

UNIVERSITA' DEGLI STUDI DI PADOVA



Facoltà di "SCIENZE STATISTICHE"

Corso di laurea in "Statistica e Tecnologie Informatiche"

**"IL POSIZIONAMENTO DEI SITI WEB NEI
MOTORI DI RICERCA"**

LAUREANDO: MATTEO SCHIAVO

RELATORE: prof. MASSIMO MELUCCI

ANNO ACCADEMICO: 2004 - 05

INDICE

1. Cos'è un sito Web aziendale e la sua importanza _____	2
2. Cos'è la visibilità e la sua importanza _____	4
3. Reperimento delle informazioni: i motori di ricerca _____	5
3.1 Spider, robot, crawler _____	8
3.2 Criteri di rilevanza dei search engine: _____	9
3.2.1 Contenuti testuali _____	10
3.2.2 Keywords _____	11
3.2.3 Popolarità _____	11
4. Come rendere visibili i siti Web _____	15
4.1 Keywords _____	15
4.2 Contenuto _____	19
4.3 Titolo _____	20
4.4 Meta tag _____	24
4.4.1 Meta tag description _____	25
4.4.2 Meta tag keywords _____	26
4.4.3 Meta tag robots _____	27
4.5 Robots.txt _____	29
4.6 Link _____	33
4.6.1 Link interni _____	34
4.6.2 Link verso l'esterno _____	36
4.7 Segnalazione, registrazione del sito nei motori di ricerca _____	38
4.7.1 Segnalazione alle directory _____	39
4.7.2 Segnalazione ai search engine _____	40
5. Problemi che si possono incontrare per posizionare un sito Web _____	43
5.1 Dinamicità delle pagine del sito Web _____	43
5.2 Uso di FRAME _____	46
5.3 Effetti speciali _____	50
5.4 Spamming / Antispamming _____	52
5.4.1 Testi, keywords e link invisibili _____	53
5.4.2 Doorway pages, Gateway pages, Doorpages _____	56
5.4.3 Cloaking _____	62
5.4.4 Scambio di link _____	66
6. Appendice: esperienza stage _____	67
7. Bibliografia _____	70

1. Cos'è un sito Web aziendale e la sua importanza

La nascita d'internet risale agli anni Sessanta ma il suo potenziale commerciale si è rivelato solo con lo sviluppo del World Wide Web.

Sostanzialmente il Web è una rete di “pagine” nelle quali aziende, ma non solo, possono condividere informazioni, comunicare con i clienti, ricevere notizie dai frequentatori, effettuare transazioni e inviare messaggi, prodotti e servizi personalizzati ai propri clienti.

Grazie all'utilizzo d'insiemi non lineari di documenti contenenti informazioni di varia natura, detti ipertesti, di collegamenti intelligenti che consentono di spostarsi agevolmente da un sito ad un altro, di motori di ricerca grazie ai quali si possono cercare informazioni digitando una o più parole chiave e di vari modi di pagamento, tra cui le carte di credito e il denaro elettronico, il Web è diventato un'area di grande interesse per il marketing e per il commercio.

Dal punto di vista tecnico, un sito Web non è altro che un insieme di documenti HTML collegati tra loro tramite “links”, cioè collegamenti diretti.

Per documento HTML s'intende un documento scritto in formato testo che comprende al suo interno delle istruzioni che definiscono il modo di visualizzare le porzioni di testo scelte, sarà poi compito del browser che riceve tali documenti ad interpretare le istruzioni nel visualizzare la pagina.

La pagina che vediamo attraverso il browser, è il risultato della lettura del file e della sua interpretazione.

Questo file è composto dal testo della pagina più le istruzioni su come deve essere visualizzata.

Quindi si può parlare di codice sorgente della pagina, poiché i dati contenuti nel file HTML sono letti, interpretati processati e visualizzati nel loro aspetto finale.

Dal punto di vista economico, possiamo sostenere che, in generale, un sito Web aziendale è uno strumento che pubblicizza e vende prodotti e servizi direttamente on

line, è un'iniziativa a supporto dell'attività commerciale di un'azienda che è svolta sulla rete Internet.

Un sito Web può migliorare di molto l'immagine di un'azienda, creare, migliorare o sviluppare un marchio, supportare le funzioni aziendali nell'esercizio dei suoi compiti, permettere una gestione ottimale del servizio d'assistenza pre e post vendita ed infine consentire un'efficace comunicazione esterna.

Principalmente un sito Web può proporsi come:

1. Un mezzo di comunicazione. Molte aziende pagano i siti Web perché facciano comparire il proprio banner di sponsorizzazione. Spesso si tratta di semplice pubblicità tradizionale ma di solito facendo clic con il mouse sull'annuncio si viene collegati al sito ufficiale dello sponsor, dove si potranno trovare informazioni più dettagliate.
2. Un canale di distribuzione. Molte aziende vendono con successo i loro prodotti via Internet in tutto il mondo.
3. Un mezzo di diffusione d'informazioni su prodotti e servizi. Quasi tutti i siti aziendali includono informazioni sulle linee di prodotti e indicano a chi rivolgersi per ulteriori informazioni; spesso vengono incluse anche le prime informazioni nella fase di lancio dei nuovi prodotti e informazioni utili d'alto genere.
4. Un supporto tecnico. Il cliente può spesso ricevere assistenza tecnica dallo staff dell'azienda inviando i propri quesiti per e-mail tramite il sito ufficiale.

Nel caso in cui un sito non sia visitato da un numero soddisfacente di clienti, le potenzialità del sito non saranno pienamente usufruite.

Indi, perché un sito sia veramente efficace deve essere conosciuto, visitato dai clienti/utenti e qui entra in gioco l'importanza della visibilità del sito nella rete.

2. Cos'è la visibilità e la sua importanza

La funzione di ricerca è la più diffusa tra le attività on line e questo mette in evidenza come l'accesso ai siti Web avvenga, nella maggior parte dei casi, attraverso i motori di ricerca.

Per questo i motori di ricerca sono uno degli strumenti migliori per riuscire ad acquisire un'alta visibilità su Internet, poiché consentono di veicolare gli utenti proprio verso quei siti Web che i clienti stessi sono interessati a trovare.

Per visibilità s'intende la posizione che occupa un sito Web nei risultati delle ricerche di un utente in un motore.

Il posto che dovrebbe occupare per avere una buona visibilità dovrebbe essere entro la terza pagina o entro il 30° posto, poiché da uno studio è scaturito che l'utente difficilmente va a vedere i risultati oltre la terza pagina, ma preferisce fare un'altra ricerca o cambiare motore.

L'obiettivo principale è quello di "catturare" un potenziale cliente, nel momento in cui sta ricercando una specifica categoria di prodotti/servizi/contenuti.

Un posizionamento tra i primi risultati dei motori di ricerca per parole e frasi chiave rilevanti e inerenti all'attività dei clienti, aumenta la visibilità di un sito e permette alle aziende di ottenere un vantaggio competitivo nei confronti dei concorrenti.

Figurare in testa ai risultati di ricerca è quindi il modo migliore per accogliere sul proprio sito utenti sicuramente interessati, disposti a dedicare tutta la loro attenzione, proprio perché il sito visitato è l'evoluzione concreta della loro ricerca.

Essere censiti in modo preminente sui motori di ricerca più popolari aggiunge spessore e attendibilità al sito stesso, prestigio e ricordo del brand aziendale, fornisce dunque un elemento di valore di marca, permette una migliore attività di branding rispetto alla pubblicità on line.

Riuscire ad aumentare la visibilità dei siti sul Web, al fine di incrementare il numero delle visite degli utenti, è il desiderio di qualunque proprietario di un sito Web.

Come rendere visibile un sito Web e da cosa dipende la posizione nelle classifiche lo vedremo nei prossimi capitoli.

Va fatta una precisazione però, fin'ora si è parlato di visibilità del sito Web, in realtà le tecniche di posizionamento sono applicate alle singole pagine del sito, quindi per essere precisi si dovrebbe parlare di posizionamento di pagine Web di un sito.

Naturalmente, poiché un sito Web è un insieme di documenti HTML collegati tra loro tramite "links", rendendo visibile una pagina del sito Web, automaticamente si renderà evidente l'intero sito, per questo si parla in generale di posizionamento di un sito Web.

3. Reperimento delle informazioni: i motori di ricerca

I motori di ricerca sono sicuramente uno dei più importanti strumenti per la promozione e la visibilità di un sito Web.

Attraverso questi, gli utenti di tutto il mondo possono trovare le informazioni in un mare sempre più affollato di siti Web.

Generalizzando possiamo considerarli grandi archivi di dati che catalogano i riferimenti ai siti Web di tutto il mondo ed in seguito ad una ricerca di un utente, riportano una lista di siti che contengono i termini cercati.

Dalle statistiche più recenti risulta che, oltre il 90% degli utenti Internet ha usato un motore di ricerca ed il 63% ne usa regolarmente uno o più di uno, questo evidenzia quanto sia importante censire un sito in questi database.

Non a caso, i siti più visitati in assoluto su Internet sono motori di ricerca, con medie che superano i dieci milioni d'accessi il giorno.

Oltre alla quantità potenziale d'utenti raggiungibili, i motori di ricerca permettono di ottenere un'elevata qualità dei contatti generati, specie per quelli provenienti dai link dei risultati di ricerca.

L'utente, infatti, avrà certamente una migliore disposizione nei confronti di un sito trovato grazie ad una propria ricerca, piuttosto che nei confronti di uno cui è arrivato in seguito ad una sollecitazione pubblicitaria, effettuata mediante banner od attraverso un altro tipo di comunicazione.

Difatti, da una ricerca effettuata da NPD Group (gennaio 2001), risulta che la predisposizione del visitatore all'acquisto on line di un prodotto o servizio è cinque volte maggiore quando accede ad un sito da un motore di ricerca, rispetto a quando arriva attraverso un banner.

Il successo di tali motori di ricerca è dovuto essenzialmente a due fattori:

1. Sono lo strumento più comodo per cercare specifici siti Web, i quali peraltro continuano a crescere in quantità, rendendo indispensabile l'uso di strumenti d'analisi e d'orientamento
2. L'utilizzo è gratuito, sia per l'utente che effettua le ricerche, sia per chi inserisce il proprio sito nel database.

Il motivo per cui s'inserisce un sito nei motori di ricerca, è naturalmente quello di attirare visitatori sulle proprie pagine.

Nel caso che l'esito di una ricerca mostra il nostro sito Web ai primi posti, ci sono molte possibilità che sia cliccato dall'utente rispetto ai successivi siti Web.

Come già detto in precedenza, raramente un utente visualizza i siti oltre le prime 20 o 30 posizioni.

Va poi considerato l'aspetto psicologico secondo il quale, un sito in testa all'elenco ha comunque un migliore appeal per l'utente e questo assicura il click.

Una cosa importante è distinguere i **search engine** dalle **directory** (in questo testo ho usato il termine "motori di ricerca" per indicare sia i "search engine", sia le "directory").

Una **directory** contiene una raccolta d'indirizzi della sola homepage dei siti Web, catalogati per tipologia dei contenuti, che sono stati espressamente selezionati da personale umano.

In pratica, i Webmaster e i proprietari di siti Web fanno richiesta alle directory per catalogare il proprio sito e quest'ultime accettano la richiesta solo se i siti Web raggiungono un certo standard qualitativo; Virgilio e Yahoo! sono delle directory.

I **search engine**, invece, scandagliano continuamente l'intero World Wide Web tramite software chiamati spiders, robots o crawlers, e includono nel proprio archivio d'indirizzi tutti i singoli documenti che compongono un sito Web che riescono ad individuare, perciò è possibile che un determinato sito sia presente più volte all'interno di questi in relazione al numero di pagine che si compone, a prescindere dalla qualità dei loro contenuti e indipendentemente dal fatto che i siti siano stati trovati per caso o siano stati proposti al search engine da un utente; Google e Altavista sono dei search engine.

Quindi, il numero di siti Web potenzialmente archiviabili dai search engine è di gran lunga superiore al numero di siti Web potenzialmente archiviabili dalle directory.

Perciò, chi vuole rendere visibile il proprio sito Web è opportuno che privilegi i search engine piuttosto che le directory, poiché sono maggiori le possibilità di figurare nei risultati attraverso le molteplici pagine del sito invece che con la sola homepage.

Però negli archivi dei search engine confluiscono anche siti di bassa qualità o con pochi contenuti.

Per risolvere questo problema e per riuscire a consigliare agli utenti siti Web di buona qualità, i migliori search engine attribuiscono un valore al sito archiviato in base a dei parametri, così da offrire, in seguito ad una ricerca effettuata dall'utente, una classifica ordinata in base al grado di qualità, partendo dal sito che presenta valore più alto a quello più basso.

A questo punto è spontaneo domandarsi quali motori scegliere per registrare il nostro sito Web.

Naturalmente la scelta cadrà sui motori di ricerca più frequentati, anche in relazione al mercato potenziale che si vuole raggiungere, quindi si cercherà di utilizzare i più importanti motori internazionali che soddisfano circa l'80% di tutte le ricerche su Internet.

Le classifiche che indicano la popolarità di un determinato motore di ricerca normalmente offrono il numero totale dei visitatori raggiunti, mentre dal punto di vista della registrazione dei siti il dato che interessa è quello della "search audience", in altre parole relativo alla quantità di ricerche effettivamente richieste.

I motori di ricerca più utilizzati sono: Yahoo!, Google, Inktomi, Lycos, FAST – AllTheWeb, Altavista, HotBot.

Nel caso che il proprio target prevede soprattutto il mercato italiano, vale la pena registrare il proprio sito Web su alcuni motori di ricerca nostrani come Virgilio e Arianna.

3.1 Spider, robot, crawler

Gli "spider, robots, crawlers" sono programmi usati dai search engine per visitare continuamente una gran quantità di siti Web.

