



UNIVERSITÀ
DEGLI STUDI
DI TERAMO



SEO

Search Engine Optimization

-4-

Prof. Romina Eramo

University of Teramo

Department of Communication Sciences

ramo@unite.it

Acknowledgement: Alessandro Celi (University of L'Aquila)

Analisi dei dati

- **Web Analytics 1.0:** raccogliere ed analizzare i dati del flusso dei click
- **Web Analytics 2.0:**
 - Flusso dei click
 - Analisi Multipla (web analytics, feedburner, twitter rank, fan facebook, ...)
 - Sperimentazione e Test (A/B test, Click Tale, ...)
 - Voce dei clienti (Sondaggi, Ricerche di Mercato, ...)
 - Competitive Intelligence (Google Ad Planner, Google Trends, ...)

Strumento di Analisi

- **Strumenti interni (log server)**
 - **Pro:** errori pagine, hit, ip
 - **Contro:** dati meno accurati, difficoltà nel settare lo strumento, conservazione dei dati
- **Strumenti esterni (analisi sessioni con javascript e cookie)**
 - **Pro:** semplice implementazione, dati accurati, analisi conversioni, campagne adv
 - **Contro:** limitazione alle sole pagine che hanno il codice incluso, privacy.

Flusso dei Click

- I dati dei click ci aiutano a valutare l'efficacia di pagine e campagne e ad analizzare tutti i tipi di comportamento del sito: visite, visitatori, tempo trascorso sul sito, pagine visualizzate, frequenza di rimbalzo, fonti e altro
- Utilizziamo Google Analytics
- Libro consigliato:
Web Analytics 2.0 – Avinash Kaushik
Hoepli Informatica

Analisi dei dati

- **Visite**
- **Frequenza di Rimbalzo**
- **Pagine Viste**
- **Percorsi di Navigazione**
- **Fonti di Traffico**
- **Errori**
- **Conversioni**

Visite

- **Visite:** una serie di richieste provenienti da un medesimo client in un certo arco temporale che solitamente è di 30 minuti
- **Visitatori Unici:** visitatore identificato una sola volta, in maniera univoca, in un determinato arco temporale
- **Frequenza di Visita/Recency:** la ricorrenza con cui gli utenti visitano un determinato sito web
- **Tempo medio di visita:** tempo trascorso sul sito per ogni visitatore
- **Nuovi Visitatori e di Ritorno**
- **Profondità della visita:** numero di pagine visualizzate per visita da un utente

Frequenza di Rimbalzo

- **Frequenza di Rimbalzo o Bounce Rate:** percentuale di visite in cui un visitatore approda sulla pagina web di un sito e abbandona subito dopo il sito senza visitare altre pagine.

Pagine Viste

- **Pagine Viste**
 - **Pagine Viste per Visita:** numero medio di pagine visualizzate per ogni visita
 - **Pagine più viste**
 - **Pagine di Ingresso ed Uscita:** pagine di destinazione principali, pagine di uscita principali
 - **Percorsi di Navigazione**

Sorgenti di Traffico

- **Fonti di Traffico (Referral)**
 - **Traffico Diretto**
 - **Motori di Ricerca**
 - **Keywords di Ricerca**
 - **Siti Referenti/Referral** (In google / referral vengono quindi raggruppati gli accessi al sito provenienti da Google Immagini, Google Gruppi, Google Blog, Google Shopping, Google News, Google Reader, ecc.)

Note: Google.it /Referral:

<http://www.riccardoperini.com/google-referral-analytics.php>

Conversioni

- **Conversioni**
 - **Numero Conversioni**
 - **Percentuale di Conversioni**
 - **Fonti di Conversione**
 - **Pagine che convertono di più**
 - **Canalizzazioni Obiettivo**

Errori

- Errori
 - Errori di pagine nei log

Google Search Console

- **Analisi Indicizzazione con Google Search Console (fino al 2015 era Webmaster tools)**
 - **Sitemap**
 - **Accesso Crawler**
 - **Impostazioni sito**
 - **Query di ricerca**
 - **Siti che ci linkano**
 - **Link interni**
 - **Parole chiave**

Google Webmaster Tools

- **Diagnostica**
 - Malware
 - Errori di scansione
 - Statistiche di Scansione
 - Visualizza come GoogleBot
 - Suggerimenti HTML
 - Anteprime istantanee
 - Prestazioni del sito

Penalizzazioni

- **Filtri Anti-Duplicazione dei Contenuti**
- **Filtri Anti-Spam**
- **Over-optimization**
- **Analisi delle penalizzazioni applicate dai motori di ricerca**

Google Search Console

- inviare la [sitemap](#) del sito web.
- Controllare le statistiche degli accessi di [Googlebot](#), dati tecnici come numero totale, dimensioni byte, tempo per il caricamento, [codici di stato HTTP](#).
- Tramite il file [robots.txt](#) controllare le pagine bloccate accidentalmente dal file stesso e le statistiche di copertura
- elencare i link al sito web delle pagine sia interne che esterne.
- elencare la lista delle pagine che Googlebot ha difficoltà a scansionare, inclusi gli errori.
- elencare le keyword cercate tramite Google nella [SERP](#), clic totali, visualizzazioni totali (impression), e la media del [CTR](#), per le diverse pagine del sito.
- impostare un dominio preferito (example.com piuttosto che www.example.com), per stabilire come il sito web viene mostrato nella SERP.
- gestione dei dati strutturati ([Schema.org](#)) usati per arricchire l'anteprima delle pagine web.
- ricevere notifiche da Google per le "azioni manuali" (penalizzazioni).
- report di eventuali problemi di sicurezza (sito hackerato o attacchi malware).
- aggiungere o rimuovere alla proprietà nuovi utenti per la gestione.
- il Tool "Controlla URL" consente di visualizzare lo stato attuale di indicizzazione, eventuali problemi tecnici e l'ultimo accesso del crawler.^[12]

Penalizzazioni

Il posizionamento è costituito da 4 fasi importanti:

- **ricerca delle chiavi**
- **indicizzazione**
- **ottimizzazione**
- **aumento popolarità**

Penalizzazioni

Nel corso degli anni si sono fatti passi avanti nel cercare di rendere queste quattro fasi sempre più alla portata di tutti

E' stato fatto un grande lavoro per diffondere questa cultura e negli Staff dei vari CMS ci sono dei SEO che hanno piano piano ottimizzato molte piattaforme.

Penalizzazioni

- Studio delle penalizzazioni
- Prevenire le penalizzazioni
- Uscire dalle penalizzazioni

La strada intrapresa da Google: aumentare il numero di penalizzazioni, renderle sempre più automatiche e dure.

Miniguia sulle penalizzazioni:

<https://www.seozoom.it/penalizzazioni-google-sito-seo-guida/>

Penalizzazioni

Consigli Utili:

- Fare siti per gli utenti, rispettando le linee guida del w3c, senza forzature o zone d'ombra
- Leggere e rispettare le linee guida di google

Penalizzazioni

Centro di Google per il supporto ai Webmaster:

<http://www.google.com/support/webmasters/>

A questo URL specifico ci sono le indicazioni su
cosa non fare:

<http://www.google.com/support/webmasters/bin/answer.py?hl=it&answer=35769>

Indicazioni di Google

Norme sulla qualità

[torna all'inizio](#)

Le presenti norme sulla qualità illustrano le forme più comuni di comportamento ingannevole o a scopo di manipolazione; Google, tuttavia, può rispondere in maniera negativa ad altri comportamenti ingannevoli non elencati nel presente documento (ad esempio la registrazione di nomi di siti con ortografia leggermente diversa da quella di siti molto noti). Non è consigliabile supporre che Google approvi una pagina solo perché in essa non sono state utilizzate tecniche ingannevoli. I webmaster che si impegnano a rispettare i principi di base sopra elencati offriranno agli utenti un servizio decisamente migliore e, di conseguenza, godranno di un miglior posizionamento rispetto a coloro che si dedicano a cercare scappatoie.

Se ritieni che un altro sito stia violando le norme sulla qualità di Google, [segnala tale sito alla pagina https://www.google.com/webmasters/tools/spamreport](https://www.google.com/webmasters/tools/spamreport). Google privilegia lo sviluppo di soluzioni scalabili e automatizzate dei problemi, cercando quindi di ridurre la lotta "manuale" contro lo spamming. I rapporti sullo spamming ricevuti vengono utilizzati per creare algoritmi scalabili che riconoscono e bloccano i futuri tentativi di spam.

Come e perché segnalare lo spam

- Google mette in campo soluzioni automatiche (penalizzazioni) per combattere lo spam, che lo rendono in grado di affrontare le principali tecniche di spam e prendere contromisure in modo scalabile.
- I sistemi di filtro non sono tuttavia infallibili e per questo da tempo è stato messo a disposizione di tutti un modulo pubblico di segnalazione dello spam, che è possibile visionare al seguente indirizzo:
<https://www.google.com/webmasters/tools/spamreportform?hl=it>

Come e perché segnalare lo spam

- Ci si potrebbe domandare a questo punto perché dovremmo effettuare una segnalazione a Google. Dopotutto Google è un'azienda a scopo di lucro e ci sta chiedendo di dargli una mano gratis
 - Potremmo contribuire alla penalizzazione di chi ha messo in atto queste pratiche scorrette, lasciando così più spazio e più visibilità a chi invece si è comportato bene, seguendo un ideale di web più pulito e senza trucchetti.
 - Potremmo anche decidere di segnalare solo contenuti che ci danneggiano direttamente, seguendo un principio di opportunità.

