

検索エンジンにおける表示順位監視システムの試作

鳥井慎太郎（沼田一道教授，松浦隆文助教）

1. はじめに

現代社会においてインターネット上に存在する広範な情報（ウェブページ、ウェブサイト、画像ファイル、ネットニュースなど）に能率よくアクセスすることは必須となっている。インターネットユーザーの多くは必要とする（興味ある）情報の存在場所（URL）を知るために Yahoo! や Google などの検索エンジンを使用するのが普通である。従って、サイト運営者は自分のサイトがこの検索エンジンの検索対象となることを期待する。

検索エンジンでキーワードを検索すると、膨大な数のホームページがヒットする。しかし、検索エンジンユーザーの70%は検索結果の2ページ目（上位20位）までしか見ていないといわれている。つまり、サイト運営者が自分のサイトを見てもらうためには、検索結果の上位20位以内に表示されることを目標とすべきである。そのため検索結果の表示順上位を上げる様々な方法が考えられており、それらは総称して SEO（Search Engine Optimization）と呼ばれている。

自分のウェブサイトに SEO を施して検索結果表示順位の向上を図っているサイト運営者は、当然のことながら、その成果を確認したくなる。しかし、自分で通常の検索を行って確認しようとしても、そのままだはうまくいかない。検索エンジンにはパーソナライズ検索という機能が存在し、ユーザーの閲覧履歴等に基づいて検索をチューニングしているため、正確な順位を反映していない可能性があるからである。検索結果で上位に表示されていたとしても他のユーザーの検索結果と違っている可能性があり SEO の効果を正しく確認することができない。

従って、どの SEO 施策が有効であったかを知るためには、パーソナライズ機能を無効にした検索順位の変動を記録することが必要となる。しかし、パーソナル検索を無効にすると通常の使用から離れた機械的な作業となり、日々の検索順位をチェックし記録し続けるのは負担が大きい。そこで、本研究ではパーソナライズ機能を無効にした検索エンジン検索結果の表示順位を自動的に取得し、記録、通知する検索エンジン監視記録システムをウェブシステムとして試作する。本システムには簡単なアクセス解析機能も持たせる。

2. システム

設計開発するシステムはユーザー認証を行うことにより、多数のユーザーで利用を可能としている。ユーザーが設定したキーワードによる各検索エンジン（Yahoo, Google, Bing）の検索結果（表示順位）を取得し、メールでユーザーへ通知する。また日々の結果を記録し一月単位で閲覧可能とする。

本システムを利用するメリットは以下の通りである。

- ・ パーソナライズ検索を無効にすることで正確な検索結果の表示順位を取得することが可能
- ・ 自動的に記録するためサイト運営者の機械的な作業をなくし負担を減らす
- ・ 検索結果の表示順位から SEO の効果を把握することが可能

・インストール不要でインターネットに接続したブラウザでどこからでも利用が可能

<システムの実体>

ウェブシステムであり作成者の用意したサーバー上で動作をさせる。サーバーの OS は Linux, PHP5 バージョンは 5.3.3, MySQL5 バージョンは 5.0.77 という環境である。検索結果表示順位の取得と記録、メールでの通知は指定時間に指定コマンド（プログラム）を自動的に実行させる Unix 系 OS の機能である「Cron」を使っている。ユーザーはインターネットに接続したブラウザからユーザー認証を行い利用する。詳しいインタフェースは次節に記述する。

<データベース設計>

データを保持するためにデータベースにテーブルを用意する。ユーザーのデータを保持するユーザーテーブル、登録されたサイトのデータを保持するサイトテーブル、どのサイトが何のキーワードを設定しているのかを保持するキーワードテーブル、検索結果の表示順位を保持するランキングテーブル、アクセス情報を保持するアクセステーブルの5つである（図1）。

