

# La nuova generazione dei motori di ricerca

Document clustering e gestione dei risultati

di Daniela Canali

**D**opo la posta elettronica, la ricerca continua a essere l'attività più diffusa tra gli utenti online. Secondo i risultati di uno studio pubblicato da Cyveillance nel luglio 2000, il web ad accesso pubblico<sup>1</sup> dovrebbe contenere più di due bilioni di pagine di informazione e continua a espandersi in modo esponenziale, con una crescita che tende a triplicarsi, stando alle stime delle Web Statistics dell'Office of Research di OCLC per gli ultimi due anni.<sup>2</sup> L'abilità di ritrovare nella rete informazione

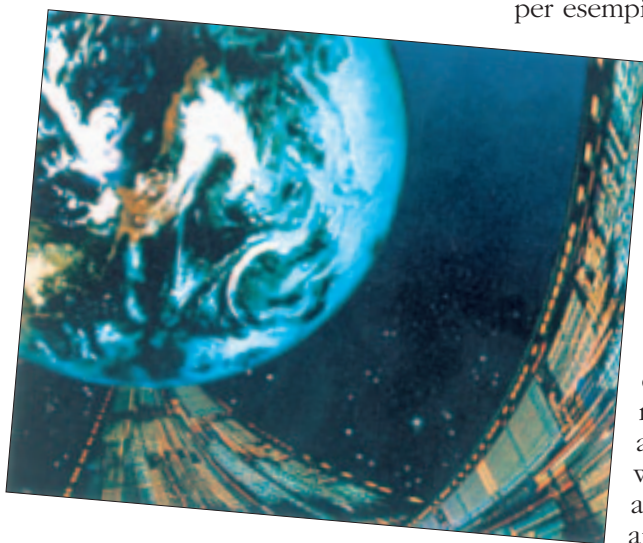
realmente utile è ancora legata al livello di precisione nell'impostare le richieste e alla capacità dell'utente di utilizzare i servizi, oltre naturalmente all'apporto degli *intelligent agents* che attualmente costituiscono il *core* dei motori di ricerca.

Per i milioni di siti che popolano Internet non esiste niente di simile agli standard di controllo bibliografico presenti nel mondo delle pubblicazioni a stampa: non esiste un equivalente dell'ISBN per identificare in modo univoco un documento; nessun sistema standard, analogo per esempio a quelli sviluppati dal-

la Library of Congress di catalogazione o classificazione, nessun indice centrale che registri il patrimonio informativo disperso nel web. Molti documenti in rete, forse la maggior parte, mancano persino dell'indicazione del nome dell'autore e della data di pubblicazione. Per trovare il famoso ago nel pagliaio (nel web, appunto) l'utente è abituato a utilizzare due approcci: il classico moto-

re di ricerca o la directory (guida o indice che dir si voglia) per soggetto. Tutti i motori effettuano in realtà ricerche per parole chiave nel proprio database ma i risultati sono influenzati da vari fattori: grandezza del database, frequenza di aggiornamento, capacità e stile della ricerca e velocità di esecuzione degli algoritmi. L'aggiunta di nuovo contenuto in formati diversi e le operazioni di restyling cui tali siti vanno soggetti, soprattutto per i frequenti cambiamenti societari, hanno portato alla nascita di nuove tipologie, i portali, basati sulla logica del *one-stop-shop*, che comportano un punto di partenza gradito all'utente e una localizzazione accentrata di tutti gli usi del web. Se pensiamo alla difficoltà in cui spesso ci troviamo nel cercare nel nostro computer un vecchio documento possiamo facilmente immaginare l'effettiva improbabilità di trovare informazione veramente pertinente e autorevole (per quanto sui parametri di autorevolezza del web ci sarebbe molto da dire) nell'hard disk globale del web: non a caso è stata coniata l'espressione *world wide anarchy*. Il lavoro dei motori consiste in pratica nel catalogare il web per noi tramite robot che instancabilmente e ininterrottamente scandagliano la rete cercando di mappare la vasta estensione del cyberspazio.

