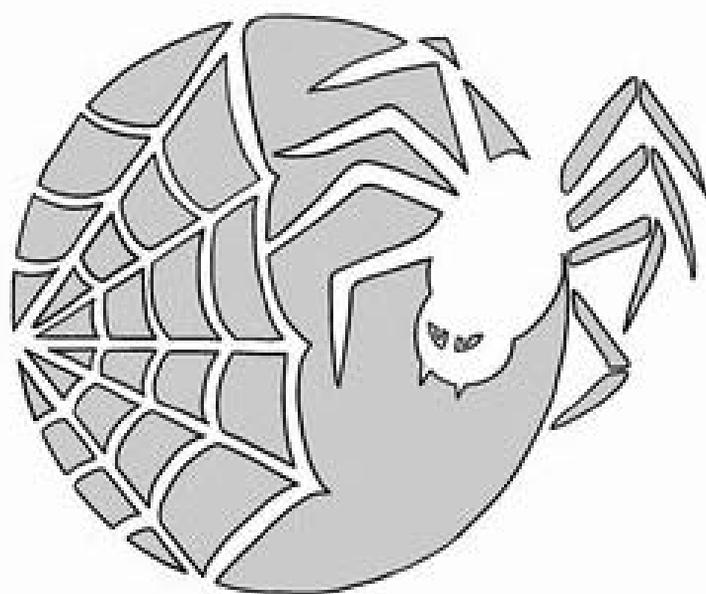

Tutto quello che gli altri non ti dicono

Posizionamento nei motori di ricerca

Francesco Gabriele Ricci



POSIZIONAMENTO NEI MOTORI DI RICERCA

Tutto quello che gli altri non ti dicono

Francesco Gabriele Ricci

Posizionamento nei motori di ricerca

Tutto quello che gli altri non ti dicono

Francesco Gabriele Ricci

Email: francesco@ricci.in

Telefono: 340-2734425

Sito Web: www.libroposizionamento.com

www.forumposizionamento.com

Tutti i diritti sono riservati a norma di legge e a norma delle convenzioni internazionali. Questo libro non può essere riprodotto in nessuna parte senza l'autorizzazione scritta dell'autore e/o editore.

Finito di stampare nel mese di Ottobre 2009

SOMMARIO

Prefazione	1
Presupposti per indicizzare un sito web	4
Come scegliere il giusto nome a dominio?	5
L'hosting incide sul posizionamento?	13
La scelta delle parole chiave del sito	25
La scelta delle parole chiave negli url	34
I sottodomini sono migliori delle sottodirectory? ..	44
Dettagli Tecnici	50
I meta tag	51
Il file robots.txt	59
I tag fondamentali	65
La sitemap	72
I feed possono aiutare nel posizionamento?	81
Come ottimizzare un sito in Flash	83
Utilizzare il redirect 301 Moved Permanently in modo corretto	88
URL Rewriting: come e perché	95
I Fattori	101
Il PageRank	102
Il Trust Rank	109

Il Sandbox	115
La link popularity	120
L'importanza dei contenuti aggiornati	124
Trucchi da lasciar perdere	128
L'importanza di DMOZ	132
Il rischio della over-optimization	135
Il mantenimento del posizionamento	139
Conclusione	142

Prefazione

Ho sempre pensato che uno scrittore debba essere il più imparziale possibile perchè non sta scrivendo qualcosa per se stesso ma per gli altri, purtroppo non è sempre così e a malincuore ho troppo spesso dovuto leggere libri di parte.

In queste pagine quindi non troverete nessun riferimento ad un motore di ricerca specifico, tutto questo sia per una questione di pluralismo, sia per dare modo al lettore di crearsi una cultura globale sull'argomento e renderlo capace di scegliere verso quale motore di ricerca puntare.

Inoltre nel momento in cui viene scritto questo libro ci troviamo di fronte al dominio assoluto di Google, all'ascesa di Bing, alle trattative di una coalizione Microsoft-Yahoo e alla nascita di nuovi motori di ricerca sulla base di piccole idee innovative, scrivere quindi parlando anche di uno solo di questi motori avrebbe creato un prodotto di scarso valore per il futuro.

La mia idea di realizzare un libro sul posizionamento nei motori di ricerca è nata dal fatto che questo è un argomento molto sentito soprattutto in quest'ultimo periodo nel quale i nomi a dominio convincenti iniziano a scarseggiare e il fattore maggiore del successo di un'attività online è l'essere in vetta nei motori di ricerca per non restare anonimi.

Gli utenti in internet effettuano ogni giorno milioni di ricerche e il gioco sta nel "farsi trovare" facendo visitare il proprio sito e non quello della concorrenza, ed è tutta una questione di ascesa verso la popolarità del sito o di declino nell'anonimato dove nessuno ti conosce e nessuno ti visita. Ho visto più volte nella mia vita dei progetti validissimi fallire solamente perchè non riuscivano a diventare popolari, e la chiave di tutto era il loro posizionamento nei motori di ricerca.

E' vero che esistono altri libri del genere, ma questo è diverso dagli altri, infatti non si ferma sui soliti discorsi teorici ma insegna cosa realmente bisogna scrivere, quali passi bisogna compiere, quali pezzi di codice possono aiutare nel posizionamento di un sito web, in modo da consentire anche ai non esperti di ottenere ottimi risultati in poco tempo.

Questo libro è rivolto a chiunque sia appassionato del settore indipendentemente dal suo livello di preparazione, infatti è una valida risorsa per essere utilizzato sia come strumento d'apprendimento per webmaster neofiti, sia come strumento d'aggiornamento per chi è già un abile professionista SEO.

Per facilitare e rendere piacevole la lettura il libro è stato strutturato con una serie di domande alle quali spero di

aver risposto nel modo più semplice e completo possibile.

Buona lettura,
Francesco Gabriele Ricci

PRESUPPOSTI PER INDICIZZARE UN SITO WEB

Come scegliere il giusto nome a dominio?

La scelta del nome a dominio è, o meglio dovrebbe essere, la prima fase per la realizzazione di un sito web anche se nella maggior parte dei casi questa fase viene tralasciata o gli si dedica troppo poco tempo.

Tutto questo accade perchè ad oggi poche persone capiscono la sua reale importanza, invece bisogna sapere che è un fattore discriminante non nel successo o nell'insuccesso ma per lo meno nel maggiore o minore successo.

Questo sarà quindi il nostro primo punto di forza verso una fantomatica concorrenza che penserà subito alla parte implementativa del sito trascurando questo ed altri aspetti che, permettetemi la metafora, dobbiamo immaginare come fondamenta di una casa che quasi non si notano ma sono importantissime per quello che ci andremo a costruire sopra.

Prima di iniziare a parlare dei nomi a dominio è meglio fugare ogni dubbio, infatti molti erroneamente pensano che `nomesito.com` sia un dominio di primo livello, ma così non è, infatti `google.it`, `yahoo.com`, `bing.it`, etc... sono tutti dominio di secondo livello, i domini di primo livello sono invece `.it`, `.com`, `.biz`, etc... ovvero le cosiddette estensioni.

E' così difficile trovare un nome a dominio libero?

A parole ci può sembrare semplice ma nella pratica non lo è di certo, e più passano le ore, i giorni, i mesi e gli anni e più diventa difficile perchè il numero di domini registrati cresce e questo fenomeno non accenna a fermarsi, anzi è in costante aumento.

Ad aggravare la situazione già di per sè critica ci si mettono anche delle società che lucrano sulla registazione e sulla successiva vendita a prezzi maggiorati degli stessi.

Su questo i pareri su internet sono discordanti tra chi condanna queste società e chi le appoggia, ma tutt'altro discorso è quando qualcuno si comporta in modo del tutto scorretto, ed è questo il caso del cosiddetto cybersquatting, ovvero il fenomeno di registrazione di nomi di dominio corrispondenti a marchi altrui o a nomi di personaggi famosi al fine di realizzare un lucro tramite la vendita.

Questa è sicuramente una pratica scorretta che ha portato per altro il governo degli Stati Uniti d'America ad emanare nel 1999 una legislazione ad hoc chiamata Anticybersquatting Consumer Protection Act.

Ma la questione che rimane aperta, è fino a che punto la registrazione di un nome a dominio può risultare corretta o scorretta, e per capire ciò potremmo far riferimento alla vita reale: Chi compra una casa, un terreno, un'automobile con

l'intenzione di rivenderla compie un reato? E' una persona scorretta?

Oggettivamente è nella cultura umana comprare qualcosa per rivenderla, ma dobbiamo applicare tali leggi anche nel web?

Per ovviare a questo problema l'ICANN (l'ente preposto alla gestione del sistema dei nomi a dominio di primo livello) ha aumentato il numero di estensioni di primo livello creando i domini .info, .name, etc... non riuscendo però a risolvere il problema perchè oramai l'estensione principale di internet è il .com, ed è quello che ogni persona vuole registrare seguita dalla propria estensione nazionale es. .it, .es, .de, etc...

Comprare estensioni come .org, .info, .net, .biz, etc... ha senso solo se per quel nome si ha a disposizione anche il .com e si vuole marcare bene il territorio, ogni altro acquisto sarebbe una spesa quasi inutile che porterebbe al disorientamento degli utenti e nei casi peggiori (ma non impossibili) portare traffico ad un sito concorrente e perdere potenziali utenti.

Dopo questa piccola introduzione al naming possiamo passare alla prima regola della scelta di un nome a dominio, questo infatti deve essere innanzitutto attraente per l'utente, e questo termine non sta a significare bello o alla moda ma semplicemente che deve contenere il numero minore di "difetti".

Ma cosa sono questi difetti?

Vi starete tutti chiedendo come può un nome avere dei difetti, la risposta è presto detta: ad esempio un nome a dominio eccessivamente lungo è un difetto e spesso questo porta anche ad una difficoltà nella memorizzazione essendo la lunghezza di una parola o frase inversamente proporzionale alla sua capacità di essere ricordata.

Ma non dobbiamo fare di tutta l'erba un fascio, infatti un dominio lungo (di ben 48 lettere) e apparentemente difficile da ricordare come sottolapancalacapracantasottolapancalacapracrepa.com, essendo uno scioglilingua conosciuto da tutti risulta molto più facile da ricordare di un dominio lungo la metà come ad esempio rossopatatemareluna.com che non ha un nesso logico fra i termini oppure asd3jss9ccadas23.com che non ha alcun senso logico.

A parte questi casi limite è ovvio che bisogna trovare il giusto compromesso.

Alcune persone hanno cercato di risolvere il problema utilizzando degli acronimi, soluzione accettabile ma non ottima in quanto a volte alcuni acronimi erano difficili da ricordare.

Tuttavia questa soluzione ha pagato solo nel breve periodo, infatti ad oggi sono stati registrati tutti i .com di 3 lettere e siamo verso la saturazione anche di quelli a 4 lettere.

Un'altro grave difetto da tenere in considerazione è il mistyping tradotta liberamente dall'inglese: scrivere in modo errato; in altre parole il problema che si viene a creare quando ci si ricorda il nome ma non ci si ricorda come si scrive, questo fenomeno è cresciuto con i dilagare degli inglesismi nei nomi a dominio italiani.

Alcuni esempi di questi difetti sono:

- **domini con la doppia “o” che si legge “u”**
- **domini con numeri al posto di parole**
es: `kisstoyou.com` -> `kiss2you.com`,
`kissforyou.com` -> `kiss4you.com`
- **domini con linguaggio degli sms**
es: `givemeone.com` -> `gimmeone.com`,
`allforyou.com` -> `allforu.com`
- **domini con trattini**
es: `ricette-cucina-italiana.com`

La tecnica del mistyping però è anche una delle più vecchie tecniche di posizionamento nei motori di ricerca, tramite questo metodo infatti il webmaster punta ad ottenere utenti confidando nelle loro ricerche errate generate per ignoranza (ad esempio "giochi flesh" invece di "giochi flash") oppure per distrazione (ad esempio “tuto tondo” invece di “tutto tondo”). Proprio a partire da questa tecnica sono iniziate le registrazioni di nomi simili a società famose come

www.google.com ; yaho.com; altavita.com; microsoft.com etc...

Questa tecnica è stata esposta per una conoscenza prettamente teorica e viene altamente sconsigliata a chiunque abbia intenzione di creare un sito serio.

Un difetto grave che un nome a dominio può presentare è la mancanza dei principali domini di primo livello. Come precedentemente detto infatti come estensione per eccellenza del web si è affermata il .com seguita per ogni nazione dal suo dominio nazionale, .it per l'italia.

Registrare un dominio .biz può avere senso se e solo se si è già proprietari del .com, infatti gli utenti tendono a ricordare i nomi dei domini e non la loro estensione, quindi digiteranno nella barra degli indirizzi del browser il nome del tuo sito seguito da .com o .it, lasciando per ultimo il .biz sperando che non si siano stufati di fare tentativi o che non siano finiti su un sito della concorrenza.

Ultimo difetto che prenderemo in considerazione è quello dei nomi a dominio con dei trattini.

La moda dei trattini nacque, a differenza di come molti pensano, non in relazione alla caccia dei nomi a dominio, bensì con la credenza (in passato del tutto plausibile) che i motori di ricerca privilegiassero i domini separati da trattini nel caso in cui quelle parole venissero ricercate in una query.

Ad oggi la questione resta aperta tra fautori e detrattori del trattino nelle url anche perchè non esiste alcuna linea guida da seguire che sia stata pubblicata ufficialmente da parte di qualche motore di ricerca. Guardando comunque le serp (Search Engine Page Results) dei motori di ricerca viene da pensare che forse una volta le cose stavano effettivamente così, ora non più.

Quest'ultimo difetto ci ha portato ad introdurre un altro aspetto importante, infatti finora ci siamo solamente focalizzati sull'attrazione del dominio sugli esseri umani, ma non è da sottovalutare l'attrazione che può esercitare verso le directory e i motori di ricerca stessi.

Nella maggior parte delle directory (tra cui DMOZ tanto per citare la più famosa) i siti vengono elencati in ordine alfabetico, ed essendo cosa nota che i primi risultati vengono sempre più cliccati rispetto agli altri, la scelta di un nome a dominio che inizia con la lettera A o con una delle prime lettere dell'alfabeto potrebbe pagare.

Per quanto riguarda i motori di ricerca invece i siti che verranno privilegiati saranno sicuramente quelli che contengono nel nome a dominio parte delle parole ricercate, ad esempio se cerchiamo “pasta artigianale” i motori di ricerca riterranno (con riserva d'errore) più attinenti nomi a dominio che contengono la parola pasta rispetto ad altri.

Non ci dobbiamo però far condizionare troppo da questo consiglio, infatti è vero che seguendo questa

tecnica possiamo essere favoriti, ma non seguirla non significa di certo venire penalizzati, basta pensare al sito barilla.it che non contiene di certo la parola “pasta” eppure si trova posizionata bene in tutte le query del settore.

In conclusione sembra chiaro che dobbiamo puntare a cercare un nome breve, mnemonico, facile da scrivere, attinente ai contenuti del sito e “attraente” per gli esseri umani e per i motori di ricerca.

L'hosting incide sul posizionamento?

I fattori che influiscono sul posizionamento di un sito web in un motore di ricerca sono molteplici, si parla di circa 200, e tra questi deve essere considerato anche l'hosting a cui abbiamo deciso di appoggiarci per pubblicare le nostre pagine web.

Nonostante sia uno dei momenti più importanti per l'avvio di una attività sono pochi i webmasters che prendono realmente in considerazione il fattore hosting, senza rendersi conto che stanno trascurando un aspetto molto importante che può determinare in maniera pesante l'indicizzazione e nel peggiore dei casi dare anche luogo a penalizzazioni.

Spesso chiunque decide di avviare un'attività su internet, mosso da ragioni economiche, si appoggia ad un hosting economico con l'intenzione di passare successivamente ad un qualcosa di migliore anche a seconda delle necessità che si verranno a creare in quanto non sempre è necessario ospitare un piccolo sito in crescita su dei server dedicati.

Questa scelta è del tutto plausibile, d'altro canto però spendere tempo e denaro per poi affidare il nostro sito ad un hosting non all'altezza non è di certo una scelta intelligente.

Inoltre dobbiamo tenere conto che uscire da una penalizzazione di un motore di ricerca è molto difficile ed è un processo molto lungo, quindi è bene scegliere dall'inizio un hosting serio ed affidabile per non

compromettere il nostro duro lavoro.

Quali fattori differenziano gli hosting?

Di seguito andremo ad analizzare 5 fattori che differenziamo un hosting ottimo da uno mediocre e le ripercussioni che possono avere sul posizionamento.

Uptime server

Un hosting non affidabile è soggetto a errori e crash che rendono irraggiungibile il sito per medio-lunghi periodi. Oltre al fatto che un server down non è un buon biglietto da visita per i propri utenti, se lo spider del motore di ricerca dopo diversi tentativi non sarà in grado di recuperare le pagine dal tuo sito web interromperà la visita e diminuirà la frequenza di scansione progressivamente fino ad arrivare addirittura ad una perdita di indicizzazione del sito se il down si protrae per qualche giorno.

Tutto questo avviene perchè i motori di ricerca hanno tutto l'interesse a proporre nelle prime posizioni ai propri utenti siti web affidabili e raggiungibili.

Entrando più nel dettaglio, lo spider durante la sua normale attività di ricerca trovando un server che non risponde riefetterà le stesse richieste a distanza di qualche decina di minuti l'una dall'altra per verificare se

si tratta di una disconnessione di routine o di un vero e proprio down. Se il server risponderà a queste richieste tutto tornerà alla normalità, in caso contrario lo spider inizierà a far visita al nostro sito dopo un tempo via via maggiore e di pari passo penalizzerà il nostro sito facendolo scendere nella SERP in quanto non riterrà proficuo proporre ai propri utenti un sito irraggiungibile. Se il down dovesse durare diversi giorni, il sito verrebbe considerato “morto” e ci sarebbe l'esclusione dal motore di ricerca. Questa è la punizione peggiore che può infliggere un motore di ricerca, infatti una volta esclusi non saremo più presi in considerazione e per essere presenti nuovamente sul motore di ricerca dovremo fare una richiesta di indicizzazione come se il sito fosse nuovo e ripartire da zero.

