



SEMINARIO DI CULTURA DIGITALE (aa 2014/2015)

Relazione di Laura Gorrieri

ORGANIZZAZIONE DELLA CONOSCENZA

**La sfida di organizzare la conoscenza e la risposta
delle ontologie**

Indice dei contenuti

Introduzione.....	1
§1 Organizzazione della conoscenza.....	3
§2 Organizzare la conoscenza: le sfide.....	6
§3 Organizzare la conoscenza: le ontologie.....	11
§3.1 I grandi pensatori del passato.....	12
§3.2 I moderni sulle spalle dei giganti.....	14
Conclusione.....	20
Bibliografia di lavoro.....	22
Bibliografia.....	22
Sitografia.....	23

Introduzione

Se organizzare la conoscenza è un compito che svolgiamo istintivamente in quanto esseri umani, che si oppongono a quella che viene chiamata *knowledge soup*, ovvero un insieme fluido di stimoli e informazioni che riceviamo dal mondo esterno in maniera frammentata ed incoerente ed in continuo divenire, formalizzare la conoscenza è invece un compito che nella vita quotidiana tralasciamo, mentre storicamente parlando la sfida è stata raccolta dai filosofi e dai logici, prima di passare il testimone agli informatici. Sarà quindi necessario considerare la complessità e le sfide poste da questo fluido in continuo divenire che è la *knowledge soup* e i problemi posti per la sua gestione a livello di sistemi informatici.

Come osserva Guarino, questo ambito di ricerca si connota con una forte interdisciplinarietà: nonostante la necessità di ordine sia insita nel nostro modo di vedere ed esaminare il mondo, le competenze necessarie per una classificazione formale sono da rintracciarsi tra diverse discipline, ponendo questa ricerca in secondo piano rispetto ai temi centrali delle stesse:

One of the reasons of the lack of interest towards ontology in **classical AI** research lies in the fact that problems like ontology and conceptual modelling need to be studied under a *highly interdisciplinary* perspective: besides the basic tools of logic and computer science, an open-minded aptitude towards the subtle distinctions of philosophy and the intricate issues of natural language and commonsense reality is necessary.

GUARINO N. 1995, p.3. Corsivo presente nel testo.

L'impulso ad organizzare il mondo, a mettere ordine nel caos nel quale ci muoviamo, ha sempre accompagnato filosofi e pensatori e più in generale qualsiasi persona che abbia compiuto un lavoro che si possa definire creativo. I sistemi creati per rispondere a questa esigenza si sono sempre focalizzati sul pensiero e sulle capacità cognitive degli esseri umani, mentre oggi il problema si ripropone specularmente per i computer. Per poter rispondere a questa necessità è necessario indagare e capire più a fondo che tipo di lavoro faccia il nostro

cervello quando organizza il mondo nel tentativo di proporre una visione coerente dello stesso. Ma è veramente possibile riuscire in questa impresa? Il nostro modo di pensare e di ordinare differisce sensibilmente da ciò che una macchina è in grado di fare, ma fortunatamente vi è un punto di incontro comune: la razionalità. Vedremo quindi come la logica - espressione formale della *ratio* - è un momento fondante indispensabile per l'organizzazione della conoscenza e ci permette di creare un dialogo con le macchine, che elaborano il mondo grazie ad una sequenza ordinata di 0 e di 1. Ma basterà la logica per ordinare completamente il mondo, o meglio, per creare una struttura che sia in grado di fornire un'adeguata rappresentazione del mondo? Se alcuni filosofi classici pensavano che questo fosse possibile, vedremo come questo pensiero subisca dei cambiamenti avvicinandosi all'epoca contemporanea e come questi diversi sistemi di pensiero portino a differenti modelli di organizzazione della conoscenza.

Il presente lavoro è stato diviso in tre sezioni. Nella prima tratteremo della disciplina dell'organizzazione della conoscenza in generale: vedremo quindi che cos'è, da che livelli è composta e quali sono gli approcci pratici più comuni. Nella seconda sezione invece discuteremo del problema della formalizzazione della conoscenza, in quanto il mondo in cui viviamo ci fornisce degli stimoli tra loro incoerenti e che sono da conciliarsi con la nostra necessità di ordinamento e in che modo la esprimiamo. Nelle terza e ultima sezione considereremo le ontologie come sistemi di organizzazione della conoscenza e vedremo come si sono evoluti dalla nascita della logica ad oggi.

§1 Organizzazione della conoscenza

Il dubbio cresce con la conoscenza.

JOHANN WOLFGANG VON GOETHE

Con il termine *organizzazione della conoscenza* (dal tedesco Wissensordnung, tradotto con knowledge organization, spesso abbreviato in KO) si fa riferimento ad un campo di studio interdisciplinare che ha diverse connessioni con altre discipline, come la filosofia, la biblioteconomia, la computer science, la linguistica, la sociologia ed indica «l'insieme delle tecniche e degli strumenti che permettono di archiviare e presentare il sapere umano in forma di sistemi ordinati e coerenti»¹. Comprende al suo interno aspetti sia teorici che pratici e raggruppa «a wide variety of topics and approaches, from highly abstract ideas to the concrete needs of users in searching and browsing information»². La KO si basa su una struttura divisa a livelli, ognuno dei quali rappresenta una fase per poter arrivare all'effettiva descrizione indicizzata dell'oggetto preso in analisi:

1. Il primo livello consiste nella *teoria di organizzazione della conoscenza*. È la fase più astratta e raccoglie teorie ed approcci alla conoscenza che provengono da discipline esterne come la filosofia, la psicologia, la filosofia della scienza, l'ermeneutica ed altre.
2. Il secondo livello è relativo al *sistema di organizzazione della conoscenza* (knowledge organization system o KOS), il quale include schemi come liste di parole chiave o vocabolari controllati, classificazioni, oppure ontologie digitali. Inizialmente i KOS erano stati pensati per scopi specifici e locali, ma con gli sviluppi successivi si è fatta strada l'esigenza di condivisione, la cui risposta è stata data dal terzo livello.
3. Il terzo livello è di natura tecnica e riguarda la creazione di *standard e formati* come MARC (*MAchine Readable Cataloguing*), RDF (*Resource Description Framework*), SKOS (*Simple Knowledge Organization System*) e OWL (*Ontology Web Language*), per la diffusione dei dati.
4. Il quarto e ultimo livello è l'applicazione pratica delle teorie e dei sistemi

