

Dig *Italia*

Anno VI, Numero 1 - 2011

Rivista del digitale nei beni culturali

ICCU-ROMA

Le Ontologie.

Numero speciale monografico di AIDAinformazione

Maria Lucia Di Geso

ICCU

La rivista *AIDAinformazioni* dell'Associazione per la documentazione avanzata (AIDA) ha inaugurato il 2010 con un numero speciale monografico¹ dedicato ad un argomento di particolare interesse per chi si occupa di web semantico, vocabolari controllati e, più in generale, di strumenti di gestione della conoscenza: le *ontologie*. Il fascicolo, curato da Maria Teresa Biagetti (Università degli studi di Roma "La Sapienza") - che apre la pubblicazione con una presentazione ed un lungo ed esaustivo intervento² - si avvale di una serie di contributi di grande spessore, di studiosi e sviluppatori, che coprono diversi aspetti, anche problematici, della riflessione sull'applicazione delle ontologie nel mondo dell'informazione. Il contributo della Biagetti si sofferma sull'ausilio che le ontologie possono offrire sia alle attività di indicizzazione che nel recupero dei documenti nell'ambito delle biblioteche digitali e mette in evidenza la loro enorme potenzialità in quanto strumenti capaci di creare "nuova conoscenza": infatti l'impiego delle ontologie a supporto della ricerca nel web e l'integrazione automatica delle informazioni provenienti da fonti diverse mediante l'utilizzo di s/w in grado di applicare modalità di ragionamento automatico può favorire, ad esempio, «l'individuazione di precise entità - persone, avvenimenti, luoghi -

ma anche di concetti e di argomenti che, pur non essendo esplicitamente citati nel documento, possono essere semanticamente implicati dal documento nella sua interezza, e ai quali i documenti possono rinviare»³.

L'articolo di Biagetti elenca e spiega i vari tipi di ontologie esistenti: dalle ontologie fondazionali che rappresentano concetti più generali, alle ontologie di dominio (che abbracciano singoli campi disciplinari) alle *task ontologies* (finalizzate ad una specifica attività) fino alle *application ontologies* (riservate a particolari attività all'interno di un campo scientifico specifico).

Si analizzano anche alcuni aspetti relativi alla costruzione delle ontologie, al loro rapporto con i thesauri (come ad esempio AGROVOC Thesaurus) e alla loro importanza per trasformare il web "da una rete di links ad una rete di significati" in grado di dare risposte immediate a domande complesse. Vengono presentate descrizioni analitiche di alcune ontologie formali, basate su linguaggi formali altamente espressivi che connettono diverse categorie e possono favorire lo scambio semantico tra diversi ambienti culturali e linguistici. Viene infine presentata la base dati lessicale WordNet, dizionario basato su una struttura semantica e realizzato assumendo come modello i thesauri linguistici.

¹ *Numero speciale monografico Le ontologie*, a cura di Maria Teresa Biagetti, «AIDAinformazioni», 28 (2010), n. 1-2, <http://www.aidaweb.it/>

² Maria Teresa Biagetti, *Le ontologie come strumenti per l'organizzazione della conoscenza in rete*, «AIDAinformazioni», 28 (2010), n. 1-2, p. 9-31, <http://www.aidaweb.it/>

³ M. T. Biagetti, *ivi*, p. 10.

Nel secondo saggio⁴, Maria Teresa Pazienza, (Università degli studi di Roma “Tor Vergata”) sostiene che, ai fini della ricerca di informazione, è divenuto ormai vitale sviluppare ontologie di dominio che siano estendibili e riutilizzabili in più contesti applicativi, vista la difficoltà di condividere uno stesso modello cognitivo.

Partendo dalla definizione del concetto di *ontologia*, l'autrice sviluppa il discorso dell'utilizzazione delle ontologie come supporto al ragionamento automatico dei sistemi intelligenti, dando particolare risalto alla cosiddetta *ontologia dei fenomeni collettivi* (collezioni, entità collettive, gruppi sociali) la cui realizzazione richiede un forte livello di complessità e granularità, poiché in essa giocano un ruolo fondamentale le relazioni tra i membri agenti del fenomeno; d'altro canto «il comportamento di un fenomeno collettivo ha un valore proprio, univoco e diverso dal comportamento dei singoli componenti»⁵.

