

Ontologie computazionali

Ontologie

- “Rappresentazione astratta di concetti e loro relazioni”
- Ontologie formali: rappresentate secondo un formalismo di rappresentazione
- Finalità: condividere una concettualizzazione comune tra individui, organizzazioni, macchine

Definizione di ontologia

An **engineering artifact**, constituted by a specific **vocabulary** used to describe a certain reality, plus a set of explicit assumptions regarding the **intended meaning** of the vocabulary itself" (Guarino, 1998).

Ontologie formali

- Le ontologie formali si basano su linguaggi che permettono di descrivere in maniera esplicita
 - Le proprietà delle classi
 - Le proprietà delle relazioni tra classi
- Questi linguaggi permettono alle macchine di fare inferenze sui concetti e a noi di avere certezza della validità di queste inferenze

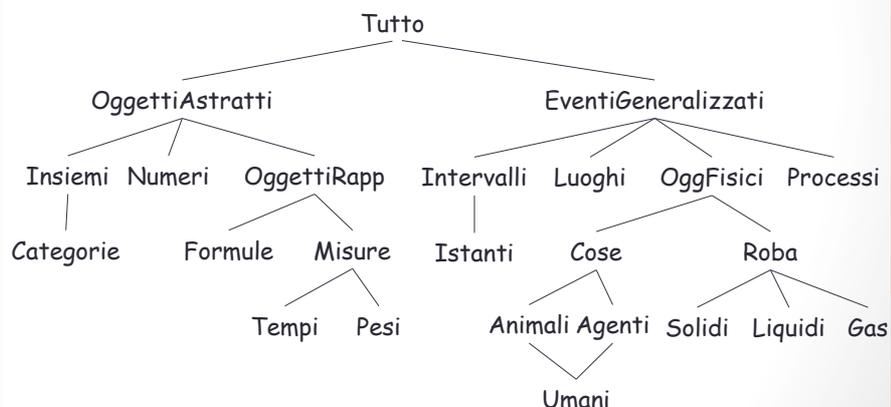
Tipi di ontologie

Una possibile classificazione delle ontologie fa riferimento al loro livello di specificità:

- Ontologie top-level: concetti fondazionali comuni a tutti i domini (spazio, tempo, ecc.)
- Ontologie mid-level: utilizzano il livello fondazionale per definire concetti generali ma non fondazionali: organizzazioni, comunicazione, stati fisici, sistemi di misurazione, ecc.
- Domain ontologies: rappresentano i concetti e le relazioni proprie di un dominio specifico

Ontologia top-level: esempio

- Russell-Norvig 2005



Il sistema CYC

- EnCYClopedic Knowledge, conoscenza enciclopedica
- Base di conoscenza finalizzata a rappresentare la conoscenza di senso comune (*commonsense knowledge*)
- Comprende circa 200.000 concetti
- Ne esiste una versione *open source* (OpenCyc)

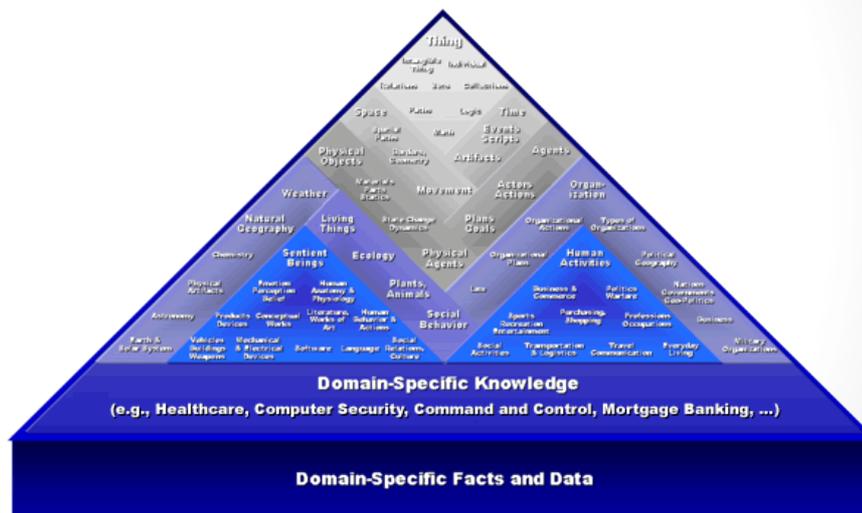
CYC

- La base di conoscenza (Knowledge Base, KB) di CYC consiste di
 - termini, il vocabolario di CYC
 - asserzioni che mettono in relazione questi termini
- Queste asserzioni includono
 - Asserzioni semplici
 - Regole

