

Professionisti della conoscenza,
www.cstadvising.com

■ Per info:
dott. Massimiliano Miglio
+39 328.91.66.816

■ E-mail: advising@cstadvising.com
■ m.miglio@cstadvising.com

Dalla filosofia all'informatica: Ontologia una concettualizzazione di un dominio condiviso

Massimiliano Miglio, CSTAdvising

Stato Documento	Draft
Autore	Massimiliano Miglio
Versione	1.0
Release	0
History	
Descrizione	Cos'è un'Ontologia, approccio alla progettazione di un'ontologia e casi d'uso

Introduzione

Nel paragrafo *1.1 Le Ontologie* viene illustrato il concetto di ontologia dandone da prima una definizione filosofica e poi una definizione in ambito informatica.

Nel paragrafo *1.2 Organizzazione dei un Ontologia: la castagna ontologica* viene illustrata la castagna ontologica, ovvero i 3 livelli ontologici:

Upper Ontology, Application Ontology, Low Domain Ontology.

Nel paragrafo *1.3 Le Ontologie* viene illustrato il concetto di Knowledge Base.

Nel paragrafo *1.4 Approccio alla progettazione di un Ontologia* vengono esposti 4 metodologie da seguire nella fase di progettazione di un ontologia.

Nel paragrafo *1.5 Casi d'uso* viene proposto un elenco molto interessante di casi d'uso di ontologie

Nel paragrafo *1.6 Ontologie vs Database: schema concettuale vs dati* viene fatto un confronto tra l'utilizzo di un Database e l'utilizzo di un'ontologia nell'organizzazione di grosse quantità di informazioni.

Nel paragrafo ***Errore. L'origine riferimento non è stata trovata. Errore. L'origine riferimento non è stata trovata.*** viene mostrato un piccolo esempio di ontologia

1. Le Ontologie

1.1 Che cos' è un'ontologia

Esistono varie definizioni di “ontologia”. Dal punto di vista filosofico essa è vista come una “spiegazione sistematica dell' essere”. La più citata nel contesto informatico è quella di **Gruber** che la definisce come una “esplicita specificazione di una concettualizzazione”. Secondo **Neches** essa “...definisce i termini di base e le relazioni che includono il vocabolario di un'area e le regole per combinare i termini e le relazioni allo scopo di definire estensioni al vocabolario.”

Borst la definisce come una specifica formale di una concettualizzazione condivisa mentre **Guarino**: “...definisce una teoria logica che da una giustificazione esplicita e parziale di una concettualizzazione.”

Secondo quanto detto nel paragrafo *“Una nuova visione del Web”*, del 1° capitolo, un'ontologia rappresenta una concettualizzazione condivisa di un certo dominio, e nasce come un accordo consensuale sulla definizione di concetti e relazioni che caratterizzano la conoscenza del dominio stabilito, garantendoci la possibilità di applicare regole d'inferenza (ragionamento) sia per stabilire nuove asserzioni deducibili (nuova conoscenza sulla base di quella a disposizione), sia per organizzare e recuperare in modo intelligente ed efficiente le informazioni presenti sul web.

Essa contiene l'insieme dei concetti (entità, attributi, processi), le definizioni e le relazioni fra concetti, le quali possono essere di vario tipo: tassonomico (*IS-A*), meronimico (*PART-OF*), telico (*PURPOSE-OF*) ecc. Dunque è possibile vedere un'ontologia come una rete semantica di concetti appartenenti ad un dominio legati tra loro dalle suddette relazioni.

Formalmente un'ontologia può essere definita come una Tripla $O = \{C, R, A\}$, dove C = insieme di concetti del dominio di interesse, R = insieme di relazioni del tra i concetti appartenenti a C , A = insieme di assiomi (Se $A = \emptyset$ l'ontologia si dice non assiomatizzata).

