

Unica pecca riscontrabile nel lavoro di Oliver è che l'entusiasmo lo porta a minimizzare i problemi – procedurali, formativi, istituzionali, tecnologici, ecc. – che derivano inevitabilmente dall'introduzione di un nuovo codice. *Introducing RDA* di Chris Oliver resta comunque un libro (dal rassicurante formato di un quaderno scolastico) che illustra molto bene il nuovo codice, a condizione di conoscere un minimo la lingua inglese, e che può soddisfare la curiosità che si sta diffondendo su questo argomento. È anche un ottimo riassunto di *FRBR* e *FRAD*, oltre che, specie nel capitolo 7, delle novità che investono la catalogazione e che derivano, in particolare, dal più ampio accesso che si vuol garantire ai dati prodotti da noi bibliotecari.

Agnese Galeffi
Scuola vaticana di biblioteconomia

Le ontologie, a cura di Maria Teresa Biagetti. «Aida informazioni», n.1-2 (2010), 140 p. ISSN 1121-0095.

Oggi non ci soddisfano più le tradizionali strategie di ricerca di documenti, perché abbiamo bisogno di accedere direttamente al contenuto per reperire l'informazione che ci interessa. Le potenzialità del web semantico fanno intravedere la fattibilità di una navigazione mirata per gli utenti, e in questo scenario diventa fondamentale il ruolo delle ontologie formali, strumenti elaborati nell'ambito dell'intelligenza artificiale per la rappresentazione della conoscenza e riproposti per facilitare l'accesso alle informazioni ed ai concetti, per definire sia i significati dei termini interpretandoli correttamente in altre lingue, sia la natura e la struttura di un dominio di conoscenza utilizzando le relazioni logiche.

Ma cosa sono queste ontologie e come possiamo utilizzarle? Per accostarsi a questa tematica e migliorarne la comprensione è fondamentale leggere il numero monografico della rivista «Aida informazioni», interamente e gratuitamente disponibile sul sito AIDA, segno evidente che tale argomento è diventato un elemento-chiave per la gestione della conoscenza, essendo così strettamente collegato al trattamento della documentazione.

Il saggio introduttivo di Maria Teresa Biagetti offre una panoramica sui nuovi criteri di gestione semantica presenti nel Web. La descrizione completa e dettagliata delle ontologie muove dalla considerazione che queste sono strumenti profondamente diversi dalle classificazione e dai *thesauri*, ma più appropriati per superare i limiti delle tradizionali strategie di ricerca nel Web e recuperare le informazioni in quanto, disponendo di molte relazioni semantiche, risolvono i problemi di ambiguità e giungono a realizzare l'interoperabilità tra diversi sistemi di organizzazione della conoscenza. Passando in rassegna alcune definizioni, Maria Teresa Biagetti ne propone una che riassume molto bene la loro natura: «le ontologie sono schemi concettuali condivisi e strutturati, definiscono i concetti e le relazioni che sono impiegati per descrivere e rappresentare un'area della conoscenza utilizzando linguaggi non ambigui, suscettibili di essere processati dalle macchine». Le diverse tipologie di ontologie (ontologie leggere, mappe topiche, ontologie di dominio e *upper ontologies*) sono messe a confronto con alcuni linguaggi utilizzati per il web semantico, come RDF (*Resource Description Framework*), che risulta ancora poco soddisfacente dal punto di vista dell'espressività logica, e OWL (*Ontology Web Language*), linguaggio fondato sulla *Description Logics*, branca della logica formale, che consente lo sviluppo di ontologie e la loro condivisione in rete attraverso la costruzione di molte relazioni tra le classi (cardinalità, eguaglianza, disgiunzione e simmetria), di cui vengono forniti due esempi per la costruzione di una sintassi. Si procede con l'analisi di *WordNet* (un dizionario dotato di struttura relazionale a modello dei *thesauri*) e delle sue successive elabo-

razioni, che permette di costruire le relazioni tra i termini di tipo lessicale e concettuale, utile soprattutto per il recupero di informazioni e la disambiguazione dei significati. Per completare la panoramica delle ontologie, vengono illustrate quelle basate sulla logica formale, come *SUMO* (*Suggested Upper Merged Ontology*), una ontologia molto vasta, che comprende concetti astratti e generali comuni a vari ambiti scientifici, e *Cyc*, un progetto di ontologia formale caratterizzato da un complesso sistema di categorie e dotato di un motore per le inferenze.

