



XML, SEMANTIC WEB E RAPPRESENTAZIONE DELLA CONOSCENZA

Massimo Parodi
Alfio Ferrara

La ricerca sugli sviluppi di XML mostra come il problema della descrizione dei dati abbia origini antiche, nel solco di tradizioni e teorie che si sono occupate di rappresentare la conoscenza. In particolare si segnalano due modelli: il modello classificatorio e il modello predicativo. Queste due forme che la discussione ha assunto fin dalla sua origine sembrano oggi riproporsi nel dibattito sul Semantic Web, ponendo, in un inedito contesto tecnico, problemi che può essere utile confrontare con i contributi della tradizione filosofica.

1. INTRODUZIONE

La diffusione e il successo di XML (*eXtensible Markup Language*), come formato di trasmissione e archiviazione delle informazioni, ha contribuito a rendere il ricorso agli strumenti informatici sempre più omogeneo anche in contesti molto diversi tra loro, nei quali pratiche e tecniche di rappresentazione della conoscenza hanno talvolta tradizioni e forme specifiche e consolidate. In particolare, il fatto che XML permetta di definire la struttura che descrive un contesto informativo all'interno del contesto stesso, precisando l'articolazione di un documento attraverso l'uso di marcatori in esso inseriti, ha prodotto due fenomeni interessanti.

In primo luogo, XML viene utilizzato spesso nella rappresentazione di documenti testuali, anche di carattere letterario o genericamente umanistico, diffondendosi quindi in un settore nel quale la rappresentazione digitale delle informazioni costituisce in larga misura un elemento di novità.

In secondo luogo, la versatilità di XML, nel generare definizioni di diverse tipologie di docu-

mento, produce la convinzione diffusa che tale linguaggio possa costituire una sorta di strumento "neutro" con il quale rappresentare la conoscenza contenuta nelle diverse fonti informative, dagli ordini di spedizione commerciali ai manoscritti latini tardo medievali. Tuttavia, il ricorso ai formalismi dell'informatica per l'archiviazione, la trasmissione e la rappresentazione dei dati che costituiscono la conoscenza può apparire neutro solo fino a che si consideri l'integrità del singolo dato la sua *identità sostanziale*, anziché la non corrottabilità delle relazioni fra i dati stessi.

Ma proprio perché tali formalismi non sono, di fatto, neutri, impongono un'organizzazione e una strutturazione alle relazioni che intercorrono tra i dati, dal momento poi che gli elementi singoli della conoscenza che si intende rappresentare vengono collocati in una struttura che dipende esclusivamente dall'applicazione e dal linguaggio utilizzato per archiviare le informazioni.

Il significato dei dati raccolti risulta quindi solo apparentemente conservato, dal momento che viene rispettata la sostanza e l'identità



di ogni dato singolarmente inteso. Ma, se si assume che il significato di un insieme di informazioni e conoscenze risieda anche, e forse soprattutto, nelle relazioni che intercorrono fra esse, allora si deve concludere che gli strumenti informativi agiscono sul significato, poiché impongono, in ogni caso, di definire le relazioni in modo congruente a quanto consentito dallo strumento stesso.

Ogni strumento informatico presuppone infatti la definizione di tipologie di relazioni possibili e delle regole che le definiscono; tali regole sono necessariamente sovrapposte ai dati che si desidera trattare attraverso lo strumento. Senza un'adeguata consapevolezza delle regole proprie di ogni strumento di questo genere, e senza l'indispensabile senso critico, si corre il rischio di assumere regole e modelli di descrizione e archiviazione non adeguati ai dati da rappresentare, alterandone i rapporti interni e, di conseguenza, il significato.

La ricerca che in questi ultimi anni si è sviluppata a partire da, e intorno a, XML è particolarmente significativa proprio per comprendere come ogni logica di descrizione dei dati non sia neutra né, tantomeno, istintiva, ma abbia, talvolta, origini antiche, nel solco di precise tradizioni e teorie che si sono occupate appunto di rappresentare la conoscenza e per mettere quindi in evidenza come talune difficoltà connesse all'uso di XML non siano affatto di natura tecnica, ma nascano da tale logica e, in ultima analisi, da quelle teorie e da quelle tradizioni.

2. STRUTTURA E SIGNIFICATO

L'elemento di forza dell'XML e, al tempo stesso, carattere decisivo della sua duttilità, è la possibilità di definire, attraverso il sistema dei marcatori incorporati nel documento, i linguaggi con cui il documento potrà successivamente essere letto e rappresentato, isolando in tal modo l'aspetto strutturale, descritto attraverso un DTD (*Document Type Definition*), da altri due aspetti fondamentali: quello della formattazione, dipendente da strumenti esterni che analizzano in modo critico il documento e, soprattutto quello semantico, cioè il significato dei marcatori.

