

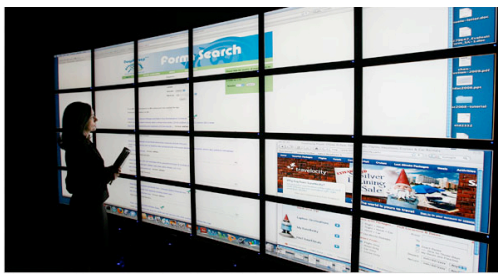
Scrivi qui per c

Ricerca realizzata con [Wordpress](#)

- [Home](#)
- [Chi siamo](#)
- [Contatti](#)
- [Redazione](#)

Esplorando il DeepWeb, i fondali della Rete dove Google non arriva

in [Internet e il web](#), [Media e potere](#)
di [Redazione](#) | 5 marzo 2009



At the University of Utah, Prof. Juliana Freire is working on DeepPeep, an ambitious effort to index every public database online.

*Le profondità inesplorate della Rete: il DeepWeb, un ambizioso tentativo di indicizzare ogni database pubblico on-line- Il cambiamento dell'esperienza della ricerca in Rete da parte dell'utente e le incredibili prospettive che si delineano per le aziende - Un articolo di Alex Wright sul **New York Times***

di Alex Wright
([New York Times](#))

La scorsa estate Google ha serenamente oltrepassato un traguardo memorabile: ha aggiunto la trilionesima pagina alla lista di pagine Web indicizzate. Ma per quanto il numero possa sembrare incredibilmente grande, rappresenta solo una frazione dell'intero Web.

Oltre questo trilione di pagine si estende, nascosto, un Web ancora più ampio: informazioni finanziarie, cataloghi commerciali, programmi di volo, ricerche mediche e ogni sorta di materiale immagazzinato in database che risultano perlopiù invisibili ai motori di ricerca.

Le sfide che i principali motori di ricerca si trovano ad affrontare nel penetrare il cosiddetto DeepWeb (Profondo Web) riguardano la loro incapacità nel trovare tutte le risposte effettivamente disponibili, anche alle domande più semplici.

Oggi una nuova razza di tecnologie sta prendendo forma ed estenderà la portata dei motori di ricerca fino agli angoli più nascosti del Web. Quando ciò sarà praticabile, non solo migliorerà la qualità dei risultati delle ricerche, ma potrebbe riformare il modo in cui molte aziende fanno affari on-line.

I motori di ricerca si affidano a programmi chiamati *crawler* (o *spider*) che raccolgono informazioni seguendo le tracce degli hyperlink che intrecciano il Web. Mentre questo approccio funziona bene per le pagine che costituiscono la superficie del Web, questi stessi programmi incontrano però maggiori difficoltà nel penetrare quei database che sono impostati per rispondere a determinate queries (interrogazioni).

“Il Web navigabile è solo la punta dell’iceberg”, sostiene Anand Rajarman, co-fondatore di Kosmix (www.kosmix.it), un’azienda start-up che si occupa di DeepWeb tra i cui investitori figura [Jeffrey P. Bezos](#), direttore generale di Amazon. Kosmix ha sviluppato un software che combina le ricerche con i database che più probabilmente raccolgono le informazioni maggiormente attinenti, per poi fornire una panoramica dell’argomento ricavata da molteplici fonti.

“La maggior parte dei motori di ricerca tenta di aiutare gli utenti a trovare un ago nel pagliaio”, dichiara Rajarman, “ma quello che stiamo cercando di fare noi è aiutarli ad esplorare il pagliaio”.

Ed il pagliaio è infinitamente ampio. Con milioni di database connessi al Web, e infinite possibili combinazioni di chiavi di ricerca, non vi è modo per nessun motore di ricerca – non importa quanto sia potente – di vagliare ogni possibile corrispondenza dei dati a disposizione.