1 GNOLI C. 2013, p 1.

2 GNOLI C. 2011, p 30.

a collezioni esistenti di istanze di conoscenza, come archivi, biblioteche, mostre o ai loro cataloghi, siano essi digitali o cartacei.³

Per quanto riguarda il primo livello è necessario sottolineare come la KO theory dipenda innanzitutto dalla visione del mondo che si vuole adottare in base al risultato che si vuole raggiungere e non vi è quindi una teoria universale, ma un insieme di teorie con differenti caratteristiche che vanno soppesate. Questo approccio si riflette anche sul livello successivo:

knowledge organization system has a bias to some philosophical position, and proposes the key is to mediate between different views and develop the system in accordance with the goals and values of the users for which the system is intended.

McILWAINE I.C., MITCHELL J.S 2008, p. 80.

I KOS si possono organizzare gerarchicamente in base alla loro capacità di descrizione semantica delle risorse, tenendo presente che ad una maggiore precisione semantica corrisponde una maggiore complessità di scrittura e gestione. Si parte quindi da strutture semplici come l'organizzazione per titolo o parole chiave. Questo approccio non è sempre auspicabile: un titolo di un libro, ad esempio, potrebbe non essere informativo sul suo contenuto, mentre le parole chiave potrebbero non dare informazioni sufficienti e si porrebbe il problema dell'ordinamento delle stesse (in ordine alfabetico? per importanza?). L'utilizzo di vocabolari controllati - come ad esempio i soggettari o i tesauri - è considerato già più informativo, poiché mette in relazione i termini che descrivono una risorsa con termini semanticamente affini, rendendo più facile l'organizzazione di contenuti simili sotto etichette simili. Vi sono poi strutture che sfruttano ancora di più le potenzialità delle relazioni tra concetti, come le tassonomie che creano annidamenti gerarchici di appartenenza oppure le classificazioni che usano lettere e simboli e descrivono una risorsa grazie ad una stringa alfanumerica in cui ogni carattere o sottostringa rimanda ad un concetto secondo una notazione precedentemente stabilita. Infine, tra i livelli più alti si collocano le ontologie, le quali esprimono, attraverso delle

³ CFR *ibidem*.

dichiarazioni formali, concetti, relazioni tra concetti e restrizioni sul tipo di relazioni possibili.

Come evidenzia Claudio Gnoli, la KO è applicata in molteplici occasioni in cui si vuole rispondere ad un'esigenza di organizzazione e coerenza e solleva diversi problemi sull'ordine e l'indicizzazione di una collezione di documenti, di qualsiasi tipo.

Questo tipo di problemi è pratico ma al contempo filosofico, perché l'organizzazione stessa del materiale conoscitivo evidentemente influenzerà la sua fruizione, contribuendo a orientarne la recezione presso i suoi utenti. Si tratta dunque, allo stesso tempo, di un campo di applicazione della filosofia - riconosciuta quale informatrice almeno implicita di qualsiasi possibile approccio all'organizzazione della conoscenza - e di un aspetto della filosofia stessa, se è vero che anche illustri filosofi si sono posti il problema della sistematizzazione delle conoscenze.

GNOLI C. 2013, p 2.

Purtroppo, per questioni di spazio, non ci è permesso approfondire il tema dell'organizzazione della conoscenza in tutti i suoi molteplici aspetti. Ci preme però segnalare al lettore che è possibile reperire informazioni e nuove tendenze, nonché le informazioni per la consultazione del journal di knowledge organization, sul sito di *ISKO (International Society for Knowledge Organization)* e di *ISKO Italia*. Il lettore troverà i rimandi ad entrambi nella sitografia.

§2 Organizzare la conoscenza: le sfide

There are more things in heaven
and earth, Horatio, than are dreamt
of in your philosophy.

Hamlet - WILLIAM SHAKESPEARE

Le persone hanno un desiderio naturale di organizzare, classificare, etichettare e definire cose, eventi e modelli della loro vita quotidiana, che si mostra a noi naturalmente non ordinata. Con il termine *knowledge soup* si intende la natura fluida e in divenire delle informazioni che le persone ricevono, sulle quali costruiscono idee e schemi mentali e tramite le quali comunicano con altri esseri umani. Come affrontare ed estrarre conoscenza da questo sistema fluido è una cosa che non viene insegnata, ma fa parte della nostra natura. Proprio per questo motivo è evidente che le macchine non possono imparare a farlo esattamente come gli esseri umani:

The reasoning ability of the human brain is unlike anything implemented in computer systems. A five-dollar pocket calculator can outperform any human on long division, but many tasks that are easy for people and other animals are surprisingly difficult for computers. Robots can assemble precisely machined parts with far greater accuracy than any human, but no robot can build a bird nest from scattered twigs and straw or wash irregularly shaped pots, pans, and dishes the way people do. For recognizing irregular patterns, the perceptual abilities of birds and mammals surpass the fastest supercomputers.

SOWA J.F. IN RAMADAS J., CHUNAWALA S. 2004, p. 55.

Prendiamo ad esempio una sfida classica che mette a confronto uomo e macchina nella loro capacità di codificare ed accedere alla conoscenza in tempo reale: un simile scenario si è verificato quando nel 1997 il campione mondiale di scacchi ha giocato e perso una partita contro un supercomputer, che si è quindi dimostrato migliore di quanto un uomo possa mai essere nel prevedere e calcolare le mosse degli scacchi, grazie ad un'implementazione dei pattern dei pezzi. Ma se invece consideriamo un gioco come Go, un antico

gioco da tavolo cinese le cui regole sono più semplici rispetto a quelle degli scacchi, si è osservato come nessun computer possa andare oltre il livello di gioco 'principiante'. Questo avviene perché per giocare a Go non è sufficiente avere una banca dati dei pattern e delle mosse possibili, ma è necessario avere la capacità di analizzare strutture visive sempre nuove formate dalla disposizione di dozzine di pedine bianche e nere sul campo di gioco.