La *definizione del tipo di documento* descrive

accuratamente i marcatori che individuano le porzioni di testo, ma prescinde del tutto da cosa essi possano significare. L'aspetto strutturale dà origine, invece, a una sorta di gerarchia di marcatori *generici*, più ampi, al di sotto dei quali si collocano altri marcatori, per così dire *speciali*, che possono, talora, fungere da generi per ulteriori specificazioni. Chi abbia un minimo di consuetudine con la storia della logica occidentale non può fare a meno di riconoscerci una specie di *albero di Porfirio*. Si tratta di una struttura che venne

proposta per interpretare la dottrina aristotelica delle *categorie*, secondo la quale all'interno di ogni tipo di predicato (le categorie appunto di sostanza, relazione, qualità, quantità, luogo, tempo ecc.) si può immaginare un collegamento a cascata di modi di predicazione, adatti a soggetti sempre meno numerosi, a partire dal genere sommo ovvero il nome della categoria (*sostanza*,

per esempio), che viene specificato attraverso differenziazioni e specie successive (sostanza vivente o non vivente; sostanza vivente dotata o non dotata di anima vegetativa, sensitiva, razionale ecc), fino a giungere alla più determinata possibile (*uomo*, per esempio) al di sotto della quale non esistono più ulteriori specie, ma solo individui.

E non a caso le principali restrizioni sintattiche imposte dalle descrizioni rappresentate dai DTD consistono quindi nella necessità che l'oggetto descritto sia un *albero* e che non si verifichi, entro tale albero, alcuna sovrapposizione dei marcatori. Il sistema non deve, inoltre, ammettere che entro una porzione di testo marcata in un certo modo si apra un ulteriore marcatore destinato a chiudersi oltre la chiusura del primo. Sarebbe come prendere in considerazione l'eventualità che, entro il *genere animale razionale*, anziché le specie individuate dalla differenza *mortale* e *non*

Porfirio (233 – 305), filosofo greco, noto soprattutto per aver raccolto gli scritti di Plotino in sei *Enneadi* e per la *Vita di Plotino*. La sua *Isagoghe* influenzò il pensiero neoplatonico e la cultura medievale nella quale propose, attraverso i commenti di Mario Vittorino e di Boezio, l'interesse per la logica aristotelica e in particolare per il problema degli *universali*.

mortale – per usare uno degli esempi comuni a tutta la tradizione aristotelica – si possano collocare le specie *alto* e *biondo*: gli scandinavi appartenerebbero ad entrambe le specie e non si saprebbe più dove mettere la divinità (razionale e non mortale).

La programmatica indipendenza del punto di vista strutturale ha portato a pensare alla descrizione DTD come del tutto indipendente da ogni implicazione di contenuto, per cui l'organizzazione dell'albero non ha a che vedere di fatto con entità come *animale*, *razionalità*, *scandinavi alti* e *biondi*, ma solo con marcatori come *animale*, *razionale*, *scandinavo* i cui rapporti, esclusivamente formali, sono semplicemente quelli descritti nel documento. Avviene dunque che un documento XML ben formato descriva perfettamente l'albero di Porfirio, ma rappresenti anche la struttura informativa della realtà descritta, priva di connotazioni semantiche. In Porfirio, invece, alla cui interpretazione di Aristotele dobbiamo una delle prime manifestazioni della rappresentazione ad albero della conoscenza, la struttura della classificazione esprime le differenze proprie della realtà, il suo aspetto semantico. Questi problemi, che hanno interessato larga parte del dibattito filosofico tardo antico e medievale, sembrano riproporsi oggi in un contesto del tutto diverso con l'idea, sviluppatasi con forza negli ultimi tempi, del **Semantic Web**: l'ipotesi cioè di arrivare a una sorta di formalizzazione – la

I tre livelli del **Semantic Web**:

Livello di descrizione della semantica

- OIL
- DAML+OIL

Livello di descrizione dello schema

- RDF Schema

Livello di descrizione del modello dei dati e dei metadati

- RDF

più generale possibile – di aspetti semantici, di contenuto, di molteplicità di documenti e soprattutto di tipologie di documento reperibili in rete.

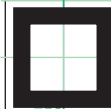
Con una mossa teorica, che ricorda altri passaggi della storia della filosofia occidentale, si sono venuti sviluppando metodi (e linguaggi) che intendono descrivere la semantica dei documenti, introducendo un impianto descrittivo definito su tre livelli [2]: un linguaggio relazionale del tipo

“soggetto – predicato – oggetto”, per descrivere metadati e modelli dei dati, l'RDF (*Resource Description Framework*); un linguaggio e una sintassi che insieme definiscano e

Ontologia è un termine carico di valenze teoretiche nella tradizione del pensiero occidentale e, per avere un punto di riferimento sintetico ma attendibile, è preferibile sfogliare le pagine del *Dizionario di filosofia* di Nicola Abbagnano: “... La seconda concezione fondamentale [di *metafisica*] è quella della metafisica come ontologia o dottrina che studia i caratteri fondamentali dell'essere: quei caratteri che ogni essere ha e non può non avere” [1]. E infatti la costruzione di una ontologia non è altro che la definizione dei caratteri fondamentali di quanto può venire accettato in un determinato mondo possibile, sottoinsieme, nel nostro caso, dei possibili mondi di significato del web semantico. Più sotto, Abbagnano nota che in questo tipo di metafisica sono normalmente implicate: “...a) una determinata teoria dell'essenza, e precisamente quella dell'essenza necessaria; b) una determinata teoria dell'essere predicativo e precisamente quella dell'inerenza; c) una determinata teoria dell'essere esistenziale e precisamente quella della *necessità*” [1]

descrivano il vocabolario delle rappresentazioni desiderate, ossia un'estensione dell'RDF pensata per la rappresentazione di strutture più generali, di carattere classificatorio (classi e sottoclassi, per esempio). Queste strutture prendono il nome di schemi e il linguaggio è denominato RDF Schema. Infine, un livello nel quale viene definita formalmente la semantica e gli strumenti di supporto per l'interpretazione automatica. A questo livello, che riguarda la descrizione della realtà di riferimento, si collocano proposte di formalismi quali OIL (*Ontology Interchange Language*) e DAML (*Darpa Agent Mark-Up Language*)+OIL [3].