Tuttavia non c'è bisogno di un down di qualche giorno per essere penalizzati, infatti anche un sito che ha down di pochi minuti ma frequenti, in quanto magari non riesce a reggere il carico nelle ore di punta, non è ben visto dai motori di ricerca perchè viene considerato non del tutto affidabile, un sito del genere non potrà di certo ambire alla prima posizione.

La quasi totalità degli hosting odierni ci garantisce un uptime del 99,9% che fino a qualche anno fa era un ottimo traguardo, ma al giorno d'oggi quasi 9 ore di down all'anno, ovvero 1 minuto e mezzo al giorno, che sembrano un'inerzia possono invece fare la differenza.

Alcuni hosters hanno quindi deciso di fare di più e garantire un uptime del 99,99% che significa solamente

52 minuti all'anno di down, con un server del genere si possono veramente dormire sonni tranquilli e sembra la scelta più giusta da fare affidandosi a degli esperti del settore.

Ma come sempre c'è chi vuole strafare arrivando a proporre un uptime del 99,999% o “a 5 nove” detto in gergo tecnico, ovvero 5 minuti e 16 secondi all'anno di downtime che sono solamente 9 decimi di secondo al giorno in media.

Tale obiettivo sembra più una trovata pubblicitaria che un vero e proprio servizio in quanto 5 minuti non sarebbero sufficienti neanche per un cambio di hard disk o ram in caso di guasto, ad ogni modo anche se tutto ciò fosse vero entrerebbe in gioco il fattore economico: un servizio del genere ha dei costi esorbitanti e ci fornirebbe un'affidabilità maggiore di quella di cui abbiamo realmente bisogno, quindi forse sarebbe più opportuno spendere questo denaro in altri modi.

Tempi di risposta del server

Fino a qualche anno fa le connessioni non erano veloci come quelle odierne, quindi non ci si è mai posti il problema della velocità di caricamento delle pagine web, oggi però le cose sono cambiate e i motori non potevano non prendere in considerazione questo parametro.

Quindi anche se un server avrà un'uptime alto, per

assurdo anche del 99,999%, ma caricherà le pagine lentamente verrà penalizzato.

I discorsi trattati precedentemente per l'uptime di un server valgono anche per un server lento che può compromettere seriamente il giudizio del motore di ricerca nei nostri confronti e può portarci dapprima ad una penalizzazione fino ad arrivare all'esclusione in quanto saremo considerati di bassa qualità e non adatti ad essere proposti agli utenti.

Per farci un'idea della qualità del nostro hosting dobbiamo andare a controllare più volte durante tutto l'arco della giornata, soprattutto nelle ore di punta, le prestazioni del nostro server tenendo d'occhio il tempo impiegato per il download di una pagina.

Per fare ciò possiamo appoggiarci ad uno dei tanti tool presenti in rete oppure su linux possiamo lanciare da shell un comando del tipo *"time wget www.sito.com"*.

Un hosting accettabile dovrà rientrare nei seguenti valori (tempi per il download di una pagina espressi in millisecondi):

Tempo massimo: non superiore a 2.000

Tempo medio: non superiore a 1.000

Tempo minimo: non superiore a 500

Se i vostri parametri non rientrano in questi range è bene correre ai ripari prima che sia troppo tardi contattando il vostro hoster e/o iniziando a pensare seriamente di cambiare hosting.

Indirizzo IP condiviso

Una domanda che dobbiamo sempre porci è: chi sono i nostri vicini?

La quasi totalità dei siti web infatti condivide lo stesso server con altri siti, infatti sulla stessa macchina vengono ospitati anche centinaia di altri siti web, questo viene chiamato hosting condiviso. Di conseguenza il nostro sito potrebbe venire penalizzato solo perchè sulla stessa macchina sono ospitati altri siti sgradevoli o di dubbia qualità ai motori di ricerca, e ne bastano solamente 2-3 per determinare una penalizzazione a livello di indirizzo IP.

In parole povere se un hoster ospita nei propri server dei siti spam e voi malauguratamente finite su una di queste macchine verrete, anche se ingiustamente, penalizzati nonostante il vostro sito web sia pulito e ottimizzato.

Per capire meglio il ragionamento fatto dai motori di ricerca possiamo usare una metafora, se voi costruite una splendida villa con piscina, giardino, campi da tennis e chi più ne ha più ne metta, ma questa si trova in un quartiere malfamato, proprio a causa dei vostri “vicini” il valore della vostra casa, e anche la vostra reputazione, verrebbero seriamente compromessi.

La soluzione più ovvia è, come detto in precedenza, appoggiarsi ad un server dedicato o virtuale, soluzione però non sempre alla portata di tutti. Quindi per ovviare a questo problema la soluzione più veloce ed efficace è quella di richiedere al nostro hoster un indirizzo IP

dedicato al nostro sito in modo che i motori di ricerca vedano il nostro sito web come un qualcosa di a sé stante.

Avere un indirizzo IP per ogni sito è altamente consigliabile anche per non incappare in penalizzazioni causate, anche se in buona fede, da noi stessi. E' del tutto naturale inserire un link fra tutti i nostri siti web per pubblicizzarli o aumentare la link popularity, ma i motori di ricerca considerano in modo negativo siti che si autolinkano che risiedono sullo stesso indirizzo IP, ma anche facente parte della stessa classe C (ovvero le ultime 3 cifre dell'indirizzo).

Quindi è consigliabile distribuire i nostri siti web su indirizzi IP completamente diversi per evitare la malaugurata situazione in cui i motori di ricerca considerino i nostri siti web uno spam network, ovvero una serie di piccoli siti facente capo ad un unico grande sito sul quale convogliare tutto il traffico. Inutile dire che un motore di ricerca si farà un'idea del genere sui nostri siti web questi verranno penalizzati e/o bannati tutti.

Se non abbiamo le capacità per gestire più indirizzi IP sui nostri siti web, una soluzione adatta a tutte le persone che non abbiano grandi capacità tecniche è quella di effettuare delle ricerche e controllare quale società offre hosting a tutti quei siti che sono ben posizionati, così avremo la certezza che l'hoster al quale si appoggiano è di qualità e non ospita siti spammer.

Localione geografica del server

La localione geografica gioca un ruolo marginale nel determinare la lingua del sito web e per il target di utenza. Acquistare quindi un hosting negli Stati Uniti per risparmiare qualche soldo potrebbe non essere una scelta del tutto vantaggiosa.

Dal punto di visto dell'utente un server hostato negli Stati Uniti o in Italia non comporta alcun cambiamento, anzi neanche ci si fa caso, ma il discorso cambia per i motori di ricerca che hanno bisogno di capire di che nazionalità sono i siti web che visitano e a qualche target di utenza fanno riferimento.

Utilizzando le versioni localizzate (ovvero entrando nel .co.uk invece che nel .it), oppure restringendo la ricerca ai solo siti inglesi o francesi, i motori di ricerca hanno bisogno di rivalutare completamente quali siti web e in quale posizione mostrarli cercando di restituire i risultati più pertinenti possibili ad ogni singolo utente in ogni singola nazione.

Il primo metodo è stato quello di basarsi sul TLD (in inglese top-level domain, è l'ultima parte del nome di dominio internet, ovvero l'estensione) del sito web, ma col passare del tempo tale tecnica è diventata inaffidabile, infatti in alcune TLD come .com, .biz, .info sono iniziati a spuntare siti in tutte le lingue del mondo.

Attualmente ogni motore applica una propria logica, condivisibile o meno che sia, ma la tecnica più gettonata è quella di dare ragione prima al TLD e poi all'IP in casi di ambiguità.

Google però si è voluto contraddistinguere da tutti gli altri e ha fatto di più dando nel proprio Webmaster Tool la possibilità ai webmasters di selezionare il target geografico dei propri siti.

Di pari passo per aiutare i motori di ricerca in questo lavoro sono stati proposti degli standard non formalizzati per rappresentare i metadati geografici all'interno delle pagine web.

Il primo proposto da GeoUrl

```
<meta name="ICBM" content="latitudine, longitudine" />
```

che tiene conto solamente della latitudine e longitudine. Una versione più avanzata invece è stata proposta da GeoTags

```
<meta name="geo.position" content="latitudine;longitudine" />  
<meta name="geo.placename" content="Nome città" />  
<meta name="geo.region" content="Codice regionale" />
```

come possiamo vedere oltre alla latitudine e longitudine contiene anche il nome della città ed il codice della nazione e provincia.

Qui di seguito proponiamo un esempio per capire meglio il funzionamento di entrambe le soluzioni ipotizzando di voler geotaggare un sito web residente a Roma.

```
<meta name="ICBM" content="41.8961, 12.48458" />  
<meta name="geo.position" content="41.8961;12.48458" />  
<meta name="geo.placename" content="Rome" />  
<meta name="geo.region" content="IT-RM" />
```

Sulla scia di questa innovazione sono stati proposti anche metodi per localizzare i feed RSS, ma come per le pagine web si devono reputare come pseudostandard ancora in fase di discussione e per questo di scarsa incidenza sul posizionamento.

Impostazioni e configurazioni del web server

Esistono purtroppo web server di scarsa qualità o configurati male che non rispondono correttamente alle richieste e non restituiscono in modo appropriato i messaggi d'errore così come previsto dal protocollo http. Tutto questo provoca un disorientamento da parte degli spiders dei motori di ricerca che spesso non sanno come interpretare tali dati e altre volte, e sono forse i casi peggiori, li interpretano in modo errato.

La prima cosa da verificare quindi è che il web server ritorni ad ogni richiesta un “codice di stato “ (o meglio status code) sia in caso di successo che insuccesso.

Può sembrare un discorso complicato, ma in realtà è molto semplice, infatti a tutti quanti sarà capitato mentre navigava su internet di trovarsi davanti un errore 404 Not Found ovvero una pagina non trovata, bene

quello era uno status code di errore.

Per capire la loro importanza basta immaginare che il nostro sito web non abbia il file robots.txt (ovvero il primo file che visita uno spider per recuperare le informazioni su come visitare le pagine), se il nostro webserver non comunicherà in modo appropriato che il file in questione non esiste, gli spiders gireranno a vuoto e poco dopo abbandoneranno il sito in quanto stanno capendo poco o nulla di quello che si trovano attorno.

Gli status code sono oltre 50 e vengono inseriti nell'header della pagina come codice di controllo tra il browser e il webserver. Essi vengono raggruppati in 5 macrocategorie

- 1xx Informazioni
- 2xx Successo
- 3xx Redirezione
- 4xx Errore Client
- 5xx Errore Server

I più famosi sono sostanzialmente 4, l'errore 404 che come precedentemente detto ci permette di segnalare una pagina non trovata, l'errore 500 che segnala un problema interno al server spesso risolvibile aspettando qualche minuto e riprovando, l'errore 403 Forbidden che ci dice che non abbiamo i privilegi per accedere a tali directory o pagine ed infine l'errore 301 il quale indica che la risorsa cercata non è più disponibile ed è stata spostata altrove, in un'altra pagina o addirittura in

un altro dominio.

Per quanto riguarda gli errori generati dai vari linguaggi di programmazione che potrebbero venir fuori utilizzando ad esempio un CMS non compatibile con il server, è bene sostituirli con dei testi personalizzati, magari che spieghino anche agli utenti cosa è accaduto, altrimenti gli spiders si ritroveranno a leggere la sintassi di tali errori dando un peso negativo alle pagine coinvolte.

La scelta delle parole chiave del sito

Quando si avvia un'attività online è molto importante scegliere le keywords giuste in quanto l'analisi dei motori di ricerca si basa proprio sui testi contenuti nel nostro sito web.

I testi presenti nel nostro sito vanno quindi redatti in maniera del tutto naturale e nel modo più semplice possibile senza incappare nella cosiddetta “sindrome da keyword”, ovvero il tentare di inserire a tutti i costi le parole chiave nel testo a discapito della qualità e della leggibilità.

Nel tempo anche i motori di ricerca si sono evoluti e così come una persona riesce a comprendere se un testo è stato scritto con naturalezza, essi tramite calcoli basati sulla frequenza delle parole e la percentuale sull'intero testo riescono a determinare la qualità di una pagina.

Oltre la qualità di un testo va considerato che la presenza di una keyword al di sopra del 5-10% del totale del testo può essere considerato spamming con tutte le conseguenze del caso.

Ciò non significa che dobbiamo diventare delle macchine per calcolare frequenze e percentuali mentre scriviamo i contenuti per il nostro sito, ricordiamoci solo la semplice ma efficace regola di scrivere i nostri testi con naturalezza e non avremo di che preoccuparci.

Come vengono scelte le parole chiave?

La scelta delle parole chiave di un sito web si divide in 3 fasi fondamentali:

1. Analisi
2. Scelta
3. Inserimento

Queste tre fasi determinano fattori importantissimi per la strategia di posizionamento e per questo dobbiamo seguire delle regole molto precise che andremo ad analizzare passo per passo.

Ci sono vari metodi per iniziare a selezionare una rosa di keywords, uno dei più semplici è quello di fare un brainstorming e verificare i risultati forniti dai vari motori di ricerca.

La tecnica semplice ma efficace consiste nell'effettuare delle query nei motori di ricerca con frasi attinenti al nostro business e nell'annotarsi quanti risultati vengono forniti singolarmente.

Ad esempio ipotizzando che la nostra attività realizzi siti web andremo a fare le seguenti ricerche (tra parentesi sono espressi i risultati)

creazione siti web (2.130.000)

creazione siti internet (1.260.000)

società che faccia un sito web (351.000)

realizzazione siti web (6.370.000)

realizzazione siti internet (3.640.000)

Dai risultati emersi possiamo dedurre che la keyphrase (ovvero la frase chiave) che conviene utilizzare è sicuramente “realizzazione siti web” mentre è consigliabile lasciar perdere “società che mi faccia un sito web”.

Questo metodo non ci permette di valutare quello che realmente gli utenti ricercano, ma si basa sulla semplice deduzione che più c'è concorrenza in una determinata keyphrase più è un terreno fertile nel quale impiantare il nostro business in quanto sicuramente molti utenti effettuano quella ricerca.

Questo è però un metodo da prendere con le pinze, infatti spesso alcune keywords o keyphrase molto gettonate sono impossibili da posizionare, se volessimo cercare di raggiungere la prima posizione per la keyword “siti” sarebbe sicuramente un'impresa con poche probabilità di successo, ma è anche vero che qui si vedono i veri talenti.

Un altro metodo sicuramente più accurato ma anche più impegnativo è quello di andare a verificare il trend dei visitatori. Questo può essere fatto attraverso siti che permettono di verificare il numero di ricerche effettuate per una data keyword in diversi periodi temporali. Questi risultati però sono poco significativi per una serie di ragioni tra cui la prima è sicuramente l'affidabilità di

tali siti.

La strada giusta da intraprendere è quella di spendere qualche decina di euro in una campagna pubblicitaria sulle piattaforme di advertising proprie dei motori di ricerca come ad esempio Adwords di Google e Search Marketing di Yahoo tanto per citarne un paio.

Tramite questi due tools riusciremo ad avere un'idea di quanta gente cerchi la stessa cosa, di quali ricerche vengono effettuate maggiormente nel nostro settore, il loro volume di traffico ed indirettamente anche quanta concorrenza attiva ci troviamo di fronte, senza scordarci che quello che leggiamo sono solamente delle stime.

La concorrenza attiva è un altro fattore da valutare adeguatamente, infatti per la keyphrase “realizzazione siti web” avevamo oltre 6 milioni di risultati, ma questi non erano tutti concorrenti, la maggior parte delle pagine infatti si trovano nell'indice perchè contengono quella frase ma non puntano a scalare la SERP o addirittura per una citazione casuale, ma nel momento in cui un webmaster paga per ricevere visite attraverso quella keyphrase diventa un avversario con i nostri stessi obiettivi.

Terzo ed ultimo metodo è quello di creare sondaggi mirati per i nostri utenti, offrendo magari un servizio gratuito a chi risponde.

Può sembrare inutile ma effettuare dei sondaggi ci permette di venire a conoscenza di come la gente esprime i propri desideri e interessi, ed è proprio questo

che noi vogliamo.

Illudendoci di sapere quali sono gli interessi del nostro target di utenti peccheremmo di presunzione, facciamocelo dire da loro!

Ora abbiamo una prima bozza di keywords relative al nostro settore, il passo successivo sarà quello di plasmarle al nostro sito attraverso i seguenti aspetti:

- **Valutare l'attinenza al sito**

Andranno selezionate le keywords che realmente rispecchiano il tema del nostro sito e gli interessi dei nostri potenziali utenti. Cercare di posizionare una keyword non molto attinente al nostro sito web sarà ancora più difficile in quanto i motori di ricerca non ci daranno un punteggio completamente positivo.

- **Immedesimarsi nell'utente**

La stragrande maggioranza delle ricerche avviene mediante frasi composte da due o più termini, dovremo quindi metterci nei panni degli utenti e pensare quali parole chiave potrebbero immettere nel campo di ricerca per trovare siti del nostro settore con tutte le possibili associazioni del caso.

- **Utilizzare keyphrase**

Oltre ad essere potenzialmente più semplice il

posizionamento di una keyphrase rispetto ad una keyword singola, l'utilizzo di frasi ci permette di raggiungere un target più specifico e di proporre contenuti più attinenti. Una buona tattica è quella di sfruttare aggettivi o aggiungere dettagli, quindi continuando il precedente esempio della nostra attività che realizza siti web, partendo dalla keyword principale “realizzazione siti web” aggiungiamo le keyphrases:

- a) realizzazione siti web economici, definendo una peculiarità
- b) realizzazione siti web a Venezia, impostando una localizzazione geografica
- c) realizzazione siti web in flash, definendo un metodo di realizzazione

Come potete vedere non bisogna fare altro che dare sfogo al nostro estro e cercare più combinazioni possibili.

- **Valutare il KEI**

Il KEY (Keyword Effectiveness Index) ovvero l'indice di efficacia di una parola chiave è un parametro fondamentale da andare a calcolare. Esso è dato dal quadrato della popolarità di una keyword moltiplicato per 1000 e diviso per il numero di siti restituiti da un motore di ricerca preso come riferimento. Più alto sarà quindi il

KEI maggiore sarà l'efficacia di tale parola chiave.