Struttura di CYC

- La KB di Cyc è divisa in molte microteorie (*microtheories*), ciascuna delle quali è costituita da un insieme di asserzioni che condividono le stesse assunzioni
 - Le microteorie si focalizzano su un particolare dominio di conoscenza
 - Attualmente sono migliaia
- Questa suddivisione permette al sistema di fare asserzioni che sarebbero apparentemente contraddittorie

Struttura della base di conoscenza



Esempi di regole

“Se Fred possiede un oggetto, possiede anche gli oggetti vicini a quello”

```
(#$forall ?X
  ($forall ?Y
    ($implies
      ($and
        ($owns #$Fred ?X)
        ($owns #$Fred ?Y))
      ($near ?X ?Y))))
```

“Se Fred possiede una bicicletta, è una bicicletta rossa”

```
(#$implies
  ($owns #$Fred #$Bike001)
  ($colorOfObject #$Bike001 #$RedColor))
```

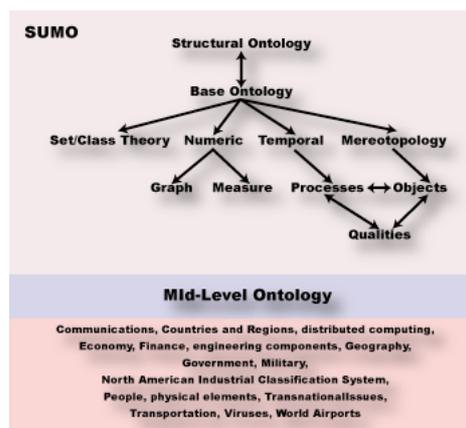
Come ragiona Cyc

- Il motore inferenziale di Cyc è in grado di effettuare la deduzione logica (*modus ponens, modus tollens, quantificazione universale e esistenziale*) e i meccanismi di inferenza tipici dell' IA (*ereditarietà, classificazione automatica*)
- La dimensione della base di conoscenza (200.000 termini e dozzine di asserzioni riguardanti ogni termine) hanno richiesto la messa a punto di tecniche speciali per affrontare la complessità

Applicazioni di Cyc

- Modellazione della conoscenza
- Apprendimento e pattern recognition
- Assistenti intelligenti
- Sicurezza delle reti
- Intelligent analysis, basi di dati, ecc.

Ontologia SuMO



IEEE
Suggested
Upper
Merged
Ontology

Scritta in
linguaggio KIF

KIF

- Knowledge Interchange Format (KIF), linguaggio in cui è scritta l'ontologia SUMO
- Contiene
 - Termini: individui e classi, con relazioni gerarchiche (instance e subclass)
(instance The82ndAirborne MilitaryUnit)
(subclass Person Animal)
 - Connettivi AND e OR
(and (instance KofiAnnan Human)
(occupiesPosition KofiAnnan
SecretaryGeneral UnitedNations))
 - Quantificazione e implicazione

Classi e relazioni

- SUMO contiene:
 - una gerarchia di classi
 - un insieme di relazioni
- Di ogni classe, SUMO descrive le caratteristiche attraverso un insieme di assiomi
 - Per esempio, un evento di mangiare comporta del cibo oppure una casa è l'abitazione di qualcuno
 - Tali assiomi sono espressi utilizzando le relazioni contenute nell'ontologia

Sumo: esempio di classe

Posizione della classe nella gerarchia: *PermanentResidence* è una sottoclasse di *Residence*

```
(subclass PermanentResidence Residence)
```

Documentazione: *PermanentResidence* è una *Residence* in cui una persona abita

```
(documentation PermanentResidence
  EnglishLanguage "A &%Residence where people
  live, i.e. where people have a &%home.")
```