Indicazioni di Google

Norme sulla qualità - Principi di base

- Progetta le tue pagine per gli utenti, non per i motori di ricerca. Non ingannare i tuoi utenti o sottoporre ai motori di ricerca contenuti diversi da quelli che mostri agli utenti, pratica comunemente nota come "cloaking".
- Evita i trucchi per migliorare il posizionamento nei motori di ricerca. Una buona regola generale è chiederti se ti sentiresti a tuo agio nello spiegare il tuo comportamento a un sito web concorrente. Un altro test utile consiste nel chiedersi se quanto stai facendo sarà d'aiuto ai tuoi utenti e che cosa faresti se i motori di ricerca non esistessero.
- Non [partecipare a schemi di link](#) progettati per migliorare la posizione del tuo sito o il punteggio PageRank. In modo particolare, evita i link a siti di spam o "cattivi vicini" poiché ciò potrebbe influire negativamente sul tuo posizionamento.
- Non utilizzare programmi non autorizzati per inviare le pagine, verificare il posizionamento e così via. Tali programmi sprecano risorse di elaborazione e violano i nostri [Termini di Servizio](#). Google sconsiglia l'utilizzo di prodotti come WebPosition Gold™ che inviano query automatiche o programmatiche a Google.

Indicazioni di Google

Norme sulla qualità - Norme specifiche

- Evita [testo o link nascosti](#).
- Non utilizzare il [cloaking o comandi di reindirizzamento non ammessi](#).
- Non inviare [a Google query automatizzate](#).
- Non [caricare pagine con parole chiave non pertinenti](#).
- Non creare più pagine, sottodomini o domini dai [contenuti sostanzialmente duplicati](#).
- Non creare pagine che installino virus, trojan o altro [badware](#).
- Evita le ["pagine doorway" create appositamente per i motori di ricerca](#) o altri approcci mirati alla soppressione dei cookie, come i programmi affiliati con [contenuto originale scarso o nullo](#).
- Se il tuo sito partecipa a un programma affiliato, accertati che vi aggiunga valore. [Fornisci contenuti univoci e pertinenti](#), che diano agli utenti un motivo per visitare il tuo sito prima di altri.

Se riscontri che il tuo sito non rispetta queste norme, puoi modificarlo in modo da renderlo conforme e [inviare una richiesta affinché venga riconsiderato](#).

Indicazioni di Google

Schemi di link



Il posizionamento del sito tra i risultati di ricerca di Google si basa parzialmente sull'analisi dei siti che contengono link alle tue pagine. La quantità, la qualità e la pertinenza dei link sono fattori che concorrono alla valutazione del sito. I siti che contengono link alle tue pagine possono fornire un contesto sull'argomento del tuo sito e indicarne la qualità e la popolarità. Tuttavia, alcuni webmaster partecipano a schemi di scambio di link e creano pagine partner utilizzate esclusivamente a fini di crosslinking, senza tenere conto della qualità dei link, delle fonti e dell'impatto a lungo termine sui propri siti. Questa pratica viola le [istruzioni per i webmaster](#) di Google e può influire negativamente sul posizionamento del tuo sito nei risultati di ricerca. Tra gli esempi di schemi di link vi sono:

- I link aventi lo scopo di manipolare PageRank
- I link a siti di spam o non conformi ai regolamenti del Web
- Numero eccessivo di link reciproci o scambio eccessivo di link ("Collegati a me e io mi collego a te")
- [Acquisto o vendita di link per aumentare la classificazione PageRank](#)

Il modo migliore per incoraggiare l'inserimento in altri siti di [link pertinenti](#) al tuo è creare contenuti univoci e peculiari, in grado di acquisire rapidamente popolarità nella comunità di Internet. Maggiore è l'utilità dei tuoi contenuti, maggiori saranno le probabilità che un altro utente li ritenga validi per i propri lettori e inserisca un link ad essi. Prima di prendere una decisione, dovresti chiederti se tutto questo sia vantaggioso per i visitatori delle tue pagine.

Il numero di link che puntano al tuo sito non è il solo criterio da prendere in considerazione; la qualità e la pertinenza di tali link sono fattori altrettanto importanti. Vale la pena creare [contenuti attraenti](#): i link sono in genere voti editoriali assegnati per scelta e la fervente comunità dei blogger può rivelarsi ottima per generare interesse presso gli utenti del Web.

Dopo avere apportato le modifiche desiderate ed esserti accertato che il tuo sito sia conforme alle nostre istruzioni, [invia il sito affinché venga riconsiderato](#).

Link a pagamento

Perché devo segnalare a Google i link a pagamento?



Google e la maggior parte degli altri motori di ricerca utilizzano i link per determinare il grado di notorietà di un sito. Il posizionamento di un sito tra i risultati di ricerca di Google si basa in parte anche sull'analisi dei siti che contengono link che rimandano ad esso. L'analisi basata sui link è uno strumento estremamente utile per misurare il valore di un sito e ha notevolmente migliorato la qualità della ricerca web. La quantità e, soprattutto, la qualità dei link sono fattori che vengono tenuti in grande considerazione in questa valutazione.

Tuttavia, alcuni SEO e webmaster adottano la pratica di acquistare e vendere link per aumentare il PageRank, senza tenere conto della qualità dei link, delle fonti e dell'impatto a lungo termine sui propri siti. Acquistare o vendere link per aumentare il PageRank è una pratica che viola le istruzioni per i webmaster di Google e può influire negativamente sul posizionamento di un sito nei risultati di ricerca.

Non tutti i link a pagamento violano le nostre istruzioni. L'acquisto e la vendita di link fanno parte normalmente dell'economia del Web, se eseguiti a fini pubblicitari e non per manipolare dei risultati di ricerca. I link acquistati a fini pubblicitari devono essere designati come tali. Ciò può avvenire in diversi modi, ad esempio:

- Aggiungendo un attributo `rel="nofollow"` al tag
- Reindirizzando i link a una pagina intermedia, il cui accesso è bloccato ai motori di ricerca mediante un file robots.txt

Google cerca di non tenere conto dei link volti a manipolare i risultati dei motori di ricerca, quali gli scambi eccessivi di link e i link acquistati per aumentare il PageRank. Se scopri un sito che acquista o vende link per aumentare il PageRank, ti preghiamo di [comunicarcelo](#). Useremo questa informazione per migliorare la nostra funzione di rilevamento algoritmico di questi link.

Testo e link nascosti



Testo e link nascosti

Il testo o i link nascosti nei tuoi contenuti possono far sì che il tuo sito venga ritenuto inaffidabile, in quanto le informazioni presentate ai motori di ricerca differiscono da quelle proposte ai visitatori. Il testo (ad esempio un numero eccessivo di parole chiave) può essere nascosto in diversi modi, ad esempio:

- Utilizzando testo bianco su uno sfondo bianco.
- Inserendo del testo dietro un'immagine.
- Utilizzando CSS per nascondere il testo.
- Impostando la dimensione del carattere su 0.

I link nascosti sono destinati all'indicizzazione da parte di Googlebot, ma non sono leggibili da persone fisiche in quanto:

- Sono costituiti da testo nascosto (ad esempio, i colori del testo e dello sfondo coincidono).
- È stato utilizzato CSS per creare collegamenti ipertestuali di altezza non superiore a un pixel.
- Il link è nascosto in un carattere piccolo, ad esempio un trattino nel mezzo di un paragrafo.

Se nel tuo sito vengono rilevati testo e link nascosti con intento ingannevole, il tuo sito può essere rimosso dall'indice di Google e non comparirà nelle pagine dei risultati di ricerca. Quando esamini il tuo sito per verificare se contiene testo o link nascosti, cerca elementi che non siano facilmente visualizzabili dai visitatori. Il testo o i link sono stati inseriti solo per i motori di ricerca anziché per i visitatori?

Se utilizzi del testo per descrivere qualcosa che non è accessibile ai motori di ricerca, ad esempio JavaScript, immagini o file Flash, ricorda che molti visitatori che fanno uso di screen reader, browser per cellulari, browser senza plug-in e connessioni lente non potranno visualizzare tali contenuti. L'utilizzo di testo descrittivo per questi elementi migliorerà l'accessibilità del sito. Puoi eseguire un test di accessibilità disattivando JavaScript, Flash e le immagini nel tuo browser o utilizzando un browser di solo testo come [Lynx](#). Di seguito sono riportati alcuni suggerimenti su come rendere il tuo sito accessibile:

- **Immagini:** utilizza l'attributo "alt" per il testo descrittivo. Consigliamo inoltre di utilizzare intorno all'immagine una didascalia e un testo descrittivo leggibili.
- **JavaScript:** inserisci lo stesso contenuto di JavaScript in un tag noscript. Se utilizzi questo metodo, accertati che i contenuti siano identici a quelli presenti in JavaScript e che vengano visualizzati anche dai visitatori che non hanno attivato JavaScript nel proprio browser.
- **Video:** inserisci testo descrittivo riguardante il video in HTML. Puoi anche utilizzare trascrizioni.