- **Scegliere keywords remunerative**

Il nostro obiettivo è sicuramente quello di massimizzare i guadagni e non sempre maggiori visite significano maggiori guadagni, dobbiamo quindi valutare quali sono le keywords più remunerative ed attraverso quali di essi noi riusciamo a trarne i maggiori profitti. Il più delle volte infatti è bene evitare keywords ad alto traffico ma troppo generiche.

- **Evitare le keywords ambigue**

Alcune keywords potrebbero facilmente confondersi con altri settori, nonostante negli ultimi tempi i motori di ricerca tramite un lavoro di semantica stanno riuscendo a differenziare tali termini, è sempre bene stare alla larga da queste keywords perchè ci ritroveremo a competere anche con siti di un altro settore, aumentando così il numero di “avversari”, e a ricevere utenti non interessati ai nostri servizi.

- **Difficoltà di posizionamento**

Ogni caso è a parte e bisogna fare una valutazione delle proprie capacità e potenzialità del sito web, ma in linea di massima non conviene mai buttarsi a capofitto in keywords altamente competitive, è sempre meglio iniziare da altre

meno competitive nell'attesa che il nostro sito web salga di popolarità.

Inoltre alcune keywords, che a mio avviso si dovrebbero definire ipercompetitive, sono quasi impossibili da raggiungere in quanto fanno parte di settori già affermati da anni. Cercare di posizionarsi in prima posizione ad esempio per la keyword “linux” a mio avviso è un'impresa senza possibilità di successo.

Possiamo considerare a questo punto conclusa la fase di analisi che si è basata su dati numerici e su considerazioni ad hoc su potenziali utenti e settore di appartenenza.

La fase successiva, ovvero quella della scelta, mira a valutare un'ulteriore ed ultima volta le keywords in modo da creare quella che in gergo viene chiamata “Piramide di Keywords”.

Cos'è una Piramide di Keywords?

Dobbiamo immaginare questa piramide come una struttura astratta che ci permetta di definire una relazione di importanza tra le nostre keywords, dove in cima abbiamo la “keyword principale” che è la keyword più importante tramite la quale ottimizzeremo tutto il

sito web, un gradino più in basso abbiamo le “keywords di secondo livello”, ovvero 3 keywords direttamente collegate alla keyword principale e che estendono la sua semantica. Scendendo ancora più in basso troviamo le “keywords di terzo livello” composte da 6 frasi o parole collegate al tema principale del sito ma di minore importanza.

Una volta stabilite le 10 keywords da inserire nella piramide le rimanenti non saranno scartate, ma dobbiamo immaginare che ci sia un'ulteriore gradino, quello delle “keywords collegate” che saranno utilizzate nei nostri testi per arricchire i contenuti e rafforzare il tema del sito.

Per capire meglio il funzionamento di una piramide di keywords andremo adesso a realizzarne una per la nostra fantomatica attività online che realizza siti web.

1. **La keyword principale** che definisce il tema del sito sarà sicuramente “realizzazione siti web”, anche per una serie di considerazioni fatte precedentemente quali il maggior utilizzo nel web e l'attinenza con il sito.
2. **Le keywords di secondo livello** direttamente collegate alla chiave principale potrebbero essere: realizzazione siti internet – creazione siti web – creazione siti internet

3. **Le keywords di terzo livello** di importanza minore potrebbero essere: realizzazione siti web dinamici - progettazione siti internet – gestione siti internet – siti web economici – sviluppo siti web – sviluppo siti internet

4. Le **keywords collegate** che altro non sono che una serie di termine collegati al tema principale da inserire nei nostri testi potrebbe essere: restyling siti web - restyling sito – realizzazione sito flash – realizzazione sito php – realizzazione forum - ristrutturazione siti - web agency - realizzazione logo aziendale - siti e-commerce – accessibilità siti internet - grafica multimedia

Questo genere di ottimizzazione richiede un lungo e laborioso lavoro di analisi, ma in pochi semplici passi siamo riusciti a realizzare una pianificazione che può portarci a risultati soddisfacenti.

Per fare in modo che l'ottimizzazione diventi massima, questo lavoro verrà ripetuto per ogni pagina del nostro sito web, infatti i motori di ricerca non indicizzano siti ma pagine di un sito, e queste pagine avranno maggiore possibilità di scalare le posizioni se saranno caratterizzate come elementi a sé stanti.

Ipotizziamo di voler posizionare un sito di elettrodomestici, al suo interno ci saranno pagine facenti

parte di varie macrocategorie come ad esempio televisori, frigoriferi, forni, ecc...

E' quindi necessario considerare tali pagine come se fossero delle entità a parte in quanto dovranno essere trovate da utenti che hanno bisogno di quegli oggetti in particolare e non di elettrodomestici in generale.

Dove conviene inserire le parole chiave?

Dopo un lungo e faticoso lavoro, siamo giunti all'ultima fase, quella dell'inserimento delle parole chiave nel sito che non può essere lasciato al caso, infatti esistono alcune sezioni della pagina che i motori di ricerca considerano particolarmente importanti. Sarà quindi nel nostro interesse cercare di far apparire le parole chiave in queste sezioni permettendogli di acquisire una maggiore rilevanza.

Di seguito abbiamo una lista delle sezioni in cui è consigliabile far apparire le keywords, ma badate bene: "consigliabile" non vuol dire "far apparire a tutti i costi".

1. Nel titolo della pagina.
2. Nei meta tags.
3. Nelle intestazioni (tags H1 e simili).
4. Nelle prime righe di una pagina.
5. Nei testi dei link interni ed esterni purché

pertinenti.

6. Tra le parole in grassetto o in corsivo.
7. Nel nome del file e/o della cartella.

Per ora bisogna fare un atto di fede e credere che sia realmente così, continuando nella lettura tutti questi aspetti verranno ampiamente descritti singolarmente fugando ogni dubbio.

Un'ultima considerazione da fare riguarda le stop words, ovvero parole e caratteri troppo comuni, che vengono scartati in automatico dalle ricerche sui motori.

Tanto per citare qualche esempio “di”, “con”, “per”, etc... non vengono considerate dalla quasi totalità dei motori di ricerca, potrebbe quindi diventare controproducente cercare di posizionarsi per la keyphrase “autobus di roma” invece di “autobus roma”.

La scelta delle parole chiave negli url

A monte di questa scelta c'è stato un dibattito molto acceso che si è portato avanti per anni nel campo SEO riguardante l'utilizzo di URL statici o dinamici e in che modo l'utilizzo di una rispetto all'altra aiutasse nel posizionamento.

Fino a qualche anno fa era plausibile affermare che assegnare un'estensione statica alle pagine, ad esempio .html, favorisse l'indicizzazione, ma ad oggi la quasi totalità dei motori di ricerca gestisce in modo del tutto corretto sia gli url statici che quelli dinamici ed hanno sviluppato una certa capacità nell'analizzare la presenza di keyword negli URL al fine di contestualizzare la pagina, dando così un posizionamento migliore al sito.

Abbiamo quindi capito che questa tecnica paga in termini di posizionamento ma non deve diventare una gara a chi infila il maggior numero di keywords in un URL, infatti la struttura del sito deve rimanere il più semplice ed omogenea possibile.

Proponiamo quindi un esempio che ci aiuterà a capire meglio quanto detto ipotizzando di avere un negozio che vende articoli per il bricolage:

www.sitodiesempio.it/negozio.php?idcat=5&idprod=24212&n=25

Un indirizzo del genere non darà nessuna indicazione ai motori di ricerca riguardo la nostra attività e la nostra

vendita.

www.sitodiesempio.it/negozio_bricolage/vendita_bricolage/articoli_bricolage_vasi.html

D'altro canto un URL come questo costituisce un'evidente forzatura nell'ottimizzazione, che il motore di ricerca potrebbe considerare come spam.

www.sitodiesempio.it/negozio_bricolage/articoli/vasi.html

Questo è sicuramente un URL ottimizzato e che non trascura l'usabilità e la chiarezza mettendo bene in risalto di cosa stiamo parlando, ovvero di articoli in un negozio di bricolage e per la precisione di vasi.

Ma la questione non sembra essere chiara a tutti, infatti molti webmasters creano in modo artificiale URL di una lunghezza spropositata e piene di parole chiave, ricevendo l'effetto contrario, ovvero una penalizzazione da sovra-ottimizzazione e non un posizionamento da ottimizzazione.

Prima di andare avanti sfatiamo un altro mito: utilizzare il trattino “-” invece dell'underscore “_” come separatore porta dei benefici?

Prima di rispondere a questa domanda vorrei precisare che i caratteri interpretati come separatori dai motori di

ricerca sono ben 5 ovvero:

1. Il trattino “-”
2. L'underscore “_”
3. La virgola “,”
4. Lo slash “/”
5. La pipe (o barra verticale) “|”

La risposta è che i motori di ricerca interpretano il trattino e l'underscore in modo identico, per la precisione come se le parole fossero divise da uno spazio.

Perchè è conveniente utilizzare l'url rewriting?

- **Non tutti i motori sono uguali**

La maggior parte dei motori riesce a gestire come abbiamo detto URL statici e dinamici senza problemi ma non bisogna fare di tuttata l'erba un fascio, infatti nel web ci sono tanti motori di ricerca che non hanno tecnologie così avanzate ed hanno problemi nell'indicizzare URL dinamici soprattutto se con molti parametri.

- **Click-through rate più elevato**

Nel corso degli anni gli utenti dei motori di ricerca non sono rimati quei “cliccatori passivi” di

risultati come erano all'inizio, ma hanno imparato a riconoscere dalla presentazione un sito con contenuti oppure un sito con l'obiettivo di essere solamente spinto. E' statisticamente provato che siti con URL espliciti e chiari nei risultati di un motore di ricerca ricevono quasi il doppio dei click rispetto agli altri siti.

- **Url più facile da ricordare**

Avendo un url del tipo
/nome_categoria/nome_prodotto/ gli utenti lo ricorderanno con più facilità e non c'è il rischio che non riescano a ritrovare la pagina in questione in una successiva visitata.

- **Indicizzazione nei browser**

I browser di ultima generazione ci permettono digitando direttamente delle parole nella barra degli indirizzi di ritrovare nella cronologia siti che abbiamo già visitato. Questo permetterà all'utente di tornare sul nostro sito digitando in locale una semplice parola che lui ricorda essere collegata al nostro prodotto.

- **Piramide dei contenuti**

Una riscrittura degli URL correttamente eseguita ci permette di generare una struttura gerarchica del sito ottenendo una vera e propria piramide dei contenuti.

Cosa bisogna evitare nell'url rewriting?

- **Evitare contenuti duplicati**
Se non abbiamo una perfetta padronanza con la tecnica potremmo in buona fede generare contenuti duplicati facendo puntare diversi URL alla medesima pagina. E' bene sempre controllare a fondo il nostro sito web immedesimandoci negli spiders.
- **Non inserire spazi tra le parole**
Creare una cartella o una pagina con degli spazi è una tecnica scorretta (es: sito.com/negozi online/articoli bricolage). Inoltre come spiegato in precedenza i motori di ricerca interpretano i trattini e gli underscore come degli spazi, quindi è bene utilizzare tali caratteri.
- **Non creare troppi livelli**
URL con troppi livello (es: sito.com/liv1/liv2/liv3/liv4/liv5/liv6/pagina.html) non vengono ben visti dai motori di ricerca che si fermeranno solitamente al 3° o 4° livello.
- **Non usare troppe parole**
Non è consigliabile inserire nomi di cartelle o di pagine troppo lunghe, un buon limite è di 5 parole. Se avete necessità maggiori è bene scindere la frase in più livelli. Ad esempio se

abbiamo un URL come
sito.com/negoziò_online_articoli_casa_bricolage
_vasi_economici.html è bene farlo diventare
qualcosa del tipo
sito.com/negoziò_online/articoli_casa/bricolage
/vasi_economici.htm che è per altro di più facile
lettura.

- **Evitare i caratteri diacritici**
E' bene non inserire nei nostri url caratteri diacritici come ad esempio à, è, ò, ù, etc... che oltre a non essere gradevoli non sono compatibili con tutti i browser e motori di ricerca.
- **Cercare di non utilizzare i SessionID in querystring**
Il sessionID è un parametro univoco per ogni utente che mantiene al suo interno informazioni riguardo la sessione dell'utente stesso. Come abbiamo precisato questo sessionID è univoco quindi se 10 crawler visitano il nostro sito, anche contemporaneamente, gli verranno forniti 10 url diversi del tipo
www.sito.com/pagina.php?PHPSESSID=931564f24f7df12e0d2eb2d7168c85ff
Gli spiders dei motori di ricerca quindi vedranno la stessa pagina decine e decine di volte, se non infinite se si viene a creare un loop, e non riuscendo a determinare in modo corretto cosa

sta accadendo potrebbero abbandonare l'indicizzazione del sito o peggio penalizzarci per contenuti duplicati.

Concludendo il consiglio è quindi quello di creare degli URL search engine friendly ma di non limitarsi a guardarli solo dal punto di vista del posizionamento ma anche dal punto di vista dell'usabilità di un visitatore.

I sottodomini sono migliori delle sottodirectory?

Innanzitutto spieghiamo bene cosa è un sottodominio, che come suggerisce il nome è un qualcosa di livello inferiore ad un dominio.

Scendendo più nei dettagli una directory si presenterà con un URL del tipo:

<http://www.sitodiviaggi.com/roma/>

un sottodominio si presenterà invece così:

<http://roma.sitodiviaggi.com>

Notiamo subito che questa tecnica ci permette di inserire parole chiave nell'URL del sito, e abbiamo imparato nei precedenti capitoli quanto questo può essere importante.

Oltre ogni discorso sull'ottimizzazione, se andiamo a valutare il fattore umano, vedremo che avere dei sottodomini invece che delle sottodirectory è meglio.

Trovandosi di fronte a una decina di risultati su un motore di ricerca, l'utente cliccherà sicuramente più un sito che contiene la keyword nell'URL e non un sito anonimo che non ci fornisce alcuna informazione di base.

Inoltre i sottodomini sono più facili da ricordare per il semplice fatto che non hanno bisogno di avere alcuna

estensione, e quante volte cimentandoci nel ricordare l'URL di un sito abbiamo digitato .html ed invece era .htm?

Per quali contenuti è bene creare un sottodominio?

I sottodomini vanno creati per tutti quei contenuti differenti e che si allontanano quindi dal tema centrale del sito web. In un sito con categorie ad esempio queste potranno avere maggiore rilevanza, ed essere quindi indicizzate meglio, assegnando ad ognuna di esse un sottodominio diverso.

L'esempio portato in precedenza riguarda un'ottima situazione nel quale è bene scindere il sito principale in più sottodomini. Infatti in un sito di viaggi è bene creare un sottodominio per ogni città in quanto le attrazioni e servizi di quella città verranno caratterizzati meglio e i visitatori potranno accedere in maniera più rapida alle risorse di cui hanno bisogno.

Ma quali sono i benefici di un sottodominio?

Il primo beneficio sostanziale è quello di poter risparmiare un livello nell'albero di profondità. Come detto in precedenza infatti gli spiders dei motori di

ricerca partono dalla radice principale e scendono in giù di livello, più in profondità verrà trovata una pagina, minore sarà la rilevanza assegnatagli.

Per capire bene il concetto dobbiamo immaginare il nostro sito web come un grattacielo ma partendo dall'alto. Al primo livello troviamo il nome del sito web, al secondo livello troviamo le directory e dal terzo livello a scendere troviamo le pagine web e le sottodirectory.

Utilizzando i sottodomini il primo ed il secondo livello verranno compressi in un'unica entità facendo di fatto risparmiare un livello.

Un altro vantaggio risulta essere la possibilità di poter registrare ogni sottodominio singolarmente nelle directory aumentando di fatto la nostra presenza nelle stesse e con essa la nostra popolarità.

Infine abbiamo un vantaggio ai quali forse pochi pensano: quello di poter creare domini strategici nel caso in cui i domini ad hoc sono già stati registrati dalla concorrenza.

Ad esempio se volessimo posizionare un dominio di hotel low cost a Roma e lowcosthotelroma.com sia già registrato potremmo ripiegare su roma.lowcosthotel.com con lo stesso potenziale successo.

Quali sono invece gli svantaggi dei sottodomini?

Un sottodominio viene visto come un qualcosa di a parte dal sito principale, quindi andranno dedicati tempo e fatica ad ognuno di essi singolarmente per il posizionamento.

Bisogna inoltre valutare bene il numero dei sottodomini per non incappare in tecniche di spamdexing. Creare 10 sottodomini infatti è una tecnica del tutto lecita a patto che questi abbiano dei contenuti, infatti creare 10 sottodomini di una pagina partendo da un sito di 3 pagine è un'evidente tecnica di spam.

E' più facile posizionare un sottodominio o un dominio?

Sfatiamo da subito il mito secondo il quale i sottodomini sono più facile da posizionare: i sottodomini e i domini vengono trattati allo stesso modo e sono quindi soggetti alle stesse richieste dai motori di ricerca.

Invece alla domanda “I motori di ricerca danno lo stesso peso ad un sottodominio ed a una sottodirectory?” dobbiamo rispondere No. Un sottodominio infatti viene considerato come un qualcosa di a sè stante, la sottodirectory invece come un qualcosa di livello inferiore e gli viene attribuita di conseguenza un peso minore.

Ultimo aspetto da considerare al riguardo è quello dei siti multilingue.

Ipotizziamo che avendo il sito www.miosito.it non possiamo registrare le estensioni .fr, .de, .es, etc..., o anche una sola di queste, perchè già prese o per l'impossibilità dell'acquisto in quanto non siamo cittadini di quel paese, come gestiamo il nostro sito multilingua?

La risposta che ci viene subito in mente è di creare qualcosa del sito:

www.miosito.it/it/

www.miosito.it/en/

www.miosito.it/fr/

e questa potrebbe essere una soluzione plausibile, ma nel caso in cui avessimo anche delle sottodirectory o dei sottodomini come ci dovremmo comportare?