La natura della conoscenza e di come viene organizzata assume quindi un ruolo centrale anche nel campo dell'intelligenza artificiale: un'organizzazione sistematica della conoscenza è ottima per rendere i concetti più semplici da imparare e da richiamare, tuttavia non è possibile esaurire la realtà incasellandola in un'organizzazione sistematica; nonostante l'ordine apparente, risulta comunque evidente che vi sarà sempre uno scarto tra un sistema e il reale: anche provando a ridurre un'entità alla sua rappresentazione, vi sarà sempre un quid che ci sfugge, il quale è formato da eccezioni, complessità, trasversalità dell'oggetto di studio; è legato ad altri aspetti impossibili da esaurire in via formale e va a comporre l'ambiguità che permea la nostra esistenza. A causa dell'esistenza e della conformazione di questa knowledge soup si pongono difficili sfide per ogni sistema che tenti di organizzare la conoscenza in modo coerente e informativo.

Il maggiore ostacolo cui vanno incontro sistemi come questi è proprio quello della complessità della knowledge soup, la quale non deriva dal modo in cui il cervello umano funziona e nemmeno da come i linguaggi naturali sono strutturati per esprimere le informazioni. Come ha osservato Whitehead, logico e matematico del XX sec, questa complessità nasce da quella 'zona di penombra' del nostro pensiero, da dove fanno capolino domande che esprimiamo in modo semi-incosciente e che guidano la nostra esperienza verso i nuovi stimoli. Possiamo individuare alcune tipologie di pensieri frammentati che fanno parte della zona di penombra:

- *Sovrageralizzazioni*: trarre delle conclusioni basate su informazioni che sono troppo generali o non abbastanza specifiche.

Gli uomini camminano tutti su due gambe. E allora gli storpi? Quelli in

sedia a rotelle? I neonati? Un uomo che dorme? Un acrobata che fa la verticale?

- *Condizioni anormali ed eccezioni*: partire da una definizione ritenuta vera ed iniziare a chiedersi che tipo di eccezioni possono occorrere.

Se hai una macchina puoi guidare da Pisa a Bologna. Ma che succede se si scarica la batteria? O se ti è scaduta la patente? Se ci fossero problemi sulla strada?

- *Definizioni incomplete*: data una definizione ci iniziamo ad interrogare se ci siano altre condizioni necessarie per descrivere l'oggetto in questione. Un pozzo di petrolio è un buco scavato nel terreno che estrae petrolio. E se il petrolio esaurisce? È comunque un pozzo anche se per qualche motivo viene chiuso? Tre scavi collegati ad uno stesso giacimento sono tre pozzi o è uno solo?

- *Conflitti nel sistemare un concetto in due categorie che si escludono*: quando vi è un ente della nostra realtà che rientra in due categorie che per definizione si escluderebbero.

Al Festival della Filosofia sono invitati scrittori e filosofi di pregio della nostra epoca. L'edizione del Festival della Filosofia del 2012 è stata aperta da Fabio Volo. Che cos'è quindi Fabio Volo? Un romanziere o un filosofo di pregio?

- *Applicazioni inaspettate*: partendo da una definizione generale si ci chiede che applicazione possa avere su casi limite.

Le parti del corpo umano sono descritte nei libri di anatomia. I capelli sono una di queste? Una parrucca? Una parrucca fatta di capelli della persona che la indossa? Un capello che si è spezzato ed è caduto? Le unghie ne fanno parte? Quelle finte? Se la pelle ne fa parte, i tatuaggi ne fanno parte? Gli impianti di pelle? Gli organi trapiantati? Le protesi?

Questi pensieri sorgono a causa della natura del mondo nel quale ricorrono eccezioni e casi limite. Nessun linguaggio formale, basato su parole o simboli discreti, potrà mai catturare la complessità di un sistema in divenire come quello in cui ci muoviamo noi esseri umani: il cambio continuo di gradazioni e

un numero indefinito di eccezioni in cui possiamo incappare, rendono impossibile creare un sistema che si basi su concetti precisi e definiti di ciò che viene imparato attraverso l'esperienza.

Un'altra differenza importante che caratterizza il nostro modo di acquisire le informazioni è quella che riguarda le inferenze possibili che il nostro cervello applica quando partendo da una serie di premesse cerca una conclusione soddisfacente. In epoca classica si riteneva che il nostro pensiero si basasse solo sulla deduzione - inferenza logica del tipo che si usa nei sillogismi classici, mentre Peirce, logico e matematico contemporaneo di Whitehead, ha evidenziato che quest'ultima è certamente importante, ma non è l'unico processo inferenziale che viene applicato dagli esseri umani. Vi sono altri metodi che andrebbero considerati, come l'induzione, che dal caso particolare ricava il generale, e l'adduzione, che funziona proponendo una nuova ipotesi che possa spiegare le premesse. Nella tabella sottostante abbiamo riportato dei semplici esempi per rendere più chiare le differenze tra queste inferenze.

	DEDUZIONE	INDUZIONE	ADDUZIONE
Premesse	I. Tutti gli uccelli volano. II. Titti è un uccello.	I. Titti, Polly e Cippi sono uccelli II. Eddie è un pipistrello IV. Titti, Polly e Cippi volano V. Eddie vola.	I. Tutti gli uccelli volano. II. Titti vola.
Conclusione	Titti vola.	Tutti gli uccelli volano.	Titti è un uccello

Tra questi metodi solo l'adduzione è quella che può proporre come conclusione una idea nuova: grazie alla valutazione di una qualche misura di distanza semantica, l'adduzione seleziona dei tasselli rilevanti dalla knowledge soup e li combina insieme per creare qualcosa di nuovo. Si può applicare in diversi modi:

- Riuso: selezionando una regola o una teoria da applicare al problema in esame, dopo una ricerca associativa per affinità semantica tra quelle già

incontrare in precedenza.

- Revisione: trovando una teoria o un frammento di una teoria che corrisponda approssimativamente al problema in esame e applicandovi una serie di procedimenti per adattarlo alla situazione corrente.
- Combinazione: ricercando frammenti sparsi o tasselli di conoscenza ed eseguire diversi passaggi per combinarli o modificarli al fine di renderli utilizzabili per il problema attuale.