Il lavoro di Maria Teresa Pazienza inizia con il riconoscimento della stretta relazione tra le ontologie e il web semantico, che può «aggiungere semantica esplicita al contenuto testuale dei documenti pubblicati», e si sofferma ampiamente nella descrizione storica e filosofica delle ontologie. Con l'ausilio di definizioni e passaggi chiave, valuta l'evoluzione delle ontologie nell'ambito dell'ingegneria della conoscenza e della *Information Technology*, apprezzandone la funzione di disambiguazione semantica a supporto del ragionamento automatico e dei sistemi di intelligenza artificiale. Dopo un cenno alle tipologie principali, l'autrice individua una nuova tipologia, quella delle ontologie dei *fenomeni collettivi* caratterizzati da proprietà specifiche, composti da molti individui e distinti in tre gruppi (collezioni, entità collettiva e gruppo sociale) e ipotizza una serie di relazioni e sottorelazioni intercorrenti tra i gruppi ed i singoli componenti, per dimostrare la complessità di rappresentazione di tale trama di collegamenti. Ma il punto focale del saggio è costituito dal percorso didattico per la costruzione di una ontologia, che prende avvio dallo stretto rapporto tra ontologie e linguaggio naturale e introduce il processo di *ontology learning*, ovvero l'estrazione (semi-)automatica di concetti e relazioni a partire da una base di conoscenza utilizzando tecniche di elaborazione del linguaggio, prerequisito essenziale per la costruzione di una ontologia. Sono individuate le fasi progressive necessarie alla costruzione: l'estrazione di termini afferenti al dominio di interesse, l'identificazione di sinonimi, la formazione e l'organizzazione dei concetti, la definizione delle relazioni ed il popolamento tramite il recupero di istanze in maniera automatica (da fonti strutturate) o manuale. Mettere in relazione i concetti non basta, occorre anche collocarli in un *framework* ontologico che include i metadati necessari a descrivere l'ontologia, sia dal punto di vista semantico per esprimere il significato dei termini ed il livello di approfondimento, sia dal punto di vista pragmatico rivolto alle finalità di utilizzo dell'ontologia.

Dopo una carrellata sui diversi criteri e relativi strumenti per la valutazione delle ontologie, si passa a illustrare il *Semantic Turkey*, prodotto dal gruppo di ricerca ART dell'Università di Roma Tor Vergata, una piattaforma che si propone di fornire soluzioni innovative per l'acquisizione della conoscenza dal Web e la sua gestione in strutture ontologiche personalizzate. Attraverso operazioni semplici (come il *drag and drop*) si recuperano informazioni dalle pagine Web che possono essere inserite in uno schema semantico realizzato con questo *editor* di ontologie, secondo proprietà e caratteristiche predefinite. Il software permette di visualizzare graficamente l'ontologia, in modo da evidenziare legami e relazioni, classi e proprietà degli individui che la compongono, mentre con il *mouse* si apre una finestra contenente l'URL delle pagine corrispondenti sul Web. Il *Semantic Turkey* è progettato per l'integrazione tra la conoscenza esplicita registrata dall'utente e la conoscenza di sistema basata su tecnologie innovative, al fine di favorire lo scambio e la condivisione della conoscenza tra ambienti diversificati.