Le soluzioni sono sostanzialmente due, ovvero:

<http://en.miosito.it/curriculum/>

oppure

<http://curriculum.miosito.it/en/>

Entrambe sono del tutto valide, ma quella che vi consiglio è sicuramente la prima.

Come vi ho spiegato in precedenza i sottodomini devono essere il più caratterizzati possibile, quindi il sottodominio sarà composto dalla lingua di riferimento

e nella sottodirectory verrà inserito il curriculum ovvero l'argomento, o meglio il contenuto, che verrà trattato in tutte le lingue.

DETTAGLI TECNICI

I meta tag

I meta tag sono dei particolari tag HTML che servono per specificare meta informazioni e non per implementare il vero e proprio sito web.

La loro storia inizia di pari passo con gli albori del World Wide Web, quando i primi spiders si limitavano a leggere solo le prime righe dei siti web.

Proprio per aiutare gli spiders a catalogare correttamente le pagine furono introdotti questi tag nei quali i webmasters potevano inserire un breve riassunto delle loro pagina web.

Si arrivò ad un punto in cui molte persone iniziarono a creare siti di bassa qualità affidandosi a servizi online che generavano i meta tag credendo che fosse sufficiente fare ciò per ricevere visite, ma i risultati furono ovviamente pessimi.

Tuttavia da strumento di cortesia divenne presto uno strumento d'inganno, alcuni webmasters disonesti infatti iniziarono ad inserire nei tag testi che non rispecchiavano i contenuti reali del sito col solo fine di ricevere visite, e non era raro in quel periodo finire su siti pornografici inserendo la parola MP3.

Si è dovuto attendere fino al 1998 quando Google , seguito a ruota da quasi tutti gli altri motori di ricerca, iniziò a sviluppare altre tecniche di indicizzazione dando un ruolo sempre più marginale ai meta tag.

Al giorno d'oggi è ancora utile inserire i meta tag?

Moltissime persone affermano con certezza che ad oggi i meta tag non hanno più alcun valore, vi posso assicurare che non c'è affermazione più sbagliata.

E' un dato di fatto che non hanno più la stessa importanza che potevano avere 10 anni fa, ma ad oggi sono ancora importanti per diversi motivi, ad esempio in siti composti esclusivamente da immagini o realizzati in flash possono essere utilizzati per categorizzare il sito.

Un sito ben realizzato inoltre lo è in tutte le sue componenti e molti motori di ricerca valutano tali tag non per carpire il contenuto del sito, bensì quanto esso sia curato e degno più di altri di essere posizionato.

Infine c'è anche il rovescio della medaglia, un sito con 1000 keywords nei meta tag verrà di certo etichettato come spam e trattato di conseguenza.

In conclusione i meta tag oggi non servono più ai fini del posizionamento ma possono “assegnare” un punteggio positivo o negativo alla reputazione di un sito.

Come inserisco questi meta tag nel mio sito?

Dopo questa breve introduzione passiamo alla parte implementativa.

I meta tag vanno inseriti nel codice HTML tra i tag

<HEAD> e </HEAD>, sono formati da una coppia nome-valore e non prevedono chiusura.

I tre attributi principali sono:

- NAME
- HTTP-EQUIV
- ROBOTS (sotto-attributo di NAME)

che andremo ad analizzare nel dettaglio.

Il meta tag NAME

L'attributo NAME contiene informazioni non importanti per gli spiders o per i browser ma di potenziale utilità per i navigatori.

- AUTHOR
Identifica l'autore della pagina.
<META NAME="author" CONTENT="Nome Cognome">
- DESCRIPTION
Contiene una breve descrizione della pagina web.
E' bene non superare i 100-150 caratteri.
<META NAME="description" CONTENT="Breve frase">

- **KEYWORDS**

Tale sezione viene utilizzata per inserire una lista di keywords relative alla pagina facendo attenzione a non superare i 10 termini. Non è mai stata specificata una regola precisa sull'utilizzo o meno delle virgole per separare i termini, è buona norma quindi metterle per una questione di leggibilità.

```
<META NAME="keywords"  
CONTENT="keyword1, keyword2, keyword3, ... ">
```

- **COPYRIGHT**

Tag con la funzione di specificare eventuali diritti d'autori sulla pagina.

```
<META NAME="copyright"  
CONTENT="Nome Cognome">
```

- **GENERATOR**

Indica il programma o script utilizzato per la realizzazione della pagina.

```
<META NAME="generator"  
CONTENT="Nome Programma">
```

- **DC LANGUAGE**

Specifica la lingua utilizzata all'interno della pagina.

```
<META NAME="DC language"  
CONTENT="ita" SCHEME="RFC1766">
```

Il meta tag HTTP-EQUIV

L'attributo HTTP-EQUIV contiene informazioni di esclusiva comunicazione tra server e browser.

- **REPLY-TO**
Indica l'indirizzo email del webmaster o dello scrittore dei contenuti.
`<META NAME="reply-to" CONTENT="username@dominio.com">`
- **EXPIRES**
Indica al browser la scadenza in formato GTM del documento. Un documento scaduto dovrebbe essere richiesto al server e non recuperato dalla cache. Molti webmasters utilizzano il trucchetto di inserire una data passata per non far memorizzare la pagina nella cache.
`<META NAME="expires" CONTENT="Data GTM">`
- **REFRESH**
Permette di aggiornare la pagina ad un intervallo di tempo prestabilito. Aggiungendo l'attributo URL permette la redirectione verso un'altro indirizzo. L'URL va inserito in modo esteso.
`<META NAME="refresh" CONTENT="10">`
`<META NAME="refresh"`

`CONTENT="10;URL=http://www.sito.com">`

- **PRAGMA**

Tag del tutto deprecato. Impedisce al browser Netscape Navigator di salvare le pagine nella cache

`<META NAME="pragma" CONTENT="no-cache">`

Il meta tag **ROBOTS**

Il meta tag robot può essere considerato un vero e proprio sotto-attributo dell'attributo NAME in quanto ha una propria struttura e viene utilizzato per dare delle direttive agli spiders su quali contenuti visitare e/o archiviare e quali no.

La sua sintassi di base è:
`<META NAME="robots" CONTENT="all | none | index | noindex | follow | nofollow">`

Se si vuole far archiviare al motore di ricerca i contenuti dalla pagina si usa la direttiva INDEX, altrimenti si usa NOINDEX.

Per consentire allo spider di seguire tutti i link presenti nella pagina dobbiamo utilizzare la direttiva FOLLOW altrimenti NOFOLLOW se non lo vogliamo.

Le due tipologie di direttive possono essere combinate senza problemi inserendole nel campo content divise da

una virgola come di seguito

```
<META NAME="robots" CONTENT="index,  
nofollow">
```

Inoltre esistono anche le 2 direttive ALL e NONE.

ALL abilita entrambe le direttive ed è quindi sinonimo di INDEX,FOLLOW

NONE disabilita entrambe le direttive ed è quindi sinonimo di NOINDEX,NOFOLLOW

E' bene sapere inoltre che in caso di mancanza del meta tag ROBOTS gli spiders si prenderanno tutte le libertà del caso interpretandolo come un INDEX,FOLLOW.

Esistono però anche dei meta tag che potremmo definire personali, Google ad esempio ha messo a disposizione due direttive ad hoc per il suo spider.

Molti webmasters si lamentavano infatti della funzione "Copia cache" che consente agli utenti di visualizzare la pagina così com'era nel momento in cui è passato lo spider di Google ad indicizzare la pagina.

In un sito aggiornato molto spesso tale funzione può essere controproducente, potremo quindi chiedere di non salvare una copia negli archivi di Google inserendo il meta tag

```
<META NAME="googlebot" CONTENT="noarchive">
```

Il secondo tag permette invece di rimuovere dalla pagina dei risultati lo snippet, ovvero la parte di testo inserita sotto il titolo che ci introduce alla pagina, in quanto non sempre molto attinente.

Il tag in questione è:

```
<META NAME="googlebot" CONTENT="nosnippet">
```

E' bene sapere che richiedendo la rimozione degli snippet, viene richiesta implicitamente anche la rimozione dell'archiviazione.

I tag ROBOTS però non sono sempre sufficienti a far operare gli spiders nel modo in cui vogliamo ed hanno inoltre la grande limitazione di dover essere inseriti in ogni pagina non potendo escludere a priori dei gruppi di contenuti o adoperare regole comuni.

Proprio da questa necessità è nato il file robots.txt, un utile strumento per controllare i comportamenti dei motori di ricerca che andremo ad analizzare nei dettagli nel prossimo capitolo.

Il file robots.txt

Nel capitolo precedente abbiamo analizzato il meta tag ROBOTS e ci siamo accorti che questo ha delle grandi limitazioni essendo uno strumento dedicato alla singola pagina web e non all'intero sito.

Per soddisfare queste nuove e crescenti esigenze è nato il Robot Exclusion Standard, ovvero un protocollo che permette ai webmasters di avere il pieno controllo sulle loro pagine web durante l'indicizzazione dei motori di ricerca.

Il file robots.txt ha iniziato la sua diffusione con Altavista, il primo popolare motore di ricerca, e recentemente ne è stata proposta anche una seconda versione: l'Extended Robot Exclusion Standard, che non ha trovato la piena approvazione di tutti e rimane ancora una proposta di standard.

Come possiamo capire è uno standard molto diffuso, ma non tutti gli spiders prendono in considerazione questo file, è bene quindi non fidarsi troppo delle direttive inserite all'interno.

Una volta creato il file robots.txt, questo va pubblicato nella root del sito web. Ad esempio se il nostro sito è www.indirizzo.com, il file robots.txt andrà pubblicato il www.indirizzo.com/robots.txt

Nel caso in cui uno spider non trova questo file si sente libero di poter agire come meglio crede nel nostro sito e

di indicizzare a sua discrezione tutti i contenuti.

Non sempre è necessario creare il file robots.txt in quanto potremmo non avere la necessità di porre limitazioni agli spiders, è però buona norma creare sempre questo file perchè darà ai motori di ricerca una sensazione di professionalità e nel frattempo ci metteremo anche al riparo da problemi indesiderati derivanti dall'errore 404.

Ma come deve essere scritto il file robots.txt?

Come abbiamo visto il file robots.txt è un semplice file di testo realizzabile quindi con un qualsiasi editor, nel quale al suo interno abbiamo una serie di record composti da uno o più comandi.

Andremo adesso ad analizzarli uno per uno:

- **User-Agent** <due_punti> <spazio> <nome>
Imposta il nome dello spider al quale ci stiamo riferendo (es: googlebot per Google, scooter per Altavista, slurp per Yahoo!, ecc...).
L'User-Agent è “case insensitive” e quindi può essere scritto indifferentemente in maiuscolo o in minuscolo.
- **Disallow** <due_punti> <spazio> <directory | file>

Comunica agli spiders quali directory o file non deve indicizzare. Il blocco può essere generale inserendo un asterisco per bloccare tutti i contenuti oppure esplicito inserendo il nome del file/cartella che si vuole escludere oppure inserendo un'espressione regolare.

- **Allow <due_punti> <spazio> <directory | file>**

Per il comando Allow valgono le stesse regole di Disallow. Viene quasi sempre utilizzato successivamente ad un blocco generale per consentire l'accesso esclusivamente ad una o più directory o file.

- **Robot-version <due_punti> <spazio> <versione>**

Viene utilizzato unitamente al comando User-Agent per impostare la versione dello spider alla quale ci stiamo riferendo.

- **Visit-time <due_punti> <spazio> <time> <trattino> <time>**

Comunica agli spider l'intervallo di tempo durante il quale le pagine possono essere visitate. Gli orari fanno riferimento al GTM.

Es: Visit-time: 0800-1230 # Indica agli spiders che le pagine possono essere visitate solo tra le 8:00 GTM e le 12:30 GTM

- **Request-rate** <due_punti> <spazio> <pagine/secondi>
Comunica agli spiders quante pagina visitare ogni X secondi. E' di scarso utilizzo e viene supportato da pochi.
- **Comment** <due_punti> <spazio> <testo>
Comando utilizzato per inserire commenti all'interno del file. Può essere sostituito dalla forma compatta # (cancelletto).
- **Sitemap** <due_punti> <spazio> <url>
Consente di impostare l'url della sitemap del sito alla quale gli spiders potranno fare riferimento. E' possibile inserire più sitemaps semplicemente ripetendo la riga in questione.

Ora passiamo a qualche esempio per capire meglio di cosa stiamo parlando.

Tutti i robot non potranno indicizzare alcuna pagina del sito

```
User-agent: *  
Disallow: /
```

Tutti i robot sono liberi di indicizzare qualsiasi pagina del sito

User-agent: *
Disallow:

oppure

User-agent: *
Allow: /

Impedire a Google Images di indicizzare la cartella foto-private ed il file foto_natale.htm

User-agent: Googlebot-Image
Disallow: /foto-private/
Disallow: /foto_natale.htm

Impedisce agli spiders di indicizzare qualsiasi pagina dinamica

User-agent: *
Disallow: /*?

Impedisce agli spiders di indicizzare qualsiasi pagina del sito eccetto curriculum.html

User-agent: *
Disallow: *
Allow: /curriculum.html

Ultima raccomandazione: una volta che avrete creato il vostro file robots.txt controllatene la correttezza tramite un analizzatore di file robot.txt (che potete trovare

online) prima di pubblicarlo online.

I tag fondamentali

Dopo aver spiegato nei precedenti capitoli quali sono le basi iniziali di un buon posizionamento, iniziamo ad addentrarci nella costruzione di una pagina web ottimizzata analizzando uno per uno i tag più importanti che determinano la bontà.

I tag fondamentali da inserire in una pagina web sono in ordine rigorosamente di importanza:

1. Il tag **TITLE**
2. Gli Header Tags
3. I tag per la formattazione del testo
4. L'attributo **ALT** delle immagini

Iniziamo ora ad analizzarli uno per uno.

Il tag **TITLE**

Il titolo è il tag più importante all'interno di una pagina web e la scelta di tale tag rappresenta il primo fondamentale passo verso l'ottimizzazione.

Nel corso degli anni sono cambiate molte cose nel posizionamento nei motori di ricerca, ma una cosa è sicuramente rimasta invariata, ed anzi è forse cresciuta, è proprio l'importanza assegnata dai motori di ricerca al titolo di una pagina.

Molti webmasters però ignorano l'importanza di questo elemento e non è raro vedere siti che hanno lo stesso

tag title ripetuto in tutte le pagine del sito, cosa che potrebbe risultare sgradevole ai motori di ricerca in quanto potrebbe essere vista come una rude tecnica per spingere una determinata keyword.

Quali caratteristiche deve avere un tag title?

Come prima cosa non deve essere un'accozzaglia di keywords ma deve essere un frase di 8-9 parole che descriva al meglio la pagina in questione, badate bene deve descrivere la pagina e non l'intero sito!

Nella decisione di un titolo bisogna fare molta attenzione alla prossimità delle parole, infatti per la keyphrase “viaggi oriente” sarà privilegiato un sito con titolo “viaggi oriente economici” al posto di uno che ha come titolo “viaggi economici in oriente”.

Altro fattore importante è la collocazione delle parole, infatti verrà dato più peso ad una parola scritta all'inizio del titolo che alla fine.

I motori di ricerca prendono anche in considerazione l'ordinamento delle parole, il titolo “viaggi europa” sarà considerato in modo diverso rispetto a “europa viaggi” e di conseguenza saranno maggiormente rilevanti in due ricerche differenti.

E' vero che le ripetizioni in un tag title sono penalizzanti?

Rispondere con sì o no sarebbe riduttivo, la risposta in questo caso è: dipende!

Ripetere due o più volte la stessa keyword è sicuramente penalizzante, ma non si può dire lo stesso in una keyphrase.

Per fare un esempio un titolo come “Viaggi in Europa: voli per viaggi” non verrà sicuramente considerato spam e potrebbe, è d'obbligo usare il condizionale, essere migliore di “Viaggi in Europa: voli”.

Il trend migliore da seguire è però sicuramente quello di ripetere una sola volta le keywords nei tag title.

Gli Header Tags

Gli header tags, o heading tags, sono quei tag che compaiono nella forma `<H1></H1>`, `<H2></H2>`, ecc... e vengono usati per definire le intestazioni di un documento.

Il più importante è sicuramente il tag `<H1>` che indica l'argomento principale della pagina, ed è quindi buona norma non ripeterlo.

Abbiamo la bellezza di 6 tag H e sarà nostro compito utilizzarli al meglio per creare una struttura discendente in cui si evidenzino al meglio le varie aree tramite intestazioni.

Esempio con struttura corretta

```
<H1>...</H1>  
<H2>...</H2>  
<H3>...</H3>  
<H4>...</H4>  
<H5>...</H5>  
<H6>...</H6>
```

Esempio con struttura errata (header non ordinati)

```
<H2>...</H2>  
<H1>...</H1>  
<H3>...</H3>  
<H4>...</H4>  
<H6>...</H6>  
<H5>...</H5>
```

Esempio complesso con struttura corretta

```
<H1>...</H1>  
  
<H2>...</H2>  
<H3>...</H3>  
  
<H2>...</H2>  
<H4>...</H4>  
<H5>...</H5>
```

```
<H3>...</H3>  
<H6>...</H6>
```

L'unico effetto collaterale nell'utilizzo di questi tag è quello di rovinare il layout grafico, problema risolvibile mediante i fogli di stile.

I tag per la formattazione del testo

I tag per la formattazione del testo ci aiutano ad evidenziare e dare enfasi a parti di testo alle quali i motori di ricerca daranno maggiore importanza.

I tag utili a questo scopo sono:

- Grassetto
 oppure
- Testo corsivo
<i></i> oppure
- Testo sottolineato
<u></u>

Questi tag possono ovviamente essere concatenati tra di loro per dare maggiore enfasi ad alcune parole, ma è bene non abusarne per evitare di creare testi poco leggibili e rischiare penalizzazioni da sovraottimizzazione.

L'attributo ALT delle immagini

Ultimo in ordine di importanza, ma di certo non da dimenticare è il tag ALT delle immagini che ci permette di descrivere il loro contenuto per i browser testuali, e gli spiders sono dei browser testuali.

