

## Tavola rotonda sul Web semantico: Sintesi e proposte Prato 26 gennaio 2005

### *Grafi e classificazione*

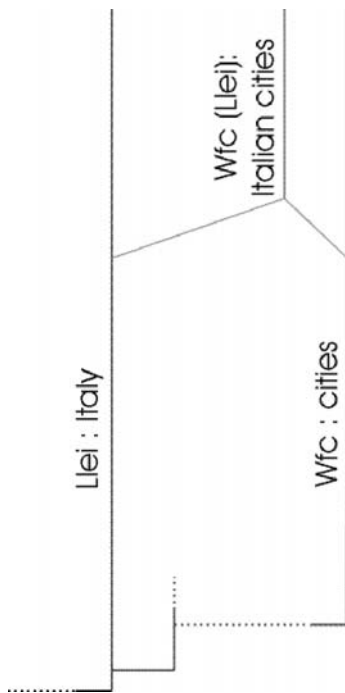
Claudio Gnoli e Viviana Doldi

(Università di Pavia. Dipartimento di Matematica)

Un recente intervento di Luca Rosati [1] osserva come il modello a grafo utilizzato nell'ambito del Web semantico, in particolare dall'iniziativa italiana Platypus Wiki, presenti interessanti analogie con l'analisi a faccette, che seppur in forma semplificata sta riscuotendo l'interesse di molti architetti dell'informazione e webmaster. La caratteristica di entrambi i modelli, infatti, è che "people can use different paths to reach the same node" [2]. In contrasto, gli schemi esclusivamente gerarchico-enumerativi, come la classificazione Dewey e molte "tassonomie" sul Web, impongono all'utente di adeguarsi più rigidamente ad essi, obbligandolo a risalire e ridiscendere l'albero classificatorio per spostarsi in rami differenti.

In verità, anche gli schemi enumerativi offrono qualche espediente per spostamenti trasversali, sottoforma di rinvii di tipo "vedi anche", di "related terms", di note d'ambito del tipo "classificare X sotto la classe N", eccetera. Nella nota tassonomia di Yahoo questo tipo di rinvii è contraddistinto dal simbolo "@", e grazie all'ambiente ipertestuale permette di saltare in modo molto rapido a un ramo semanticamente affine ma strutturalmente lontano. Risalendo la storia dell'organizzazione delle conoscenze potremmo trovare molti esempi di questa parziale ipertestualità (nella lingua filosofica di John Wilkins, nell'Encyclopédie di Diderot e d'Alembert, ecc.). Il problema di fondo è sempre lo stesso: ridurre a una sequenza lineare, per gli scopi del browsing, una rete di relazioni a molte dimensioni, molto più complessa ma più aderente alla realtà del mondo.

L'idea di modellizzare queste relazioni in forma di grafo è molto interessante, anche perché prospetta la possibilità di trattarle matematicamente, e quindi automaticamente. I *grafi* sono oggetti matematici [3] definiti a un livello della matematica assai fondamentale, vicino alla logica e appartenente alla combinatoria (e qui torna alla mente Leibniz...). In particolare, gli *alberi* sono grafi connessi, cioè in cui nessun nodo è isolato dagli altri, e aciclici. I grafi che rappresentano schemi di classificazione sono *orientati*, ossia hanno una radice e una direzione di riferimento (in quanto le relazioni tra le classi più generali e quelle più specifiche sono direzionali).



Se a qualche nodo del grafo si può arrivare per due (o più) percorsi diversi, come nel modello del Platypus Wiki e nelle architetture a faccette, si parla di un grafo con *assorbimento*. Così, da un albero gerarchico ci avviciniamo a un modello a rete, che conserva però delle direzioni preferenziali. Spostarsi attraverso una rete permette maggiore libertà di movimento, come abbiamo detto; nel contempo però pone dei problemi alla linearizzazione delle classi finalizzata al browsing. Supponiamo che il concetto di città sia rappresentato nel nostro schema dal simbolo *Wfc*, e quello di Italia da *Llei*. Come possiamo rappresentare il soggetto “città italiane”? Un’ovvia possibilità è quella di combinare i due simboli. Ma in che ordine? Se adottiamo l’ordine *Wfc Llei*, nella successione linearizzata (il menù di un sito o gli scaffali di una biblioteca) il documento si troverà insieme ad altri documenti sulle città, se invece scegliamo *Llei Wfc* il documento finirà vicino ad altri sull’Italia. Per gestire questo problema occorre stabilire delle regole per l’*ordine di citazione*.

