特集 1
SPECIAL FEATURE

Webアプリケーションへの応用

HTTPモジュールでコード変換処理を横取りして混在を整理

大澤 文孝　OSAWA, Fumitaka

Level

1 2 3 4 5

Technology Tools

- [x] Visual Basic
- [ ] Visual C#
- [ ] Visual C++
- [ ] SQL Server
- [ ] Oracle
- [ ] Access
- [ ] Excel
- [x] ASP.NET
- [ ] Other:

Samples

この記事で取り上げたソースコードおよびサンプルプログラムは、
http://www.shoeisha.com/mag/windev/
からダウンロード可能です。

はじめに

Blogには、検索エンジンが訪れたときに、その検索語句を自身のページに埋め込んで表示する機能をもつものがあります。

本稿ではそのようなBlogに似た仕組みとして、静的なHTMLに、「<!-- #SearchResult -->」という特殊なコメントを埋め込んでおくと、検索された語句がHTML化されて埋め込まれるというアプリケーションを作っていきます(図1)。

検索語句は、HTTPのヘッダ情報から取得できますが、ヘッダに記載された文字コードとHTMLを構成する文字コードとは異なることがあるので、文字コードの変換が必要になります。

文字コードの変換には、前稿「文字コードの自動判別機能実装」で作成したKanjiLib.Convertクラスを用いることにします。

Refererを使った検索語句の集計

図1のようなアプリケーションを作る場合、まず必要となるのが、検索語句の抽出と集計です。

本稿では、yahoo.co.jpやgoogle.co.jp、

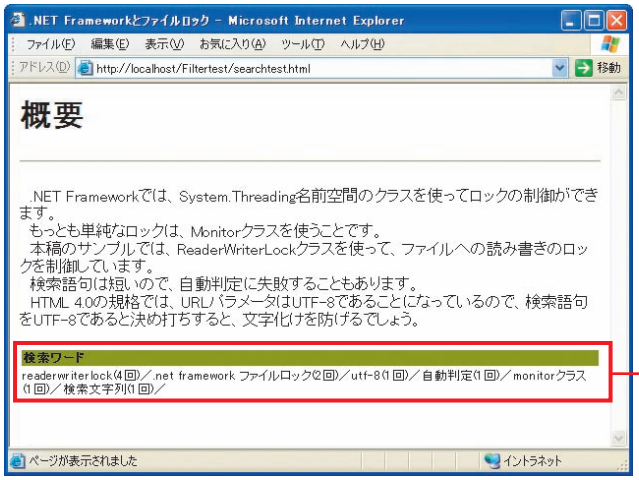

図1：検索語句を静的なHTMLに埋め込む

静的なHTMLファイルに「<!-- #SearchResult -->」と書いておくことで、ここにそのページが検索されたときの語句一覧が埋め込まれる。

msn.co.jpといった検索エンジンで入力された検索語句を対象とします。

ユーザーが検索エンジンを使ってWebページを検索するときの流れは、図2のようになります。

検索エンジンに検索語句を入力すると、その結果が一覧として表示されます。結果ページにはリンクが付いており、そのリンクをクリックすれば、目的のページへと訪れることができます。

この流れにおいて、次の2点に着目すると検索語句を抽出できます。

・結果ページのURLには検索語句が含まれる
・リンク元はRefererヘッダで示される

結果ページのURLには検索語句が含まれる

結果ページのURLには、検索語句がURLエンコードされて付随します。

たとえばgoogle.co.jpで「翔泳社」を検索したときには、「q=%E7%BF%94%E6%B3%B3%E7%A4%BE」というURLパラメータが付きます。

この「%E7%BF%94%E6%B3%B3%E7%A4%BE」という値は、「翔泳社」という文字をUTF-8で示し、それをURLエンコードしたものです[注1]。

検索語句のパラメータ名は、検索エンジンによって異なります。google.co.jpの場合には「q=」となりますが、yahoo.co.jpの場合には、「p=」というパラメータ名となります（表1）[注2]。