Se sul tuo sito individui testo o link nascosti rimuovili, oppure, se sono pertinenti per i visitatori del sito, rendili facilmente visualizzabili. Se il tuo sito è stato rimosso dai nostri risultati di ricerca, consulta le [istruzioni per i webmaster](#) per ulteriori informazioni. Dopo avere apportato le modifiche desiderate ed esserti accertato che il tuo sito sia conforme alle istruzioni per i webmaster [invia una richiesta di riconsiderazione per il tuo sito](#).

Cloaking

Cloaking

Il cloaking è la pratica di presentare agli utenti contenuti o URL diversi da quelli presentati ai motori di ricerca. La restituzione di risultati diversi a seconda dello user-agent può provocare l'identificazione del tuo sito come ingannevole e la sua conseguente rimozione dall'indice di Google.

Tra gli esempi di cloaking vi sono:

- Offerta di una pagina di testo HTML ai motori di ricerca, visualizzando però una pagina di immagini o Flash per gli utenti.
- Offerta ai motori di ricerca di contenuti diversi rispetto a quelli visualizzati per gli utenti.

Se il tuo sito contiene elementi non sottoponibili a scansione da parte dei motori di ricerca (ad esempio file multimediali diversi da file Flash, JavaScript o immagini), è consigliabile non fornire ai motori contenuti mascherati. Dovresti invece considerare i visitatori del tuo sito che non riescono a loro volta a visualizzare tali elementi. Ad esempio:

- Fornisci testo alt descrittivo delle immagini per i visitatori che utilizzano screen reader o che non hanno attivato le immagini nei propri browser.
- Fornisci i contenuti testuali di JavaScript in un tag noscript.

Accertati di fornire gli stessi contenuti in entrambi gli elementi (ad esempio, lo stesso testo in JavaScript e nel tag noscript). L'inserimento di contenuti sostanzialmente diversi nell'elemento alternativo può provocare un intervento di Google sul sito.

Reindirizzamento Javascript

Comandi di reindirizzamento JavaScript non ammessi

Durante l'indicizzazione di una pagina contenente JavaScript, Googlebot indicizza tale pagina ma non può seguire o indicizzare alcun link nascosto in JavaScript. L'utilizzo di JavaScript è una pratica web totalmente legale, a differenza dell'uso di JavaScript con l'intento di ingannare i motori di ricerca. Ad esempio, l'inserimento in JavaScript di testo diverso da quello inserito in un tag noscript viola le nostre [istruzioni per i webmaster](#), in quanto visualizza contenuti differenti per gli utenti (che vedono il testo basato su JavaScript) rispetto a quelli visualizzati per i motori di ricerca (che vedono il testo basato sul tag noscript). Secondo questi criteri, la violazione delle istruzioni per i webmaster consiste nell'incorporare in JavaScript un link che reindirizza intenzionalmente l'utente a una pagina diversa da quella visualizzata dal motore di ricerca. Quando un link di reindirizzamento è incorporato in JavaScript, il motore di ricerca indicizza la pagina originale anziché seguire il link, mentre gli utenti vengono inviati all'obiettivo del reindirizzamento. Come il cloaking, questa pratica è ingannevole in quanto visualizza per gli utenti contenuti diversi rispetto a quelli visualizzati per Googlebot e può indirizzare un visitatore altrove rispetto alla sua destinazione originale.

Tieni presente che il semplice inserimento di link in JavaScript non è una pratica ingannevole. Quando esamini JavaScript sul tuo sito per verificare che quest'ultimo sia conforme alle nostre norme, tieni presente lo scopo.

Non dimenticare che, poiché i motori di ricerca in genere non possono accedere ai contenuti di JavaScript, è probabile che anche i link attendibili in JavaScript non siano ad essi accessibili (come del resto ai visitatori i cui browser non sono abilitati per JavaScript). È preferibile mantenere i link all'esterno di JavaScript o replicarli in un tag noscript.

Pagine Doorway

Pagine doorway

Le pagine doorway in genere sono un insieme di pagine di scarsa qualità dove ciascuna pagina è ottimizzata per una determinata frase o parola chiave. Spesso, le pagine doorway sono scritte per ottenere un posizionamento per una determinata frase e quindi incanalare gli utenti verso una data destinazione.

Sia che siano distribuite su più domini sia che siano presenti in un solo, le pagine doorway tendono a infastidire gli utenti e costituiscono una violazione delle istruzioni per i webmaster di Google.

L'obiettivo di Google è offrire ai propri utenti i risultati di ricerca migliori e più pertinenti. Pertanto, Google disapprova le pratiche volte a manipolare i motori di ricerca e a ingannare gli utenti indirizzandoli a siti diversi da quelli selezionati e che forniscono contenuti unicamente a vantaggio dei motori di ricerca. Google può decidere di intraprendere delle azioni nei confronti dei siti con pagine doorway e degli altri siti che si avvalgono di queste pratiche ingannevoli, tra cui la rimozione di questi siti dall'indice di Google.

Se il tuo sito è stato rimosso dai nostri risultati di ricerca, consulta le [istruzioni per i webmaster](#) per ulteriori informazioni. Dopo avere apportato le modifiche desiderate ed esserti accertato che il tuo sito sia conforme alle nostre istruzioni, [invia il sito affinché venga riconsiderato](#).

aggiornato 5/12/2009

Query Automatizzate

Query automatizzate



I Termini di servizio di Google non consentono l'invio al sistema di query automatizzate di alcun tipo senza il previo consenso esplicito di Google. L'invio di query automatizzate assorbe risorse e prevede l'utilizzo di software (ad esempio WebPosition Gold™) per stabilire il posizionamento di un sito o una pagina web nei risultati di ricerca di Google per varie query.

aggiornato 5/12/2009

Parole Chiavi in Eccesso

Uso di parole chiave in eccesso

 [Stampa](#)

L'uso di parole chiave in eccesso è una pratica che consiste nel sovraccaricare una pagina web di parole chiave nel tentativo di modificare il posizionamento di un sito nei risultati di ricerca di Google. Riempire le pagine con parole chiave determina un peggioramento del servizio offerto agli utenti e può danneggiare il posizionamento del tuo sito. Cerca di creare contenuti utili e ricchi di informazioni che utilizzino le parole chiave in modo appropriato e nel contesto.

Per risolvere questo problema, esamina il tuo sito per individuare un eventuale utilizzo errato delle parole chiave. In genere, si tratta di elenchi o paragrafi di parole chiave, spesso ripetute a caso. Controlla attentamente, in quanto le parole chiave possono spesso presentarsi sotto forma di [testo nascosto](#) oppure essere nascoste nei tag "title" o negli attributi "alt".

Dopo avere apportato le modifiche desiderate ed esserti accertato che il tuo sito sia conforme alle nostre istruzioni, [invia il sito affinché venga riconsiderato](#).

aggiornato 5/12/2009

Contenuti Duplicati

Contenuti duplicati



Il termine "contenuti duplicati" si riferisce in genere a blocchi di contenuti importanti all'interno o tra i domini che sono identici o molto simili. In genere questa pratica non è all'origine ingannevole. Tra gli esempi di contenuti duplicati non dannosi vi sono:

- Forum di discussione che possono generare sia pagine regolari che pagine "abbreviate" destinate a dispositivi mobili
- Contenuti archiviati visualizzati o collegati tramite più URL distinti
- Versioni unicamente stampabili delle pagine web

Se il tuo sito contiene più pagine con contenuti pressoché identici, esistono vari modi per indicare il tuo URL preferito a Google (questa pratica è detta "canonicalizzazione"). [Ulteriori informazioni sulla canonicalizzazione.](#)

In alcuni casi, tuttavia, i contenuti vengono deliberatamente duplicati tra i domini nel tentativo di manipolare il posizionamento nei motori di ricerca o conquistare un traffico maggiore. Pratiche ingannevoli come questa possono peggiorare il servizio offerto agli utenti, i quali vedono sempre gli stessi contenuti ripetuti in un insieme di risultati di ricerca.

Google cerca di indicizzare e visualizzare pagine contenenti informazioni distinte. Ad esempio, se il tuo sito contiene una versione "regolare" e una "stampabile" di ciascun articolo e nessuna di esse è bloccata in robots.txt o con un metatag noindex, ne sceglieremo una sola. Nei rari casi in cui ci accorgiamo che i contenuti duplicati possono essere visualizzati con l'intento di manipolare il posizionamento e ingannare i nostri utenti, apportiamo le necessarie modifiche all'indicizzazione e al posizionamento dei siti interessati. Di conseguenza, il posizionamento dei siti può subire effetti negativi oppure il sito può essere rimosso completamente dall'indice di Google, in modo che non compaia più tra i risultati di ricerca.

Di seguito riportiamo le operazioni da eseguire per risolvere attivamente i problemi di contenuti duplicati ed essere sicuri che i visitatori visualizzino i contenuti a loro destinati.

