

Le mappe topiche Come imparai a non preoccuparmi e ad amare i metadati

di Federico Meschini

Introduzione

L'annuale conferenza "unificata" dell'Association for Literary and Linguistic Computing (ALLC)¹ e dell'Association for Computers and the Humanities (ACH)², denominata comunemente ALLC/ACH, è uno degli appuntamenti più importanti nel settore dell'applicazione delle nuove tecnologie alle scienze umane, disciplina in Italia ormai definita informatica umanistica³ e nei paesi anglosassoni da tempo nota come *humanities computing*.

Organizzata alternativamente negli Stati Uniti e in Europa, l'edizione del 2004 è stata ospitata dall'Università di Göteborg in Svezia⁴, mentre l'ALLC/ACH 2005 avrà luogo in Canada a Victoria, nel British Columbia⁵. Da un punto di vista sia qualitativo sia quantitativo, il livello degli interventi è molto elevato e l'evento risulta estremamente utile per avere una visione generale delle linee di ricerca attualmente in corso nelle varie aree disciplinari, dalla linguistica computazionale alle biblioteche digitali, dalle applicazioni dei linguaggi di marcatura all'analisi testuale, dai libri elettronici fino all'influenza delle teorie narratologiche nello sviluppo dei videogiochi⁶.

All'interno della comunità che si occupa di informatica umanistica un ruolo sempre più rilevante, sia per il peso scientifico sia per il numero di persone coinvolte, è svolto dalla Text Encoding Initiative⁷ (TEI), un'associazione internazionale e inter-

FEDERICO MESCHINI, Università della Tuscia, Facoltà di lingue e letterature straniere moderne, largo dell'Università, 01100 Viterbo, e-mail fmeschini@tin.it.

L'ultima consultazione di tutti i siti Web citati risale al 20 marzo 2005.

1 ALLC: *Association for Literary and Linguistic Computing*, <<http://www.allc.org>>.

2 ACHWeb, <<http://www.ach.org>>.

3 Per una definizione di informatica umanistica si veda Gino Roncaglia, *Informatica umanistica: le ragioni di una disciplina*, «Intersezioni» 23 (2002), n. 3, p. 353-376. L'articolo è disponibile online all'indirizzo <http://www.merzweb.com/testi/saggi/informatica_umanistica.htm>.

4 ALLC/ACH 2004, <<http://www.hum.gu.se/allcach2004>>.

5 ALLC/ACH conference 2005, <<http://web.uvic.ca/hrd/achallc2005>>.

6 ALLC/ACH 2004, sessione plenaria di chiusura: Espen Aarseth, *Old, new, borrowed: blue? What can the humanities contribute to the games field?*

disciplinare attiva dal 1987. L'ALLC/ACH stessa contiene al suo interno una sorta di "convegno parallelo" della TEI, un incontro totalmente informale ma che risulta un'ottima occasione per scambiare idee, confrontare progetti e opinioni e fare in un certo senso il punto della situazione⁸.

Lo scopo principale della TEI è la definizione di uno standard per la codifica elettronica dei testi letterari o, come vengono definite nel sito Internet ufficiale, delle «Guidelines for electronic text encoding and interchange». Tali linee guida sono espresse tramite un linguaggio di marcatura aderente nelle prime versioni alle specifiche SGML e dal 2002, come naturale evoluzione, a XML, con l'edizione P4⁹.

La codifica TEI è ormai uno standard sempre più utilizzato nei progetti di digitalizzazione *full text*, non solo per la validità in sé delle specifiche e delle concrete possibilità di utilizzo¹⁰, ma anche per la vitalità e disponibilità dei partecipanti, che tra gruppi di lavoro, gruppi d'interesse¹¹ e *mailing list* offrono un valido ed efficiente supporto a chiunque voglia adottare questo linguaggio per i propri scopi. Chi ha avuto esperienze simili sa benissimo che la marcatura di un testo è solo il primo dei tanti passi che portano alla creazione di una biblioteca digitale. L'aspetto della visualizzazione (pubblicazione) di un testo, per esempio, è attualmente di primaria importanza. Prima dell'avvento del Web nel campo delle risorse testuali elettroniche l'enfasi era posta sull'analisi e la pubblicazione era vista come un fattore secondario.

L'avvento del WWW ha cambiato tutto, portando al primo posto la visualizzazione, seguita dalle possibilità di ricerca generali e infine dalle analisi testuali specifiche, come concordanze ed estrazione di *token*. Il fatto che il linguaggio TEI sia attualmente un "dialetto XML" rende più facilmente risolvibili tutti questi aspetti, perché è possibile utilizzare diversi programmi *open source* in grado di manipolare questo particolare formato. Quindi tra adattamenti specifici di software nati in contesti più propriamente "informatici" e programmi sviluppati direttamente in ambito accademico e istituzionale¹² le possibilità sono decisamente più numerose di quanto non fossero solo qualche anno fa, quando la scelta obbligata era l'utilizzo di applicazioni sì efficienti, ma proprietarie, complesse e costose come Dynatext/Dynaweb.

Naturalmente in questo campo, così come in altri settori legati al trattamento delle informazioni, appena un problema è non risolto ma in via di risoluzione, si pongono subito altre questioni. Una delle più sentite riguarda il rapporto tra il testo elettronico e

7 TEI: *Text Encoding Initiative*, <<http://www.tei-c.org>>.

8 Il *meeting* ufficiale si svolge invece annualmente in autunno: <<http://www.tei-c.org/Publicity/baltimore.html>>.

9 *TEI P4: guidelines for electronic text encoding and interchange, XML compatible edition*, edited by Michael Sperberg-McQueen and Lou Burnard 2004. Si veda <<http://www.tei-c.org/P4X>>.

10 Gli abstract inviati dai vari partecipanti all'edizione del 2005 del convegno ALLC/ACH vengono "marcati" in TEI P4 così da ottenere dallo stesso documento sia il PDF per l'edizione a stampa sia l'HTML per il sito Web, tutto in un unico flusso di lavoro grazie alla tecnologia dei fogli di stile di trasformazione e di formattazione, XSLT e XSLFO.

