

Semantic Web

Anna Lamandini

Insegnante

anna.lamandini@istruzione.it

Abstract

The semantic Web is a technology at the service of knowledge which is aimed at accessibility and the sharing of content; facilitating interoperability between different systems and as such is one of the nine key technological pillars of TIC (technologies for information and communication) within the third theme, programme specific cooperation of the seventh programme framework for research and development (7^oPQRS, 2007-2013). As a system it seeks to overcome overload or excess of irrelevant information in Internet, in order to facilitate specific or pertinent research. It is an extension of the existing Web in which the aim is for cooperation between and the computer and people (the dream of Sir Tim Berners – Lee) where machines can give more support to people when integrating and elaborating data in order to obtain inferences and a global sharing of data. It is a technology that is able to favour the development of a “data web” in other words the creation of a space in both sets of interconnected and shared data (Linked Data) which allows users to link different types of data coming from different sources. It is a technology that will have great effect on everyday life since it will permit the planning of “intelligent applications” in various sectors such as education and training, research, the business world, public information, tourism, health, and e-government. It is an innovative technology that activates a social transformation (socio-semantic Web) on a world level since it redefines the cognitive universe of users and enables the sharing not only of information but of significance (collective and connected intelligence).

Parole chiave: Internet; sistemi semantici; Expert System

L'Europa verso il Web del futuro

i2010 (“società europea dell’informazione per la crescita e l’occupazione”, 2010) è il nuovo quadro strategico della Commissione Europea che definisce gli orientamenti strategici per incoraggiare la conoscenza, l’innovazione e sostenere la crescita, dando priorità ai principali pilastri tecnologici del “Settimo programma quadro per la ricerca e lo sviluppo tecnologico” (7° PQRS 2007-2013). All’interno del 3° Tema, Programma specifico “Cooperazione” del suddetto 7° PQRS, le TIC sono definite catalizzatori d’innovazione e di competitività e perciò costituiscono il nucleo della società basata sulla conoscenza. I progressi realizzati nelle TIC garantiranno la trasformazione della conoscenza in applicazioni concrete e benefici per tutti i cittadini europei. Le priorità delle attività di ricerca sulle TIC si incentreranno su nove pilastri tecnologici chiave, tra i quali i sistemi semantici (metodi e tecniche di acquisizione, interpretazione e recupero pertinente delle conoscenze).

L’Europa intende porsi alla guida verso il Web del futuro o Web 3.0, mezzo globale per lo scambio dei dati, in cui tutti gli utenti saranno costantemente collegati (Internet degli oggetti: collegamento senza fili di macchine, di veicoli, di dispositivi, di sensori e di molti altri oggetti attraverso Internet). *“L’Internet del futuro cambierà radicalmente la nostra società”*, ha dichiarato Viviane Reding, Commissario per la Società dell’informazione e i media, *“Web 3.0 significa potersi dedicare a qualsiasi attività (commerciali, sociali e ricreative) in qualsiasi momento e in qualsiasi luogo, senza soluzioni di continuità, grazie a reti veloci, affidabili e sicure”*.

Nell’ambito del Settimo Programma Quadro per la Ricerca e lo Sviluppo Tecnologico (7°PQR) del 3°tema “Tecnologie dell’informazione e della comunicazione” (TIC), l’Unione Europea ha finanziato il progetto SEALS (Semantic evaluation at large scale) con lo scopo di sviluppare una piattaforma europea per la semantica, al fine di valutare e di confrontare le tecnologie disponibili più adatte al Web Semantico.

Web 2.0 e limiti dei Motori di Ricerca

Il Web 2.0 ha trasformato i siti Web da contenitori di informazioni (Web 1.0), in risorse aperte di funzionalità e contenuti. Si parla di Web 2.0 come piattaforma a disposizione dell’utente, al quale vengono offerte svariate applicazioni (*Social Media* o *Social Software*) che consentono di:

- creare, elaborare, selezionare, pubblicare e gestire contenuti condivisi,
- interagire e collaborare in rete, creando interconnessioni, reti di contatti e comunità online (*Social Networking*).

