

Verso la biblioteca digitale

Giorgio Poletti

CARID –Università degli Studi di Ferrara

"Il Web è stato progettato come spazio delle informazioni, con l'obiettivo di essere utile non solo per la comunicazione umano-umano, ma anche con l'intento che le macchine possano parteciparvi ed aiutare. Uno degli ostacoli principali a questo è che la maggior parte delle informazioni sul Web sono progettate per essere consultate da umani ed anche se derivassero da una base di dati con i significati bene definiti (almeno in alcuni termini) per ciascun campo, la struttura dei dati non è evidente ad un robot che passa in rassegna il Web. L'approccio al Web semantico è lo sviluppo di lingue per esprimere le informazioni in forma processabile dalla macchina. Il Web semantico è il Web dei dati, per alcuni versi come una base di dati globale."

Tim Berners-Lee

Il concetto di biblioteca digitale, nel contesto che qui si intende analizzare, è da correlare strettamente al concetto di metadato.

Il concetto di metadato e l'elaborazione del concetto di "supermetadato"¹ porta anche a riflettere su una nuova accezione del concetto di Digital Preservation che allarga i suoi ambiti di ricerca da un tecnicismo stretto ad un'implicazione di strutturazione dell'informazione in funzione dell'accessibilità del documento stesso, rielaborando in chiave digitale e di standard il concetto di consultazione.

La digitalizzazione di documenti cartacei o comunque analogici è sicuramente un importante filone di ricerca nell'ottica di non perdere informazioni, e quindi cultura, per l'estrema caducità dei supporti documentali (basti pensare alle pergamene).

Tale ricerca per essere efficace va supportata da un filone di ricerca parallelo che si occupi di una preservazione dei documenti in un'ottica di struttura, cioè indipendentemente dagli strumenti che dovranno accedere ai documenti.

Queste linee di lavoro hanno avuto inizio dalla consapevolezza che un processo di conservazione è significativo nella misura in cui determini un meccanismo di estrazione dei dati in relazione al loro scopo di utilizzo, per creare sistemi di management di metadati flessibili finalizzati alla determinazione di standard per la conservazione e l'accesso a lungo termine ai documenti.

Le biblioteche sono sicuramente il primo e più immediato campo applicativo di queste linee di pensiero, la dove il sapere è archiviato, la dove il valore aggiunto non è la mole di sapere è la strutturazione del sapere che vi è conservata.

In questo contesto è interessante affrontare il concetto di ontologia in considerazione del fatto che ontologia e web semantico stanno gradualmente diventando l'obiettivo degli studi di chi si occupa della condivisione dei documenti.

¹ Il concetto di "supermetadato" è da mettere in relazione con la definizione di **Digital Data Object (DDO)**, un oggetto che include le informazioni relative allo schema che permette di leggere e comprendere i metadati, gestire le informazioni relative all'oggetto non comprese nei metadati (ad esempio versione dell'oggetto) e poter erborare e strutturare informazioni provenienti da documenti diversi per contenuto, struttura e tipo di fonte.

Un linguaggio specifico si occupa di affrontare queste tematiche, OWL (Web Ontology Language) divenuto meno di un anno fa "Candidate Recommendation" del W3C, cioè un invito esplicito per le implementazioni, indicando che la specifica è stabile e adatta all'implementazione.

OWL è un linguaggio che permette di consolidare le radici del Web semantico, "OWL rappresenta un importante passo avanti per rendere i dati sul Web maggiormente elaborabili dalla macchina e riusabili tra applicazioni" (Tim Berners-Lee).

In pratica si possiede un linguaggio in grado mettere a disposizione un linguaggio per definire ontologie strutturate basate sul Web, strumenti in grado di progettare ed implementare strutture in grado di garantire un'integrazione ed un'interoperabilità di dati tra comunità che descrivono il loro dominio di conoscenza.

Il Web Semantico permetterà

- ricerche efficaci, cercheremo per concetto e non più per termine; oggi quando in un motore di ricerca inseriamo un termine in realtà pensiamo di trovare documenti correlati al concetto che noi associamo a quel termine. Se cerchiamo il termine "piano" magari intendiamo lo strumento musicale ma il WEB ci indicherà anche i siti che trattano di geometria;
- semplice condivisione dei dati che potranno rimanere aggregati mantenendo la loro caratteristica di informazioni in modo da poter correlare i documenti e i contenuti per generare nuovi documenti e nuovi contenuti, operare nella direzione del knowledge management.

La biblioteca digitale il Web semantico sono realtà simili concretizzazioni diverse di un'unica idea, quasi a dire con Calvino, immaginando la biblioteca una città le cui meraviglie sono i libri:

"D'una città non godi le sette o settantasette meraviglie, ma la risposta che dà ad una tua domanda." (Italo Calvino).