Gli spider inseriscono nuove pagine agli archivi e aggiornano le informazioni su quelle già esistenti, leggendo il testo contenuto nelle pagine ed estraendo quelle parole/termini che rappresentano al meglio i contenuti del sito.

Lo spider è in grado di ottenere una quantità enorme d'indirizzi di siti e pagine Web da visitare perché per ogni pagina letta, lo spider cerca al suo interno e memorizza ogni link ad altri siti da visitare.

Attraverso questo processo a catena un search engine riesce ad incrementare il numero di siti conosciuti molto più di quanto possa essere fatto dalle directory.

Infatti, le directory si basano sull'iscrizione e valutazione dei siti operata dagli esseri umani.

Dato che per ogni pagina letta lo spider cerca e memorizza i link ad altri siti, bisogna fare in modo che da qualunque pagina del sito sia possibile risalire direttamente o indirettamente a tutte le altre pagine.

Quindi bisognerebbe inserire in ogni pagina del sito un link all'homepage, in modo che in qualunque pagina del sito lo spider si trovi, quest'ultimo possa risalire alla pagina principale e, da lì, trovare le altre seguendo tutti i link incontrati.

Lo spider legge soltanto il testo ASCII, contenuto nei tag del codice HTML, quindi le immagini (gif, jpg o altro formato) ed animazioni sono ignorate.

Quindi i search engine prediligono i siti Web che contengono buone quantità di testo, naturalmente in tema con gli argomenti trattati dal sito.

Una cosa da non dimenticare è che, ogni search engine usa spider che classificano i siti secondo diversi criteri che cambiano frequentemente per impedire la manipolazione dell'indice.

Nella tabella seguente sono riportati gli spider dei search engine principali:

Motore di ricerca	Spider
Google	googlebot
Fast - AlltheWeb	fast
Inktomi - Yahoo! - HotBot	slurp
Altavista	scooter mercator
Lycos	lycos_spider
Arianna	arianna.iol.it

3.2 Criteri di rilevanza dei search engine

I search engine hanno come obiettivo quello di presentare una lista di siti, ordinata per “indice di qualità”, che trattino temi riguardanti i termini inseriti dall'utente nella ricerca.

Per creare questa lista, ogni search engine usa algoritmi in grado di stabilire quanto ogni sito presente nei propri archivi sia idoneo alla ricerca effettuata dall'utente in base alle parole chiave cercate, ma non solo.

Possiamo immaginare che ad ogni sito i search engine assegnino un punteggio.

I siti che avranno punteggio più alto, avranno un posizionamento migliore nella classifica, in altre parole saranno riportati ai primi posti delle liste.

Alcuni motori tendono a dare maggior punteggio anche ai siti che sono considerati più "importanti" o "autorevoli".

Possiamo sostenere che i search engine classificano i siti Web in conformità a tre parametri:

- Contenuti testuali
- Keywords
- Popolarità

Da ricordare però che ogni search engine prende in considerazione questi tre fattori in modo diverso, attribuendo a ciascuno importanza e peso differenti.

Per esempio, alcuni motori non tengono per niente conto della popolarità di un sito mentre altri, come Google, hanno fatto della popolarità uno degli elementi chiave per garantire agli utenti ricerche di buona qualità.

3.2.1 Contenuti testuali

Con certezza, il fattore che più d'ogni altro incide sul posizionamento di un sito Web nei search engine è rappresentato dai contenuti testuali del sito stesso.

Come già detto in precedenza, gli spider leggono soltanto il testo ASCII, contenuto nei tag del codice HTML.

Quindi più l'argomento ricercato dall'utente è trattato sul sito, e più il search engine spingerà quest'ultimo verso i primi posti della lista.

Dunque è consigliato trattare qualunque argomento in maniera estesa e approfondita.

3.2.2 Keywords

Questo secondo parametro è legato al primo.

L'utente che effettua una ricerca su un search engine, inserisce alcuni termini che ritiene attinenti all'argomento di suo interesse.

Per questo, quando si realizzano testi, bisogna prestare attenzione che siano presenti anche quelle parole chiave che presumibilmente gli utenti useranno come termini di ricerca sui motori (come creare un buon testo con parole chiave lo vedremo nei cap. 4.1 e 4.2).

Nel caso che le keyword rappresentino una buona percentuale del testo complessivo di una pagina, il search engine tenderà a far salire il sito nelle liste.

È importante ricordare che i search engine penalizzano le pagine ed i siti cui esse appartengono, che presentano percentuali troppo alte di keyword, poiché sono considerate pagine “NON veritiere”, in altre parole pagine create apposta per guadagnare posizioni nelle classifiche dei search engine.

Indi è meglio cercare di creare testi di senso compiuto nei quali le keyword appaiono in quantità non eccessiva.

3.2.3 Popolarità

Un fattore che ultimamente sta acquisendo sempre più importanza nei criteri usati dai search engine per stabilire il posizionamento dei siti, è quello della loro popolarità sul Web, chiamata “Link Popularity” (abbreviata al più corto LP).

Difatti, i siti Web ritenuti più "popolari" da parte del search engine, occupano nelle liste posizioni più alte rispetto a siti meno conosciuti o stimati.

Per Link Popularity s'intende il numero di link presenti sul Web che puntano all'home del sito in analisi.

I link contati sono sia quelli provenienti da siti esterni, sia quelli dalle pagine interne.

La Link Popularity, in realtà, è un parametro della singola pagina, ma a tutti gli effetti si considera quella che fa riferimento all'home poiché è questa che normalmente si riscontra in modo rilevante.

Parlare, quindi, di Link Popularity del sito non è rigorosamente corretto, anche se comunemente accettato.

La Link Popularity esprime, secondo la logica dei motori, il fatto che nel mondo i Webmaster dei siti hanno ritenuto il nostro sito rilevante al punto da essere "linkato" nelle loro pagine.

Ciò porta ad affermare che il nostro sito, maggiori referenze ha sul Web, più si suppone sia importante.

La Link Popularity misura per i motori l'importanza di un sito, indipendentemente dal suo tema.

Per calcolare l'indice di popolarità i search engine si basano sul numero di link sparsi per il Web che puntano alla pagina Web.

La misurazione della LP è fattibile, ma con un limite, poiché si possono contare solo i link presenti sui siti conosciuti ai search engine.

Più sono i link che puntano al sito e più il sito è considerato popolare.

Quindi qual è il vantaggio di avere il valore più alto?

- 1) Più siti sono collegati al sito e più traffico si riuscirà a generare.
- 2) I search engine posizioneranno il sito più in alto se si avranno molti link esterni.
- 3) Più link esterni si avrà e più facilmente gli spider dei search engine indicizzeranno il sito Web.

Inoltre i link non possiedono tutti ugual peso, un link presente sul sito di un'importante e conosciuta società ha peso maggiore rispetto ad un link presente su una semplice homepage personale.

Da ricordare inoltre che sono validi solo i link presenti in forma leggibile per gli spider.

Javascript, per esempio, essendo ignorato dagli spider, non produce link conteggiati. Allo stesso modo, possono essere ignorati i link a pagine che poi effettuano un redirect.

Il livello di popolarità di un sito è diventato molto importante perché è utilizzato da alcuni dei principali search engine per influenzare in modo marcato, i criteri di ranking; un esempio è Google ed Altavista.

In linea teorica i search engine che utilizzano quest'elemento cercano di valutare un sito nell'ambito del suo ambiente Web, individuando i temi che legano le singole pagine alle altre presenti on line.

E' fondamentale che i link che si ricevono siano coerenti con gli argomenti del sito, perché ciò che è verificata è l'attinenza tra i siti e non la quantità assoluta di link.

Infatti, link provenienti da pagine con lo stesso tema di quella in esame sono ritenuti di maggior contributo.

Il valore della Link Popularity resta lo stesso, ma la tematizzazione (altro fattore importante per i search engine) è positivamente influenzata.

Nel caso dei link puri in HTML, il fatto che il testo attivo contenga parole a tema per la pagina d'arrivo dello stesso (e meglio ancora se anche per quella di partenza) aumenta ulteriormente la tematizzazione.

Per questo motivo, è preferibile avere pochi link, ma a tema, piuttosto che molti da pagine qualunque.

Molto importante è anche essere presenti in directory, fatto interpretato positivamente nella valutazione della popolarità, così come l'eventuale recensione da parte di siti importanti.

Anche lo scambio di link è un fattore migliorativo del posizionamento.

Due pagine di siti diversi che si “linkano” tra loro, oltre al vantaggio della LP e quello della tematizzazione (se presente) ricevono un’ulteriore spinta in virtù del fatto che si suppone vi sia stato un lavoro coordinato dei due Webmaster e, quindi, congruenza di contenuti.

Occorre prestare attenzione, però.

Porre un link su una propria pagina, vuol dire assumersi una responsabilità davanti al motore.

Link a pagine non più esistenti o a siti “bannati” può implicare penalizzazioni e perdita di posizione tra i risultati.

La LP essendo un parametro svincolato dall’ottimizzazione delle pagine, rappresenta un modo per aumentare il posizionamento del sito senza fare alcuna modifica ai suoi contenuti, anche se i maggiori risultati si ottengono lavorando sulla tematizzazione.

Il suo effetto è positivo per le ricerche effettuate su qualunque parola chiave, anche se i maggiori incrementi si misurano in corrispondenza dei temi della pagina e dei link.

Aumentare i link presenti sul Web, quindi, deve essere parte di un processo che una volta iniziato non si conclude più.

Avere molti link su Web, inoltre, significa consentire molte strade agli spider per arrivare al nostro sito.

Ciò implica maggiore frequenza d’aggiornamento e velocità d’inserimento di nuove pagine.

Indubbiamente, per i nuovi siti, lo sviluppo della popolarità è qualcosa che si costruisce con il tempo e in generale occorrono diverse settimane per vedere premiato il lavoro di sviluppo dei link per ciò che riguarda il posizionamento.

4. Come rendere visibili i siti Web

Le tecniche volte a far essere un sito visibile sui search engine richiedono innanzi tutto un'attenta progettazione del sito, che deve possedere caratteristiche tali da essere appetibile sia agli utenti sia ai search engine.

Si tratta dunque di un processo che andrebbe attentamente pianificato e messo in atto già in fase di progettazione del sito Web.

Le parti principali da tener conto per un buon posizionamento sono:

- *Le keywords*: la scelta delle parole chiave inerenti agli argomenti trattati nella pagina Web
- *Il contenuto*: contenuti testuali del sito
- *Il titolo*: la scelta di un titolo adeguato alla pagina Web
- *I Meta tag*: particolari tag HTML attraverso i quali è possibile specificare una varia serie d'informazioni della pagina Web
- *I link*: creare buone strutture di navigazione sia interne sia esterne

4.1 Keywords

Se vogliamo che gli utenti trovino il nostro sito, le pagine di cui si compone devono contenere le parole che potrebbero essere oggetto di ricerca da parte degli utenti.

I search engine, infatti, classificano i siti soprattutto in base al loro contenuto, e forniscono i risultati delle ricerche dando priorità a quei siti nei quali le parole richieste sono più frequenti e più in evidenza rispetto al resto del documento.

La prima operazione da compiere quindi, è quella di mettersi nei panni dell'utente e pensare a quali termini potrebbe usare nei search engine per cercare siti Web come il

nostro, quindi definire le keyword più appropriate per descrivere il contenuto del proprio sito internet e stilare una lista su cui pianificare il lavoro di posizionamento.

Un elemento da tenere in considerazione, è quello di non prevedere keyword che contengano denominazioni protette come marchi registrati.

Nello stilare la lista ci possono venire in aiuto alcuni search engine “pay per click”, i quali dispongono di una funzione che mostra la quantità di richieste relative ad una determinata Keyword o riguardanti un particolare argomento.

Per esempio per le parole in inglese potrebbe essere utile il servizio offerto da “GoTo”, per quelle in italiano lo strumento offerto da “Godado” chiamato “Keyword Lookup” oppure da “WorldTracker” uno dei tools più versatili e completi.

Queste informazioni non servono solo a soddisfare la curiosità ma consentono di conoscere meglio gli utenti della rete, i loro interessi e il modo con cui cercano di avvicinarsi ai temi che li riguardano.

Se il nostro è un sito Web di un concessionario che si occupa della vendita d’auto nuove ed usate, una prima lista sommaria potrebbe essere:

auto, km0, usato, vettura, automobile, veicolo, semestrale, ecc.

Se il nostro spazio Web è destinato ad utenti di più nazioni, si possono comporre liste separate nelle diverse lingue.

La seconda operazione è di creare delle frasi composte con le parole del precedente elenco, per esempio:

auto usate, veicoli semestrali, veicoli usati, chilometri zero, auto km0, auto seminuove, ecc.

La lista andrà duplicata mettendo tutti i nomi nella forma singolare ed in forma plurale.

Se l'azienda opera in un area geografica specifica o si occupa di settori particolari, questi vanno inseriti, per esempio:

auto usate Vicenza, auto d'importazione, noleggio auto, auto straniere, auto estere, ecc.

Inoltre potrebbe essere il caso di inserire delle parole che rappresentino alcuni interessi del nostro potenziale visitatore, legati in qualche modo al nostro sito, per esempio:

auto d'occasione, autoaffari, finanziamenti auto, veicoli d'occasione, auto scontate, usato garantito, ecc.

Questo perché non sempre l'utente parte da parole che descrivono ciò che sta cercando ma va dritto ai vantaggi che conta di ottenere o punta su caratteristiche specifiche.

Esistono alcune sezioni della pagina che i search engine considerano particolarmente rilevanti.

Cercare di far apparire una o più parole chiave in queste sezioni permette di infondere alle stesse una maggiore rilevanza.

