità di progettazione **modulare basata su template** e di **chiara definizione di interfacce**

Dove prendo il vocabolario?



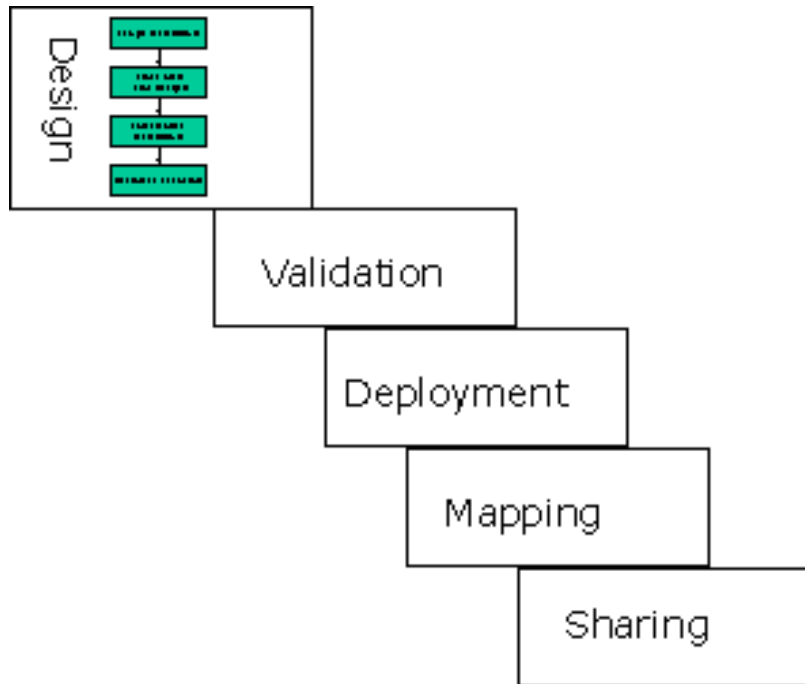
<http://www.cogsci.princeton.edu/~wn>

Passi del progetto di ontologie



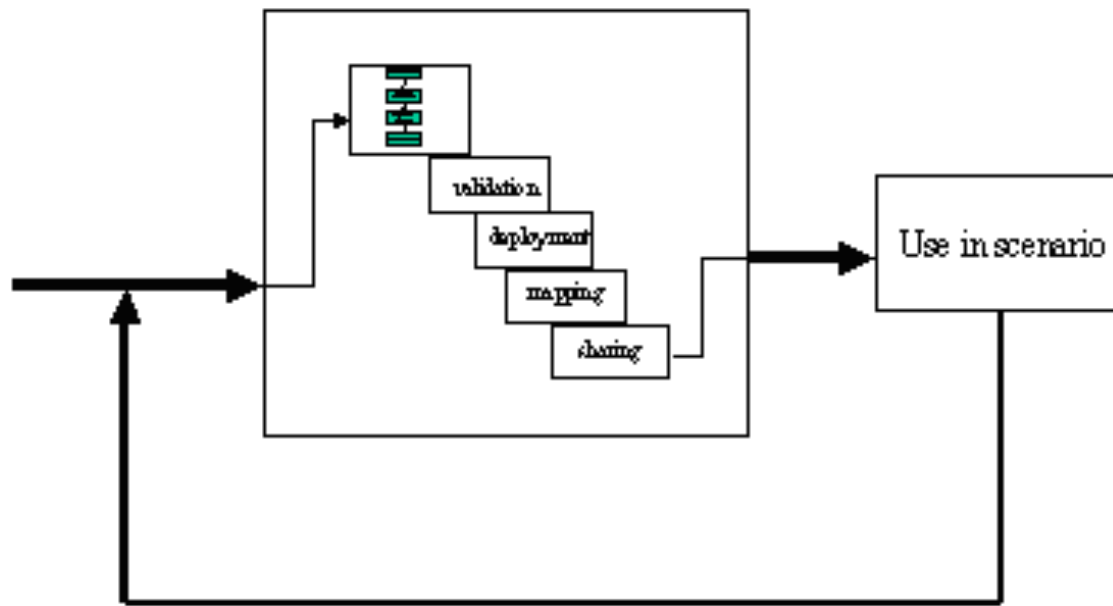
- Uso di strumenti visuali (e.g. UML)
- Ambienti specializzati per l'ontology design (e.g., OntoMaker)

Waterfall ontology lifecycle



- Ontologie come *software artifact*
- Passi di uno sviluppo industriale
- Tecniche di software engineering

Iterative ontology lifecycle

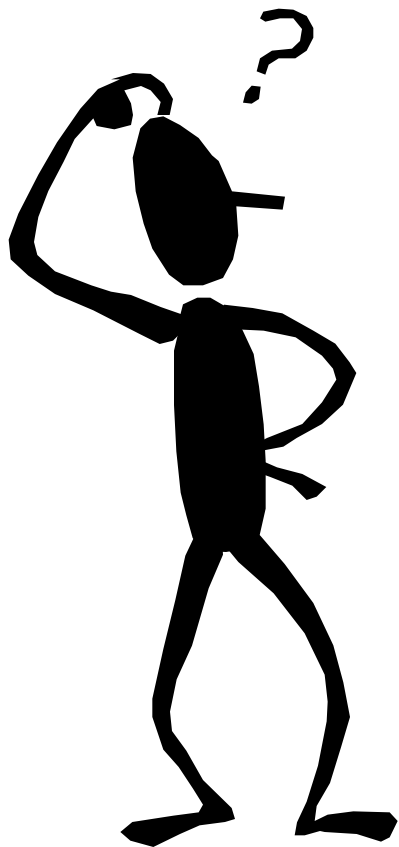


Problema aperto:
Necessità di far evolvere l'ontologia **mentre la si usa**

Conclusioni



- ❑ Web of Knowledge=Web + metadati
- ❑ Shared Web of Knowledge: Web + metadati espressi in un formato standard di riferimento Basato sullo spazio delle URI
- ❑ Schemi interessanti già disponibili
- ❑ Primi passi verso grandi ontologie di dominio
- ❑ Ci stiamo lavorando, **Crema+Lecce**
- ❑ Per saperne di più: <http://bingo.crema.unimi.it/ontology/>



Ci sono domande?