Molti erroneamente credono che la query lanciata in un motore di ricerca sia elaborata sulla globalità della rete Internet, senza rendersi conto che in realtà si sta usando solo un *subset* di ciò che è disponibile.<sup>3</sup> Il problema principale non è dunque la scarsità, ma l'abbondanza dei risultati. Le tecnologie di ricerca offerte dai motori, in costante evoluzione quanto a velocità, sistemi di ranking e capacità dei database, sono in fondo molto vicine ai sistemi di *text mining* tipici dei *knowledge management systems*. A livello di architettura logica, i motori si dividono in *search*





*architecture* (software di *spidering* connessi al web per intercettare nuove pagine da indicizzare nei database relazionali tramite indice invertito o *word-based*) e in *directories* (specie di sottocategorie di argomento comune per website, piuttosto che per singole pagine). Le pagine web indicizzate nel database sono accessibili all'utente tramite un comune browser grafico: l'indicizzazione avviene di solito in base alle parole che la pagina contiene, attraverso i tradizionali indici invertiti o il modello di vettore spaziale di Salton, sviluppato da Gerald Salton, padre dell'information retrieval: questo modello crea un punteggio delle parole chiave "pesate" in base alla relazione tra la query dell'utente e una qualsiasi pagina web. La maggior parte dei motori usano inoltre oggi l'informazione *inlink* per fornire un ranking di qualità per ogni pagina web recuperata. Basandosi sul concetto di "popolarità", una pagina web "linkata" da molti siti ritenuti di qualità è considerata a sua volta più rilevante di altre.<sup>4</sup>

La nuova generazione dei motori di ricerca è rappresentata dai motori che utilizzano tecniche di visualizzazione dei risultati quali la *document clustering* e criteri di ranking completamente nuovi. Un esempio è Teoma,<sup>5</sup> della Teoma Technologies, acquisita nel settembre 2001 dalla società californiana AskJeeves che, nonostante l'intenzione di implementarne la tecnologia nelle proprie funzioni di ricerca, ne ha lasciato online la versione beta come autonomo motore di ricerca.<sup>6</sup> Il motore, il cui nome significa "esperto" in lingua gaelica, applica algoritmi altamente avanzati nell'organizzare in tempo reale siti web ritenuti significativi per le chiavi di ricerca. Applicando una nuova misura di rilevanza, la *subject-specific popularity*, si basa su un modello matematico di struttura del web per generare query dina-

miche. Ai risultati delle ricerche vengono applicati criteri di popolarità e di analisi del testo, un clustering dinamico per argomento e un'analisi specifica dei link,<sup>7</sup> ed infine, ed è questo l'aspetto più insolito, c'è il coinvolgimento di esperti. La *dynamic topic clustering*, guardando al web da una prospettiva locale, sembra dotare il motore della capacità di comprendere l'argomento di cui trattano le pagine web che sta analizzando, a differenza dei normali metodi di ricerca che rischiano di segnalare siti di scarsa attendibilità. Stando agli esperti, Teoma rappresenta la naturale evoluzione dei criteri di ricerca, nel rispetto della libertà dell'utente che può così scegliere di utilizzare "un'esperienza guidata": la vera peculiarità del motore è rappresentata dalla capacità di esaminare il numero di pagine sullo stesso argomento che fanno riferimento a un sito, per determinarne il livello di autorevolezza in aggiunta a quello di *subject-specific popularity*. Teoma impiega inoltre l'analisi identificativa di esperti per fornire liste di risorse "esperte" su un argomento. La pagina dei risultati presenta tre opzioni basate su tecniche proprietarie: "Refine", "Results" e "Resources". I siti risultano organizzati in *web communities* naturali relative all'argomento di ogni query, presentate sotto l'intestazione "Refine", la funzione che permette di restringere o di raffinare la ricerca. Dopo l'identificazione di queste *web communities*, viene analizzata la relazione dei siti nell'ambito di una comunità, assegnando il punteggio in base al numero delle pagine sull'argomento che vi fanno riferimento. In altre parole, viene determinata la miglior risposta per una ricerca chiedendo ad esperti nell'ambito di una comunità web quale sia la ri-