Ecco di seguito un elenco dei principali posti dove è consigliato far apparire le parole chiave:

- *Nel titolo (tag TITLE):* i search engine assegnano al titolo della pagina un'importanza molto alta poiché sono spinti a credere che i suoi contenuti siano molto credibili e che riassumano bene i reali contenuti della pagina. Comunque di questo punto ne parleremo più approfonditamente nel cap. 4.3.
- *Nei meta tag:* i meta tag sono particolari tag HTML attraverso i quali è possibile specificare informazioni di vario genere e tipologia su una pagina

Web. Anche di quest'argomento ne parleremo più approfonditamente nel cap. 4.4.

- *Nelle intestazioni (tag H1 e simili):* un'intestazione può contenere una o più parole chiave inerenti all'argomento trattato dal testo. Una buona intestazione deve essere breve ed i contenuti devono riassumere bene i temi della pagina o del paragrafo cui esse fanno riferimento.
- *Nelle prime righe di testo di una pagina:* le prime righe del testo di una pagina sono quelle che per prime catturano l'attenzione del lettore e sulle quali l'utente si basa per decidere di continuare o no la lettura del testo, quindi sarebbe opportuno che le prime righe di una pagina contengano una buona introduzione agli argomenti da essa trattati ed alcune delle keyword più rappresentative dei temi discussi.
- *Nei testi dei link:* il testo che fa parte di un link ad un'altra pagina gioca un ruolo molto importante. Far apparire una keyword nel testo dei link aiuta a formare un'attinenza tra il termine usato e la pagina verso cui il link punta ed anche con la pagina su cui il link risiede.
- *Tra le parole evidenziate in neretto o in corsivo:* alcuni search engine possono dare particolare importanza a parole evidenziate in diversi modi all'interno del testo (tag B, STRONG, I, EM, ecc.). L'importante è non eccedere poiché può portare un abbassamento della leggibilità del testo e soprattutto perché i search engine sono in grado di rilevare eventuali abusi.
- *Nel nome del file:* Benché non si tratti di una sezione della pagina, diversi search engine controllano il nome del file per individuare al suo interno eventuali keyword. Quindi sembrerebbe opportuno far sì che il nome del file corrisponda o contenga una o più delle parole chiave inerenti ai contenuti della pagina cercando però di mantenere il nome della pagina corto e di facile memorizzazione per le persone. Tuttavia, possiamo affermare che la considerazione che i search engine assegnano al nome del file è destinata a diminuire sempre più nel tempo, poiché i responsabili dei search engine si

stanno rendendo conto che c'è un abuso di questa caratteristica proprio per aumentare artificialmente l'attinenza di una pagina con le keyword.

- *Nei commenti:* Durante la costruzione di una pagina Web in HTML si possono scrivere dei commenti, utili per il Webmaster. Questi commenti sono racchiusi tra: `<!... commento...>`. Il commento non è visualizzato nella pagina Web dall'utente che visita il sito ma è letto dagli spider, per questo si potrebbero usare i commenti per inserirci all'interno parole chiave. Bisogna prestare attenzione però a non esagerare poiché potrebbero essere considerati dagli spider come spamming e quindi far perdere posizioni al sito Web nella classifica del search engine.
- *Nelle immagini:* Nel comando opzionale per la visualizzazione delle immagini che consente di mostrare una descrizione della figura quando il puntatore del mouse si trova sull'immagine, per esempio:
``

Una cosa fondamentale nella selezione delle keyword è di mantenere una totale aderenza ai contenuti delle pagine Web.

Cercare di posizionare sui search engine termini che poi non corrispondono al sito a cui è indirizzato l'utente, non solo è considerato spam ma genera visite da parte d'utenti che probabilmente non sono interessati agli argomenti del sito o peggio ancora, sono infastiditi dall'essere stati indirizzati su un documento diverso da quello prospettato.

4.2 Contenuto

I search engine analizzano in dettaglio i contenuti della pagina Web verificandone i termini più rilevanti.

I contenuti testuali della pagina vanno prodotti nella maniera più "naturale" e semplice possibile.

Nella prima fase di redazione del testo, è consigliato trattare l'argomento a cui la pagina è dedicata senza pensare alla necessità di inserirvi le parole chiave.

Questo approccio permette di ottenere testi di qualità e di leggibilità che altrimenti non si avrebbero se si cercasse d'inserire a tutti i costi, le keyword.

Una volta che si sarà terminato la stesura del testo, i suoi contenuti potranno essere riletti per individuare se le keyword inerenti all'argomento trattato sono già presenti all'interno del testo oppure se è necessario aumentarle, diminuirle o variarle.

Le keyword vanno usate all'interno del testo in maniera moderata e opportuna, sia per la leggibilità e la qualità e sia per i search engine, poiché quest'ultimi riescono a determinare se le keyword sono state usate con naturalezza oppure se sono state usate in maniera artificiosa, in tal caso si potrebbe essere penalizzati sul posizionamento della classifica.

La maggior parte degli esperti concorda nel sostenere che una buona percentuale di parole chiave si aggira tra il 5% e il 10% del testo complessivo della pagina.

Comunque la cosa importante è scrivere con semplicità i testi, producendo così per vie naturali sinonimi e termini correlati alle parole chiave di nostro interesse, che hanno decisamente un effetto positivo per il posizionamento della pagina ed un effetto positivo per il lettore per la leggibilità e la qualità del testo trasmettendo così un'immagine professionale del sito.

4.3 Titolo

Il testo definito per il titolo è quello che compare nella prima riga del programma di navigazione.

I search engine assegnano al titolo della pagina un'importanza molto alta poiché sono spinti a credere che i suoi contenuti siano veritieri e che riassumano bene i reali contenuti della pagina.

La ragione della fiducia riposta nel titolo dai search engine sta nel fatto che il titolo di una pagina Web è presentato in primo piano all'utente che ha effettuato la ricerca (*vedi fig. 4.1*).

Il titolo costituisce così il fattore su cui l'utente baserà il suo primo giudizio sulla pagina Web; se il titolo contiene informazioni che l'utente giudica interessanti e compatibili con ciò che sta cercando, allora la visita della pagina Web è assicurata.

I search engine ritengono quindi piuttosto improbabile che nel titolo siano inserite informazioni fuorvianti circa i reali contenuti della pagina.

Per questo motivo, una parola contenuta nel titolo è considerata realmente attinente al tema della pagina e costituisce uno dei più importanti parametri in base ai quali i search engine determinano l'attinenza di una pagina Web con i termini di ricerca specificati dall'utente.

In sintesi, se una determinata keyword non compare nel titolo, quel documento avrà certamente meno probabilità di figurare tra i primi siti in funzione di una ricerca effettuata sui termini in questione.

Teoricamente non esiste una lunghezza massima del titolo ma in generale i search engine leggono non più di 70 caratteri e quindi è indispensabile definire un titolo che sia sintetico e che riporti nei primi 70 caratteri le keyword su cui si vuole ottimizzare la pagina.

Dovendo impostare un numero limitato di caratteri, il titolo della homepage non permette di gestire tutte le keyword previste per il sito.

È necessario quindi che ogni pagina del sito abbia un titolo ottimizzato in funzione delle parole chiave più idonee e che si scelgano per ogni documento i termini più rilevanti in base ai suoi contenuti.

L'ideale sarebbe evidenziare, per ogni pagina, i termini più ripetuti e delineare il titolo su questi.

È importante che nel titolo appaiano due essenziali informazioni:

1. Il nome del sito o dell'azienda a cui la pagina appartiene
2. I temi trattati nella pagina a cui il titolo fa capo

Anche se il nome del sito o dell'azienda non è necessario ai fini del posizionamento sui search engine, è utile per identificarsi al lettore ed una chiara identificazione può comunicare all'utente trasparenza e professionalità, aumentando le probabilità che il link alla pagina Web sia seguito.

Per quanto riguarda il secondo punto, basta indicare gli argomenti principali, cioè quelli trattati di più nella pagina.

Un esempio di titolo per l'homepage del sito Web di un concessionario d'auto potrebbe essere:

```
<HTML>
```

```
<HEAD>
```

```
<TITLE>
```

```
“nome concessionario”: auto nuove ed usate, km 0 e semestrali
```

```
</TITLE>
```

```
</HEAD>
```

```
</HTML>
```

FIG. 4.1

The screenshot shows a Google search interface. At the top left, there is a button that says "Vai alla pagina principale di Goole". To the right, there are navigation links: "Web", "Immagini", "Gruppi", "Directory", and "News Novità!". Below these is a search bar containing the text "auto usate" and a "Cerca" button. To the right of the search bar are links for "Ricerca avanzata" and "Preferenze". Below the search bar, there are radio buttons for "Cerca nel Web" (selected) and "Cerca solo le pagine in Italiano".

Below the search bar, a blue bar indicates the search results: "Web Risultati 1 - 10 su circa 428,000 pagine in Italiano per auto usate. (0.09 secondi)".

The main content area is divided into two columns. The left column contains search results, with arrows pointing to the "TITOLO", "DESCRIZIONE", and "URL" fields. The right column contains "Collegamenti sponsorizzati".

Search Results (Left Column):

- TITOLO:** [AutoScout24 - Il grande Automercato europeo dell'usato](#)
- DESCRIZIONE:** AutoScout24 - il database, per le **auto** nuove e **usate**, presenta la sua ampia offerta di **auto**. Cerca l'**auto** nel nostro grande database ...
- URL:** www.autoscout24.it/ - 40k - 27 giu 2004 - [Copia cache](#) - [Pagine simili](#)
- TITOLO:** [AutoSuperMarket](#)
- DESCRIZIONE:** Versione on line del mensile di annunci di **auto usate**. Possibilità di ricerca per marca e modello.
- URL:** www.autosupermarket.it/ - 2k - [Copia cache](#) - [Pagine simili](#)
- TITOLO:** [Auto Usate: Prezzi, Sconti E Offerte Di Auto Usate Con Kelkoo](#)
- DESCRIZIONE:** ... **Auto usate**. ...
- URL:** it.kelkoo.com/b/a/c_173601_auto_usate.html - 61k - [Copia cache](#) - [Pagine simili](#)
- TITOLO:** [Quattroruote portale motori](#)

Sponsorized Links (Right Column):

- [Autoworld.it](#)
Auto Usate, Km0 e Nuove garantite. Noleggio e Scandenze **Auto** via SM
www.autoworld.it
- [Nuovi Servizi Auto](#)
Ampia scelta usato garantito, camper **auto** moto **usate** garantite.
www.nsa1992.it/
- [auto usate](#)
il mercato dell'**auto** in Liguria oltre mille annunci on line
www.autoinriviera.it
- [Mercato auto occasioni](#)
Le migliori occasioni del Ticino 1500 veicoli
upsa-ti.ch/occasioni
- [Vendita auto privati](#)
Pubblica il tuo annuncio gratuito on line per vendere l'**auto** usata.
www.autoprontaconsegna.com

At the bottom left of the results area, there is a link "Cerca per provincia".

Gli elementi del sito che sono riportati nella lista dei risultati sono il Titolo, la Descrizione, presa in questo caso dal meta tag, e l'indirizzo Url. I siti che compaiono a destra sotto la colonna "Collegamenti sponsorizzati", sono quei siti che pagano per comparire nei risultati delle ricerche con determinate parole o frasi.

4.4 Meta tag

I meta tag sono particolari tag HTML attraverso i quali è possibile specificare informazioni di vario genere e tipologia su una pagina Web.

I search engine decidono la posizione di una pagina basandosi sul testo contenuto nel foglio stesso, sulla popolarità del sito Web e su diversi altri parametri.

I meta tag hanno perso quasi del tutto la loro efficacia come strumento per il posizionamento delle pagine Web a causa dello spamming, tuttavia rimangono a volte utili per rafforzare un minimo il tema della pagina e spesso servono anche come strumento d'usabilità nei confronti degli utenti.

Essenzialmente, i meta tag dedicati ai search engine sono tre:

1. **DESCRIPTION**

Questo meta tag permette di inserire una descrizione dei contenuti della pagina Web in cui esso appare.

2. **KEYWORDS**

Questo meta tag consente di specificare una lista di parole chiave che riguardano gli argomenti trattati nella pagina Web.

3. **ROBOTS**

A differenza dei due meta tag di cui sopra, il cui scopo è includere informazioni supplementari sui contenuti della pagina, quest'altro meta tag contiene delle istruzioni di "lettura" per gli spider che preleveranno e analizzeranno la pagina.

4.4.1 Meta tag Description

Questo meta tag ha lo scopo di fornire allo spider una breve descrizione dei contenuti della pagina.

Per breve descrizione s'intende che quest'ultima non dovrebbe mai superare le due righe di testo, circa 150 caratteri.

Questa descrizione deve ospitare un piccolo sunto degli argomenti trattati nella pagina, dovrebbe anche enfatizzare le materie che più di altre possono attirare l'attenzione del lettore, poiché il contenuto di questo meta tag è spesso riportato dai search engine nelle ricerche effettuate dagli utenti (*vedi fig. 4.1*).

I contenuti di tale tag devono dunque servire per dare all'utente dei search engine un'idea di ciò che troverà nella pagina, prima ancora che essa sia visitata.

La descrizione deve essere semplice, efficace e di tipo discorsivo, non banalmente sfruttata per includere parole chiavi.

Un esempio del meta tag description dell'homepage di un sito internet di un concessionario d'auto nuove ed usate potrebbe essere:

```
<HTML>
```

```
<HEAD>
```

```
<META NAME="description" content="Concessionaria della città di Vicenza, con  
usato, km 0, semestrali, auto italiane e straniere, con finanziamenti, assicurazioni">
```

```
</HEAD>
```

```
</HTML>
```

La descrizione va dunque sfruttata per includere una breve presentazione discorsiva dei contenuti della pagina Web come nell'esempio; il linguaggio è semplice e comunica al lettore cosa troverà nel sito ancor prima di visitarlo e fornisce informazioni utili come l'area geografica d'interesse ed inoltre contiene alcune parole chiave.

