

Sistemi per il recupero delle informazioni

Gabriele Pozzani

A.A. 2015/2016

**Corso di Laurea Magistrale in
Editoria e Giornalismo**

**Sistemi per il recupero
dell'informazione sul Web:
Search Engine Optimization**

SEO (I)

- Oltre a comportamenti scorretti per aggirare gli algoritmi dei motori di ricerca e migliorare il ranking del proprio sito, esistono tecniche lecite per migliorare la propria “reperibilità” e visibilità
- Tali tecniche ricadono sotto il nome di Search Engine Optimization (SEO)
- Esistono aziende apposite per l'ottimizzazione dei siti web con tecniche SEO
 - Ma attenzione, ne esistono anche molte “meno serie” che usano tecniche non lecite

3

SEO (II)

- Spesso il termine SEO non è ben distinto dalle tecniche di posizionamento illecite
- A tal fine può essere utile parlare anche di
 - Black hat SEO: insieme delle tecniche illecite (=spamdexing)
 - White hat SEO: insieme delle tecniche di ottimizzazione lecite
 - Hanno l'obiettivo di aumentare effettivamente la qualità del sito e quindi di creare “valore”

4

SEM (I)

- Il SEO è una delle attività del Search Engine Marketing (SEM)
- Obiettivi del SEM sono:
 - Benchmarking online: studio dello scenario competitivo sui motori di ricerca (in termini di saturazione del canale e qualitativi dei messaggi/contenuti proposti)
 - Online branding: incremento della notorietà attraverso il presidio dei risultati dei motori di ricerca
 - Brand monitoring: monitoraggio del sentiment nei risultati dei motori di ricerca
 - Database building: avvio di ulteriori attività di marketing che sfruttino il motore come porta di ingresso verso le iniziative promozionali
 - Vendita diretta (E-commerce) attraverso i motori di ricerca e i comparatori prezzi (motori verticali)
 - Online customer support: assistenza clienti online

5

SEM (II)

- Le principali attività nel SEM sono:
 - Search results monitoring: monitoraggio dei risultati proposti per specifiche query
 - Posizionamento sui motori di ricerca
 - Posizionamento naturale tramite la SEO
 - Posizionamento a pagamento sui motori di ricerca
 - inserzioni a pagamento (link sponsorizzati) per le quali l'inserzionista acquista le parole per le quali vuole comparire
 - Search engine advertising (SEA): attività di gestione di campagne di link a pagamento su siti e portali che maggiormente consentono di raggiungere il target d'impresa (detta anche Keyword advertising)

6

SEO (III)

- Negli ultimi anni i motori di ricerca hanno aggiornato i propri algoritmi di indicizzazione
- Alla parte basata sull'analisi dei link (e.g., PageRank) hanno aggiunto l'analisi dei contenuti
- Le tecniche SEO puntano ad ottimizzare i contenuti di un sito al fine di migliorarne l'”aspetto”, il valore e la qualità agli occhi del motore di ricerca
- L'ottimizzazione avviene rispetto alle linee guida (best practice) dettate dai motori di ricerca

7

Linee guida (I)

- Le linee guida puntano a migliorare la qualità di un sito sotto diversi aspetti
 - Struttura
 - Contenuti
 - Comportamento nei confronti dei crawler
- Una parte delle linee guida nasce da studi di usabilità ed interazione uomo-macchina
 - Sono utili in generale per costruire un buon sito, indipendentemente dai motori di ricerca

8

Linee guida (II)

- Vediamo ad esempio le linee guida suggerite da Google
 - Istruzioni per webmaster
 - Ottimizzazione per motori di ricerca
 - Guida introduttiva di Google all'ottimizzazione per motori di ricerca (SEO)

9

SEO base, titolo (I)

- Titoli
 - Il titolo comunica gli argomenti della pagina sia agli utenti che ai motori di ricerca
 - Il titolo va indicato nel tag `<title>` all'interno del tag `<head>`
 - Il contenuto del tag `<title>` comparirà, di norma, nella prima riga dei risultati
 - Buone norme
 - Possibilmente, usare un titolo univoco per ogni pagina del sito
 - Titolo che comunichi efficacemente l'argomento del contenuto
 - Titoli brevi ma descrittivi