Secondo Peirce, affianco a deduzione, induzione e adduzione si colloca un quarto metodo, l'analogia, che combina le caratteristiche dei primi tre:

Analogy is more primitive than logic because it does not require language or symbols. Its only prerequisite is *stimulus generalization* — the ability to recognize similar patterns of stimuli as signs of similar objects or events. In Peirce's terms, logical reasoning requires symbols, but analogical reasoning could also be performed on image-like signs called *icons*.

SOWA J.F. IN RAMADAS J., CHUNAWALA S 2004, p. 66. Corsivi nel testo

È quindi chiaro che differenze tra noi e i computer sono molteplici e si legano sia al modo in cui gli esseri umani ragionano, ma soprattutto al modo in cui le informazioni sono distribuite nel fluido di azioni-reazioni che è la realtà. Questo non significa che è necessario desistere dal tentativo di organizzare e formalizzare la nostra conoscenza, bensì ci mette in guardia sulle sfide alle quali andremo incontro nel farlo e ci spinge a migliorare costantemente. Questo sforzo va nella direzione di creare un sistema di organizzazione della conoscenza funzionale ed efficiente che impara da come gli esseri umani ragionano, ma non potrà mai sostituire del tutto le nostre capacità specifiche.

§3 Organizzare la conoscenza: le ontologie

La scienza è ciò che sappiamo e la
filosofia è ciò che non sappiamo.

BERTRAND RUSSELL

Abbiamo visto cosa si intenda per organizzazione della conoscenza e quali sfide sia necessario fronteggiare a causa della natura delle informazioni che riguardano la realtà. Andiamo quindi a vedere nel dettaglio i tentativi di risposta a queste sfide, concentrandoci in particolare su quegli esperimenti riguardanti le ontologie. Come si è accennato in precedenza, le ontologie sono uno dei KOS di livello più alto. Ma che cos'è esattamente un'ontologia?

La parola *ontologia* deriva dal greco ὄντος, ontos, genitivo singolare del participio presente del verbo essere e da λόγος, logos, ovvero 'parola, discorso, pensiero'. *L'ontologia* come disciplina studia le categorie delle cose che esistono o possono esistere in un certo dominio del reale. Il prodotto di questo studio è *un'ontologia* ovvero un catalogo di cose che si pensa esistano o possano esistere in un dominio **D**, descritte in un linguaggio **L**, solitamente un linguaggio formale espresso tramite la logica. La logica da sola non dà nessuna informazione, ma quando viene combinata con un'ontologia si crea un linguaggio **L** in grado di esprimere relazioni a proposito delle entità nel dominio di interesse **D**.⁴

È quindi utile delineare un percorso di come lo studio delle ontologie è nato e come si è evoluto durante il tempo. Per questioni di sintesi non potremmo soffermarci su ogni aspetto come esso meriterebbe. Il mio scopo, in questa sede, è quello di mostrare come sia stato affrontato il tema di organizzazione della conoscenza in ontologie, dai classici alla modernità e in che modo abbia risposto alle sfide poste nel corso dei secoli.

4 SOWA J.F. 2001, p. 2.

§3.1 I grandi pensatori del passato

Nonostante il termine 'ontologia' venga dal greco, è stato introdotto relativamente di recente, ovvero nel XIX secolo dai filosofi tedeschi per distinguere lo studio dell'essere in filosofia dallo studio che compie la scienza degli esseri viventi. Il termine tradizionale per lo studio delle diverse tipologie di entità era quello di *categoria*, introdotto per la prima volta da Aristotele che ne descriveva dieci, secondo lui in grado di raggruppare e classificare tutto ciò che esiste: sostanza, qualità, quantità, relazione, il dove, il quando, l'aver, l'agire, il giacere, il subire. Il sistema di Aristotele è rimasto il più importante per molto tempo ed ancora oggi troviamo degli echi della sua filosofia nello studio dell'organizzazione della conoscenza:

For over two thousand years, Aristotle's categories and his system of syllogisms for reasoning about the categories were the most highly developed system of logic and ontology.

SOWA J.F. IN RAMADAS J., CHUNAWALA S. 2004, p. 57.

Il passo decisivo che ha reso il suo sistema fondamentale sembra essere la nascita della logica e in particolare quella del *sillogismo*, ovvero di un'inferenza logica che da due premesse arriva ad una conclusione preservandone il valore di verità. «La condizione perché il sillogismo sia corretto è che [...] nelle due premesse siano presenti e collegati due termini detti estremi e un terzo termine, detto termine medio» ; la conclusione quindi «sarà costituita dal collegamento fra due estremi»⁵. I sillogismi sono spesso applicati a frasi espresse con il linguaggio naturale, ma in una forma piuttosto stilizzata, che viene anche chiamata *linguaggio naturale controllato*: viene letto come se fosse linguaggio naturale, ma le persone che lo scrivono devono avere una preparazione sulla forma da utilizzare; ad un occhio inesperto potrebbe non essere subito evidente che la struttura della frase è condizionata dallo scopo di creare un sillogismo formalmente corretto. È il caso del classico sillogismo

5 PALLADINO D. 2010, p.6.

- I. Tutti gli uomini sono mortali
- II. Socrate è un uomo
- Conclusione: Socrate è mortale

Nel medioevo vengono create nuove figure di sillogismo che si distinguono per struttura e inferenza usata: modus ponendo ponens, modus tollendo tollens, modus ponendo tollens, modus tollendo ponens. Nonostante l'apparente semplicità, il sillogismo e le sue forme inferenziali sono le fondamenta di un sottoinsieme della logica del primordine, detta *logica descrittiva*, usata ad esempio in linguaggi come OWL.