Una cosa importante è quella di dedicare una buona descrizione ai contenuti d'ogni singola pagina che la dovrà ospitare e non all'argomento generale di un sito Web. Quindi per ogni pagina del sito è necessario produrre una descrizione diversa e sempre attinente ai suoi contenuti.

4.4.2 Meta tag Keywords

Il meta tag KEYWORDS ha lo scopo di indicare ai search engine una lista di parole chiave inerenti ai contenuti della pagina Web in cui il tag appare.

La scelta di tali parole rappresenta una fase molto importante della stesura di una pagina Web, non tanto perché alcuni motori si basano ancora sui contenuti di questo meta tag per cercare di individuare il tema principale della pagina, quanto perché le keyword individuate dovrebbero poi essere utilizzate nel testo contenuto nella pagina. Come creare una buona lista di parole chiave l'abbiamo già visto nel cap. 4.1.

La lista di keyword che andrà a costituire il contenuto del meta tag KEYWORDS conterrà sia singole parole sia piccoli gruppi di parole come nel seguente esempio:

```
<HTML>  
<HEAD>  
<META NAME="keywords" content="auto usate,usato auto,usato,Km 0,km  
zero,chilometri zero,chilometri 0,usato garantito,autoaffari,auto semestrali,auto  
d'occasione,auto straniera,auto estere,auto d'importazione,auto scontate,sconti  
auto">  
</HEAD>  
</HTML>
```

Le keyword sono separate da virgole, ma ci sono vari pareri su quest'argomento.

C'è chi consiglia di non usare virgole, chi di separare le parole con una virgola ed uno spazio, ecc.

Il metodo usato nell'esempio è di separare le parole solo con virgole, evitando così gli spazi, in modo da non introdurre caratteri superflui.

Anche la posizione delle parole all'interno della lista possiede la sua importanza: le parole elencate ai primi posti sono considerate dai search engine più rilevanti rispetto a quelle che seguono.

Per cui è bene inserire tra le prime parole quelle che presumibilmente saranno usate più frequentemente nelle ricerche degli utenti.

Per quanto riguarda la quantità di parole da inserire nei contenuti del meta tag Keywords possiamo affermare che un numero ragionevole è compreso tra le 15 e 30 parole.

4.4.3 Meta tag Robots

Differentemente dai precedenti due, il meta tag Robots non ha lo scopo di fornire informazioni inerenti ai contenuti della pagina Web.

Al contrario, si tratta di un modo per comunicare con lo spider del search engine e invitarlo a seguire alcune direttive circa l'uso della pagina Web prelevata.

Attualmente, le direttive impartibili allo spider sono due ed hanno i seguenti scopi:

1. Richiesta di includere (o non includere) i contenuti della pagina nell'archivio del search engine
2. Richiesta di seguire (o non seguire) tutti i link presenti nella pagina, al fine di individuare altre pagine del sito Web

Per chiedere allo spider di indicizzare i contenuti della pagina si usa la direttiva INDEX.

Se, invece, non si desidera che i contenuti della pagina siano archiviati si usa la direttiva NOINDEX.

Per chiedere allo spider di seguire tutti i link contenuti nella pagina si usa la direttiva FOLLOW.

Al contrario, per impedire che lo spider segua i link contenuti nella pagina si usa la direttiva NOFOLLOW.

I due tipi di direttive possono essere combinati a piacere e inclusi nel contenuto del tag ROBOTS separandoli con una virgola.

Il seguente esempio invita lo spider a catalogare la pagina e a seguirne tutti i link trovati al suo interno:

```
<HTML>  
<HEAD>  
<META NAME="ROBOTS" CONTENT="INDEX,FOLLOW">  
</HEAD>  
</HTML>
```

L'esempio successivo chiede allo spider di non includere la pagina negli archivi del search engine.

Tuttavia, s'invita lo spider a seguire i link presenti nella pagina:

```
<HTML>  
<HEAD>  
<META NAME="ROBOTS" CONTENT="NOINDEX,FOLLOW">  
</HEAD>  
</HTML>
```

Le due direttive possono essere combinate tra loro a discrezione e compatibilmente con le nostre esigenze.

È meglio non indicizzare pagine dai contenuti temporanei e soggette ad una prossima cancellazione, oppure si può chiedere allo spider di non seguire i link ad alcune pagine private del sito, non dedicate al pubblico lettore.

Una raccomandazione importante da fare circa i contenuti del meta tag ROBOTS è di non indicare direttive in conflitto tra loro (ad esempio INDEX e NOINDEX assieme) e quella di non ripetere le direttive.

Esistono due ulteriori direttive: ALL, che "accende" tutte e due le direttive e che quindi è sinonimo di "INDEX,FOLLOW", e NONE che "spegne" tutte e due le direttive e che è dunque sinonimo di "NOINDEX,NOFOLLOW".

Alcune importanti precisazioni finali:

- In mancanza del meta tag ROBOTS, gli spider si comportano come se avessero trovato INDEX,FOLLOW
- I search engine che riconoscono e rispettano il meta tag ROBOTS sono la maggior parte, ma non tutti.

4.5 Robots.txt

Il file “robots.txt” (Robots Exclusion Standard) è un sistema che permette ai Webmaster di avere un controllo maggiore su quante e quali pagine HTML far prelevare da quali spider.

Lo “standard per l'esclusione dei robot o spider” si avvale dell'utilizzo di un normale file di testo, da creare attraverso un qualunque text editor.

Tale file va chiamato "robots.txt" e contiene, in un particolare formato, delle istruzioni che possono impedire a tutti o alcuni spider il prelievo di alcune o tutte le pagine del sito.

Una volta creato il file robots.txt, esso va pubblicato on line nella directory principale del sito Web.

Ad esempio, se il sito ha indirizzo <http://www.NomeDelSito.com>, il file dovrà essere accessibile all'indirizzo <http://www.NomeDelSito.com/robots.txt>.

Tutti gli spider di quei search engine che hanno deciso di aderire a questo standard e di seguire le indicazioni del file robots.txt, ogni volta che accederanno al sito Web, per prima cosa andranno a cercare nella directory principale il suddetto file e, se lo troveranno, seguiranno le direttive contenute al suo interno.

Il file robots.txt contiene una lista di pagine e directory che gli spider non possono prelevare.

Per di più non esiste un'istruzione per dire allo spider: "Puoi prelevare questa pagina" ma esistono solo istruzioni per dirgli: "Non puoi prelevare questa pagina".

Non a caso lo standard si chiama Standard per l'esclusione dei robot.

Quindi se non abbiamo la necessità di impedire ai search engine di prelevare alcune o tutte le pagine del nostro sito, allora il file robots.txt non ci serve.

Il file robots.txt contiene dei record, ognuno dei quali comprende due campi: il campo "User-agent" ed uno o più campi "Disallow".

Il campo User-agent serve ad indicare a quale robot/spider le direttive successive sono rivolte.

La sua sintassi è: *User-agent* <duepunti> <spazio> <nome_dello_spider>

Ogni User-agent è identificato solitamente da un nome univoco che lo rende riconoscibile al Web server del sito.

Il nome dell'User-agent è "case insensitive", ossia può essere scritto indifferentemente in maiuscolo o minuscolo senza che ciò interferisca con l'efficacia.

Il campo Disallow serve ad indicare a quali file e/o directory non può accedere lo spider indicato nel campo User-agent.

La sintassi di questo campo è: *Disallow* <duepunti> <spazio> <nome_del_file_o_directory>

Solitamente ad essere escluse sono le directory che contengono contenuto che non si vuole rendere pubblico: le pagine d'amministrazione, le pagine ad uso interno ecc.

Esempio di un record:

User-agent: googlebot

Disallow: /testi.html

Disallow: /mp3/

Il suddetto record dice a Google ("googlebot" è il nome dello spider di Google) che non gli è permesso prelevare il file testi.html né accedere alla directory mp3 e ai suoi contenuti, sottodirectory comprese.

Notate come il nome del file è preceduto da un carattere "/" (che indica la directory principale del sito) e come il nome della directory fa uso di un carattere "/" anche in coda.

Il campo User-agent può contenere un asterisco "*", sinonimo di "qualunque spider". Per cui l'esempio seguente dice a tutti gli spider di non prelevare il file temporaneo.html:

*User-agent: **

Disallow: /temporaneo.html

Il campo Disallow può contenere un carattere "/" ad indicare "qualunque file e directory".

L'esempio che segue impedisce a scooter (lo spider d'Altavista) di prelevare qualunque cosa:

User-agent: scooter

Disallow: /

Infine, il campo Disallow può essere lasciato vuoto, ad indicare che non ci sono file o directory di cui si vuole impedire il prelievo.

L'esempio seguente mostra come dire a tutti i search engine di prelevare tutti i file del sito:

*User-agent: **

Disallow:

In questo caso si ha lo stesso risultato di non aver inserito un file robots.txt.

Come già detto prima, il file robots si compone di uno o più record ognuno dei quali prende in esame spider differenti, quindi gli esempi fatti poc'anzi possono essere contenuti in un unico file robots.txt come qui di seguito:

User-agent: googlebot

Disallow: /testi.html

Disallow: /mp3/

User-agent: scooter

Disallow: /

*User-agent: **

Disallow:

In questo caso allo spider di google diciamo di non prelevare il file testi.html e di non accedere alla directory mp3 ed ai suoi contenuti, allo spider d'Altavista impediamo di prelevare qualunque cosa e tutti gli altri search engine hanno libero accesso a tutto.

Si possono anche inserire all'interno del file robots.txt alcuni commenti, scrivendo righe di testo il cui primo carattere è un "cancellino" #.

Tali righe saranno completamente ignorate dagli spider dei motori.

Per esempio:

*User-agent: **

Non voglio che i search engine eccedano

nella mia directory personale.

Disallow: /personale/

La maggiore utilità del file Robots.txt è nella possibilità di escludere specifici robots "maligni" dalla visita del sito.

Per robots maligni s'intendono tutti quegli User-agent che:

- Creano traffico inutile sul sito
- Scandagliano il sito alla ricerca d'indirizzi e-mail per utilizzarli nello spam
- Eseguono troppe richieste e sovraccaricano il server
- Non sono utili per il nostro sito

Il metodo migliore per identificarli è quello di verificare sui log d'accesso i nomi degli User-agent che giornalmente passano sul sito, verificare a che programma appartengono cercando il nome su un search engine e se la funzione del robot non è gradita, inserire il nome dell'User-agent nel file robots.txt.

Per esempio due robot maligni potrebbero essere:

- *EmailCollector*: software che ricerca sulla rete indirizzi d'e-mail per poi utilizzarli nello spam.
- *Teleport*: offline browser noto per le sue capacità di sovraccaricare i server.

4.6 Link

All'origine d'ogni potenziale buon risultato di posizionamento, esiste una condizione basilare: le pagine del sito devono essere archiviate dai search engine e per ottenere ciò è indispensabile che gli spider dei motori siano in grado di individuare e accedere alle suddette.

Diversi dei siti Web esistenti sono di fatto non navigabili da parte degli spider, che non riescono ad individuare le pagine che compongono i siti a causa di mancanza o

carenza dello strumento principale attraverso cui avviene la normale navigazione: i link HTML.

Negli ultimi anni si sono affermate tecnologie che offrono, tra le altre cose, anche la possibilità di creare menu o strutture di navigazione senza far uso di link HTML, in altre parole del tag "A" e del relativo suo attributo "HREF".

Purtroppo gli spider dei motori non sono stati istruiti per interpretare codice diverso dall'HTML, perciò tutti quei link non-HTML o più generalmente quei sistemi di navigazione che non fanno uso del tag standard "A", possono risultare una vera e propria barriera architettonica sia per gli spider dei search engine sia per gli utenti che fanno uso di browser poco aggiornati o browser particolari.

Quindi è sempre meglio creare link in linguaggio HTML.

Esistono due tipi di link:

1. Link interni
2. Link verso l'esterno

4.6.1 Link interni

Per "link interni" intendiamo quelli che risiedono su una delle pagine del sito e che puntano ad un'altra pagina del medesimo sito.

Tali link devono assolvere almeno tre essenziali esigenze:

1. Fornire al visitatore una struttura principale di navigazione del sito, logica e il più possibile semplice
2. Permettere agli spider di individuare tutte le pagine del sito
3. Collegare tra loro pagine che sviluppano concetti simili o correlati

Gli spider dei search engine possono approdare ad un sito Web seguendo un qualunque link trovato su altri siti.

Non v'è certezza che tali link puntino proprio all'homepage e quindi è opportuno accertarsi che spider e utenti abbiano modo di raggiungere una qualunque pagina del sito indipendentemente dalla pagina da cui sono inizialmente entrati.

Una delle soluzioni più comuni consiste nel dotare ogni pagina del sito di un menu di navigazione, attraverso il quale risalire alle altre pagine o, in caso di un sito composto di molte pagine, a delle sezioni principali da cui poi possono diramarsi sottosezioni e sottopagine.

Va da sé che una pagina essenziale, che deve necessariamente apparire tra quelle considerate più importanti, è l'homepage, perciò è bene che ogni pagina possieda sempre un link alla suddetta.

I siti Web che utilizzano i FRAME e che usano uno di essi per ospitare menu di navigazione, possono portare problemi, poiché, il FRAME è a tutti gli effetti, una pagina a sé stante e se un menu di navigazione è contenuto esclusivamente in un FRAME ciò implica che le altre pagine saranno prive di link di navigazione.

La soluzione a tale problema consiste nell'eliminare del tutto i FRAME, preferendo ad essi soluzioni basate su "include" del Web server o di linguaggi "lato server", oppure dotare ogni pagina del sito di un link all'homepage, dove assieme al FRAMESET va ospitato un tag NOFRAMES contenente una copia del menu di navigazione usato sul sito.