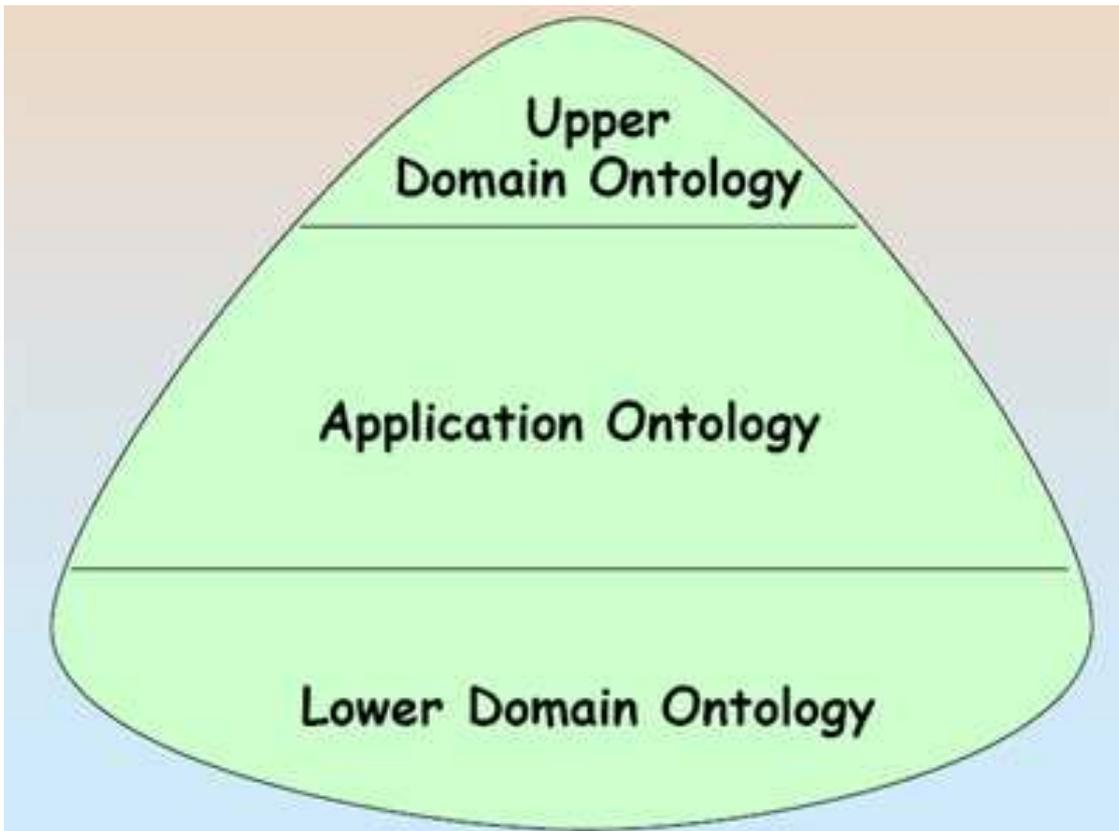
Gli insiemi C ed R individuano un Grafo $G = \{(V,E)\}$, tale che: $V \equiv C$, $E = \{(c1, c2) \in C \times C\}$.

Un'ontologia può presentare vari livelli di formalizzazione, ma deve necessariamente includere un vocabolario di termini (*concept names*) con associate definizioni (*assiomi*), e relazioni tassonomiche. Essa è una sorta di “stadio preliminare” di una “base di conoscenza”, il cui obiettivo è la descrizione dei concetti necessari a “parlare” di un certo dominio.

Una base di conoscenza include la conoscenza necessaria a modellare ed elaborare un problema, a derivare nuova conoscenza, a provare teoremi, a rispondere a domande concernenti un certo dominio.

La presenza di ontologie e la capacità di ragionare sulla conoscenza che esse ci permettono di rappresentare, consentono di emulare la logica umana basata sulla cognizione di causa in una determinata realtà.

1.2 Organizzazione dei un Ontologia: la castagna ontologica



Upper Domain Ontology: qui si collocano i concetti generali comuni a più domini applicativi, concetti dunque astratti, di livello filosofico.

