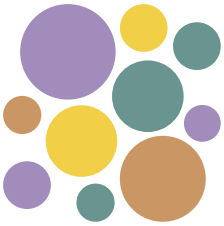
Thread | Esperienze

Chi ha paura dei metadati?

#2 • IL DATO È TRATTO

Laura Di Martile

Hyperborea



Continua da **Chi ha paura dei metadati? #1**

Il dato è tratto

Calcolo e memoria sono le doti con cui la tecnologia informatica supporta e potenzia calcolo e memoria umani, elaborando masse informative tali da non poter essere concepite dalle persone, ma da esse disciplinate e normalizzate. Componenti semantici, segni morfologici e sintassi sono alla base del linguaggio, anche informatico.

All'aprirsi del XXI secolo il concretizzarsi del cosiddetto **web semantico** affina e raffina i modelli interpretativi secondo cui la combinazione e la relazione tra dati univoci e riconoscibili possono essere interpretati in maniera automatica da applicazioni e da motori di ricerca e ricostruire contesti, anche complessi, significativi e pertinenti rispetto alle ricerche dell'utente. Il ricorso a elementi che identificano i contenuti di documenti, file, immagini, etc., rende manifeste le relazioni tra fonti, beni, persone ed eventi, costituendo il caposaldo e la logica dell'architettura semantica.

Per setacciare le informazioni utili, con profitto e senza dispersione, in oceani sempre più grandi di particolari significativi si fa ricorso alla composizione di dettagli di contorno che restituiscano il contesto, lo scenario semantico, dei dati nudi e crudi. Si fa ricorso alla **metadatazione**: ogni risorsa informativa può essere connotata come appartenente idealmente ad una classe ed essere caratterizzata da una o più proprietà. Questo kit di categorie e attributi può essere definito a priori, come set di notazioni legate al dispositivo tecnico con cui vengono prodotte o realizzate le risorse culturali (come, ad esempio, le impostazioni e il pacchetto di dati della macchina fotografica per gli scatti digitali) oppure essere modellato a seconda degli obiettivi di chi usa, pubblica o rende disponibili i dati (gli **open data** istituzionali ne rappresentano un caso).

Si potrebbe pensare al metadato come il dato secondario che incarta quello primario; a seconda dell'involucro una stessa informazione formale può appartenere a contesti diversi: la parola ***pesca*** potrebbe riferirsi all'atto del pescare in mare o a quello di estrarre numeri a sorte per l'assegnazione di premi. I connotati della confezione ci diranno se si tratti dell'uno o dell'altro caso.

Le informazioni di contesto non sono un'invenzione di oggi (si pensi banalmente ai dati rilevati in un qualsiasi ufficio di Anagrafe e alle voci sotto cui sono declinati: "Nato a:…", "il giorno…", "il mese", "l'anno…", etc…). In ambito di banche dati è stata un'intuizione quella di riconoscere e restituire il contesto di un elemento attraverso la codifica partecipata e condivisa di informazioni su di esso; come nei puzzle, nei bilanciamenti di chimica e nelle costruzioni Lego: se un incastro non ha le giuste caratteristiche non può far parte del quadro e va escluso. Attingere al sacchetto delle possibilità vuol dire pescare e riconoscere i dati che calzano e quelli che non calzano. Non potendo fare una verifica minima e diretta (come per i piedi delle sorelle di Cenerentola), l'unica soluzione sembra quella di contrassegnare i dati con attributi specifici, ma uguali e validi per più contesti che ne definiscano il bordo concettuale. Per un libro saranno il titolo, l'autore, l'editore, l'anno di pubblicazione etc.; per un documento lo scrivente, il destinatario, la data, il luogo, etc.; per una scultura, il nome dell'artista, la data dell'esecuzione, il materiale, la tecnica… Una casistica di "caselle" da riempire con i dati particolari: Calvino, Italo (*autore/voce di autorità*), Le cosmicomiche / Italo Calvino (*titolo*), 7. Ed (*edizione*), Torino: Einaudi, c1965 (stampa 1973) (*pubblicazione*).

La contestualizzazione delle informazioni viene fornita all'utente dall'intreccio delle informazioni su di essa, canovaccio tessuto e messo in campo, di volta in volta, al momento della ricerca, in funzione dei parametri immessi dall'utente. Ogni metadato rappresenta un corredo informativo sulla risorsa "dato" che può essere utilizzato, in tutto o in parte, per essere riconosciuto come di interesse per chi esplora. Più saranno i metadati in circolazione più le necessità degli utenti potranno avere riscontri e risposte efficaci. La metadatazione è infatti un meccanismo di connotazione e concatenazione tramite cui possono dialogare piattaforme di raccolte dati, anche locali e magari isolate. A certi patti.

Un italiano, un americano e un belga...

Si metta il caso in cui si trovino a dover condividere la dispensa di una cucina tre persone di lingua e nazionalità diverse e ciascuno voglia utilizzare del riso per realizzare un piatto tipico: riusciranno a capire quale sia la confezione che serve, se sia adatta alla pietanza da mettere nel piatto, se sia sufficiente? Le infor-

mazioni ci sono e sono sintetizzate in forma letterale e simbolica sull'imballaggio: tipo di riso, luogo di origine, peso, tempi di cottura, stabilimento di lavorazione e data di confezionamento, scadenza, modalità di conservazione...
Queste categorie di informazioni possono considerarsi quelle di contesto, i contenuti che descrivono e dettagliano queste "etichette" corrispondono al dato stesso: i metadati sono le caselle/capsule entro cui sono definiti i dati; la somma degli uni e degli altri dà vita alla descrizione di un "articolo". *35, 100, 9* non saranno più solo numeri o valori assoluti, ma saranno riconosciuti come elementi legati alla natura di un ingrediente per la cucina.

Se la metadatazione usa un linguaggio noto ai tre individui il problema non si pone; se ciò non si verifica si dovrà trovare un terreno comune di comunicazione, in parte appoggiandosi alla consuetudine di trovare le informazioni in un determinato ordine, in parte utilizzando formule trasversalmente note e comprensibili a tutti. In forma astratta, la possibilità di leggere e comprendere delle informazioni da parte di soggetti diversi si appoggia alla capacità di trovare un terreno comune di dialogo, all'**interoperabilità** concettuale ed operativa del sistema di comunicazione. Le prospettive con cui si guarda ad una informazione possono essere diverse ma la natura del dato deve essere definita e inequivocabile per chi ne dovrà fare uso. Il passaggio da una vista ad un'altra o da una piattaforma ad un'altra non può prescindere da uno scambio che mantenga l'integrità del corredo informativo.
Ma ci potrebbero essere dei dati "fuoristrada" rispetto all'area di interpretazione stabilita come comune: come la si mette per esempio con l'unità di misura che indica il famoso riso? Grammi, once? Si può utilizzare una equivalenza? Ecco far capolino una piccola piega, apparire un lieve solco, anche se non ancora una ruga.

[continua]

Hyperborea
TIME IS ON MY SIDE