

[Una svista rivela i segreti di Inktomi \(e di molti altri ...\)](#)

Lucio Miranda

Lucio.Miranda@inetitalia.com

www.inetitalia.com

copyright © 2002. Tutti i diritti riservati. Vietata la riproduzione, anche parziale, senza il consenso scritto dell'autore.

Il mondo dei motori di ricerca e' notoriamente permeato di mistero. Gli algoritmi di analisi e il loro impiego sono tra i segreti che vengono difesi con maggior tenacia da parte dei MR. Fa parte dell'eterno antagonismo esistente tra il mondo dei MR e "gli altri", i webmaster, coloro che sviluppano i siti e le societa' di SEO (Search Engine Optimization).

Di conseguenza e' molto difficile, se non impossibile, capire quali sono i criteri che i MR utilizzano nel lavoro di analisi dei siti. Quali sono gli elementi che vengono valutati? Che peso viene dato a ciascuno di questi elementi? Esiste un rapporto tra i vari elementi sottoposti a valutazione di modo che uno stesso elemento (la lunghezza del testo, per esempio) assume una valenza diversa in funzione degli altri elementi cui si accompagna (la numerosita' delle keyword per esempio)?

Sono domande destinate a non ricevere mai risposte concrete e dirette e ci si deve accontentare delle dichiarazioni carpite ai vari funzionari/ingegneri in occasione di interviste o convegni dedicati alle tecniche di SEO. Si tratta sempre di mezze parole, di frasi che lasciano piu' dubbi che certezze e che si prestano a interpretazioni le piu' varie.

Insomma, esiste un muro di omerta' difficile da penetrare che protegge il *modus operandi* dei vari MR. A volte, raramente, questo muro mostra delle crepe e nel lasso di tempo che occorre ai MR per riparare il muro e' dato sbirciare all'interno del giardino ... E' quello che e' successo in po' di tempo fa ad Inktomi. Un documento che riportava nel dettaglio cio' che Inktomi considera pratiche di spamming e' stato reso accessibile online. Brett Tabke, di Search Engine World, lo ha trovato casualmente mentre su FAST cercava documenti relativi all'argomento "editorial database".

Ne e' scaturito un articolo che ha fatto il giro della comunita' dei SEO in un lampo. A seguito della fuga di notizie Inktomi ha subito ritirato il documento che non e' al momento piu' visibile pubblicamente. In pratica ha riportato il documento nel sicuro ovile dell'intranet aziendale dopo la sua scappatella nel libero universo di Internet.

Quello che segue e' un riassunto di quanto scoperto da Brett. Incidentalmente (e questo getta luce su come agiscono i MR) il documento conteneva anche una lista di societa' di SEO che erano state "bandite" da Inktomi. Dal documento si desumeva che, in pratica, se Inktomi veniva a scopriva che un certo sito poteva essere penalizzato per il semplice fatto di essersi fatto aiutare da una delle societa' di SEO che erano state bandite. Questo indipendentemente dal fatto che il sito avesse posto in essere delle pratiche di spamming.

D'altronde il coltello dalla parte del manico ce l'hanno i MR ed e' chiaro che spostando sempre di piu' i criteri di indicizzazione su parametri *off page* i MR hanno anche spostato l'ago della bilancia pesantemente a loro favore. Non

rimane molto da fare quindi se non tentare di capire ed adattarsi alle regole del gioco. (oppure dichiarare di *non* capire ed adeguarsi comunque .. ;-).