Application Ontologies: qui si collocano concetti specifici ma non elementari delle singole applicazioni che caratterizzano uno specifico dominio.

Lower Domain Ontology: qui si collocano i concetti elementari che non possono essere ulteriormente rifiniti o decomposti.

Secondo tale suddivisione possiamo dire che il livello Upper Domain Ontologies è l'anello di congiunzione delle varie ontologie del livello Application Ontologies, essendo queste una specializzazione dell'Upper Domain Ontologies.

1.3 Knowledge base

Un'ontologia, come visto nei paragrafi precedenti, si pone l'obiettivo di modellare un dominio d'interesse, definendone lo schema concettuale.

Questa però non è sufficiente a descrivere in modo esaustivo la conoscenza di dominio, ed è dunque necessario aggiungere altri componenti.

Per arrivare ad avere una Knowledge Base completa, ovvero la completa base di conoscenza che rappresenta fedelmente ed esaurientemente un determinato dominio d'interesse, dobbiamo aggiungere un insieme di fatti ed un insieme di regole.

Dunque possiamo definire Knowledge Base (KB) l'insieme costituito da un:

- ✓ Ontologia, schema concettuale del dominio in questione, in cui si definiscono degli assiomi sui concetti da rappresentare.
- ✓ Insieme di fatti, ovvero di istanze dei concetti definiti nell'ontologia.
- ✓ Regole, insieme di regole sul quale inferire nuova conoscenza, ovvero sul quale ragionare per aggiungere nuovi fatti.

1.4 Approccio alla progettazione di un'Ontologia

1.4.1 Approccio Bottom-Up

Nell'approccio Bottom-Up la progettazione avviene dal basso verso l'alto, ovvero si individuano prima tutti i concetti più specifici, poi quelli più generali. Tale procedimento va iterato fino al raggiungimento di un livello di astrazione (generalizzazione) sufficiente per gli scopi per cui viene progettata l'ontologia. I concetti saranno organizzati secondo una tassonomia che rappresenta la gerarchia degli stessi.

Come per il livello di generalizzazione, il livello di specializzazione dei concetti individuati deve tener conto degli scopi dell'Ontologia.

1.4.2 Approccio Top-Down

Nell'approccio Top-Down la progettazione dell'ontologia avviene invece dall'alto verso il basso, ovvero individuando prima i concetti più generali, e poi, via via quelli più specifici, arrivando ad un livello di specializzazione adeguato per gli scopi dell'ontologia.

Anche qui i concetti saranno organizzati in una tassonomia che rispetta la gerarchia degli stessi.

Come per il livello di generalizzazione, il livello di specializzazione dei concetti individuati deve tener conto degli scopi dell'Ontologia.

1.4.3 Approccio Misto

Nell'approccio misto, come suggerisce il nome, si procede individuando prima tutti i concetti che si ritengono importanti per il dominio da rappresentare, poi i legami gerarchici che legano i concetti per organizzare gli stessi dal più generale al più specifico.

Nota: in tutti e 3 gli approcci (Bottom-Up, Top-Down, Approccio Misto), dopo aver individuato una tassonomia (struttura gerarchica di concetti), lo step ultimo necessario, è individuare tutti gli altri legami che legano i concetti della tassonomia. Questa infatti arricchita dalle nuove relazioni costituisce un'ontologia.

1.4.4 Metodologia semi-automatica: OntoLearn

L'approccio OntoLearn, presentato dalla prof.ssa Paola Velardi del dipartimento di Scienze dell'Informazione dell'Università di "Roma La Sapienza", si colloca come un approccio totalmente nuovo, con l'idea di fornire un supporto semi-automatico nella fase di progettazione di un'ontologia. Questa fase infatti potrebbe essere molto gravosa, ed è soprattutto molto delicata, quindi un supporto semi-automatico, gestito da un Ingegnere della conoscenza, affiancato da un esperto di dominio, vuole ridurre i tempi, la fatica e gli errori di progettazione di un'Ontologia.