Comunque riguardo ai FRAME ne parleremo più approfonditamente al Cap. 5.2.

E' consigliato aggiungere una mappa del sito, contenente su una sola pagina o comunque su un numero ristretto di pagine una lista di link alle principali risorse del sito, possibilmente accompagnati da una breve descrizione.

Tali mappe risulteranno particolarmente utili sia agli utenti, per individuare velocemente una risorsa specifica, sia agli spider dei search engine.

Inoltre le mappe che includono brevi descrizioni delle pagine del sito a cui i link puntano possono risultare un ottimo modo per tematizzare ulteriormente i contenuti delle pagine linkate.

L'ultimo ma non meno importante obiettivo di un'attenta progettazione dei link interni è quello di usarli per collegare tra loro pagine che sviluppano concetti simili o correlati.

Questi collegamenti tra pagine diverse rompono lo schema "statico" offerto dai menu di navigazione e rappresentano un utilissimo strumento per permettere sia all'utente sia allo spider del search engine di muoversi agilmente tra testi correlati, fornendo una chiara indicazione sul legame esistente tra il testo in cui il link appare e il testo verso cui il medesimo punta.

Questa correlazione è estremamente importante per i search engine e diviene ancora più marcata nel momento in cui il testo usato all'interno del link comprende una o più parole chiave usate nella pagina alla quale il link conduce.

Un consiglio, come il solito è di non esagerare con il numero di link contenuti in una pagina, in particolare evitare di superare i cento link per pagina, questo perché un foglio con troppi link potrebbe somigliare ad una "link farm", in altre parole a quelle pagine malviste dai search engine il cui unico scopo è di tentare di incrementare artificialmente la popolarità dei siti aumentando il numero di link che puntano ad essi.

4.6.2 Link verso l'esterno

I link che puntano a pagine d'altri siti sono considerati "verso l'esterno" e sono chiamati spesso "outbound link".

Un outbound link, seguendo il buonsenso, dovrebbe nascere per migliorare l'esperienza di navigazione al visitatore del proprio sito.

In teoria, se un Webmaster individua in un sito esterno alcuni contenuti che possono ritenersi utili e interessanti all'utente, non dovrebbe avere dubbi nel creare un link verso quel sito, consigliandone la visione.

Facendo ciò, si ottiene un importante vantaggio: si tematizzano maggiormente i contenuti del sito che offre il link.

Consigliando la visione di siti esterni che trattano argomenti simili o correlati a quelli esposti nella pagina che offre il link, si contribuisce a rafforzare maggiormente agli occhi dei search engine il tema della propria pagina e, più generalmente, del proprio sito.

Nella pratica, tuttavia, la creazione di link a siti esterni non sempre è frutto di una scelta dettata dal desiderio di migliorare la navigazione del visitatore o la qualità del proprio sito.

Al contrario, essendo i link ricevuti da altri siti uno dei tanti fattori preso in considerazione dai principali search engine, i link a siti esterni sono divenuti negli ultimi anni una vera e propria merce di scambio.

I Webmaster che creano liste di link ai classici "siti partner" o "siti amici" lo fanno spesso solo per ottenere in cambio un link che da quei siti punti al proprio.

Questa pratica di per sé non è negativa, ma può avere delle controindicazioni se abusata o se effettuata con poco criterio.

La prima di queste controindicazioni consiste nel danno d'immagine che un link può recare se offerto a siti di dubbio gusto o semplicemente di cattiva qualità.

Un link è a tutti gli effetti, agli occhi degli utenti ed a quelli dei search engine, un consiglio che il Webmaster offre ai visitatori, e il solo atto di consigliare un sito esterno di bassa qualità, influisce sull'opinione che gli utenti possono farsi del sito che ne suggerisce la visione.

La seconda controindicazione concerne la tematizzazione del proprio sito agli occhi dei search engine: offrire link a siti che trattano tematiche estremamente lontane da quelle discusse sul proprio sito non aiuta i search engine ad individuare il tema trattato dalla pagina su cui risiede il link verso l'esterno.

E il tema di una pagina o di un intero sito è uno di quei fattori che nel tempo acquisiranno sempre più importanza nel posizionamento.

Per ultimo, sussiste un teorico rischio di penalizzazione per quei siti che offrono link a siti che usano tecniche di posizionamento considerate scorrette dai search engine.

Per risolvere tutti questi potenziali problemi, è sufficiente stabilire una regola e seguirla scrupolosamente: offrire link solo a siti di qualità, possibilmente in tema con la pagina su cui risiede il link, facendo bene attenzione che il sito beneficiario non usi tecniche di spam nei confronti dei search engine.

4.7 Segnalazione, registrazione del sito nei motori di ricerca

Per effettuare la registrazione di un sito nei motori è spesso sufficiente recarsi sui siti dei motori e segnalare direttamente a loro l'indirizzo del sito Web da registrare.

Far registrare un sito nei motori di ricerca rappresenta una delle più importanti ed indispensabili operazioni da effettuare.

Se il sito non è inserito negli archivi dei motori di ricerca, non vi è alcuna possibilità che nei risultati delle ricerche degli utenti possa apparire un riferimento ad esso.

Una prima, essenziale precisazione va fatta circa il significato dei termini "registrare" e "segnalare".

In realtà l'operazione di "registrazione" del sito negli archivi del motore è d'esclusiva competenza del motore stesso.

Ovvero, è il motore che decide se e quando registrare un sito Web nei propri archivi e non il Webmaster.

Infatti oggi l'attività di registrazione è chiamata "search engine optimization", proprio per evidenziare che non è più sufficiente segnalare l'homepage del proprio sito ma va pianificata con cura tutta una serie d'azioni che permettano al sito stesso di

essere effettivamente censito, per poi puntare all'ottimizzazione delle singole pagine Web al fine di ottenere posizioni di testa nei risultati di ricerca.

Una "segnalazione" al motore di ricerca è in realtà l'unica cosa che un Webmaster può fare per indurre il motore ad interessarsi del sito Web: segnalare al motore che il sito esiste e poi attendere che il motore lo visiti.

4.7.1 Segnalazione alle Directory

E' bene precisare che nelle directory la richiesta di segnalazione passa in mano ad un redattore, al contrario dei search engine dove le pagine Web sono analizzate da alcuni software specifici (spider, robot, crawler).

Innanzitutto va analizzata con cura la categoria su cui richiedere la registrazione.

È opportuno selezionare una categoria sufficientemente specifica in relazione al proprio sito ma nel frattempo non troppo analitica poiché potrebbe essere poco visitata.

Su alcune directory è possibile indicare anche più di una categoria, funzione utile, per quei siti che presentano differenti attività dell'azienda.

Occorre quindi sviluppare il titolo e la descrizione da attribuire al sito.

Bisogna ricordare però che è sempre facoltà del redattore della directory non solo accettare o no il sito ma anche modificarne o stravolgere del tutto i dati richiesti.

Vanno quindi banditi i superlativi privilegiando, invece, fatti e caratteristiche evidenti e concrete, facilmente riconoscibili nel sito.

Un altro elemento da considerare è la brevità della descrizione, altrimenti aumenta il rischio che sia sostituita da un'altra.

Su alcune directory è ammesso l'inserimento di una lista di parole chiave in base alle quali si vorrebbe essere rintracciati, in questo caso è fondamentale concentrarsi su termini che effettivamente corrispondano al sito proposto.

Completata la segnalazione è opportuno memorizzare tutti i dati inviati, la data della richiesta ed eventuali estremi forniti dalla directory, questo perché potrebbero rivelarsi necessari in futuro per richiedere eventuali modifiche ai dati immessi.

Per i siti aziendali, i tempi sono mediamente più lunghi, così come maggiore è la difficoltà ad essere censiti.

Infatti alcune directory come Yahoo! e Looksmart prevedono per i siti business una registrazione a pagamento.

4.7.2 Segnalazione ai search engine

Mentre le directory soddisfano gli utenti in cerca di destinazioni on line nell'ambito di macro-argomenti, i search engine sono la risposta a tutte le richieste più mirate e dirette.

Il search engine non fornisce al Webmaster alcuna garanzia sull'effettivo inserimento di un sito in archivio, ad esclusione di quei search engine che offrono tali generi di garanzie dietro pagamento di somme di denaro.

Esistono essenzialmente due modi per informare il search engine dell'esistenza di un nuovo sito Web:

1. Recarsi sul sito Web del search engine e comunicare l'indirizzo (URL) del nuovo sito compilando un apposito modulo
2. Fare in modo che gli spider del search engine, durante la visita ad altri siti Web, trovino un link al nuovo sito (è dimostrato che le pagine così indicizzate hanno un grado di rilevanza più alto delle pagine segnalate direttamente al search engine)

Quando ci si appresta a segnalare un sito al search engine attraverso l'apposito modulo sul suo sito Web, bisogna distinguere tra i motori che richiedono solo la

segnalazione della pagina principale del sito ed i motori che richiedono la segnalazione d'ogni singola pagina appartenente al sito da segnalare.

Nel primo caso, quando il search engine visiterà il sito, provvederà autonomamente ad individuare tutte le pagine di cui esso è composto, seguendo tutti i link che è in grado di trovare dalla prima pagina.

È quindi essenziale che il sito sia stato progettato per essere navigabile attraverso normali link HTML, facenti uso dell'attributo "HREF" del tag "A".

Nel secondo caso, è necessario fornire al search engine l'indirizzo d'ogni singola pagina del sito Web.

È un procedimento che, secondo le dimensioni del sito, potrà risultare tedioso, ma è l'unica maniera per far indicizzare tutte le pagine a quei search engine che non seguono da soli i link.

Inoltre, con certi search engine è consigliato segnalare un numero ridotto di pagine il giorno (sulle 5 pagine), poiché un numero di segnalazioni superiore potrebbe essere interpretato come un tentativo di "spamming".

La richiesta di registrazione può avvenire anche in maniera automatizzata tramite l'uso di software, così da eliminare quel processo tedioso di segnalazione delle pagine Web del sito.

Tuttavia la segnalazione automatizzata non risulta molto produttiva poiché alcuni search engine penalizzano le pagine segnalate attraverso software automatici al fine di scoraggiare l'uso di tali sistemi che spesso sono impiegati in modo improprio o esagerato.

Da ricordare che i programmi di segnalazione automatici riguardano solo i search engine, mentre per le directory è indispensabile effettuare la segnalazione manualmente, sia perché sono richieste maggiori informazioni, sia perché l'operazione è espletata in più sessioni, cosa che un software non può eseguire automaticamente.

Una volta che il search engine è in possesso dell'indirizzo del sito Web o delle singole pagine, il compito del Webmaster si conclude: l'unica cosa che rimane da fare è

attendere che lo spider del motore venga a visitare il sito, al fine di inserire le pagine che lo compongono in un database.

Sempre ricordando il fatto che non è da escludere che il search engine non visiti per niente il sito segnalato, dal momento in cui il motore è entrato in possesso dell'indirizzo del sito, può trascorrere anche molto tempo fino a quando lo spider visiterà il sito Web.

Alcuni search engine impiegano qualche giorno, altri search engine impiegano settimane o persino qualche mese.

Se tutto va per il verso giusto, lo spider del search engine farà visita al sito segnalato. Dopo aver letto il file robots.txt e aver preso nota d'eventuali file e directory a cui gli è negato accedere, lo spider inizierà a leggere le pagine del sito, impiegando tempi che variano da motore a motore.

L'inserimento nel database delle pagine lette dallo spider avviene quasi subito dopo la fase d'individuazione delle stesse, nonostante un primo inserimento in archivio non coincide necessariamente con la possibilità che il sito appaia sin da subito tra i risultati delle ricerche.

Per ottenere quest'ultimo risultato, è necessario anche in questo caso attendere ulteriore tempo, che varia da un paio di giorni a qualche settimana, secondo il search engine.

Il motivo d'attesa di tempi lunghi è dovuto principalmente a due fattori:

- La complessità tecnologica nella gestione d'aggiornamenti frequenti degli archivi, che ormai vanno assumendo dimensioni ciclopiche
- Il tentativo di scoraggiare le tecniche più aggressive di registrazione, che puntano alla registrazione di un gran numero di pagine per poi analizzarle e modificarle in base ai risultati di posizionamento.

Trascorso questo periodo, le pagine saranno definitivamente registrate nel motore, che le mostrerà nelle ricerche degli utenti e che provvederà a visitarle saltuariamente, in cerca di cambiamenti nella struttura del sito o nei suoi contenuti.

5. Problemi che si possono incontrare per posizionare un sito Web

Molti siti Web attualmente pubblicati in rete presentano caratteristiche che ne limitano o addirittura impediscono una corretta catalogazione da parte dei search engine.

E' l'esempio di quei siti Web che fanno uso di pagine Dinamiche, di FRAME e d'Effetti speciali (cioè pagine contenenti elementi prevalentemente grafici).

Di seguito vedremo come ovviare a questi problemi.

5.1 Dinamicità delle pagine del sito Web

Generalizzando, possiamo sostenere che i search engine censiscono in modo adeguato esclusivamente le pagine "statiche" presenti in un sito Web.

Per pagine "statiche" s'intendono quei file con un nome di documento univoco memorizzati nel server sul quale risiede il sito Web.

Il fatto è che molti siti utilizzano piattaforme tecnologiche che costruiscono le pagine in modo "dinamico", in altre parole sviluppando i contenuti solo al momento in cui un visitatore vi accede, dietro passaggio di alcuni parametri attraverso l'URL, dal server Web in collaborazione con un database, quindi senza che esista un vero e proprio documento stabile.

Un esempio è la lista di risultati proposta dai search engine, composta al momento in base alla richiesta del visitatore.