L'articolo di Stefania Costantini, Davide Lanti e Alessio Paolucci affronta il tema delle ontologie dal punto di vista della *Web Intelligence*, un'espressione di intelligenza artificiale che riconosce l'importante ruolo delle macchine per manipolare e gestire la conoscenza estraendola dal linguaggio naturale. In quest'ambito, i concetti sono tradotti in formato logico per disambiguare i termini e rappresentarli con l'ausilio di regole gram-

maturali, che analizzano sintatticamente una frase e ne interpretano le componenti. Le ontologie sono definite «a contract for mining», un contratto sul significato, per individuare univocamente le entità e le proprietà di un insieme lessicale di conoscenza. È interessante la visione della struttura del web semantico articolata in vari livelli a partire dal linguaggio RDF, che rende la sintassi leggibile dalle macchine, e la sua successiva evoluzione in RDF Schema applicabile ad uno specifico dominio. Al di sopra dei linguaggi si trovano le ontologie, originate da processi di codifica e generazione della conoscenza, che a loro volta possono essere interrogate con l'apposito linguaggio SPARQL, basato su corrispondenze sintattiche.

Ma cosa sono gli agenti citati nel titolo e come si rapportano alle ontologie? Sono entità software in grado di sfruttare gli elementi semantici delle pagine web al fine di guidare l'utente verso le informazioni desiderate, ma per comunicare devono usare la stessa lingua, quindi basarsi su un dizionario comune, visto come una forma semplificata di ontologia. Se poi gli agenti operano in un particolare dominio di conoscenze tecniche, ecco che le ontologie offrono il loro supporto, sia nel caso di un dominio statico (non soggetto a cambiamenti) prevedibile a priori, sia nel caso di un dominio dinamico, in cui gli agenti operano su regole generiche ma possono interrogare l'ontologia per recepirne le peculiarità.

Un'applicazione di questi studi è la piattaforma DALI, un linguaggio logico sviluppato per definire agenti che interagiscono con il mondo esterno e svolgono un vero processo di ragionamento (*modulo di meta ragionamento*) basato sulle ontologie e sulle regole logiche e sintattiche. L'importanza delle ontologie è ulteriormente evidenziata da due aree di ricerca emergenti: il *Data integration*, la possibilità di accedere e gestire dati provenienti da sorgenti eterogenee, confrontabili con procedure di mappatura semantica dei concetti; il *Natural language processing*, un metodo per integrare le tecnologie di elaborazione del linguaggio con le ontologie intese come sorgenti di conoscenza. Il discorso si chiude tornando al tema della navigazione sul Web tramite il *Linked data*, un paradigma che, progettando ontologie su conoscenza formalizzata, moltiplica i legami semantici tra ontologie e sintassi diverse, quasi a trasformare in realtà il sogno di Tim Berners sulla cooperazione tra uomo e macchina parzialmente intelligente.

Fausto Giunchiglia e Vincenzo Maltese approfondiscono il discorso sulle ontologie leggere, caratterizzate da una struttura gerarchica ad albero e utilizzate per la classificazione di oggetti, ed in particolare sulle ontologie leggere a faccette che permettono una migliore progettazione di pagine web, l'organizzazione e la navigazione tra i dati. Le faccette – intese in questa sede come aspetti della conoscenza che formalizzano un dominio individuandone entità, proprietà e azioni – sono trattate su un piano didattico, ricollegandosi al sistema di indicizzazione ideato da Ranganathan. Dopo una chiara distinzione tra ontologie descrittive (usate per descrivere oggetti del mondo reale) e ontologie di classificazione (usate per organizzare oggetti in sottoinsiemi), si esaminano le ontologie leggere a faccette «dove ciascun termine e corrispondente concetto che occorre nelle etichette dei suoi nodi devono corrispondere ad un termine e corrispondente concetto nella base di conoscenza di supporto, modellata come schema di classificazione a faccette».

È interessante seguire il percorso suggerito per la costruzione di una tale ontologia: individuato uno specifico dominio, si crea una struttura di faccette basate sul metodo DEPA (disciplina, entità, proprietà, azione), una variazione proposta da Bhattacharyya della formula PMEST di Ranganathan. A queste faccette viene aggiunto un modificatore che introduce elementi utili per qualificare i concetti e che può essere universale (modificatore spaziale e temporale) o speciale (relativo a manifestazioni e categorie tipiche di un dominio). La struttura sarà utilizzata per classificare i documenti, associandoli a una etichetta (o nodo dell'ontologia leggera) e attribuendo i concetti presi dallo schema di