Per estrarre dati significativi dal DeepWeb, i motori di ricerca devono le chiavi di ricerca e poi stabilire come indirizzare ogni determinata ricerca ad uno specifico database. Questo approccio potrebbe risultare immediato in teoria, ma in pratica l’incredibile varietà di strutture database e le possibili combinazioni dei termini costituiscono una sfida informatica alquanto spinosa.

“Si tratta della più interessante sfida circa l’integrazione dei dati che si possa immaginare”, afferma Alon Halevy, ex professore di scienze informatiche presso la [University of Washington](#) che attualmente guida un team di Google impegnato nel tentativo di risolvere l’enigma del DeepWeb.

La strategia di Google circa il DeepWeb include un programma che analizza i contenuti di ogni database che intercetta. Ad esempio, se il motore di ricerca individua un pagina con un contenuto correlato all’arte, allora inizia a vagliare ogni plausibile chiave di ricerca – “Rembrandt”, “Ricasso”, “Vermeer” e così via – finché uno dei termini non fornirà una corrispondenza. A questo punto il motore di ricerca analizza i risultati e sviluppa un modello anticipativo del contenuto del database.

Sulla stessa linea, la Prof.ssa Juliana Freire della University of Utah sta lavorando ad un ambizioso progetto chiamato DeepPeep (www.deepeeep.org) che mira a scovare e indicizzare ogni database presente pubblicamente sul Web. Estrarre i contenuti di così tanti database sparpagliati richiede sofisticate tecniche di intuizione informatica.

“Il modo più ingenuo sarebbe quello di interrogare ogni parola del dizionario” sostiene Freire. DeepPeep, invece, inizia ponendo un numero ridotto di interrogazioni, “che poi useremo per comprendere il funzionamento dei database e scegliere le parole da ricercare”, in modo da raggiungere oltre il 90% del contenuto di ogni database.

Mentre i principali motori di ricerca stanno sperimentando l’incorporazione di contenuti DeepWeb tra i propri risultati, devono anche capire come presentare differenti tipi di informazioni senza complicare eccessivamente le proprie pagine. Questa è una sfida specialmente per Google, che ha resistito a lungo alla tentazione di apportare significativi cambiamenti al format dei propri risultati, basato sul principio “provato e appurato”.

“Google affronta una vera e propria sfida”, sostiene Chris Sherman del sito Search Engine Land. “Vogliamo migliorare l’esperienza della ricerca, ma devono stare estremamente attenti nel proporre cambiamenti che potrebbero alienare i propri utenti”.

Oltre al regno delle ricerche degli utenti-consumer, le tecnologie DeepWeb potrebbero anche far sì che il mondo del business sfrutti le informazioni in modi differenti. Ad esempio, un sito dedicato alla salute potrebbe incrociare i dati delle case farmaceutiche con gli ultimi ritrovamenti della ricerca medica, o un sito di notizie locali potrebbe stendere la propria copertura lasciando che gli utenti attingano ai file custoditi nei database governativi.

Questo livello di integrazione dei dati potrebbe infine portare a qualcosa di simile al Web Semantico, la visione – molto promossa ma tuttora irrealizzata – di un Web fatto di dati interconnessi. Le tecnologie DeepWeb avanzano la promessa di raggiungere risultati simili a costi molto inferiori, automatizzando i processi di analisi delle strutture dei database e del controllo incrociato dei risultati.

“La cosa incredibile è la capacità di connettere fonti di dati disparate”, sostiene Mike Bergman, scienziato informatico che avrebbe coniato il termine DeepWeb. Bergman afferma che l’impatto a lungo termine del DeepWeb sarà più volto a modificare più il modo di fare affari piuttosto che soddisfare i capricci dei navigatori. Aspettare per credere.

