

Le ontologie mediche

Giulia Benotto
Informatica Umanistica

7 Dicembre 2005

In generale: Le ontologie

Ontologie: Cosa sono?

- **Definizione filosofica:** Area della metafisica che studia come è realmente fatto l'universo che ci circonda attraverso lo studio dell'essere, le categorie fondamentali e le relazioni fra esse.
- **Definizione informatica:** Tentativo di formulare uno schema concettuale esaustivo e rigoroso nell'ambito di un dominio dato.
- Guarino e Giarretta (1995) propongono di usare la parola "ontologia" con la "O" minuscola per indicare le ontologie formali usate nel settore informatico e la parola "Ontologia" con la lettera maiuscola per indicare quelle attinenti al settore filosofico.

Basi di conoscenza: Cosa sono?

- Per **basi di conoscenza** si intendono corpi di informazione che formalizzano un universo descrivendo fatti e asserzioni considerati sempre veri all'interno del particolare dominio.
- Contengono informazioni generali sul dominio, dette di **background**.
- Le informazioni vengono catturate attraverso il dominio specifico, per progettare software specifici, realizzati per prendere in input il background di conoscenza e processarlo.
- Sono programmi che non processano **dati** ma processano **conoscenza**.

Differenze fra ontologie e basi di conoscenza

- Differenza principale è nel **dominio applicativo**.
- Gli schemi basati sui database hanno **task specifici** e sono orientati all'**implementazione**.
- Le ontologie sono maggiormente **indipendenti dai task**.
- Nella costruzione dell'ontologia si deve tenere conto che i dati in essa contenuta debbano essere **interscambiati**.
- Per tanto si devono utilizzare termini **condivisi**.
- I dati contenuti nelle basi di conoscenza sono impossibili da intercambiare.

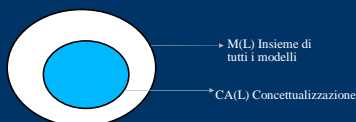
Definizione di Studer (1997)

"L'ontologia è una specificazione formale ed esplicita di una concettualizzazione condivisa"

- Per **concettualizzazione** ci si riferisce all'ottenimento di un **modello astratto** di un certo fenomeno del **mondo reale** attraverso l'identificazione dei **concetti rilevanti** che lo caratterizzano.
- Con **Esplicita** intendiamo che i concetti, le proprietà, i vincoli che caratterizzano il fenomeno siano **esplicitamente definiti** (per esempio, nel dominio medico il concetto di malattia e di sintomo sono legati da precise relazioni causali, una malattia provoca dei sintomi, ma non provoca sé stessa).
- **Formale** indica che un'ontologia sia machine-readable
- **Condivisa** in quanto la conoscenza su cui si basa l'ontologia non è frutto delle elucubrazioni di un singolo, ma condivisa da un gruppo che esprime il suo consenso su essa

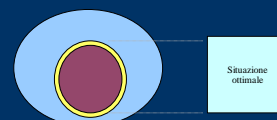
Definizione di concettualizzazione

- Studer (1997) definisce la concettualizzazione come la possibilità di specificare il modello astratto di alcuni fenomeni, costituito avendo identificato i concetti rilevanti del fenomeno stesso.
- Dato L come il linguaggio base composto da una serie di simboli V che costituiscono il vocabolario di tale linguaggio, avremo che quando un agente A ricorre a L, i modelli di L intesi da A costituiscono un sottoinsieme di tutti i possibili modelli definiti da L. Tale sottoinsieme è la concettualizzazione di V rispetto ad A.



Proprietà delle ontologie

- Le ontologie presentano due proprietà, dettate da quanto detto prima:
- **Coverage**: la copertura dei modelli intesi
- **Precisione**: La copertura dei modelli che sono nell'ontologia. La precisione è tanto maggiore quanto più i modelli si riducono a coincidere con i modelli intesi.
- **Coverage** e **precisione** possono diventare misure quantitative nel caso in cui il dominio sia **finito**



Componenti fondamentali

- **Tassonomia:** un set di termini, che insieme alle loro definizioni e alle relazioni fra i termini stessi, forma l'ontologia.
- **Concetti o classi:** può essere qualsiasi cosa su cui si può dire qualcosa e perciò può essere la descrizione di un task, una funzione, un'azione, una strategia o un processo di ragionamento. E' caratterizzato da un termine, da un'estensione e da un'intensione.
- **Estensione:** Insieme degli oggetti a cui il concetto può essere applicato.
- **Intensione:** Insieme di proprietà, caratteristiche e attributi usati per specificare la semantica del concetto, cioè l'insieme di caratteristiche che questi oggetti hanno in comune.
- **Relazioni:** Rappresentano un tipo di interazione fra i concetti del dominio.
- **Assiomi:** Affermazioni **sempre vere** sul dominio definito dell'ontologia e servono per specificare la semantica dei concetti. Generalmente specificano il modo in cui il vocabolario concettuale (concetti e relazioni) può essere utilizzato.