11 Definiti come *working group* e *special interest group*.

12 Per un elenco di partenza si veda la *TEI/CMS tool list* all'indirizzo <http://miro.acs.its.nyu.edu/tei_cms/show.php>.

le informazioni che lo descrivono, i cosiddetti metadati. Questo discorso è estendibile a qualsiasi oggetto fatto di bit e non solo di natura testuale, anzi nel caso di materiale multimediale una dimensione testuale è ancora più importante per consentire il dialogo tra le applicazioni e i computer, per le funzioni di organizzazione, catalogazione, condivisione, ricerca e reperimento. Con l'aumentare del numero di risorse digitali disponibili, di qualsiasi tipo esse siano, non solo culturali o testuali, aumenta l'importanza dei metadati associati. L'importanza è anche il ruolo, non più statico e di secondo piano, ma fondamentale e dinamico. Cambiando il ruolo, cambia di conseguenza anche il tipo. Se per attività "tradizionali", come la classificazione o la ricerca per campi, è più che adeguato il concetto di "metainformazione" cui siamo abituati - al limite aggiornato con dati specifici al contesto digitale¹³, un tipo di informazione "gestionale-amministrativa-descrittiva-strutturale"¹⁴ - si avverte adesso anche l'esigenza di un altro concetto di metadato, più legato alla sostanza che alla forma dell'oggetto: un metadato semantico, relativo ai contenuti, che permetta la creazione di reti di conoscenza, di collegamenti basati sul significato tra i vari oggetti digitali, e che possa trascendere l'appartenenza "fisica" di una risorsa, in modo da potersi estendere virtualmente su tutta la rete.

Il motivo di tale esigenza è espresso perfettamente da Jean-François Lyotard in *La condizione postmoderna*. Lyotard è dell'opinione che nei giochi a informazione incompleta, come le società precedenti la nostra, i vantaggi competitivi si ottengano tramite l'acquisizione di un nuovo supplemento d'informazioni. Al contrario, nei giochi a informazione completa o con un quantitativo d'informazione sufficiente, come la società contemporanea, una migliore performance si ottiene per lo più attraverso il collegamento di una serie di dati ritenuti fino a un certo momento indipendenti¹⁵.

Il problema è quale tecnologia possa portare questi vantaggi competitivi alle ormai numerose collezioni di testi digitali: occorre qualcosa che sia comprensibile dalla macchina oltre che leggibile dall'uomo e perfettamente compatibile in prima istanza con le specifiche della TEI ed eventualmente con altri tipi di formati, sia testuali sia multimediali. La codifica TEI, che segue la struttura ad albero tipica dei linguaggi basati su XML, prevede due sottosezioni principali: l'intestazione, contenente i metadati bibliografici, amministrativi e gestionali del documento, e il corpo, con il contenuto vero e proprio. Tali sezioni sono rappresentate rispettivamente dagli elementi <teiHeader>¹⁶ e <text>. Nella mia esperienza ho notato che chi proviene da un corso di studi incentrato sulle materie letterarie in fase di codifica tende a sottovalutare il <teiHeader>, compilandolo quel tanto che basta e concentrandosi piuttosto sul contenuto del documento, rischiando così di tralasciare informazioni importanti¹⁷. D'altro canto queste stesse persone sono estremamente interessate a un concetto di metadato semantico, il quale però per motivi sia tecnici sia di omogeneità strutturale non può collocarsi all'interno dell'intestazione né del contenuto.

13 Come per esempio il tipo di file, il quantitativo di memoria occupata, l'applicazione con cui è stato generato e quelle con cui può essere utilizzato.

14 E a questo scopo rispondono alla perfezione standard come METS <<http://www.loc.gov/standards/mets/mets-home.html>>, MODS <<http://www.loc.gov/standards/mods/>>, o l'italiano MAG <<http://www.iccu.sbn.it/schemag.htm>>.

15 Jean-François Lyotard, *La condizione postmoderna*, Milano: Feltrinelli, 1981, p. 95.

16 Per una descrizione dell'elemento <teiHeader> e del suo contenuto si veda <<http://www.tei-c.org/P4X/HD.html>>.

17 E questo sottolinea ulteriormente l'importanza del ruolo dei bibliotecari nei progetti di codifica e digitalizzazione.

Questo era uno degli argomenti più attuali e discussi all'ALLC/ACH 2004, non solo durante gli interventi ufficiali ma anche e soprattutto nelle chiacchierate informali che hanno luogo durante le pause tra una presentazione e l'altra e nei *break* per il pranzo. Chiunque abbia un minimo di esperienza di convegni sa quanto questi scambi informali d'informazioni siano importanti per la collaborazione e la ricerca scientifica. In queste discussioni si finiva sempre per parlare delle tecnologie collegate al Web semantico, con le immancabili citazioni dei vari RDF¹⁸, DAML¹⁹ e OIL²⁰, ma a essere nominata con maggiore frequenza e interesse non era una sigla astrusa, bensì un'espressione suggestiva e facile da ricordare: *topic map*, letteralmente traducibile come mappa di argomenti²¹.

Mappe di argomenti e territori d'informazione

Ma cosa sono queste mappe di argomenti, perché suscitano tanto interesse in ambienti non tradizionalmente interessati ai metadati e cosa offrono di nuovo rispetto alle altre specifiche? Volendo dare una definizione più concisa possibile, le *topic map* forniscono funzionalità combinate di indici, glossari e thesauri, creando così potenti meccanismi di navigazione tra vaste collezioni di risorse digitali interconnesse tra loro, laddove questo tipo di interconnessione non debba essere necessariamente fisica bensì solamente concettuale. Ciò grazie al salto di astrazione che viene compiuto: queste mappe non si pongono allo stesso livello del documento o della risorsa, ma sono sovrapponibili, si posizionano a un livello superiore²² e formano un comune sovrastato semantico agli oggetti cui fanno riferimento e che vengono "mappati"²³. In questo modo più mappe possono essere applicate alle stesse informazioni oppure la stessa mappa può essere applicata a gruppi di informazioni diverse, permettendo un elevato livello di flessibilità e personalizzazione. La struttura proposta è di tipo reticolare e a più livelli, utilizzando uno schema che si presta molto di più al sistema della ricerca scientifica e alle modalità di organizzazione del pensiero²⁴, superando i limiti delle strutture lineari e ad albero imposte dai formati di memorizzazione dei supporti informatici.

Una prima distinzione che va fatta è tra il modello concettuale di questo paradigma e la sua concreta implementazione, essendo il primo qualcosa di più stabile, mentre il secondo è più influenzato dal susseguirsi delle tecnologie²⁵.

18 *Resource Description Framework*, <<http://www.w3.org/RDF/>>.

19 *DARPA Agent Markup Language homepage*, <<http://www.daml.org/>>.

20 *Ontology Inference Layer*, <<http://www.ontoknowledge.org/oil/>>.

21 Il termine greco *topos* significa sia "luogo" sia "argomento".

22 Questo "salto di livello" risolve i problemi legati alla struttura e alla compatibilità con i vari formati.

23 Mentre i database relazionali catturano le relazioni tra gli oggetti informativi, spesso inglobandoli al loro interno, le *topic map* creano dei collegamenti tra questi oggetti senza alterare la loro posizione. Un articolo interessante che confronta i due modelli è Alexander Johannesen, *Here is a how to topic maps, Sir!*, <<http://shelter.nu/art-007.html>>.

