

Mappa dell'analisi di un sito web



Enrico Altavilla

Proud Member of
SearchBrain
Marketing with Search in Mind

Nihil difficile volenti.

Sommario

Introduzione.....	6
La mappa degli argomenti.....	7
Standard HTML.....	8
Regione e linguaggio.....	8
Paginazione.....	8
Architettura.....	9
Distribuzione tra pagine o sezioni dell'importanza/PageRank.....	9
Coerenza con quanto gli utenti cercano.....	9
Contenuti.....	10
Canonicalizzazione.....	10
Testi.....	10
Immagini e video.....	10
Contenuti generati dagli utenti.....	11
Meta informazioni.....	11
Tag TITLE e META "description".....	11
Usabilità (desktop e mobile).....	12
Design.....	12
Interazioni.....	12
Percorsi degli utenti.....	12
Configurazione della piattaforma di analytics.....	13
Motore di ricerca interno.....	13
Integrazione con AdWords e GSC (Google Search Console).....	13
Sicurezza e malware.....	14
Protocollo HTTPS.....	14
CMS e versione di plugin/estensioni.....	14
Check di sicurezza.....	14
Check sui malware.....	14
Contenuti inaspettati.....	15
Parole su Google Search Console.....	15
Ricerca manuale di parole sospette.....	15
Spam e penalizzazioni.....	16

Tendenze del traffico organico	16
Messaggi su GSC.....	16
Cambiamenti durante un update di Google	16
Popolarità	17
Backlink.....	17
Query brandizzate	17
Citazioni del brand	17
Comparazione con i competitor.....	17
Dati strutturati	18
"Briciole di pane"	18
Box di ricerca nei Sitelink.....	18
Riguardo i contenuti	18
Recensioni	18
Aziende e organizzazioni.....	19
Prodotti.....	19
Ricette.....	19
Eventi.....	20
Video.....	20
Articoli	20
Web Performance	21
Ottimizzazione dei trasferimenti	21
Cache dei browser	21
Unione di file multipli CSS o JavaScript.....	21
CDN pubblici.....	22
Risparmiare byte	22
Supporto di HTTP 304	22
Codice pulito	22
Minificazione	22
File JS inutilizzati.....	22
Regole CSS inutilizzate	23
Compressione.....	23
Cambio di formato delle immagini	23

Compressione lossless delle immagini.....	23
Rendering dei browser	23
JavaScript che bloccano il rendering.....	23
CSS che bloccano il rendering.....	24
Browser reflow.....	24
Dichiarazione del charset	24
Dimensioni delle immagini in pixel	25
Regole CSS inutilizzate	25
Web font.....	25
Predizione.....	25
Prefetch dei DNS	26
Subresource hint.....	26
Prefetch	26
Prerender.....	26
Protocolli.....	27
SPDY	27
Supporto di HTTP/2	27
Dispositivo.....	27
Indicizzazione	28
Direttive noindex	28
Statistiche di indicizzazione su GSC	28
Richieste di indicizzazione su GSC	28
Richieste di rimozioni su GSC	28
Indicizzazione della app	28
Crawling.....	30
Navigazione possibile senza cookie e referrer.....	30
robots.txt.....	30
Direttive nofollow	30
Parametri degli URL.....	31
Metodi dei form HTML	31
Sitemap XML.....	31
Log del web server	32

Dati di crawling di GSC	32
Statistiche di blocco su GSC	32
HTTP	33
Codici di stato 4xx / 5xx	33
Redirezioni	33
Web server pronto per HTTP/2	33
Risposte condizionali	34
Spider e robot bloccati	34
Anti flood e protezioni da attacchi DoS	34
Invii basati sull'user agent	34
Invii basati su IP e associazioni geografiche	34
Invii basati sul linguaggio del browser	34
Conclusione.....	35
Glossario.....	36
Archivio.....	36
Crawler	36
GSC / Google Search Console.....	36
Knowledge Graph	36
Indicizzazione.....	36
Query	36
Rich snippet	36
SERP	36

Introduzione

Questo **e-book gratuito** è nato per condividere con altri professionisti SEO e di marketing quali aspetti possono essere tenuti in considerazione quando si deve svolgere l'analisi di un sito web per finalità di **ottimizzazione**.

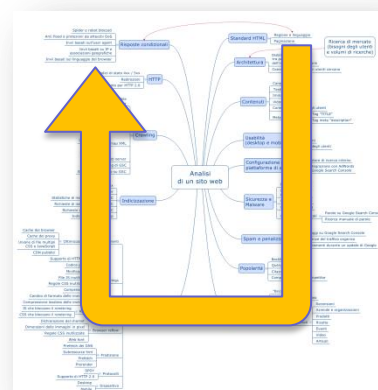
L'e-book prende spunto da una mappa divulgata in passato attraverso il mio profilo Google+, ma aggiunge una descrizione/commento per ciascuno degli aspetti elencati, citando il tipo di valutazione che bisognerebbe svolgere.

Ho incluso dei box, quelli con la lente di ingrandimento, che contengono **link di approfondimento** alle tematiche discusse e a fine documento è disponibile un piccolo **glossario**, che chiarisce il significato di alcuni termini.

Ovviamente, non ho la pretesa di proporre questa mappa come se fosse completa di tutto ciò che può essere valutato in sede di analisi di un sito. Questa è semplicemente la **mia** personale mappa di argomenti, ovvero ciò che mi è utile per le attività di consulenza che svolgo.

La trattazione degli argomenti inizia dalla zona in alto a destra della mappa (dedicata agli standard HTML) e prosegue in senso **antiorario**, concludendosi con la zona in alto a sinistra. L'ultimo argomento è dunque quello delle risposte condizionali.