Nell'epoca classica viene sviluppata un'altra struttura importantissima per l'organizzazione della conoscenza, ovvero il diagramma ad albero. Il primo viene disegnato nel III secolo da Porfirio per descrivere le categorie aristoteliche e riprendendo i concetti di *genus* e *differentia*: partendo dal *genus* -il più alto dei quali è la sostanza- si possono creare degli archi che collegano questa radice ai suoi nodi, ovvero le *differentiæ*, che ad esempio dividono la sostanza in materiale ed immateriale e successivamente queste sono collegate ad altri nodi, i *subordinate genera*, ognuno dei quali può nuovamente essere

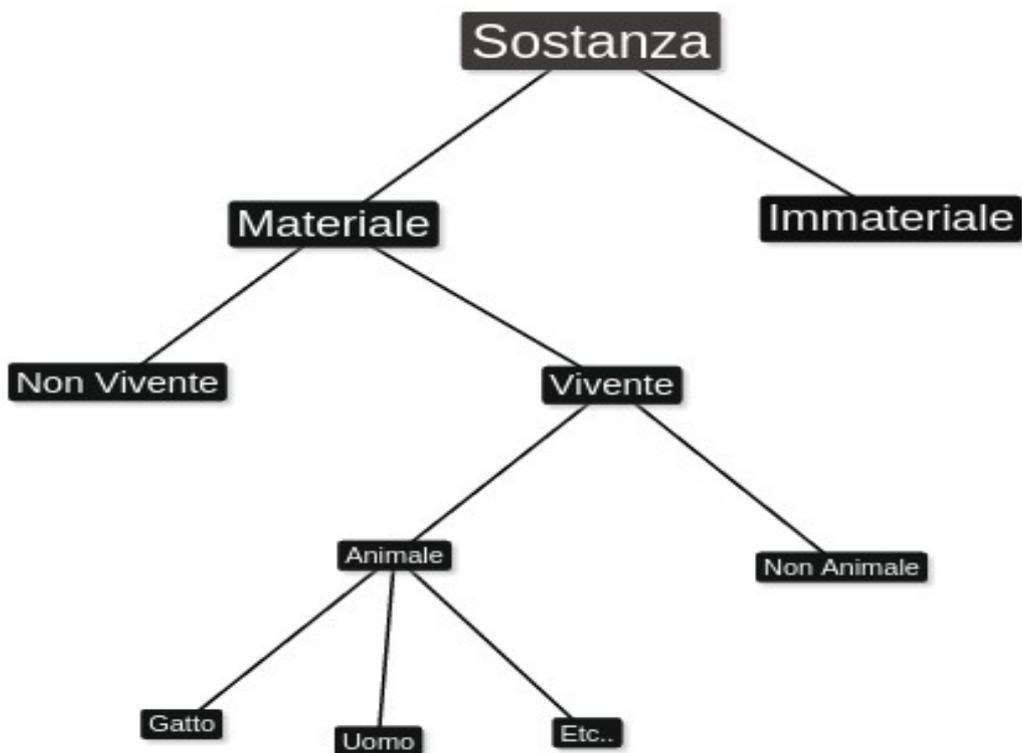


Illustrazione 1: Albero di Porfirio

divisibile in diversi sottoinsiemi secondo le differentiae. Questa struttura diventa particolarmente interessante quando vi si applica il concetto di ereditarietà che permette ai subgenera di ereditare le caratteristiche dal genus di provenienza. In figura si può vedere l'*Albero di Porfirio*, riproposto nella versione disegnata nel XIII secolo da Pietro di Spagna.

In questa direzione si mosse anche Leibniz nel XVII secolo che applicò una base matematica a questa struttura di categorizzazione, assegnando un numero primo ad ogni categoria ritenuta primitiva e considerando ogni entità come il prodotto dei numeri primi dei genera o subgenera di cui fa parte. Se ad esempio assegniamo il numero 2 alla sostanza, il numero 3 all'attributo materiale e il 5 a immateriale, otterremo che ogni multiplo di 2 e 3 descriverà una sostanza materiale, mentre un multiplo di 2 e 5 una sostanza immateriale. Il principale limite è però che si arriva velocemente a dei numeri piuttosto grandi e compiere operazioni su di essi in un'epoca senza computer non era semplice. Successivamente Leibniz provò a modificare questo sistema cercando un modo di esprimerlo con due elementi principali, un paio di numeri che indicavano positivo e negativo, ma non riuscì mai a creare un sistema ben formato.

§3.2 I moderni sulle spalle dei giganti

I filosofi spesso costruiscono le loro ontologie con l'approccio top-down, dalla definizione più generale alle istanze sottostanti, dando concetti generici di ciò che può abitare cielo o terra. Ai giorni nostri invece i programmatori sono più inclini a lavorare con un approccio bottom-up, partendo da un numero limitato di casi e iniziando a costruire generalizzazioni sempre maggiori. Per i database o i sistemi di intelligenza artificiale costruiscono inizialmente piccole ontologie o *microworld*, caratterizzati da pochi concetti e da un'applicazione particolarmente ridotta. Negli anni '80 uno degli approcci più diffusi era quello di creare *mondi a blocchi*: si caratterizzano da ontologie a blocchi e creazione di piramidi con gli stessi.

Chat-80.⁶ Un esempio di microworld è Chat-80, un sistema a forma domanda-risposta implementato in ProLog e creato da David Warren e Fernando Pereira nel 1982, che si basava su un insieme controllato di concetti relativi alla geografia. Lo scopo dei creatori era quello di trovare «techniques to be easily adaptable to a variety of applications, with as much of the implementation code as possible being application independent» e per fare ciò si sono chiesti che approccio fosse preferibile:

One way to try to articulate the meaning of a sentence is to paraphrase it into some standard, unambiguous form of English. Since these standard forms are likely to be stilted and long-winded, it will probably be convenient to represent them in a more concise notation. This is essentially the logician's approach, and we will call such meaning representations "logical forms".

Warren D., Pereira F. 1982, pp. 110-111.

La struttura dei dati di Chat-80, in figura, è ad albero; utilizza l'ereditarietà per trasmettere caratteristiche da supertipi a sottotipi; per le query usa un database dove sono riportate indicazioni geografiche; per l'analisi del linguaggio si basa su combinazioni e limitazioni di forme nome, verbo e aggettivo.