図2：検索エンジンから結果を辿るまでの流れ

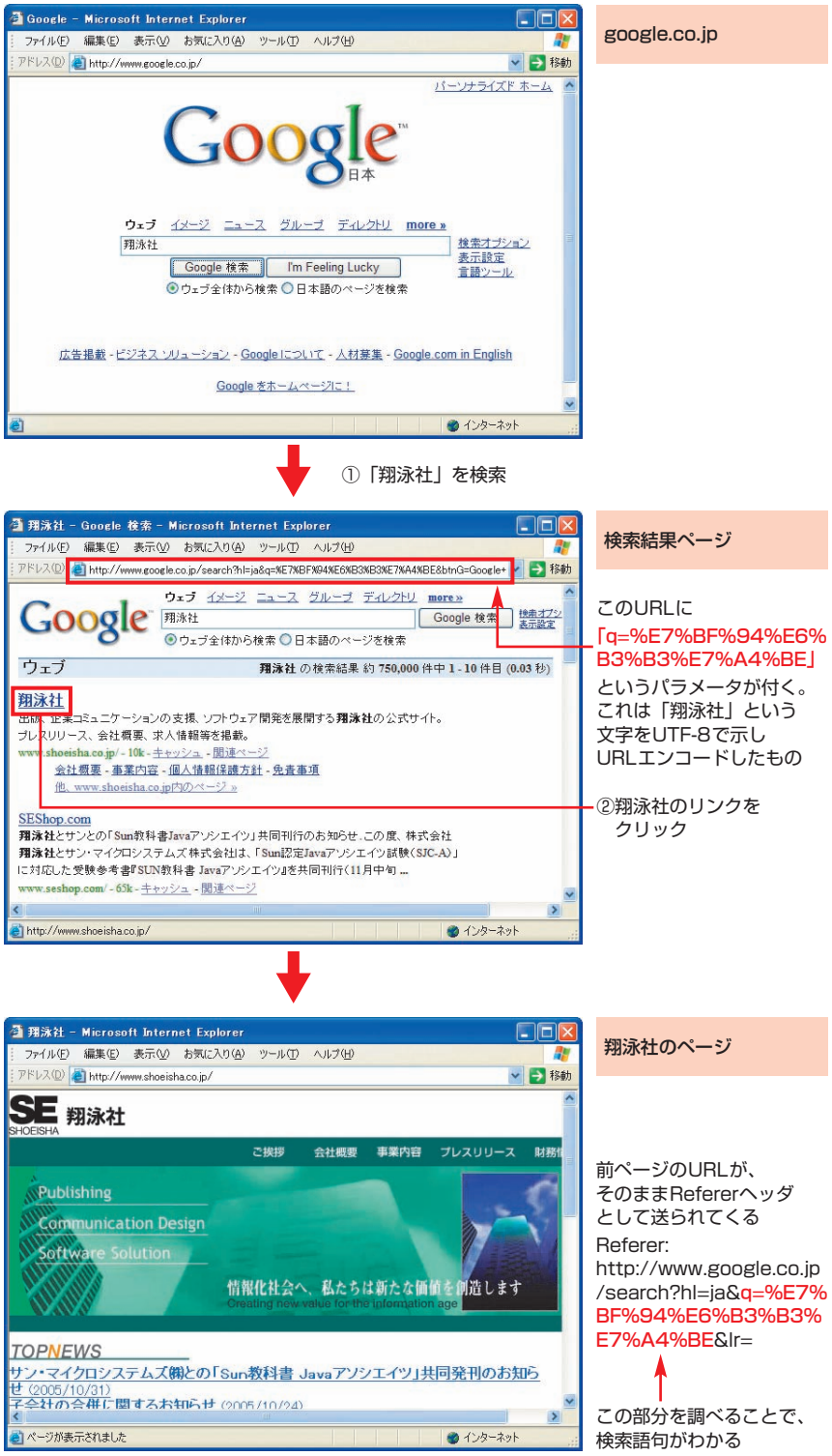

表1：主な検索エンジンの検索パラメータ

検索エンジン	ホスト名	パラメータ名
yahoo.co.jp	search.yahoo.co.jp	p=
google.co.jp	www.google.co.jp	q=
msn.co.jp	search.msn.co.jp	q=

注1）URLエンコードとは、非アスキー文字を「%XX」の16進数に、また、空白を「+」に変換したもののことです（RFC2396を参照）。
注2）表1は、実際に検索エンジンで検索してパラメータ名を調べたものにすぎず、仕様として定められたものではありません。将来的にホスト名やパラメータ名が変更される可能性もあります。