10

SEO base, titolo (II)

```
<html>  
<head>  
<title>Le figurine dei calciatori di Marco - vendita online di figurine, guida ai prezzi,  
articoli e novità</title>  
<meta name="description" content="Le figurine dei calciatori di Marco: vasta selezione di figurine di tanti famosi protagonisti del mondo del calcio">  
</head>  
<body>
```



Figurine dei calciatori

Cerca con Google Mi sento fortunato

Le figurine dei calciatori di Marco 

Le figurine dei calciatori di Marco, vasta selezione di figurine di tanti famosi protagonisti del mondo del calcio.
www.figurinedeicalciatoridimarco.it - [Copia cache](#) - [Simili](#)

SEO base, descrizione (I)

- Le descrizioni (o sommari) possono essere specificati per ogni pagina
 - Il meta tag `description` di una pagina fornisce ai motori di ricerca un sommario dell'argomento della pagina
 - Potrebbero essere usati come snippet (frase di descrizione) per le pagine
 - Buone norme
 - Scrivere una descrizione che possa informare ed interessare gli utenti
 - Usare una descrizione unica per ogni pagina

SEO base, descrizione (II)

```
<html>
<head>
<title>Le figurine dei calciatori di Marco - vendita online di figurine, guida ai prezzi,
articoli e novità</title>
<meta name="description" content="Le figurine dei calciatori di Marco: vasta selezione di figurine di tanti famosi protagonisti del mondo del calcio">
</head>
<body>
```



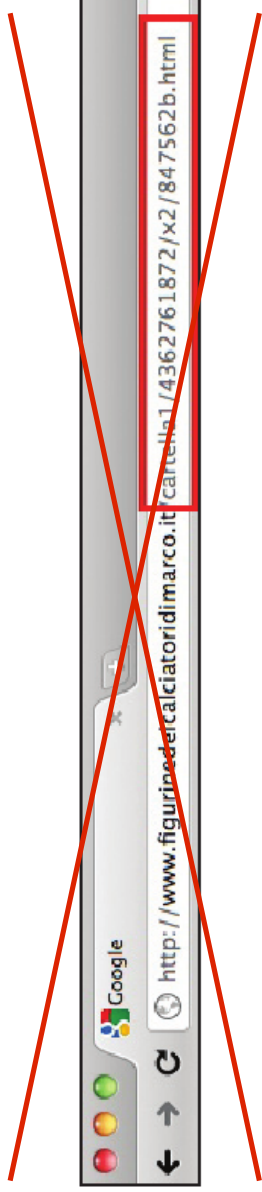
13

Struttura, URL (I)

- URL semplici da capire comunicano meglio i contenuti
 - Se l'URL contiene parole rilevanti, fornisce agli utenti e ai motori di ricerca più informazioni sulla pagina
 - l'URL di un documento viene mostrato come parte di un risultato di ricerca, sotto il titolo e la snippet
 - Buone norme:
 - Utilizzare delle parole nell'URL, anziché codici o sigle
 - Semplificare la struttura delle tue directory
 - Fornire una versione unica dell'URL per raggiungere un dato documento

14

Struttura, URL (II)



15

Struttura, semplicità di navigazione (I)

- Pianificare la navigazione basandosi sulla pagina principale
 - Pensare al modo in cui un visitatore passerà dalla pagina principale ad una con contenuti più specifici
- Utilizzare le liste “breadcrumb”
 - lista di link in linea in cima o in fondo alla pagina che permette ad un visitatore di tornare indietro verso la sezione precedente o verso la home page

Le figurine dei calciatori di Marco > Articoli > Le dieci figurine più rare del mondo

16

Struttura, semplicità di navigazione (II)

- Preparare due sitemap: una per gli utenti
 - pagina che mostra l'architettura del sito e che, generalmente, consiste in una lista gerarchica delle pagine.
 - Un visitatore potrebbe usarla per orientarsi meglio sul sito e trovare quello che sta cercando una per i motori di ricerca
 - rende la scansione del sito più semplice per i crawler
 - Dice ai crawler ogni quanto rileggere il sito perché potrebbe essere cambiato