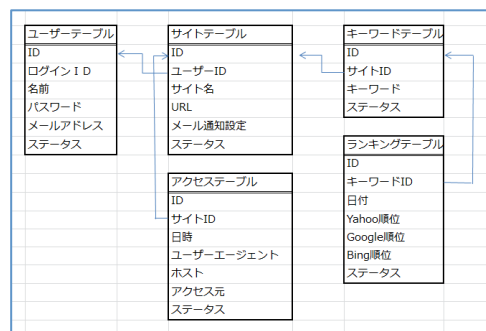


図1：DB リレーション図

<検索結果の表示順位の取得と記録>

「Cron」によって毎日 00 時 00 分にプログラムを実行のコマンドが実行され、各検索エンジン(Yahoo, Google, Bing)でのパーソナライズ検索無効の検索結果表示順位を取得、記録する。

以下文中の ①～⑤ は、図2中の番号に対応する。

- ① キーワードテーブルとサイトテーブルをステータスが有効（削除されていない状態）を条件に検索する。
- ② データベースから返された検索結果からキーワードとサイトの URL を保持する。

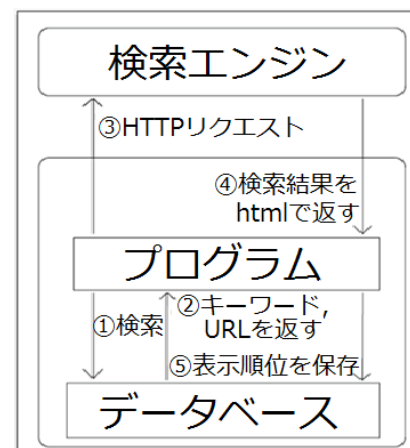


図2：検索結果表示順位取得と記録

パーソナライズ検索機能を無効にするパラメータ「&pws=0」、文字コード UTF-8 に URL エンコード (URL において使用できない文字を使う際に行われるエンコード) をしたキーワードを「?q=キーワード」という形で、ベースとなる URL の後ろにつけることで問い合わせ先の URL となる。Google の場合の URL は「http://www.google.co.jp/search?q=キーワード&pws=0」という形である。

- ③ 問い合わせ先の URL へ HTTP リクエストを送信する。
- ④ リクエストに対して検索エンジンから返ってきた html を文字列として保持。

1 ページに検索結果が 10 件表示されるので③、④の処理を自分のサイトを探す順位の最大値÷10 回繰り返す。2 回目以降の繰り返しでは問い合わせ先 URL へ「&start= (10×現在の繰り返し回数)」を付加する。

そして繰り返し得られた html を連結し 1 つの文字列として扱う。

図 4 は検索エンジン (Google) から取得した html の一部分、図 3 がブラウザ上で見た検索結果であり、反転している部分が図 3 の html で <cite> から </site> までにあたる。html の文字列の中から「<div class="f kv">」から「</cite>」の間を全て正規表現で抽出し配列に入れる。

そして、取得して配列に入れた文字列の中にはサイトの URL が入っているが図 3 のように「seo」と検索された文字列を含んでいた場合に太字にするタグが URL の間に含まれてしまう。そのため「」、」という部分文字列を削除し、URL の入った配列をループさせ自分のサイトの URL を発見した際に添字を返す。返された添字に 1 を足した物が検索結果の表示順位である。



図 3 : Google 検索結果 ("SEO"で検索)



図 4 : Google 検索結果の HTML ("SEO"で検索)

⑤ データベースに検索エンジンでの表示順位、日付、サイト・キーワードを識別する ID を保存。

<アクセス解析 >

準備として解析対象のサイトの HTML にシステム上で生成された JavaScript を貼付ける。解析対象サイトにアクセスがあった際に張り付けた JavaScript から、本システムがアクセス情報を受け取りデータベースへ保存する。受け取るアクセス情報は日時、アクセスされたページの URL、アクセスしたユーザーの情報 (ホスト名、IP アドレス利用中のブラウザ、デバイスの名称)、解析対象のサイトへのアクセス元 URL (アクセス元が検索エンジンであった場合には検索されたキーワードが含まれる) である。

3. ユーザインタフェース

ユーザーはインターネットに接続したブラウザからユーザー認証をすることで利用することが可能となる。検索結果の表示順位を自動取得させるためには、準備としてシステム上で自分のウェブサイトの URL とキーワードを登録する必要がある。

<ユーザー認証>

ログインフォームから入力されたログイン ID とパスワードでユーザーテーブルを検索する。一致する行があればセッションにユーザー情報を保持しサイト一覧画面へ遷移し、一致する行が見つからない場合エラーメッセージを表示する。ログアウトは保持しているユーザー情報を破棄しログイン画面へ遷移する。

