

ABBONAMENTI

ARCHIVIO

PIÙ VISTI

SOCIAL

METEO

TUTTOAFFARI

LAVORO

LEGALI

NECROLOGIE

SERVIZI



LA STAMPA.it

TECNOLOGIA

Cerca...

TORINO - CUNEO - AOSTA - ASTI - NOVARA - VCO - VERCCELLI - BIELLA - ALESSANDRIA - SAVONA - IMPERIA e SANREMO

VOCI DI MILANO

ATTUALITÀ OPINIONI ECONOMIA SPORT TORINO CULTURA SPETTACOLI MOTORI DONNA CUCINA SALUTE VIAGGI EXTR@ FOTO VIDEO

HOME POLITICA ESTERI CRONACHE COSTUME TECNOLOGIA SCIENZA AMBIENTE LAZAMPA I TUOI DIRITTI MARE MONTAGNA SPECIALI

Tweet

0



Rimuovi dalla timeline di Facebook

TECNOLOGIA

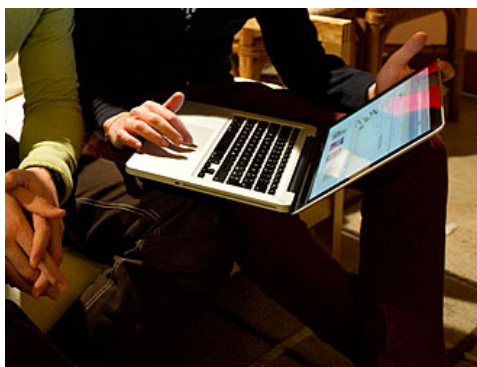
19/05/2011

Tecnologie del linguaggio per archivi "intelligenti"

Così la semantica ci aiuta a scovare documenti ricchi di informazioni

GIUSEPPE FUTIA

Se gli esseri umani si esprimono con le parole e le macchine ragionano sulla base di numeri, la natura del linguaggio rappresenta la differenza più profonda che esiste tra questi due mondi. L'ambizione di aziende specializzate in tecnologie semantiche, in grado di estrarre il significato dai testi, consiste nel ridurre questa distanza in modo tale che a partire da una specifica richiesta dell'utente, il computer sia in grado di rispondere alle sue esigenze nel miglior modo possibile.



Il Web attuale è un ambiente "generalista" che raccoglie un insieme di informazioni, il cui valore risulta molto variabile e non sempre adeguato a ciò di cui siamo alla ricerca. All'interno di un contesto di questo genere, i meccanismi sfruttati da Google, che estraggono le pagine Web più autorevoli a partire da una ricerca testuale, risultano ancora oggi strumenti molto efficaci, perché occorre risolvere in breve tempo un basso livello di complessità linguistica. Tuttavia, non tutte le collezioni di documenti sono vaste e generiche come il web e nel caso di archivi digitali che trattano di argomenti specifici, diventa importante fornire risultati migliori di quelli ottenuti dalla ricerca basata su parole chiave. In domini più ristretti, come gli archivi documentali storici, le tecnologie di analisi linguistica consentono di fornire risposte più mirate all'utente in cerca di informazioni.

Ad esempio, se l'utente scrive nella finestra di ricerca "immagini del Tintoretto" e poi "vita del Tintoretto", le sue richieste verranno effettivamente soddisfatte soltanto se, nel primo caso, i risultati ottenuti saranno foto digitalizzate di opere del Tintoretto e nel secondo se otterrà saggi di storia dell'arte. Oltre a questo, l'uso di strumenti semantici permette di effettuare le ricerche non sulla base delle parole digitate, ma sulla base di concetti: in un contesto architettonico si potrà associare il termine "cupola" al termine "volta", che assumerà dunque un significato ben preciso, benché alla parola "volta" si possano attribuire accezioni differenti. In entrambi i casi, la semantica consente una comprensione del significato della richiesta dell'utente e fornisce quindi, potenzialmente, un risultato più soddisfacente.

A questo proposito, l'**archivio storico de La Stampa** è una risorsa estremamente ricca e interessante su cui sviluppare un sistema di accesso semantico. Per l'archivio storico, **CELI**, azienda torinese che sviluppa tecnologie per offrire questo tipo di servizi, incaricata dal Comitato per la Biblioteca digitale dell'Informazione Giornalistica e in collaborazione con gli altri partner del progetto, sta sviluppando un sistema di analisi che consentirà una fruizione più efficace dei contenuti, aumentandone il valore d'uso per l'utente non specialista. La chiave di questa analisi, vista la natura delle notizie giornalistiche, è l'identificazione automatica nel testo di riferimenti a luoghi, persone ed eventi. In questo modo, sarà possibile per l'utente trovare tutte le notizie in cui viene citato un personaggio pubblico, o una azienda o un paese, e creare percorsi di ricerca personalizzati incrociando informazioni di diverso tipo sul contenuto degli articoli.

Rispetto alla semplice ricerca su parole chiave, il vantaggio principale è legato a una maggiore precisione:

Ultimi Articoli

+ Tutti gli articoli

OPEN LETTER TO SKYPE

FROM CONSUMERS PROTECT ASSOCIATES, INTERNET ACTIVISTS, JOURNALISTS & OTHER ORGANIZATIONS

+ Appello a Microsoft e Skype su privacy e trasparenza



+ Google guida la lotta per limitare l'accesso delle autorità ai dati personali



+ Web Movies in streaming dal 31 gennaio



+ Il Web "svela" la Corea del Nord



+ È il Dna l'hard disk del futuro



+ Google, Apple, The Pirate Bay Wikileaks: il cinema racconta il Web

Condividi gli articoli con i tuoi amici

Con l'app Facebook LaStampa.it puoi condividere immediatamente le notizie e gli approfondimenti che hai letto. Attiva l'app sul tuo profilo e segnala a tutti i tuoi amici le tue news preferite! Scopri di più su facebook.lastampa.it!

Accedi a Facebook

se si cercano gli articoli in cui viene citato Giovanni Leone (l'ex Presidente della Repubblica), non si vorrebbero ottenere risultati in cui compaiono i singoli termini – quelli in cui ad esempio è presente la parola leone – ma quelli in cui effettivamente si parla del personaggio in questione. Il progetto sull'archivio storico de La Stampa sostiene l'ipotesi secondo cui al crescere della quantità di informazioni, deve corrispondere un incremento della qualità degli strumenti con cui intendiamo esplorare le immense risorse che abbiamo a disposizione, perché il valore dei contenuti è spesso nascosto, difficile da individuare. E realtà multidisciplinari come le imprese che si occupano di semantica, nelle quali linguisti ed informatici lavorano fianco a fianco, rappresentano il mezzo migliore con cui affrontare tale complessità.

SPECIALE ELEZIONI



NEWS /
MULTIMEDIA



SCOPRI LE
NOSTRE APP



LE VOCI DELLA
POLITICA



PREMIO 1APP
4DEMOCRACY



I RISULTATI DEL
2008



GOOGLE
ELEZIONI

Annunci PPN



Aiuta una bambina
Le bambine soffrono di
pesanti discriminazioni
sessuali
[Adotta una bimba a
distanza](#)



**Conto Corrente
Arancio**
Zero spese, carta di credito
gratis. Scopri i vantaggi!
[contocorrentearancio.it](#)



Cerchi l'hotel ideale?
[trivago™](#) - Compara 600.000
Hotel in tutto il mondo
[trivago.it](#)