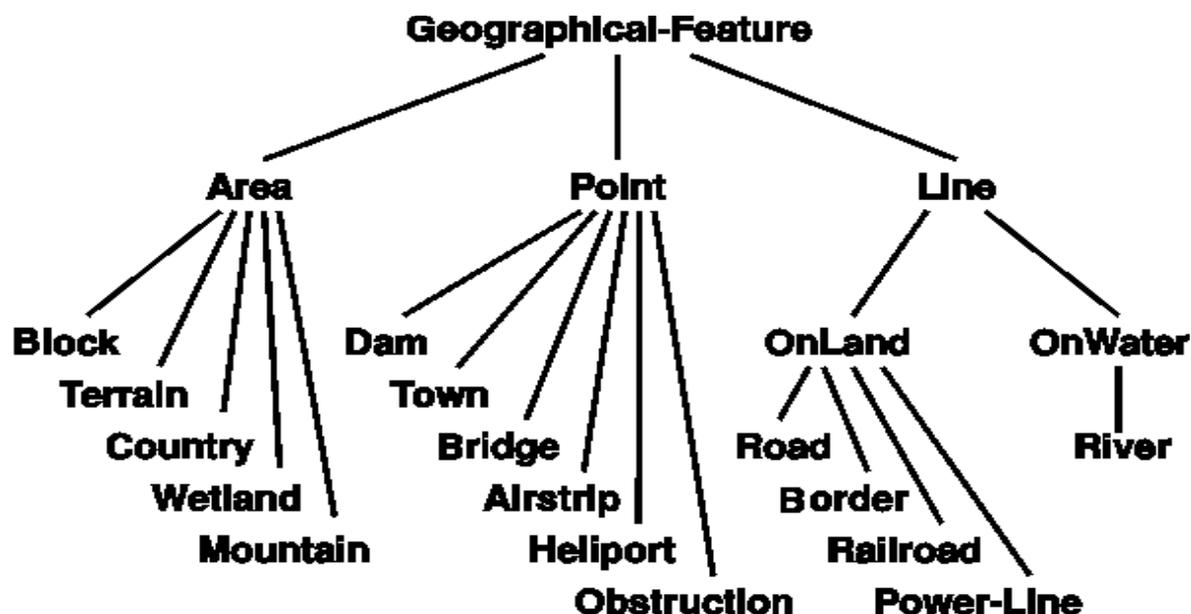


Illustrazione 2: Struttura ad albero dei dati di Chat-80, da [Sowa](#).

6 CFR WARREN D., PEREIRA F. 1982.

Uno dei limiti maggiori cui vanno incontro sistemi come Chat-80 è che, data la loro struttura e le limitazioni imposte, non è possibile fondere questi sistemi con altri che non siano esattamente dello stesso tipo, rendendo ontologie e microworld isolati rispetto ad altri esperimenti simili. Questo problema si verifica in tutti i casi in cui non è garantita l'interoperabilità di un sistema e nel corso del tempo si è evidenziata la necessità della creazione di uno standard.

Standard ANSI del 1978.⁷ Questa esigenza era chiara fin dagli anni '70 e vi rispose l'ANSI (*American National Standard Institute*) nel 1978 creando uno standard detto *conceptual schema* che fa parte della struttura a tre livelli proposta da ANSI/SPARC (*American National Standards Institute/Standards Planning and Requirements Committee*). Questo schema si propone come l'intersezione tra tre aree: il database, il programma e l'interfaccia utente. Il conceptual schema fornisce inoltre definizioni comuni delle applicazioni possibili e delle relazioni tra le diverse aree. In figura si può vedere, in maniera stilizzata, la rappresentazione di questo standard.

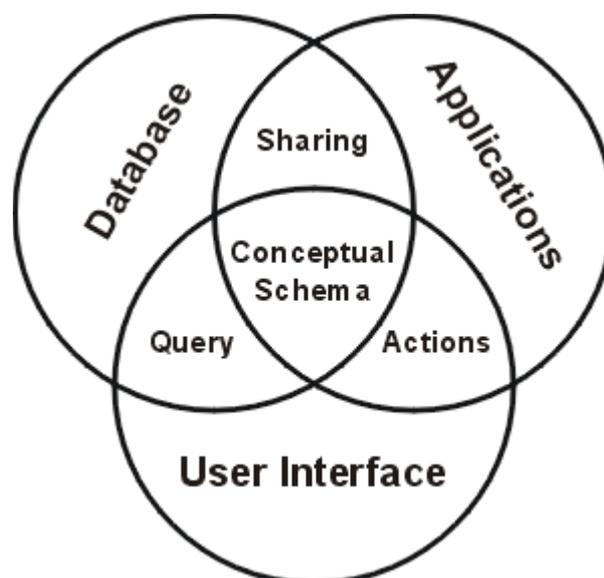


Illustrazione 3: Conceptual schema, da [Sowa](#).

7 CFR TSICHRITZIS, DIONYSIOS C., ANTHONY KLUG 1978.

Nonostante questo schema sia stato usato per molto tempo, non c'è mai stata un'implementazione totale, ma vi sono state varie implementazioni parziali che hanno portato alla fondazione di alcuni sviluppi informatici importanti, come i linguaggi di quarta generazione (fourth generation languages, detti anche 4GLs) oppure la programmazione orientata agli oggetti (Object-oriented programming system o OOPS). Nessuno di questi ha tuttavia raggiunto lo scopo ultimo e auspicabile di poter dialogare con gli altri in uno schema generale unificato. Questa aporia è sintetizzata perfettamente dalla frase del programmatore Terry Longstreth in un incontro sui database nel 1980: «Any of those tools by itself it's a tremendous aid to productivity. But any of two together will kill you».

Cyc.⁸ I tre approcci di pensiero classici che abbiamo analizzato nella sezione precedente - quindi Aristotele, Porfirio e Leibniz, sono stati ripresi nel 1995 quando è stato nato *Cyc*, un sistema che si compone di un dizionario e di un'ontologia tra le più grandi mai costruite. Il progetto è stato fondato nel 1984 da Doug Lenat e prende il proprio nome dalla sillaba della parola inglese 'encyclopedia', poiché il suo scopo iniziale era quello di riproporre tutta la conoscenza presente nella Columbia Desk Encyclopedia e di creare una struttura che potesse raccogliere al proprio interno tutto lo scibile. In figura vengono mostrati ventiquattro categorie generali, oltre alle quali *Cyc* contiene circa altri 100.000 concetti usati per la creazioni di regole e fatti che sono codificati nel sistema.