Il valore aggiunto sta nell’organizzazione, nell’indicizzazione e nella categorizzazione collaborativa (*folksonomy*) dei contenuti generati dagli utenti (*User generated Content* o UGC) e dalla loro condivisione con altri utenti (Intelligenza collettiva- P. Lévy e connettiva- D. De Kerkhove). Tutto ciò ha portato a una crescita esponen-

ziale dei dati, tale per cui, oggi, il Web rappresenta la più grande collezione pubblica di documenti strutturati e non. Parallelamente alla proliferazione di contenuti, sono emerse crescenti difficoltà nei processi di gestione e recupero dei contenuti. I motori di ricerca, basati sull'uso di parole chiave (*keywords*), presentano limitazioni nel loro uso:

1. effetto rumore: derivante da grandi quantità di documenti non rilevanti e non pertinenti a causa di polisemie, omonimie, sinonimie;
2. effetto silenzio: assenza o scarsa quantità di risultati rilevanti e pertinenti;
3. falsi negativi: nei risultati della ricerca non compaiono pagine che sarebbero state di nostro interesse, ma che non contenevano esattamente la parola chiave da noi immessa, magari contenevano un suo sinonimo;
4. un 80% di "hidden Web" o "deep Web" (la rete nascosta e profonda), cioè i contenuti non rintracciabili, pagine periferiche, immagini, files audio, files video, file flash, archivi zippati, informazioni contenute in basi di dati, contenuti dinamici che cambiano in tempo reale, etc...;
5. mancanza di integrazione di informazioni, di cooperazione e interoperabilità tra sistemi diversi per rispondere a domande complesse (es. l'organizzazione di un viaggio).

Il Semantic Web (SW) o Web Semantico si propone come ambiente che, dando rilevanza al significato dei termini, consente:

1. una ricerca più precisa, mirata e pertinente;
2. la riduzione di risposte ridondanti;
3. il reperimento, l'integrazione e l'utilizzo delle informazioni.

Semantic Web o Web Semantico

Il Semantic Web (SW) è un'estensione del Web attuale, il cui scopo è la cooperazione tra computer e persone, affinché le macchine possano essere di maggior supporto alle persone nell'esecuzione automatica di compiti e nel rispondere a domande complesse. La realizzazione del Semantic Web è possibile solo attribuendo un significato ben definito all'informazione (attraverso linguaggi standardizzati: RDF e ontologie), cosicché le macchine possano:

1. attivare funzionalità di "comprensione" semantica (modalità machine-understandable);
2. integrare i dati e metterli in relazione, attraverso operazioni logiche;
3. elaborarli automaticamente per ricavare delle inferenze e delle deduzioni, cioè attivare processi, attraverso i quali, partendo da conoscenze note, si giunge ad ottenere nuove conoscenze (modalità "machine-processable").

La connessione fra i testi e le parole, infatti, va al di là del legame sintattico e si basa sul significato dei documenti pubblicati e sulla possibilità di metterli in relazione fra loro, mediante parole-chiave che "sostengono" il testo scritto (i meta-

dati). Perciò, è necessario passare dalla sintassi dei documenti, alla rappresentazione formale dei loro contenuti, attraverso linguaggi standardizzati, che sono in continua evoluzione. Il modello RDF (*Resource Description Framework*, definito dal W3C) è il primo passo verso un'interoperabilità semantica. RDF codifica il significato dell'informazione in insiemi di *triple* o "frasi" elementari, costituite da tre elementi: soggetto (risorsa), predicato (proprietà) e oggetto (valore). Con RDF possiamo collegare una risorsa Web ad un'altra e specificare che tipo di relazione possiedono: ad es. la pagina ufficiale della "Ferrari" è collegata alla pagina della categoria "auto sportive" di Wikipedia. RDF, pur essendo un linguaggio potente per la rappresentazione della conoscenza, non può operare inferenze o deduzioni sulle relazioni stabilite tra le risorse. Per dare un significato a queste relazioni è necessario passare al livello superiore: l'ontologia, "una definizione formale e condivisa di concetti e delle relazioni che sussistono tra di essi, per formare una rete semantica" (Gruber, 1995). Attraverso un'ontologia si può specificare che, l'affermazione "la Ferrari è un tipo di auto sportiva" significa anche dire che la Ferrari è un'auto e che, perciò, ha un motore, un prezzo e così via. Un'ontologia, tra le altre cose, consente di condividere concetti e relazioni con altre ontologie. Una delle ontologie più utilizzate è quella di DBPedia, costruita utilizzando concetti e relazioni trovati all'interno di Wikipedia. Le ontologie arricchiscono il funzionamento del Web: migliorano l'accuratezza delle ricerche e permettono lo sviluppo di software intelligenti in grado di integrare i dati, provenienti da fonti diverse, rispondendo così a domande complesse, la cui risposta non si trova in una singola pagina Web, ma richiede inferenze tra set d'informazioni diversificate.