Questa situazione genera problemi quando si chiede ai search engine di registrare le pagine del sito, proprio perché questi richiedono l'esistenza di un indirizzo Web definito (URL) e pur accettando indirizzi complessi, raramente le censiscono.

A volte le pagine che generano contenuti in base a parametri dinamici possono anche indurre il motore ad entrare in un circolo vizioso, in un loop continuo.

Per esempio, una pagina con il seguente indirizzo:

www.NomeDelSito.it/ciao.php?s=7a7894&ciaoid=78

dove il parametro s (session) indica in questo caso il valore della sessione di collegamento utilizzato dal sito per tenere traccia del passaggio di un singolo utente, cambiando il numero di sessione, e dunque l'URL, la pagina non cambia, quindi un motore non potrebbe tenere traccia del cambiamento di questa variabile e dunque si troverebbe ad indicizzare centinaia di pagine identiche con URL differenti, oppure, peggio, continuerebbe ad indicizzare il sito all'infinito.

Per risolvere questo problema i search engine o ignorano le pagine dinamiche oppure impongono dei limiti ben precisi nella loro indicizzazione.

Ma c'è di più, all'aumentare dei parametri utilizzati negli URL aumenta la difficoltà di indicizzazione dei motori.

In parte questo problema dell'indicizzazione delle pagine dinamiche è risolvibile.

Alcune aziende per ovviare a questo problema, utilizzano software in grado di cristallizzare i contenuti delle pagine dinamiche in singoli file statici curando la richiesta di registrazione, ma l'impegno necessario, in termini di tecnologia e d'infrastrutture, non è quasi mai giustificato dai risultati che si possono ottenere.

La migliore soluzione è di procedere alla riscrittura degli URL.

La maggior parte delle difficoltà sono riscontrate dalle pagine che nell'indirizzo Web hanno il carattere “?” o “&”.

Con opportune modifiche alla configurazione software dei server è possibile sostituire il “?” o “&” con un altro carattere di divisione.

Il concetto è semplice: utilizzare nel sito URL il più possibile vicini agli standard.

Per esempio:

Indirizzo originale: www.NomeDelSito.it/ciao.asp?id=789&idcat=45

Indirizzo “standardizzato”: www.NomeDelSito.it/ciao/id_789/idicat_45.html

Esistono due soluzioni gratuite e automatizzate sia per siti costruiti in ASP sia per siti costruiti in PHP.

Il loro utilizzo è vincolato ad alcuni passaggi:

- Bisogna avere accesso al pannello di controllo di IIS per ASP oppure al file httpd.conf (o anche .htaccess) per Apache
- Bisogna modificare tutti gli indirizzi all'interno del sito Web per farli puntare al nuovo tipo d'indirizzo

Per chi utilizza pagine dinamiche e non può intervenire direttamente nella configurazione del Web server non resta che affidarsi ad un trucchetto.

Alcuni search engine indicizzano le pagine dinamiche solamente se linkate da pagine statiche.

Non ci resta dunque che preparare all'interno del nostro sito alcune pagine statiche con link verso le pagine dinamiche che c'interessa indicizzare.

Queste pagine dovranno a loro volta essere linkate da più punti del sito Web (magari da tutte le pagine), in modo da essere sempre reperibili dal motore di ricerca.

5.2 Uso di FRAME

Molti siti Web usano FRAME, il loro successo è dato da diversi motivi:

- Dal punto di vista dell'utente: evitare di ricaricare le parti comuni e mantenere in vista alcuni punti del layout
- Dal punto di vista del Webmaster: includere il layout comune in pochi files

Tuttavia gli svantaggi che comporta un uso scorretto di un layout a FRAME sono superiori ai vantaggi che possono derivare dal loro utilizzo.

L'uso di FRAME può portare a problemi di posizionamento, poiché sono ignorati dagli spider di diversi search engine perché quest'ultimi si comportano in maniera molto simile ai vecchi browser testuali in uso diversi anni fa, ignorando così molte caratteristiche tecniche avanzate.

Indi diversi spider saltano i tag FRAMESET e tutti i loro contenuti, ignorando così gli URL ai file HTML contenuti nei tag FRAME.

Questo può portare lo spider a non trovare le altre pagine appartenenti al sito Web visitato.

La struttura dei FRAME, infatti, prevede l'impiego di singoli file HTML per ognuno dei FRAMESET e un file master che ne descrive le dimensioni e le caratteristiche.

Attualmente i search engine leggono solo la pagina che definisce i FRAMESET, ignorando il contenuto vero e proprio di cui si compone il sito, oppure andando a prelevare i singoli elementi della pagina che rimangono però scollegati dal resto del documento.

È frequente allora che una struttura a FRAME sia inserita all'interno di un search engine in modo errato: a volte viene catturato solo un menu, altre volte compare soltanto la parte interna con il contenuto del FRAME e dunque viene perso ogni menu di navigazione.

Per evitare problematiche di questo genere, è meglio evitare di utilizzare una struttura a FRAME o nel caso in cui si desideri utilizzare è bene:

1. prevedere sin da subito dei metodi che ricostruiscano il FRAMESET, nel caso in cui sia catturata soltanto una pagina interna
oppure
2. fare in modo di far trovare allo spider gli URL alle altre pagine del sito.

Per quanto riguarda il primo punto, come già detto in precedenza, i search engine non sempre riescono ad indicizzare perfettamente un sito diviso in FRAME.

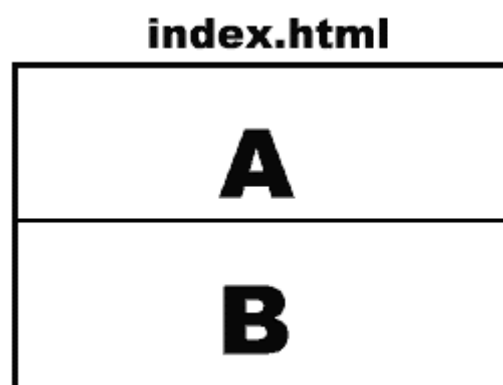
Molto spesso invece di linkare il file che crea il FRAMESET (solitamente index.html), richiamano i singoli FRAME che tale struttura vanno a costruire.

Il risultato è spesso disastroso, con pagine senza menu o link di navigazione.

In ultima analisi una ricerca condotta in questi termini rischia seriamente di compromettere i risultati di un Website e di far perdere potenziali visitatori.

Facciamo riferimento al seguente grafico:

Index.html è il nome del file principale che costruisce il FRAMESET e richiama i due documenti: A nel FRAME superiore e B nel FRAME inferiore.



Se il search engine risponde alla ricerca dell'utente con la pagina index.html non c'è alcun problema.

Il problema si crea quando il search engine trova il documento A o B, in questo caso ci viene in aiuto Javascript con uno script semplice.

In ogni pagina che fa parte del FRAME basta aggiungerci il seguente script:

```
<HTML>
<HEAD>
<script>
    if(self.name==top.name){
        location.href="index.html"
    }
</script>
</HEAD>
<BODY BGCOLOR="black">
</BODY>
</HTML>
```

In questo modo il FRAME verrà visualizzato dall'utente in maniera corretta ed in tutta la sua completezza.

Naturalmente per gli utenti che non fanno uso di browser visuali il problema di visualizzare correttamente la pagina rimane.

Per risolvere questo problema e per assolvere il punto 2 ci viene in aiuto il tag NOFRAMES.

Il tag NOFRAMES è stato introdotto per tutti quei browser che non supportano i FRAME e quindi anche gli spider.

Difatti, gli utenti che usano browser testuali rischiano di non poter accedere alle pagine del sito qualora il progettista del sito Web non abbia usato correttamente il tag NOFRAMES.

Nel tag NOFRAMES è bene inserirci una breve descrizione del sito e dei suoi contenuti, link ad un menu delle pagine del sito od a qualsiasi pagina interna al sito da cui si possa risalire a tutte le altre, collegamenti a versioni testuali delle principali pagine del sito e infine recapiti postali, telefonici e di posta elettronica per coloro che preferissero entrare in contatto con una persona invece che continuare a navigare sul sito.

Ciò detto fin'ora, oltre a rendere il proprio sito Web usufruibile da quante più persone possibili, comporta la produzione di contenuti, cosa molto appetibile da parte degli spider dei search engine.

Il linguaggio HTML richiede che, i tag NOFRAMES ed i loro contenuti vadano inseriti all'interno dei tag FRAMESET ed immediatamente prima o dopo la lista dei tag FRAME, come nell'esempio:

```
<HTML>
<HEAD>
  <TITLE>Titolo</TITLE>
</HEAD>

<FRAMESET>

  <NOFRAMES>
    Questo testo viene letto solo dai browser testuali e dagli spider
  </NOFRAMES>

  <FRAME>
  <FRAME>
</FRAMESET>

</HTML>
```

Ultimamente gli spider dei search engine si sono evoluti ed alcuni di essi supportano i FRAME, così possono trovare senza problemi i link interni ai tag FRAME.

Tuttavia, questo non costituisce realmente un buon cambiamento.

Difatti, esistono almeno quattro buone ragioni per continuare a preferire un sito privo di FRAME:

- Non tutti i search engine possiedono spider col supporto dei FRAME
- Un sito con i FRAME possiede una quantità di testo inferiore ad una versione senza FRAME dello stesso sito
- Il sito con i FRAME diviene, a causa del tipo d'indicizzazione dei motori, meno usabile per l'utente
- Il sito con FRAME non può essere visualizzato da quegli utenti che usano browser che non supportano i FRAME

In sintesi possiamo affermare che dal punto di vista dell'indicizzazione del sito sui search engine, è preferibile fare a meno dei FRAME in tutti i casi possibili.

Nel caso che non sia possibile fare a meno dei FRAME, è opportuno ricorrere all'utilizzo del tag NOFRAMES inserendovi contenuti testuali e link, cosa molto utile per gli spider e per gli utenti con browser testuali oppure usare Javascript ricordando però che gli utenti con browser testuali non potranno leggerlo.

5.3 Effetti speciali

Per effetti speciali s'intendono quelle pagine contenenti parti non-testuali, prevalentemente grafiche, i cui scopi principali consistono nell'arricchire l'esperienza visiva dell'utente e nel fornire nuove tecniche per navigare un sito.

La tecnologia maggiormente utilizzata a questo scopo è la tecnica Flash di Macromedia, che permette di realizzare animazioni ed effetti visivi altrimenti impossibili da ottenere col normale codice HTML: filmati, animazioni, interfacce utente, giochi, sperimentazioni artistiche, ecc.

Le ragioni per limitare l'uso di questi effetti speciali sono principalmente 2:

- La prima è perché, affollando la pagina Web d'elementi grafici sia statici sia in movimento, va a diminuire il grado di leggibilità dei testi contenuti nella pagina.
- La seconda e importante ragione ai fini del posizionamento riguarda appunto al fatto che, gli effetti speciali non sono testi e quindi, non possono venire letti da tutti coloro che preferiscono o che sono costretti ad utilizzare browser testuali e questo significa che neanche gli spider possono leggerli e quindi indicizzarli nei search engine.

Ci sono casi in cui il sito Web da realizzare deve possedere caratteristiche difficilmente implementabili con tecniche differenti da Flash.

In tali casi, la cosa da fare è progettare il sito Web in modo che possa essere navigabile anche da coloro che accedono alle pagine con browser che non supportano il prodotto di Macromedia.

Nel caso che il sito faccia uso di link costituiti esclusivamente da elementi Flash, è consigliato produrre un sistema parallelo di link basato sui normali tag "A" del linguaggio HTML.

Pure per i contenuti testuali del sito che sono interamente inglobati in oggetti Flash, è opportuno inserirli all'interno delle normali pagine HTML.

Nei casi in cui un sito sia interamente sviluppato in Flash, è consigliato produrre una seconda versione del sito, sviluppata interamente in pagine HTML.

Tutto questo perché, così, le pagine possono essere lette da utenti con browser testuali e così essere lette da quante più persone possibili e soprattutto perché possono essere lette dagli spider dei search engine che così possono indicizzare il sito negli archivi dei search engine.

5.4 Spamming / Antispamming

Lo spam consiste nell'utilizzare delle tecniche poco ortodosse per cercare di ottenere posizioni più alte nei search engine.

Ad esempio la ripetizione estrema di parole chiave o l'utilizzo di parole che non c'entrano con il contenuto effettivo del sito, sono considerate spam.

Spam significa anche l'uso di parole, di codice, di script e d'ogni altra tecnica che al fronte di un miglioramento nelle posizioni nelle ricerche dei motori, non porta nessun beneficio agli utenti.

Il termine spam è usato di sovente anche per definire tutta la posta indesiderata che si riceve nella propria casella di posta elettronica.

I search engine hanno delle tecniche raffinate per "smascherare" i siti che ne fanno uso e quando ne trovano uno arrivano spesso anche alla cancellazione del medesimo dal loro indice.

L'attenzione sullo spam da posizionamento è talmente forte che, a volte, sembra talmente eccessivo l'accanimento dei search engine contro i possibili abusi, da far perdere di vista l'obiettivo principale nei confronti dell'utente finale, quello di fornire risultati di qualità.

L'uso di queste tecniche è portato dal fatto che non sempre è possibile produrre un sito con i tre parametri in base ai quali i search engine attribuiscono un punteggio ai siti Web (qualità dei contenuti, presenza di keyword nel testo, popolarità da link).

Ecco dunque che le tecniche che tratteremo agiscono proprio per aumentare artificialmente quelle caratteristiche di cui, in condizioni normali, il sito Web non sarebbe dotato o sarebbe provvisto solo in quantità limitata.

Le tecniche principalmente usate e che andremo a trattare sono:

- Testi, keywords e link invisibili
- Doorway pages, Gateway pages, Doorpages
- Cloaking

➤ Scambio di link

E' bene ricordare di non fare uso di spam in nessun caso.