Ma soprattutto, è a questo punto che compare un termine che in ambito informatico assume un significato quasi tecnico e che suscita invece una grande attenzione da parte di chi si accosta a questi temi da una prospettiva filosofica. La definizione infatti dei possibili oggetti di un mondo in base ai quali articolare la descrizione semantica di una realtà di interesse viene definita **ontologia**. Non è difficile da parte nostra osservare che un'ontologia, anche nel senso informatico, ha la pretesa di determinare con precisione quali predicati definiscano un soggetto, ne precisino cioè l'essenza, quali predicati siano possibili di un soggetto e, infine, quali oggetti necessariamente esistano nel mondo possibile descritto.



Di fronte dunque a una molteplicità di documenti e di tipologie di documento, l'ontologia svolge una funzione unificatrice così come "... ogni scienza è, come tale, studio della sostanza in qualcuna delle sue determinazioni, per esempio, della sostanza in movimento la fisica, della sostanza come quantità la matematica; la metafisica [come ontologia] è la teoria della sostanza in quanto tale" [1]. Ponendosi per una volta, finalmente, nel ruolo di *principio* (Dio, l'Uno, il Demiurgo, o come altro di preferisca chiamarlo), l'uomo definisce quale sia la sostanza in quanto tale nel mondo che vuole unificare e interrogare, costruendo in tal modo quella descrizione semantica, interrogando la quale si potranno al tempo stesso interrogare tutte le fonti da essa descritte.

Ma allora viene da chiedersi se, trattandosi di metafisica, il mondo che si va costruendo assomigli all'albero, di Porfirio e dell'XML, o al reticolo di relazioni fra classi e proprietà descritte in un'ontologia. La risposta sembra ovviamente la seconda, in quanto, come si è più volte ricordato, l'XML pretende di non avere implicazioni semantiche al di fuori dei termini usati per i marcatori. La questione però sta proprio qui e riguarda essenzialmente la possibilità che la struttura ad albero che definisce quella di un documento possa davvero non implicare alcun aspetto semantico.

Gli studi sull'albero di Porfirio hanno mostrato chiaramente che pur pretendendo di essere un *dizionario*, per usare la terminologia con cui Umberto Eco [6] indica un insieme di definizioni che dovrebbero in linea di principio sorreggersi a vicenda senza necessità di alcun ricorso all'esperienza esterna al dizionario, esso si rivela in realtà un'*enciclopedia* e cioè un insieme di definizioni che ricorre all'esperienza per definire un mondo che in qualche modo si assume come esistente all'esterno dell'enciclopedia stessa. Nel caso dell'albero di Porfirio il problema nasce dalla pretesa di arrivare a definire l'essenza delle cose, basando quindi la distinzione che separa le specie all'interno dei generi su caratteri che siano appunto costitutivi dei caratteri essenziali delle specie individuate. Risulta dunque evidente che tutto il peso teorico della struttura poggia proprio sulla scelta

delle differenze che devono risultare non puramente accidentali ma per l'appunto essenziali e il fatto che possano esistere, all'interno di un genere, differenze di questo tipo è in realtà solo suggerito dalle differenze che vengono osservate nella realtà empirica, rispetto alle quali le differenze essenziali risultano sintomi, indizi. Ma se le cose stanno così, di fatto si ammette che la costruzione dell'albero poggia su dati empirici e non è la struttura puramente formale di essenze subordinate le une alle altre. Dovrebbe essere, dunque, il modello metafisico della realtà, o almeno della nostra conoscenza, e si rivela invece una imitazione della realtà o del modo in cui la conosciamo¹.

Un secondo problema sta nella pretesa assenza di sovrapposizione fra i marcatori, ovvero nella univocità dei nodi dell'albero: nella struttura dell'albero porfiriano non si può escludere che le medesime differenze compaiano in posizioni diverse, al di sotto di nodi diversi dell'albero. Ci si può chiedere ad esempio se solo le sostanze viventi sensitive vegetative e razionali siano divisibili in mortali e non mortali o se invece non potrebbe risultare più opportuno dividere le sostanze viventi sensitive vegetative mortali in razionali e non razionali, oppure ancora se la stessa coppia di differenze non debba necessariamente comparire in punti diversi della struttura, creando esattamente il fenomeno della sovrapposizione tra definizioni diverse, per cui una parte di una specie cadrebbe all'interno di un'altra, anche se non totalmente.