Assiomi (regole) che valgono per il concetto: Data una certa *PermanentResidence* (?RESIDENCE), esiste una persona (?PERSON) per cui quella ?RESIDENCE è una casa

```
(=>
  (instance ?RESIDENCE PermanentResidence)
  (exists (?PERSON)
    (home ?PERSON ?RESIDENCE)))
```

Relazione part-of: proprietà e utilizzo per descrivere una classe

```
(instance properPart AsymmetricRelation)
(instance properPart TransitiveRelation)
(subrelation properPart part)
```

```
(=>
  (instance ?ROOM Room)
  (exists (?BUILD)
    (and
      (instance ?BUILD Building)
      (properPart ?ROOM ?BUILD))))
```

(documentation Room EnglishLanguage "A &%properPart of a &%Building which is separated from the exterior of the &%Building and/or other &%Rooms of the &%Building by walls.")

Ontologie lightweight

- Ontologie ‘leggere’
 - Normalmente semplici tassonomie, senza assiomi e con poche relazioni
 - Facile standardizzazione
- Esempi:
 - WordNet (lessico strutturato in classi di sinonimi, iponimi e iperonimi)
 - Tesauri e dizionari controllati
 - Iconclass
 - Soggettari
 - Dizionari controllati

Ontologie large-scale

- Sono ontologie di grandi dimensioni
 - CYC, YAGO, YAGO2, Dbpedia (e altre)
 - Possono essere ottenute tramite l'estrazione automatica di concetti da testi (Dbpedia e YAGO)
 - Dall'integrazione di risorse (YAGO e YAGO2) diverse, incluse risorse lessicali
 - Dal lavoro di una comunità di utenti (CYC), via *crowd sourcing*
- Date le dimensioni, gli strumenti di indicizzazione e di accesso acquisiscono grande importanza
 - Spesso vengono pubblicate online e come tali integrate in altre applicazioni

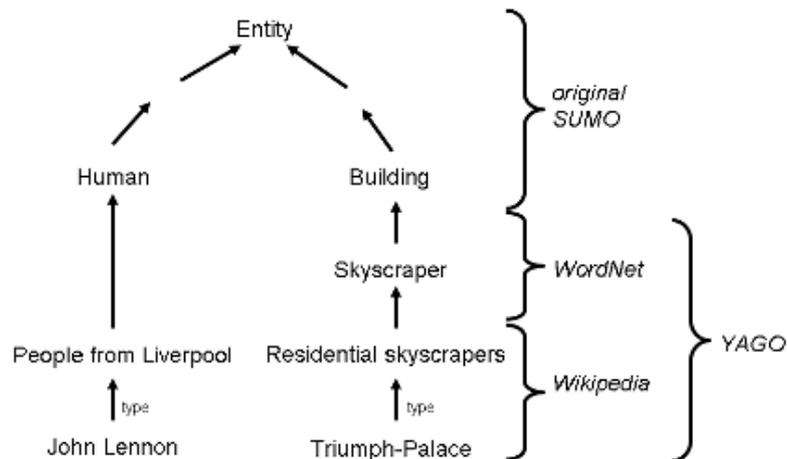
Interfaccia tra ontologie e linguaggio naturale

- Per l'accesso ai concetti dell'ontologia, è importante l'integrazione tra di essi e il linguaggio naturale
- L'integrazione avviene
 - attraverso la documentazione, cioè associando a ogni concetto una sua descrizione informale in linguaggio naturale
 - Associando ai concetti un equivalente lessicale in una risorsa linguistica esterna
- Esempi:
 - SUMO: Accesso mediato da WordNet e documentazione multilingua
 - YAGOSUMO e YAGO2: integrazione con WordNet

YAGOSUMO

- The YAGO-SUMO integration incorporates **millions of entities** from YAGO, which is based on Wikipedia and WordNet, into the Suggested Upper Merged Ontology (SUMO).
- With the combined force of the two ontologies, an enormous, unprecedented corpus of formalized world knowledge is available for automated processing and reasoning, providing information about millions of entities such as people, cities, organizations, and companies.
- Compared to the original YAGO, more advanced reasoning is possible due to the axiomatic knowledge delivered by SUMO. A reasoner can conclude e.g. that a child of a human must also be a human and cannot be born before its parents, or that two people sharing the same parents must be siblings.