Difatti, anche se si riesce ad evitare i controlli dei search engine e conseguentemente sfuggire alla penalizzazione nelle classifiche di ricerca, potremo essere "denunciati" da nostri concorrenti, specialmente in quei campi con alta competitività.

5.4.1 Testi, keywords e link invisibili

L'obiettivo di molti Webmaster è di inserire all'interno dei testi del sito molte parole chiave differenti, poiché i risultati di una ricerca su un motore sono relativi ai siti che contengono le parole chiave cercate ed utenti diversi usano keywords diverse nel tentativo di individuare i siti che trattano uno specifico argomento.

Per i Webmaster è difficoltoso introdurre molte parole chiave all'interno dei testi, poiché quest'ultimi poi appaiono eccessivamente carichi di parole chiave e meno leggibili agli occhi dei lettori.

Dall'esigenza di introdurre più parole chiavi all'interno delle pagine senza influire sulla quantità e qualità dei testi dedicati agli utenti, sono nati dunque diversi trucchi, che hanno l'obiettivo comune di creare testi visibili solo ai search engine e non ai visitatori del sito.

Uno di questi trucchi consiste nello scrivere del testo con lo stesso colore del fondo della pagina.

In questo modo l'utente non può individuare tale testo essendo dello stesso colore dello sfondo della pagina, ma il motore di ricerca, leggendo direttamente il codice HTML della pagina, è in grado di leggere il testo senza badare al colore in cui è scritto.

Questo testo "invisibile", è solitamente usato per trattare più approfonditamente un argomento in quelle pagine dove, per necessità di design, è necessario mantenere bassa la quantità di testo visibile, ma non sempre è così.

Di solito questi testi invisibili sono usati per scrivere testi contenenti molte parole chiave o addirittura soltanto una lista di parole chiave, senza che le stesse siano introdotte in frasi di senso compiuto.

Una seconda tecnica per includere nel codice HTML testi non visibili agli utenti, consiste nell'inserire gli stessi tra quei tag HTML espressamente ideati per non mostrare il testo a video o per mostrarlo solo in alcune occasioni, che nell'uso comune dei browser si presentano raramente.

Per esempio i tag HTML usati per inserire commenti all'interno del codice HTML.

Il loro principale uso è di introdurre brevi annotazioni al codice HTML utili ai Webmaster.

Naturalmente non sono visibili agli utenti che visitano la pagina Web e per questo spesso sono usati per far trovare ai search engine un po' di testo in più rispetto a quello mostrato agli utenti.

La sintassi dei commenti è la seguente:

```
<!-- Questo è un commento -->
```

Un altro tag usato per questi scopi è il tag NOFRAMES, usato di solito nei siti Web che fanno uso di FRAME.

Il tag NOFRAMES è stato ideato per mostrare testi a quei browser che non possiedono il supporto dei FRAME o che hanno temporaneamente disabilitato tale caratteristica, come già spiegato più approfonditamente nel cap. 5.2.

L'uso improprio di questo tag consiste nello scriverci al suo interno parole chiavi o frasi preparate apposta per i search engine:

`<NOFRAMES>`

Testo visibile solo ai browser che non supportano i FRAME ed agli spider.

`</NOFRAMES>`

Un altro tag è il tag NOSCRIPT che è usato per fornire contenuti alternativi per supportare gli utenti dei browser testuali o dei browser meno aggiornati che non supportano il linguaggio indicato nel corrispondente tag SCRIPT.

Spesso questo tag è utilizzato in modo improprio per fornire ai search engine testi non visibili alla maggior parte degli utenti:

`<NOSCRIPT>`

Testo visibile solo ai browser che non supportano il linguaggio indicato in un precedente tag SCRIPT e naturalmente visibile agli spider.

`</NOSCRIPT>`

Le tecniche appena citate sono alcune di quelle esistenti, infatti, vi sono altri tag dove inserire parole chiave, come il tag INPUT “HIDDEN” o nell’attributo ALT del tag IMG.

Queste tecniche possono essere utilizzate, oltre che per aumentare artificiosamente la quantità di testi e keyword, anche per incrementare il numero di link.

Spesso è utile inserire in una pagina sia link ad altre pagine dello stesso sito, sia link a pagine di siti affiliati, sia alcuni link a siti esterni che trattano i medesimi argomenti.

Ci sono almeno tre buone ragioni per non usare queste tecniche:

- La prima ragione per non usare questi trucchi è che gli stessi possono causare problemi agli utenti di browser testuali. Infatti, i tag NOFRAMES e NOSCRIPT, così come l’attributo ALT del tag IMG sono stati creati per facilitare l’accesso al Web agli utenti che non possono o vogliono usare browser grafici; destinarli ad altri scopi è molto sconsigliato, poiché, l’utente

privo di browser grafico andrà a visualizzare i contenuti dei tag appena citati, visualizzando così contenuti privi di significato.

- La seconda ragione è che, aumentare il testo in una pagina può essere controproducente poiché "diluisce" la rilevanza di tutte le parole in esso contenute. E poi i search engine tengono conto della percentuale di keyword all'interno del testo e non della loro quantità.
- L'ultima ragione è che oggi la maggior parte dei search engine è in grado di individuare i testi invisibili e penalizzare i siti che ne fanno uso, diminuendo il loro "punteggio" e facendoli conseguentemente apparire, in posizioni più arretrate nei risultati delle ricerche. I motori più "rigorosi" possono persino arrivare a bandire in modo definitivo un sito dal proprio archivio.

5.4.2 Doorway pages, gateway pages, doorpages

Le "Doorway pages" o "Gateway pages" o "Doorpages" sono delle pagine Web esterne al sito Web vero e proprio e dall'impostazione semplice ed essenziale, appositamente costruite per cercare di figurare nei risultati di ricerca e che, successivamente, indirizzano l'utente al sito vero e proprio.

Sono semplici pagine Web che, generalmente, sono ottimizzate in funzione di una serie di termini con l'obiettivo di apparire in testa ai risultati delle ricerche effettuate con tali keyword.

Il loro uso è utile e talvolta indispensabile per inserire i siti nei search engine per le diverse motivazioni/situazioni:

- Per quei siti che hanno contenuti difficilmente registrabili, per esempio quei siti realizzati con tecnologia Flash o basati su contenuti dinamici o che contengono FRAME

- Le doorway permettono di figurare nei risultati di ricerca in base a più parole chiave. Le singole pagine del sito, che principalmente devono invece rispondere ad obiettivi estetici e di comunicazione, raramente si riescono ad ottimizzare in modo efficiente
- In relazione ai diversi criteri con i quali i search engine classificano le pagine Web, le doorway consentono di indirizzare la registrazione in funzione delle differenti regole, poiché ciascuna di esse realizza una versione diversa per ogni search engine

Molti Webmaster però impiegano le doorway per un'attività non proprio adeguata, realizzando centinaia di pagine Web che non hanno alcun valore in termini di contenuti, tentando di posizionarle in testa ai risultati su argomenti non attinenti al sito vero e proprio.

Queste doorway, quindi, assumono le seguenti caratteristiche:

- I contenuti testuali di tali pagine sono pensati in maniera specifica per i search engine e non per essere mostrati agli esseri umani. Le pagine contengono spesso solo una gran quantità di parole/frasi chiave e nessuna o poche frasi di senso compiuto.
- I testi d'ogni pagina doorway sono di solito focalizzati su una "frase-chiave", in altre parole un insieme di parole chiave ben preciso.
- Essendo un insieme di parole chiave del tutto inutili per un essere umano, spesso le pagine doorway sono accompagnate da tecniche di "ridirezione automatica", il cui compito è quello di ridirigere il visitatore che accidentalmente dovesse raggiungerle verso una pagina "normale", pensata per gli esseri umani e che generalmente coincide con la pagina principale del sito Web.

Come vedremo più avanti nell'esempio, l'uso delle doorway dal punto di vista tecnico non è difficile.

Il Webmaster deve creare un discreto numero di queste pagine, ognuna focalizzata su una differente parola o frase chiave riguardante i temi del sito Web, e volendo, focalizzarla ad uno specifico search engine.

Per focalizzare la pagina su un dato motore è sufficiente fare un attento studio delle pagine migliori classificate, e così definire con una buona approssimazione quali parametri sono stati utilizzati dal motore per definire la classifica.

Variabili come la lunghezza complessiva della pagina, i contenuti del titolo e delle intestazioni, i link presenti sulla pagina, la frequenza e la distribuzione delle parole chiave nel testo e così via possono essere adattate per uno specifico motore e per una specifica keyword.

Inoltre, il Webmaster deve fare in modo che le pagine contengano del codice di ridirezione che entri in azione solo quando le stesse sono visitate da una persona e non da un search engine.

Questo può essere ottenuto con del codice scritto in Javascript o usando un meta tag HTML, meta tag REFRESH.

I browser visuali leggono senza problemi il codice Javascript o il meta tag di ridirezione mentre gli spider dei motori solitamente non supportano né il linguaggio Javascript né il meta tag citato.

In questo modo gli spider leggono i contenuti della pagina doorway senza subire nessun ridirizzamento, mentre i browser visuali degli utenti eseguono la ridirezione prima ancora che gli utenti si rendano conto che la pagina contenga testi chiaramente usati per falsare i risultati delle ricerche.

Il motivo per cui non usare le pagine doorway in questo modo è che quest'ultime non sono infallibili e molti utenti possono casualmente entrare in una di esse senza subire la ridirezione automatica.

Così a queste persone appariranno contenuti privi di significato ed inoltre sarà evidente il tentativo di falsare i risultati delle ricerche sui motori.

L'effetto negativo può essere anche maggiore nel momento in cui il sito che fa uso delle doorway è un sito aziendale; in questi casi si rischia seriamente di infondere all'azienda un'immagine di scorrettezza.

Come accadeva per i testi nascosti, le pagine doorway possono costituire un problema ai browser testuali o a quelli che non supportano il linguaggio Javascript, poiché si comporteranno come gli spider e quindi non subiranno nessun ridirizzamento.

Inoltre le pagine doorway possono far penalizzare un sito poiché i search engine sono in grado, nella maggior parte dei casi, di individuare la differenza tra una pagina normale ed una doorway.

Anche se gli spider dei motori non subiscono la ridirezione automatica, sono in grado di accorgersi del sistema di ridirezione.

Quindi usare un sistema di ridirezione spinge lo spider ad indagare più a fondo sul sito ed a cercare di capire se il sistema di ridirezione è usato per scopi leciti oppure no.

Le penalizzazioni in cui si può incorrere sono le stesse relative ai testi invisibili.

Praticamente, ai fini del posizionamento nei risultati di ricerca, la realizzazione delle doorpage ottimizzate in funzione di termini specifici non deve badare più di tanto alla forma o alla qualità dei contenuti, ma deve concentrarsi su come superare i filtri dei search engine; questi agiscono, in particolare, proprio su quei documenti costruiti su argomenti precisi.

Qui di seguito sarà presentato un esempio di doorpage, ottimizzata per le chiavi di ricerca “auto usate” e “auto semestrali”:

Iniziamo il tag con un commento, invisibile al navigatore ma non allo spider, in cui definiamo le due principali chiavi che vogliamo ottimizzare

```
<!--auto usate: auto semestrali -->
```

```
<HTML>
```

Lo Javascript che segue dovrà ridirizzare il navigatore dopo cinque secondi, alla pagina principale del sito. In questo caso la doorpage sarà visibile all'utente, basterà abbassare il tempo per fare in modo che non lo sia più.

```
<script language="JavaScript">
<!--
setTimeout("location.href='http://www.NomeDelSito.it'",5000);
// End -->
</script>
<HEAD>
```

Normalmente le chiavi vanno ripetute anche nei tag meta e soprattutto nel titolo del documento.

```
<TITLE>auto usate: auto semestrali</TITLE>
<META name="description" content=" Concessionaria con auto usate ed auto
semestrali di tutte le marche on line">
<META name="keywords" content="auto usate,autosemestrali ">
</HEAD>
```

Inizia il corpo del messaggio. Come primo paragrafo, inserito tra i tag H1 che donano importanza alla frase, troviamo nuovamente le chiavi di ricerca che abbiamo specificato poco sopra.

```
<BODY bgcolor="#000000" link="#005AFF" vlink="#005AFF" alink="#005AFF"
text="#005AFF">
<CENTER>
<H1> <FONT color="#005AFF" size="2">auto usate ed auto semestrali</FONT>
<br>
<br>
```

Una ripetizione delle keyword va inserita anche nei tag ALT delle immagini e in due link che puntano al sito madre.

```
<A href="http://www.NomeDelSito.it"></A>
<A href="http://www.NomeDelSito.it"><IMG src="logo.gif" width="100"
height="100" border="0" alt="auto semestrali"></A></H1>
<H1><font face="Verdana, Arial, Helvetica, sans-serif" size="2">
<A href="http://www.NomedelSito.it">Auto usate ed auto semestrali di tutte le
marche</A>
<br>
<A href="http://www.NomeDelSito.it">Auto usate ed auto semestrali on line</A>
<br>
</font>
<br>
</H1>
<H1><B><FONT face="Verdana, Arial, Helvetica, sans-serif" size="2"
color="#FFFFFF">
<A href="mailto:auto_usate@NomeDelSito.it">auto_usate@NomeDelSito.it</A>
</FONT></B></H1>
</CENTER>
</BODY>
</HTML>
<!--auto usate: auto semestrali -->
```

Il codice scritto sopra sarà letto dai search engine visualizzando la sola struttura testuale ad esclusione dei meta dati, quindi risulterà:

auto usate auto semestrali auto usate auto semestrali concessionaria con auto usate ed auto semestrali di tutte le marche on line auto usate auto semestrali auto usate ed auto semestrali auto usate auto semestrali auto usate ed auto semestrali di tutte le

marche auto usate ed auto semestrali on line auto_usate@NomeDelSito.it auto usate auto semestrali

Si noti che la parola chiave auto è ripetuta ben 18 volte su 55 quindi in una percentuale del circa 33% e come avevamo già detto, una buona percentuale per non dare nell'occhio ai search engine si aggira sul 5-10%.