È naturalmente vero che una struttura descritta dai marcatori XML non ha le stesse pretese della struttura porfiriana, non mira cioè a produrre una definizione, una somma di differenze che coincidano esattamente con l'oggetto che si intende descrivere e tuttavia il richiamo alla centralità delle differenze ci riporta a considerare che l'articolazione

¹ "Il dizionario ... si dissolve necessariamente, per forza interna, in una galassia potenzialmente disordinata e illimitata di elementi di conoscenza del mondo. Quindi diventa un'enciclopedia e lo diventa perché di fatto era un'enciclopedia che s'ignorava ovvero un artificio escogitato per mascherare l'inevitabilità dell'enciclopedia" [6].

dei marcatori non sembra poter essere del tutto priva di portata semantica. Si può pretendere che i loro nomi siano puramente convenzionali e non portino con sé alcun significato, ma è la scelta delle relazioni di subordinazione a reintrodurre il dubbio e, soprattutto, la loro possibile comparsa in porzioni diverse del documento da descrivere².

La questione che nasce è una questione di fondo: sembra impossibile isolare completamente l'aspetto strutturale da quello semantico o, forse è meglio dire, l'aspetto strutturale porta inevitabilmente con sé delle componenti semantiche, perché a queste si fa inevitabilmente riferimento per risolvere i dubbi sui tipi di subordinazione, sulle ripetizioni dei marcatori, sul modo in cui evitare le sovrapposizioni. Dunque, quando si pretende che i DTD definiscano esclusivamente tipi di documento, lasciando ancora non risolte le questioni semantiche, affidate solo al livello della ontologia, si pretende qualcosa che non è compatibile con la struttura descritta. Tutti i rapporti tra i vari livelli su cui si articola la gestione dei documenti XML vengono quindi in qualche misura sconvolti da implicazioni che la tradizione occidentale della rappresentazione della conoscenza ci aiuta a individuare e ci ricorda di tenere sempre presenti.

3. DALLA STRUTTURA ALL'ONTOLOGIA

Se dunque da un lato scopriamo che la struttura di un documento XML non può essere del tutto priva di portata semantica e che l'aspetto semantico e quello strutturale non possono essere chiaramente distinti se non sulla base di una convenzione, dall'altro verificiamo il sorgere dell'esigenza di arricchire la rappresentazione XML con costrutti capaci di descrivere le conoscenze in nostro possesso in termini pienamente semantici.

A questo proposito, può risultare interessante considerare come alcune delle soluzioni proposte per il Semantic Web ripercorrono

un dibattito che ha attraversato ampiamente la storia della filosofia occidentale. La diffusione dell'XML ha delineato uno scenario, ancora non attuale ma possibile, nel quale il web diviene un immenso serbatoio di informazioni e documenti che condividono lo stesso modello di rappresentazione. In questo ampio insieme di alberi di Porfirio si presentano da subito due problemi.

Il primo è il fatto che la struttura dell'albero è contemporaneamente portatrice di un valore semantico, per le ragioni appena esposte, ma al tempo stesso non è sufficiente per determinare univocamente i significati associati a ogni documento. Come nel caso di Porfirio, il valore semantico della struttura è determinato dalla capacità classificatoria dell'albero, ma quest'ultimo non è sufficiente ad esprimere senza ambiguità l'insieme di relazioni che ogni nodo può avere con gli altri. In altri termini, non associa al documento alcuna regola di interpretazione.

L'interpretazione semantica rimane così un'attività lasciata al fruitore del documento. L'idea del Semantic Web è invece propriamente quella di aggiungere ad ogni contenuto informativo una descrizione e delle regole di interpretazione univoche, comprensibili e processabili da una macchina. Il problema dell'interpretazione e dell'univocità fu ciò che spinse il pensiero occidentale a discostarsi progressivamente dal modello porfiriano della rappresentazione, esattamente come le stesse motivazioni portano oggi la ricerca sul Semantic Web a definire linguaggi e modelli di rappresentazione basati su logiche diverse da quelle dell'albero XML, che diviene sempre di più un modello esclusivamente sintattico.

Il secondo problema riguarda la necessità di strumenti di integrazione fra le diverse fonti informative. Questo problema esula dalle finalità di questo intervento, ma è utile ricordarlo dal momento che esso mette in luce due aspetti interessanti. Il primo è che le ontologie vengono utilizzate a questo scopo propriamente in virtù dell'istanza di unitarietà che portano con sé. L'accezione informatica e quella filosofica del termine ontologia non sono, sotto questo profilo, molto distanti. Il secondo, consiste nel fatto che nel momento in cui si rende necessaria la rap-

² "...ciò che costituisce la vera differenza non è né l'uno accidente né l'altro, è il modo in cui li raggruppiamo riorganizzando l'albero" [6].



presentazione di una realtà di interesse e non di un dato, di un mondo e non di un individuo, in informatica come in filosofia, viene meno il modello gerarchico e classificatorio dell'albero, in favore di un modello relazionale e predicativo. Potremmo dire, forse, che le specificazioni che possiamo dare di una sostanza, nella definizione di un albero e di una classificazione, non sono sufficienti a descrivere un mondo nel quale si diano più alberi e più classificazioni e che la realtà non può essere descritta nei termini del concetto di sostanza, quanto piuttosto in quelli del concetto di relazione.

