

REAL SEO

検索エンジンの視点から～ページ更新と差分

差分から得られる、増加分をYahooは過大評価する？ページ単位でインデックスを考えてみる。

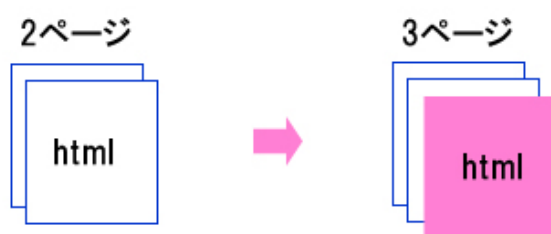
『ブログを更新した』

このような言い方をしますが、正確には、

『ブログのページを追加した』

でしょう。

(多くのブログシステムの場合、トップページは更新となるが)



『更新頻度の高いサイトは検索エンジンのクローラーの巡回頻度が上がる』

この場合も概ね、ページの増加を指す場合が多いようです。

今回はページ単位での更新を、検索エンジンがどのように見ているか、考えていきたいと思います。

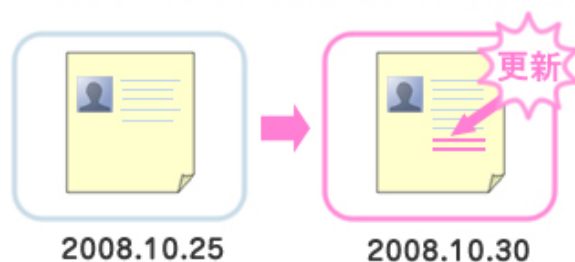
差分～ファイルを比較して、更新を抜き出す～

検索エンジンは、ページごとにインデックスをし、その内容を記録しています。

この時、新規のページ（初回のインデックス）でない場合、過去のインデックスと比較していると考えられます。

過去の情報と、現在の情報を比較して、『追加されたもの、削除されたもの』を判別して記憶していくようです。

差分：2つのファイルを比較して更新部分を抜き出す



Wikipediaで『差分』と調べてみると、

差分を調べることで、更新前と更新後とのファイルの差異を調べることができ、前の状態に戻すことや、差分からパッチを作成することもできる。

例えば、

[old.txt](#)

[new.txt](#)

という2つのファイルがあったとします。

この2つのファイルの情報の増減を視覚化すると、

<http://webapi.sakura.ne.jp/tex/index.php>

青の打ち消し線が、削除された情報で、赤の下線が増加した情報です。

検索エンジンは、このような表記はしませんが、これに似た機能を実装しているようです。

稀にですが、過去の情報を現在のインデックスに引用することがあります。

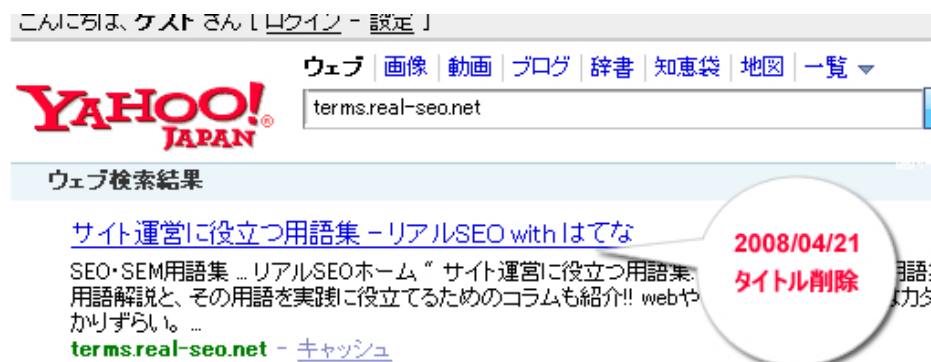
このサイトの『用語集ページ』 (<http://terms.real-seo.net/>) はタイトルを取ってはずしています。

過去 (2008/4/21) に『[タイトルを消すとどうなる??](#)』という実験をしました。

『Hタグ』から、タイトル (検索結果表示上の) を引っ張ってきたりしていました。

([タイトルを消すとどうなる??Vo.4](#))

そして、現在yahooの検索結果 (URLで検索した場合) では、下記のように、4/21の時点でのタイトルに戻しています。



すでに半年経過していますが、過去の履歴から情報を抜き出しているのです。

Yahoo!は少なくとも半年以上はインデックスを保管するようです。

(*用語の使い方が間違っているかもしれませんが、あしからず)

差分から得られる、増加分をYahoo!は過大評価する？

検索エンジンからすると、そのサイトのどのページにどれくらいの情報追加や削除が行われたかは、インデックス

されていれば、一目瞭然のはずです。

検索エンジン側は、より効率を上げるため、更新の多いページヘクローラーを優先的に廻しています。

特に、Yahoo!はその差分から得られる『増加』をかなり評価するようです。

(ブログのコメントや口コミ掲示板など)

ページを増大させるという方法は、かなり多く行われているようですが、増大させた各ページの内容を成長させるシステムを導入している方はまだ少ないような気がします。

これからWEBビジネスに参入する人は、この辺りの設計を十分に検討すると良いのではないかと思います。

先日来、『はてな』や『食べログ』の事例をしつこいほど紹介していますが、そうした『仕組み』を導入した優秀なサイトだと考えているからです。

参考記事

[はてなのSEO戦略Vo.2](#)

[比較サイトの王様のSEO戦略Vo.2](#)

印刷用：[検索エンジンの視点から～ページ更新と差分【PDF】](#)