24 Sono notevoli le somiglianze con modelli nati in altri ambienti legati per esempio alle scienze cognitive, come le *mind map*, le "mappe mentali", teorizzate dallo psicologo Tony Buzan, <<http://www.mind-map.com/>>.

25 *HyTime*, <<http://www.hytime.org/>>, la sintassi precedentemente usata per descrivere il modello delle *topic map*, utilizzava SGML, mentre XTM (*XML Topic Maps*, <<http://www.topicmaps.org/xtm/1.0/>>), quella attualmente più diffusa, segue le regole di XML. Esiste inoltre un altro tipo di notazione, LTM (*Linear Topic Map notation*, <<http://www.ontopia.net/download/ltn.html>>), sempre basata su un formato testuale ma che non utilizza la sintassi dei linguaggi di marcatura.

L'ormai classica introduzione a questo modello è l'articolo *The TAO of topic maps* di Steve Pepper²⁶. Il gioco di parole nel titolo è contemporaneamente un riferimento al termine cinese "tao"²⁷, traducibile con via, sentiero, e l'acronimo formato da *topic*, *association* e *occurrence*, definiti come i concetti di base principali²⁸. Pepper definisce in maniera originale le mappe topiche come «the GPS of the information universe», ribadendo l'idea che «[they] are also destined to provide powerful new ways of navigating large and interconnected corpora»²⁹ e sottolineando il problema principale nell'organizzazione delle risorse digitali: la mancanza degli strumenti tradizionali esistenti nel mondo analogico, come indici, glossari e thesauri. Infatti il procedimento effettuato dai motori di ricerca è un'indicizzazione indiscriminata di tutte le parole presenti, senza tener conto per esempio dei problemi relativi alla presenza di sinonimi (più termini per lo stesso argomento) o di omonimie (più argomenti per lo stesso termine), restituendo numerosi risultati, tra i quali probabilmente, ma non è detto, ci sarà l'oggetto della nostra ricerca. Lo scopo principale delle *topic map* è di avviare a questa limitazione proponendo un altro approccio che trae spunto, tra le varie discipline, dalle scienze dell'informazione e dalla *knowledge representation*.

Tra i vari metodi utilizzati nelle risorse tradizionali, l'indice in particolare può essere considerato come una guida alle informazioni contenute in un documento. Un esempio³⁰ può aiutarci a capirne le caratteristiche fondamentali:

BIBLIOFILIA, III, 87 c.
 BIBLIOGRAFIA, III, 88 c.
 Biblioláthas. Vedi Didimo.
 BIBLIOTECA, III, 89 b; * 88-9 (tav.).
 Architettura, III, 91 b; * 88-9 (tav.); * 90.
 Biblioteconomia, III, * 90 c; * 88-9.
 BIBLO (o Biblos, Biblus). III, * 92 b; I, 411 c; 413 c; VII,
 677 d; XI, * 248-249; XIV, 106 c; XVI, 686 d; IV F.

In questa breve porzione di un indice sono elencati in ordine alfabetico i nomi di alcuni argomenti (*topic*) e per ognuno di essi una o più indicazioni del luogo in cui l'argomento viene trattato (*occurrence*), utilizzando il numero del volume e delle pagine, insieme a

26 Steve Pepper, *The TAO of topic maps: finding the way in the age of infoglut*, 2002, <<http://www.ontopia.net/topicmaps/materials/tao.html>>. Per un elenco di altri articoli e materiale introduttivo si veda *Topic map articles*, <<http://www.topicmap.com/topicmap/resources.html#introductions>>.

27 Il termine è alla base della filosofia cinese taoista, teorizzata da Lao Tzu nel suo *Libro della via e della virtù*. Edizione in lingua italiana di riferimento Lao Tzu, *Tao te ching*, Milano: Mondadori, 2001. Il testo è disponibile in edizione elettronica sul Web sia in italiano, <http://www.liberliber.it/biblioteca/l/lao_tzu/index.htm>, sia in cinese <<http://www.chinapage.com/laozi724.html>>.

28 Sam Hunting puntualizza che nel *reference model*, il modello concettuale di riferimento, è più corretto parlare di rapporti tra *assertions*, *topic*, *occurrence*, *topic*, *base name*, *base name* e *variant name*, ammettendo però che ATOTBBV ha molto meno potere evocativo di TAO: Sam Hunting, *The rise and rise of topic maps*, in *XML topic maps: creating and using topic maps for the Web*, Jack Park and Sam Hunting editors, Boston: Addison-Wesley, 2002, p. 65.

29 Steve Pepper, *The TAO of topic maps* cit.

30 Tratto da *Grande dizionario enciclopedico UTET*, v. *Indice – Atlanti*, Torino: UTET, 1975, p. 110.

delle convenzioni tipografiche e delle sigle che servono a specificare ulteriormente il tipo di occorrenza³¹, come per esempio la presenza di un'immagine tramite l'utilizzo della sigla "tav.". Un'altra informazione importante è quella relativa ai collegamenti tra due o più *topic* (*association*): alla voce "Biblioláthas" la parola chiave "vedi" indica un collegamento con "Didimo", anche se non vengono specificati né la natura del collegamento né i diversi ruoli che hanno i due argomenti. È prevista anche la presenza di alcune varianti per il nome, vedi "Biblos" e "Biblus" per "BIBLO", e la possibilità di un'organizzazione gerarchica, come nel caso di "Architettura" e "Biblioteconomia" per "BIBLIOTECA"³².

Naturalmente l'indice per nomi è solo una delle tante possibilità, poiché potrebbe essere organizzato anche per soggetti e luoghi. Le componenti fondamentali di un indice sono gli argomenti, identificati da uno o più nomi, le associazioni tra di essi e le occorrenze di un argomento. Ulteriori informazioni possono essere aggiunte specificando la tipologia di ognuno di questi componenti. *Topic*, *association* e *occurrence* sono per Steve Pepper anche le componenti chiave del modello delle mappe di argomenti, a cui aggiunge però anche la possibilità per un argomento di avere una definizione, come nei glossari, e la tipizzazione delle relazioni presente nei thesauri³³.

Il concetto di *topic* è quanto mai generico e viene definito come la rappresentazione all'interno di una *topic map* di una qualsiasi cosa su cui possano essere fatte delle asserzioni. L'unica limitazione è che un *topic* può rappresentare uno e un solo argomento, qualunque esso sia. Data una definizione di questo genere è evidente come sia possibile creare una mappa a partire da qualsiasi tipo di informazione. Consideriamo la seguente frase: «Sul Bollettino AIB è pubblicato *Le mappe topiche* di Federico Meschini». A prima vista gli argomenti (e di conseguenza i *topic*) che si possono individuare sono tre: "Bollettino AIB", "Le mappe topiche" e "Federico Meschini", o meglio questi sono i nomi, le etichette da utilizzare come riferimento. Potrebbero però essere presenti delle varianti, come per esempio "Bollettino AIB: rivista italiana di biblioteconomia e scienze dell'informazione", se prendiamo in considerazione anche il sottotitolo, oppure "Mappe topiche (Le)" e "Meschini Federico" ai fini di un'eventuale indicizzazione alfabetica. Da questo possiamo dedurre che:

a) i *topic* sono identificati tramite i nomi,

b) per ogni *topic* oltre al nome principale di riferimento possono essere presenti una o più varianti.