sorsa migliore. Valutandone le opinioni, Teoma stabilisce l'autorevolezza del risultato della ricerca. I risultati di ricerca rilevanti assegnati secondo la *subject-specific popularity* vengono presentati sotto l'intestazione "Results" nella pagina dei risultati. Le cosiddette risorse esperte su un argomento contengono liste di altri siti autorevoli e link relativi all'argomento di ricerca. Dunque i risultati sono presentati in forma di *web page* (link a siti individuali che soddisfano la query dell'utente), *web page* raggruppate per argomento (folder che contengono tutti i siti che soddisfano la query raggruppati per sottoargomenti, le *web communities*) ed *expert links* (link ai siti web che forniscono liste di siti autorevoli su un particolare tema). L'aumento della concorrenza, secondo gli analisti, non può che migliorare la qualità delle ricerche in linea: Danny Sullivan, editore di SearchEngine Watch.com, definisce Teoma come una scelta interessante,<sup>8</sup> ma dubita fortemente del fatto che, come si legge spesso in rete, si trasformi nel killer di Google, che elabora più di 150 milioni di query di ricerca al giorno (calcolate sul proprio e su altri siti, come Yahoo!, che ne utilizzano la tecnologia).<sup>9</sup> Inoltre indicizza tre miliardi di documenti web, rispetto ai 200 milioni attuali di Teoma<sup>10</sup> che entro l'anno dovrebbe raggiungere quota 500 milioni. Anche l'homepage presenta una somiglianza fra i due siti nella scelta dello sfondo ►

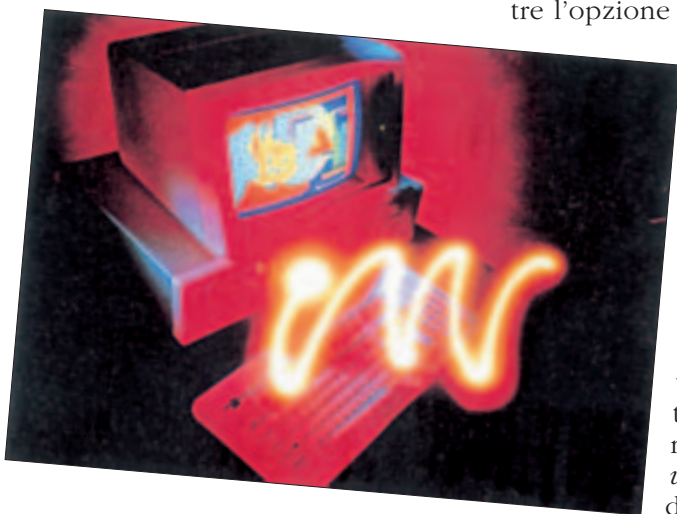
bianco con pochi tratti di colori luminosi. Il paragone proposto da Kieren McCarthy in "The Register" è però significativo:

Google asks about a certain expert in a field and then goes and takes a poll from people in the street over which one they think is best. Teoma would ask all the experts in the field which one of themselves they think is best.<sup>11</sup>

Altri punti giocano decisamente a favore di Google: in Teoma non esiste la ricerca avanzata (operatori booleani ecc.), manca la funzione di *cache*, molto utilizzata in Google, (mentre Wisenut, molto simile a Teoma, come vedremo, ha la funzione "Sneak-a-Peek" che consente il preview delle pagine). Agli amanti di questo tipo di strumenti di ricerca, dal settembre 2001 la rete offre il sito di Wisenut,<sup>12</sup> motore che, con lo slogan "Search. Exactly.", dichiara di migliorare la rilevanza dei risultati rispetto a Google attraverso l'algoritmo di *context-sensitive ranking*, in grado di esaminare e confrontare link e testo di una pagina, ponendo al primo posto il risultato più rilevante. Nella "guerra dei motori" Wisenut dichiara di possedere la tecnologia di ricerca più veloce ed efficace, anche dal punto di vista dei costi, sostenendo di poter indicizzare 50 milioni di pagine al giorno utilizzando solo