3.1. Un esempio

A prima vista dunque un insieme di conoscenze sembrerebbe rappresentabile attraverso una gerarchia di concetti e un insieme di proprietà e attributi di tali concetti. Ma ad un esame più attento questo modello, il modello ad albero, si rivela insufficiente allo scopo, non perché sia privo di contenuto semantico, ma piuttosto perché di tale semantica non viene definita alcuna regola di interpretazione, perché il modello appare insufficiente di fronte all'insieme delle possibili relazioni fra i concetti. La struttura di un documento XML, come l'albero di Porfirio, esempi entrambi di alberi, non bastano a risolvere la ricchezza delle possibili relazioni: in entrambi i casi necessitiamo di linguaggi e modelli di rappresentazione più ricchi. Si cercherà attraverso un esempio di mostrare come lo stesso patrimonio informativo possa venire rappresentato in modo diverso in un documento XML e in un'ontologia, costituendo così un insieme diverso di conoscenze. Si vedrà anche come i limiti della rappresentazione XML siano gli stessi limiti del modello porfiriano.

Si immagini dunque di dover descrivere una realtà nella quale vi siano *animali* e *piante*. Delle piante facciano parte gli *alberi*, composti da *rami*, composti a loro volta da *foglie*. Gli animali si dividono in *carnivori* ed *erbivori*: i primi mangiano animali, i secondi piante [5]. Un possibile albero che rappresenti questa situazione potrebbe essere quello riportato nella figura 1.

Il modello è naturalmente molto semplice, ma si presta da subito ad alcune considera-

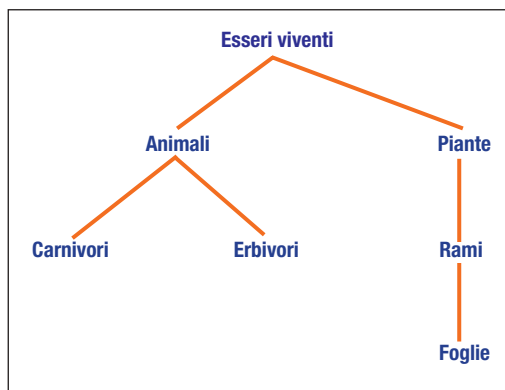


FIGURA 1

Esempio di modello di classificazione

zioni. Dal punto di vista strutturale non vi è alcuna differenza fra la relazione che intercorre, ad esempio, fra *Animali* e *Carnivori* e quella che intercorre fra *Piante* e *Rami*: in entrambi i casi si è di fronte a due relazioni di subordinazione. Il fatto che *Rami* sia l'unico sottonodo di *Piante* è irrilevante (è sempre possibile per completezza aggiungere il nodo *NonRami* come sottonodo di *Piante*).

Tuttavia ci si potrebbe chiedere se, dal punto di vista semantico, il concetto di ramo stia a quello di pianta come quello di carnivoro a quello di animale.

Ora si supponga che si voglia intendere che il ramo è una parte della pianta, mentre i carnivori sono un sottoinsieme degli animali. Fra il concetto di parte e quello di sottoinsieme vi è una chiara differenza (un carnivoro è un animale, un ramo non è una pianta) ma l'albero non offre alcuna possibilità di segnalare questa differenza. In altri termini, a partire dalla rappresentazione data non è possibile definire alcuna regola che dica in che modo si debba interpretare le relazioni che intercorrono fra nodi e sottonodi. Il fatto che all'albero non sia associata alcuna euristica non significa però affermare che tale rappresentazione sia totalmente priva di valore semantico. Non sarebbe infatti possibile sostenere, per esempio, che un erbivoro è una pianta.

La struttura dell'albero, nel caso di Porfirio come di un documento XML, pone dunque dei vincoli all'interpretazione semantica, non rappresenta in modo neutro le informazioni fornite, ma, al tempo stesso, non definisce nemmeno regole di interpretazione univoche per le relazioni descritte. Una seconda osservazione riguarda il fatto che