Con una percentuale del 33% è facile che i search engine capiscano che si tratta di una pagina creata appositamente per il posizionamento e quindi è facile subire una penalizzazione.

Comunque basterà scrivere dell'altro testo per abbassare la percentuale e quindi diluire le parole chiave.

Si può fare a meno dell'uso delle doorpage facendo in modo di progettare il sito Web includendo delle "pagine d'approfondimento", cioè pagine incentrate su un argomento ben preciso e collegate alle altre pagine del sito attraverso i normali link HTML.

Naturalmente queste pagine saranno visibili sia agli utenti sia ai search engine, senza sistemi di ridirizzamento.

Queste pagine di solito appaiono in posizioni prominenti nelle ricerche riguardanti quella frase chiave, e divengono le prime pagine attraverso cui l'utente accede al sito Web, per questo vengono anche chiamate "Welcome pages".

5.4.3 Cloaking

Il "cloaking" o "IP delivery" è una tecnica che consente di creare una pagina Web "dinamica", che mostra contenuti diversi secondo chi la visita.

Per esempio: un sistema di cloaking può automaticamente individuare la nazionalità dell'utente e mostrare una pagina nella sua lingua di riferimento.

In questo modo, pur se diversi utenti accedono alla medesima pagina, la stessa può mostrare contenuti differenti, secondo la loro nazionalità.

Naturalmente come tutte le tecniche viste nel cap. 5.4 anche il cloaking può essere utilizzato in modo scorretto.

Può essere usato come un'evoluzione delle doorway pages, poiché il criterio di fondo rimane comunque quello di mostrare delle pagine diverse ed ottimizzate ai search engine e ai loro spider.

Per esempio: un sistema di cloaking può capire se la pagina è richiesta da un utente o dallo spider di un search engine, e mostrare contenuti normali nel primo caso e una pagina piena di keyword nel secondo.

Indi, ai search engine possono essere proposte delle doorpage, mentre ai visitatori normali possono essere mostrate le pagine standard del sito, oppure, con il cosiddetto "page-jacking", l'applicazione propone allo spider il contenuto di una pagina Web di un sito qualsiasi che però figura in testa ai risultati, mentre all'utente Web è proposta la pagina standard.

Per cui, quest'uso ha lo scopo di migliorare il ranking nei search engine nel quale un singolo indirizzo Web ha associati diverse pagine, una per ogni search engine e una per l'utente finale.

I pro nell'uso di questa tecnica sono essenzialmente tre:

- Il cloaking permette di raggiungere le vette di più search engine usando una sola URL; è difficile posizionarsi in alto su tutti i search engine con una sola pagina poiché ogni motore utilizza un algoritmo diverso per indicizzarla.
- Nascondendo il codice HTML usato per i motori all'utente finale, si tengono nascoste agli occhi dei competitori le strategie usate.
- Poiché la pagina per il motore non è vista dagli utenti comuni, non deve essere "bella", pertanto può essere ottimizzata al massimo.

Il rovescio della medaglia di questa tecnica è che i maggiori search engine stanno attuando delle contromisure per contrastare l'uso del cloaking.

Molti search engine bandiranno permanentemente dai loro database i siti che lo utilizzeranno.

Il cloaking è una tecnica con la quale le richieste ad uno stesso indirizzo sono servite da pagine diverse a seconda che il visitatore sia un utente o un search engine.

In altre parole, i browser tipo Internet Explorer o Netscape vedono una pagina, mentre i motori un'altra, pur visitando lo stesso indirizzo (Es. <http://www.dominio.com/>)

Le fasi sono dunque due distinte: quella del riconoscimento, e la produzione di codice "ottimizzato".

Ci sono due metodi di riconoscimento.

Quello che fa uso dell'Agent Name cioè del nome che il browser dichiara e quello che fa uso dell'indirizzo IP.

Per quanto riguarda l'Agent Name, applicato ai search engine, il riconoscimento può essere effettuato ricavando, attraverso il linguaggio di scripting che vi è più congeniale, l>User Agent del visitatore, per capire se chi sta visitando il vostro sito è il browser di un utente o lo spider di un motore.

Riconoscendo in questo modo lo spider, e conoscendo come ragiona ogni search engine nella classificazione dei siti Web, è possibile produrre codice personalizzato ed ottimizzato per la visita di ciascuno spider, mentre ai normali utenti saranno forniti i contenuti predefiniti del sito.

Tutti i browser hanno un nome e anche gli spider dei search engine.

Per esempio quello d'Altavista è chiamato "Scooter".

Sapendo questo, inviare una pagina diversa a seconda del nome è piuttosto semplice. Si deve utilizzare qualche script da parte server che controlla il nome e si comporta di conseguenza.

Qualcosa del tipo: se Agent Name è uguale ad A o B o C, servi pagina A altrimenti servi pagina B.

Il problema di questo approccio però, è che si può falsificare facilmente il nome che il browser manda al server, pertanto non è bene utilizzare questa tecnica se si vuole tenere il codice della pagina segreto il più possibile.

L'indirizzo IP, invece, è l'indirizzo numerico che identifica i nodi collegati alla rete. Tutti hanno un indirizzo IP quando sono su Internet, e pertanto anche gli spider.

Il metodo è analogo al precedente, l'unica differenza è che invece di controllare l'Agent Name lo script controllerà l'IP.

Purtroppo però le cose sono molto più complesse poiché si deve mantenere un'enorme lista di indirizzi IP.

Inoltre, questi indirizzi possono cambiare e nuovi IP essere aggiunti.

Il grande vantaggio è che diventa impossibile per ognuno vedere il codice presentato ai motori.

Uno script che tenga in considerazione entrambe le variabili, cioè sia Agent Name sia indirizzo IP, sarà senza dubbio più accurato nel controllo di quanto non possa essere uno script fondato solamente su una delle due variabili in gioco.

Questo però non vuol dire che un search engine è sempre identificabile dall'indirizzo IP o dal testo nell'User-agent.

Infatti, alcuni motori, proprio per individuare quelle pagine che forniscono contenuti differenti tra quelli richiesti da un search engine e quelli richiesti dai normali utenti, accedono spesso alle pagine Web da più indirizzi e con user-agent diversi.

Indi, per far uso del cloaking bisogna stare continuamente aggiornati su quali indirizzi IP sono usati dai search engine, poiché quest'ultimi utilizzano molteplici indirizzi cambiandoli molto spesso.

Indi i search engine possono in qualunque istante accedere alle pagine che fanno uso del cloaking mostrando un indirizzo IP e un User-agent differenti da quelli conosciuti dal sistema d'identificazione e questo comporta una penalizzazione del sito nelle classifiche dei search engine.

Le penalizzazioni per chi fa uso di questa tecnica possono essere molto aspre, possono addirittura arrivare a cancellare definitivamente i siti Web dall'archivio del search engine.

Qui di seguito sarà proposto un esempio di script che riconosce lo spider dall'utente usando l'Agent Name.

```
<?
```

```
$spider = array("Spider", "ArchitextSpider", "MaxBot.com", "allothers",  
"ESISmartSpider", "FAST-WebCrawler", "xcrawler", "AllOthers",  
"GreatWhiteCrawl", "altavista", "Scooter", "G.R.A.B.", "scooter", "Mercator",  
"crawler", "inktomi", "Slurp", "Googlebot", "google", "Gulliver", "NorthernLight",  
"lycosidae", "lycos", "Lycos_Spider", "T-Rex", "InfoSeek", "infoseek",  
"ArchitextSpider", "excite");
```

```
$i = "0";  
while( $i < count($spider)){  
if (ereg("$spider[$i]", "$HTTP_USER_AGENT")) {  
$spider_check='1';  
}  
$i++;  
}
```

```
if (isset($spider_check)){
```

```
}  
?>
```

5.4.4 Scambio di link

L'obiettivo dello scambio di link è di aumentare la "popolarità da link" di un sito.

Lo scambio di link consiste nella possibilità di mostrare un link al proprio sito su diversi altri siti nella rete, in cambio però si è tenuti ad ospitare sul proprio sito alcuni link ad altri siti che hanno aderito all'iniziativa.

Ed in questo non ci sarebbe niente di male, anzi, è consigliato scambiare qualche link con siti Web che trattano argomenti correlati a quelli discussi sul proprio sito.

Il problema è che questi scambi di link sono mal visti dai search engine perché vengono prodotte decine di link e diventano un evidente tentativo d'evidenziare la rilevanza di un sito agli occhi dei search engine, anche perché si attuano scambi di link tra siti che non hanno alcun argomento in comune.

Inoltre, i link inseriti da qualche parte senza motivo o resi visibili solo agli spider servono a ben poco, poiché i search engine più avanzati danno preferenza a quei link che appaiono introdotti in un contesto discorsivo.

Naturalmente, come tutte le tecniche viste in questo capitolo, anche questa è soggetta ad una penalizzazione nelle classifiche dei search engine.

Per concludere, è meglio produrre siti Web con contenuti interessanti in modo che diversi Webmaster siano portati a consigliare un sito interessante e dai buoni contenuti, oppure chiedere uno scambio di link tra siti che trattano gli stessi argomenti in modo da migliorare l'esperienza di navigazione di molti utenti, e di apparire più rilevanti ai search engine.

6. Appendice: esperienza stage

Dal 8 marzo 2004 al 22 giugno 2004 ho svolto uno stage presso la Nuova Fergia s.p.a., un concessionario del gruppo Fiat, la quale comprendeva auto Fiat, Lancia ed Alfa Romeo ed in più un marchio proprio, Gi Auto, per la vendita d'auto usate.

Il progetto di stage consisteva nella gestione di pagine Web con particolare attenzione all'usability ed al content management.

Il sito da seguire era appunto quello di Nuova Fergia, un sito nel quale venivano catalogate la maggior parte delle auto usate presenti e disponibili nell'azienda.

Tramite un sistema di ricerca presente nel sito era possibile cercare l'auto della marca e del modello desiderato ma non solo, era possibile effettuare la ricerca anche in base al prezzo, ai kw, al tipo d'alimentazione, ai km, ecc.

Il mio compito nell'azienda ben presto si dimostrò non essere quello delineato dal progetto di stage, ma si basò soprattutto sull'aggiornamento e sulla gestione del database per il sito Web, quindi soprattutto un lavoro di data entry.

Il sito di Nuova Fergia veniva seguito in sostanza da tre aziende:

- La Nuova Fergia s.p.a. che gestiva il database del sito (IBM as400).
- La Telemar s.r.l. che seguiva il sito dal punto di vista tecnico (per esempio se c'erano problemi nella ricerca delle auto il problema era di sua responsabilità), in più il sito era costruito su una piattaforma progettata direttamente dalla Telemar (Sitengine).
- La Caval Service s.r.l. che seguiva il sito dal punto di vista dei contenuti (articoli, foto, ecc.). Inoltre aveva anche un ruolo di consulente marketing.

Il mio ruolo, quindi, oltre ad essere quello d'aggiornamento e gestione del database era anche quello di "coordinatore", in altre parole punto di riferimento per Caval Service e Telemar.

Oltre a ciò, dovevo tenere controllato il sito ed in caso di problemi prendere contatto con l'azienda competente oppure se erano problemi imputabili al database, risolverli. Le auto inserite nel database dovevano essere anche accompagnate da una fotografia e questo era un altro mio compito.

Una volta fotografate le auto, venivano nominate con il proprio numero di targa, oppure nel caso d'auto estere con le ultime sei cifre del numero di telaio e spedite via e-mail alla Caval Service, la quale avrebbe ridimensionato e pubblicato le foto nel sito.

Alla fine d'ogni mese, dovevo spedire un report alla Caval Service riguardante le vendite dell'usato e la provenienza dei clienti.

Questi dati li estraevo tramite query dal database.

I report poi erano oggetto di discussione e di decisione nell'assemblea che si svolgeva ogni inizio mese, per delineare strategie commerciali e di marketing.

Data la mancanza di un progetto di stage e la difficoltà di trovare un accordo su un progetto di tesi, tra azienda, professore e il sottoscritto, d'accordo con il docente ho deciso di sviluppare la tesi su un argomento non riguardante direttamente lo stage ma a mio avviso di molto interesse ed attualità: "Il posizionamento dei siti Web nei motori di ricerca".

Dal mio punto di vista, questa esperienza di stage è stata positiva per quanto riguarda l'inserimento in un ambiente lavorativo e la relazione con i colleghi di lavoro.

Ma per l'accrescimento di abilità riguardanti il mio percorso di studio è stata un'esperienza povera; non ho imparato e approfondito argomenti riguardanti il mio corso di laurea.

Per di più non ho potuto basare la mia prova finale sull'esperienza di stage e quindi ho dovuto sviluppare un lavoro cercando e studiando materiale su libri e siti Web senza poterlo mettere in pratica nella realtà.

7. BIBLIOGRAFIA

- “MINING THE WEB” Soumen Chakrabarn
- “MOTORI DI RICERCA E VISIBILITA’ SUL WEB” Mauro Lupi
- “MARKETING MANGEMENT” Russell S. Winer
- “MODERN INFORMATION RETRIEVAL” Ricardo Bazza-Yates
Berthier Riberlo-Neto

- WWW.MOTORIDIRICERCA.IT
- WWW.MOTORICERCA.INFO
- WWW.HTML.IT
- WWW.FATTITROVARE.IT
- WWW.SUBMISSION.IT
- HTTP://PRO.HTML.IT
- WWW.STUDIOCAPPELLO.IT
- WWW.POSIZIONAMENTO-MOTORI.COM
- WWW.ITISBARSANTI.ORG