100 server: attualmente nel suo database raccoglie 800 milioni di pagine. Il suo fondatore, Yeogirl Yun, sostiene che il criterio di rilevanza utilizzato da Google ha creato grande popolarità tra gli utenti, ma il ciclo di leadership nelle tecnologie di ricerca sembra essere di due o tre anni, come già dimostrato dall'avvicinarsi nei favori degli utenti di Yahoo!, AltaVista e quindi Google.<sup>13</sup> Anche per Wisenut si parla di *more user-friendly format* dei risultati di ricerca:<sup>14</sup> il sito offre un semplice box di ricerca mentre il valore aggiunto, per così dire, è nella pagina dei risultati, grazie all'utilizzo delle categorie "WiseGuide", sezione in cui sono automaticamente ed estemporaneamente organizzati i risultati, in modo da fornire all'utente una guida rapida alla ricerca di risultati, specifici: questa funzione genera infatti categorie semanticamente relative ai termini della query.<sup>15</sup> In caso di query molto generiche o che utilizzano termini dal significato molteplice, questa caratteristica risulta in effetti molto utile. La funzione "WiseWatch" permette di scegliere quanti risultati si vogliono visualizzare per pagina (10, 20, 30 o 50), se mostrare un sommario di ogni pagina o solo titolo e url, e se raggruppare i risultati in modo che siano ordinate tutte le pagine web rilevanti dello stesso sito. Offre inoltre l'opzione di decidere in quale lingua, tra una scelta di oltre 25 idiomi, effettuare la ricerca e se eventualmente ritornare alla *WiseGuide*. La tecnologia proprietaria *WiseWatch* è anche un filtro opzionale di protezione contro pagine potenzialmente offensive per l'utente. WisenutBot è il nome di ZyBorg, il *web crawler* utilizzato da Wisenut per scan-

dagliare il web e indicizzarne tutte le pagine pubblicamente disponibili, nel rispetto dei controlli di accesso. Attualmente Wisenut ha una versione americana, una coreana ed è online anche la versione giapponese, con un database di 78 milioni di pagine tutte ricercabili utilizzando il *context-sensitive-ranking-system*. Il motore è utilizzabile in Lycos Japan che ha realizzato una partnership con Wisenut per potenziare il proprio motore, avendolo preferito a Google grazie alla funzione di *clustering* dei risultati. La versione giapponese include anche la funzione "Sneak-a-Peek" che permette un preview del sito senza dover lasciare la pagina dei risultati.<sup>16</sup> Il team di Wisenut tiene a precisare che, a differenza della maggior parte dei motori che si basano sul sistema operativo Microsoft come piattaforma per tutte le operazioni di ricerca, ha dedicato molto tempo a riscrivere intere parti del loro sistema operativo, nell'ottica della scalabilità del motore e del servizio.<sup>17</sup> Quando Wisenut si accorge di avere nel suo indice più pagine di uno stesso sito, la lista dei risultati presenta un link tipo "See 51 more pages from this site!"; la funzione equivalente in Teoma appare preferibile poiché la prima lista appartenente a ciascuno dei *topic groups* mostrati al top della pagina presenta il link "Related topic & experts' links" che permette di "saltare" direttamente al *topic group* di interesse. In entrambi i motori i *topic groups* generati automaticamente e presentati al top della pagina (Wisenut li chiama "WiseGuide") sono una caratteristica chiave che presenta spesso una via veloce per raggiungere i documenti più rilevanti senza perdersi in centinaia o migliaia di risultati. Quanto alle somiglianze con Teoma, Wisenut presenta il principio ispiratore di Google *keep-it-simple look*, offre una sintassi di ricerca simile (*phrase matching*, "+" e "-", non supporta