(traduzione di Andrea Fama)

« [PeaceReporter, parte nuovo sito](#)
[Il canto del cigno della stampa Usa](#) »

Articoli correlati

- [Il canto del cigno della stampa Usa](#) (6 marzo 2009)
- [PeaceReporter, parte nuovo sito](#) (4 marzo 2009)
- [Benvenuti a New Media City](#) (3 marzo 2009)
- [La carta igienica 2.0? Non sarà di carta](#) (2 marzo 2009)
- [The Newspaper sulle tracce di Internet](#) (27 febbraio 2009)

Un commento

Sottoscrivi per questo articolo [gli RSS ai commenti](#) o lascia un [trackback](#)

Pingback e trackback

Pingback da [Il ReteGiornale - la Tua Voce in Rete» Libertà d'informazione » Esplorando il DeepWeb, i fondali della Rete dove Google non arriva](#)

il marzo 5th, 2009 at 09:16

Commenta

Username:

Indirizzo mail:

Indirizzo web:

Lascia la tua opinione:

post my comment

Dossier

- [Un "Progetto Einaudi-Albertini" per l' indipendenza dei media](#)
- [The Viewspaper sulle tracce di Internet](#)
- [Internet come arma di classe](#)
- [Tra leggi e diritti: la Rete nella rete](#)
- [Lunga vita al giornalismo partecipativo](#)

[Vedi tutti i dossier](#)

Categorie

- [Approfondimenti](#)
- [Attualità](#)

- [Censura e sangue](#)
- [Editoria](#)
- [Fotogiornalismo](#)
- [Giornali](#)
- [Internet e il web](#)
- [La professione](#)
- [La pubblicita](#)
- [La Tv e le tv](#)
- [Media e potere](#)
- [Nuovi giornalismo](#)
- [Radio e suoni](#)

Newsletter

Per iscriverti alla newsletter, invia una mail dalla [redazione](#). Ricordati di sostituire "[at]" con "@".

Partner



Licenza sui contenuti




[feeds](#)

Articoli recenti

- [Il canto del cigno della stampa Usa](#)
- [Esplorando il DeepWeb, i fondali della Rete dove Google non arriva](#)
- [PeaceReporter, parte nuovo sito](#)
- [Politkovaskaja: l' istinto di conservazione di una "enorme macchina repressiva corrotta"](#)
- [Benvenuti a New Media City](#)
- [La carta igienica 2.0? Non sarà di carta](#)
- [I quotidiani si rifanno il look. Un vezzo mentre la barca affonda?](#)
- [Un "Progetto Einaudi-Albertini" per l' indipendenza dei media](#)
- [R.I.P.: Rocky Mountain News, quotidiano USA attivo da 150 anni](#)

- [The Viewpaper sulle tracce di Internet](#)

Archivi

Network

- [Digg](#)
- [FaiNotizia](#)
- [Google News Italia](#)
- [Il gruppo di discussione Lsdi](#)
- [NewAssignment.Net](#)
- [Newsvine](#)
- [OhmyNews International](#)
- [Popurls](#)
- [Reporters](#)
- [Wikio News](#)

Pubblicità

Microfinance Empowers

Join us in
enabling the
poorest of the
poor to improve
their own lives

Public Service Ads by
Google

I tag più popolari

Articoli recenti

- [Il canto del cigno della stampa Usa](#)
- [Esplorando il DeepWeb, i fondali della Rete dove Google non arriva](#)
- [PeaceReporter, parte nuovo sito](#)
- [Politkovaskaja: l'istinto di conservazione di una "enorme macchina repressiva corrotta"](#)
- [Benvenuti a New Media City](#)
- [La carta igienica 2.0? Non sarà di carta](#)
- [I quotidiani si rifanno il look. Un vezzo mentre la barca affonda?](#)
- [Un "Progetto Einaudi-Albertini" per l'indipendenza dei media](#)
- [R.I.P.: Rocky Mountain News, quotidiano USA attivo da 150 anni](#)

- [The Viewpaper sulle tracce di Internet](#)

© [LSDI](#) | [Torna su](#) | [WPTP-Chocolate](#)