Tecniche considerate essere di spamming da parte di Inktomi

Tecnica di spamming	Definizione
DNS Spamming	Livelli multipli di domini di terzo livello che incorporano keyword. Esempio: keyword1.dominio.com; keyword2.dominio.com etc.
Keyword stuffing dei subdominii	Domini di terzo livello con piu' di una keyword che riproducono pagine sostanzialmente simili. Esempio: keyword1.keyword2.dominio.com. Il concetto e' che spesso tecniche di questo tipo sono usate per creare pagine doorway duplicate l'una dell'altra ad eccezione di poche keyword. Se pensate di usare una struttura del genere fatelo con pagine che hanno del contenuto rilevante e che non abbiano duplicati.
Keyword stuffing delle directory	Esempio: www.dominio.com/keyword. Inktomi prende in considerazione i numeri IP e analizza i siti che presentano un numero troppo elevato di subdirectory
Redirect	Pagine che ridirigono su altre pagine. Il monitoraggio e' effettuato analizzando indirizzo IP e DNS server della pagina che re-indirizza e di quella re-indirizzata.
Keyword di riflesso	Esiste la possibilita' per il server di creare al volo delle pagine di testo sostanzialmente uguali ma dove le keyword cambiano in base alla richiesta fatta al server da parte del MR.
Cloaking	<p>Le pagine cloaked vengono riviste manualmente e se il contenuto della pagina cloaked differisce significativamente da quello delle pagine visibili dal pubblico, il sito viene penalizzato. Inktomi analizza le pagine cloaked usando un IP non-Inktomi.</p> <p>Questa e' un'indicazione importantente che conferma quello che si sospettava da tempo senza averne mai avuto conferma: le pagine cloaked vanno bene e sono tollerate. L'importante e' non abusarne per fare spamming dell'indice. In altre parole: usare le tecniche di cloaking rimanendo fedeli al contesto del sito per generare pagine che siano friendly nei confronti dei MR></p>
Pagine duplicate	La produzione di centinaia di pagine essezialmente l'una uguale all'altra eccetto per poche keyword.
Falsi MetaTags	L'uso di meta tags che di fatto non trovano poi riscontro sul testo della pagina HTML.
Testi Body	Quando il testo include delle keyword che di fatto non riflettono il contenuto del resto del testo, la pratica e' considerata spamming. Idem nel caso che il testo sia invisibile (qui c'e' addirittura il dolo ..)
Link stuffing	Creare pagine di soli link che, usando keyword diverse, puntano tutte pero' alla stessa pagina nel tentativo di aumentarne la link popularity.

Tecniche di monitoraggio dello spamming attuate da Inktomi

Indirizzo IP	Analisi dell'indirizzo IP e dell'IP subnet per eliminare tentativi di spamming ripetuti da parte dello stesso trasgressore.
Indirizzi DNS	Indirizzo IP, DNS server e dati di registrazione del dominio vengono usati in congiunzione per monitorare le azioni di trasgressori abituali.
Servizi di redirect e link farm	Alcuni siti, tipici i ".to", sono spesso usati da chi fa spamming dell'indice. Altri siti sono invece delle link farm usate solo per generare link popularity a favore di altri siti (le link farm sono pagine html con migliaia di link a pagamento).
MD5	MD5 sta per "message digest number 5". Si tratta di un algoritmo di controllo che viene usato da Inktomi per scovare pagine duplicate (quelle in cui il testo non cambia ad eccezione di poche keyword).
Elevato numero di richieste di registrazione	Se si esagera col numero di richieste di registrazione all'indice si rischia di essere penalizzati. Stiamo parlando di centinaia di richieste al mese, per cui non dovrebbero esserci problemi per tutti coloro che registrano il proprio sito in buona fede.

Altre scoperte interessanti

Il documento scappato dall'intranet rivelava altri particolari. Eccone alcuni tra i piu' interessanti:

- Inktomi puo' decidere di penalizzare siti che giudica poco importanti/interessanti (decidono loro ..) quando la categoria cui appartengono e' troppo affollata.
- "I cavalieri Bianchi". Sono siti che a giudizio di Inktomi sono di particolare valore. In quanto tali non vengono mai penalizzati.
- Partner di Inktomi quali HotBot o MSN possono chiedere la penalizzazione o la rimozione di particolari siti a loro insindacabile giudizio. Che si stia formando un cartello? O una mafia?