L'autrice si sofferma poi sulle complicate metodologie relative all'identificazione dei concetti da inserire in un'ontologia, *l'ontology learning*, e *l'ontology population*: la prima fase (*ontology learning*), si basa sull'acquisizione di un modello di dominio direttamente dai dati, dalle loro definizioni e dalle loro relazioni: schemi di basi di dati, tassonomie, dizionari, collezioni di documenti correlati (*corpora*) possono offrire ai ricercatori utili strumenti per la derivazione di ontologie, dopo particolari elaborazioni: nel caso XML-DTD, diagrammi UML o schemi di data base, *l'onto-*

logy learning consiste nell'effettuare un *mapping* tra le definizioni degli schemi e le corrispondenti definizioni ontologiche. Fanno parte dell'*ontology learning* le fasi di: estrazione dei termini, identificazione dei sinonimi, formazione dei concetti⁶, organizzazione gerarchica degli stessi, identificazione delle sue relazioni, del suo dominio e range di validità. Una volta fornita una rappresentazione del dominio di interesse, subentra la fase dell'*ontology population* che consiste nel popolare l'ontologia di istanze, definendo le specifiche istanze delle classi e indicandone valori e restrizioni. Il popolamento di un'ontologia avviene mediante l'annotazione semantica⁷ che può essere completamente manuale, oppure servirsi di *tool* di annotazione automatica, utilizzati per popolare un'ontologia e capaci di suggerire all'utente un insieme di annotazioni semantiche.

L'ultima parte del saggio analizza le metriche di valutazione dell'attività di sviluppo di ontologie (alcuni criteri generali sono: copertura, precisione, dettaglio assiomatico, sostenibilità, flessibilità) e si conclude con la presentazione di un *framework* per la creazione e gestione delle ontologie: il Semantic Turkey (ST), sviluppato dal gruppo di ricerca ART dell'Università di Roma Tor Vergata, con l'intento di «fornire all'utente un ambiente unificato per il Web, la ricerca di informazioni rilevanti di interesse, e la loro memorizzazione in uno schema concettuale personale all'interno di un'ontologia importata o autonomamente definita»⁸.

⁴ Maria Teresa Pazienza, *Ontologie e Web semantico: proprietà e problematiche connesse al loro uso diffuso*, «AIDAinformazioni» cit., p. 33-61, <http://www.aidainformazioni.it/pub/pazienza122010.pdf>.

⁵ M. T. Pazienza, *Ontologie*, cit., ivi, p. 40.

⁶ Un concetto è definito come «una tripla costituita da: una definizione intenzionale del concetto (descrizione in linguaggio naturale del significato intuitivo di un concetto, così come una collezione di attributi), una estensione del concetto, tutte le espressioni lessicali (termini) del concetto in un corpus»: cfr. M. T. Pazienza cit., ivi, p. 47.

⁷ Si definisce annotazione semantica “la rappresentazione formale di un contenuto (istanze) attraverso i concetti, le relazioni e le istanze come sono descritte in un'ontologia e collegate alla risorsa annotata”: cfr. M. T. Pazienza, *Ontologie* cit., p. 49.

⁸ M. T. Pazienza, *Ontologie* cit., p. 56.

Il saggio di Costantini, Lanti e Paolucci⁹ approfondisce il ruolo che le ontologie possono svolgere nell'ambito di un "sistema multiagente", cioè un «insieme di entità *software* chiamate *agenti*»¹⁰. Gli agenti web, intesi come agenti in grado di sfruttare gli elementi semantici della pagine web al fine di guidare l'utente verso le informazioni desiderate, non possono prescindere dalle ontologie per risolvere i problemi derivanti, ad esempio, dall'integrazione dei sistemi *legacy*, problemi di linguaggi diversi, di utilizzazione di formati diversi per i messaggi, etc. In questo contesto vengono analizzate due applicazioni fondamentali delle ontologie in un sistema multiagente: l'ausilio alla comunicazione tra agenti e la formazione di una base di conoscenza specifica per un particolare dominio di interesse. Nel primo caso, si mette in evidenza la capacità di "traduzione" che le ontologie offrono agli agenti nel loro scambio di informazioni e, per le ontologie in formato RDF, di creazione e scambio di regole grammaticali comuni; nel secondo caso si sottolinea la possibilità di utilizzare ontologie di dominio già esistenti sul web, per impadronirsi dei dettagli del dominio in cui l'agente opera e di cui ha bisogno.