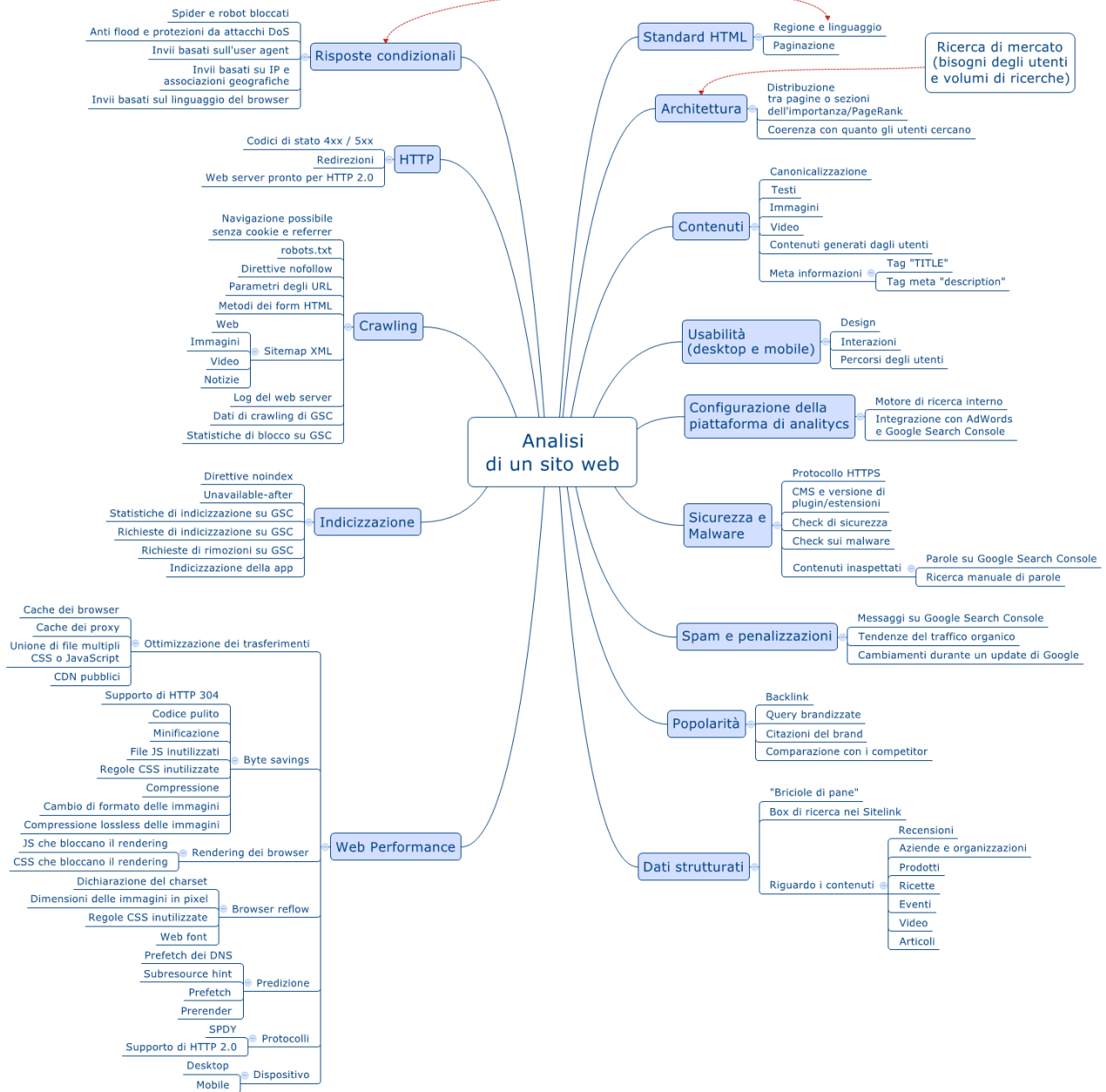
Questo e-book nasce anche per dare visibilità al corso "[SEO per eCommerce](#)", di cui sono realizzatore e docente: si tratta di una full-immersion di un giorno, dedicata esclusivamente all'ottimizzazione dei siti di eCommerce e alla gestione delle criticità SEO più tipiche negli shop online.



Mi auguro che la mappa commentata ti sia utile, ti invito ad iscriverti alla [mia newsletter](#) se vuoi ricevere in email altri miei approfondimenti SEO e ti auguro buona lettura!

Enrico Altavilla

La mappa degli argomenti



Standard HTML

È possibile aiutare i motori di ricerca a classificare correttamente i contenuti di un sito facendo uso di appositi markup HTML.

Regione e linguaggio

Google chiede ai webmaster di siti multilingua o multiregione di usare dei tag LINK con attributi `rel="alternate"` e `hreflang` per indicare, di una specifica pagina, quali varianti linguistiche o regionali esistono sul sito web.

Queste stesse informazioni possono in realtà essere fornite al motore di ricerca anche attraverso le sitemap XML.



Utilizzo di hreflang per gli URL locali e la lingua ►

Paginazione

Gli attributi `rel="next"` e `rel="prev"` dei tag LINK aiutano i motori di ricerca a comprendere che una pagina fa parte di una serie. Gli usi più tipici di questi attributi avvengono nelle pagine che elencano i risultati di una ricerca oppure nelle pagine catalogo di prodotti o servizi.



Indicazione dei contenuti suddivisi in pagine ►

Architettura

La definizione dell'architettura di un sito si può basare, tra le altre cose, anche su una ricerca di mercato volta a scoprire quali esigenze i consumatori esprimono attraverso le ricerche sui motori.

Queste informazioni sono utili per stabilire quali contenuti il sito web deve ospitare e come organizzarli.

Distribuzione tra pagine o sezioni dell'importanza/PageRank

È possibile fare un'analisi del linking interno per osservare come l'importanza di un sito si divide tra le pagine o sezioni che lo compongono. Idealmente, le pagine più strategiche per il business dovrebbero essere quelle con importanza più alta, ovvero quelle meglio linkate.



Come calcolare la distribuzione del "PageRank" tra le pagine di un sito ►

Coerenza con quanto gli utenti cercano

L'architettura di un sito può essere progettata tenendo conto di ciò che gli utenti cercano e *quanto* lo cercano. Analizzare i volumi di ricerca per le query che sono interessanti per il business può aiutare a stabilire quali contenuti produrre e quanta importanza attribuire a ciascuno di essi.