Cyc non è l'unica ontologia di grandi dimensioni che esiste - si segnalano infatti ontologie come EDR (*Electronic Dictionary Research*) o WorldNet, però è quella con gli assiomi più definiti e dettagliati. Ontologie con un elevato numero di assiomi sono anche definite *ontologie assiomatizzate*. *Cyc* risulta essere il prodotto perfetto dei sistemi di organizzazione della conoscenza classici e ne evidenzia i limiti, che sono resi tangibili da questa gigantesca ontologia:

Cyc represents the culmination of the Aristotelian approach. Its hierarchy of 600,000 concept types is the largest extension of the Tree of Porphyry ever

8 CFR LENAT D.B. 1995.

implemented, and its automated reasoning is the fulfillment of Leibniz's dream. [...]Unfortunately, "ranking things in genera and species," as Leibniz said, has not proved to be profitable. The Cyc project has survived for over twenty years, but only with the infusion of large amounts of research funds.

SOWA J.F. IN RAMADAS J., CHUNAWALA S. 2004, p. 61.

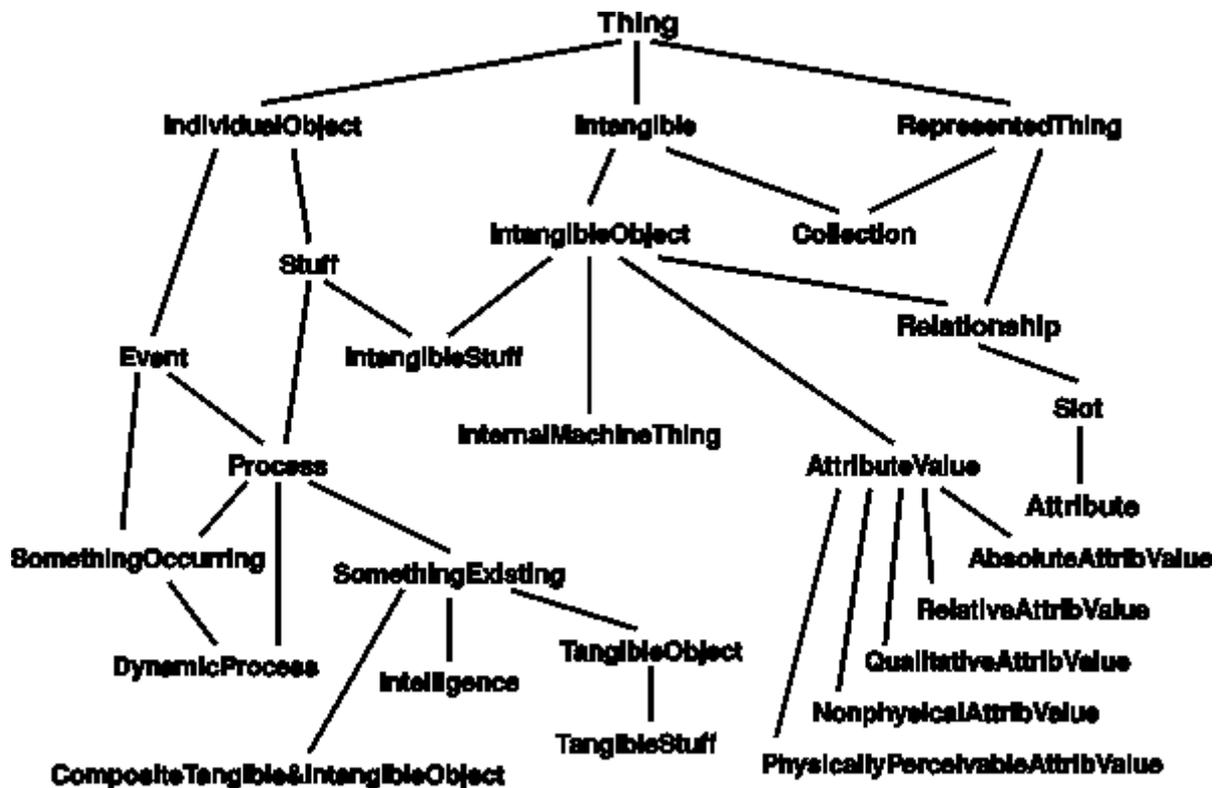


Illustrazione 4: Categorie di livello più alto di Cyc, da [Sowa](#).

VivoMind.⁹ Un'ontologia come Cyc quindi si basa su metodi classici come quello aristotelico, di tipo assiomatico-deduttivo, mostrandone tuttavia i limiti. La deduzione è molto efficace in matematica, mentre quando si tratta di un approccio empirico gli assiomi possono essere generati solo dall'induzione che è guidata dagli stessi principi decisionali dell'analogia. Riprendendo quindi il concetto di analogia e la linea di pensiero di Peirce, si possono creare sistemi a *ragionamento basato sui casi* o CBR (*Case Based Reasoning*), ovvero un processo di risoluzione di problemi che si basa su una raccolta di soluzioni di problemi precedenti. Il funzionamento a grandi linee è il seguente: si parte con una domanda o un obiettivo **Q** che riguarda il problema **P** in esame; il sistema

9 CFR SOWA J.F., MAJUMDAR A.K. IN DE MOOR A., LEX W, GANTER B. 2003.

recupera dalla memoria, tramite analogia, casi precedenti di problemi che assomiglino a **P** che vengono ordinati in base a criteri di similarità semantica rispetto al caso in esame. Il caso ritenuto più simile a **P**, e quindi semanticamente più vicino, è considerato il più rilevante e la soluzione abbinata la più probabile per rispondere a **Q** viene quindi selezionata.

Un'applicazione di CBR è *VivoMind Analogy Engine* o VAE, che è stato usato per valutare risposte aperte a word problems di algebra dopo che già due approcci avevano fallito: quello logico e quello statistico. Il funzionamento di VAE è semplice ma efficace: trasforma la domanda scritta in inglese in una rappresentazione concettuale a grafo e lo confronta con quelli generati dai problemi risolti che ha in memoria, precedentemente risolti da insegnanti di matematica; seleziona poi la soluzione con distanza semantica minore e la propone. È necessario l'intervento di un insegnante nel caso in cui non vi fossero problemi analoghi, ma dopo i primi 30 o 40 casi in cui vi è stato un aiuto dall'esterno il sistema non ne ha più avuto bisogno, dimostrandosi funzionale anche nell'imparare dai casi nuovi.