L'articolo si sofferma poi sulla piattaforma multiagente DALI (Linguaggio logico orientato agli agenti) sviluppato da Stefania Costantini e utilizzato per la definizione dei componenti dei software agenti; gli agenti DALI per mezzo delle ontologie sono stati messi in grado di ragionare sui messaggi in ingresso al fine di interpretarli (mediante un modulo di metaragionamento), di reagire agli eventi esterni e di interrogare tutte le ontologie presenti, scritte in linguaggio OWL (Linguaggio ontologico per il web).

Il quarto saggio¹¹ di Fausto Giunchiglia e Vincenzo Maltese, analizza l'uso delle ontolo-

gie in un ambito più vicino al mondo bibliotecario, cioè quello della classificazione di oggetti, quali ad esempio libri, foto e pagine web. A tale scopo vengono utilizzate le cosiddette *ontologie leggere*, dotate di una struttura gerarchica ad albero nella quale ciascun nodo è identificato da un'etichetta in linguaggio naturale. Tali ontologie possono trasformarsi in *ontologie leggere formali* quando le etichette originarie sono tradotte in formule matematiche espresse in linguaggio formale, quale ad esempio quello della logica descrittiva. Gli autori sostengono che, sebbene l'impiego delle *ontologie leggere* possa avere un ruolo fondamentale nella classificazione automatica dei documenti, nei sistemi di ricerca automatici e nella soluzione dei problemi di eterogeneità semantica tra differenti ontologie, tuttavia non hanno avuto finora la diffusione che ci si sarebbe aspettata, forse a causa del notevole impegno ed investimento richiesto dalla loro adozione. L'obiettivo degli autori è dunque quello di proporre un particolare tipo di *ontologie leggere*, le *ontologie leggere a faccette*, quale soluzione ai problemi delineati. Nella costruzione delle faccette gli autori si sono rifatti all'approccio del bibliotecario indiano Ranganathan ed al successivo sviluppo nel sistema di indicizzazione POPSI. Le faccette rappresentano differenti aspetti della conoscenza e formalizzano le principali caratteristiche di ciascun dominio (ad esempio medicina, letteratura, sport, etc.), le entità che vi appartengono (ad esempio, parti del corpo in medicina), le proprietà di tali entità (ad esempio le malattie per le parti del corpo) e le azioni che possono essere intraprese (ad esempio la prevenzione, la cura in medicina).

Gli autori mettono in evidenza i vantaggi di una rappresentazione a faccette che permet-

⁹ S. Costantini – D. Lanti – A. Paolucci, *Agenti ed Ontologie: verso la Web Intelligence*, «AIDainformazioni», cit., p. 63-86.

¹⁰ S. Costantini – D. Lanti – A. Paolucci, *Agenti ed Ontologie*, cit., p. 69.

¹¹ Fausto Giunchiglia – Vincenzo Maltese, *Ontologie leggere a faccette*, «AIDainformazioni» cit., p. 87-106, <http://www.aidainformazioni.it/pub/giunchiglia122010.pdf>.

te, ad esempio, sia la post-coordinazione sia la costruzione di soggetti complessi (in campo biblioteconomico) combinando termini da faccette differenti sia per l'indicizzazione che per la classificazione ed il recupero dei documenti. L'ultima parte del saggio presenta la metodologia POPSI per derivare i soggetti partendo dai titoli associati ai documenti da indicizzare.