Contenuti

Contenuti

Dai testi ai contenuti multimediali, è necessario determinare se il sito propone i contenuti in modo chiaro o se il suo linking interno induce i motori ad acquisire le stesse risorse ad URL sempre diverse.

Canonicalizzazione

Nel caso in cui siano presenti sul sito contenuti duplicati, causati da parametri negli URL o da un linking interno non pianificato (risorse linkate in versione www e non-www, per esempio), è necessario porvi rimedio.

I segnali per indicare ai motori quali versioni delle risorse sono quelle "canoniche" sono molteplici.



Utilizzo di URL canonici ►

Testi

È importante determinare sia se esistono pagine dai contenuti eccessivamente simili tra loro sia se i testi appaiono anche in altri siti sul web.

Immagini e video

Come per i testi, bisognerebbe evitare che gli spider trovino la stessa immagine o lo stesso video sotto URL differenti.

È possibile impartire la direttiva "canonical" attraverso le intestazioni HTTP.



Specifica un link canonico nella tua intestazione HTTP ►

Contenuti

Contenuti generati dagli utenti

Che siano contenuti supplementari, per esempio commenti e recensioni, o contenuti principali, come i post di un forum di discussione, è importante valutare quanto i contenuti generati dagli utenti sono utili per i visitatori del sito e se sono distanti da pratiche di spamming.



[Istruzioni sullo spam generato dagli utenti ▶](#)
[Spam nei commenti ▶](#)

Meta informazioni

I metadati possono fornire informazioni strutturate (es: Schema.org) oppure fornire descrizioni di altri dati (es: tag HTML META "description").

Tag TITLE e META "description"

I testi prodotti per questi due tag devono seguire le linee guida definite dai motori di ricerca.



[Controllo dei titoli e degli snippet di pagina ▶](#)

Usabilità (desktop e mobile)

Spesso ciò che è poco usabile per gli utenti crea problemi anche ai motori di ricerca.

Design

Attribuire ai contenuti secondari molta enfasi e relegare i contenuti primari a zone meno prominenti può danneggiare l'usabilità del sito e rendere più difficile ai motori la comprensione di quali siano i contenuti primari.

Un esempio su tutti: l'uso improprio di un "carousel".



[Designing Effective Carousels: Create a Fanciful Amusement, Not a House of Horrors ▶](#)

Interazioni

Che si tratti di click sui link, compilazioni di un modulo o semplice scorrimento di una pagina, l'interfaccia utente rappresenta un elemento che va valutato per accertarsi che le possibili interazioni appaiano evidenti ai visitatori.

Tranne eccezioni, se qualcosa risulta evidente agli utenti, risulterà evidente anche ai motori di ricerca.

Percorsi degli utenti

Un utente deve sempre sapere in che zona del sito si trova e dove presumibilmente arriverà quando seguirà un link.

La creazione di link che puntano a risorse duplicate o semplicemente errate genererà confusione sia negli utenti sia ai motori di ricerca.

Configurazione della piattaforma di analytics

Motore di ricerca interno

La piattaforma di analytics va configurata affinché sia possibile monitorare le query che gli utenti inviano al motore di ricerca del sito web. Le informazioni che è possibile estrarre da queste query sono numerose ed estremamente utili; per esempio, le query mettono in evidenza quali contenuti del sito sono difficilmente raggiungibili usando i link.

Integrazione con AdWords e GSC (Google Search Console)

Se la piattaforma di analytics lo permette, è sempre suggerito collegare la stessa ad AdWords, al servizio GSC (Google Search Console, precedentemente chiamato "Google Webmaster Tools") e ad altri servizi che possono fornire informazioni su come gli utenti si comportano sul sito.

Sicurezza e malware

Nessun motore di ricerca è contento di suggerire ai propri utenti di visitare un sito web che potrebbe danneggiarli.

Protocollo HTTPS

Fornire agli utenti del sito uno scambio di dati protetto dalla criptazione del protocollo HTTPS è un buon modo per apparire scrupolosi nei confronti dell'utenza e per ingraziarsi Google.



[HTTPS as a ranking signal ▶](#)

CMS e versione di plugin/estensioni

I CMS usati ed i plugin che essi sfruttano vanno sempre aggiornati alle ultime versioni disponibili per evitare di rendere il sito vulnerabile ad attacchi esterni.

Check di sicurezza

È sempre suggerito fare un check periodico del sito con uno strumento per la scansione delle vulnerabilità. Uno di questi è [Sucuri](#).

Check sui malware

Google monitora costantemente una grande quantità di siti per individuare eventuali malware installati o di cui i siti si fanno tramite. L'azienda fornisce anche indicazioni su come gestire un caso di infezione.

È possibile controllare se Google ha individuato malware su un sito visitando il seguente URL, inserendo il nome del sito da valutare al posto di "nomesito.it":

`https://www.google.com/safebrowsing/diagnostic?site=nomesito.it`



[Come sistemare un sito infetto ▶](#)

Contenuti inaspettati

La qualità di un sito dipende anche dai contenuti che esso ospita. A volte alcuni contenuti indesiderati vengono pubblicati sul sito da utenti esterni o da software che sfruttano falle di sicurezza.

Parole su Google Search Console

Su GSC è possibile dare un'occhiata a quali sono le parole più frequentemente citate nei testi del sito. Se emergono parole sospette, non in linea con quelli che dovrebbero essere i temi trattati dal sito, è suggerito approfondire il fenomeno.



[Parole chiave per i contenuti ►](#)

Ricerca manuale di parole sospette

Il buon vecchio operatore "site:" di Google è uno strumento eccellente per comprendere quali sezioni di un sito sono state oggetto di pubblicazioni indesiderate.