Metodologie per la costruzione delle ontologie

- **Metodologia:** un approccio sistematico per condurre un processo che implica attività che devono essere realizzati ad alcuni stati del processo di sviluppo delle ontologie.
- Si deve rendere possibile il riutilizzo e la condivisione della conoscenza.

Differenti approcci

- **Gruber (1998):** Necessità di separare la conoscenza dai programmi. La conoscenza viene descritta attraverso un linguaggio di rappresentazione dichiarativo. Identificazione di classi generali e relazioni che sottolineano fatti specifici dell'applicazione e che organizzano la conoscenza. Tali costrutti devono essere **ereditabili**.
- **Enterprise Ontology project (1998):** **Ontology capture** utilizzato per identificare i concetti più importanti piuttosto che i più generali o i più particolari, seguito da un processo di generalizzazione o specificazione al fine di ottenere il resto della gerarchia.
- **Tove project (1995):** Catturare degli **scenari**, cioè esempi che emergono da una situazione data. Motivano lo sviluppo delle ontologie e suggeriscono le possibili soluzioni che danno una semantica informale per gli oggetti e le relazioni che formeranno l'ontologia.
- **Framework di Uschold (1998):** Framework che in cinque fasi illustra i procedimenti per la costruzione delle ontologie.
- **Noy e Mc Guinness (2001):** Metodo basato sull'ingegneria della conoscenza, usato come tutorial nella costruzione delle ontologie.

Sistemi di informazione Ontology driven

- L'uso delle ontologie nel **processo di sviluppo** comporta alcuni vantaggi.
- Aiutano il progettista a migliorare la qualità dell'analisi concettuale, migliorano l'utilizzo e la mantenibilità del sistema di informazione.
- Forniscono un ampio livello di riutilizzo, perché permettono il riutilizzo di conoscenza, invece del tradizionale riutilizzo del software, e un aumento del riutilizzo e della condivisione delle applicazioni nel dominio di conoscenza usando un comune vocabolario attraverso piattaforme software eterogenee.
- Usare le ontologie al momento dell'utilizzo del sistema di informazione, fa sì che l'ontologia non sia una parte separata, accessibile a un IS, ma un componente dell'IS che collabora con tutte le altre componenti per realizzare il principale obiettivo.
- Sono usate per permettere la comunicazione fra agenti dei software fornendo un significato condiviso per gli agenti che si riferiscono ad essi.

Le ontologie in un loro settore di utilizzo: Le ontologie mediche

Perché il settore medico?

- La ricerca medica è afflitta da un problema di comunicazione.
- Comunità distinte di ricercatori si servono di tecnologie diverse e spesso incompatibili, questo causa problemi di integrazione.
- Attraverso l'uso delle ontologie sarebbe possibile unificare tutti i dati medici e renderli comprensibili a chiunque ne abbia accesso.
- Settore medico interessante anche per i progettatori di ontologie.
- Il dominio medico necessita di ontologie che permettano l'applicazione simultanea di prospettive distinte a una medesima realtà.
- Utili nella sperimentazione di nuovi farmaci

Tipologie di ontologie mediche

- All'interno del dominio medico, si può applicare la distinzione fra ontologie top-level (usate raramente), ontologie di dominio (UMLS) e ontologie basate sulla terminologia.
- Le ontologie top-level hanno a che fare con le categorie impiegate in un altro dominio come gli oggetti, gli eventi, le parti e il tutto, le istanze e le classi.
- Le ontologie di dominio, come abbiamo già visto, applicano il sistema top-level a un dominio particolare, nel caso medico possono occuparsi di terapie, disagi, cellule, geni.
- Un'ulteriore specificazione nella terminologia, si avrà ricorrendo alle ontologie basate sulla terminologia. Queste costituiscono un ampio sistema di livello ancora più basso, prendendo in considerazione termini quale, ad esempio, "noduli senza contratture".