Linked Data e i vantaggi applicativi del Semantic Web

Gli obiettivi del Semantic Web sono raggiungibili se i computer avranno accesso a una collezione di dati interconnessi (*Linked Data*), per poter creare un "Web of Data" e non una semplice collezione di dati. Perciò, Sir Tim Berners-Lee ha lanciato una nuova sfida: costruire un Web di dati aperti e collegati fra loro (LOD: ***open linked data***). Un'applicazione di questo nuovo modo di concepire i dati potrebbe essere nell'industria farmaceutica, dove parti di ricerche, inizialmente non collegate, riguardanti una malattia o un farmaco possono essere congiunte e unite.

Se le risorse saranno descritte in modo formale, attraverso l'uso di ontologie, specificando concetti e relazioni che legano i contenuti, sarà facilitata la navigazione per concetti, lungo i link che collegano il significato delle informazioni. In fase di ricerca diretta, un motore di ricerca semantico può analizzare le interrogazioni ("*query*") degli utenti, identificare gli elementi ontologici collegati alle richieste e presentare loro i possibili significati identificati, permettendo, così, di raffinare la ricerca ("disambiguazione") e /o di prendere in considerazione concetti correlati. L'inserimento di informazioni sarà perfezionato

da un “*tagging concettuale*”, (non più con stringhe di testo) che utilizzi etichette che reindirizzino al concetto cui si riferiscono. La condivisione di ontologie, infine, permetterà la realizzazione di un’interoperabilità semantica, ovvero, l’interscambio di dati fra sistemi diversi, basata sul significato dell’informazione. Inoltre, si auspica un’interoperabilità di interconnessioni personali (“*Global Giant Graph*” Grafo o network gigante e globale), se si adotteranno formati standardizzati (come per i documenti), tra social network diversi, per trasformare il Web 2.0 o Web sociale, in un grande social network in grado di coniugare l’intelligenza connettiva con quella delle macchine (Tim Berners-Lee).

Applicazioni semantiche: il caso Cogito di Expert System

La ricerca semantica promette di “trovare tutto, prima e meglio”(M. Varone, CTO di Expert System, Modena 2008); consente, perciò di semplificare la gestione della conoscenza non strutturata, pur non essendo perfetta. Tuttavia, occorreranno alcuni anni affinché la ricerca semantica diventi pervasiva e orizzontale. Allo stato attuale, consente di offrire strumenti per compiti precisi:

1. ricerche ristrette a settori specifici della conoscenza (intranet, domini specifici di conoscenza, gestione di servizi di *customer care*);
2. applicazioni che consentono di ottenere informazioni di pubblica utilità, in tempo reale, via SMS (disponibilità dei servizi di trasporto pubblico; notizie sul meteo, su eventi culturali; farmacia di turno e medici disponibili etc.).

La semantica serve anche a questo, a rendere semplice l’utilizzo delle informazioni e ridurre il tempo necessario ad avere una risposta.

Alla base di tali applicazioni c’è Cogito, software semantico della modenese Expert System, che analizza testi digitali e ne estrapola concetti e relazioni rilevanti.

Infatti, l’obiettivo di Cogito è di consentire alle imprese un immediato recupero dei dati sia strutturati in database, sia quelli non strutturati: documenti, articoli, libri, e-mail, blog, pagine Web, SMS ecc...

Funzionalità principali

1. Indicizzazione semantica di documenti diversi in base alle impostazioni dall’utente;
2. Ricerca semantica con logica concettuale: ad esempio, se si cerca “*ambiente*”, saranno individuati anche i testi che parlano di *habitat naturali, ecosistemi*;
3. Estrazione e correlazione: dai testi analizzati, estrae tutte le relazioni tra gli elementi (es. tutti i testi che parlano di “*territorio*” e “*Unione europea*”).
4. Ricerca soggetto-azione-oggetto: Ad es.se il soggetto è il *governo*, il sistema mostrerà tutti i verbi relativi (*decidere, promulgare* ecc...) e tutti i complementi oggetto (*leggi, cittadini, decreto*).
5. Visualizzazione e integrazione con mappe per rappresentare in modo grafico e intuitivo le relazioni semantiche fra concetti.