Indicazioni di Google

- **Blocco ed esclusione delle pagine dall'indicizzazione:** anziché lasciare che siano gli algoritmi di Google a stabilire la versione "migliore" di un documento, puoi aiutarci a individuare la tua versione preferita. Ad esempio, se non desideri che sottoponiamo a indicizzazione le versioni stampabili degli articoli del tuo sito, disattiva tali directory o utilizza espressioni regolari nel tuo file [robots.txt](#).
- **Usa reindirizzamenti 301:** se hai ristrutturato il tuo sito, utilizza reindirizzamenti 301 ("indirizzamento permanente") nel file .htaccess per reindirizzare gli utenti, Googlebot e altri spider. In Apache, puoi eseguire questa operazione con un file .htaccess, mentre in IIS puoi utilizzare la console di amministrazione.
- **Mantieni la coerenza:** cerca di mantenere coerenti i link interni. Ad esempio, non creare link a <http://www.example.com/pagina/> e <http://www.example.com/pagina> e <http://www.example.com/pagina/index.htm>.
- **Utilizza domini di primo livello:** per consentirci di visualizzare la versione più appropriata di un documento, utilizza domini di primo livello, ove possibile, per gestire contenuti specifici per Paese. Siamo più propensi a pensare che il sito www.example.de presenti contenuti specifici sulla Germania rispetto a www.example.com/de o de.example.com.
- **Presta particolare attenzione alla diffusione di contenuti su altri siti:** se pubblichi i tuoi contenuti su altri siti, Google visualizzerà sempre la versione ritenuta più appropriata per gli utenti in ciascuna ricerca. Tale versione può essere o meno quella da te preferita. È utile accertarsi, tuttavia, che ciascun sito su cui i contenuti sono diffusi includa un link che rimanda al tuo articolo originale. Puoi anche chiedere a chi utilizza il tuo materiale diffuso di bloccare la versione che compare sul loro sito mediante il file robots.txt.
- **Utilizza gli Strumenti per i Webmaster per comunicarci in che modo desideri che il tuo sito venga indicizzato:** puoi comunicare a Google il tuo [dominio preferito](#) (ad esempio, www.example.com o <http://example.com>).
- **Riduci al minimo la ripetizione di testo standard:** ad esempio, anziché inserire un lungo testo sul copyright nella parte inferiore di ogni pagina, includi un breve riepilogo, quindi inserisci un link a una pagina contenente ulteriori dettagli.
- **Evita la pubblicazione di pagine incomplete:** gli utenti non amano visualizzare pagine "vuote"; se possibile, evita quindi l'utilizzo di segnaposto. Ad esempio, non pubblicare pagine per le quali non disponi ancora di contenuti effettivi. Se decidi comunque di creare pagine segnaposto, utilizza [robots.txt](#) per bloccarle ed evitare che vengano sottoposte a scansione.
- **Acquisisci familiarità con il tuo sistema di gestione dei contenuti:** accertati di conoscere bene le modalità di visualizzazione dei contenuti sul tuo sito web. Blog, forum e sistemi correlati mostrano spesso gli stessi contenuti in più formati. Ad esempio, una voce di blog può comparire con la stessa etichetta nella home page di un blog, in una pagina di archivio e in una pagina di altre voci.
- **Riduci al minimo i contenuti simili:** se disponi di molte pagine simili, prova a espandere ciascuna di esse o a consolidarle tutte in una sola pagina. Ad esempio, se il tuo sito di viaggi contiene pagine distinte per due città ma le informazioni sono le stesse su entrambe le pagine, potresti unire le due pagine in una sola pagina riguardante entrambe le città o espandere ciascuna di esse in modo che presenti contenuti unici su ciascuna città.

Contenuti Duplicati

La presenza di contenuti duplicati su un sito non giustifica l'adozione di misure particolari, a meno che l'intento non sia ingannevole e volto a manipolare i risultati dei motori di ricerca. Se sul tuo sito sono presenti contenuti duplicati e tu non segui i consigli sopra riportati, noi provvediamo a scegliere una versione dei contenuti da visualizzare nei nostri risultati di ricerca.

Tuttavia, se con la nostra verifica abbiamo rilevato l'uso di pratiche ingannevoli e il tuo sito è stato rimosso dai nostri risultati di ricerca, esamina attentamente il tuo sito. Se il tuo sito è stato rimosso dai nostri risultati di ricerca, consulta le [istruzioni per i webmaster](#) per ulteriori informazioni. Dopo avere apportato le modifiche desiderate ed esserti accertato che il tuo sito sia conforme alle nostre istruzioni, [invia il sito affinché venga riconsiderato](#).

Se riscontri che un altro sito duplica i tuoi contenuti appropriandosene indebitamente e ripubblicandoli, è improbabile che ciò influisca negativamente sul posizionamento del tuo sito nelle pagine dei risultati di ricerca di Google. Se rilevi un caso particolarmente evidente, non esitare a [presentare una richiesta DMCA](#) per rivendicare la proprietà dei contenuti e richiedere la rimozione dell'altro sito dall'indice di Google.

aggiornato 5/12/2009

Scarsa presenza di contenuti originali

Assenza o scarsa presenza di contenuti originali



Una delle fasi più importanti per migliorare il posizionamento del tuo sito nei risultati di ricerca di Google è accertarti che esso contenga grandi quantità di informazioni utili comprendenti parole chiave pertinenti e utilizzate in modo appropriato, che indicano l'argomento dei contenuti.

Alcuni webmaster, tuttavia, tentano di migliorare il posizionamento della propria pagina e attirare visitatori creando pagine contenenti molte parole ma contenuti autentici scarsi o assenti. Google prende provvedimenti nei confronti dei domini che tentano di ottenere posizioni migliori semplicemente visualizzando pagine di cui si sono indebitamente appropriati o pagine generate automaticamente, che non sono di alcun valore per gli utenti. Tra gli esempi vi sono:

- **Siti affiliati poveri di contenuti:** questi siti raccolgono entrate pay-per-click (PPC) inviando i visitatori ai siti di programmi affiliati, pur fornendo agli utenti contenuti o servizi caratterizzati da un valore aggiunto scarso o assente. Di norma, tali siti non presentano contenuti originali. Può trattarsi infatti di siti mirati alla soppressione dei cookie o di modelli privi di contenuti univoci.
- **Pagine doorway:** pagine create solo per i motori di ricerca.
- **Contenuti generati automaticamente:** contenuti generati a livello di programmazione. Si tratta spesso di paragrafi di testo casuali che non hanno alcun senso per il lettore ma che possono contenere parole chiave di ricerca.
- **Contenuti di altri siti:** alcuni webmaster utilizzano contenuti tratti da altri siti più affidabili, supponendo erroneamente che aumentare il volume di pagine web con contenuti casuali non pertinenti rappresenti una buona strategia a lungo termine. I contenuti di cui ci si appropria, anche se tratti da fonti di alta qualità, non forniscono alcun valore aggiunto agli utenti se non vengono utilizzati ulteriori servizi utili o contenuti forniti dal tuo sito. È opportuno creare contenuti originali che contribuiscano a distinguere il tuo sito da altri. In tal modo, i visitatori ritorneranno e i risultati di ricerca saranno di grande utilità.

Non è un problema se sei un affiliato, a condizione però che crei valore aggiunto per i tuoi utenti e offri loro contenuti validi che li inducano a visitare il tuo sito. Ad esempio, potresti creare recensioni, valutazioni e confronti di prodotti.

Se il tuo sito è stato rimosso dai nostri risultati di ricerca, consulta le [istruzioni per i webmaster](#) per ulteriori informazioni. Dopo avere apportato le modifiche desiderate ed esserti accertato che il tuo sito sia conforme alle nostre istruzioni, [invia il sito affinché venga riconsiderato](#).

Sito Compromesso

Il mio sito è stato compromesso

 [Stampa](#)

Vogliamo che i nostri utenti si sentano sicuri quando eseguono una ricerca sul Web e siamo costantemente al lavoro per identificare pagine pericolose e aumentare la protezione dei nostri utenti. Google utilizza algoritmi automatici e il feedback degli utenti per compilare elenchi di siti che possono risultare pericolosi. I due principali tipi di pagine pericolose nei nostri elenchi sono le pagine di phishing e le pagine di malware. Tali elenchi potrebbero non essere in grado di proteggere perfettamente gli utenti da tutti i siti rischiosi presenti sul Web. Inoltre, esiste sempre la possibilità che un sito sicuro possa essere identificato erroneamente come rischioso, tuttavia aggiorniamo gli elenchi regolarmente per tenere conto di eventuali cambiamenti.

Phishing	Ulteriori informazioni...
Malware	Ulteriori informazioni...
Richiesta di un controllo del tuo sito	Ulteriori informazioni...



Indicazioni di Google

Phishing

[torna all'inizio](#)

Il "phishing" o "spoofing" è un tipo di attacco in cui una pagina ti chiede dati personali o finanziari in modo ingannevole. Per ulteriori informazioni sul phishing, consulta il sito [Antiphishing.org](#).

Se ritieni che stiamo includendo nel nostro elenco una pagina che non è di phishing, inviaci una [segnalazione di falso phishing](#). Puoi anche segnalarci una pagina di phishing se ne individui una non inclusa nel nostro elenco.

