

Canonical URL e contenuti duplicati

di Roberto Ciri (SEO Specialist 1mas) – 31/07/2009

Un nuovo standard concordato fra i tre principali motori di ricerca Google, Msn e Yahoo permette di eliminare il problema della duplicazione i contenuti fra siti web ed indicare quale è la pagina principale da cui proviene un contenuto originario.

Google, Msn e Yahoo si sono accordati su uno standard relativamente al tag "canonical", un formato che permette di specificare pubblicamente quale è la versione principale ed originaria di un URL (indirizzo di una pagina web). Ciò permetterà ai webmasters la possibilità, e responsabilità, di scegliere quali contenuti proposti ai motori di ricerca vadano considerati come principali e di scartare quindi i duplicati.

Se nel proprio sito web ci sono contenuti identici o molto simili raggiungibili tramite indirizzi differenti delle pagine (URL), questo tag permette di **assegnare una priorità alla pagina che presenta il contenuto originale**.

La presenza di questo nuovo tag permette quindi di ridurre i problemi di indicizzazione legati alle pagine ripetute e ai contenuti duplicati, permettendo di avere un maggior controllo sugli indirizzi delle pagine riportate nei risultati di ricerca. Di conseguenza aiuta anche ad assicurare il consolidamento della link popularity di una pagina.

Riassumiamo di seguito in cosa consiste il problema dei contenuti duplicati, rimandando l'approfondimento dell'argomento ad un prossimo articolo su "Google e i contenuti duplicati" e vediamo come il tag "canonical" lo risolve.

Cosa sono i contenuti duplicati?

Il termine "contenuti duplicati" si riferisce a corposi contenuti testuali (e non) all'interno di uno stesso dominio o tra domini diversi che sono identici o molto simili con altri contenuti.

In genere questa pratica non ha un'origine intenzionale o ingannevole verso i motori di ricerca: è il caso di portali che mostrano i contenuti di altri siti o linkano altre pagine, di blog che riportano porzioni di testo di altri siti come citazione, di versioni stampabili delle pagine web o di forum che generano pagine dinamiche prelevate da altri siti. Ad esempio non viene considerato come contenuto duplicato uno stesso articolo scritto in lingue diverse.

In altri casi i contenuti duplicati vengono creati con l'intento di agire sul posizionamento nei motori di ricerca o aumentare il traffico per parole chiave meno popolari. Il risultato è che gli utenti del motore di ricerca vedono sempre gli stessi contenuti ripetuti in risultati di ricerca differenti, peggiorando quindi il servizio offerto dal motore di ricerca stesso.

Google come gestisce i contenuti duplicati?

Quando l'algoritmo di Google analizza un sito web, prova a indicizzare e mostrare nei risultati di ricerca pagine con diversi contenuti. Pertanto se ad esempio nel sito sono presenti degli articoli in versione "pagina visualizzata" e "da stampare" e non c'è nessuna limitazione nel file robots.txt o attraverso il meta tag noindex, Google sceglie solo una delle due versioni da mostrare.

Nei casi in cui Google presume che il contenuto duplicato vuole essere mostrato per agire sul ranking di una pagina o di un sito (e di conseguenza sul posizionamento) al fine di deviare gli utenti del motore di ricerca, allora può effettuare una **penalizzazione sul posizionamento** dell'intero sito coinvolto o addirittura rimuovere completamente il sito dal proprio indice in modo che non compaia più tra i risultati di ricerca.

Ciò comporta che la pagina che viene meglio posizionata può essere una diversa rispetto a quella che il webmaster malizioso intendeva "spingere".

Come Google riconosce i contenuti originali tramite il tag "canonical"?

Entra in gioco quindi il tag "canonical" che permette di comunicare al motore di ricerca quale è l'indirizzo in cui si trova la pagina originale che presenta un determinato contenuto.