Spam e penalizzazioni

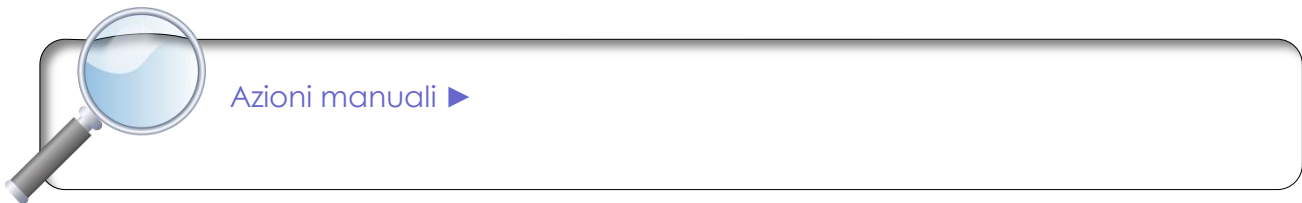
È importante comprendere quale sia la "fedina penale" di un sito. A volte essa è immacolata, altre volte è sporca e altre volte ancora mostra una sorta di limbo grigio dal quale sarebbe opportuno uscire.

Tendenze del traffico organico

Improvvisi crolli del traffico organico possono essere imputati a molteplici cause, da un problema tecnico del sito ad un cambio di algoritmi dei motori ad una penalizzazione automatica inferta al sito. Il fenomeno va ovviamente approfondito.

Messaggi su GSC

Se Google ha deciso di penalizzare manualmente un sito, l'avviso del provvedimento apparirà nella sezione messaggi di Search Console.



Cambiamenti durante un update di Google

Se il traffico organico ha subito forti modifiche in concomitanza con un update di Google, sale la probabilità che il fenomeno sia imputabile ad una modifica dei criteri di ranking del motore o ad un aggiornamento repentino dei dati su cui i criteri si basano.

Un elenco degli aggiornamenti di Google più importanti è visualizzabile con [SEO Memory](#), un tool sviluppato da SearchBrain.

Popolarità

Popolarità

La popolarità di un sito, la sua autorevolezza e più in genere la sua reputazione online, rimangono segnali cardine che i motori di ricerca sfruttano per proporre agli utenti le risorse migliori.

Backlink

La valutazione di un profilo di link fornisce un'indicazione chiara di quanto il sito sia popolare, quanta di quella popolarità deriva da azioni spontanee degli utenti o da campagne di link building e se il sito sta rischiando penalizzazioni. Oltre ai link osservabili su GSC, esistono anche strumenti appositi, come [Majestic](#).

Query brandizzate

Un indice di popolarità del brand a cui un sito web è dedicato è percepibile dalla presenza di query brandizzate, ovvero di ricerche in cui i consumatori esplicitano il nome del brand.

Una valutazione di queste query può essere effettuata con lo strumento [Keyword Planner](#) di AdWords o con strumenti appositi per l'acquisizione dei suggerimenti di Google Autocomplete, come [Keyword Researcher](#) e [Übersuggest](#).

Citazioni del brand

Più un brand è popolare, più sarà citato in rete. Una semplice ricerca su Google è un buon primo passo per avere percezione della popolarità di un brand e di come essa appare confrontata con quella dei competitor.

Comparazione con i competitor

Che si tratti di link o di presenza del brand in query di ricerca o nei testi sul web, comparare questi dati con quelli dei competitor può essere utile per avere una percezione più chiara di quanto un sito sia popolare.

Dati strutturati

Sfruttando le strutture dati [Schema](#) è possibile fornire ai motori di ricerca delle informazioni chiare e senza rischio di ambiguità sui contenuti del sito.



[Schema.org Documentation](#) ►

[Promote Your Content with Structured Data Markup](#) ►

"Briciole di pane"

I breadcrumb ("briciole di pane") servono per chiarire all'utente in che punto del sito si trova e per risalire facilmente la gerarchia di categorie che conduce alla home page.



[About breadcrumbs](#) ►

Box di ricerca nei Sitelink

Per alcuni siti, Google mostra dei Sitelink che includono un campo di ricerca. Le ricerche fatte usando questo campo vengono inoltrate direttamente al sito web.



[Sitelinks Search Box](#) ►

Riguardo i contenuti

I dati strutturati servono soprattutto a fornire ai motori delle informazioni sui contenuti del sito, dai prodotti agli eventi. Anche se presenti nel codice HTML della pagina, non c'è tuttavia garanzia che i motori di ricerca li sfruttino per modificare l'apparenza dei risultati di ricerca.

Recensioni

Le recensioni che gli utenti possono pubblicare sul sito, riguardanti un prodotto o servizio, vengono strutturate attraverso un apposito schema.



[Enabling Rich Snippets for Reviews and Ratings ▶](#)

Aziende e organizzazioni

Per alcune ricerche riguardanti aziende o organizzazioni, Google può aggiungere alle informazioni già presenti nel Knowledge Graph quelle che i gestori dei siti delle aziende forniscono attraverso schema.



[Corporate Contacts ▶](#)

Prodotti

È possibile fornire informazioni dettagliate sui prodotti venduti da un sito inserendo nel codice HTML della pagina il codice schema per questa tipologia di elementi.



[Enabling Rich Snippets for Products ▶](#)

Ricette

Anche le ricette sono tra i tipi di contenuti che possono giovare di una fornitura di dati strutturata in maniera chiara. Google possiede un indice specifico per le pagine web dedicate alle ricette e negli snippet che appaiono sui risultati di ricerca possono apparire informazioni supplementari.



[Enabling Rich Snippets for Recipes ▶](#)

Eventi

Usando la struttura Schema dedicata agli eventi, si forniscono al motore di ricerca delle informazioni riguardanti le date e i luoghi degli eventi citati. Questi dati possono apparire nei risultati delle ricerche degli utenti, arricchendo lo snippet di informazioni utili.



[Enabling Rich Snippets for Events ▶](#)

Video

Se il sito web ospita all'interno delle proprie pagine dei video, è possibile fornire ai motori delle informazioni strutturate su questo genere di contenuti, includendo un'immagine da mostrare sulle SERP.



[Enabling Rich Snippets for Videos ▶](#)

Articoli

Anche gli articoli dovrebbero essere sempre accompagnati da dati strutturati, a cominciare dall'autore, spesso citato sulle pagine in questione solo come testo.



[Enabling Rich Snippets for Articles ▶](#)

Web Performance

Per "Web Performance" si intende la capacità di un sito di apparire veloce agli utenti e reattivo alle loro interazioni. Sotto questo grande cappello, trovano posto molteplici tecniche per l'ottimizzazione della velocità.