Malware

[torna all'inizio](#)

Il "malware" è un tipo di software che tenta di carpire informazioni personali o di utilizzare il computer per eseguire operazioni da te non richieste o indesiderate. Le pagine di malware sono pagine web contenenti codice dannoso che potrebbe essere scaricato e installato sul tuo computer senza il tuo consenso. Per ulteriori informazioni sui contenuti web potenzialmente dannosi, tra cui virus e altri software dannosi, e su come proteggere il tuo computer, consulta il sito [StopBadware.org](#).

Per scoprire se il tuo sito è stato identificato come sito web che potrebbe ospitare o diffondere software dannoso (un tipo di "badware"), consulta la pagina "Panoramica" relativa a questo sito negli Strumenti per i Webmaster di Google. Per poter visualizzare queste informazioni, è necessario effettuare la verifica della proprietà del sito. [Altre informazioni sulla verifica del sito](#). Inoltre, inviamo messaggi di notifica ai webmaster dei siti interessati ai seguenti indirizzi email:

- abuse@
- admin@
- administrator@
- contact@
- info@
- postmaster@
- support@
- webmaster@

Questa identificazione si basa in parte sulle linee guida stabilite da StopBadware.org. Tuttavia, Google utilizza criteri, procedure e strumenti propri per identificare i siti che ospitano o diffondono badware. In alcuni casi, è possibile che il codice dannoso venga aggiunto a siti genuini da terzi, originando il messaggio di avviso di Google. Se ritieni che il tuo sito sia stato identificato erroneamente o se apporti le modifiche necessarie affinché non ospiti o non distribuisca più software dannoso impedendo che sia possibile inserire badware, puoi richiedere che il tuo sito venga riesaminato (vedi sotto).

Indicazioni di Google

Richiesta di un controllo del tuo sito

[torna all'inizio](#)

Dopo aver esaminato il tuo sito e rimosso tutti gli eventuali contenuti dannosi, puoi inviare una nuova **richiesta di esame**. Tieni presente che devi confermare la tua identità di proprietario del sito prima di poter richiedere che il sito venga riesaminato.

1. Nella home page degli Strumenti per i Webmaster, seleziona il sito desiderato.
2. Nel messaggio secondo cui **parti del sito potrebbero distribuire malware**, fai clic su **Altri dettagli**.
3. Fai clic su **Richiedi un controllo**.

Il tuo sito verrà sottoposto a un nuovo esame. Qualora venga appurato che il tuo sito non ospita o non diffonde badware, provvederemo a rimuovere l'identificazione dai risultati di ricerca.

Penalizzazioni

- **Penalizzazione:** comporta il declassamento del nostro sito agli occhi dei motori di ricerca, facendogli perdere posizioni nelle *SERP*.
- **Ban:** comporta l'esclusione da tutti i risultati del motore di ricerca del nostro sito. Non sempre è permanente, ma per lunghi periodi il sito non apparirà più in nessuna *SERP*.
- **Filtri:** il nostro sito non appare in alcuni risultati di ricerca; non abbiamo subito né un *ban*, né un vero e proprio declassamento da *penalizzazione*, ma le *SERP* filtrano il nostro sito per alcune chiavi.

Penalizzazioni algoritmica e manuale:

- Penalizzazione **algoritmica**: non è altro che il frutto di una perdita di ranking dovuta ad un aggiornamento dell'algoritmo. Dal 2012, tale update ha subito diversi rinnovamenti, fino ad arrivare a quello che è conosciuto come Fred.
- Non saremo avvisati da nessuno e tecnicamente è solo un cambio di rotta, non una punizione specifica per il tuo sito.

Penalizzazioni 2005 – WordPress

- Causa: contenuti di scarsa qualità
- Effetto: riduzione ranking / visibilità
- Durata: 2 giorni

Wordpress è il sistema di gestione dei contenuti web più noto al mondo che nel 2005, fu vittima del suo stesso successo. Una startup pubblicò più di 100.000 articoli su WordPress, ognuno dei quali con contenuti irrilevanti, sfruttando l'autorevolezza del dominio, allo scopo di generare guadagni.

Scrivi contenuti utili e approfonditi, non copiare, scrivi per i lettori umani, non per Google.

Penalizzazioni 2007 – Forbes, Washington Post e molti altri

- Causa: vendita di link
- Effetto : riduzione ranking / visibilità
- Durata: non definita, ma probabilmente breve

Nell'ottobre 2007, Google annunciò un aggiornamento del PageRank, e diversi siti di riviste autorevoli persero svariati punti di colpo. Tra questi figuravano siti del calibro di washingtonpost.com, Forbes.com, suntimes.com e molti altri meno popolari.

Le riviste online spesso guadagnano vendendo spazi pubblicitari. Bisogna ricordarsi però di aggiungere il nofollow a quei link.

Trustrank, Badrank, Topical Trustrank

Trustrank: una risorsa valida difficilmente linka un sito spam.

Badrank: Le pagine di spam linkano molte volte altre pagine di spam. E' facile quindi assegnare un punteggio negativo a queste pagine (vedi scambio link in automatico, posti dove inserire link facilmente, ecc...)

Topical Trustrank: link a risorse che non trattano lo stesso argomento della pagina linkante. Link poco inerenti e non a tema.

Algoritmi di Google

- **Panda** (Febbraio 2011): aveva come target contenuti duplicati e siti di bassa qualità
Top Heavy (Gennaio 2012): colpiva siti con eccessivi banner e annunci pubblicitari, che, si sa, vanno a discapito della user experience.
- **Penguin** (Aprile 2012): coinvolgeva siti con link building aggressiva, sovraottimizzazione di anchor text e altre pratiche.
- **Pirate** (Agosto 2012): limitava le posizioni di siti che avevano ricevuto vari DMCA (report di violazione copyright).
- **EMD** (Settembre 2012): preveniva che siti di bassa qualità raggiungessero alte posizioni solo per il fatto di avere parole chiave esatte nel nome del dominio.
- **Payday** (Giugno 2013): aveva ripulito i risultati di ricerca dalle “spammy queries” come viagra, pornografia e altre che erano state abusate finora.

Algoritmi di Google

- **Hummingbird** (Settembre 2013): aveva a che fare con la rilevanza di una query con il contenuto delle pagine del sito. Invece di applicare la corrispondenza solo ad una o più parole, la rilevanza era dell'intera query di ricerca.
- **Pigeon** (Luglio 2014): era un aggiornamento per migliorare le ricerche locali, introducendo nell'algoritmo il concetto di distanza territoriale.
- **Mobile Friendly** (Aprile 2015): era stato sviluppato per spingere le pagine ottimizzate per l'accessibilità da dispositivi mobili.
- **RankBrain** (Ottobre 2015): un'intelligenza artificiale era stata implementata per comprendere le relazioni tra parole e misurare la user experience.
- **Fred** (Marzo 2017): limitava i favoritismi a siti che prediligevano la monetizzazione rispetto all'user experience.

Penalizzazioni Manuali

- Penalizzazione **manuale**: accade quando il Google Webspam Team (quality rater) decide di analizzare il profilo backlink di un sito internet, e nel caso in cui si valuti che le Google Guidelines non siano rispettate. Tale penalizzazione può far retrocedere nelle pagine di risposta ma anche costare l'intera rimozione dagli indici di Google per i siti web segnalati.
- Te ne accorgerai perché riceverai una notifica nell'account Google Search Console con i dettagli del caso.
- Google fornire liste di casi sospetti ai propri incaricati, oltre a consentire ad utenti qualsiasi di inviare una notifica quando si imbattono in qualcuno che gioca sporco, magari un concorrente sleale.

Principali motivi di penalizzazione

Le situazioni più comuni, considerate da Google, sono:

- **Backlinks** non naturali da o verso il proprio sito (magari pagati e senza `rel="nofollow"`).
- Sito vittima di **hacking**: gli hacker potrebbero modificare le pagine o caricare intere cartelle di contenuti irrilevanti al solo scopo di danneggiarti.
- **Cloaking**: mostrare all'utente una versione di pagina diversa da quella fornita a Google.
- Contenuto di bassa qualità o **SPAM**: testi autogenerati o copiati altrove. **Keyword stuffing**: inserimento di keywords dappertutto per sovraottimizzare la pagina.

Strumenti di Google per controllare ban e penalizzazioni

- Google Search Console
- Google Analytics
- Google Trends e Google Insights for Search
(utili per capire l'andamento e stagionalità delle ricerche; il traffico del tuo sito rispetto a quello di altri siti web nel tuo stesso settore)

Penalizzazioni più diffuse

Sovraottimizzazione

Esempio:

- inserire una serie di parole chiavi ripetute e senza senso nell'h1
- utilizzare un footer per creare dei link fatti appositamente per spingere alcune pagine interne sui motori
- riempire tutti i tag con la nostra chiave (title, h1, alt, b, strong, i, em, keywords, description etc ...).
- Creare un menù ripetendo sempre la keyword da spingere
- Creare un testo innaturale super ottimizzato ripetendo sempre la nostra keyword

Over Optimization Penalty -950

Keyword stuffing e testo nascosto

- Con questa penalizzazione il sito finisce intorno agli ultimi 50 risultati mostrati da Google. In pratica intorno la 950° posizione.
- In passato era facile recuperare da questa penalizzazione: bastava ripulire tutte le pagine del sito web per riprendere in SERP la posizione occupata in origine dal sito.