Ontologie mediche

Strategie di costruzione

- Parte 1: Fornire una visione globale delle ontologie e delle ontologie top level che definiscono implicitamente.
- Parte 2: Mostrare come i principi di classificazione e le definizioni derivate dall'ontologia top-level possano aiutare nell'assicurare la qualità delle ontologie basate sulla terminologia e nell'allineamento delle ontologie.

Classificazione delle entità

Continuants Indipendenti	Continuants Dipendenti	Occorrenze (sempre dipendenti)
ORGANISMI MOLECOLE CELLULE	REGOLE FUNZIONI CONDIZIONI (Malattie)	PROCESSI STORIE VITE (decorsi delle malattie)
Classi		
Istanze		

Regole che governano i livelli

- I termini nelle gerarchie classificatorie devono essere divisi in livelli predeterminati (analogo alla classificazione biologica in regno, classe, ordine ecc.).
- I termini nella gerarchia partonomica devono essere suddivisi in predeterminati livelli di granularità (organismi, organi, cellule, molecole ecc.).
- E' possibile effettuare diverse partizioni della realtà a differenti livelli di granularità (proprio come possiamo suddividere l'organismo in cellule od organi)
- Si devono possedere le risorse per supportare, ad esempio, non solo un'ontologia anatomica a livello degli organi, ma anche a risoluzioni sempre più fini: dal livello cellulare a quello delle proteine
- Si devono supportare classificazioni di *processi a diverse risoluzioni, inclusi i processi chimici e biologici che avvengono nel corpo umano*

Ia-a overloading e uso di eredità multipla

- Il successo dell'allineamento delle ontologie dipende dal grado in cui relazioni basilari come *is_a* o *part_of* possono essere relazionate, dal grado in cui i significati si sovrappongono per permettere alle ontologie di essere allineate.
- L'uso di **eredità multipla** comprende l'assegnamento alla relazione *is_a* di una pluralità di significati diversi in una singola ontologia.
- Il melange risultante permette l'integrazione coerente attraverso ontologie ottenibili solo sotto la guida di esseri umani con una conoscenza biologica rilevante

Regole per le definizioni

- **Intelligibilità:** I termini usati in una definizione devono essere più semplici (più intelligibili) dei termini da definire, altrimenti la definizione non è d'aiuto né alla comprensione umana, né alla processazione da parte della macchina.
- **Sostituibilità:** sono anche chiamati contesti estensionali. Sono sostituibili, per definizione, in un modo che il risultato sia grammaticalmente corretto e abbia lo stesso valore di verità della frase di partenza.
- **Modularità (non si può definire tutto):**
- Isola i termini primitivi (=livello 0)
- Definisci i termini al livello $n+1$ per ogni $n \geq 0$ usando soltanto:
 - Termini resi dal livello n e oltre
 - Costanti logiche e ontologiche quali "e", "tutto", "is_a", "part_of"
 - Se queste regole non sono soddisfacenti, allora la ricerca degli errori e l'allineamento dell'ontologia può essere fatto soltanto con l'intervento dell'uomo

Esempi di ontologie mediche

- Negli anni ottanta gli obiettivi erano diversi da quelli odierni: le ontologie mediche erano finalizzate a costruire sistemi in grado di aiutare il singolo operatore sanitario a prendere decisioni ottimali.
- La realizzazione di sistemi di intelligenza artificiale legati al settore medico dovrebbe portare a cure più efficaci nei confronti del paziente, dato che tutti coloro che operano all'interno dei processi di cura dovrebbero essere in grado di utilizzare le conoscenze mediche oggi disponibili.
- **Modello fondazionale dell'Anatomia**: assicura consistenza e standard ultimi nella rappresentazione delle classi anatomiche. E' una risorsa bioinformatica per sviluppare contenuti anatomici in applicazioni per gruppi specifici.
- **IFOMIS**: approccio realista e idea che sia necessario, nella costruzione di un database, prestare attenzione al mondo a cui quel sistema deve essere applicato. Obiettivo dell'IFOMIS è quello di realizzare una teoria delle partizioni e delle interrelazione tra entità nella realtà stessa.

Costruzione di una base di dati riguardante il settore medico

Scopo del progetto

- La base di conoscenza è stata progettata con l'intenzione di essere d'aiutizio non solo ai medici, ma anche ai pazienti che si trovano a dover leggere una cartella clinica o affrontare una diagnosi senza saper interpretare i termini utilizzati per formularle.
- Attraverso la base di conoscenza si possono vedere e distinguere le relazioni fra i vari fenomeni e le loro interconnessioni, così che anche una persona che non conosce i fenomeni medici possa capire quali sintomi hanno condotto a una certa diagnosi.
- Lo scopo è quello di essere il più comprensibile possibile, usabile da tutti quelli che ne necessitano il supporto, siano essi membri dalle categoria medica o meno.