OntoLearn si basa su 4 step e su un set di documenti che rappresenta in modo esaustivo il dominio da rappresentare:

- ✓ Creazione di un Lessico in modo semi-automatico, a partire da un set di documenti di dominio
- ✓ Creazione di un Glossario, a partire dal lessico creato al punto 1. Il glossario conterrà le definizioni dei termini appartenenti al lessico individuato al punto 1.
- ✓ Creazione di una tassonomia, organizzazione gerarchica dei termini individuati al punto 1. In questa fase si utilizzerà non solo il lessico individuato al punto 1 e le definizioni individuate al punto 2.
- ✓ Arricchimento della tassonomia con relazioni di vario genere che trasformano la tassonomia in una ontologia.

La parola semi-automatico, indica che ad ogni fase, l'intervento di un ingegnere della conoscenza e di un esperto di dominio avrà il compito di validare l'output della fase corrente, per avere un input corretto da sottoporre alla fase successiva.

In particolare, nella fase 1, mediante tecniche di NLP (Processamento del linguaggio naturale) il set di documenti dato in input viene elaborato con lo scopo di dare in output una lista di termini ritenuti rilevanti per il dominio in questione. Un esperto di dominio, validerà la lista di termini proposta, approvando ciò che ritiene davvero importante, e rigettando i termini ritenuti poco rilevanti.

A questo punto il set di termini approvati (SRT: Set Relevant Term) , costituirà l'input per la fase 2. L'output della fase 2, sarà costituito da una serie di definizioni associate a ciascun termine presente dell'SRT. Anche qui l'intervento umano, ed in particolare di un esperto di dominio, validerà le definizioni estratte, approvando ciò che ritiene giusto e rigettando ciò che non ritiene utile. L'output di questa fase è dunque un Glossario di dominio, che costituirà l'input della fase 3.

Nella fase successiva dunque il glossario verrà elaborato con tecniche di NLP allo scopo di individuare una serie di legami gerarchici tra i termini del lessico.

L'output della fase 3 propone una tassonomia, ovvero un'organizzazione dei termini del lessico e del glossario, secondo relazioni gerarchiche (Relazioni IS-A).

A questo punto, un esperto di dominio, ed un ingegnere della conoscenza, collaboreranno per validare la tassonomia, aggiungendo o eliminando concetti e relazioni.

La tassonomia validata, insieme al glossario, costituirà l'input per la fase 4, in cui il software, sempre attraverso tecniche di NLP, tenterà di proporre in output una ontologia, ovvero tenterà di arricchire la tassonomia data in input con relazioni rilevanti tra i concetti tassonomici (relazioni part-of, purpose-of, ecc..). L'output di questa fase sarà poi validato ed ampliato manualmente dal lavoro congiunto di un esperto di dominio e da un ingegnere della conoscenza.

Alla fine della 4° ed ultima fase l'ontologia è pronta per l'uso.

1.5 Casi d'uso

1.5.1 Portali Web

Un portale web è un sito web che fornisce contenuto informativo per un dominio di interesse. Un portale web permette alle persone interessate al dominio ricevere notizie, trovare altre persone e parlare con loro, costruire una comunità e trovare materiale utile.

Nel caso specifico di un portale che intende fornire materiale informativo, è importante offrire un servizio ad alta fruibilità, per garantire agli utenti di accedere alle informazioni desiderate nel minor tempo possibile e soprattutto nel modo più semplice ed efficiente possibile. Molto spesso, il materiale informativo che il portale web vuole offrire è organizzato in modo manuale dai responsabili del portale stesso, secondo categorie e sottocategorie. Altre volte invece sono i fornitori stessi del materiale, che lo etichettano mediante semplici metatag, che identificano il contenuto informativo del materiale stesso.

In ogni caso comunque, un'organizzazione manuale dei contenuti non è sempre ottimale, e spesso con grandi moli di dati risulta essere del tutto inadeguata ai fini di una buona fruibilità.

Una soluzione a questi problemi è quella di definire un'ontologia di dominio che permette una classificazione intelligente delle risorse gestite nel portale.