Penalizzazione da Boilerplate

- Boilerplate è una penalizzazione algoritmica con possibilità di recupero pressoché immediata (24/48 ore).
- Boilerplate scatta quando ci sono elementi ripetuti e superflui tra le pagine web del tuo sito come ad esempio testi troppo concentrati e ripetuti in tutte le pagine o un eccesso di link ottimizzati nel footer.
- Boilerplate analizza anche altri elementi: la presenza di testi duplicati tra pagine web, il click through che le pagine ricevono nelle SERP, etc...

Penalizzazione da -30 a -80

All'inizio questa penalizzazione era inflitta manualmente.

Oggi è completamente automatica.

- Le posizioni perse nelle SERP possano variare da -30 a -80, di 10 posizioni in 10 posizioni, a seconda della gravità.
- Penalizzazione dovuta ad **anchor text dei link interni troppo ottimizzati**, eccesso di link concentrati nelle home page, link esterni verso siti spam.

Eccesso di Backlink

- backlink da pagine interne del tuo sito web verso la homepage.
- backlink provenienti da risorse esterne al tuo sito web tutti con anchor text identico e acquisiti troppo rapidamente.

Google è in grado di individuare (e penalizzare) pagine e siti web che ricevono backlink troppo “simili” tra loro.

Eccesso di Backlink

Se il tuo sito web parla di giochi online e i backlink che ricevi da risorse esterne riportano come anchor text:

- “giochi online”
- “giochi online flash”
- “giochi online di carte”

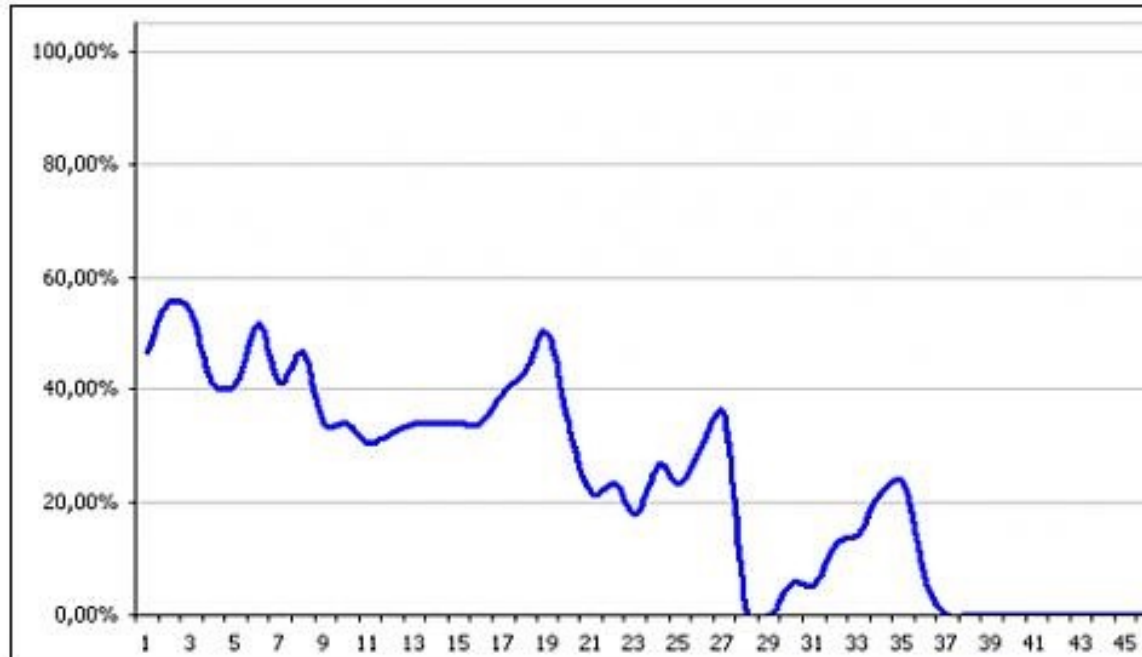
Cerca di variare, se possibile, i backlink in questo modo:

- “i migliori giochi online”
- “giochi online flash”
- “giochi di carte online”

Eccesso di Tag

- E' una penalizzazione da sovraottimizzazione di tag o pagine di ricerche indicizzate o di liste di contenuti troppo simili.
- Creano inutili contenuti duplicati anche in quantità elevata.
- Si esce da questa penalizzazione facendo deindicizzare queste pagine. Escludendo la cartella dei tag ai motori di ricerca attraverso il robots.txt o tramite l'indicazione noindex, follow (soluzione migliore).

Fluttuazioni Sinusoidali



Fluttuazioni Sinusoidali

- Dipendono dal punteggio che Google dà ad un sito per quanto riguarda il livello di spam.
- Se il punteggio è al limite dello spam per Google il sito inizia a “oscillare”, per un periodo è ben posizionato nelle SERP, per un altro periodo no.
- Quando il sito scende al di sotto della soglia minima di spam sparisce dalle SERP o arretra.
- Quando invece sale sopra la soglia, torna in SERP nelle posizioni originali o quasi.

Contenuti Duplicati

Sostanzialmente, un contenuto si dice duplicato quando compare più volte su URL diversi di blog, e-commerce o altri siti.

- **Motori di ricerca:** Google non sa quale versione debba includere nell'indice né quale mostrare agli utenti nei risultati di ricerca.
- **Webmasters:** a livello di link building, alcuni potrebbero linkare il contenuto su fonti non ufficiali, provocando una perdita di backlinks importanti.
- **Plagio:** un malintenzionato potrebbe forzare l'indicizzazione della propria copia di un contenuto altrui, prima che Google visiti la versione ufficiale, giovandosi degli sforzi del reale proprietario.

Contenuti Duplicati

- Circa il 29% del web, secondo Raven Tools, contiene contenuti duplicati.
- Google ha introdotto lo strumento del Digital Millennium Copyright Act (DMCA), il legittimo proprietario può richiedere la cancellazione, dall'indice del motore di ricerca, di tutte pagine copiate.
- Per sfatare alcuni miti, Google ufficialmente non penalizza i contenuti duplicati, perché ormai ha vari modi per riconoscerli, ma interviene quando verifica che qualcuno li utilizza per provare deliberatamente a manipolare i risultati di ricerca oppure per svolgere altre attività di questo tipo.

Contenuti Duplicati

Algoritmo molto potente in grado di individuare contenuto duplicato:

- solo prendendo alcune parti di un documento.
- per testi, immagini, audio e video.

I contenuti duplicati possono portare al ban del sito.

Contenuti Duplicati

Ci sono vari livelli di duplicazione dei contenuti:

- Quanta percentuale di un testo è stata duplicata?
- Sono state copiate solo alcune frasi o testi interi?
- La struttura del sito è stata anch'essa duplicata?

Contenuti Duplicati

- I contenuti ritenuti duplicati vengono filtrati (il sito non è né bannato né sotto l'effetto di una penalizzazione)
- Le pagine “duplicate” finiscono nei risultati supplementari.
- Quando la quantità di contenuti duplicati cresce il sito viene penalizzato o bannato.

Contenuti Duplicati

- Soluzioni?
 - Rielaborare i testi duplicati
 - Aumentare il trustrank e la linkpopularity
- Il nostro sito potrebbe essere sottoposto anche ad una penalizzazione semi-automatica, cioè che prevede sia l'uso degli algoritmi, sia l'intervento di un umano per la penalizzazione.
- Questo dipende dalla misura in cui il documento è duplicato, nel caso in cui un sito appaia totalmente il duplicato di un altro sito potrebbe subire anche il ban.

Contenuti Duplicati

- Google, quando si presentano contenuti duplicati in siti diversi predilige le pagine che presentano un TrustRank maggiore, spesso penalizzando le altre.
- Proviamo ad effettuare una ricerca:
“Silvia, rimembri ancora quel tempo della tua vita mortale” (A Silvia - Giacomo Leopardi)

Contenuti Duplicati

- Nelle SERP vengono riportati siti che presentano lo stesso risultato, ovviamente non originale trattandosi della celebre poesia, ma al quale il motore riconosce comunque una sua validità, in quanto la struttura di quei siti appare fortemente differenziata; pagine statiche, forum, blog, social, Google Libri.

Dove si trovano i contenuti duplicati?

- **URL equivalenti:** alcuni CMS possono generare URL diversi per il medesimo contenuto (es. `.it/articolo-1?p=true` vs `.it/articolo-1?p=false`)
- **Versioni del sito:** se hai una versione separata del sito che gira su `http` o `https`, `www` o senza `www`, può accadere che si generino vari URL per i medesimi contenuti.
- **Copie:** molte persone ancora copiano e pubblicano articoli altrui spacciandoli per propri.

Quali sono le soluzioni a queste situazioni?