L'ultimo saggio¹² è un contributo a più mani che, partendo dalle ontologie formali (dette anche computazionali), distingue all'interno due tipi diversi di ontologie: le già citate ontologie "leggere" e le ontologie fondazionali. Le prime, sviluppate per domini specifici di applicazione o per compiti predeterminati, hanno spesso la forma di tassonomie e facilitano l'accesso ai dati web prodotti da una determinata comunità che condivide un certo vocabolario. Il limite di queste ontologie si riscontra appena la comunità si allarga e gli utilizzatori del sistema non condividono le stesse assunzioni sui termini del vocabolario utilizzato; ciò accade quando occorre scambiare dati e servizi tra comunità multiculturali e plurilinguistiche. Infatti quando un'ontologia non è «sufficientemente ricca di nozioni generali che possono essere usate per comprendere le differenti assunzioni di significato»¹³, rischiando ambiguità nella comprensione dei messaggi fra gli agenti del sistema, decadono anche i servizi che il web è in grado di offrire.

Al contrario nelle ontologie fondazionali, che forniscono schemi più generali e sono destinate a scambiare informazioni tra diversi domini, acquista una rilevanza fondamentale il problema dell'ambiguità terminologica. Esse sono nate proprio allo scopo di «facilitare la negoziazione dei significati dei termini di un

certo dominio attraverso diverse comunità scientifiche o linguistiche»¹⁴ perché il vocabolario che offrono descrive fenomeni che possono essere anche fuori dal dominio ordinario di applicazione. Fondandosi su concetti generali quali quelli di *oggetto*, *evento*, *qualità*, *ruolo* e su relazioni del tipo *parte/tutto*, *dependenza*, *partecipazione*, etc. possono andare incontro a varie esigenze applicative, possono essere riutilizzabili per modellare situazioni differenti e sono ben fondate da un punto di vista filosofico-concettuale.

Date queste premesse, gli autori presentano l'ontologia fondazionale e descrittiva DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*), sviluppata nell'ambito del progetto internazionale WonderWeb, con lo scopo di "catturare" le categorie ontologiche che emergono nel linguaggio naturale. Gli obiettivi del progetto si possono riassumere nei seguenti punti:

- «sviluppare una "libreria" di ontologie fondazionali capaci di coprire un ampio raggio di domini applicativi per permettere, da una parte, lo sviluppo di ontologie di dominio più precise, dall'altra, l'integrazione semantica delle ontologie;
- sviluppare una famiglia di linguaggi con cui descrivere le ontologie estendendo i linguaggi web *standard* e allo stesso tempo mantenendo la massima compatibilità, al fine di offrire la necessaria flessibilità e massimizzare l'interoperabilità tra linguaggi diversi;
- sviluppare una serie di tecniche e metodologie per offrire un approccio ingegneristico allo sviluppo e all'uso delle ontologie, in particolare per permettere l'integrazione semantica e la condivisione di ontologie differenti;

¹² Silvia Gaio - Stefano Borgo - Claudio Masolo - Alessandro Oltramari - Nicola Guarino. *Un'introduzione all'ontologia DOLCE*, «AIDAinformazioni» cit., p. 107-125, <http://www.aidainformazioni.it/pub/gaio122010.pdf>.

¹³ S. Gaio - S. Borgo - C. Masolo - A. Oltramari - N. Guarino, *Un'introduzione*, cit., p. 110.

¹⁴ *Ibidem*.

¹⁵ Ivi, p. 112.

- sviluppare un supporto tecnico per le richieste delle applicazioni del Semantic Web»¹⁵.

L'ontologia DOLCE ha avuto due sviluppi principali: da un lato, è stata utilizzata come punto di partenza per sviluppare nuove ontologie, come ad esempio, nell'ambito del *design* ingegneristico, dove ha offerto una caratterizzazione della nozione di "funzione" in ingegneria, nell'ambito dei servizi pubblici e dell'*E-Government* dove ha contribuito alla

modellizzazione concettuale, in ambito sanitario, dove è stata usata per categorizzare le immagini mediche e i dati di test di laboratorio.

L'altro utilizzo ha riguardato il contributo di DOLCE all'integrazione ed al miglioramento di ontologie già esistenti e in questo ambito ha contribuito a produrre LRI-Core, ontologia sviluppata nel settore legale e, in linguistica, all'interno del progetto OntoWordNet, si è cercato di allineare i concetti del database lessicale WordNet a quelli di DOLCE.