17

Contenuti

- Offrire contenuti e servizi di qualità
- Pensare alle parole che un utente potrebbe utilizzare per trovare una parte del contenuto
 - Utenti diversi possono usare ricerche diverse per cercare un sito
 - Anticipare tali differenze di approccio alla ricerca e tenerne conto durante la stesura del contenuto (usando una buona combinazione di frasi chiave)
- Creare testi facili da leggere
- Organizzare il contenuto a seconda dell'argomento

18

Contenuti, anchor text

- L'anchor text è il testo cliccabile dei link
 - Dice qualcosa di utile sulla pagina linkata sia per gli utenti che per i motori di ricerca
 - Scegliere testi descrittivi
 - Usare un testo conciso
 - Formattare i link in modo che siano riconoscibili dal resto del testo

19

Contenuti, immagini

- Usare nomi di file e testo alternativo brevi ma descrittivi
 - Il testo alternativo è
 - Visualizzato quando per qualche ragione l'immagine non può essere visualizzata
 - Usato dai motori di ricerca come “descrizione dell'immagine” al fine di indicizzarla
- Raggruppare tutte le immagini in un'unica cartella
- Usare formati di immagini comunemente supportati
 - JPEG, GIF, PNG, BMP

20

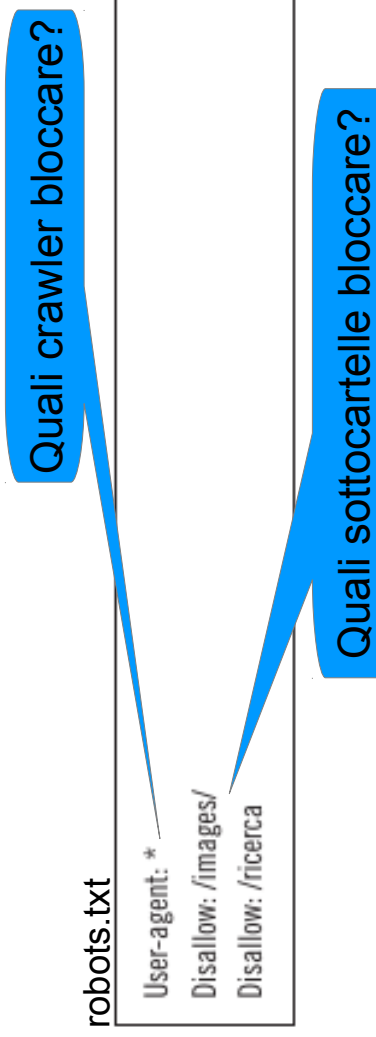
Contenuto, suddivisione del testo

- Organizzare il testo in paragrafi
- Evidenziare in modo appropriato titoli e sotto titoli
 - Con gli appositi attributi HTML

21

Comunicare con i crawler

- Ogni crawler prima di accedere ad un sito vi cerca un particolare file, **robots.txt**
 - Questo file comunica ai motori di ricerca se debbano o meno accedere e indicizzare le varie sezioni del sito



22

Combattere il web spam in blog e forum (I)

- Un modo di spamdexing consiste nell'inserire nei commenti di blog/forum il link ad una pagina che si vuole “pubblicizzare”



- È possibile dire ai crawler di non seguire alcuni link
 - Attributo `rel="nofollow"` dei link
- Applicando in automatico questo attributo ad eventuali link inseriti nei commenti si evita che questi vengano considerati e creino spam
 - Si evita che il proprio blog/forum venga considerato spam e quindi venga penalizzato dai motori di ricerca

23

Combattere il web spam in blog e forum (II)

- Inserire un CAPTCHA in modo da bloccare commenti automatici



24

Strumenti online

- Google e Microsoft mettono a disposizione strumenti online gratuiti
 - aiutano ad avere maggiore controllo su come i motori di ricerca interagiscono con i siti web e forniscono informazioni utili a riguardo
- Permettono
 - vedere quali parti di un sito il crawler non riesce a raggiungere
 - invio di una Sitemap XML
 - analizzare e generare file `robots.txt`
 - rimuovere URL indicizzati
 - identificare problematiche relative a titoli e descrizioni delle pagine web
 - conoscere le principali query di ricerca che portano a un sito
 - dare un'occhiata a come il crawler vede le pagine
 - rimuovere elenchi di link indesiderati che potrebbero essere utilizzati nei risultati
 - ricevere notifiche in caso di violazioni delle linee guida della qualità e inoltrare una richiesta di reconsiderazione