

Web 検索エンジン

理工学部知能情報システム学科 渡辺 義明

1 はじめに

インターネット上には極めて多様な情報が公開されるようになってきました。有効に利用すれば強力な情報取得手段になります。皆さんは、必要とする Web(World Wide Web または WWW) ページへどのようにして辿りついていきますか。スタートページからリンクを辿ることだけでは、暇つぶしには良いでしょうが必要ページへ到達することは困難です。

必要とするページをすばやく探し出すには検索エンジン(検索エンジン)が便利です。この記事では、この検索エンジンの使い方を概説します。

2 検索エンジン

世界中に存在する Web ページの情報を収集整理しておき、ユーザーからの問い合わせに対して該当するページ一覧を提示するサービスが存在します。これを検索エンジンと呼びます。表 1 に国内外の代表的な検索エンジンの例を挙げます。これ以外にも多数あります。また盛衰も激しいので、この表はあくまでも執筆時点での代表例として理解下さい。

情報の収集整理法によって、サービスは大きく二種類に分かれます。一つは、Web ページを分野(カテゴリ)別に分類して保存し、一覧表として提示する形式のサービスです。ディレクトリ型(またはカテゴリ型)と呼びます。分類登録を行うのは、そのサービスの管理者および登録を希望するページの管理者です。人手を介しますので、比較的整理された情報が得られますが、情報の量は少なくなります。

もう一つは、Web ページの情報をロボットと呼ばれる自動プログラムで定期的に自動収集してデータベースに保存しておき、キーワードを用いた問い合わせに対して、データベースから該当ページを探し出して提示するサービスです。ロボット型と呼びます。こちらは情報の収集から検索まで全自動で行

表 1: 検索エンジンの例
国内ディレクトリ型中心サービス

Yahoo! Japan	http://www.yahoo.co.jp/
ACARA NAVI	http://www.acara.com/
CSJ Index	http://www.csj.co.jp/csjindex/
DRAGON	http://www.dragon.co.jp/

国内ロボット型中心サービス

goo	http://www.goo.ne.jp/
Open Text	http://www.jp.opentext.com/
RCAAU	http://kuamp.kyoto-u.ac.jp/labs/infocom/mondou/
AltaVista	http://altavista.dec-j.co.jp/

国内併用型サービス

Infoseek Japan	http://www.infoseek.co.jp/
Excite Japan	http://jp.excite.com/
NTT Directory	http://navi.ntt.co.jp/
Hole-in-One	http://hole-in-one.com/
InfoNavigator	http://infonavi.infoweb.or.jp/
Netplaza	http://netplaza.biglobe.or.jp/
Lycos Japan	http://www.lycos.co.jp/

海外ディレクトリ型中心サービス

Yahoo!	http://www.yahoo.com/
Web Crawler	http://webcrawler.com/

海外ロボット型中心サービス

Alta Vista	http://www.altavista.com/
Open Text	http://index.opentext.net/
Northern Light	http://www.nlsearch.com/

海外併用型サービス

HotBot	http://www.hotbot.com/
Excite	http://www.excite.com/
Lycos	http://www.lycos.com/
Magellan	http://www.mckinley.com/
Infoseek	http://www.infoseek.com/

いますので、未整理ですが大量の情報が手に入ります。これを狭義のサーチエンジンと呼ぶこともあります。

ディレクトリ型の例としては、下図の Yahoo! Japan があります。提示されたカテゴリ分類項目を選択することにより、その分野のより細かい分類項目が表示されます。また、キーワードで登録情報を検索することもできます。誰もが興味を持つような一般的な情報であり、そのカテゴリ分類項目が存在する場合は、有用な情報を効率良く得ることができます。しかし、特殊な情報を得ようとすると情報不足になります。



ロボット型では、入力窓に入れたキーワードを元に収集した情報を検索します。特殊なものでも探せますが、その反面一般的なキーワードの場合には余りに多数のページが提示され戸惑うことになりま。この時は複数のキーワードを組み合わせて絞り込む必要があります。

表1は、どちらの方式を中心にしているかで分類しています。しかし、各サイトとも両方のサービスを徐々に融合してきていますので、この分類は大体的目安とと考えてください。

3 検索の実例

一般的に述べても分かりにくいので、ここでは具体的な探索要求例について見ていきましょう。以下の例は、一般的な事から特殊な事へ順に並べていきます。

1. サーチエンジンについて知りたい。
2. イッキ飲みの危険性について知りたい。

3. ワインのフルボディとは何の事か知りたい。
4. 家計の可処分所得の年次推移を、佐賀と全国で比較したい。
5. Lin-Kernighan アルゴリズムの実装プログラムを探したい。

3.1 「サーチエンジンについて知りたい」

誰もが関心を持つようなキーワードの場合は、ディレクトリ型のサーチエンジンで試すことを薦めます。カテゴリ分類項目と一致すれば整頓された一覧が手に入ります。ここでは、まず、Yahoo! Japan で「サーチエンジン」と指定して検索してみます。