In pratica, Cogito, dopo aver fatto l'analisi grammaticale, logica e lessicale del testo, dopo aver organizzato il testo in triplette ed aver fatto la "disambiguazione" tra parole con più di un significato, identifica le tre frasi più importanti del testo, estrae le entità più rilevanti (personaggi, città, luoghi, aziende, date, indirizzi) e, fra esse, costruisce una rete di riferimenti, utilizzando delle azioni come collegamento: ad es. le relazioni "Ken Follett -nato - 5 giugno 1949" e "Ken Follett -nato Cardiff (GB)" sono messe in relazione per cui, se chiedessi, attraverso un'interfaccia di ricerca, -quando è nato Ken Follett? Il sistema mi darebbe subito un'unica risposta "il 5 giugno 1949 a Cardiff (GB)"; non solo, ma mi potrebbe suggerire le correlazioni riguardanti i libri che lo scrittore ha scritto.

Cogito è una tecnologia generale e personalizzabile, che può essere utilizzata per elaborare contenuti, appartenenti a un qualsiasi settore di attività: documentazione tecnica specialistica, documentazione accademica di ricerca, documentazione aziendale, materiale enciclopedico, informazioni di servizi, articoli, messaggistica ecc... Penso che potrebbe essere utile anche per l'indicizzazione, la ricerca, la correlazione e la gestione di documentazione Didattica ed Educativa, per supportare docenti e studenti nello sviluppo e nella condivisione di nuova conoscenza. Infatti, anche in ambito formativo c'è l'esigenza di un reperimento efficiente delle risorse educativo - didattiche, in base agli stili cognitivi e di apprendimento. Uno dei modi per ottimizzare un'accessibilità mirata e pertinente di tali risorse e/o di "Learning Objects" (LO: unità di apprendimento, aggregabili, combinabili e riutilizzabili in diversi percorsi didattici), mediante software intelligenti, è quello di corredarli di una serie di metadati che ne diano una descrizione semantica relativa alle possibili applicazioni (LOM: Learning Objects Metadata). Inoltre, lo sviluppo di ontologie potenzierà la ricerca e la navigazione tra collezioni LOM e di risorse Web, poiché consentirà di navigare attraverso link che collegano i concetti espressi al loro interno e le loro relazioni concettuali, presenti in altri documenti sullo stesso argomento, collegando diverse fonti di informazioni. Il Semantic Web può offrire delle prospettive molto interessanti, non solo per un uso strumentale finalizzato alla ricerca e alla gestione del materiale didattico, ma anche in un contesto più ampio, che coinvolga gli aspetti cognitivi legati alla struttura delle discipline, al curriculum e allo sviluppo delle competenze (ad esempio: mettere in grado gli studenti di ricostruire la mappa reticolare delle varie discipline).

Il Progetto europeo AquaRing¹

È un progetto europeo che riguarda il patrimonio di conoscenza culturale delle scienze marine e acquatiche e che ha realizzato il più grande portale online europeo multilinguistico sulle risorse marine. In AquaRing è stata progettata

¹ Programma eContentplus 2005-2008

un'infrastruttura semantica (basata su vocabolari controllati e ontologie di dominio) e un'ontologia educativa che consente:

1. l'accesso distribuito ai contenuti;
2. l'esplorazione e la navigazione efficace;
3. la ricerca mirata, permettendo percorsi di analisi e di auto-apprendimento diversificati per i diversi profili di utilizzatori.

Nella prima fase è stato avviato un processo di profilazione del target, per identificare le tipologie di potenziali utenti e le loro esigenze informative.