gli operatori booleani) e presenta un set di gruppi tematici generati semiautomaticamente alla sommità della pagina dei risultati. Tra le differenze la più ovvia è la grandezza del database, in quanto Wisenut dichiara di avere già indicizzato oltre 1,5 bilioni di pagine, cosa abbastanza insolita per un nuovo motore appena apparso in rete: per quanto opinabile, il suo indice è sicuramente più grande di quello di Teoma.<sup>18</sup> Nella visualizzazione dei risultati ci sono invece notevoli diversità tra i due motori: Wisenut sembra usare come descrizione la prima parte di testo che trova in una pagina e ciò spesso produce risultati non proprio utili (per esempio una lista di link). Teoma invece usa, quando esiste, la meta tag <description> della pagina oppure estrae un frammento dal testo. Ad ogni modo, la differenza fondamentale tra i due motori consiste nel fatto che Teoma focalizza l'attenzione su un contesto puramente semantico, mentre Wisenut privilegia l'aspetto strutturale e statistico. In questo contrasto tra approccio strutturale *vs* contestuale, Wisenut genera i raggruppamenti della WiseGuide conteggi delle parole nell'ambito dei risultati di ricerca, mentre Teoma usa i *topic groups* per rappresentare concetti derivanti da un'analisi più sofisticata delle pagine dei risultati.

Nell'ambito degli strumenti che in rete attuano il clustering dei risultati, non si può non ricordare Vivísimo,<sup>19</sup> fondato nel giugno 2000 per sviluppare un prodotto commerciale da un algoritmo definito presso la Carnegie Mellon University.<sup>20</sup> Nato anch'esso dall'emergere in campo scientifico del problema della sovrainformazione in risposta alle query di ricerca, ha presentato un approccio nuovo al problema del raggruppamento automatico dei risultati.<sup>21</sup> In realtà, Vivísimo non è un motore di ricerca, poiché non registra né indicizza quanto

viene quotidianamente pubblicato sul web ma organizza i risultati forniti dagli altri motori sulla base delle seguenti procedure: in pochi secondi interroga uno o più motori di ricerca, analizza le pagine dei risultati per estrarre i documenti (titoli, url e brevi descrizioni), raggruppa i documenti in base a questi elementi e non all'intera pagina web, ordina i gruppi e i documenti all'interno di ogni gruppo,<sup>22</sup> mostra le categorie gerarchiche. Oltre che in lingua inglese, Vivísimo raggruppa documenti anche in olandese, francese e italiano, grazie all'inclusione di una *language-specific stoplist* (lista di parole non significative) e di un processo che riconosce significati simili tra varianti sintattiche. Il numero di gruppi fornito non è fisso ma viene di volta in volta determinato dalle relazioni che l'algoritmo trova all'interno del documento. Per citare il sito:

Vivísimo wants to change the way search results are displayed on computer screens everywhere. Organized search results with document clustering. No more long, tedious lists.<sup>23</sup>

In pratica la pagina dei risultati è divisa in due sezioni: la parte sinistra mostra un menu gerarchico di argomenti e sottoargomenti, mentre la parte destra mostra la normale lista dei documenti trovati che possono essere visionati aprendo la pagina in una nuova finestra del browser, a schermo pieno o in un riquadro che anticipa i risultati del sito. La parte alta della pagina presenta una barra di ricerca che consente di cercare all'interno dei risultati e il link Details che recupera una tabella riassuntiva dei risultati recuperati da ogni motore di ricerca, dei doppi e ulteriori informazioni. La funzione di ricerca avanzata offre molte opzioni,

inclusi singoli bottoni per selezionare/deselezionare i motori di ricerca, quale lingua utilizzare, quali informazioni visualizzare con i risultati e altro ancora.<sup>24</sup> Vivísimo applica la sua tecnologia ai principali motori di ricerca, quali Yahoo, AltaVista, MSN, Hotbot e Lycos<sup>25</sup> e inoltre accede a siti governativi come FirstGov.gov, che compila informazioni da tutte le agenzie federali, PubMed e diversi altri siti societari: le demo presenti nel sito mostrano come funziona il *clustering* con i motori, inclusi, oltre a quelli già menzionati, Stanford, IEEE, Scientific American, IBM, NASA, New York Times, eBay, e molti altri. Gary Price, coautore del libro *The invisible web* (2001) dalle pagine del magazine "Searcher", sostiene che