Esempio di implementazione del tag ALT

```

```

Lo sapevi che anche le immagini possono aiutarti nel posizionamento?

Può sembrare un paradosso ma è pura verità.

L'uso del tag ALT in modo opportuno per descrivere le principali immagini contenute in una pagina web ci permetterà di ricevere maggiore considerazione dai motori di ricerca che avranno “accesso” ad ogni contenuto del sito web, e nel frattempo aumenteremo anche l'accessibilità del nostro sito.

E' consigliabile utilizzare il tag ALT esclusivamente per immagini attinenti alla pagina e non per gli elementi grafici.

Inoltre per ricevere dei benefici vale la solita regola di descrivere l'immagine con una frase attinente e non

mettere un elenco di parole chiave che potrebbero provocare l'effetto contrario.

La sitemap

La sitemap è nata originariamente come una pagina html in cui erano elencati gerarchicamente o meno tutte le pagine di un sito web. Nel corso degli anni ha iniziato ad avere le attenzioni dei motori di ricerca che l'hanno cominciata ad utilizzare come base di partenza per i contenuti.

La svolta si ebbe nel giugno 2005 quando Google introdusse il servizio Google Sitemaps, dove da un semplice file di testo si passò ad un vero e proprio file XML, era l'inizio di una nuova era.

Nel novembre del 2006 anche Msn e Yahoo annunciarono il supporto alla nuova sitemap.

Ultimo passo nella storia della sitemap si ebbe nel maggio del 2007 quando anche Ask annunciò il supporto per il nuovo formato sitemap e introdusse la novità della segnalazione mediante file robots.txt .

Il protocollo sitemap è stato rilasciato sotto licenza Attribution-ShareAlike Creative Commons License, e proprio grazie ad essa ne è stato possibile l'uso in tutti i motori di ricerca.

Cosa contiene un file sitemap?

Per prima cosa è bene ricordare che un file sitemap è innanzitutto un file XML, quindi nella prima riga va

inserita una stringa che dichiara che stiamo lavorando con dati XML con codifica dei caratteri in UTF-8.

Nella seconda riga va specificata un'ulteriore stringa obbligatoria, ovvero l'<urlset> che specifica lo schema sitemap che stiamo utilizzando.

Dalla terza riga in poi sarà nostra cura inserire uno per uno gli url del nostro sito mediante i seguenti tag:

- **<url>**

Il tag in questione raggruppa le informazioni relative ad ogni singolo url del sito web. Al suo interno è possibile inserire una serie di tag facoltativi e il tag obbligatorio <loc> che vedremo in seguito.

- **<loc>**

All'interno di questo tag va inserito il vero e proprio URL la cui lunghezza non deve superare i 2048 caratteri.

- **<lastmod>**

Tag facoltativo che contiene la data di modifica del documento nel formato YYYY-MM-DD.

- **<changefreq>**

Tag facoltativo, e poco preso in considerazione, per indicare la frequenza di aggiornamento della pagina. I valori possibili sono: always, hourly, daily, weekly, monthly, yearly e never.

- **<priority>**

Tag facoltativo che indica l'importanza di una pagina. Varia da un minimo di 0.0 ad un massimo di 1.0 . Se non specificato di default viene considerato come 0.5 .

Esempio reale di una sitemap

Per comprendere meglio il funzionamento della sitemap ne costruiremo una insieme.

Per semplificare il tutto ipotizziamo che il nostro sito web abbia solamente 3 pagine web:

1. <http://www.sito.com/> - Homepage
2. <http://www.sito.com/prodotto.html> – Pagina dei prodotti
3. <http://www.sito.com/contatti.html> – La pagina dei contatti

Una sitemap completa sarà quindi la seguente:

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset
xmlns="http://www.sitemaps.org/schemas/sitemap/0
.9">
<url>
<loc>http://www.sito.com/</loc>
<lastmod>2009-08-05</lastmod>
<changefreq>weekly</changefreq>
```

```
<priority>1.0</priority>
</url>
<url>
  <loc>http://www.sito.com/prodotti.html</loc>
  <lastmod>2009-07-01</lastmod>
  <changefreq>monthly</changefreq>
  <priority>0.9</priority>
</url>
<url>
  <loc>http://www.sito.com/contatti.html</loc>
  <lastmod>2009-01-20</lastmod>
  <changefreq>never</changefreq>
  <priority>0.1</priority>
</url>
</urlset>
```

Non dimentichiamoci che la sitemap è una “proposta di indicizzazione”, lo spider infatti agirà come meglio credo ed è quindi inutile falsare i valori impostando tutte le pagine come priority 1.0 e changefreq always sperando di avere migliori risultati.

Quali sono i limiti di una sitemap?

Una sitemap è affetta da due grandi limitazioni. La prima è che ogni sitemap non può superare i 50.000 URL contenuti, la seconda riguarda la dimensione

massima che è limitata a 10 Mb.

Queste due limitazioni potrebbero frenare molto siti web di grandi dimensioni come ad esempio siti di e-commerce con decine di migliaia di prodotti, ma non costituiscono un problema reale in quanto è stato realizzato un secondo formato: la sitemap index.

La sitemap index come possiamo intuire dal nome è una vera e propria raccolta di tutte le sitemaps pubblicate sul sito web in grado di farci superare qualsiasi limite di contenuti o spazi.

Si potrebbe pensare che la sitemap index abbia l'unico limite, seppur assurdo, di poter contenere al massimo 1.000 sitemap con 50.000 URL ognuna per un totale di 50 milioni di pagine, ma in realtà non esiste neanche questo limite, infatti è possibile inserire in una sitemap index altre sitemap index, e non direttamente delle semplici sitemap, rendendo la capacità di archiviazione teoricamente infinita.

Passiamo ora alla parte implementativa analizzando il codice di una sitemap index di esempio:

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset
xmlns="http://www.sitemaps.org/schemas/sitemap/0
.9">
<sitemap>
  <loc>http://www.sito.it/sitemap\_pagine.xml</loc>
  <lastmod>2009-08-05</lastmod>
</sitemap>
```

```
<sitemap>
<loc>http://www.sito.it/sitemap\_prodotti.xml</loc>
  <lastmod>2009-07-01</lastmod>
</sitemap>
</urlset>
```

Come possiamo vedere, la sitemap index segue quasi le stesse linee guida di una normale sitemap con la differenza che non sono previsti tag `priority` e `changefreq`, il tag `lastmod` rimane facoltativo.

Come realizzo una sitemap?

Ora che conosciamo la sitemap sotto tutti i punti di vista rimane aperta la questione di come realizzarla. I metodi di creazione sono sostanzialmente tre: a mano, tramite software e automaticamente.

Creare la sitemap a mano

Per questo metodo c'è poco da dire, tranne che è sicuramente il meno conveniente sia in termini di tempo che di affidabilità. Ma se non abbiamo altre risorse non ci resta che aprire un editor di testo e iniziare a scrivere. E' altamente consigliabile a fine lavoro verificare la correttezza del file appena scritto tramite uno dei tanti sitemap analyzer reperibili in rete.

Creare la sitemap tramite software

Questa è la soluzione intermedia, ci sarà sempre bisogno del nostro intervento manuale ma almeno la sitemap verrà redatta da un software scritto appositamente per fare ciò.

Esistono sia software per ambienti desktop che strumenti online per generare la sitemap, ma la meccanica al loro interno è la medesima: una sorta di spider partendo dalla homepage visita ogni singola pagina del nostro sito estraendo una lista di URL che successivamente inserirà in un file XML.

Creare la sitemap automaticamente

Come si può ben capire questa è la soluzione più comoda ed efficace, è necessario però che il proprio sito sia realizzato mediante un CMS che si preoccupi di generare in modo automatico il file XML ad ogni cambiamento o inserimento di contenuti.

Tale soluzione aumenta di gran lunga anche l'efficacia perchè crea una nuova sitemap ad ogni modifica del sito web, permettendo ai motori di ricerca di carpire ogni variazione.

I più diffusi CMS che permettono di fare ciò sono sicuramente Wordpress, Drupal e Mambo.

Come si segnala una sitemap ai motori di ricerca?

Ci siamo, abbiamo capito come è fatta una sitemap e l'abbiamo creata, ora rimane l'ultimo passo: avvisare i motori di ricerca.

Esistono sostanzialmente tre metodi per segnalare una sitemap ai motori di ricerca, di difficoltà crescente ma anche di efficacia crescente:

1. Robots.txt
2. Registrazione
3. Ping

Robots.txt

Come detto nel precedente capitolo una sitemap può essere segnalata (per il codice vi rimandiamo al capitolo precedente) all'interno del file robots.txt .

Non ci sarà bisogno di effettuare alcuna segnalazione manuale, infatti lo spider prima di iniziare la visita controllerà questo file e accorgendosi della sitemap prenderà le dovute decisioni in merito.

Registrazione

Alcuni motori di ricerca hanno messo a disposizione dei webmasters dei veri e propri pannelli di controllo tramite i quali eseguire segnalazioni manuali.

Oltre a permetterci la segnalazione ai motori di ricerca,

questi strumenti effettueranno anche un'analisi sintattica del file e ci avvertiranno di eventuali errori 404 degli URL al loro interno.

Sono strumenti potentissimi che vi consiglio di utilizzare.

Ping

Ultimo nella lista, ma primo nell'efficacia è il ping che in parole povere è un avviso rivolto ai motori di ricerca per segnalare la presenza di un nuovo aggiornamento della sitemap.

La sintassi cambia da motore a motore, e per questo vi rimando alle loro rispettive guide in linea, ma il funzionamento è il medesimo per tutti quanti, sarà sufficiente infatti effettuare una chiamata, tramite browser o script appositi, ad un URL del tipo www.motorediricerca.com/ping?sitemap=URL_DELLA_SITEMAP per effettuare la segnalazione ai motori di ricerca che poi gestiranno a loro discrezione.

I feed possono aiutare nel posizionamento?

Un feed è un formato simile all'XML per la distribuzione di contenuti web. Se ci soffermiamo un attimo, ci accorgiamo che i feeds altro non sono che una sitemap sempre aggiornata.

I motori di ricerca hanno quindi iniziato a dare la possibilità di utilizzare i feeds come valida alternativa alla sitemap per quei siti che non ne avevano la possibilità di crearne una o a quei webmasters che non ne avevano le competenze.

Ad oggi praticamente tutti i motori che accettano la sitemap XML accettano anche il formato alternativo di feeds in RSS 2.0 o Atom 1.0, alcuni addirittura accettano anche il formato file di testo, ma questa è un'altra storia. Senza troppo scendere nei dettagli implementativi dei feeds per i quali ci servirebbe un libro intero, possiamo dire che i motori di ricerca vi estraggono le informazioni nel seguente modo:

1. Recuperano l'URL della pagina dal campo `<link>`
2. Recuperano la data di modifica dal campo `<pubDate>` se si tratta di un feed RSS o dal campo `<modified>` se si tratta di un feed Atom.

L'uso della data di modifica rimane ovviamente facoltativo, mentre non è possibile specificare un valore per la priorità o per la frequenza di aggiornamento.

I feeds contengono solo gli URL recenti, non è controproducente?

La domanda è del tutto comprensibile, ma la risposta è NO.

Sicuramente avere una sitemap completa ed aggiornata è una soluzione migliore rispetto ai feeds che contengono solo gli URL più recenti, ciò non toglie però che gli spiders possano utilizzare queste pagine come base di partenza per poi raggiungere le altre pagine mediante i link presenti.

Ma utilizzare i feeds invece che la sitemap in alcuni casi porterebbe addirittura dei benefici.

Ci sono stati dei casi in cui dei webmasters hanno messo online per la prima volta delle sitemaps molto grandi contenenti qualche migliaio di URL e questa azione a volte è stata considerata come spam e i siti in questione sono stati penalizzati seppur per brevi periodi.

Potremmo chiamare questa situazione sovraottimizzazione da feeds o Feeds Bombing, e si viene a presentare quando viene caricata per la prima volta una sitemap per un sito web già avviato e che ha quindi già molte pagine.

In questi casi la soluzione migliore, provata sul campo, è quella di segnalare dapprima le pagine ai motori di ricerca mediante feeds e, passato del tempo caricare la sitemap seppur di grandi dimensioni.

Come ottimizzare un sito in Flash

Macromedia Flash, d'ora in poi Flash, permette di creare pagine web accattivanti e graficamente stupende, ma ciò potrebbe crearci dei problemi ai fini del posizionamento.

L'uomo è attratto da immagini, animazioni e suoni ma uno spider può leggere, o meglio estrapolare, solamente dei testi, ed ecco qui il primo grande limite.

La funzione degli spiders è quella di estrapolare grandi quantità di testo dal web per darle in pasto agli algoritmi dei motori di ricerca, è ovvio quindi che creare siti web incentrati esclusivamente sulla grafica significa navigare controcorrente.

Alcuni motori di ricerca, ma veramente pochi che si contano sulle dita di una mano, negli ultimi anni hanno sviluppato, anche grazie al rilascio dell'SDK da parte di Macromedia, tecnologie atte all'indicizzazione dei contenuti Flash, riuscendo attualmente ad estrapolare testi e link esterni a patto che non siano dinamici.

Questa è sicuramente una grande evoluzione ma non ha di certo risolto il problema completamente.

Perchè un sito in Flash si posiziona con più difficoltà?

Prima di tutto sfatiamo il mito secondo il quale i siti in Flash sono impossibili da posizionare come troppo

spesso si legge sul web. Sicuramente è più difficile, questo è un dato di fatto, ma la parola impossibile è fuori luogo in quanto con un'appropriata campagna di posizionamento si riesce a indicizzare un qualsiasi sito web realizzato in Flash.

Altri importanti limiti tecnici che pregiudicano il posizionamento di un sito realizzato con la tecnologia Flash sono i seguenti:

- **URL inesistenti**

Un sito web Flash infatti non ha URL che cambia man man che si visitano i contenuti. Noi vedremo sempre la stessa pagina web e all'interno dell'animazione Flash ci verranno mostrati i contenuti.

- **Impossibilità di essere linkati**

I contenuti non avendo degli URL propri, come spiegato sopra, non potranno essere linkati e quindi avremo grandi difficoltà a fare crescere la popolarità del nostro sito web.

- **Testi non formattati e non ordinati**

I motori di ricerca riescono ad estrapolare testi dai siti in Flash, è pur vero però che questi non sono formattati e non hanno un ordine preciso, rendendo arduo il lavoro dei motori di ricerca nel determinare i temi fondamentali della pagina web.

- **Contenuti non supportati**

Molti contenuti in Flash come ad esempio i testi dinamici non vengono visti dagli spiders e per questo non verranno indicizzati con la conseguenza di una grande perdita di informazioni.

Quali sono i metodi migliori per posizionare un sito in Flash?

La soluzione di base è quella di “generare” dei contenuti da dare in pasto agli spiders. Per fare questo abbiamo due possibilità: duplicare i contenuti o creare un sito parallelo.

La duplicazione dei contenuti dal nome potrebbe farci pensare ad una tecnica ai limiti della correttezza, ma così non è.

Flash ci permette di visualizzare un testo alternativo ai quei browser che non supportano il loro plugin, ed è proprio al posto di questo testo alternativo che noi andremo ad inserire il testo dei nostri contenuti. Gli spiders possono essere considerati infatti dei browser che non supportano il plugin Flash e quindi durante la navigazione andranno a leggere questi contenuti, indicizzandoli.

Il codice da utilizzare è il seguente:

```
<object type="application/x-shockwave-flash"  
data="file_flash.swf" width="100" height="100">  
<param name="movie" value="file_flash.swf">
```

[CODICE HTML ALTERNATIVO]

```
</object>
```

Questa tecnica va utilizzata con molta attenzione in quanto non possiamo prevedere le reazioni dei motori di ricerca essendo un campo ancora in via di sperimentazione, è quindi buona norma inserire poco testo e molto attinente al sito in questione.

Evitate di utilizzare i tag <EMBED> e <NOEMBED> in quanto sono stati deprecati dall'XHTML 1.0 Transitional.

Una volta sistemato il testo, possiamo dedicarci ai link. Come già detto il Flash potrebbe mettere in ombra varie sezioni del sito in quanto lo spider potrebbe non riuscire a risalire ai loro URL.

Ne consegue che dovremo essere noi a istruire gli spider, e potremo farlo sia creando una mappa del sito in html, sia mettendo nelle pagine che contengono le animazioni in Flash dei link HTML alle altre sezioni del sito.

Dovrà essere nostra cura poi controllare che i link presenti nel sito permettano di raggiungere qualsiasi pagina da noi creata.

Lavoro terminato! Ora il nostro sito sarà compatibile con qualsiasi spider, infatti anche se non abbiamo eliminato tutte le limitazioni abbiamo per lo meno reso il sito più search engine friendly.

La soluzione esposta in precedenza più che una vera e propria soluzione è una toppa da applicare ad un sito web realizzato con la tecnologia Flash.

La strada sicuramente migliore da seguire, anche se più radicale, è quella di creare un sito web parallelo mettendo in homepage l'utente davanti alla scelta tra il sito in Flash graficamente accattivante o la versione HTML veloce e snella. Studi di settore hanno dimostrato che l'80% degli utenti sceglierà la versione HTML, ma a noi questo poca importa.

Aggiornare costantemente un sito web in parallelo può essere un lavoro snervante, ma trarremo tutti i benefici possibili dai motori di ricerca che indicizzeranno la versione testuale del sito e nel frattempo offriremo un servizio in più ai nostri utenti.

In conclusione penso che tutti noi abbiamo capito che la tecnologia Flash per quanto realizzi siti web di indubbia bellezza grafica, non è di certo un terreno fertile per puntare a scalare le posizioni di un motore di ricerca.

Utilizzare il redirect 301 Moved Permanently in modo corretto

A volte abbiamo la necessità di dover spostare delle pagine, o addirittura un intero sito, verso una nuova locazione. Per fare ciò senza perdere il lavoro di posizionamento fatto sulle vecchie pagine è possibile utilizzare il redirect 301.