Molte di queste tecniche di ottimizzazione si basano su caratteristiche del protocollo HTTP 1.0 e 1.1. La versione 2.0 del protocollo è stata proposta nel maggio 2015 ma le basi del protocollo rimangono quelle definite delle versioni precedenti.



[Il protocollo HTTP 1.1 ►](#)

Ottimizzazione dei trasferimenti

Tra i browser degli utenti e i siti visitati esiste un continuo transito di file: dalle risorse HTML, alle immagini che compongono le pagine web. Attraverso molteplici tecniche, è possibile ridurre il numero di trasferimenti e quindi di accorciare i tempi che gli utenti sono costretti ad attendere.

Cache dei browser

Il protocollo HTTP permette di erogare le risorse web associando loro una "data di scadenza" o altre informazioni che possono aiutare i browser a decidere se è il caso di richiedere risorse già ricevute precedentemente. È importante che i web server siano configurati in modo da associare tali informazioni alle risorse spedite.



[Sfrutta la memorizzazione nella cache del browser ►](#)

Unione di file multipli CSS o JavaScript

Per ridurre la quantità di risorse erogate ai browser si possono accorpate in un'unica risorsa i file CSS (o JavaScript) usati dalla stessa pagina. In questo modo il numero di trasferimenti si ridurrà ad uno solo e questo risultato è benefico perché ogni nuova richiesta necessita di tempo in più. Con l'avvento della versione 2 del protocollo HTTP, questa necessità di accorpamento è meno sentita.



[Combine external CSS ▶](#)

[Combine external JavaScript ▶](#)

CDN pubblici

Alcune aziende molto popolari, come Google, mettono a disposizione server pubblici (CDN) per ospitare risorse web usate da un'alta percentuale di siti, per esempio alcune librerie JavaScript come jQuery o alcuni web font.

Per più di una ragione, un sito che usa tali risorse potrebbe avere una leggera convenienza a richiamare nel codice HTML delle pagine la versione offerta dal CDN invece che una versione locale.

Risparmiare byte

Anche quando uno scambio di informazioni tra i browser e i server è inevitabile, esistono molte tecniche per ridurre il più possibile la quantità dei byte in transito.

Supporto di HTTP 304

Il protocollo HTTP mette a disposizione dei browser/spider un meccanismo per farsi spedire una risorsa web dai server solo al presentarsi di precise condizioni.

Per esempio, un browser può chiedere una risorsa solo se è stata modificata dalla data in cui l'aveva richiesta al server la prima volta. Nel caso in cui la risorsa non è cambiata, il server non invia l'intero file ma solo l'informazione che la risposta non è cambiata dalla data specificata.

Codice pulito

Un codice HTML pulito e semplice permette di risparmiare preziosi byte.

Minificazione

Per minificazione si intende un processo di riduzione della lunghezza di file HTML, CSS e JavaScript. Esso consiste nel rimuovere tutti i caratteri non indispensabili (spazi, commenti, ecc.) e a riscrivere i contenuti in una sintassi più breve.

File JS inutilizzati

Spesso succede che una pagina HTML richiami diversi file JavaScript, anche se non li utilizzerà tutti. Dotarsi di un sistema che permetta di far richiamare a ciascuna risorsa HTML solo i file JavaScript realmente necessari può ridurre considerevolmente la quantità di byte in transito.

Regole CSS inutilizzate

Accorpate molte regole CSS in un unico file è un approccio ordinario ma ha un effetto collaterale negativo: è possibile, a seconda di come un sito è stato sviluppato, che molte pagine possono usare solo una piccola percentuale delle regole contenute nel file.

Controllare quali regole sono superflue per una specifica pagina e pianificare una metodologia per dotarla del solo codice CSS che le serve è un buon modo per risparmiare byte e ridurre il carico di attività che il browser deve svolgere.

Compressione

Sul web è molto usata la compressione gzip, che è suggerito applicare ad ogni risorsa testuale prima che venga erogata al richiedente.

Cambio di formato delle immagini

A volte è possibile ridurre la dimensione in byte di un file immagine semplicemente cambiando il suo formato. Per esempio, un'immagine GIF potrebbe misurare meno byte se convertita in formato PNG.

Compressione lossless delle immagini

Non tutti i programmi di grafica creano immagini JPEG, PNG e GIF cercando di minimizzarne la dimensione in byte. Esistono software e servizi specifici che permettono di riscrivere il codice interno al file in modo che occupi meno spazio e lasciando i contenuti visuali inalterati (i pixel dell'immagine saranno identici a quelli del file originario).

Rendering dei browser

I browser effettuano il rendering (disegno) di una pagina web via via che ricevono ed interpretano i file che la compongono. Purtroppo in alcuni casi il rendering necessita di essere interrotto, per ragioni tecniche.

È importante ridurre al minimo queste interruzioni, in modo che all'utente venga mostrato nel più breve tempo possibile la pagina definitiva, o quantomeno la versione definitiva della zona di pagina che sta osservando.

JavaScript che bloccano il rendering

Durante l'interpretazione del codice HTML, il browser è costretto a fermarsi ogni qualvolta incontra del codice JavaScript, eseguire il codice e riprendere l'interpretazione dell'HTML da dove si era fermata.

Se il codice JavaScript viene fornito in un file esterno, all'interruzione sopra descritta si aggiunge anche l'attesa del download del file.

Per evitare queste interruzioni, si può minimizzare la quantità di codice presente in cima al file HTML o usare tecniche di caricamento asincrone.



Rimuovi i file JavaScript che bloccano il rendering ►

CSS che bloccano il rendering

Così come avviene per i file JavaScript, anche la presenza di file CSS esterni impone al browser di interrompere il rendering e scaricare il file per interpretarlo.

Se gli stili presenti nel file CSS esterno non sono tutti necessari al disegno della sezione della pagina che sta in cima, il webmaster può estrarre dal file le regole CSS strettamente necessarie ad effettuare il rendering della cima della pagina, incorporarle nel codice HTML e far caricare il resto delle regole da un file esterno, richiamato dal fondo del codice HTML.