Unlike Northern Light folders, which are pre-mapped to a fixed taxonomy structured by librarians, Vivísimo creates its clusters on the fly using terms taken from the brief descriptions or snippets in the search results.<sup>26</sup>

Al termine di questo breve excursus sugli strumenti di rete che operano secondo nuovi approcci all'informazione, possiamo definire il processo di *clustering* un valore aggiunto per facilitare la vita agli utenti. Si tratta di un primo passo verso la gestione integrata dei documenti, con tutti i limiti del riconoscimento automatico del contenuto. Nello sviluppo di tali sistemi si può comunque leggere la tendenza a un'analisi più profonda dell'informazione disponibile in Internet, anche se tale approccio ➤

cio non è in grado di risolvere di per sé il problema di portare alla luce quanto resta sommerso nei website ricchi di informazione diversamente strutturata: le risorse sono considerate materiale informativo di natura stabile come e-magazine, data collection, e-book e documenti PDF, difficili da rintracciare nei grandi siti. ■

## Note

<sup>1</sup> <<http://www.cyveillance.com/web/newsroom/releases/2000/2000-07-10.htm>>. Per informazioni più dettagliate e aggiornate è interessante il sito di Security Space all'url: <[http://www.securityspace.com/s\\_survey/data/man.200203/growth.html](http://www.securityspace.com/s_survey/data/man.200203/growth.html)>, che in data 1 aprile 2002 ha pubblicato numerosi report relativi alle rilevazioni statistiche più disparate nel web.

<sup>2</sup> OCLC. OFFICE OF RESEARCH, *Web Statistics. Web Characterization Project*. <<http://wcp.oclc.org/stats.htm>>.

<sup>3</sup> HSICHUN CHEN, *Knowledge management systems, a text mining perspective*, Tucson Arizona, Knowledge Computing Corporation, 2001, p. 11, "None of the search engines completely covers the web... In fact each search engine may cover only 15% of the web at any given time".

<sup>4</sup> L'algoritmo di Page Rank usato da Google è un buon esempio di tecnica di analisi dei link: il punteggio è assegnato alle pagine in base ai link da altre pagine secondo il criterio "più link, più alta rilevanza".

<sup>5</sup> <<http://www.teoma.com/>>. Teoma nasce da un progetto di ricerca iniziato nel 1998 a Rutgers, The State University of New Jersey in New Brunswick, NJ, da un team di ricercatori guidati da Apostolos Gerasoulis, professore di Computer Science, con il supporto dell'Office of Corporate Liaison and Technology Transfer di Rutgers, Gerasoulis nel 1999 ha fatto sì che la Teoma Technologies sviluppasse la nuova generazione delle tecnologie di ricerca web sotto la licenza dell'università. Il professore Tao Yang dell'University of California, Santa Barbara ha aderito a Teoma nel 2000 come *chief scientist* per co-dirigere lo sviluppo della tecnologia.

<sup>6</sup> AskJeeves, famoso provider di tecnologie per la ricerca, ha chiuso il 1° aprile il sito web di Direct Hit, motore di ricerca che basava il suo meccanismo di ranking sulla popolarità *clickthrough*, misurando per esempio cosa i visitatori preferivano cliccare da una lista e per quanto tempo restavano nel sito (se tornavano a Direct Hit, ciò indicava una scarsa qualità o rilevanza). AskJeeves intende integrare la tecnologia Direct Hit in Ask.com e Teoma.

<sup>7</sup> Come in Google, in Teoma l'uso della *link analysis* per la produzione dei risultati è piuttosto massiccio, ma il sistema di Teoma misura i link solo all'interno di collezioni di documenti ritenuti rilevanti per una particolare ricerca, e non su tutte le pagine raggiunte nell'intera rete, come fa Google.

<sup>8</sup> <<http://digitalmass.boston.com>>.