YAGOSUMO: architettura e sorgenti di conoscenza



YAGO2

- YAGO2s is a huge semantic knowledge base, derived from Wikipedia WordNet and GeoNames. Currently, YAGO2s has knowledge of more than 10 million entities (like persons, organizations, cities, etc.) and contains more than 120 million facts about these entities.
- YAGO is special in several ways:
 - The accuracy of YAGO has been manually evaluated, proving a confirmed accuracy of 95%. Every relation is annotated with its confidence value.
 - YAGO combines the clean taxonomy of WordNet with the richness of the Wikipedia category system, assigning the entities to more than 350,000 classes.
 - YAGO is an ontology that is anchored in time and space. YAGO attaches a temporal dimension and a spacial dimension to many of its facts and entities.
 - In addition to a taxonomy, YAGO has thematic domains such as "music" or "science" from WordNet Domains.

YAGO2

- Interfaccia di interrogazione
- <https://gate.d5.mpi-inf.mpg.de/webyagospotlx/WebInterface>

DBpedia

- Iniziativa di ricerca iniziata nel 2007.
- DBpedia (from "DB" for "database") is a project aiming to extract structured content from ... the Wikipedia project.
 - This structured information is then made available on the World Wide Web.
- DBpedia allows users to query relationships and properties associated with Wikipedia resources, including links to other related datasets.
 - Parte del Linked Data effort. Scritta in linguaggio RDF.

<http://en.wikipedia.org/wiki/DBpedia>

DBPedia: esempio

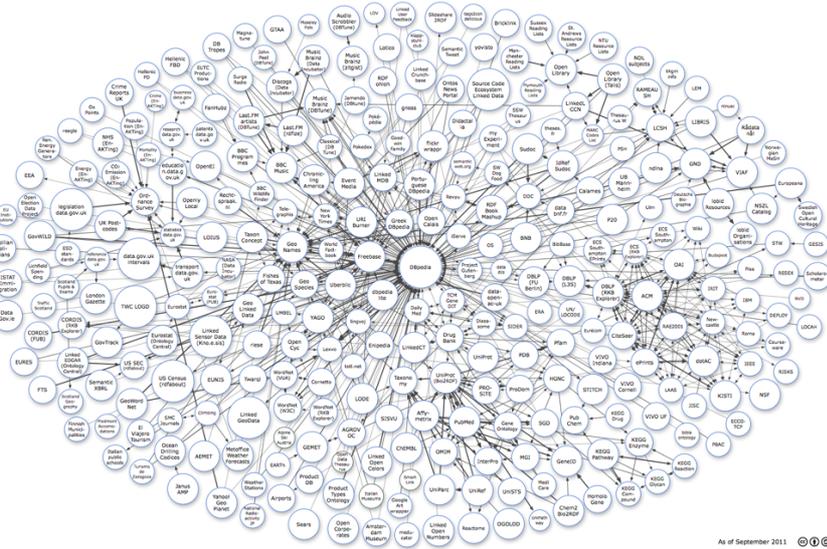
```
SELECT ?who, ?WORK, ?genre WHERE {  
  db:Tokyo_Mew_Mew dbprop:author ?who .  
  ?WORK dbprop:author ?who .  
  OPTIONAL { ?WORK dbprop:genre ?genre } .  
}
```

Trova autore, opera e genere del manga Tokio Mew Mew

Linked Open Data

- Dati pubblicamente disponibili (Open), pubblicati secondo il paradigma dei Linked Data:
 - Ogni dato si trova a un indirizzo web (http ...) che ne costituisce l'identificativo. Per es:
<http://dbpedia.org/ontology/Place>
 - La struttura interna del dato può essere descritta tramite il linguaggio RDF (Resource Description Framework)
- Un insieme di dati collegati tra loro è denominato *dataset* (per es. dbpedia.org)
- I dataset risiedono nella rete, formando il Linked Open Data (LOD) Cloud

Linked Open Data (LOD) Cloud



As of September 2011 ©

Linguaggi per descrivere ontologie

Concetti e relazioni

Processo di codifica

Ontologia



Contenuti mentali

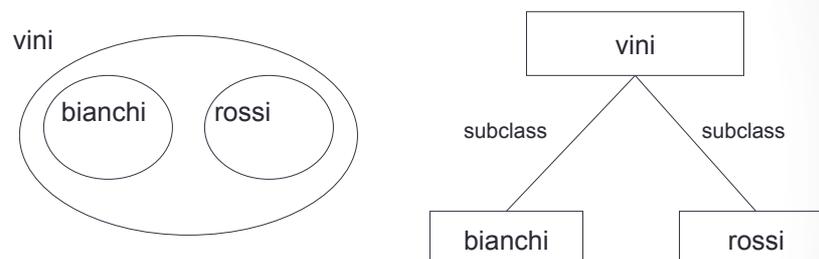
Primitive: classe, sottoclasse, relazione, ecc.

