



@domani/search

Sempre più di frequente in Internet si discute di information overload, un eccesso di informazione in cui gli utenti spesso si perdono. Nell'ormai lontano 1994, in Rete vi erano più di **5.000 Gopher** server e circa **3.500 siti World Wide Web**; oggi i Gopher sono scomparsi, ma il numero dei server Www non è più valutabile, e le pagine Web sono ormai più di **10 miliardi**. Questa complessa struttura consente all'utente di passare da una risorsa all'altra senza bisogno di ricordare indirizzi e comandi, ma lo costringe anche a muoversi in un ambiente dove la moltiplicazione di uno stesso accesso ad una stessa risorsa è la regola, mentre i menu dei World Wide Web non sempre sono abbastanza descrittivi.

Numerosi documenti, singoli testi o veri e propri archivi indicizzati, sono stati digitalizzati e sono disponibili in Internet. Molti di questi, prima o dopo la loro disponibilità in Rete, vengono anche stampati su carta. Altri no.

È inevitabile, da anni, parlare di una vera e propria catalogazione delle risorse Internet, dal momento che la rete Internet, nel suo complesso, può essere paragonata ad una sola, enorme biblioteca virtuale che invece di contenere soltanto libri e periodici in formato digitale mette a disposizione anche oggetti molto diversi tra loro e non tutti di tipo strettamente documentario (si pensi ad Opac, banche dati, strumenti di ricerca, archivi di immagini, file audio, filmati, software, gruppi di news, liste di discussione, persone raggiungibili attraverso indirizzi di posta elettronica, e via dicendo). Una biblioteca dove una stessa risorsa è spesso presente in numerosissime copie, alcune più vicine e di più facile accesso, altre più lontane.

Il Cern di Ginevra, per primo, ed in particolare Tim Berners-Lee nell'ambito del progetto World Wide Web, ha di fatto imposto come standard l'oggetto, chiamato **Url (Uniform resource locator)**, in grado di puntare in modo univoco qualsiasi risorsa Internet.

Oltre all'**Url**, che in una biblioteca potrebbe essere paragonato alla collocazione, è da tempo in via di definizione un altro oggetto, l'**Urn (Uniform resource name)**, finora però privo di applicazioni concrete. Urn è il nome univoco che un'autorità centrale dovrebbe assegnare ad ogni risorsa Internet, esattamente come un Isbn viene assegnato ad una pubblicazione. L'Urn quindi è del tutto indipendente dalla collocazione della risorsa stessa. Un determinato file, ad esempio, avrebbe un solo Urn che lo identifica, ma a quel solo Urn potranno corrispondere più Url, uno per ciascuno dei server che lo contengono.



@domani/search

Dato che l'Urn (il corrispondente dell'Isbn) e l'Url (il corrispondente della collocazione) non comprendono la descrizione della risorsa stessa, sarà necessario definire anche un **Urc (Uniform resource characteristics)**, un oggetto costituito da un insieme di meta informazioni sulla risorsa e che potrebbe corrispondere alla descrizione bibliografica.

Se una risorsa si sposta o cambia indirizzo, infine, potrebbe risultare molto comodo l'uso di un **Purl**, ovvero un Persistent Url che resti immutato anche quando una risorsa si sposta nel cyberspace, grazie alla «triangolazione» garantita (per ora solo in modo sperimentale e limitatissimo) da agenzie incaricate di tenere sotto controllo queste frequenti migrazioni, rendendole irrilevanti per gli utenti.

In futuro, grazie ad un accurato lavoro di catalogazione e a una maggiore attenzione all'interoperabilità fra sistemi informativi eterogenei, gli strumenti per la ricerca delle informazioni in Rete potrebbero evolversi ed operare con meccanismi molto più sofisticati di quelli esistenti oggi. Ma il condizionale è sempre più obbligatorio e la realtà di oggi, tutta tesa verso il business e la ricerca di un profitto difficile da ottenere, non pensa a fornire agli utenti dei validi servizi di catalogazione semantica delle risorse Internet; al contrario, lavora con motori di ricerca per parola che sempre più spesso vengono accusati non solo di essere poco efficienti e di vedere soltanto una parte della Rete, ma anche e soprattutto di presentare i risultati delle ricerche in base al denaro che ricevono dai loro sponsor. Chi paga è reperibile, insomma, mentre le pagine Web di chi non paga il motore, per quanto interessanti, restano invisibili.

Un progetto che suscitava molte aspettative e che tutt'ora è molto discusso, sebbene non ancora applicato su scala massiccia, è il **Dublin core metadata**, nato nel 1995. L'obiettivo del progetto Dublin core <<http://dublincore.org>> è la definizione di uno standard che permetta la descrizione di ogni risorsa disponibile in Rete da parte degli autori stessi, in modo sia da garantire un livello minimo di autocatalogazione per ciascun documento sia di costituire la premessa necessaria per realizzare eventuali indicizzazioni successive più sofisticate, con sistemi automatici o con catalogatori professionisti.

Lo standard del Dublin core individua alcuni elementi descrittivi del documento, di tipo bibliografico (titolo, autore, soggetto, parole chiave, abstract, editore, data, lingua, e via dicendo); la traduzione italiana degli elementi essenziali è disponibile sul Web dell'Associazione italiana biblioteche, <<http://www.aib.it/aib/lis/std/t9503.htm>>. Ogni elemento del Dc può essere incluso nel file Html della risorsa catalogata nella parte dedicata all'header, invisibile ai browser ma utilizzata dai principali motori di ricerca, con il tag <Meta>, già compreso nello standard Html.



@domani/search

Occorre tenere presente anche che alcuni degli standard già approvati o in corso di elaborazione da parte del W3c, il consorzio diretto dall'inventore del Www, Tim Berners-Lee, sono destinati ad influire non poco sulle future tecniche di catalogazione delle risorse Internet.

Il **Resource description framework (Rdf)**, in particolare, un lavoro ormai diventato raccomandazione ufficiale del consorzio W3C <<http://www.w3c.org/RDF>> stabilisce il modo di descrivere qualsiasi tipo di documento disponibile in Internet mediante Xml, in modo da potervi incorporare diversi tipi di standard descrittivi, tra i quali anche quelli definiti dal Dublin core. Tra il 2003 e il 2004 è stato approvato anche un altro importante standard legato alla costruzione del Web semantico (su cui si tornerà fra poco), il **Web ontology language (Owl)** <<http://www.w3c.org/2004/OWL>>, cioè il linguaggio da utilizzare per scrivere i dizionari (o ontologie, o reti di conoscenza).

### Testo adattato da:

Metitieri, F. Ridi, F, *Biblioteche in rete. Istruzioni per l'uso*. Laterza, 2005  
([www.laterza.it/bibliotecheinrete/index.htm](http://www.laterza.it/bibliotecheinrete/index.htm))