Ottimizza la pubblicazione di file CSS ►

Browser reflow

Ci sono casi in cui il browser è costretto a ricalcolare la posizione degli elementi presenti in una pagina web, perdendo tempo prezioso. Questa attività viene chiamata "reflow" e alcune caratteristiche della risorsa HTML aumentano il rischio che avvenga.



Minimizing browser reflow ►

Dichiarazione del charset

Spesso il charset del documento HTML viene dichiarato attraverso un tag META. È suggerito spostare la dichiarazione del charset nelle intestazioni HTTP, in modo che il browser inizi ad effettuare il parsing dell'HTML sapendo già quale charset dovranno applicare all'analisi del codice.



[Specify a Character Set Early ▶](#)

Dimensioni delle immagini in pixel

È bene specificare sempre le dimensioni di un'immagine attraverso gli attributi `width` e `height` del tag `img` o attraverso le regole CSS.

Regole CSS inutilizzate

Dell'utilità di non inviare al browser le regole CSS non usate da una pagina si è già discusso nella sezione "Risparmiare byte" di questo e-book e precisamente al paragrafo "Regole CSS inutilizzate".

Il suggerimento appare anche in questa sezione perché questa attività di ottimizzazione permette di evitare il fenomeno del reflow.

Web font

Una pagina web può caricare font da file CSS esterni, per poi usarli nei testi. Purtroppo, come avviene per tutti i file CSS esterni da caricare, nascono problemi di blocco del rendering del browser e di reflow: a seconda dei casi il testo della pagina può venire scritto in un font pre-installato sul dispositivo dell'utente e solo successivamente ridisegnato col font caricato dall'esterno.

È pertanto suggerito non eccedere nell'uso di web font e ridurre al minimo il loro utilizzo per evitare rallentamenti e sgradevoli effetti di ridisegno.

Predizione

Esiste un nutrito elenco di suggerimenti che è possibile dare ai browser, in modo che si "preparino" a gestire le risorse che presumibilmente incontreranno nel valutare il codice HTML di una pagina.



[Preconnect, prefetch, prerender... ▶](#)

Prefetch dei DNS

Tutti i nomi di dominio (o, più correttamente, gli "host") che fanno riferimento a risorse usate da una pagina web vengono trasformati dal browser in indirizzi IP. Siccome questa traduzione richiede del tempo, è suggerito esplicitare in cima al codice HTML l'elenco degli host che il browser troverà successivamente nel codice HTML.



[DNS Prefetching \(Chrome\) ▶](#)
[Controlling DNS prefetching \(Mozilla\) ▶](#)
[Supporto di prerendering e prelettura \(IE\) ▶](#)

Subresource hint

Un altro tipo di suggerimento dato al browser dalla cima del codice HTML di una pagina consiste in quali altri file dovrà scaricare da lì a poco. Si tratta in pratica di un suggerimento ad alta priorità del download di un file.



[LINK rel=subresource \(Chrome\) ▶](#)

Prefetch

Il suggerimento di effettuare un "prefetch" di una risorsa induce il browser a scaricare la risorsa dopo aver effettuato il rendering della pagina. Un modo per usare questi suggerimenti è quello di fare il prefetch di una pagina che *presumibilmente* l'utente vorrà visitare dopo quella attuale.



[Link prefetching FAQ \(Mozilla\) ▶](#)
[Supporto di prerendering e prelettura \(IE\) ▶](#)

Prerender

Il suggerimento "prerender" è uno step successivo a quello di prefetch e consiste nel suggerire al browser di scaricare ed effettuare in background il rendering di una pagina web, in modo che la stessa appaia subito all'utente non appena verrà cliccato su un link che conduce ad essa.



[Preconnect, prefetch, prerender... ▶](#)

[Supporto di prerendering e prelettura \(IE\) ▶](#)

Protocolli

Le performance di un sito dipendono anche dal protocollo usato dal browser e dal web server quando si scambiano informazioni. Storicamente il protocollo usato per il web è stato l'HTTP ma esistono protocolli complementari e versioni di HTTP più recenti.

SPDY

SPDY (si legge "speedy") è un protocollo di trasmissione progettato da Google ed è un complemento al protocollo HTTP. Configurare un server web affinché supporti SPDY, permette di velocizzare il transito di dati con i browser che supportano tale protocollo.



[SPDY: An experimental protocol for a faster web ▶](#)

[SPDY ▶](#)

Supporto di HTTP/2

Un server web che supporta l'ultima versione del protocollo HTTP, la versione 2, permetterà ai siti gestiti di godere di trasferimenti più veloci con i browser degli utenti.



[HTTP/2 ▶](#)

[HTTP/2 Implementations ▶](#)

Dispositivo

Nello svolgere le analisi sulle performance, è necessario valutare come il sito si comporta sia quando viene richiesto da utenti che usano computer desktop/laptop, sia quando viene visitato da utenti su dispositivi mobile.

Google mette a disposizione lo strumento [PageSpeed Insights](#), che può essere usato per valutare sia le prestazioni sia parte dell'esperienza utente, da dispositivi mobili e da computer desktop.

Indicizzazione

Esattamente come avviene in un libro, i motori di ricerca tengono un indice dei documenti tenuti nei propri archivi, per trovarli velocemente quando gli utenti fanno ricerche. Affinché un documento sia incluso nei risultati di ricerca è strettamente indispensabile che esso sia indicizzato dal motore.

Direttive noindex

La direttiva `noindex` è l'unica direttiva "ufficiale" per chiedere ad un motore di eliminare una risorsa dai propri indici. Vale sempre la pena di controllare in che modo le direttive `noindex` vengono usate da un sito, attività che è facile svolgere usando un crawler come [Screaming Frog](#).

Statistiche di indicizzazione su GSC

Google Search Console mette a disposizione dei gestori dei siti delle statistiche sulla quantità di risorse indicizzate dal motore di ricerca. Pur non venendo aggiornate di frequente, è facile osservare se ci sono stati fenomeni di improvvise indicizzazioni o de-indicizzazioni, eventi che andrebbero approfonditi per essere consapevoli della causa.