Successivamente, per soddisfare le esigenze di insegnanti, di studenti e di genitori, è stata progettata un'area educativa del portale, nella quale sono state raccolte e rese fruibili risorse didattiche digitali (bibliografie, giochi educativi, animazioni, esercizi, glossari, lezioni, lesson plans, simulazioni, etc), grazie allo sviluppo di un'ontologia educativa specifica che ha integrato gli schemi ontologici che coprono diversi aspetti del dominio di conoscenza degli ambienti acquatici di AquaRing. Lo scopo di tale ontologia è:

1. descrivere ed arricchire il valore pedagogico delle risorse;
2. organizzare e recuperare i materiali in modo efficiente e significativo, attraverso l'uso di un motore di ricerca semantico, che supporta una ricerca orientata all'uso didattico delle risorse.

Ad esempio un docente, cercando informazioni sull'"*Aurelia aurita*," potrebbe restringere la ricerca, specificando che desidera visualizzare solo le risorse destinate agli studenti della formazione primaria, aventi un preciso tipo di obiettivo didattico.

Alla base di tale indicizzazione e di tale ricerca è stato sviluppato un modello ontologico centrato sul concetto di "Resource" e basato su cinque classi principali e relative sotto-classi:

1. il contesto didattico entro cui si intende usare la risorsa didattica;
2. gli obiettivi didattici cognitivi (Knowledge, Comprehension, Application, Analysis, Synthesis, Evaluation);
3. le risorse educative e le loro caratteristiche (livelli di difficoltà, modalità diverse di fruizione, tempi di fruizione, interattività, tipo di oggetto didattico);
4. gli utenti (learner e mediator).

Si possono evidenziare i seguenti aspetti innovativi:

1. tale ontologia potrebbe essere applicata in scenari diversi, poiché non prevede alcuna specifica restrizione sulla tipologia di risorse;
2. l'uso di questa ontologia apre prospettive future di attività di ricerca per supportare servizi educativi diversificati, orientati agli utenti; ad esempio la progettazione di "*digital libraries*"(librerie digitali) per bambini, che adeguino le modalità di accesso ai contenuti alle capacità cognitive e

- motorie dei bambini 3-6 anni (organizzati per macrocategorie di facile comprensione e ricche di grafica);
3. A livello di annotazioni semantiche delle risorse sono necessarie competenze attinenti sia al dominio di conoscenza, sia al dominio educativo oppure è necessaria una consulenza di esperti di pedagogia;
 4. In un'ottica di integrazione europea, le tecnologie semantiche permettono la condivisione di contenuti pertinenti a differenti domini, in un contesto multilinguistico e multiculturale che, in caso contrario, rimarrebbero settorializzate e patrimonio delle singole strutture.

Verso il Web Socio Semantico

Il Web Semantico acquisisce un valore specifico anche come tecnologia innovativa che innesca una trasformazione sociale, poiché ridefinisce l'universo cognitivo dei suoi utilizzatori, distribuendo e facendo condividere non solo informazioni, ma significati.

Si prospetta un'evoluzione verso un Web Socio-Semantico, inteso come un sistema di intelligenze connettive, che integrerà ed utilizzerà le migliori sinergie tra i contenuti e le migliori metodologie tecnologiche per trovarli e riutilizzarli.

Una "conoscenza collettiva e connettiva" (Pierre Lévy. e D.De Kerckove) emergerà solo se questi dati potranno essere interpretati, modificati, aggregati, ricombinati e condivisi, indipendentemente dal luogo in cui si trovano e indipendentemente dal formato di pubblicazione, in modo da creare nuova conoscenza e nuove modalità d'apprendimento.

Bibliografia e sitografia

Adorni G., Coccoli M., Vercelli G., Vivanet G. (2007), *Topic Maps e XTM per l'elearning*. Journal of e-Learning and Knowledge Society (3).

Alvino S., Forcheri P., Ierardi M., Sarti, L. (2008), *Describing learning features of reusable resources: a proposal*. Int. Journal of Social Science , 2 (3), pp. 156-162.

AquaRing Website. URL:

http://www.aquaringweb.eu/http://srvvirt2.softeco.it/aquaring_site/ (verificato in data 20.10.10).

Acquaviva M., Benini M. "VICE: E-Learning nell'era del Semantic Web", 2005.

Centro di Ricerca Informatica Interattiva. Dipartimento di Informatica e Comunicazione, Università degli Studi dell'Insubria.

Berners Lee, T., Hendler, J., & Lassila, O.; "The Semantic Web" in Scientific American 284, 2001, pp. 34-43.