<サイト、キーワードの管理>

登録されたサイトを一覧で出力する。一覧から登録、編集、削除、詳細表示へ遷移することができる。登録、編集はフ

ホームから行う。管理可能な項目はサイト名、サイト URL、メール通知設定、キーワードの4つであり、1つのサイトに対して複数のキーワードを登録することが可能である。

<データの出力>

キーワードごとの検索結果の表示順位とアクセス解析結果を月別で出力。ページャーがあり前の月と次の月の結果へ飛ぶ事ができる。各検索エンジンでの表示順位を表と折れ線グラフ(図5)、アクセス解析結果はアクセス数を表と棒グラフ、アクセス元と検索されたキーワードを一覧、ユーザーのデバイスとブラウザを円グラフで出力する。

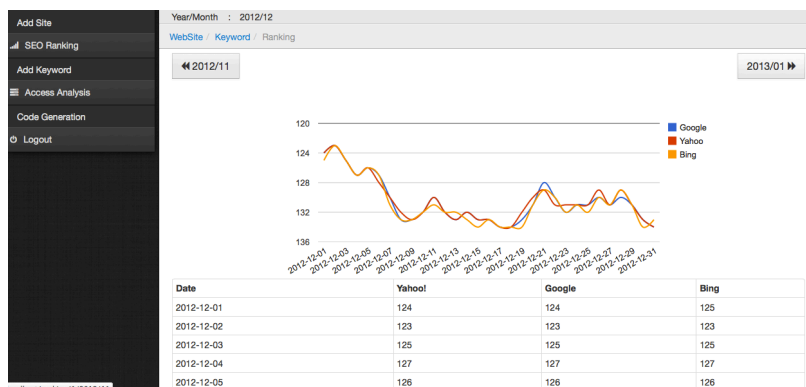


図5：表示順位の月別出力画面

<メール通知>

サイトのメール通知設定が有効になっている場合に、表示順位取得と同じ時刻(00時00分)にCronで実行される。サイト名と前日の各検索エンジンでの表示順位をユーザーへメールで送信する。

4. 評価

サイトを運営する4名に本システムを1ヶ月間利用してもらい評価をしていただいた。それぞれ5個のサイトをシステムに登録し、1つのサイトにつき2~4個のキーワードを設定した。

「毎日のデータを確認する事が可能なことによって、行っているSEOの効果を把握することができた」、「たくさんあるサイトの管理の負担が減ったので今後も利用し続けたい」という評価を全員からいただくことができた。また、機能の改善として「毎日のメール通知の間隔を1週間ごと、1ヶ月ごとなどにまとめて結果を送られるように設定をしたい」、「メールの通知先を複数設定したい」、「検索結果の表示順位を月別だけではなく期間を設定して3ヶ月間の結果や1年間の結果を見られるようにしたい」等の意見を得た。

5. まとめと今後の課題

今回開発したシステムによってサイト運営者のウェブでの集客を効果的に増やすための負担を減らすことができたのではないと思う。

今後の課題として機能の改善として「通知メールの設定(通知先、通知間隔、通知時刻)」、「データ出力期間の設定」がある。他に利用者の見えない部分として自動で順位を行っている部分では検索エンジンへのリクエストから返ってくるまでの処理時間が長く感じ、順位の取得を行うサイトの数が増えるほどサーバーの負担が増えていだろう。表示順位の低いところまで見ればみるほど処理が増えるため確認をする順位の設定を考えなければならないであろう。また、「設定しているキーワードの出現数のカウント」、「無効リンクのチェック」、「被リンク数の確認」などサイト自体の解析を行えるとよりサイトを管理しやすくなると考えられる。改善や機能の追加を行いより便利で長期的に利用のできるシステムとして今後も稼働させていきたい。

参考文献

[1] 西田圭介: Googleを支える技術 巨大システムの内側の世界: 技術評論社

[2] JavaScriptist (<http://javascriptist.net/>): 最終閲覧日 2012/11/19

[3] SEOの必要性とメリット (http://www.stc-77.com/seo_2.htm): 最終閲覧日 2012/12/14