Richieste di indicizzazione su GSC

Un'occhiata alla pagina di GSC delle richieste di indicizzazione fatte a Google è utile per avere uno storico ed una panoramica completa delle attività di indicizzazione che i gestori del sito hanno avviato in passato.

Richieste di rimozioni su GSC

Le richieste di rimozione di risorse dall'indice che i gestori di un sito possono fare a Google vengono elencate in un'apposita pagina di Google Search Console. Spesso si fa un abuso di questo tool ed è pertanto suggerito valutare se sono state fatte in passato richieste di rimozioni e perché.



Blocco della visualizzazione dei contenuti del sito web nei risultati della Ricerca Google ►

NON utilizzare lo strumento di blocco degli URL per motivi di manutenzione ►

Indicizzazione della app

Se un sito web ha da proporre ai propri utenti un'app per dispositivi mobile, è possibile collegare sito e applicazione affinché gli utenti che hanno installato l'applicazione sul

proprio dispositivo possano ritrovarsi nei risultati di ricerca di Google un link che lancia l'applicazione.



Una migliore esperienza di ricerca per app e utenti ►

Crawling

Un prerequisito affinché il sito venga analizzato al meglio dai motori di ricerca è che i loro spider abbiano accesso alle risorse che lo compongono. Esiste una nutrita lista di caratteristiche tecniche che è necessario controllare per accertarsi che il sito non stia creando "barriere architettoniche" che frenano o bloccano il crawling.

Navigazione possibile senza cookie e referrer

Gli spider dei motori di ricerca, nella maggior parte dei casi, non richiedono risorse ai siti presentandosi con un cookie o comunicando una pagina di provenienza attraverso l'intestazione HTTP "Referer".

Alcuni siti e CMS richiedono che gli utenti accettino i cookie o comunichino ad ogni richiesta una pagina di provenienza, altrimenti viene generato un errore. È dunque importante che il/la SEO effettui un crawling del sito simulando lo spider di un motore di ricerca, senza presentare alcuna intestazione di cookie o Referer, per osservare se il sito genera errori.

robots.txt

Il file robots.txt di un sito può contenere direttive "Disallow", che vengono usate per chiedere agli spider dei motori di non richiedere una o più risorse.

È importante stabilire a monte se esiste la necessità di limitare l'azione degli spider sul sito e bisogna fare attenzione a non impedire l'accesso a file che sono indispensabili ai motori per costruirsi un'immagine oggettiva del sito web.



[Informazioni sui file robots.txt ►](#)

[Specifiche robots.txt ►](#)

Direttive nofollow

Una direttiva nofollow può essere usate per chiedere ai motori di ricerca di ignorare alcuni link, impedendo ad essi di trasferire qualsiasi segnale positivo o negativo alle pagine di destinazione. La direttiva può essere applicata a specifici link oppure a tutti i link esistenti in una pagina.



Utilizzo di rel="nofollow" per link specifici ►

Parametri degli URL

I parametri presenti negli URL servono a comunicare al sito quali contenuti e caratteristiche la pagina richiesta dovrà possedere.

L'uso esteso di parametri negli URL e il modo in cui possono essere combinati tra loro può tuttavia portare alla produzione di risorse molto simili, che non aggiungono molto ai contenuti del sito che i motori di ricerca già possiedono. È necessario dunque accertarsi che i parametri non stiano inducendo i motori ad indicizzare e archiviare risorse duplicate o superflue.



Conoscere l'impatto degli URL duplicati ►

Metodi dei form HTML

Gli "invii" di form (moduli) HTML può avvenire usando due diversi metodi: GET e POST. Il primo genera un URL con parametri associati ai campi del form, mentre il secondo non genera alcun URL.

Gli spider dei motori di ricerca possono cercare di "inviare" form che usano metodo GET nel tentativo di scoprire nuovi URL del sito, difficilmente raggiungibili attraverso i normali link. Queste perlustrazioni possono comunque indurre i motori ad indicizzare risorse di bassa qualità, a seconda del tipo di risorse prodotte dall'invio dei form. È dunque importante monitorare l'attività degli spider per controllare se stanno acquisendo risorse di qualità dubbia attraverso l'uso di form.

Sitemap XML

I file sitemap XML consentono di fornire ai motori di ricerca le liste complete delle risorse "canoniche" del sito, che potranno essere richieste dagli spider ed eventualmente essere indicizzate nel corso del tempo.

Esistono diversi formati di file sitemap, ciascuno dedicato ad una tipologia di risorse diverse. I formati più comuni sono: risorse generiche web (HTML, PDF), immagini, file video e notizie.



[Informazioni sulle Sitemap ►](#)

Log del web server

Il web server che gestisce l'erogazione delle pagine di un sito conserva in file di log un elenco di tutte le richieste ricevute nel corso del tempo, incluse le richieste effettuate dagli spider dei motori.

Analizzare quali risorse sono state chieste dagli spider è un modo per controllare se gli stessi stanno prestando attenzione alle pagine più strategiche del sito, se essi richiedono le risorse più nuove o se stanno perdendo tempo a richiedere risorse secondarie o di qualità bassa.

Dati di crawling di GSC

Google Search Console fornisce statistiche sulla quantità di risorse richiesta giornalmente dagli spider di Google, sulla quantità di byte scaricati e sulla velocità con cui il web server ha risposto agli spider del motore.

GSC propone anche una pagina che elenca gli eventuali errori che sono stati generati in risposta alle richieste che Googlebot ha fatto al sito.



[Scansione ►](#)

Statistiche di blocco su GSC

Su GSC è possibile prendere nota di quali risorse risultano bloccate agli spider di Google, ovvero quali risorse sono in Disallow nel file robots.txt oppure sono inaccessibili perché risiedono in directory del sito a cui si può accedere solo autenticandosi con username e password.

HTTP

Il protocollo HTTP consiste nell'insieme di regole da seguire affinché il transito dati tra i richiedenti (client) e i server che gestiscono i siti avvenga senza errori.