Il 301 è uno status code appartenente alla macrocategoria delle redirezioni che indica una redirezione permanente, esiste anche il redirect 302 che indica una redirezione temporanea che non prenderemo in esame in quanto non importante ai fini del posizionamento.

Quando conviene usare un redirect 301?

Oltre al più banale caso di utilizzarlo durante un trasferimento di pagine ci sono altri casi in cui questa tecnica può portarci dei vantaggi.

- **Cambiare il dominio**

A causa di trasformazioni aziendali, o nel caso in cui abbiamo acquisito un dominio maggiormente strategico, potremo cambiare radicalmente sito senza rischio di perdere la popolarità acquisita fino ad allora.

- **Utilizzare la radice www invece della http://**
Anche se raramente, capita che vengano vengano indicizzate pagine della radice http:// a discapito della www, è uno scenario che potrebbe presentarsi e quindi è meglio correre subito ai ripari per evitare contenuti duplicati.
- **Traffico da diversi domini**
Può essere una buona strategia quella di registrare diverse estensioni oppure di registrare differenti nomi a dominio attinenti il nostro settore. In questi casi per portare traffico dai domini secondari verso il dominio principale senza generare contenuti duplicati possiamo utilizzare la tecnica del redirect permanente.

Quali vantaggi può portarci un redirect 301?

Come abbiamo detto in sintesi precedentemente, un redirect 301 ci porta principalmente il vantaggio di non perdere il lavoro di posizionamento fino ad allora eseguito ma nello specifico avremo tre grandi vantaggi:

- **Errori HTTP assenti**
Le vecchie pagine in assenza di un redirect restituirebbero un errore 404 per la pagina

mancante, e questo non è da considerarsi un fattore di qualità per un sito.

- **Link in ingresso intatti**

Come già spiegato la nuova pagina riceverà tutti i benefici della vecchia versione compresa la sua posizione nel ranking, ma anche nel caso in cui la pagina non sia posizionata riceveremo tutti i suoi links in ingresso che ci aiuteranno ad aumentare il pagerank globale del sito.

- **Indicizzazione della nuova versione veloce**

Immaginiamo di avere una pagina “vaso.html” che cambia nome in “prodotto_vaso.html”, tale pagina è linkata dalla pagina “lista_prodotti.html” che è a sua volta linkata dall'homepage.

Senza redirect lo spider dovrà prima visitare la homepage, poi la pagina “lista_prodotti.html” (o direttamente quest'ultima se sia fortunati) e successivamente riuscirà ad indicizzare la nuova pagina, nel frattempo però saranno stati generati una serie inimmaginabile di errori 404 che non si fermeranno neanche con la nuova indicizzazione.

Con il redirect invece non appena lo spider visiterà nuovamente la pagina “vaso.html” si accorgerà del redirect ed indicizzerà la nuova pagina eliminando la vecchia dal suo archivio.

E' bene sapere che la sostituzione delle pagine non avviene in modo istantaneo ma bisogna sempre attendere qualche giorno durante il quale è lecita una breve e temporanea penalizzazione nel posizionamento.

Come si effettua un redirect 301?

Essendo un metodo universale di redirezione è stato implementato da tutti i maggiori linguaggi di programmazione web. Vediamo ora per ogni singolo linguaggio come effettuare il redirect 301 in modo corretto:

.htaccess (n.b.: codice su singola riga)

```
Redirect permanent  
http://www.sito.com/vecchia\_pagina.html  
http://www.sito.com/nuova\_pagina.html
```

PHP

```
<?  
Header( "HTTP/1.1 301 Moved Permanently" );  
Header( "Location:  
http://www.sito.com/nuova\_pagina.html" );  
?>
```

ASP

```
<%  
Response.Status="301 Moved Permanently"  
Response.AddHeader      "Location",      "  
http://www.sito.com/nuova_pagina.html"  
%>
```

ASP .NET - C#

```
private      void      Page_Load(object      sender,  
System.EventArgs e)  
{  
Response.Status = "301 Moved Permanently";  
Response.AddHeader("Location","http://www.sito.co  
m/nuova_pagina.html");  
}
```

ASP .NET - VB

```
Private      Sub      Page_Load(object      sender,  
System.EventArgs e)  
Response.Status = "301 Moved Permanently";  
Response.AddHeader("Location","http://www.sito.co  
m/nuova_pagina.html");  
End Sub
```

Coldfusion

```
<CFHEADER statuscode="301" statustext="Moved  
Permanently">  
<CFHEADER name="Location"  
value="http://www.sito.com/nuova_pagina.html">
```

Perl

```
#!/usr/bin/perl -w  
use strict;  
print "Status: 301 Moved Permanently\n";  
print "Location:  
http://www.sito.com/nuova_pagina.html\n\n";  
exit;
```

Posso eseguire un redirect in html?

Sì. Il redirect in html viene chiamato anche ironicamente “redirect dei poveri” in quanto viene adottato da quelle persone che non hanno le possibilità o le capacità di effettuarne uno mediante i linguaggi di programmazione citati sopra.

La redirezione in questo caso avviene mediante un meta tag refresh settato a 0 secondi.

Questo non è un vero e proprio redirect 301 ma lo emula, e i motori di ricerca nel corso degli anni hanno

accettato questa soluzione paragonandola alle altre a patto che:

1. Il tempo del redirect sia settato a zero secondi
2. La pagina non deve contenere nessun altro tag a parte il meta tag del redirect

Non seguendo queste due semplici regole si potrebbe rischiare l'esclusione dai motori di ricerca per un tentativo di cloaking.

HTML

```
<META HTTP-EQUIV=Refresh CONTENT="0;
URL=http://www.sito.com/nuova_pagina.html">
```

URL Rewriting: come e perchè

Abbiamo visto nei capitoli iniziali come i motori di ricerca privilegino le pagine web con un URL chiaro ed esplicito rispetto a quelle anonime e piene di variabili, quindi ora dalla teoria passeremo alla pratica andando a toccare con mano questa tecnica.

Prima di avviare una campagna di URL rewriting è bene fare un planning per decidere quali saranno le strategie da adottare in quanto è bene riscrivere gli URL sin dall'inizio e non apportare più delle modifiche.

Come è strutturato un file .htaccess?

Per ben capire la struttura di questo file riportiamo un piccolo esempio che trasforma una pagina da home.php a index.html

```
RewriteEngine On  
RewriteRule ^index.html$ home.php [L]
```

La prima riga attiva la modalità di rewriting.

La seconda riga dichiara una regola composta da:

- **Un URL da riscrivere**

Di cui avremo la stringa tra ^ (inizio) e \$ (fine), in questo caso abbiamo un nome statico ma quasi

sempre troveremo un'espressione regolare.

- **L'URL al quale inviare la richiesta**

Anche in questo caso abbiamo un nome statico ma in casi più articolati potremmo addirittura passare dei parametri al file.

- **Parametri**

Tra parentesi quadre separate da virgole potremo inserire dei parametri:

1. L : svolge le stesse funzioni del break e comunica all'engine di fermarsi non appena trova un regola adatta
2. QSA : acronimo di “accoda query string” (query string append) ci permette di accodare parametri se l'url richiesto contiene più parametri di quanti specificati.
3. R=301 : invia un header con status code 301 Moved Permanently
4. R=302 : invia un header con status code 302 Moved Temporarily

Come si effettua una riscrittura di URL?

C'è da premettere che la conoscenza delle espressioni regolari è propedeutica all'utilizzo un pò più avanzato

del l'URL rewriting. Non essendo le espressioni regolari argomento di questo libro non ne staremo qui a parlare, ma per rendere questo capitolo alla portata di tutti commenterò in modo dettagliato gli esempi qui riportati sperando di renderli di facile comprensione.

Riscrittura semplicemente

```
RewriteRule      ^prodotto-([0-9]+)\.html$
prodotto.php?id=$1 [L]
```

URL iniziale:

<http://www.sito.com/prodotto.php?id=123>

URL finale: <http://www.sito.com/prodotto-123.html>

- ([0-9]+) indica uno o più caratteri compresi tra 0 e 9
- \.html sta per .html, essendo . (punto) un metacarattere ne dobbiamo effettuare l'escape
- \$1 viene sostituito con il valore restituito dall'espressione regolare

Riscrittura con più parametri

```
RewriteRule      ^(.+)-([0-9]+)\.html$
prodotto.php?id=$2&nome=$1 [L]
```

URL iniziale:
<http://www.sito.com/prodotto.php?id=123&nome=affettatrice>

URL finale: <http://www.sito.com/affettatrice-123.html>

- (.+) indica uno o più caratteri qualsiasi
- ([0-9]+) indica uno o più caratteri compresi tra 0 e 9
- \.html sta per .html, essendo . (punto) un metacarattere ne dobbiamo effettuare l'escape
- \$1 viene sostituito con il valore restituito dalla prima espressione regolare
- \$2 viene sostituito con il valore restituito dalla seconda espressione regolare

Riscrittura con associazione ID e stringhe

```
RewriteRule ^([^\/]+)\.html$ reparto.php?nome=$1  
[L]
```

URL iniziale:
<http://www.sito.com/reparto.php?nome=bricolage>

URL finale: <http://www.sito.com/bricolage.html>

- ([^\/]+) indica uno o più caratteri qualsiasi, eccetto lo slash
- \.html sta per .html, essendo . (punto) un

metacarattere ne dobbiamo effettuare l'escape

- \$1 viene sostituito con il valore restituito dall'espressione regolare

A questo punto è necessario creare uno script in PHP che dato un nome ci restituisca l'id corrispondente.

```
<?
$reparto = $_GET["nome"];
$resultato = mysql_query("SELECT id FROM reparti
WHERE nome='$reparto'");

if (mysql_num_rows($resultato) == 0)
{
    // Nessuna corrispondenza trovata. Gestire l'errore
    [...]
}
else
{
    $id_reparto = mysql_fetch_row($resultato);

    $resultato2 = mysql_query("SELECT * FROM
prodotti where id_reparto=".$id_reparto[0]);
    // Continuare con gestione dei dati estrapolati dalla
base di dati
    [...]
}
?>
```

E' possibile ottimizzare un file .htaccess?

Non esiste una vera e propria ottimizzazione di un file .htaccess ma ci sono dei piccoli espedienti per aumentare le prestazioni del sistema.

Se abbiamo accesso root al nostro server una buona strategia è quella di caricare le regole nel file di configurazione cosicché vengano analizzate una volta sola, infatti ad ogni accesso il file .htaccess viene aperto, parsato e chiuso con un'ingente spreco di risorse.

Un'altra buona abitudine da prendere è quella di inserire (tranne casi specifici) il parametro [L] in modo da non far testare al sistema le regola successive.

Infine l'ultima norma è come sempre il buon senso evitando di caricare inutili regole che ad esempio trasformino le pagine con estensione da .php a .html .

I FATTORI

Il PageRank

Quasi tutti sanno o pensano che il PageRank sia un valore numerico assegnato ad una pagina web, ma in realtà è un algoritmo di analisi che assegna un valore numerico ad ogni elemento di un insieme di documenti in relazione tra loro mediante collegamenti ipertestuali.

Il PageRank come lo conosciamo noi nel World Wide Web è solo una delle possibili applicazioni, infatti nessuno ci vieterebbe di applicare il PageRank ai file presenti nella nostra cartella documenti per quanto possa essere un'idea bizzarra.

L'algoritmo del PageRank prende parte del nome dal suo autore Larry Page (uno dei fondatori di Google) ed è stato brevettato dalla Stanford University ed è attualmente un marchio di Google.

Come funziona il PageRank?

Il PageRank può essere considerato una versione più completa della “Link Popularity”, infatti a differenza della seconda tiene conto anche di diversi altri fattori tra cui il più importante è sicuramente il peso del sito linkante.

La formula dell'algoritmo inizialmente sviluppata è:

$$PR[A] = (1 - d) + d \left(\frac{PR[T1]}{C[T1]} + \dots + \frac{PR[Tn]}{C[Tn]} \right)$$

- $PR[A] \Rightarrow$ Valore PageRank della pagina A che vogliamo calcolare
- $T1...Tn \Rightarrow$ Pagine che contengono almeno un link verso A
- $PR[T1] \dots PR[Tn] \Rightarrow$ Valori di PageRank delle pagine $T1 \dots Tn$
- $C[T1] \dots C[Tn] \Rightarrow$ Numero complessivo di link contenuti nella pagina che offre il link
- d (damping factor) \Rightarrow è un fattore deciso da Google che nella documentazione originale assume valore 0,85. E' soggetto a modifiche da parte di Google per decidere la percentuale di PageRank che deve passare da una pagina all'altra.

Il PageRank classico che conosciamo assume dei valori oscillanti fra 0 e 10, ma questo è solo un valore simbolico infatti si ipotizza che il PageRank assuma una scala del genere:

PageRank 1 \Rightarrow 100

PageRank 2 \Rightarrow 1.000

PageRank 3 \Rightarrow 10.000

PageRank 4 => 100.000

PageRank 5 => 1.000.000

I numeri sono solo un'ipotesi ma ci accorgiamo benissimo che passare da PageRank 0 a PageRank 4 è quasi una passeggiata, ma passare da PageRank 4 a PageRank 5 inizia a diventare molto più difficoltoso. Ottenere un PageRank 6 o 7 non è un gioco da ragazzi e per alcuni sarà un obiettivo inarrivabile, raggiungere un PageRank 10 è impossibile a meno di non essere siti di rilevanza internazionale e popolarissimi come Google.com o Yahoo.com, basti pensare che Google.it arriva solamente ad un “misero” PageRank 9.

Tutti i link hanno lo stesso peso?

Non tutti i link hanno lo stesso peso e questo a causa di almeno 4 fattori, anche se poi come sappiamo i fattori presi in considerazione dagli algoritmi dei motori di ricerca sono molteplici.

L'algoritmo ci fa capire che i primi 2 fattori sono:

1. PageRank del sito linkante
2. Numero dei link in uscita

Poi abbiamo altri due fattori legati all'analisi fatta dal motore di ricerca:

3. Attinenza fra i siti

4. Testo del link

Se verremo linkati da una pagina con PageRank 0 sicuramente non avremo alcun beneficio, invece se a linkarci sarà una pagina con PageRank potremmo avere degli influssi positivi.

Il condizionale è d'obbligo perchè il PageRank acquisito dipende in modo consistente dal numero di link in uscita che ha quella pagina.

Possiamo fare un esempio per capire meglio come funzionano le cose: supponiamo di ricevere due link, uno da una pagina con PageRank 1 e un link in uscita, ed un altro da un PageRank 5 con 8 link in uscita.

Per semplificare, ma preciso che tale esempio non rispecchia la realtà dei fatti, possiamo ipotizzare che il PageRank venga diviso in modo equo tra tutti i siti linkati, quindi nel primo caso riceveremo $1/1 = 1$, nel secondo caso $5/8=0.6$.

L'esempio non è molto corretto ma ci ha fatto capire che non sempre venire linkati da una pagina con PageRank alto porta più benefici.

All'analisi quantitativa va però affiancata anche l'analisi qualitativa di un link.

Tra due pagine di due settori diversi ma con lo stesso PageRank e numero di link in uscita, quella che ci porterà più benefici sarà sicuramente quella che ha un contenuto attinente con il nostro.

Questo avviene sia per evitare che scambi di link reciproci tra utenti possano portare a scalare di SERP senza merito, sia perchè è senza dubbio più autorevole una segnalazione da parte di un sito del proprio settore che una di un settore differente.

Per fare un esempio stupido: l'opinione di un contadino sul raccolto di un altro contadino è sicuramente più attendibile di quella di un meccanico.

Altro aspetto importante è ovviamente il termine sul quale viene inserito il link, anche per questo vale in discorso della correlazione fra i siti.

Quanto è importante il PageRank?

Da diversi anni la maggior parte dei webmasters fa sforzi abnormi per far crescere il PageRank dei propri siti, alcuni perchè credono siano l'unico fattore che li aiuterà nel posizionamento, altri semplicemente perchè è diventata una moda.

In realtà come ben sappiamo il PageRank è solo uno dei tanti fattori che incidono sul posizionamento, e non è neanche il principale come molti altri credono.

Per molti anni la posizione nelle SERP di Google è stata fortemente determinata dal PageRank, ma Google per evitare che i risultati delle ricerche fossero per un certo verso taroccati da scambio link e acquisto di link ha successivamente apportato delle modifiche ai propri

criteri di posizionamento delle pagine riducendo di fatto l'influenza del PageRank.

Ma il PageRank non è morto, ed è ancora largamente utilizzato in molti algoritmi come quello che decide tra due pagine “doppioni”, ad esempio due pagine con lo stesso capitolo della divina commedia, quale considerare di peso maggiore.

Concludendo se prima non era una buona mossa affidarsi solo al PageRank per posizionare un sito web, oggi è una mossa ancora più sbagliata.

Conviene comprare o scambiare dei link per alzare il PageRank?

Google ha dovuto affrontare molte sfide per sconfiggere i tentativi fraudolenti di aumentare il PageRank, primo fra tutti lo scambio link correndo ai ripari dando un peso praticamente nullo ai link reciproci e dando un peso minore ai link non a tema.

Negli ultimi anni però sono spuntati come funghi siti che vendono e comprano links, o meglio li affittano, e questo ha messo in grande difficoltà Google che pensava che un PageRank alto non avrebbe mai e poi mai linkato un sito di bassa qualità.

L'inganno era fatto, un webmaster che ne aveva la disponibilità economica poteva comprare tutti i links da PageRank alti che voleva e far schizzare il suo sito in alto nelle SERP, come risolvere?

Google nella metà del 2007 mise online un form di segnalazione per link a pagamento che dava l'impressione di una sorta di “denuncia online”.

Dopo un disorientamento iniziale con buona parte dei webmasters che si chiedevano se delle false segnalazioni li avrebbero potuti penalizzare, arrivò la smentita ufficiale di Google, le segnalazioni servivano solo per acquisire esempi per meglio realizzare un algoritmo di valutazione.

Una delle ultime soluzioni adattate è quella di una sorta di “limbo” nei quali i link restano per un certo periodo di tempo dopo il quale inizieranno a produrre dei benefici.