Questo però non è sufficiente. Il passo successivo da compiere è la "tipizzazione" ossia la classificazione in tipi, categorie generali. Rispettivamente per "Bollettino AIB", "Le mappe topiche" e "Federico Meschini" questi tipi potrebbero essere "rivista", "articolo" e "autore"³⁴, e a loro volta devono essere definiti all'interno di una *topic map* come altrettanti *topic*³⁵. E il procedimento potrebbe continuare: "pubblicazione", "opera intellettuale" e "persona" possono essere le categorie ancora più generali cui appartengono "rivista", "articolo" e "autore"³⁶, e così via. Oltre che ampliarsi verticalmente, la mappa topica può espandersi anche in

31 Utilizzando un'espressione più appropriata e mutuata dall'informatica si potrebbe parlare di tipologia dell'istanza.

32 Sempre utilizzando il gergo informatico in questo caso si parla di rapporto tra classe e proprietà della classe.

33 Organizzate in base alla tipologia del nome (come per esempio verbo, sostantivo e aggettivo) oppure su base semantica (la relazione per contrari).

34 In una relazione di tipo classe-istanza.

35 Potendo così utilizzare questo modello per aggiungere informazioni su di loro.

36 Anche queste naturalmente vanno definite come *topic*.

orizzontale, con l'introduzione di nuovi *topic* che si vanno a posizionare a un livello già esistente. Per esempio "Zeno Tajoli" e "DOI: uno strumento per costruire la biblioteca digitale" si andrebbero rispettivamente ad affiancare a "Federico Meschini" e "Le mappe topiche".

Definiti gli argomenti si passa alle relazioni (*association*). Se i *topic* sono le isole d'informazione che abbiamo individuato nella nostra mappa, allora le *association* sono i ponti che mettono in collegamento queste isole tra di loro. Partendo sempre dall'enunciato «Sul Bollettino AIB sono pubblicati *Le mappe topiche* di Federico Meschini e *DOI: uno strumento per costruire la biblioteca digitale* di Zeno Tajoli», possiamo estrapolare le seguenti affermazioni:

- a) sul Bollettino AIB è pubblicato *Le mappe topiche*³⁷,
- b) sul Bollettino AIB è pubblicato *DOI: uno strumento per costruire la biblioteca digitale*,
- c) *Le mappe topiche* è scritto da Federico Meschini,
- d) Zeno Tajoli ha scritto *DOI: uno strumento per costruire la biblioteca digitale*.

Com'è evidente le *association* non sono generali bensì di un tipo particolare, in questo caso "è pubblicato" e "è scritto da". Anche le categorie in base alle quali sono classificate le relazioni devono essere definite come *topic* a sé, e questo permette di effettuare dei raggruppamenti. Per a) e b) avremo una sola *association* in cui da un lato c'è il *topic* "Bollettino AIB" e dall'altro sia "Le mappe topiche" sia "DOI: uno strumento per costruire la biblioteca digitale", mentre per c) e d) avremo due *association* diverse, in quanto condividono sì lo stesso tipo di relazione, ma non hanno nessun altro *topic* in comune. Questo ci porta a parlare dei diversi ruoli (*association role*) presenti in una relazione. È chiaro come nell'*association* "è pubblicato" i *topic* "Bollettino AIB" e "Le mappe topiche" abbiano due ruoli diversi, mentre "Le mappe topiche" e "DOI: uno strumento per costruire la biblioteca digitale" condividano lo stesso rapporto nei confronti di "Bollettino AIB". Anche i ruoli possono essere definiti come *topic* e spesso, ma naturalmente non sempre, è possibile utilizzare le categorie già definite precedentemente: considerando l'*association* di tipo "è pubblicato", il *topic* "Bollettino AIB" ha l'*association role* "rivista" mentre "Le mappe topiche" e "DOI: uno strumento per costruire la biblioteca digitale" hanno il ruolo di "articolo"³⁸. Per c) avremo l'*association* "scritto da", con "Federico Meschini" nel ruolo di "autore" e "Le mappe topiche" in quello di "articolo". Essendo le relazioni in una mappa topica bidirezionali avremo lo stesso tipo di *association* e di *association role* per d), anche se naturalmente cambiano i *topic* coinvolti. Arrivati alla fine del TAO, nel nostro percorso nel modello della mappe topiche, troviamo le *occurrence*. L'occorrenza di un *topic* è una risorsa informativa, di qualsiasi genere essa sia, in qualche modo rilevante o collegata all'argomento rappresentato da quel dato *topic*. Le occorrenze per "Bollettino AIB" potrebbero essere sia la copia cartacea della rivista sia l'edizione elettronica pubblicata sul Web, entrambe relative al numero corrente. In qualche modo all'interno della *topic map* bisogna poter fare riferimento in maniera univoca e non ambigua a queste risorse, ma questo aspetto è lasciato all'implementazione concreta del modello e naturalmente al tipo di formato cui si sta puntando³⁹. È fondamentale sottolineare come la separazione tra le mappe topiche e gli oggetti cui fanno riferimento, e tutte le potenzialità che da essa ne derivano, avvenga proprio qui, nella distinzione tra *topic* e *occurrence*, elementi che di fatto sono su due livelli distinti: i primi nell'empireo, nel regno intangibile delle idee platoniche, i secondi nel mondo della natura, della concretezza fisi-

37 Consideriamo solo la semantica di queste affermazioni e non la struttura grammaticale. Quindi sarebbe equivalente scrivere "Le mappe topiche è pubblicato sul Bollettino AIB".

38 Nulla vieta di utilizzare altre categorie e di conseguenza altri *topic*, magari di ambito più generico, come per esempio "contenitore" e "contenuto".