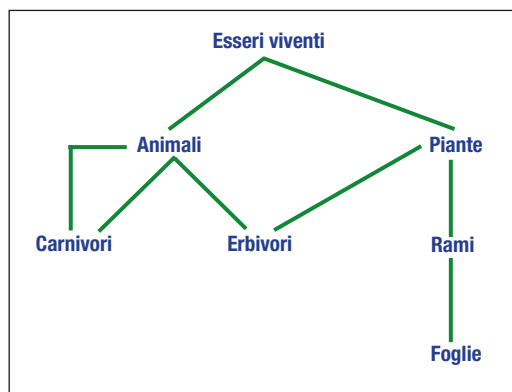


FIGURA 2
Inserimento
delle relazioni
Carnivori/Animali
ed Erbivori/Piante

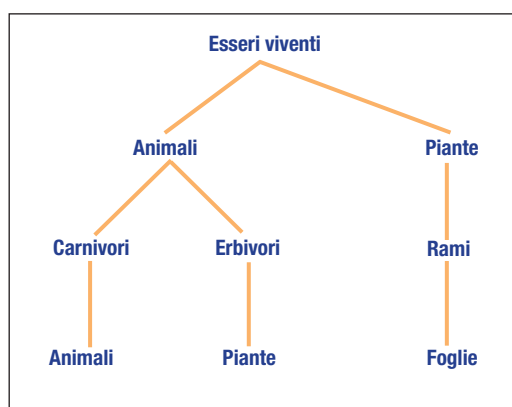


FIGURA 3
Ripetizione dei nodi
Animali e Piante

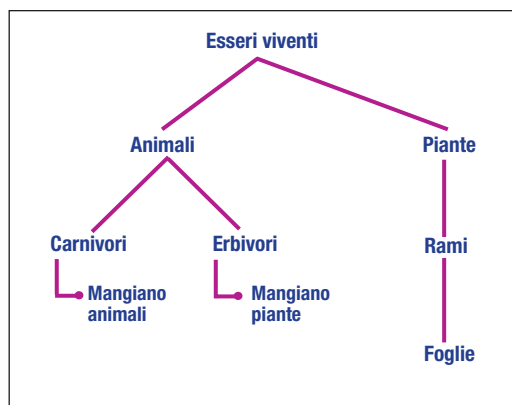


FIGURA 4
Le caratteristiche
di Carnivori ed
Erbivori come
attributi

nell'albero proposto non vi è traccia del fatto che i carnivori siano tali in quanto mangiano animali, mentre gli erbivori mangiano piante.

Per esprimere queste caratteristiche delle due specie di animali si hanno tre soluzioni possibili: la prima consiste nel definire una relazione fra il nodo Carnivori e il nodo Animali e una fra il nodo Erbivori e il nodo Piante, come mostrato dalla figura 2. Questa soluzione, però, non è ammessa né dall'albero porfiriano né da XML, perché dissolve la struttura ad albero, dal momento che

Animali è sia radice che nodo di Carnivori e che l'elemento Erbivori ha due radici.

Una seconda soluzione consiste nel definire come sottonodo di Carnivori un nuovo elemento Animali e come sottonodo di Erbivori un nuovo elemento Piante, come mostrato in figura 3.

Questa soluzione non è ammessa dall'albero di Porfirio, in cui i nomi dei nodi indicano generi e specie univoci, mentre è ammessa, sul piano sintattico, da XML. Il problema è però se tale soluzione esprima correttamente le definizioni proposte di Carnivori ed Erbivori. Se si assume che le definizioni si basino sull'identità e univocità dei termini utilizzati, allora affermare nel contesto iniziale che gli Erbivori mangiano Piante, significa anche affermare che gli Erbivori mangiano oggetti non animali, il che non è però univocamente deducibile dalla rappresentazione di figura 3.

Ne segue che se si rappresenta una definizione in cui vale l'univocità dei termini utilizzati questa soluzione non rappresenta la definizione proposta ed è inadeguata indipendentemente dalla sua ammissibilità sintattica. La terza soluzione consiste nel non considerare le caratteristiche di Carnivori e Erbivori come loro proprietà, ovvero come specifiche e caratteristiche relazioni fra essi ed altri elementi, ma considerarle piuttosto come loro attributi, ovvero come specificazioni degli elementi Carnivori ed Erbivori in quanto tali, indipendentemente dal contesto in cui sono collocati. Avendo tralasciato gli aspetti sintattici, un'illustrazione intuitiva di questa soluzione può essere quella di figura 4.

Ciò che interessa in questo caso è il fatto che piante e animali diventano i valori di una caratteristica propria degli elementi Carnivori ed Erbivori. In questo modo viene salvaguardata la struttura ad albero, ma quella che nell'albero porfiriano si sarebbe definita differenza specifica, smette di essere una relazione, in questo caso fra mangiante e mangiato, per divenire parte integrante dell'elemento definito.

Si tratta del trionfo di un'impostazione che pone al centro la sostanza a scapito della relazione. In questa soluzione un erbivoro non è un individuo che ha una relazione di



appartenenza con la classe Animali e una relazione “mangia” con la classe Piante, bensì un individuo che appartiene alla classe Animali e che ha come caratteristica propria il mangiare piante. Per piante non si intende più una classe di oggetti della realtà, ma semplicemente un valore distintivo per quell’attributo, più o meno accidentale. Questa soluzione è la più corretta al fine di salvaguardare la struttura dell’albero ma valgono per essa le osservazioni fatte per la seconda soluzione, dal momento che perdendo quella che porfirianamente si potrebbe definire differenza specifica come relazione e descrivendola solo come attributo, si perdono per esempio delle informazioni sull’elemento Piante, che non è più oggetto di alcun appetito.

Oltre alle varie osservazioni già fatte, rimangono due fondamentali ragioni per le quali l’esigenza di avere una adeguata rappresentazione della conoscenza in nostro possesso impone di abbandonare, o almeno di arricchire, il modello ad albero, sia che si tratti di XML che dell’albero porfiriano.

La prima è rappresentata da questo interrogativo e si tratta di un problema già caratteristico dell’albero porfiriano: si immagini di dover inserire nella descrizione una pianta carnivora.

Il problema è simile a quello affrontato in precedenza e sono possibili diverse soluzioni, ma la domanda, dal punto di vista ontologico, è: animali e piante sono differenti specie del genere *carnivoro*, oppure carnivoro è una specie dei generi animali e piante? Tentare di rispondere a questa domanda nei termini dell’albero di Porfirio è come tentare di definire una gerarchia fra i vertici di un triangolo equilatero. Questo perché lo strumento adeguato ad affrontare il problema è il concetto di relazione, ammettendo che vi possano essere relazioni non gerarchiche.