<sup>9</sup> Ask Jeeves e Teoma offrono ora un *paid inclusion program*, che prevede che i siti vengano rivisitati ogni 7 giorni, intervallo eccessivo rispetto alle 48 ore suggerite da Pandia Search Central (<<http://www.pandia.com/searchworld/index.html>>), e mantenuti nel database per 15 mesi al prezzo di 18 US\$ per il primo url e 18 US\$ per il successivo. Strategicamente, tale scelta non appare conveniente dal momento che il traffico generato da Ask Jeeves non è quello di Yahoo! o di Google; inoltre, se Teoma comincia a lasciarsi sfuggire i siti che non sono disposti a pagare, la qualità dei risultati di ricerca ne risente. Il database gratuito resta ovviamente funzionante. (Pandia Search Engine Weblog, Feb 16 2002).

<sup>10</sup> OneStat.com (<<http://www.onestat.com>>), provider leader nei software di analisi del web in tempo reale, il 15 aprile 2002 definiva Google "currently the leading search engine on the web. Google has a global usage share of 46.5 percent" mentre Yahoo! è il secondo con uno share d'uso del 20.6%.

<sup>11</sup> <<http://www.theregister.co.uk/content/6/20614.html>>.

<sup>12</sup> Ha debuttato in versione beta nel luglio 2001 e nel settembre successivo nella versione definitiva (<http://www.searchengineshowdown.com/features/Wisenut/review.html>). La tecnologia di ricerca è ancora in attesa di brevetto.

<sup>13</sup> <<http://www.wired.com/news/technology/0,1282,45905,00.html>>.

<sup>14</sup> "San José Mercury News", 09 giugno 2001, <<http://www0.mercurycent>

[ter.com/premium/business/docs/tech-news06.htm](http://www0.mercurycenter.com/premium/business/docs/tech-news06.htm)>.

<sup>15</sup> Yeogirl sostiene che i risultati delle WiseGuide "are built 100% on-the-fly", senza alcun utilizzo di liste di keyword predefinite.

<sup>16</sup> <<http://searchenginewatch.com/searchday/01/sd1217-sesuccess.html>>.

<sup>17</sup> <<http://www.searchengineworld.com/newsletter/2001/Wisenut.htm>>.

<sup>18</sup> ZyBorg, lo spider di Wisenut era già in funzione nel giugno del 2000 e dunque potrebbe aver avuto il tempo sufficiente per costruire un indice molto esteso.

<sup>19</sup> <<http://www.vivisimo.com>>.

<sup>20</sup> <<http://dev.web.cs.cmu.edu:6666/testReleases/demo/42.html>>.

<sup>21</sup> L'organizzazione automatica di documenti in gruppi o cluster differisce dalle altre tecniche (classificazione, tassonomie, Northern Light ecc.) in quanto è completamente automatizzata: non c'è alcun intervento umano tranne nella scrittura dell'algoritmo di base. Il problema principale della *document clustering* è trovare velocemente gruppi significativi da annotare concisamente: Vivisimo presenta un nuovo algoritmo euristico in grado di farlo, già operativo su pagine web, abstract di brevetti, newswire e altre tipologie di formati documentali.

<sup>22</sup> Al livello gerarchico più alto viene conservato il ranking presentato dal motore di ricerca, se viene interrogato un solo motore; se ne vengono interrogati diversi, vengono presentate più in alto le pagine recuperate da più di un motore.

<sup>23</sup> <<http://websearch.about.com/library/weekly/aa070901a.htm>>.

<sup>24</sup> <<http://www.post-gazette.com/businessnews/20010628vivisimo0628bnp3.asp>>.

<sup>25</sup> Tra i motori utilizzati da Vivisimo non c'è Google: per vedere i risultati di Google per raggruppamenti piuttosto che per rilevanza, è possibile scrivere direttamente a: [suggestions@google.com](mailto:suggestions@google.com).

<sup>26</sup> <<http://www.infotoday.com/newsbreaks/nb020325-2.htm>>. Il 25 marzo 2002 The Institute of Physics Publishing (IOPP; <<http://www.iop.org>>), società editrice inglese che vanta record innovativi nel campo dell'editoria elettronica, ha annunciato di avvalersi di Vivisimo per migliorare il suo servizio di Electronic Journals, ed è possibile vedere l'applicazione in funzione.