39 Nel caso delle risorse elettroniche presenti in Internet si utilizzano gli URI/URL.

ca. Naturalmente anche le occorrenze sono caratterizzate dall'appartenenza a un particolare tipo, che è rappresentato come al solito da un *topic*. La copia fisica e la versione elettronica dello stesso numero "Bollettino AIB" riguardano evidentemente lo stesso *topic*, ma altrettanto evidentemente appartengono a due tipi diversi⁴⁰. Un aspetto fondamentale è la possibilità di assegnare uno *scope*, un contesto, un ambito ben definito e delimitato di validità, alle caratteristiche di un *topic*: i suoi nomi, associazioni e occorrenze. Proseguendo con il nostro esempio, un contesto per l'*association* esistente tra "Bollettino AIB" e "Le mappe topiche" potrebbe essere la data di pubblicazione, marzo 2005, oppure la lingua italiana. Così come per le associazioni e le occorrenze anche agli *scope* deve essere assegnato un tipo⁴¹. Per verificare che a un *topic* corrisponda uno e un solo argomento è previsto l'utilizzo della *subject identity*⁴². Se due *topic* hanno la stessa identità semantica allora fanno riferimento allo stesso soggetto, devono essere uniti e le rispettive caratteristiche, nomi, associazioni e occorrenze, vengono sommate tra loro. Ciò è molto utile quando si effettua il *merge*, la fusione tra due o più *topic map*, in modo da poter coprire un più vasto settore di informazioni, per non avere ridondanza nella definizione dei *topic*. Per questo motivo la *subject identity* deve essere una risorsa informativa identificabile in maniera univoca⁴³. Se questo non è possibile, nel caso si debba far per forza riferimento a qualcosa di astratto, come il concetto di Italia o di lingua italiana, si può utilizzare un *subject indicator*, definito come un «positive, unambiguous indication of the identity of a subject»⁴⁴. Essendo una risorsa, un indicatore semantico ha un indirizzo, solitamente un URI, che va a costituire la *subject identity*. Nel caso l'indicatore sia liberamente disponibile e destinato a un uso pubblico, viene definito *published subject indicator* (PSI) e il suo scopo principale è quello di facilitare l'interoperabilità e gli scambi tra le varie mappe topiche, costituendo una comune base di conoscenza. Una rappresentazione grafica può essere d'aiuto per comprendere meglio i concetti espressi finora:

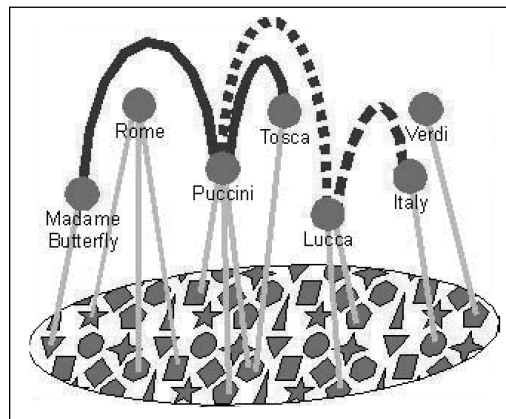


Fig. 1: Diagramma della struttura di una *topic map*

⁴⁰ Che possono essere identificati a loro volta dai *topic* "edizione cartacea" ed "edizione elettronica".

⁴¹ A sua volta identificato da un *topic*.

⁴² Traducibile come "identità di argomento".

⁴³ Come nel caso delle risorse dotate di URI. Per il *topic* "Bollettino AIB" potrebbe essere la pagina principale del suo sito Web: <<http://www.aib.it/aib/boll/boll.htm>>.

⁴⁴ Steve Pepper, *The TAO of topic maps* cit.

Nella figura 1⁴⁵ i *topic* sono i cerchi in alto contraddistinti da dei nomi⁴⁶ e sovrapposti a un gruppo di risorse informative, disegnate in basso come cerchi, quadrati, stelle e triangoli racchiusi in un'ellisse. Le linee rette che collegano le figure geometriche in basso con i cerchi in alto di conseguenza rappresentano le occorrenze di un argomento, mentre le parabole tra i vari *topic* sono le relazioni esistenti tra di loro.

Le possibilità offerte da una *topic map* sono notevoli: può essere progettata su qualsiasi cosa ed è in grado di farlo anche uno studente delle elementari⁴⁷. Però, come per ogni altra tecnologia, la diffusione e il successo dipendono soprattutto da altri fattori, sia tecnici, come la facilità d'implementazione e la disponibilità di applicazioni, sia politici, ossia l'adozione e l'utilizzo da parte delle varie entità e istituzioni.

XML Topic Maps

Lo sviluppo di una sintassi concreta per questo modello ha avuto inizio nei primi anni Novanta utilizzando come base HyTime⁴⁸, uno standard ISO pubblicato nel 1992 con lo scopo di fornire a SGML⁴⁹ funzionalità multimediali e collegamenti ipertestuali avanzati. Il lavoro è continuato in questa direzione finché le specifiche *topic map* non sono divenute a loro volta uno standard ISO nel 2000, con la sigla 13250⁵⁰.

La necessità di applicare gli stessi principi al più grande contenitore di documenti e risorse digitali mai esistito prima, il WWW, ha reso necessaria la creazione di un nuovo paradigma, conforme alle specifiche di XML⁵¹, divenuto ormai la lingua franca per l'interscambio di documenti e dati tra i vari programmi, in particolare quelli che comunicano tra di loro utilizzando i protocolli di Internet.

Il lavoro su questo linguaggio, denominato XTM (*XML Topic Maps*), è iniziato nel 2000 da parte dell'organizzazione indipendente TopicMaps.org⁵² e nel marzo del 2001 è stata resa disponibile la versione 1.0⁵³ tramite una DTD⁵⁴. Lo scopo perseguito era la creazione di un qualcosa che fosse direttamente usabile in Internet, compatibile con lo standard ISO 13250, in grado di supportare ed essere suppor-

45 Steve Pepper, *The TAO of topic maps* cit, <<http://www.ontopia.net/topicmaps/materials/tao-asstypes.jpg>>.

46 Madame Butterfly, Roma, Puccini, Tosca, Verdi, Lucca ecc. Il riferimento è naturalmente all'opera. Questo esempio è poi sviluppato da Pepper sino a divenire un'applicazione completa, *The Italian opera topic map*, di cui si parlerà in seguito.

47 Si veda Jack Park-Nefer Park, *Topic maps in the life sciences*, in *XML topic maps: creating and using topic maps for the Web* cit., p. 149-166.

48 HyTime, <<http://www.hytime.org>>.

49 *Standard Generalized Markup Language*, il linguaggio di marcatura da cui ha avuto origine HTML <<http://www.w3.org/MarkUp/SGML>>.

50 Si veda il documento *ISO/IEC 13250 topic maps*, disponibile all'indirizzo <http://www.y12.doe.gov/sgml/sc34/document/0322_files/iso13250-2nd-ed-v2.pdf>.

51 *eXtensible Markup Language*, <<http://www.w3.org/XML>>.

52 *TopicMaps.org*, <<http://www.topicmaps.org>>.

53 *XML Topic Maps (XTM) 1.0*, <<http://www.topicmaps.org/xtm/index.html>>.