La seconda ragione è che nessuna delle soluzioni proposte offriva la possibilità di definire un’interpretazione delle relazioni rappresentate, mentre occorre uno strumento che permetta di definire non soltanto relazioni di tipo non gerarchico, ma anche di associare a queste un’interpretazione, distinguendo per esempio la relazione “esse-

re parte di” dalla relazione “essere sottinsieme di”.

4. IL MODELLO PREDICATIVO

La caratteristica dunque che più di ogni altra costituisce un limite alle capacità rappresentative di XML, ma soprattutto dell’intera tradizione del modello ad albero, è l’assenza di relazioni interpretabili. Nello sforzo di definire un modello per la rappresentazione della conoscenza, che necessita come si è visto di essere un modello semanticamente ricco, non è dunque rilevante solo la centralità della relazione sulla sostanza, ma anche, e forse soprattutto, la questione cruciale dei tipi di relazioni, e la possibilità stessa di distinguere e definire tipologie diverse di relazione, sia dal punto di vista del significato che ad esse si associa, sia dal punto di vista delle proprietà di tali relazioni e della definizione di un dominio e di un *range* di applicabilità.

Per queste ragioni la riflessione su questi temi, ha lentamente superato il modello porfiriano per generare un modello predicativo e logico che potesse essere sia criterio di rappresentazione della conoscenza che supporto a quello che si definisce ragionamento, ovvero alla capacità di costruire su tale rappresentazione un insieme di possibili deduzioni che ne verificano la coerenza. Un lontano tentativo in questa direzione, ancora in un contesto del tutto dissimile da quello logico-predicativo, ma nel quale è evidente uno sforzo deduttivo è fornito dalla teoria della *resolutio-compositio* di Roberto Grossatesta, che nel XIII secolo propone un approccio per alcuni versi straordinariamente simile a quello delle moderne ontologie: Grossatesta descrisse un duplice procedimento che chiamò *resolutio-compositio*.

Questi termini erano derivati dagli studiosi greci di geometria e da Galeno, ed erano, ovviamente, la traduzione latina delle parole greche *analisi* e *sintesi*. Grossatesta in sostanza derivò da Aristotele il principio centrale del proprio metodo, ma lo sviluppò con maggior completezza. Il metodo seguiva un ordine ben definito. Per mezzo del primo procedimento, *la resolutio*, egli mostra-

va come ricavare e classificare, per somiglianze e differenze, i principi componenti o elementi costituenti il fenomeno.

Questo gli dette ciò che chiamò la definizione nominale.

Cominciava col raccogliere esempi del fenomeno in esame e col notare tutti gli attributi che avevano in comune, finché arrivava alla “formula comune” che enunciava la relazione empirica osservata; quando gli attributi venivano trovati di frequente associati si poteva sospettare una connessione causale.

In seguito, mediante l’opposto procedimento della *compositio*, risistemando le proposizioni in modo che le più particolari risultassero derivate deduttivamente da quelle più generali, dimostrava che il rapporto tra generale e particolare era un rapporto di causa ed effetto: metteva cioè le proposizioni in ordine causale. Illustrò il suo metodo mostrando come arrivare al principio comune in base al quale certi animali hanno le corna; nel capitolo 4 del libro III del suo commento agli Analitici Secondi egli disse che la presenza di corna “dipende dalla mancanza di denti nella mascella superiore in quegli animali ai quali la natura non fornisce altri mezzi di sopravvivenza che le corna”, analogamente a quanto fa per il cervo con la sua velocità e con il cammello con la sua mole.

Negli animali provvisti di corna, la materia che doveva servire a formare i denti andava invece a formare le corna. Aggiungeva Grossatesta: “Il non avere i denti nelle due mascelle è anche la ragione per cui quegli animali hanno più di uno stomaco”, correlazione che egli faceva risalire alla deficiente masticazione negli animali provvisti di una sola fila di denti” [4]. Quella istanza deduttiva che in Grossatesta è accennata diverrà centrale secoli dopo con lo sviluppo delle logiche predicative, anche come strumento di rappresentazione della conoscenza. Questo spostamento è quanto avviene anche nel Semantic Web che, prima nel trattare i metadati, con l’RDF, poi nell’ultimo livello, quello ontologico, si avvale esattamente di linguaggi e formalismi di natura logico-predicativa.

Uno di questi linguaggi, da cui si trarrà il

successivo esempio, si chiama *Ontology Interchange Language (OIL)* [7].