Berners-Lee, Heath and Bizer, “*Linked Data: The Story so Far*”, “International Journal on Semantic Web and Information Systems”, 2009.

Berners-Lee, “*The Next Web of Open, Linked Data*”, TEDtalk 2009..

Bianchi, S.; Mastrodonato, C.; Vercelli, G.; & Vivanet, G.; “*Use of ontologies to annotate and retrieve educational contents: the AquaRing Approach*” in Journal of e-Learning and Knowledge Society - Vol. 5, n. 1, february 2009, pp 211 - 220

D. C. (1998), “*Dublin Core Metadata Element Set - Version 1.1.*” Retrieved September 1, 2008 from <http://dublincore.org/documents/dces/>

Della Valle E.-Celino I.-Cerizza D. (2008), “*Semantic Web. Modellare e condividere per innovare*” Pearson-Addison Wesley.

Della Valle E.-Celino I.-Cerizza D. (2009), “*Semantic Web. Dai fondamenti alla realizzazione di un’applicazione*” Pearson-Addison Wesley.

González Rodríguez M. (2008) “*Formalised AquaRing Domain Ontologies*”. Aquaring Project: Deliverable number 3.2.

EUN. (2007), “*The EUN Learning Resource Exchange Metadata Application Profile*”. Retrieved September 1, 2008 from <http://fire.eun.org/LRE-AP-3.0.pdf>

Gruber T. (2008), “*Ontology*”. In L. L. (Eds.), Encyclopedia of Database Systems.

Gruber T. (2006). “Where the Social Web Meets the Semantic Web” Keynote presentation at ISWC, The 5th International Semantic Web Conference, November 7, 2006.

Hendler J. (2001), “*Agents and the Semantic Web*”. IEEE Intelligent Systems , 2 (16), pp. 30-37.

IEEE. (2002), “*IEEE Standard for Learning Object Metadata*”. Retrieved September 1, 2008 from IEEE Learning Technology Standards Committee: <http://ltsc.ieee.org/wg12/>

Lassila O., Swick R.R., “*Resource Description Framework (RDF) Model and Syntax Specification*”, W3C Recommendation, 1999.

Semanticamente-Il Sole 24ore-nova100-blog di M.Varone

<http://marcovarone.nova100.ilsole24ore.com/>
(verificato in data 10.11.10).

Progetto SEALS: “Semantic evaluation at large scale”, URL:<http://www.seals-project.eu/> (verificato in data 20.10.10).

Progetto SevenPro URL: <http://www.sevenpro.org/>
http://ec.europa.eu/research/infocentre/article_en.cfm?id=/research/headlines/news/article_09_05_14_en.html&item=Infocentre&artid=11273
<http://cordis.europa.eu/ictresults/index.cfm> (verificato in data 20.10.10).

Noy N., McGuinness D. (2001), “*Ontology Development 101: A Guide to Creating Your First Ontology.*” Retrieved September 1, 2008 from: <http://www-ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html>

Signore O. (2008), “*Il supporto delle ontologie nella ricerca dell'informazione*” URL: <http://www.w3c.it/papers/comTec2008.pdf>

Signore O. (2008), “*Introduzione al Semantic Web*” URL: <http://www.w3c.it/papers/wsb08.pdf>

Smith M.K., Welty C., McGuinness D.L.,” *OWL Web Ontology Language Guide*”, W3C Recommendation, 2004.

Social SemanticWeb: http://en.wikipedia.org/wiki/Social_Semantic_Web

Times Online (16.03.2008) Intervista integrale “*Web 3.0: il futuro di Internet secondo Tim Berners-Lee.*”
<http://www.timesplus.co.uk/tto/news/?login=false&url=http://www.thetimes.co.uk/tto/technology/>
<http://www.webmasterpoint.org/speciale/2008mar16-futuro-del-web-secondo-il-suo-creatore.html> (verificato in data 25.10.10).

TED Talks, 2009 (TED, Ideas worth spreading): traduzione integrale “Tim Berners-Lee on the next Web: Raw Data Now”!
http://www.ted.com/index.php/talks/tim_berners_lee_on_the_next_web.html
(verificato in data 25.10.10).

Torrigiani C., Valettini B. (2007),” *AquaRing User Segments, Profiles and Needs*”. AquaRing Project: Deliverable number D2.1.