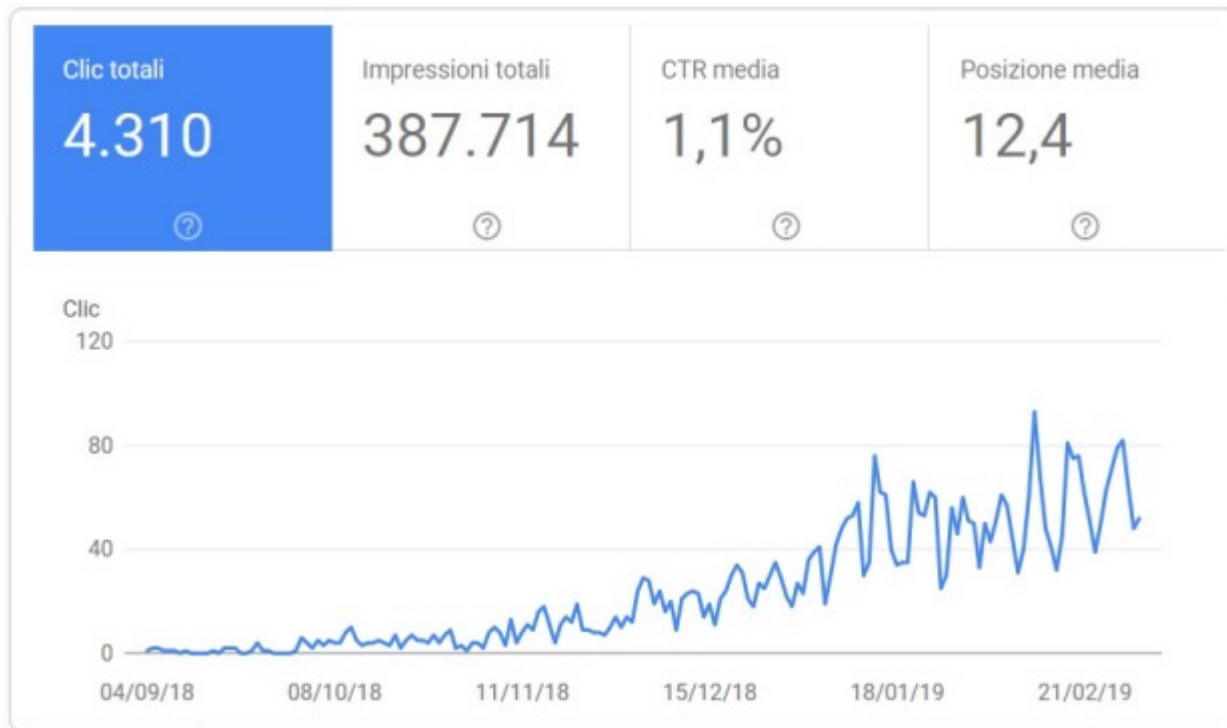
- **Canonicalizzazione:** nella sezione di tutte le pagine duplicate inserisci per indicare a Google qual è la versione ufficiale.
- **Redirect:** utilizza un reindirizzamento 301 per portare chiunque visiti una pagina duplicata verso quella corretta. Questa pratica vale anche per http vs https e www vs non-www.
- **Noindex:** usa il meta robots con attributo noindex sulle pagine duplicate per evitare che vengano indicizzate.
- E se sono copiato?
 - Copyscape.com individua le copie
 - Cerca manualmente su Google una frase tra virgolette

SandBox

- Periodo iniziale di un sito nei risultati di ricerca nel quale il posizionamento (ranking) non è perfettamente performante. Dopo un certo periodo di valutazione da parte degli spider di google la pagina acquisisce una certa rilevanza e posizionamento all'interno della SERP. Può durare 3-12 mesi.
- Google sottopone i nuovi siti ad un periodo di prova prima di farli comparire tra i risultati forniti all'utente dal Motore di Ricerca ma non sempre... solo ai siti con forte ottimizzazione, numero repentino di backlink, ... altre forzature. A volte si verificano fluttuazioni nelle SERP.
- La ragione di questa fase può anche essere condivisibile, perché consente ai nuovi siti di avere del tempo per ottenere autorità, attraverso contenuti e relazioni con altri siti del settore, senza essere premiati troppo facilmente.

SandBox

- Ecco un esempio dove si può vedere un notevole incremento di traffico dopo 4 mesi dal lancio, nonostante non sia stata fatta alcuna attività di link building né di pubblicazione contenutistica



SandBox - Fattori

- **Nicchia di mercato:** alcuni settori sono più sensibili, perché facile bersaglio di marketer o spammer agguerriti (ad esempio, guadagnare online) oppure perché trattano argomenti delicati (ad esempio, medicina).
- **Aggressività:** se già dal primo giorno inserisci banner pubblicitari, call to action e spingi molto sulla monetizzazione.
- **Contenuti scarsi:** un blog con pochi articoli e di scarso contenuto difficilmente sarà preso in considerazione da subito.
- **Qualità dei links:** non limitarti ad avere solo links dai social, ma cerca anche fonti autorevoli con l'uso del guest blogging o altre metodiche di link building.

Nel frattempo, Google raccoglie le metriche sulla user experience per «accreditarci»

Server: dove ospitare il sito?

- **Tipo di applicativo:** se si tratta di una applicazione o di qualche script che utilizza Java, probabilmente dovrai acquistare un hosting con particolari risorse e librerie a disposizione. In caso di utilizzo di CMS WordPress like, va bene una soluzione standard con hosting Linux.
- **Garanzie di uptime:** sito deve essere sempre accessibile, altrimenti il tuo pubblico potrebbe non visitarti più e andarsene.
- **Pannello di controllo dell'hosting:** hosting che offrono cPanel oppure Plesk come pannelli di controllo perché sicuri, semplici da usare e hanno tutte le funzioni di cui un classico sito WordPress potrebbe aver bisogno, come antivirus, backup automatici e molto altro.
- **Hardware dedicato:** spazio disco e banda a disposizione. Meglio host con hard disk SSD. Numero di Gigabytes a disposizione.

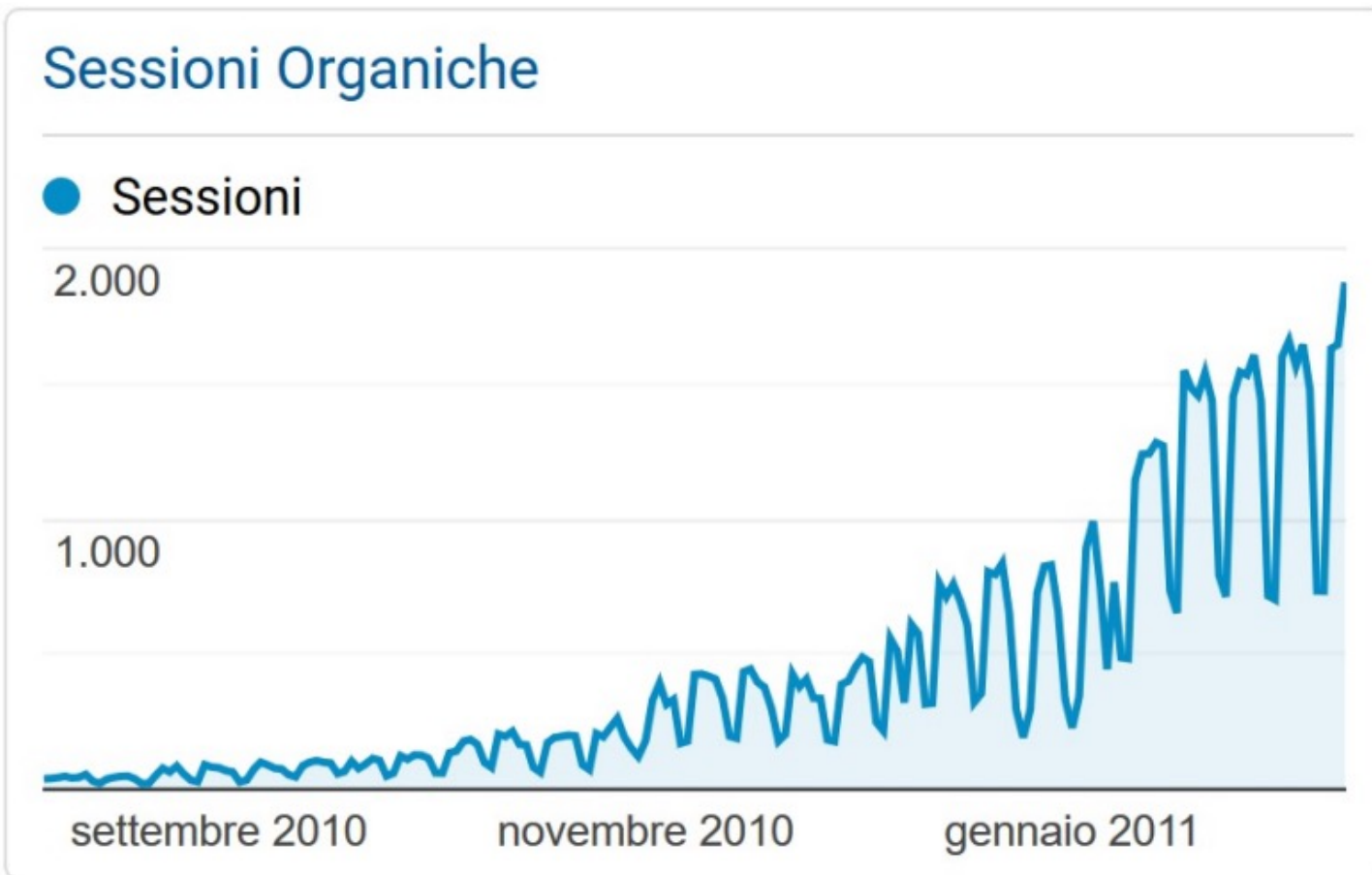
Server: dove ospitare il sito?