54 Document Type Definition, uno dei modi con i quali è possibile stabilire la grammatica, le regole di sintassi, di un linguaggio basato su XML. La DTD di XTM è disponibile in linea: <<http://www.topicmaps.org/xtm/xtm1.dtd>>.

tato facilmente da un vasto numero di applicazioni, con il minimo possibile di funzioni opzionali, *human-readable*, semplice da apprendere e da utilizzare e dal *design* formale e conciso. Vengono inoltre utilizzate, anche se a un livello base, altre tecnologie collegate a XML: XPath, XLink ed XPointer⁵⁵. Da un'analisi della DTD e delle specifiche lo scopo sembra raggiunto: XTM 1.0 è composto da soli diciannove elementi⁵⁶, un attributo⁵⁷ e ha una curva di apprendimento relativamente semplice, tanto più veloce quanto più sono chiari i concetti teorici del modello di riferimento, oltre naturalmente a una conoscenza dei principi fondamentali di XML⁵⁸.

Facendo una veloce rassegna degli elementi disponibili in XTM, senza considerare naturalmente la loro struttura e la grammatica, ossia i rapporti interni e le combinazioni⁵⁹, risulta evidente la stretta aderenza al modello teorico. La radice di una XML topic map è l'elemento <topicMap>. *Topic*, *association*, *occurrence* e *subject identity* hanno le loro controparti omonime: <topic>, <association>, <occurrence> e <subjectIdentity>. Il nome principale di un *topic* viene memorizzato in <baseName>, che a sua volta si serve di <baseNameString>. Per gli eventuali altri nomi sono presenti <variant>, <variantName> e <parameters>, ossia i criteri in base ai quali è previsto l'utilizzo al posto del nome ufficiale. Per indicare l'appartenenza e il ruolo vengono utilizzati rispettivamente <member> e <roleSpec>. Il contesto, l'ambito di validità, viene indicato con <scope> e l'organizzazione in categorie generali con <instanceOf>. Per i riferimenti a *topic*, risorse e indicatori di argomento sono previsti rispettivamente <topicRef>, <resourceRef> e <subjectIndicatorRef>. Con <resourceData> è possibile inserire descrizioni e informazioni testuali e infine per l'unione di due o più mappe si può utilizzare <mergeMap>. Risulta evidente come questo vocabolario sia estremamente essenziale se paragonato ad altri linguaggi, come la stessa TEI. Utilizzando questa sintassi la TopicMaps.org rende disponibili sul Web dei PSI per i codici ISO delle lingue e delle nazioni⁶⁰, formando così una sorta di XTM *knowledge base*.

55 XPath, <<http://www.w3.org/TR/xpath>>, viene utilizzato per selezionare parti di un documento XML, rappresentato come una struttura ad albero. XLink, <<http://www.w3.org/XML/Linking>>, permette di creare collegamenti tra le risorse, dai semplici link ipertestuali unidirezionali a funzionalità più avanzate. XPointer, <<http://www.w3.org/TR/xptr>>, basato in buona parte su XPath lavora in coppia con XLink per puntare a sezioni specifiche di un documento XML.

56 L'elemento (o marcatore) è alla base del concetto di linguaggio di marcatura. Si tratta di un elemento etichetta che contemporaneamente delimita le informazioni racchiuse al suo interno.

57 Un attributo aggiunge informazioni all'elemento di cui fa parte. In XTM è presente l'attributo *id* che serve a identificare in maniera univoca un elemento all'interno di un documento XML. Sono presenti altri due attributi, *xlink:href* e *xlink:type*, ma sono relativi ad XLink.

58 Sul Web, oltre naturalmente alle specifiche ufficiali, sono disponibili dei *tutorial*, Anitta Altenburger, *Authoring XTM topic maps*, 2000, <<http://topicmaps.it.bond.edu.au/docs/6/toc>>, e Michel Biezunski, *The XTM guide: a beginner's guide to the XTM syntax*, 2005, <<http://www.infoloom.com/xtmguide.htm>>.

59 Tutte queste informazioni sono espresse nella DTD.

60 <<http://www.topicmaps.org/xtm/1.0/language.xtm>> e <<http://www.topicmaps.org/xtm/1.0/country.xtm>>.

E nel mondo reale?

Lo scopo delle mappe topiche e di XTM è di rendere il Web un posto più ordinato e i programmi per ricercare informazioni più intelligenti, di conseguenza il loro utilizzo rimane il più delle volte nascosto all'utente finale, a meno che non sia espressamente dichiarato.

Nonostante questo è possibile trovare sulla rete diversi esempi "in chiaro" che danno l'idea delle reali possibilità di utilizzo. Il primo caso preso in considerazione è Omnigator⁶¹ realizzato dalla Ontopia⁶² con tecnologia J2EE, definito un *knowledge browser* e scaricabile gratuitamente dal sito. Sulla pagina iniziale troviamo sia una descrizione del programma sia diverse *demo* già pronte. Omnigator permette di navigare un qualsiasi file di *topic map* gli venga sottoposto, nei formati XTM, HyTime e LTM⁶³. Uno degli esempi più completi e più noti è incentrato sull'opera italiana: *The Italian opera topic map*⁶⁴. Navigando *The Italian opera* con Omnigator nella versione XTM⁶⁵ è possibile scegliere tra diverse modalità di visualizzazione. La prima, *ontology role types* e *occurrence types*. La successiva, *master index*, mostra tutti i *topic* (ben 1339) in ordine alfabetico. Le due ultime opzioni, *index of individuals* e *index of themes*, mostrano rispettivamente il primo un'organizzazione per *topic* "reali", ossia quelli che non rappresentano una categoria astratta ma un argomento concreto, qualsiasi esso sia, e il secondo invece per motivi musicali, le arie. Indifferentemente dalla modalità scelta, per ogni *topic* è possibile vedere i nomi, le occorrenze interne ed esterne alla mappa e le associazioni esistenti.

Un chiaro esempio di come una mappa topica sia fondamentalmente un *semantic middleware* da utilizzare su un gruppo di informazioni, è dato dall'applicazione OperaMap⁶⁶, una *demo* realizzata sempre dall'Ontopia, ma in questo caso impiegando due suoi prodotti commerciali: la Knowledge suite⁶⁷ e il Navigator framework⁶⁸. In quello che sembra a prima vista un normale sito Web tutte le pagine HTML sono generate automaticamente tramite questi software a partire sempre dallo stesso file *opera.xtm*, contenente l'*Italian opera topic map* vista precedentemente con Omnigator. Stessi dati di partenza ma il *presentation layer*, lo strato di presentazione, è totalmente diverso⁶⁹.