Il protocollo mette a disposizione dei webmaster dei meccanismi per comunicare correttamente nei confronti di qualsiasi richiedente, inclusi gli spider dei motori.

Codici di stato 4xx / 5xx

È suggerito effettuare un crawling del sito per analizzare l'eventuale presenza di codici di status HTTP di famiglia 4xx (errori sulla specifica risorsa) o 5xx (errori del server).

Attraverso software quali [Screaming Frog](#) è possibile svolgere i suddetti crawl, presentandosi ai server anche come Googlebot, per intercettare eventuali risposte del sito dedicate in maniera esplicita agli spider di Google.

Redirezioni

Idealmente, tutti i link di un sito che puntano a pagine del sito stesso dovrebbero condurre utenti e spider direttamente alle risorse di destinazione, senza usare URL intermedi che producono redirezioni. È quindi suggerito effettuare un crawling per controllare se alcuni link puntano a URL che effettuano redirezioni non indispensabili.

Le redirezioni sono invece suggerite nel caso in cui alcune risorse abbiano cambiato URL nel corso del tempo.

Web server pronto per HTTP/2

Nel momento in cui questo e-book viene scritto, la nuova versione del protocollo HTTP, la 2, è in fase di diffusione. Visto che HTTP/2 mette a disposizione dei siti web delle nuove tattiche per migliorare le prestazioni dei siti web, nel prossimo futuro sarà importante controllare che il web server che gestisce il sito possieda il supporto della nuova versione del protocollo.

Risposte condizionali

A volte una richiesta fatta ad un server riceve una risposta "personalizzata", a seconda delle informazioni che il richiedente ha comunicato al server.

È importante ricordare che i motori di ricerca desiderano ricevere risorse dedicate a tutti gli utenti, non ad utenti specifici.

Spider e robot bloccati

Alcuni siti o CMS possono essere stati configurati per bloccare specifici spider. È bene controllare con un software di crawling che questo blocco non riguardi anche gli spider dei motori di ricerca da cui si desidera ricevere traffico.

Anti flood e protezioni da attacchi DoS

In alcuni casi, un CMS o server può decidere di bloccare richiedenti che effettuano troppe richieste tutte in una volta. È teoricamente possibile che tali blocchi avvengano anche nei confronti di spider di motori di ricerca popolari.

Riguardo a Googlebot, è possibile scoprire se lo spider ha trovato difficoltà ad accedere al sito dando un'occhiata alla pagina di GSC dedicata agli errori ricevuti da Googlebot.

Invii basati sull'user agent

Tranne rare eccezioni, i motori di ricerca desiderano che i siti web inviino loro esattamente le stesse risorse che vengono erogate agli utenti.

Tentativi di inviare contenuti speciali per gli spider potrebbero essere confusi dai motori come tentativi di spam.

Invii basati su IP e associazioni geografiche

Alcuni siti o CMS modificano dinamicamente i contenuti di una pagina in base in base alla nazione/località presumibilmente associata al suo indirizzo IP.

Questo comportamento è ammissibile solo se tali contenuti personalizzati vengono inviati esclusivamente ad utenti che si identificano al sito attraverso un cookie di sessione. I richiedenti che non fanno uso di cookie (come gli spider dei motori) dovrebbero ricevere i contenuti "standard" di una pagina, mai contenuti personalizzati.

Invii basati sul linguaggio del browser

È possibile personalizzare testi o altri elementi in base alla lingua di preferenza indicata dall'utente al proprio browser. La maggior parte degli spider dei motori non comunica una lingua di preferenza e quindi il sito/CMS deve essere in grado di erogare comunque contenuti anche a quei richiedenti che non specificano alcuna lingua di preferenza.

Conclusione

Conclusione

La trattazione dei temi della mappa termina qui.

Al termine "ottimizzazione", più volte citato in questo e-book, ho voluto attribuire una definizione volutamente ampia, che attinge molto alla cultura SEO ma che estende le valutazioni del sito ad aspetti non necessariamente legati ai motori di ricerca.

Dal mio punto di vista "ottimizzare" è semplicemente un sinonimo di "migliorare" e la mappa che ho fornito potrebbe essere estesa con tanti altre attività che possono andare a beneficio della qualità e funzionalità di un sito.

Trattare gli argomenti di questa mappa è servito anche a ricordare quanto complesse possano essere le attività SEO. In assenza di una definizione chiara e condivisa di "SEO", la mappa è un modo per evidenziare che esistono molti modi per svolgere questo mestiere, alcuni dei quali vanno al di là delle visioni più semplicistiche di "titoli, descrizione, testi e link building".

Mi auguro di avervi fornito uno strumento utile ed una traccia per approfondire vari aspetti della SEO.



Glossario

Archivio

È il database nel quale un motore di ricerca conserva le risorse acquisite attraverso gli spider. Non va confuso con l'indice, che è una struttura che serve in fase di ricerca dell'utente per velocizzare l'individuazione delle risorse più affinenti.

Crawler

È sinonimo di spider. Si tratta del software che il motore usa per fare il download delle risorse dei siti web.

GSC / Google Search Console

È il nuovo nome di Google Webmaster Tools, un pannello di statistiche e impostazioni che Google mette a disposizione dei webmaster per gestire la relazione tra il sito web ed il motore di ricerca.

Knowledge Graph

È la struttura usata da Google per immagazzinare informazioni semantiche su vari concetti. Contiene informazioni su milioni di concetti teorici e include le relazioni che esistono tra essi.

Indicizzazione

È l'attività del motore di ricerca che crea l'indice, ovvero un elenco di termini, frasi o concetti associati con ciascuna delle risorse web archiviate dal motore.

Query

È l'interrogazione che l'utente fa al motore di ricerca. Può avvenire scrivendo delle parole in una buca di ricerca oppure pronunciandole ad un dispositivo provvisto di microfono.

Rich snippet

Nelle pagine che contengono i risultati di una ricerca, un rich snippet è una citazione di risorsa che mostra informazioni aggiuntive al semplice titolo e descrizione.

SERP

È l'acronimo di "Search Engine Result Page", ovvero la pagina di risultati creata dal motore di ricerca a seguito di una query dell'utente.