Progettazione dell'ontologia – Materiale utilizzato

- E' stato utilizzato anche un archivio di cartelle cliniche, dalle quali sono stati ricavati i dati utili, di ambito cardiologico.
- Per la progettazione dell'ontologia è stato utilizzato il programma DictEditor della ditta Synthema.
- Il programma è dotato di dizionari precaricati con regole grammaticali, sui quali si basano i dizionari specifici che vengono sviluppati in seguito. Le regole stabiliscono i metodi di flessione delle parole.



Progettazione dell'ontologia – Materiale utilizzato

- I contesti rappresentano macroaree di significato all'interno delle quali si organizzano i diversi termini. Ad ogni contesto possono appartenere molti termini, oppure molto pochi, secondo il livello di granularità che il progettatore ha deciso di dare alla base di conoscenza.



Metodologia di costruzione

- Analisi della terminologia delle cartelle cliniche attraverso due criteri fondamentali:
 - Conteggio delle occorrenze
 - Costruzione delle definizioni
- Costruzione delle relazioni svolta col supporto del testo di biologia.
- Suddivisione della terminologia in contesti diversi, che fornivano differenti livelli di granularità.
- Immissione delle relazioni.

Terminologia definita

- Tutti i termini usati per le definizioni dovevano essere conosciuti.
- La terminologia è stata selezionata in base al numero delle occorrenze e all'importanza che si assumeva avere all'interno del dominio.
- La terminologia è stata suddivisa in macroaree di significato, ovvero i contesti.
- I contesti sono stati creati secondo criteri di granularità.
- Il nome dei contesti di appartenenza è stato stabilito in base all'area di studio della medicina che riguarda il termine in questione, oppure l'area del corpo umano o l'apparato dello stesso, in cui l'oggetto cui il termine fa riferimento si trova.
- A questi vanno aggiunti tutti i processi chimici, le patologie, le possibili cure, i termini fisici.
- Si è scelto di non approfondire il livello di granularità, lasciando un contesto più generale.

Contesti

- Spettro ampio
 - Fisico
 - Chimico
 - Biologico
 - Patologia
 - Risoluzione_patologie
- Spettro ristretto
 - Apparato_circolatorio
 - Apparato_digerente
 - Apparato_respiratorio
 - Apparato_riproduttore
 - Sistema_nervoso
 - Sistema_neurologico
 - Urologia
 - Ematologia
 - Endocrinologia

Relazioni

- Il programma fornisce delle relazioni predefinite, fra le quali si deve scegliere per la costruzione della base di conoscenza.
- Relazioni come "membro di" o "parte di" erano adatte al nostro scopo. Relazioni non sono state usate in modo intercambiabile.
- "materiale componente di" è stata usata per sottolineare come alcuni organi siano composti da sostanze che a loro volta erano presenti nella base di conoscenza.
- La relazione "appartiene alla classe" è stata usata per sottolineare che le ci sono termini i quali a loro volta appartengono a classi più generali.

Descrizione del sistema

- Il sistema genera un dizionario indipendente, nel quale vengono raccolte tutte le entrate immesse in una struttura ben precisa, interpretabile dal sistema stesso e da altri programmi della ditta Synthema che interpretano lo stesso tipo di codifica.
- Esempio di entrata, codice generato per "cuore":
 - n#100002: NOUN DOMAIN(CARDIOLOGIA) DESCR(@Organo muscolare, cavo, che costituisce il centro motore dell'apparato circolatorio dei vertebrati) ALIAS (cuore: NOUN FEAT() IRULE(30) MEANING(0)) HASHOLOPART(n#100001: NOUN)**

Problemi avuti con il sistema

Versione A		Versione B	
Vantaggi	Svantaggi	Vantaggi	Svantaggi
<ul style="list-style-type: none">Modifica delle entrateInterfaccia dotata di tasto per l'accesso veloce alla visualizzazione e delle relazioni	<ul style="list-style-type: none">Poca stabilità del sistemaPossibilità di modificare le entrate parzialeQuantità limitata di entrate supportate (max. 50)	<ul style="list-style-type: none">Stabilità del sistema sufficienteTestato, se pur parzialmente	<ul style="list-style-type: none">Modifica delle entrate salvate non supportataSupporto di un numero di p.o.s. limitateSuperate le 500 entrate, bloccata la cancellazione