Passando a un settore più specifico e a noi più vicino, una delle prime e più interessanti dimostrazioni dell'uso combinato di *topic map* e risorse testuali codificate in TEI è stata effettuata da John Walsh durante il *Digital resources for the humanities*

61 Omnigator, <<http://www.ontopia.net/omnigator/models/index.jsp>>.

62 Ontopia: the topic map company, <<http://www.ontopia.net>>.

63 Nell'ultima versione supporta anche RDF. È stato inserito anche il Vizigator, un visualizzatore grafico di *topic map*.

64 Il codice è disponibile sia in XTM <<http://www.ontopia.net/omnigator/docs/navigator/opera.xtm>> sia in HyTime <<http://www.ontopia.net/omnigator/docs/navigator/opera.hytm>>.

65 <http://www.ontopia.net/omnigator/models/topicmap_complete.jsp?tm=opera.xtm>.

66 OperaMap: the Italian opera topic map, <<http://www.ontopia.net/operamap/index.jsp>>.

67 The Ontopia Knowledge Suite (OKS), <<http://www.ontopia.net/solutions/products.html>>.

68 Ontopia solutions: Navigator framework, <<http://www.ontopia.net/solutions/navigator.html>>.

69 E in quest'ultimo caso decisamente più amichevole.

tenutosi a Newcastle nel settembre 2004⁷⁰. La dimostrazione mostrata durante la presentazione non è purtroppo disponibile *online*, ma l'abstract ci dà una chiara idea del tipo di ricerca svolta⁷¹ e dei possibili sviluppi futuri in questo campo. Utilizzando come base le poesie digitalizzate di Algernon Swinburne dello *Swinburne project*⁷², Walsh ha identificato le numerose allusioni e i riferimenti presenti in queste opere che, data la vasta cultura del poeta, spaziano dai temi classici agli eventi politici contemporanei, passando per il ciclo arturiano. La *Swinburne topic map* così realizzata includeva forme poetiche, riferimenti a personaggi ed eventi storici, figure bibliche, mitologiche e arturiane, così come alle opere di altri artisti e poeti. Il file XTM in questione è stato visualizzato con Omnigator, ma Walsh sta lavorando alla creazione di uno strumento *open source* denominato Topic maps for digital humanities (TM4DH), che verrà presentato durante l'ALLC/ACH 2005.

Un'applicazione scritta personalmente, sia come "palestra" sia come base per un ulteriore sviluppo, è stata realizzata a partire dalle versioni elettroniche delle relazioni di viaggio e delle guide turistiche su Roma e nel Lazio rese disponibili *online* dal progetto Avirel⁷³, diretto da Vincenzo De Caprio. I testi⁷⁴, disponibili nella sezione *Biblioteca digitale*⁷⁵, contengono le descrizioni di numerosi siti e monumenti celebri di Roma, come la Basilica di San Pietro, Piazza del popolo o il Colosseo, ma per forza di cose mantengono la struttura stabilita dai loro autori al momento della pubblicazione. Con un file XTM è stato possibile registrare per ogni luogo il punto esatto in cui compariva all'interno dell'opera e riorganizzare queste informazioni facendo in modo che il punto di partenza fosse stavolta il luogo. Sulla pagina iniziale dell'applicazione⁷⁶ è presente l'elenco dei posti, ognuno un link attivo, e selezionando per esempio "Basilica di San Pietro", compaiono tutti i testi che ne contengono una descrizione. Anche le voci di questo elenco sono dei link e scegliendone uno si andrà nel punto esatto in cui compare all'interno dell'opera la relativa descrizione⁷⁷. Si può inoltre visualizzare direttamente il codice sottostante⁷⁸.

I campi possibili di utilizzo sono decisamente numerosi, spaziando dalla creazione di ontologie, a questioni di *knowledge management*, fino alla "semplice" riorganizzazione di risorse e documenti già esistenti, come nel caso di un portale.

70 DRH 2004, <<http://drh2004.ncl.ac.uk>>.

71 John A. Walsh, *Topic maps and TEI-encoded literary texts*. Abstract disponibile sul Web: <<http://drh2004.ncl.ac.uk/abstract.php?abstract=177>>.

72 *The Swinburne project*, <<http://www.lettrs.indiana.edu/swinburne>>.

73 *Avirel: archivio viaggiatori italiani a Roma e nel Lazio*, <<http://www.avirel.it>>.

74 Codificati nel formato TEI e resi disponibili in HTML tramite XSLT, i fogli di stile di trasformazione di XML, <<http://www.w3.org/TR/xslt>>.

75 *Biblioteca digitale Avirel*, <<http://www.avirel.it/bd/index.htm>>.

76 <http://www.avirel.it/cocoon/avirel_tm/avirel_tm.html>

77 Si tratta dello stesso documento contenuto nella *Biblioteca digitale*, evitando qualsiasi tipo di ridondanza.

78 All'indirizzo <http://www.avirel.it/cocoon/avirel_tm/avirel_tm.xml>. La generazione di tutte le pagine HTML è effettuata utilizzando i fogli di stile XSLT, applicati dinamicamente al file XML tramite Cocoon, un potente *framework* di pubblicazione *open source* sviluppato dal gruppo Apache, <<http://cocoon.apache.org>>.

Il numero di società che offrono prodotti e servizi basati su questo standard, anche di livello *enterprise*⁷⁹, è in aumento⁸⁰, testimoniando così la crescita di interesse e diffusione. Accanto ai pacchetti commerciali, una soluzione *open source* interessante è costituita da TM4J⁸¹, un contenitore di diversi progetti, come lascia intuire il nome, tutti realizzati in Java, per la creazione, manipolazione e pubblicazione delle mappe topiche.

Topic map vs. resto del semantic Web

Una domanda da non rimandare e porsi subito è quella sul rapporto e le possibilità d'interazione tra le varie soluzioni per il Web semantico, per evitare il rischio, veramente paradossale, di trovarsi con tanti mondi perfettamente coerenti e organizzati internamente ma che non siano poi in grado di comunicare tra loro. Dublin Core per esempio può essere espresso perfettamente tramite la *topic map*⁸², ma il vero confronto è naturalmente con RDF. Il modello ISO del *topic - association - occurrence* esaminato in questo articolo e quello *resource - property type - value*, proposto dal W3C con il Resource description framework, sono tra di loro solamente diversi, concorrenti, oppure forse si può addirittura parlare di complementarità? La differenza fondamentale è che la struttura TAO consente di costruire un numero di rapporti potenzialmente infinito tra i suoi membri, mentre RDF esprime relazioni binarie tra le risorse⁸³. Indubbiamente la sintassi delle mappe topiche è più prolissa, ma anche più facilmente leggibile, e una mia personale impressione è che le mappe di argomenti siano riuscite a suscitare gli interessi e a scuotere settori che sentivano il bisogno di lavorare in qualche modo sui metadati semantici, ma non pensavano minimamente all'utilizzo di RDF, giudicandolo troppo complicato e difficilmente utilizzabile per i loro scopi⁸⁴. Ciò è dimostrato anche dalla maggiore vivacità nello sviluppo e disponibilità di specifiche tecniche, strumenti e interfacce di programmazione⁸⁵. Superato un primo momento in cui sembravano essere in competizione, l'intenzione è adesso di una convergenza e cooperazione tra le due specifiche, in modo da ottenere «the best of both worlds». Durante la progettazione di XTM sono state tenute in estrema considerazione le caratteristiche di RDF e il W3C ha istituito una

79 Così vengono definite le applicazioni aziendali di una certa entità.

80 Per un elenco completo si veda *Topic map vendors*, <<http://www.topicmap.com/topicmap/vendors.html>>.

81 Acronimo di *Topic Maps for Java*, <<http://tm4j.org>>.

82 Esempi di questa corrispondenza sono presenti sia nel codice di *The Italian opera topic map* sia di *Swinburne topic map*.