Documento di testo contenente l'ontologia

Relazioni tra classi

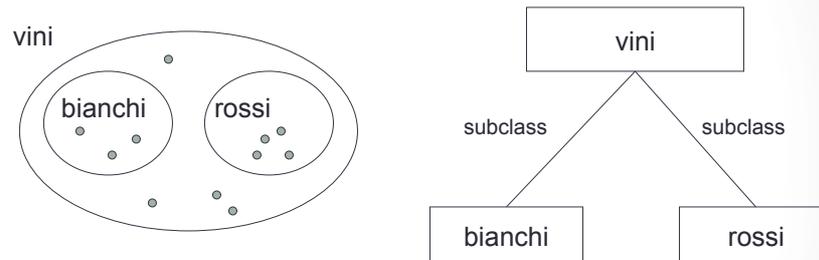
- Sottoclasse
 - Tutti gli elementi della sottoclasse sono elementi della (sovra)classe, ma non viceversa
- Due classi sono disgiunte se non hanno elementi in comune
 - Maschi e femmine, Vini bianchi, rossi e rosè

Classi disgiunte



Non esistono elementi di 'rossi' che appartengano anche a 'bianchi' e viceversa.

Classi disgiunte



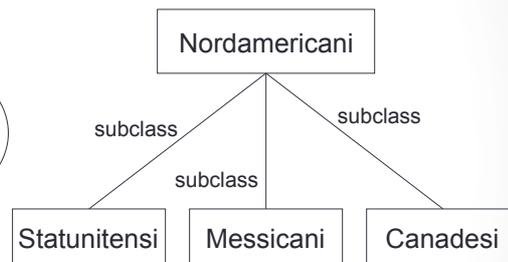
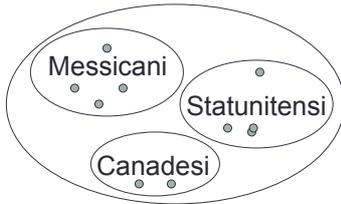
Ma i 'rosé' non appartengono a nessuna delle due classi!

Scomposizione esaustiva

- Due o più classi sono scomposizione esaustiva di una classe se tutti gli elementi della classe appartengono a una di esse
- {Statunitensi, Canadesi, Messicani} sono la scomposizione esaustiva di Nordamericani

Scomposizione esaustiva

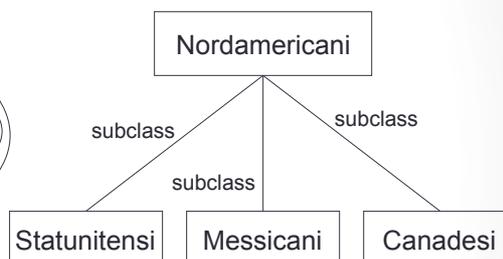
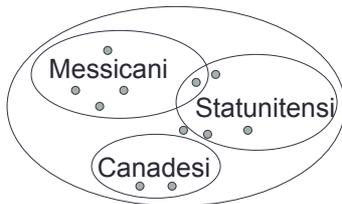
Nordamericani