Mediante l'ontologia si mette a disposizione dei fruitori una terminologia che descrive in modo efficiente ed efficace i contenuti del portale web. Ad esempio, un'ontologia potrebbe includere terminologia come "Giornale", "pubblicazione", "persona" e "autore". Quest'ontologia può includere definizioni ed assiomi che dichiarino, ad esempio, "tutti i giornali sono pubblicazioni" oppure "gli autori di tutte le pubblicazioni sono persone". Quando queste definizioni si combinano con fatti, permettono di inferire altri fatti, ovvero di arricchire la base di conoscenza, inferendo nuova conoscenza di dominio. Attraverso queste inferenze gli utenti possono effettuare ricerche intelligenti ed efficienti, ottenendo risultati impossibili attraverso convenzionali sistemi di ricerca. Questo perché l'ontologia non solo si rivolge ai fruitori del portale, ma costituisce un mezzo formale che mette gli agenti software nelle condizioni di comprendere la semantica delle risorse. OntoWeb è un esempio di portale basato in un'ontologia.

1.5.2 Collezioni Multimedia

Le ontologie si possono usare per annotare semanticamente collezioni d'immagini, audio o altri oggetti non testuali. La semantica di tali oggetti infatti è difficilmente estraibile dalle macchine.

Le ontologie multimedia possono essere di due tipi: media-specific e content-specific. Le ontologie media-specific presentano tassonomie per i diversi media type e descrivono le proprietà dei diversi media. Ad esempio, il video può includere proprietà che identifichino la durata del clip e degli intervalli.

Le ontologie content-specific descrivono invece il tema della risorsa, ovvero il suo contenuto.

Essendo queste ontologie non dipendenti dal formato della risorsa che si intende descrivere, possono essere riutilizzate per la descrizione di qualsiasi oggetto digitale (html, pdf, word, jpg, gif, ecc.), il cui contenuto informatico sia coerente con il dominio in questione.

Qualora invece è necessario far riferimento ad uno specifico media-type, è possibile combinare le due ontologie media-specific e content-specific.

1.5.3 Management Aziendale.

Le aziende importanti è spesso necessario gestire numerose pagine web relative ai comunicati stampa, offerte di prodotti e studi, procedure aziendali, informazioni interne sui prodotti, descrizioni di processi, individuazione di esperti in un determinato dominio, documentazione tecnica, documentazione di progetto, ecc.. Si possono usare ontologie per indicizzare questi documenti ed in generale queste informazioni e fornire mezzi di ricerca migliori. Sebbene molte organizzazioni abbiano già una tassonomia per organizzare l'informazione, spesso questa risulta insufficiente. Infatti, una singola tassonomia o una singola ontologia risulta poco adeguata perchè non adatta a descrivere in modo esauriente tutti i domini di interesse aziendale; lavorare simultaneamente con molte ontologie aumenterebbe invece la precisione e la ricchezza informativa nella fase di descrizione dei domini.

All'interno di un'azienda, ad esempio potrebbe essere necessario effettuare ricerche efficienti per individuare un venditore con particolari caratteristiche, un esperto di domini con un profilo ben delineato per poterlo inserire all'interno di un progetto, documentazione relativa a progetti passati per riutilizzare esperienze e conoscenze acquisite internamente all'azienda, ecc..

Utilizzare più ontologie dunque mette a disposizione una terminologia accessibile a tutte le figure presenti nell'azienda, ed inoltre dà la possibilità di creare delle viste semantiche dello stesso dominio. Infatti, mentre un sistemista può avere bisogno di cercare materiale utile alla manutenzione di un sistema di clustering, uno sviluppatore di software, della stessa azienda, potrebbe avere bisogno di recuperare materiale utile alla progettazione di software su ambienti clusterizzati. Dunque sulla base di differenti ontologie di interesse un base documentale potrebbe assumere diverse viste semantiche in modo da soddisfare tutti i tipi di ricerca.