- **Banda:** dipende da quanto traffico riceverai e da quanto pesano i files che i tuoi utenti scaricano. In base al tipo di applicativo, dovrai anche considerare RAM e CPU per mantenere performance elevate.
- **Certificato SSL/TLS:** molti hosting ormai includono il certificato SSL gratuito di Let's Encrypt per usare il sito con maggiore sicurezza, grazie al protocollo HTTPS, invece del classico HTTP.
- **Backup automatici:** Il backup periodico è una copia di sicurezza di tutti i files, del database e delle impostazioni del sito. Alcuni hosting automaticamente effettuano i backup del tuo sito a intervalli regolari, in modo che, se dovesse avere malfunzionamenti oppure subire un attacco hacker, ti basterà contattare il supporto e verrà tutto ripristinato gratuitamente ed esattamente allo stato in cui si presentava qualche giorno prima.

Tipi di hosting

- **Hosting Condiviso:** Ideale per iniziare un nuovo progetto. L'hosting condiviso è semplicemente un server utilizzato da decine di clienti diversi. Vantaggio è il risparmio. Svantaggio è che spesso non si hanno risorse dedicate e, in momenti di picchi generati da altri siti, il tuo sito ne risentirà in termini di minori risorse a disposizione per far navigare i tuoi utenti.
- **VPS o Virtual private server:** Ideale per siti con traffico. Sono virtual machine montate su un server dedicato. Con questa configurazione non ci saranno decine o centinaia di clienti ad usare il medesimo server, ma solo alcuni. Risorse dedicate, come la RAM o la CPU, che nessun altro "vicino" potrà occupare nei momenti di picco.
- **Server dedicato:** Ideale per siti con traffico e professionali. Il grande vantaggio è la possibilità di avere un IP dedicato. Con il server esclusivo e dedicato, prezzi alti, ma carichi di traffico elevati con alte velocità di navigazione.

2011, passando dall'hosting condiviso al server dedicato



Server

- Tempo di risposta del server
- Percentuale di tempo che il sito appare down (sito non raggiungibile)

Fattori che impediscono ai motori di ricerca di arrivare sul sito.

Questo fattore può portare alla de-indicizzazione del sito web.

Speed test

- <https://pagespeed.web.dev/?hl=IT>

PageSpeed Insights

<https://www.scenarioabruzzo.it/> Analizza

Dispositivi mobili Desktop

Scopri com'è l'esperienza dei tuoi utenti reali Nessun dato


Diagnostica i problemi di prestazioni

53

Prestazioni

I valori sono delle stime e potrebbero variare. Il punteggio relativo alle prestazioni viene calcolato direttamente in base a queste metriche. Vai al calcolatore.

▲ 0-49 ■ 50-89 ● 90-100



METRICHE Espandi visualizzazione

▲ First Contentful Paint 3,9 s	■ Time to Interactive 5,4 s
▲ Speed Index 11,0 s	■ Total Blocking Time 270 ms
▲ Largest Contentful Paint 6,3 s	● Cumulative Layout Shift 0,022

Captured at 7 luq 2022, 16:49 CEST Moto G4 emulato with Lighthouse 9.6.2 Caricamento singola pagina

Case study

- <http://www.giorgiotave.it/forum/consigli-seo-e-penalizzazioni/101393-sito-per-bambini-penalizzato.html>

Pagine duplicate e inutili

Case study

- <http://blog.imevolution.it/134/penalizzazione-per-troppi-link-in-entrata-ottenuti-rapidamente-ecco-come-uscirne/>
Troppi link in entrata ottenuti troppo velocemente

Case study

- <http://www.giorgiotave.it/forum/consigli-seo-e-penalizzazioni/98723-da-130mila-3mila-visitatori-al-mese-da-google-aiuto.html>
Contenuti Duplicati, Testo scarso negli articoli

Case study

- <http://www.giorgiotave.it/forum/consigli-seo-e-penalizzazioni/112235-penalizzazione-siti-eliminati-definitivamente-da-google.html> Sovraottimizzazione, Network di Siti, Stesso Codice AdSense -> Mai più reintegrati

Case study

- <http://www.wmtools.com/news/posizionamento-motori/questo-sito-e-penalizzato-ecco-perche-ed-ecco-il-vincitore>
Perdita PageRank

Case study

Siti non penalizzati per Contenuti Duplicati

<http://www.ultimecalciomercato.com>

<http://www.donazione sangue.altervista.org>

<http://www.pharmacies-online.com>

Sito Bannato per contenuti duplicati (ora recuperato)

Case study

- <http://www.giorgiotave.it/forum/consigli-seo-e-penalizzazioni/106777-aiutatemi-penalizzazione-google-improvvisa.html>
- Poco contenuto, Contenuti Duplicati, Link a siti di casinò e non tematici

Case study

- <http://blog.imevolution.it/166/ban-di-un-sito-la-causa-possono-essere-i-dns/>
 - Irraggiungibilità del sito. DNS?

Case study

- <http://www.giorgiotave.it/forum/consigli-seo-e-penalizzazioni/106873-penalizzazione-su-google.html>
- Keywords Stuffing, Testo nascosto, Doorway Page

Case study

- <http://www.giorgiotave.it/forum/consigli-seo-e-penalizzazioni/108204-perdita-posizioni.html>
Cambio Sito senza Redirect 301

Case study

- http://www.masternewmedia.org/it/online_marketing/penalizzazione-di-Google/possibili-cause-penalizzazione-di-Google-20070904.htm
Testo nascosto, duplicazione di pagine e di siti, spam nei commenti, text links non a tema

Case study

- <http://www.giorgiotave.it/forum/consigli-seo-e-penalizzazioni/85523-presunta-penalizzazione-dopo-redirect-e-cambio-hosting.html>
Homepage raggiungibile con url diversi

Case study

- <http://www.giorgiotave.it/forum/consigli-seo-e-penalizzazioni/98618-penalizzazione-sito.html>
Megamenù

Case study

- <http://www.giorgiotave.it/forum/consigli-seo-e-penalizzazioni/104791-sito-fase-di-penalizzazione.html>
Tag Excess

Case Study: Google MayDay

- <http://www.webster.it>
- E-commerce in genere

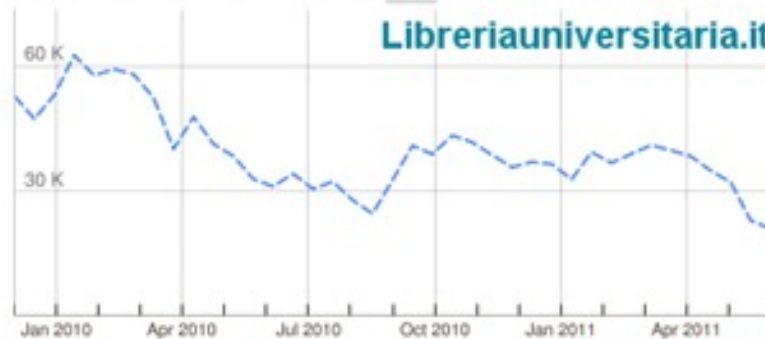
Case Study: Google MayDay

Grafici dal 1 Dicembre 2009 al 14 Maggio 2011

Libreriadelsanto.it



Libreriauniversitaria.it



Webster.it



Siamo penalizzati e ora?

Analisi sito web

- Assicurati che nelle pagine web del tuo sito non ci sia né testo nascosto né parole chiave ripetute troppo (**keyword stuffing**)
- Controlla che la **struttura di navigazione** del sito sia semplice, permetta di raggiungere tutte le pagine in pochi click, linking interno strutturato in modo da linkare contenuti correlati,...
- Verifica che non ci siano link rotti né verso le tue pagine interne né verso l'esterno

Analisi sito web

- Controlla la **qualità delle risorse che linki** al di fuori del tuo sito: è tua responsabilità non spedire gli utenti verso siti spam o near-to-spam
- Verifica i **backlink** che ricevi
- Controlla le istruzioni del file **robots.txt**
- Utilizza **pagine di errore 404 personalizzate**

Analisi sito web

- Se usi una o più [sitemap](#), verifica che non contengano pagine inesistenti o errori sugli url
- Naviga il tuo sito web disattivando javascript, cookies e fogli di stile CSS
- Verifica la [velocità del sito](#)
- Verifica che nella prima parte del sito non ci sia più pubblicità che contenuti

Analisi sito web

- Verifica i contenuti eliminando dall'indicizzazione inutili pagine duplicate, liste di dati ripetuti, articoli copiati ad altri siti
- Verifica che altri siti non copino i tuoi contenuti. In questo caso intervieni subito contattando i siti e facendo eliminare i tuoi contenuti
- Attento a chi ripubblica i tuoi contenuti tramite RSS