Non esistono Nordamericani che non appartengano a uno dei tre insiemi

Scomposizione esaustiva

Nordamericani

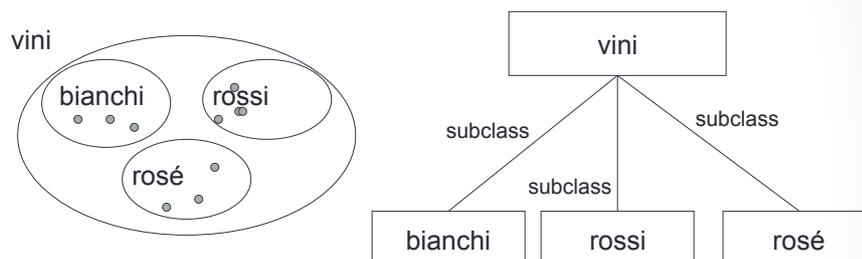


Ma esistono Nordamericani con la doppia cittadinanza

Partizione

- Partizione:
 - scomposizione esaustiva
 - classi disgiunte
- Esempi:
 - Maschi e femmine
 - Vini bianchi, rossi e rosé

Partizione



Tutti i vini sono bianchi, rossi o rosé
e non ci sono vini che appartengano a più tipi
contemporaneamente

Relazioni tra classi

- Problema: limitare la complessità del ragionamento
 - Soluzione: limitare l' espressività del linguaggio
- Relazioni tra classi:
 - Solo alcune relazioni possono essere espresse
 - Solo alcune proprietà delle relazioni possono essere espresse
 - Tipicamente non si rappresentano direttamente disgiunzione e negazione

Logiche descrittive

- Orientate alla classificazione.
- Basate sulla relazione di sottoclasse (sussunzione)
- Completezza e trattabilità computazionale
 - Esempi: sistema LOOM (87), BACK (88), linguaggio CLASSIC (91)
 - Oggi sistemi di ragionamento: JENA, RACER, OWLIM
- Sono la base del Progetto Web Semantico

T-Box e A-Box

- Nelle logiche descrittive si distinguono le definizioni di concetti dalle asserzioni sugli individui fatte utilizzando quei concetti
 - T-Box: Terminologia, cioè definizione di concetti generali
 - A-Box: Asserzioni su singoli individui

Terminologia vs asserzioni

- La terminologia è un insieme di assiomi logici su classi e proprietà
 - Sussunzione tra concetti (subclass)
 - Relazioni generiche in cui gli elementi di determinate classi rivestono un ruolo
- Data la terminologia, si fanno asserzioni su un insieme di individui
 - Le asserzioni devono essere coerenti con la terminologia (non si può dire che uno scapolo è sposato)

CLASSIC

Terminologia:

Scapolo = And (NonSposato, Adulto, Maschio).

Uomo = And (Adulto, Maschio).

And (Uomo, AtLeast(3, Figlio), AtMost(2, Figlia),
All(Figlio, And(Disoccupato, Sposato, All(Congiuge, Dottore))),
All(Figlia, And(Professore, Fills(Dipartimento, Fisica)))).

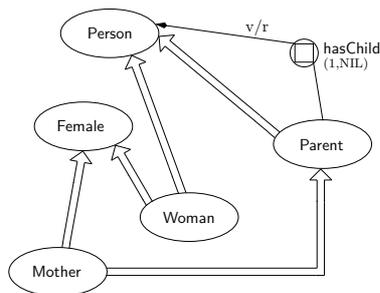
Asserzioni:

Luigi = OneOf (Scapolo).

RUOLI

- I ruoli costituiscono il mezzo per mettere in relazione i concetti
- In CLASSIC, il predicato che permette di esprimere il concetto di ruolo è FILLS
- Tipicamente, si possono porre restrizioni numeriche sui ruoli (Atleast, Atmost)
- Un motore inferenziale (reasoner) usa queste informazioni per effettuare ragionamenti

Esempio: ruoli in KL-One



Da *Introduction to Description Logics*
Nardi and Brachman, 2003

Le frecce bianche indicano relazioni di sussunzione tra le classi (classi e sottoclassi).

La relazione hasChild caratterizza la classe Parent
Parent è un tipo di Person che ha (almeno un figlio)

Il numero di figli è vincolato (restriction) a essere almeno 1 (ha un range da 1 a non specificato)

La classe Madre è definita come Mother e Parent

KL-One è una delle prime logiche descrittive

Cosa si chiede a un sistema basato su DL

- instance checking: verificare se un certo individuo (nella A-Box) appartiene a una classe
- relation checking: verificare se vale una certa relazione tra classi
- Subsumption: verificare se una classe è un sottoinsieme di un'altra classe
- concept consistency: verificare che le definizioni e le loro conseguenze non siano contraddittorie