83 *RDF topic map mapping*, <<http://www.w3.org/2002/06/09-RDF-topic-maps>>.

84 Durante l'ALLC/ACH 2005 tre interventi saranno incentrati sull'utilizzo delle *topic maps*: Federico Meschini, *Classifying the Chimera*; John A. Walsh, *TM4DH (Topic Maps for Digital Humanities): examples and an open source toolkit*; Christian Wittern, *From text to topics: zigzagging towards the knowledgebase of tang civilization*. Informazioni <http://web.uvic.ca/hrd/achallc2005/abstract_list.htm>.

85 Come il progetto *Topic Map Application Programming Interface (TMAPI)*, <<http://www.tmapi.org>>, o lo sviluppo dei linguaggi *Topic Map Query Language (TMQL)*, <<http://www.isotopicmaps.org/tmql>>, e *Topic Map Constraint Language (TMCL)*, <<http://www.isotopicmaps.org/tmcl>>. Un elenco degli strumenti è disponibile all'indirizzo *Topic map tools*, <<http://www.topicmap.com/topicmap/tools.html>>.

task force per la RDF/topic maps interoperability⁸⁶, costituita da esponenti di entrambe le comunità. Gli articoli sull'argomento sono numerosi e di un certo livello⁸⁷, lasciando ben sperare per il futuro.

Conclusioni

Utilizzare una *topic map* in un sito o in un progetto può portare indubbiamente numerosi vantaggi, ma questo è solamente il primo passo. Internet sta diventando consapevole di questa tecnologia, come dimostrato da un recente articolo pubblicato su *Wired*⁸⁸. Il punto è se tramite le mappe di argomenti si riuscirà a rendere il World Wide Web un posto più ordinato. Gli scenari immaginabili sono numerosi, dalla creazione di reti di conoscenza interconnesse tra di loro che migliorino le funzioni dei motori di ricerca, all'utilizzo di speciali interfacce di navigazione integrate nei tradizionali *browser*. Per concretizzarsi, la realizzazione di una cartografia il più completa possibile del mondo digitale deve avvenire sia dall'alto, con l'intervento delle organizzazioni ufficiali e delle multinazionali dell'informazione, come Google e Yahoo, sia dal basso, nei vari settori disciplinari, con la creazione di tante piccole mappe, ognuna relativa al proprio campo di conoscenza specifico.

Vorrei concludere citando Sam Hunting⁸⁹, uno dei creatori di XTM, che, proponendo una sua visione particolare del futuro, invita a «[...] let's imagine that we live in a world much like today's world, except that is our topic map future, and topic maps are everywhere. Hundreds of thousands of J.R.R. Tolkien fans have merged their shrines into a single representation of Middle Earth».

86 SWBPD: *RDF/Topic maps interoperability task force description*, <<http://www.w3.org/2001/sw/BestPractices/RDFTM>>.

87 Si veda per esempio Eric Freese, *Topic maps and RDF*, in *XML topic maps: creating and using topic maps for the Web* cit. p. 283-324; Lars Marius Garshol, *Living with topic maps and RDF*, <<http://www.ontopia.net/topicmaps/materials/tmrdmf.html>>; Martin S. Lacher – Stefan Decker, *On the integration of topic maps and RDF*, <<http://www.semanticweb.org/SWWS/program/full/paper53.pdf>> e gli articoli presenti nella sezione *The relationship between RDF and topic maps* <<http://www.techquila.com/topicmaps/tmworld/11902.html>> del sito *Techquila: standards-based information management*, <<http://www.techquila.com>>.

88 John Gartner, *Searching smarter, not harder*, 2004, <http://www.wired.com/news/technology/0,1282,65840,00.html?tw=newsletter_topstories_html>.

89 Sam Hunting, *How to start topic mapping right away with the XTM specification*, in *XML topic maps: creating and using topic maps for the Web* cit. p. 82-83.

Topic maps

How I learned to stop worrying and love metadata

by Federico Meschini

Topic Maps are a technology for representing and organizing information, and they are constantly raising more and more interest in different sectors, thanks to their versatility. The main feature of a Topic Map is the creation of a semantic layer which can be applied to an existing group of resources, so to have new logical links and paths of exploration between different set of information. The logical structure of this kind of map is very simple. Each argument is represented by something called topic. Between topics is possible to create associations, which can be of many types, and each topic has a different role. Finally, there are the occurrences, which are the concrete embodiments of the abstract arguments represented by topics. There is no limit to the number or nature of the occurrences of a topic, they could be both digital than analogical, and in any format. The power of this model resides in this separation between these two levels, the abstract and the real one.

The development of a language for the description of a Topic Maps began with SGML and what has become the HyTime ISO standard. Currently there is also a specification which utilizes the XML syntax and is called XTM - XML Topic Maps. On the web is now possible to find many examples and applications which shows the vast potentialities of this paradigm, and the most famous is perhaps the Italian Opera Map, developed by Steven Pepper, one of the coauthors of the XTM specifications. In the field of Humanities Computing a very interesting application has been realized by John Walsh, who showed it during the DRH 2004 in Newcastle: the Swinburne Topic Map, about the famous Victorian poet and built upon the digital version of his poems encoded in TEI.

Topic Maps are growing in interest and they can really help in making the web a place more ordered; moreover the integration and interoperability with the other technologies related to the semantic web, such as Dublin Core and RDF, are also being considered, and there is potentially no limit to the possibilities of use. Of course, the real diffusion that the Topic Maps will have depends not only by their potentialities, but also from the adoption of this technology by the big players in the information business.

FEDERICO MESCHINI, Università della Tuscia, Facoltà di lingue e letterature straniere moderne, largo dell'Università, 01100 Viterbo, e-mail fmeschini@tin.it.

L'ultima consultazione di tutti i siti Web citati risale al 20 marzo 2005.