Voci di corridoio quantificano questo periodo in 6 mesi, ma non c'è troppo da fidarsi e non è detto che questo tempo sia una costante.

Il Trust Rank

Il TrustRank è una tecnica di analisi dei link di un sito web che ha lo scopo di distinguere i siti web “utili” da quelli spam.

Il 16 marzo del 2005 Google ha registrato presso la USPTO il marchio TrustRank.

Per un computer è difficile riconoscere un sito spam da uno buono soprattutto quando ci si trova davanti ad un network fortemente linkato, al contrario per uno staff di tecnici è un compito assai più semplice ma che richiede più tempo e maggiori spese col minor risultato in termini di quantità di controlli.

L'algoritmo del TrustRank può essere usato sia per suggerire ad uno staff di tecnici quali siti sospetti controllare, sia per generare una sorta di punteggio che funzioni da feedback durante l'ordinamento dei risultati.

Come funziona l'algoritmo del TrustRank?

L'algoritmo si basa sulla semplice deduzione che le pagine spam vengono create per ingannare i motori e non hanno contenuti, quindi nessun webmaster che ha dei siti “buoni” ha dei validi motivi per inserire link verso tali siti.

E' anche vero però che siti spam possono ricevere link da siti “buoni” mediante guestbook e chat non

moderate oppure degli spammers con maggiori risorse potrebbero creare una rete di siti con contenuti validi con il solo scopo di linkare i loro siti spam.

Per risolvere questa situazione è stato quindi messo a punto il cosiddetto “Trust Attenuation” per diminuire la fiducia passata alle altre pagine via via che ci si allontana dall'insieme delle pagine buone che sono sicuramente libere da spam di qualsiasi tipo.

L'algoritmo lavora in 3 fasi fondamentali:

1. Viene scelto un insieme relativamente piccolo di pagina seme (seed pages) delle quali non si conosce ancora la qualità;
2. Un essere umano divide l'insieme in due sottoinsiemi: quello della pagine buone e quello della pagine di spam;
3. L'algoritmo processa tutte le pagine contenute nel suo indice e assegna un punteggio ad ognuna di esse basandosi sulla vicinanza delle pagine seme.

Una volta divise le seed pages in pagine buone (S+) e pagine di spam (S-) la funzione assegnerà un punteggio di 1 alle pagine S+, 0 alle pagine S- e 0,5 alle pagine che non appartengono all'insieme S e delle quali non ha quindi alcuna informazione. Ci troviamo nella fase chiamata Ignorant Trust in quanto l'algoritmo non ancora tiene conto delle relazioni fra pagine.

Successivamente l'algoritmo inizia a trasmettere punteggi di fiducia alle pagine direttamente collegate alle seed pages secondo questa formula: se una pagina ha un punteggio di fiducia "T" e contiene "n" links ad altre pagine, ad ognuna di esse verrà trasmesso un punteggio di fiducia pari a T/n .

Facciamo un esempio pratico:

- La pagina 1 e la pagina 2 appartengono all'insieme S+
- La pagina 3 non appartiene all'insieme S
- La pagina 1 linka 2 pagine tra cui anche la pagina 3
- La pagina 2 linka 4 pagine tra cui anche la pagina 3

La pagina 1 assegnerà ad ognuna delle pagine che linka un punteggio di Trust di 0,5 (la metà della sua fiducia), mentre la pagina 2 assegnerà un punteggio di 0,25 (un quarto della sua fiducia).

La pagina 3 essendo linkata sia dalla pagina 1 che dalla pagina 2 riceverà un punteggio totale di $0,5 + 0,25 = 0,75$.

Dopo aver invocato l'algoritmo un numero di volte sufficiente a coprire l'intero indice delle pagine web si deve decidere un valore di soglia per il quale tutte le pagine che supereranno tale valore saranno da considerare buone e potranno essere ordinate per essere

contenute nell'indice del motore di ricerca.

Come vengono scelte le seed pages?

Finora abbiamo trascurato questo aspetto dando per scontato che l'insieme S delle seed pages esistesse già.

La scelta delle seed pages è invece un passaggio molto delicato che può determinare pesantemente la buona riuscita dell'algoritmo.

Per determinare le seed pages sono stati proposti ed analizzati due approcci.

Il primo chiamato PageRank inverso deriva dalla stessa formula del PageRank con l'unica differenza che il punteggio non viene determinato dai link in ingresso ma da quelli in uscita. Un sito con molti link quindi, a patto che sia di qualità, potrebbe aiutare molto nella determinazione degli altri siti di qualità avendo molti collegamenti ad essi.

Il secondo approccio invece si basa totalmente sul PageRank ed assegna punteggi maggiori ai siti con il PageRank più alto. Essendo inoltre i siti di qualità con PageRank alto spesso collegati fra di loro questo aiuterebbe ad accrescere ancora di più la loro autorevolezza.

E' stato condotto uno studio fra entrambi i metodi su

una porzione ridotta del web, ed anche se lo scarto è stato davvero minimo, il sistema che ha dato i migliori risultati è stato proprio il PageRank inverso.

A questo punto è partita la prima vera e propria esecuzione dell'algoritmo su scala globale per determinare i siti web che sarebbero entrati nella rosa dei seeds.

La prima fase è stata il calcolo del PageRank inverso, avvenuta utilizzando un fattore di attenuazione di 0,85 e reiterando l'algoritmo per ben 20 volte.

Dopo aver ordinato i risultati in ordine decrescente sono stati scelti i primi 25.000 siti che hanno subito un'ulteriore filtraggio automatico eliminando i siti cloni e quelli che non erano presenti in nessuna della maggiori web directory.

Al termine di questa prima fase i siti sono rimasti solo 7900 siti, e di questi solo i primi 1.250 in ordine di PageRank inverso sono andati a formare l'insieme S dei seeds e sono stati controllati manualmente.

Di questi 1.250 siti solo 178 hanno avuto l'onore di essere considerati S+ e di entrare nella rosa dei seed sites.

Ma non pensate che su 25.000 siti solamente 178 sono risultati non spam, infatti i criteri di scelta di questi 178 siti sono stati molto rigorosi e per garantire una durata a medio-lungo termine di questo insieme di siti seme sono stati scelti solamente siti di istituzioni stabili come quelle governative, universitarie, militari, ecc...

Ulteriori studi hanno dimostrato però che questi seed sites per essere davvero ottimi devono coprire uniformemente tutti gli argomenti del web e le varie aree geografiche.

Il Sandbox

Il Sandbox, e non “la Sandbox” come molti erroneamente scrivono, è una teoria molto accreditata riguardo una sorta di purgatorio virtuale nel quale rimangono per un periodo di tempo i siti nuovi.

Sul Sandbox si sono creati i migliori miti ed inventate le migliori fiabe, ma non essendoci mai state conferme ufficiali si possono fare solo supposizioni e alcuni SEO addirittura affermano che il realtà non esista.

Tale termine ha fatto il suo ingresso nel mondo SEO nel marzo del 2004 quando si iniziarono ad intravedere degli strani cambiamenti dopo i due aggiornamenti di Google chiamati “Florida” e “Austin”, ma solo un anno dopo nell'aprile del 2005 con l'aggiornamento “Allegra” si notarono i primi significativi cambi di ranking per i siti nuovi e il mondo SEO iniziò a tremare.

Da qui si scatenarono supposizioni ed invenzioni varie, ma le teorie più accreditate sono due.

La prima vuole che Google abbia creato questa parte di algoritmo chiamato Sandbox per contrastare l'operato di alcuni spammer che creando decine e decine di siti web si assicurano una buona quantità di traffico anche se il sito venisse bannato dopo soli 30 giorni. Tale filtro ovviamente ha avuto ripercussioni su tutti i nuovi siti web che vengono indicizzati con la solita rapidità senza dargli però troppa visibilità nelle SERP.

La seconda teoria invece ipotizza che Google non abbia mai implementato l'algoritmo Sandbox e che questo

altro non è che un effetto generato dall'interazione fra alcune parti dell'algoritmo come quella relativa all'età del dominio e al traffico generato.

Ad avvalorare questa teoria ci sono anche le testimonianze di molti webmasters che per qualche settimana o qualche mese hanno visto letteralmente crollare i loro siti nel posizionamento per poi vederli tornare nuovamente dove erano prima, ma in questi casi è difficile capire se si tratti di questo effetto Sandbox o di una vera e propria penalizzazione.

Entrambi le teorie sono più che valide ma essendo solo supposizioni vanno prese con le dovute precauzioni.

Quali osservazioni hanno scatenato le supposizioni sul Sandbox?

A partire da Marzo 2005 si sono iniziati a studiare da vicino gli effetti di questo fantomatico Sandbox che sembra affliggere il 99% dei siti web di nuova creazione.

Si è notato che i siti web nuovi vengono indicizzati con la solita rapidità di Google ma poi per un periodo di tempo variabile ricevono scarsissimi risultati nelle SERP indipendentemente dai loro contenuti e dalla loro popolarità.

Tale Sandbox sembra colpire maggiormente, e durare per un periodo di tempo più lungo, i siti ottimizzati per parole chiave molto competitive e quelli che ricevono da

subito troppi link in ingresso o presentano da subito una quantità di contenuti esagerata.

Dopo un certo periodo i nuovi siti vengono propriamente classificati e riescono a concorrere per il posizionamento.

Quanto tempo dura il Sandbox?

Si può stimare un tempo medio tra i 3 ed i 6 mesi, ma la verità è che non esiste un tempo prestabilito oppure un metodo per poterlo determinare.

Possiamo dire che più il settore è competitivo e più lunga sarà la nostra permanenza nel Sandbox, ma addirittura c'è chi non ne è mai uscito...

Il Sandbox è un problema che ti ritroverai ad affrontare volente o nolente, quindi invece di restare passivi ad aspettare la fine è una buona strategia utilizzare questo periodo per pubblicizzare il tuo sito, trovare link di qualità da altri siti, stabilire partnerships, ecc...

Si può uscire prima dal Sandbox?

Ovviamente no, ma si possono usare dei piccoli accorgimenti per ridurre i tempi d'attesa.

Si ipotizza che il Sandbox dipenda principalmente dall'Age Factor e del Traffic Factor, ovvero

rispettivamente dall'anzianità di un dominio e dal traffico che riceve in proporzione al suo settore.

Alcuni quindi consigliano di comprare domini scaduti prima che questi vengono cancellati dagli indici di Google in modo da sfruttare la loro “anzianità”, ma questa è una strategia da adottare con le pinze in quanto bisogna verificare per bene cosa conteneva quel dominio e chi era il proprietario per evitare spiacevoli sorprese in futuro.

Per quanto riguarda il secondo fattore invece devi vedere tutto in quest'ottica: “Non dovrai posizionarti per ricevere traffico, ma dovrai ricevere traffico per posizionarti”, sembra infatti che per essere considerato una risorsa utile, ed essere quindi inserito nelle SERP, un sito deve generare del traffico esterno per non far sembrare che sia Google dipendente.

Una raccomandazione è quella di non eccedere, soprattutto nei primi tempi, con strategie troppo aggressive come riempire il sito di centinaia e centinaia di pagine ed iscriversi a decine e decine di directory per ricevere link, tutto dovrà essere fatto lentamente e nel modo più naturale possibile.

Come vedo se il mio sito è nel Sandbox?

Stabilire se un sito è nel Sandbox è abbastanza semplice, basta prendere una frase abbastanza complessa

direttamente dalla homepage ed effettuare la ricerca.

Se ad esempio cerchiamo “Vendiamo bellissimi oggetti dell'artigianato abruzzese da decine di anni” ci saranno molti altri siti che hanno queste parole nel testo delle loro pagine, ma siccome solo il tuo sito riporta esattamente quella frase dovrai essere in prima posizione, se così non fosse purtroppo sei nel Sandbox.

Molti effettuano la verifica del Sandbox inserendo il nome del proprio sito web, ma questa è una pratica errata che non ci dice nulla sull'uscita o meno dal Sandbox.

La link popularity

La link popularity, in italiano popolarità da link, rappresenta il numero di links in ingresso di una pagina web tenendo conto sia dei links provenienti da siti web esterni sia dei links provenienti dalle pagine interne dello stesso sito.

I fondatori di Google Larry Page e Sergey Brin notarono che i testi accademici più prestigiosi non venivano determinati in base alle copie vendute, bensì in base alle citazioni ricevute dai loro pari che gli attribuivano una certa importanza e affidabilità, i due ragazzi pensarono che tale concetto potesse essere esteso anche al web, infatti un webmaster linka un sito web solo nel caso in cui lo reputa una risorsa valida ed utile per i suoi utenti.

Proprio da questa deduzione nacque un'idea di successo e Google iniziò a fornire ai proprio utenti come risultati quei siti maggiormente linkati che venivano quindi proposti dal web stesso come contenuti di qualità.

La link popularity viene tutt'oggi utilizzata dalla quasi totalità dei motori di ricerca anche se nel corso degli anni ci sono stati notevoli evoluzioni a causa di molti webmasters scorretti che per apparire tra i primi posti nei motori di ricerca creavano centinaia di siti web con il solo scopo di linkarsi tra di loro per aumentare la loro popolarità.

La link popularity intesa come numero di link in

ingresso era quindi soggetta a manipolazioni e per questo motivo i motori di ricerca iniziarono ad utilizzare la link analysis.

Con un sistema di analisi dei links è più facile individuare tecniche di spam come l'artificial linking, ovvero la creazione di siti con l'unico scopo di linkare un sito di destinazione, e quantificare la qualità degli stessi.

Non tutti i links hanno la stessa importanza e quindi non tutti i links devono portare gli stessi benefici, è ovvio che un link spontaneo da un sito a tema deve avere più valore di uno scambio link con un sito che non ha nulla a che fare con il nostro.

Per puntare ad un buon posizionamento nei motori di ricerca è quindi meglio ricevere pochi links da siti a tema che molti links da siti che non trattano i nostri stessi argomenti, inoltre è importantissimo essere linkati dalle più importanti web directory come Dmoz.

Quali links sono considerati validi?

Essendo la link popularity valutata dagli spider, sono validi tutti i links leggibili da essi.

Links in Javascript, sottoforma di popup o redirect non saranno presi in considerazione.

Anche la maggior parte dei links nelle animazioni Flash, come spiegato nei capitoli precedenti non verranno

presi in considerazione.

Come posso aumentare la link popularity?

Il miglior metodo per aumentare la link popularity è quello di comportarsi come se non i motori di ricerca non esistano e dedicarsi alle seguenti attività:

1. Costituire delle partnership con altri webmasters del settore
2. Selezionare nuovi siti sui quali registrarsi
3. Dedicarsi alla scrittura di articoli

Essendo però questo un processo che una volta iniziato non si conclude, durando nel corso degli anni, è bene stabilire un quantitativo di energie da dedicargli per non sprecare tempo a discapito magari dei contenuti.

E' possibile comprare dei links per aumentare la link popularity?

Comprare dei links per aumentare la link popularity non è di certo la soluzione migliore e nei precedenti capitoli abbiamo anche parlato di quanto questa tecnica non sia ben vista dai motori di ricerca, in primis da Google. Tuttavia ci sono casi e casi, comprare ad esempio dei

links a caso da uno dei tanti servizi online che offrono questa opportunità ritengo sia una scelta scellerata, invece pagare un sito del proprio settore per inserire un link in homepage o per una recensione potrebbe essere un'idea migliore.

Tuttavia se vi ritrovate a pagare per ricevere dei links ritengo che avete sbagliato tutto fino a questo punto e la domanda che vi dovrete porre è: perchè non riesco a ricevere links spontanei?

La risposta a mio avviso è: perchè il tuo sito non è di qualità.

Un sito web di qualità che offre contenuti e servizi non avrà bisogno di pagare per essere segnalato, quindi il mio consiglio è di investire il vostro budget nella realizzazione di contenuti e non nell'acquisto di links.

L'importanza dei contenuti aggiornati

L'aggiunta o l'aggiornamento dei contenuti di un sito web possono essere fattori determinanti ai fini del posizionamento nei motori di ricerca.

Un sito aggiornato di frequente non sempre è di qualità, allo stesso modo un sito poco aggiornato non significa che sia di scarsa qualità, ma il trend attuale dei motori di ricerca è quello di premiare i siti che propongono ai loro utenti contenuti sempre nuovi e al passo con i tempi.

Google ad esempio tramite il suo GoogleBot effettua due tipi di indicizzazione, la Fresh-Crawl che esegue un controllo veloce e mirato sulle pagine web per verificare eventuali aggiornamenti o correzioni e la Deep-Crawl che altro non è che l'indicizzazione classica che va più in profondità e indicizza i contenuti del nostro sito web.

Grazie a questo approccio, non molto dissimile da quello utilizzato dagli altri motori di ricerca, gli spiders visiteranno i siti web molto aggiornati anche a distanza di pochi minuti verificando la presenza di nuovi contenuti per indicizzarli successivamente.

I contenuti aggiornati aiutano anche il posizionamento?

Molti pensano che l'inserimento di contenuti nuovi aiuti solo l'indicizzazione e non abbia alcun effetto sul posizionamento, ma con l'avvento dei blog, i motori di

ricerca hanno cambiato il loro modo di posizionare le pagine, infatti blog che postano notizie su un argomento nuovo raggiungono facilmente le prime posizioni, ma succede spesso che altri blog che a distanza di qualche giorno postano la stessa notizia si ritrovano a scavalcare le concorrenza in quanto vengono considerati contenuti freschi e migliori.

Da qui nasce la necessità per i siti web di mantenere costantemente aggiornati i loro contenuti postando degli approfondimento e con l'aiuto degli utenti che inserendo commenti aumentano il “rumore” della pagina.

Questo dimostra quanto possa essere determinante per il posizionamento e non solo per l'indicizzazione la pubblicazione di nuovi contenuti quotidianamente.

Come fanno gli spiders a riconoscere i contenuti aggiornati?

La frequenza e l'entità delle modifiche di un sito web possono essere determinate dai motori di ricerca mediante l'utilizzo di “copie” che sono delle rappresentazioni delle pagine web che permettono di non archiviare l'intero documento con un grande risparmio in termini di spazio e banda.