Senza entrare nei dettagli di tale linguaggio, ciò che ci interessa qui da vicino è il modo in cui viene risolta la rappresentazione delle relazioni e quale modello rappresentativo ha origine da essa. OIL consiste in definizioni di classi (*class-def*) attraverso le quali rappresentare i concetti della realtà da descrivere, e da definizioni di slot (*slot-def*) i quali risolvono appunto l’esigenza di rappresentare relazioni binarie fra classi, e vincoli (*slot-constraint*) su esse, mentre nel modello ad albero non vi è alcun costrutto per rappresentare le relazioni in se stesse. Per esemplificare si può osservare come sia descritta in OIL la situazione esaminata precedentemente:

```
class-def animale
class-def pianta
  subclass-of NOT animale
class-def ramo
  slot-constraint parte-di
  has-value pianta
class-def foglia
  slot-constraint parte-di
  has-value ramo
class-def carnivori
  subclass-of animale
  slot-constraint mangia
  value-type animale
class-def erbivori
  subclass-of animale, NOT carnivori
  slot-constraint mangia
  value-type pianta
OR
(slot-constraint parte-di
  has-value pianta)
```

Appare evidente il cambiamento: le relazioni, pur non rinunciando a definire relazioni gerarchiche come nel caso delle sottoclassi, sono varie e diverse fra loro. Si possono distinguere l’essere sottoinsieme dall’essere parte di una classe precedente e vi sono vincoli sulle relazioni che ne determinano anche le regole di applicabilità. In altri termini, alle diverse relazioni è associata un’interpretazione univoca. Questo modifica totalmente il modello porfiriano ed aumenta le capacità espressive della rappresentazio-

ne. Dovendo rappresentare graficamente questo modello non si avrà più un albero, ma piuttosto una rete di relazioni, come mostrato in figura 5.

5. CONCLUSIONI

In questo intervento si è tentato di mostrare, per quanto sommariamente, i punti di contatto e i rapporti fra lo sviluppo della discussione sui modelli di rappresentazione della conoscenza in momenti della storia della filosofia anche molto lontani dai nostri giorni e il dibattito contemporaneo che coinvolge XML e più in generale il Web.

Tale parallelismo mirava soprattutto a mettere in luce il fatto che frequentemente l'informatica ripropone, amplifica e rimette in discussione temi sui quali insistono lunghe e rilevanti tradizioni di pensiero. Per questa ragione occorre, a parere di chi scrive, sottolineare come la sola dimensione tecnica della discussione informatica non sia sufficiente, soprattutto a proposito della rappresentazione della conoscenza, per esaurire e comprendere appieno il senso del dibattito in corso e delle soluzioni proposte.

Le difficoltà nell'utilizzare la semplice codifica XML come strumento di rappresentazione della conoscenza, ad esempio, non nascono da problemi di carattere tecnico o sintattico, ma piuttosto da limiti propri del modello concettuale utilizzato.

Non è però solo l'informatica che può trarre spunto da alcune delle riflessioni proposte dalla tradizione filosofica, ma è anche quest'ultima, che spesso si è trovata alle prese con questi problemi, a poter trovare materia di stimolo e terreno di applicazione in alcuni dei settori di ricerca delle discipline infor-

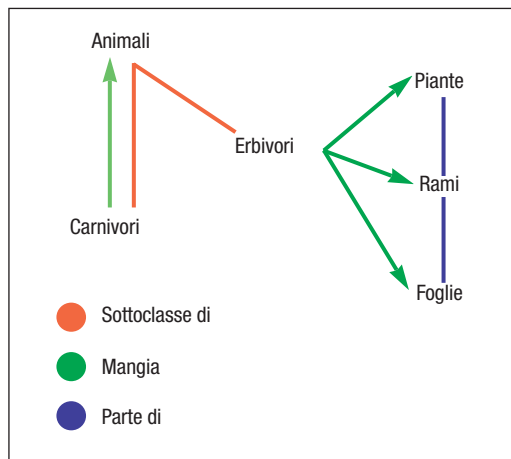


FIGURA 5

Una rappresentazione del modello OIL

matiche, ritrovando problemi e questioni antichi, ma oggi di grande attualità ed importanza, dal punto di vista della ricerca teorica come della applicazione pratica per l'informazione e la comunicazione.

Bibliografia

- [1] Abbagnano N: *Dizionario di filosofia*. UTET, Torino 1968, voce Metafisica, p. 561.
- [2] Berners-Lee T: *Semantic web road map, Internal note*. World Wide Web Consortium, 1998.
- [3] Broekstra J, et al.: *Adding formal semantics to the Web, building on top of RDF Schema*. 2000
- [4] Crombie AC: *Da Agostino a Galileo. Storia della scienza dal V al XVII secolo*. Feltrinelli, Milano 1970, p. 221-222.
- [5] Decker S, et al.: *The Semantic Web: The Roles of XML and RDF*. IEEE, set-ott 2000.
- [6] Eco U: *L'albero di Porfirio, in Semiotica e filosofia del linguaggio*. Einaudi, Torino 1984.
- [7] Horrocks I, et al.: *The Ontology Interchange Language OIL*. Tech. Report, Free Univ. of Amsterdam, 2000.
<http://www.ontoknowledge.org/oil/>.

MASSIMO PARODI è professore di Storia della Filosofia Medievale presso l'Università degli Studi di Milano. Fa parte del gruppo di docenti della Facoltà di Lettere e Filosofia che ha promosso il Master in Metodologie dell'informatica e della comunicazione per le scienze umanistiche.
e-mail: massimo.parodi@unimi.it

ALFIO FERRARA è laureato in filosofia e collabora attualmente col Dipartimento di Scienze dell'Informazione dell'Università degli Studi di Milano, dove si occupa di problemi connessi all'integrazione di sorgenti di dati eterogenee e al Semantic Web.
e-mail: ferrara@dsi.unimi.it