Gli approcci quindi anche alle ontologie sono molteplici ed è necessario provare nuovi strumenti ed attuare idee innovative per superare i limiti che esperimenti precedenti hanno incontrato. Dalla deduzione di Aristotele all'analogia di Peirce, questo percorso ci aiuta a capire come formare dei sistemi migliori e ci da importanti indizi sulla natura stessa del conoscere come azione nell'uomo.

CONCLUSIONE

Che cos'è l'organizzazione della conoscenza e come si applica? Quali sono le sfide e gli approcci che si incontrano in questo campo? Con questo lavoro speriamo di aver risolto questi interrogativi e di aver mostrato al lettore come problemi simili siano stati affrontati nel corso dei secoli, da pensatori come Aristotele a programmatori moderni. Abbiamo percorso insieme un sentiero che ci ha portato dalla definizione più generale di organizzazione della conoscenza ai limiti imposti dalla knowledge soup e infine ci siamo soffermati sui tentativi di superamento attuati da filosofi del passato e programmatori contemporanei.

In particolare, nella prima parte del lavoro abbiamo definito il campo dell'organizzazione della conoscenza e ne abbiamo riconosciuto i quattro livelli, che portano dalla teoria all'applicazione pratica. Abbiamo descritto brevemente vari knowledge organization system, ordinandoli in base al loro potere di descrizione semantica e quindi anche per complessità di creazione e fruizione.

Nella seconda parte ci siamo concentrati sulla nozione di knowledge soup e sulle sfide che pone all'apprendimento umano e dei computer; abbiamo visto come per gli esseri umani venga naturale organizzare e ordinare i concetti nella propria mente, differenziando l'approccio tipicamente umano da quello dei computer; nel dettaglio, abbiamo analizzato la nozione di 'zona di penombra' introdotta da Whitehead e le inferenze di induzione, abduzione ed analogia di Peirce che si affiancano alla deduzione classica.

La terza sezione, infine, è organizzata in due parti: dopo una breve definizione del termine ambiguo ontologia, abbiamo considerato gli esempi pratici dei filosofi e dei moderni. Nella prima parte della terza sezione ripercorriamo quindi parte del pensiero di Aristotele, concentrandoci su categorie, sillogismo e deduzione; poi passiamo a Porfirio, che realizza il primo diagramma ad albero; ed infine Leibniz, che pone solide basi matematiche agli alberi di Porfirio e all'ereditarietà tra radice e nodi. Nella seconda parte consideriamo approcci moderni al problema dell'organizzazione della conoscenza: dall'ontologia geografica di Chat-80, con relativi limiti di interoperabilità, alla risposta a questi ultimi dello standard ANSI del 1978, ad

una delle ontologie più grandi mai create, Cyc, e infine ad un sistema come VivoMind che sfrutta il principio dell'analogia come proposto da Peirce.

Ci pare comunque doveroso sottolineare che alcuni temi non sono stati trattati con la profondità che meritano, come ad esempio altri KOS diversi dalle ontologie oppure la formazione di sistemi gnoseologici in filosofia e in psicologia nel corso del tempo.

Bibliografia di lavoro

Bibliografia

- DE MOOR A., LEX W, GANTER B. 2003, *Conceptual Structures for Knowledge Creation and Communication, Proceedings of ICCS 2003*, pp. 16-36, Berlino: Springer-Verlag.
- GUARINO N. 1995, *Formal ontology, conceptual analysis and knowledge representation*, in 'Computer Studies', pp.624-640.
- MCILWAINE I.C., MITCHELL J.S 2008, *Preface to Special Issue "What is -Knowledge Organizazion"* in 'Knowledge Organization', n.2-3, pp. 79-81.
- PALLADINO D. 2010, *Corso di logica. Introduzione elementare al calcolo dei predicati*, Roma: Carocci editore.
- RAMADAS J., CHUNAWALA S. 2004, *Research Trends in Science, Technology and Mathematics Education: Review talks delivered at epiSTEME-1, an international conference to review research on Science, Technology and Mathematics Education*, pp. 55-91.
- TSICHRITZIS, DIONYSIOS C., ANTHONY KLUG 1978, *The ANSI/X3/SPARC DBMS framework* in 'Information Systems', n. 3, pp. 173-191.
- WARREN D., PEREIRA F. 1982, *An Efficient Easily Adaptable System for Interpreting Natural Language Queries* in 'Computational Linguistics', n. 8, pp. 110-122.

Sitografia

Tutti i siti nella sitografia si intendono alla loro versione online del 22/07/2015.

- Gnoli C. 2011, *Ontological foundations in knowledge organization: the theory of integrative levels applied in citation order*, URL: <http://www.ibersid.eu/ojs/index.php/scire/article/viewFile/4001/3639>
- GNOLI C. 2013, *Per una demarcazione ontologica dei concetti nell'organizzazione della conoscenza*, URL: <http://www-dimat.unipv.it/gnoli/demarcazione-atti.pdf>
- ISKO URL: <http://www.isko.org/>
- ISKOI URL: <http://www.iskoi.org/>
- LENAT D.B. 1995, *Cyc: A large-scale investment in knowledge infrastructure*, URL: <http://www.cyc.com/>
- SOWA J.F. (2001), *Building, Sharing and Merging an Ontology*, URL: <http://www.jfsowa.com/ontology/ontoshar.htm>
- Sowa J.F, *Guided Tour of Ontology*, URL: <http://www.jfsowa.com/ontology/guided.htm>
- WARREN D., PEREIRA F. 1982, *An Efficient Easily Adaptable System for Interpreting Natural Language Queries*, URL: http://delivery.acm.org/10.1145/980000/972944/p110-warren.pdf?ip=79.53.127.41&id=972944&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&CFID=694723041&CFTOKEN=32578400&__acm__=1437397811_7be0c53572edc35ea69b05ab68a13f84