Sulla base di queste copie i motori di ricerca sono in grado di determinare il peso di un aggiornamento, etichettando ad esempio come non importante il

cambiamento degli script pubblicitari, ed anche di generare un “codice di somiglianza” che potrà essere usato per rilevare doppioni e documenti simili.

Su che base lo spider decide con che frequenza visitare un sito web?

Attraverso le copie dei siti web i motori di ricerca sono in grado di generare una sorta di profilo per ogni sito web e di analizzarne il ritmo di crescita.

Se questo parametro sarà cresciuto rispetto al periodo precedente di conseguenza dovrà essere incrementata anche la frequenza di indicizzazione, allo stesso modo se un sito web ha subito una battuta d'arresto la frequenza di scansione dello spider potrà essere diminuita a favore di altri siti web.

Cos'è il brevetto sui Dati Storici?

Il 31 Marzo 2005 Google ha registrato un brevetto dal titolo Information retrieval based on historical data (Reperimento di informazioni basato su dati storici) risalente al 31 Dicembre 2003.

Nella sezione principale del testo intitolata Exemplary history data, ci sono la bellezza di 99 paragrafi suddivisi

in 14 sottocategorie, nei quali vengono descritti molteplici algoritmi utili per l'archiviazione e la restituzione di dati e informazioni.

In tale brevetto nell'analisi dei dati storici sono presenti fattori scontati come l'aggiornamento dei contenuti di un documento ma anche alcuni inaspettati come le ricerche effettuate dagli utenti, le loro scelte nella pagina dei risultati, peso e “freschezza” di un link, ecc...

Dal numero di fattori presenti nel brevetto è facile intuire quanto possano essere importanti, e oserei dire necessari, gli aggiornamenti di una pagina web.

Perchè i blog sono così efficaci nell'indicizzazione?

I blog sono, almeno nella maggior parte dei casi, aggiornatissimi e pieni di contenuti, sono sempre linkati da altri blog o siti grazie alla pratica diffusa dei blogroll ed infine raccolgono le discussioni degli utenti mediante i commenti.

Tutti questi fattori determinano una crescita ed un andamento naturale dei siti in questione che sono molto ben visti dai motori di ricerca.

Una scelta vincente può quindi essere quella di inserire sempre un blog nei nostri siti o meglio di modellare il nostro sito affinché abbia le stesse caratteristiche dei blog ovvero dinamicità e ricchezza di contenuti.

Trucchi da lasciar perdere

Nell'ambito SEO parlando di spam si fa riferimento a tutte quelle tecniche che puntano a far scalare le SERP ad un sito web in modo artificioso.

I motori di ricerca negli ultimi anni stanno spendendo una quantità di risorse sempre maggiore per combattere questo fenomeno ed offrire contenuti sempre più di qualità, arrivando ad un buon livello di affidabilità che rende difficile ingannarli.

Ma difficile non significa impossibile e la battaglia è ancora aperta.

Da questo ambito nasce la figura del black hat SEO, che si differenzia da un normale SEO per l'utilizzo di tecniche illecite per quanto riguarda le linee guida dei motori.

Di seguito andremo ad analizzare alcune di queste tecniche più famose focalizzandoci esclusivamente sulla descrizione e tralasciando la loro implementazione per ovvi motivi.

1. **Key Stuffing**

E' una delle tecniche più vecchie e consiste nel ripetere con un'ossessione quasi maniacale una parola chiave all'interno di un testo, senza curarsi dalla leggibilità. Questa tecnica è stata utilizzata dapprima nel testo, successivamente nei meta tag ed oggi nel tag title che essendo una delle zone

del sito più importanti è il bersaglio preferito dagli spammers.

2. **Testo Nascosto**

Questa tecnica permette di ingannare sia gli utenti che i motori di ricerca. Utilizzando uno dei tanti metodi per nascondere del testo ottimizzato all'interno di una pagina web si può mostrare agli utenti un sito web normale e far indicizzare nel frattempo agli spiders il testo nascosto traendone vantaggi ai fini del posizionamento.

3. **Doorways**

Questa tecnica consiste nel creare delle pagine ottimizzate che non offrono alcun tipo di informazione o servizio ma hanno il solo scopo di reindirizzare l'utente in modo automatico verso un'altra pagina o sito web.

4. **Cloaking**

Questa è una delle tecniche più raffinate che ha portato e porta più problemi ai motori di ricerca essendo lato server e quindi più difficilmente riconoscibile. Il cloaking si basa sulla semplice, almeno a parole, tecnica di identificare gli spiders attraverso IP, user agent, ecc... per mostrargli una pagina ottimizzata a differenza dei normali utenti ai quali verrà mostrata una pagina del tutto differente e magari piena di pubblicità.

5. **Artificial Linking**

Questa tecnica consiste nella creazione di siti web che hanno il solo scopo di linkare un sito web di destinazione. Questa tecnica di spam ha due implementazioni: una quantitativa e una qualitativa. La prima implementazione consiste nel creare un elevato numero di siti web senza curarsi del loro PageRank e del peso dei loro link, puntando esclusivamente sulla loro quantità. La seconda implementazione invece punta alla qualità creando solo qualche decina di siti che abbiano però dei contenuti ed una certa importanza da trasmettere al sito di destinazione.

Come abbiamo visto le tecniche di spam sono davvero molto facili da realizzare, efficaci e allo stesso tempo difficili da riconoscere per un computer, per questo molti motori di ricerca hanno affiancato al lavoro già eccellente dei loro spiders dei “Quality Rater” che sono delle persone istruite per riconoscere i siti spam e con il compito di assegnare loro delle penalità in base alla gravità riscontrata.

Oltre ogni discorso etico e morale riguardo il non usare questi trucchetti c'è da dire che una tecnica di spam potrà portarvi anche dei risultati nel breve termine, prima o poi però verrete beccati e non vi rimarrà nulla

in mano in quanto un sito web che non appare nei motori di ricerca è destinato a morire nell'anonimato.

L'importanza di DMOZ

DMOZ è un Open Directory Project (ODP) ed è stata la prima directory di siti web, divenuta importante grazie alla sua serietà ed organizzazione.

DMOZ è costruita grazie alle segnalazioni degli utenti che vengono processate dai suoi editori, un team di oltre 10.000 volontari sparsi in tutto il mondo.

A differenza delle altre web directory, essendo un ODP, DMOZ permette l'utilizzo del proprio database a qualunque altro sito web, e basta vedere solamente quanti siti in Italia utilizzano i dati di Dmoz per accorgerci di quanti links si possano ricevere grazie al semplice inserimento in questa directory.

Il vantaggio sta quindi nell'isciversi nella directory più famosa al mondo per ricevere centinaia e centinaia di links, anche di un certo peso, da altri siti web a tema piuttosto che iscriversi a centinaia di directory differenti di dubbia qualità.

Perchè è così importante essere presenti in DMOZ?

Abbiamo già spiegato a grandi linee perchè è importante essere presenti in DMOZ, ma i motivi sono molteplici, eccone alcuni:

- **DMOZ ha un PageRank alto**
Attualmente di 8/10, nel passato è stato anche di 9/10. Riuscendo ad apparire in una categoria alta, quindi con un buon PageRank, riceveremo un link che ci potrà portare dei vantaggi inimmaginabili.
- **Moltissimi siti utilizzato i dati DMOZ**
Come già detto essendo un ODP, i dati di DMOZ sono presenti in centinaia e centinaia di altri siti web che porteranno altrettanti link al nostro sito web. Il sito più celebre che utilizza questi dati è la directory di Google.
- **Visibilità ai motori di ricerca**
Molti motori di ricerca utilizzano i dati DMOZ per cercare al loro interno nuovi siti web da inserire nei propri indici.
- **Link a tema**
Essendo una directory strutturata in categorie, tutti i links generati da DMOZ e i loro siti satellite saranno a tema.

Dmoz ha perso importanza per il posizionamento?

Sicuramente fino a qualche anno fa la presenza in

DMOZ dava una bella spintarella al posizionamento di un sito web e veniva considerata un must per ottenere dei risultati.

Una svolta si ebbe nell'ottobre del 2008 quando Google eliminò dalle proprie linee guida la seguente dicitura:

Segnala il tuo sito alle principali directory, quali Open Directory Project e Yahoo!, così come ad altri siti specializzati in settori specifici.

Infatti oggi le cose sembrano cambiate e la presenza in tale directory non sembra dare vantaggi visibili ad occhio nudo.

Il web è cambiato e anche i motori di ricerca si adeguano e si rinnovano di conseguenza, ora non ci resta che aspettare e vedere quale sarà la prossima fonte di popolarità che alcuni SEO ritengono essere i social network o gli aggregatori di notizie.

Il rischio della over-optimization

Come possiamo intuire dal nome stesso l'over-optimization è un'eccessiva ottimizzazione ovvero una serie di accorgimenti e comportamenti forzati ed innaturali che hanno l'unico fine di migliorare il posizionamento a discapito però dell'usabilità.

E' bene capire però che sovraottimizzazione e spam sono due cose totalmente differenti, la prima infatti utilizza delle tecniche legali seppur innaturali, la seconda invece va ben oltre le linee guida dei motori di ricerca con delle tecniche che hanno l'unico scopo di ingannare i motori di ricerca.

Quando un'ottimizzazione diventa penalizzante?

Un'ottimizzazione diventa penalizzante quando è spinta, infatti i motori di ricerca non penalizzano l'ottimizzazione in sè per sè ma l'abuso che se ne fa.

Alcuni comportamenti che possono portare ad una penalizzazione sono i seguenti:

- **Keyword density alta**

Una keyword density normale si aggira tra il 3% e il 5%, una parola chiave che arrivi al 10% può indicare un tentativo di keywords stuffing.

- **Keyword ripetuta**

Ripetere più volte la stessa keyword non è consigliabile. Ad esempio invece di scrivere “vendita auto di lusso, vendita auto economiche vendita auto usate” è più opportuno scrivere “vendita auto di lusso, economiche ed usate”.

- **Keyword enfaticizzata**

E' sconsigliabile anche utilizzare il grassetto e/o il testo sottolineato ogni qual volta appare la parola chiave nel testo. In questi casi è evidente la volontà del webmaster di evidenziare sempre la stessa parola.

- **Abuso del tag ALT**

Questo TAG come ben spiegato in precedenza viene utilizzato ai fini dell'usabilità nel caso in cui l'immagine non possa essere visualizzata, ma molti webmasters non perdono occasione per infilarci la stessa keyword, o nei casi peggiori elenchi di keywords, in tutte le immagini del sito. Tale tecnica è ovviamente penalizzante.

Ottimizzare una pagina affinché raggiunga la vetta di una SERP è lecito, ma non bisogna eccedere per evitare di incappare in penalità da sovraottimizzazione (OOP Over Optimization Penalty) producendo così l'effetto contrario a quello desiderato.

Gli inbound links possono generare una OOP?

Sappiamo che più un sito è linkato più sarà meglio posizionato, ma questa non è una regola assoluta, questi links infatti non dovranno apparire artificiali per evitare di vedere azzerati i loro benefici.

Uno dei maggiori fattori presi in considerazione è sicuramente il tempo, infatti se un sito appena nato riceve sin da subito qualche centinaia di links o registra un improvviso aumento dei links in entrata quasi sicuramente sta mettendo in atto un'attività artificiale per spingere il sito web.

Ma anche nel caso in cui i links vengano ricevuti in un lasso di tempo maggiore ma presentino caratteristiche comuni come codice html simile, stesso indirizzo IP, ecc... i motori di ricerca li classificheranno come artificiali.

Tali comportamenti provocheranno l'assegnazione di una penalità con una conseguente caduta, più o meno significativa, dalle prime posizioni delle SERP.

Come si esce da questa penalizzazione?

Uscire da questo genere di penalizzazioni non è semplice perchè non sempre è ben chiaro il motivo per il quale si è stati penalizzati.

In questi casi bisogna quindi analizzare il sito con senso

critico e capire dove si è osato troppo per porvi rimedio rientrando nei limiti delle regole SEO.

Inoltre non tutte le penalizzazioni sono uguali, infatti troppi inbound links in un breve periodo possono determinare una penalizzazione temporanea che può essere trascurata aspettando che scompaia da sola.

Discorso differente per il keyword stuffing, tale penalizzazione infatti è permanente e può portare addirittura ad un banning del sito nel caso in cui non si ponga rimedio al problema.

Il mantenimento del posizionamento

Con molta fatica e grazie ad un ottimo lavoro di ottimizzazione e promozione il nostro sito web ha raggiunto un buon posizionamento nei motori di ricerca, ma non è finita qui.

Il posizionamento di un sito web infatti non termina con il raggiungimento della posizione desiderata ma prosegue per tutto il ciclo vitale del sito web per mantenere le posizioni raggiunte e cercare di migliorarle sempre più col passare del tempo.

Inoltre l'aggiornamento degli algoritmi dei motori di ricerca e la nascita di nuovi siti web determinano una variazione della visibilità del nostro sito web che può portarlo alla scomparsa dalle posizioni utili in caso di mancanza di interventi di manutenzione.

Quali interventi di manutenzione bisogna attuare?

Innanzitutto non è del tutto corretto chiamarli interventi di manutenzione, ma è molto meglio parlare di interventi di crescita e consolidamento.

Questi interventi sono divisi in due categorie: quelli di monitoraggio del posizionamento e quelli di aggiornamento del sito web.

Gli interventi di monitoraggio del posizionamento

servono per analizzare i risultati ottenuti ed elaborare nuove strategie in linea col trend attuale, tra questi troviamo:

- **Monitoraggio periodico**
Utile per controllare periodicamente il posizionamento del sito web nei diversi motori di ricerca per le keywords scelte e di nicchia.
- **Analisi delle statistiche di accesso,**
Tramite tali statistiche possiamo controllare le parole chiave più ricercate, i siti web dai quali provengono i visitatori, il tempo di permanenza nel sito, le pagine più visitate, ecc...
- **Monitoraggio dei web competitors**
E' buona norma infatti controllare il posizionamento dei nostri concorrenti, la nostra visibilità rispetto ad essi, la loro crescita, ecc... per poter prendere le adeguate contromisure per difendere la nostra posizione.

Gli interventi di aggiornamento del sito web invece servono per mantenere i risultati ottenuti mediante modifiche e aggiunte delle pagine web. Alcuni di questi interventi da attuare sono:

- **Incrementare la link popularity**
Cercando di ricevere sempre nuovi e maggiori links da siti a tema ed aumentare l'autorevolezza del sito per ricevere links spontanei.
- **Creare nuovi contenuti**
Mantendo un sito web sempre aggiornato e con contenuti sempre freschi e nuovi.
- **Trovare nuove keywords**
Analizzando anche il trend dei visitatori bisogna cercare nuove parole chiave a tema che possano portare visitatori al proprio sito web.

Inoltre per mantenere un buon posizionamento sui motori di ricerca bisogna essere sempre al passo con i tempi, ed i forum dedicati al posizionamento sui motori di ricerca sono uno strumento indispensabile per rimanere sempre aggiornati sulle tematiche SEO, per scambiare opinioni con SEO professionisti ed anche per partecipare a test di sperimentazione sulle nuove tecniche.

Conclusione

Un sito web ben posizionato nei motori di ricerca riceverà molte visite tramite le quali potrà generare consistenti guadagni, viceversa un sito che si trova in posizioni di retrovia riceverà poche visite generando di conseguenza scarsi guadagni, ed è proprio il fattore economico ad aver scatenato sin dagli albori del web la lotta serrata fra motori di ricerca e spammers.

Per diversi anni è stato facile ingannare i motori di ricerca in quanto i fattori da gestire erano pochi e facilmente frodabili, ma col tempo le cose sono cambiate.

Quando i motori di ricerca si limitavano a valutare i meta tag, attraverso il keyword stuffing gli spammers iniziarono ad alterare i risultati delle SERP accaparrandosi le prime posizioni.

Quando i motori di ricerca cominciarono ad analizzare i testi delle pagine web, gli spammers iniziarono ad inserire testi invisibili agli utenti utilizzando lo stesso colore dello sfondo, formattazioni speciali tramite css e javascript, ecc...

Quando i motori di ricerca cominciarono a valutare la link popularity, la risposta degli spammers non si fece attendere creando vere e proprie fabbriche di links (links

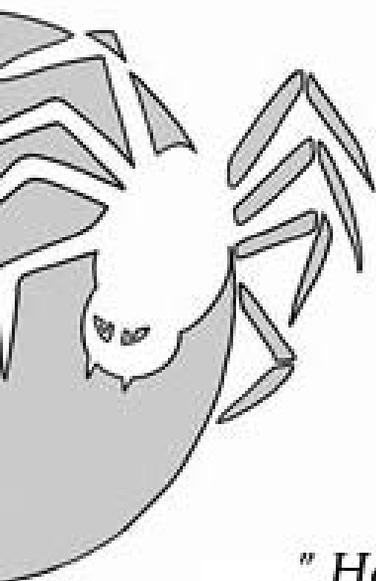
farm) che avevano l'unico scopo di linkarsi fra di loro ed aumentare la popolarità dei siti.

Quando i motori di ricerca introdussero la link analysis, gli spammers iniziarono a scambiare e vendere links tra siti a tema ingannando di fatto gli algoritmi.

Ad oggi i motori di ricerca stanno seguendo diverse strade per rimuovere lo spam delle loro SERP in modo da fornire risultati rilevanti ai propri utenti e non perderne la fiducia, ma nell'accozzaglia di pagine web ce ne sono alcune che hanno l'unico scopo di ingannare i motori di ricerca, e la sfida continua...

Con questo libro spero di avervi fatto capire che la strada giusta per avere un sito di successo, anche se comporta un dispendio maggiore di tempo e fatica, è quella di creare siti web con contenuti e dedicargli costantemente del tempo per la crescita e la promozione.

A volte si può anche perdere la voglia non vedendo risultati a breve tempo, ma la pazienza è la virtù dei forti e i mezzucci per posizionare un sito web rapidamente hanno vita breve.



" Ho visto più volte nella mia vita dei progetti validissimi fallire solamente perchè non riuscivano a diventare popolari, e la chiave di tutto era il loro posizionamento nei motori di ricerca. "