1.5.4 Agenti e servizi: decision making

Il Web Semantico può fornire agenti software che abbiano la capacità di comprendere e integrare diverse risorse d'informazione. Un esempio specifico è quello di un pianificatore d'attività sociali,

che può ricevere le preferenze di un utente (come il tipo di film che preferisce, che tipo di cibi mangia, ecc.) ed usare quest'informazione per pianificare le attività serali di un utente in una sera.

Questo tipo d'agenti software richiede la presenza di ontologie di dominio i cui concetti rappresentano e descrivono i servizi che si mettono a disposizione di un utente (ristoranti, hotel, B&B, night-club, discoteche, cinema, ecc..) e di ontologie di servizi.

1.5.5 Anagrafica di persone: Realtà azienda, Web Community, ecc..

Capita molto spesso di avere l'esigenza di dover categorizzare e ricercare in modo efficiente delle persone interne ad una ben definita realtà. Ad esempio internamente ad una grande realtà aziendale è spesso necessario individuare delle figure professionali che rispettino un determinato profilo, per poterle poi inserire all'interno delle attività azienda, oppure nelle web community, oggi molto di moda tra i giovani come luogo di incontro e di socializzazione, è necessario individuare quei partecipanti alla community che hanno uno specifico profilo, in modo da poter selezionare un insieme di persone a noi più congeniali. Un ontologia di dominio potrebbe fornirci i mezzi necessari alla categorizzazione dei profili di un dipendente azienda, o di un partecipante della web community, così che le ricerche potrebbero essere più mirate ed intelligenti.

Le ricerche infatti sarebbero basate sulla formulazioni di query semantiche che descrivono in modo preciso il profilo che stiamo cercando ed il sistema software sottostante sarebbe in grado di restituirci risultati più precisi.

1.5.6 Piattaforme e-learning

Le ontologie ed il Web Semantico sono oggi un mezzo molto importante anche per il supporto alle piattaforme e-learning. Un'azienda che vuole realizzare una tale piattaforma non può prescindere da un'organizzazione ottimale ed efficiente dei corsi proposti e del materiale didattico ed integrativo proposto ai learner. Una o più ontologie fornirebbero la giusta terminologia per l'organizzazione di tutto il materiale e di tutte le attività formative proposte.

Inoltre attraverso le ontologie è oggi possibile creare piattaforme e-learning che mettono a disposizione dei learner percorsi formativi personalizzati, sulla base della conoscenza pregressa dello studente e degli obiettivi che si pone il corso stesso.

1.6 Ontologie vs Database: schema concettuale vs dati

Se pur molto evidente per gli esperti di rappresentazione della conoscenza, la differenza tra Ontologie e Database, risulta essere non sempre molto chiara.

Molti infatti potrebbero obiettare che entrambi servano ad immagazzinare dati e che l'utilizzo di ontologie nella gestione di grosse quantità di dati non porti nessun valore aggiunto rispetto all'utilizzo di un comune database.

Per comprendere a fondo la differenza tra l'utilizzo di un ontologia e di un database, è necessario comprendere che utilizzare un ontologia significa adottare un approccio orientato allo schema concettuale, che descrive il dominio in cui ci muoviamo, mentre utilizzare un Database Relazionale, significa affrontare il problema con un approccio orientato ai dati.

Si potrebbe obiettare che anche un database relazionale ha un proprio schema concettuale che descrivere la realtà in cui ci troviamo, ma il potere espressivo dei linguaggi utilizzati per definire i due schemi concettuali (ontologia e database relazionale) è ben differente, e dunque le informazioni che questi ci offrono sono enormemente differenti.

Un ontologia infatti è molto più ricca di informazioni rispetto ad un semplice schema concettuale di un database. Inoltre, uno schema concettuale di un database, ha lo scopo unico di descrivere le entità (i dati), che andremo ad immagazzinare, mentre un ontologia ha lo scopo di descrivere i concetti.

Dunque, mentre le entità di uno schema concettuale di un DB sono descritte solamente attraverso una lista di attributi e di relazioni, i concetti di un ontologia presentano una descrizione molto più ricca.