Tale tag va inserito all'interno della sezione "head" di una pagina, tramite il seguente codice HTML:

```
<link rel="canonical" href="http://www.nomesito.it/nomepagina.php?id=xx" />
```

Tale indicazione suggerisce al motore di utilizzare la pagina "nomepagina.php?id=xx" come pagina principale ed originaria nel caso in cui vengano trovate pagine simili ripetute.

Nel caso in cui il motore in fase di indicizzazione non incontri una pagina impostata come "canonical", ne definirà una secondo gli algoritmi normalmente utilizzati.

Consideriamo il seguente esempio di un ipotetico sito che vende cartucce per stampanti, in cui la pagina originale di un determinato modello di cartuccia abbia come indirizzo (URL):

```
www.cartucceperlatuastampante.com/prodotto.php?cod=0
```

I visitatori e Google possono giungere a tale pagina anche attraverso altri indirizzi, ad esempio legati ad una ricerca effettuata all'interno del sito stesso o a parametri collegati all'ID di sessione o di tracking della navigazione, come:

www.cartucceperlatuastampante.com/prodotto.php?cod=0&marca=canon

oppure:

www.cartucceperlatuastampante.com/prodotto.php?cod=0&sessionid=123&trackingid=456

Anche se le informazioni chiave presentate nelle pagine che rispondono a queste URL sono le stesse della versione principale, esse potrebbero avere delle piccole variazioni di contenuto legate a informazioni sul prodotto estratte in base alla marca o al riconoscimento dell'utente o del suo percorso di navigazione.

Tramite il tag "canonical" si può **specificare quale è la versione originale della pagina**, inserendo il codice:

```
<link rel="canonical" href="http://www.cartucceperlatuastampante.com/prodotto.php?cod=0" />
```

all'interno della sezione <head> delle URL con i contenuti duplicati.

In tal modo Google capirà che i contenuti duplicati si riferiscono tutti all'URL canonico "http://www.cartucceperlatuastampante.com/prodotto.php?cod=0" e trasferirà di conseguenza altre proprietà dell'URL come il PageRank.

In ultima analisi il tag "canonical" rappresenta una indicazione che si inserisce nel codice HTML di comunicazione tra sito web e motore di ricerca definendo l'ordine di importanza da seguire nel posizionamento di una pagina nei risultati di ricerca.

L'utilizzo del tag "canonical" come suggerimento per il motore di ricerca permette quindi al webmaster di guidare un utente che sta effettuando una ricerca verso la pagina che potrebbe essere più interessante.

Inoltre in questo modo si alleggerisce il motore di ricerca della presenza di pagine duplicate nei suoi indici, aumentando la qualità ed utilità dei risultati forniti.

L'adozione del riconoscimento del tag "canonical" da parte dei principali motori di ricerca, sebbene sia uno standard non del tutto ufficiale, presto determinerà dei cambiamenti nelle SERP (risultati di ricerca naturali), man mano che i vari motori

inizieranno ad assorbire le indicazioni presenti nei tag canonical inseriti nel codice delle pagine indicizzate.

Ciò determinerà anche delle nuove valutazioni da fare da parte degli specialisti SEO per quanto riguarda il posizionamento nei motori di ricerca.

1mas srl

www.1mas.it

V.le S. Eufemia, 34

25135 Brescia (BS)

+39 030 369 6400

Tutte le informazioni ed immagini presenti in questo articolo potranno essere utilizzate esclusivamente per scopi personali e non commerciali. L'adattamento, diffusione e pubblicazione, sotto qualsiasi forma, anche parziale, dei contenuti di questo articolo è ammessa solo se esplicitamente autorizzata da 1mas srl. In questo caso le informazioni divulgate devono obbligatoriamente essere accompagnate dalla dicitura "© 1mas srl " e dall'indicazione esplicita della URL del sito da cui sono state tratte "www.1mas-posizionamento-motore-ricerca.it". Per qualsiasi richiesta d'autorizzazione di riproduzione e diffusione rivolgersi per e-mail a 1mas srl - seo@1mas.it

Canonical URL e contenuti duplicati

di Roberto Ciri - 31/07/2009 - www.1mas-posizionamento-motore-ricerca.it