classificazione con i termini collegati, mentre con le categorie DEPA si può disambiguare il significato. Tale metodo stabilisce le relazioni semantiche tra i concetti e il contesto di riferimento e permette di identificare anche le informazioni contestuali mancanti, ottenendo un maggiore controllo sul linguaggio. Per l'indicizzazione dei concetti, si consiglia di utilizzare la *Postulate-based Permuted Subject Indexing System* (POPSI), sviluppato attraverso l'interpretazione logica del linguaggio di indicizzazione, che prescrive l'uso di un apparato di parole (preposizioni, congiunzioni e participi) per comunicare l'esatto significato del soggetto. La metodologia POPSI prevede alcuni passi già noti agli indicizzatori, come l'analisi dell'espressione indicativa del soggetto, la standardizzazione del soggetto e l'ordinamento alfabetico.

Il contributo di Silvia Gaio, Stefano Borgo, Claudio Masolo, Alessandro Oltramari e Nicola Guarino chiarisce che le ontologie formali hanno lo scopo di esplicitare il significato di insiemi di informazioni (come le classificazioni) e che la loro importanza è cresciuta con il problema della gestione dei dati presenti in rete. In questo scenario si collocano le ontologie formali fondazionali, che forniscono la descrizione di concetti generali e risultano adatte allo scambio di informazioni tra domini diversi, dove sono più numerosi i problemi di ambiguità terminologica. Ne è un esempio l'ontologia DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*), sviluppata nell'ambito del progetto *Wonder Web*, che ha lo scopo di catturare le categorie ontologiche che emergono dal linguaggio naturale e dalla conoscenza comune, presentandosi come modulo di riferimento per confrontare ontologie fondazionali. DOLCE offre una categorizzazione tra continuanti (enduranti, entità che esistono nel tempo come gli oggetti) ed occorrenti (perduranti, entità che accadono nel tempo come gli eventi); qualità (entità e proprietà come forme, colori, suoni) e *qualia* (variabilità nello spazio qualitativo) ed entità astratte (regioni qualitative, spazi che raccolgono entità diverse); permette inoltre di specificare alcune relazioni di base intuitive, immediate e applicabili a domini diversi. Quindi, DOLCE ha una duplice applicazione, che può essere impiegata sia per sviluppare nuove ontologie in un determinato dominio caratterizzandone i concetti, sia per integrare ontologie già esistenti allo scopo di migliorarle e ampliarne le possibilità applicative.

Nel suo insieme, il numero monografico di «AIDA Informazioni» è un testo valido, che offre un apprezzabile aggiornamento sul tema delle ontologie e che, presentando una impostazione rigorosa, risulta essere un utile supporto per applicazioni volte ad organizzare il patrimonio concettuale di domini scientifici e tecnici. Un altro merito di questa pubblicazione è lo sforzo di introdurre in Italia un approccio scientifico alle ontologie e all'organizzazione della conoscenza, ancora poco diffuso, tanto da concordare pienamente con il commento conclusivo di Cladio Gnoli nella rubrica *Organizzazione della conoscenza*: le classificazioni bibliografiche sembrerebbero in concorrenza e addirittura superate dai nuovi strumenti di rappresentazione della conoscenza che si basano sulla semantica e su linguaggi computerizzati, ma bisogna riconoscere che tutti questi nuovi sistemi di indicizzazione vanno considerati come integrazioni di quelli tradizionali per ottenere una migliore semantizzazione del patrimonio documentario.

Infatti, l'applicazione delle ontologie, realizzata inizialmente nei processi di elaborazione linguistica e di rappresentazione della conoscenza, ora sta rispondendo al bisogno di condividere schemi semantici all'interno di architetture connesse a un dominio per comunità ristrette di utenti, ma si ha l'impressione che occorra percorrere ancora molta strada per un loro impiego effettivo nelle biblioteche al fine di organizzare le risorse afferenti a un dominio e renderle confrontabili e usabili da utenti esperti.

Alessandra Ensoli

Istituto superiore per la protezione e la ricerca ambientale, Roma