Come visto nei paragrafi precedenti infatti, OWL, Ontology Web Language ci mette a disposizione una serie di costrutti appartenenti alle logiche descrittive, e dunque ci consente di andare ben oltre il semplice elenco di attributi. La descrizione di un concetto infatti, avviene mediante la composizione ricorsiva di costrutti logici che ci permettono di definire un concetto come l'intersezione, l'unione, il complemento, la disgiunzione o l'enumerazione di più concetti.

Le relazioni tra i concetti, possono essere definite come relazioni funzionali (invertibili e non), e/o come relazioni transitive, e/o simmetriche.

Infine, possiamo anche definire una serie di restrizioni sui singoli attributi di un concetto.

Tutto ciò dunque, rende chiaro che il potere espressivo di un linguaggio come OWL, sia di gran lunga maggiore a quello di un linguaggio SQL Like, e che lo schema concettuale di un DB porta con se un minor numero di informazioni rispetto ad un ontologia.

A questo punto, ci si potrebbe chiedere a cosa serve avere un potere espressivo così alto, se un DB relazionale è in grado farci immagazzinare prima e recuperare dopo le informazioni che vogliamo trattare.

Un potere espressivo più alto corrisponde ad una migliore organizzazione dei contenuti e dunque ad una migliore fruibilità degli stessi. I contenuti che immagazziniamo sono infatti catalogati secondo lo schema concettuale in uso. La fruibilità dei contenuti stessi dipende ovviamente dal tipo di catalogazione. Dunque più lo schema concettuale che utilizziamo per catalogare i nostri contenuti è elevato, più la catalogazione che possiamo fare è efficiente.

Una catalogazione fatta attraverso un semplice DB relazionale, pur se ben progettata, rispecchia in termini di efficienza la poca ricchezza di informazioni che porta con se lo schema concettuale del DB. La catalogazione fatta attraverso uno schema concettuale ricco come quello di un ontologia, ci consente di sfruttare in fase di recupero delle risorse catalogate, la ricchezza delle informazioni che l'ontologia stessa ci offre, e dunque la fase di recupero delle risorse catalogate, risulta essere più efficiente.

Un database relazione infatti è accessibile attraverso query SQL like, ovvero espressioni logiche più tosto semplici, in grado di farci recuperare solo ciò che abbiamo realmente e fisicamente immagazzinato.

L'accesso ad un ontologia invece, può essere fatto utilizzando dei reasoner, ovvero dei software capaci di ragionare. Attraverso il ragionamento artificiale di questi reasoner, siamo dunque in grado di effettuare ricerche molto più complesse ed efficienti (la complessità delle ricerche dipende come detto prima dalla complessità dello schema concettuale in uso, e dunque dall'ontologia).

Un reasoner infatti è in grado di inferire conoscenza, ovvero di apprendere nuova conoscenza sfruttando lo schema concettuale in uso (ontologia) ed i fatti conosciuti al momento della ricerca.

Un reasoner è dunque in grado di “emulare” il ragionamento umano, andando a recuperare quelle informazioni non esplicitamente espresse, ma comunque presenti in modo implicito.

Ad esempio, affermiamo che:

1. Mario è padre di Giuseppe
2. Giuseppe è fratello di Andrea,
3. Bruno è zio di Andrea.

Attraverso l'interrogazione di un semplice DB relazionale non potremmo sapere che Bruno è anche lo zio di Giuseppe. Un reasoner invece, attraverso un processo inferenziale, è in grado di darci questa informazione, che pur se non espressa in modo esplicito, è *vera* e presente nel nostro dominio.

Dunque, in un contesto in cui è necessario immagazzinare e catalogare una grande quantità di informazioni, un approccio ai dati (DB relazionale) risulta essere alla lunga meno efficiente di un approccio concettuale (ontologia), con ovvia conseguenza di una peggiore fruibilità de contenuti, a discapito degli utenti che impegnano il loro tempo e le loro energie per raggiungere risultati non sempre efficienti.