La matematica offre strumenti per trattare formalmente queste situazioni. La *teoria dei codici* tratta l’attribuzione di un codice a ciascun nodo di un grafo: il codice non sarà altro che il simbolo nella notazione del nostro schema di classificazione, che ci servirà per stabilire l’ordine di *successione conveniente* in cui disporre le classi (automaticamente, in ambiente digitale). Un nodo in cui si verifichi assorbimento può essere tradotto, mediante l’algoritmo di Huffman, in due nodi ciascuno corrispondente a uno dei due percorsi possibili. Si sceglierà poi una regola per privilegiare l’uno o l’altro percorso, e quindi la notazione corrispondente. Una possibile regola è “citare per prima la classe che viene per ultima nello schema”; questa scelta è in accordo con il *principio di inversione* adottato nelle classificazioni a faccette: le classi che compaiono per ultime nelle tavole devono invece comparire per prime nella rappresentazione di un singolo soggetto composto.

Questi concetti possono essere applicati anche a uno schema basato sui *livelli di integrazione*, nel quale cioè le classi siano elencate secondo livelli di complessità crescente: i fenomeni atomici, molecolari, chimici ecc. nelle prime classi, quelli

biologici in classi intermedie, quelli tecnologici e culturali nelle ultime classi [4]. Ad esempio le città sono definite al livello tecnologico, per cui il loro simbolo comincia per W, mentre l'Italia è definita al livello inferiore del territorio geografico, per cui il suo simbolo comincia per L; nella classe composta quindi  $Wf_C$  dovrà precedere  $Llei$ . Un sistema di questo tipo, abbozzato dal Classification Research Group negli anni Sessanta, fu definito da Derek Austin *freely faceted classification*, e appare particolarmente idoneo all'indicizzazione in ambiente digitale.

L'applicazione della teoria dei codici alla classificazione per livelli di integrazione presenta ancora molti problemi aperti, ai quali speriamo di dedicarci nel prossimo futuro. Ad esempio: non è realistico rappresentare tutti i concetti per composizione di concetti più primitivi, in quanto di questo passo si otterrebbero simboli infinitamente lunghi; ma allora, quando è giusto essere analitici (come in "città italiane") e quando invece sintetici (come in "acqua", composto di "idrogeno" e "ossigeno" ma a un livello di integrazione superiore)? Inoltre, come rappresentare fenomeni quali le rocce sedimentarie, che pur essendo a un livello inorganico derivano dalla decomposizione di organismi, ossia di un livello più elevato? Il recente concetto di *profondità logica* può fornire un criterio per individuare i livelli secondo cui disporre i fenomeni?...

## Riferimenti

- 1: Le faccette, il web semantico e l'ornitorinco (Trovabile) – [http://www.trovabile.org/faccette\\_web\\_semantico\\_ornitorinco](http://www.trovabile.org/faccette_web_semantico_ornitorinco)
- 2: Platypus wiki: a semantic wiki wiki web / Stefano Emilio Campanini, Paolo Castagna, Roberto Tazzoli (SWAP: Semantic Web applications and perspectives : 1st Italian Semantic Web workshop : Ancona : 10th December 2004) – <http://semanticweb.deit.univpm.it/swap2004/cameraready/castagna.pdf>
- 3: Graph theory / Frank Harary – Addison-Wesley : Reading (Mass') : 1969
- 4: Integrative level classification (ISKO Italia. Documenti) – <http://www.iskoi.org/ilc> : 2004-

Pavia, 24 gennaio 2005