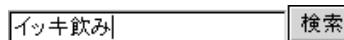
やはり、該当するカテゴリ分類項目がありました。



そこから辿れば、サーチエンジンの使い方や評価、比較などのページを見ることができます。

3.2 「イッキ飲みの危険性について知りたい」

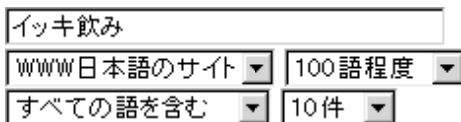
これも、まずは Yahoo! Japan で「イッキ飲み」のキーワードで探してみます。



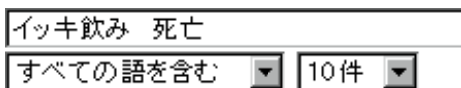
カテゴリ分類にはありませんが、防止の活動をしている団体のページが見つかりました。



しかし一つだけしか見つからなかったため、今度はロボット型の検索エンジンとして代表的な goo で探してみます。



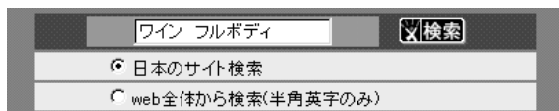
ここでは、300件余りが見つかりました。少し多いので、その中でも切実な死亡事故を扱ったものに絞ることにします。キーワードを複数指定して、それらを両方とも含むページを検索します。



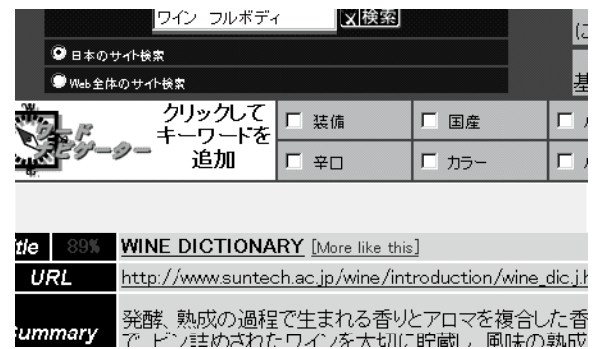
その結果、28件が見つかりました。

3.3 「ワインのフルボディとは何の事が知りたい」

この例では、検索エンジン Excite Japan で、二つのキーワードを指定して探してみましよう。



結果として得られたページ一覧の最初に、Wine Dictionary なるタイトルのページがあります。なおこの検索エンジンでは、関連キーワード追加や類似ページ検索 [More like this] ができることが図から分かります。

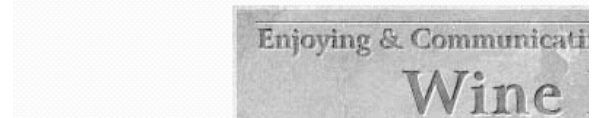


Wine Dictionary なら説明がありそうです。しかし、これを選択すると、ページが見つからないとのメッセージが表示されました。



404 Not Found

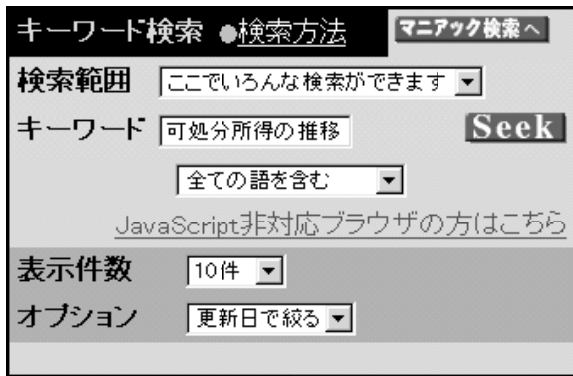
このような場合は別の候補ページを当てることになりません。しかし、この例の場合には、各種ワイン銘柄の味のページが殆どであり、確かにフルボディという用語は使っているのですが、それ自体の説明はなかなか見つかりません。そこで先ほどのページに戻り、URL (アドレス) の末尾を削って上位のページを見てみます。



すると該当ページは、無くなったのではなく、他へ引っ越したことが分かりました。Web ページの構成はその管理者によって少しずつ変更が加えられています。該当ページは名前やリンク構造上の位置が変えられただけでもかもしれません。末尾を削ってトップページに近いところから辿りなおせば見つかることがあります。

3.4 「家計の可処分所得の年次推移を、佐賀と全国で比較したい」

これはキーワードの選択が難しい例です。どのキーワードも一般的で、多数のページと一致します。ここでは、さまざまな入力方式が提示される Infoseek Japan で調べてみます。



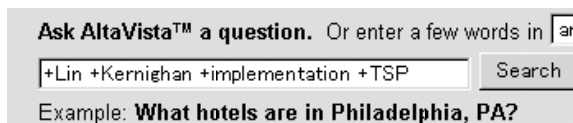
キーワードをいろいろ組み合わせても期待するページがなかなか得られません。このような場合は良いキーワードを思いつけるかどうかが鍵です。今回は「統計」をキーワードに加えることが成功に結びつきました。



上の入力で佐賀のデータが得られ、「佐賀」を除き「全国」を入れれば、全国版のデータが得られました。両者とも解析に使えるデータ形式が取得できました。

3.5 「Lin-Kernighan アルゴリズムの実装プログラムを探したい」

最後は特殊な情報を探す例です。殆どの人には何の興味も無い情報でも、自分が有用と思うなら、世界中には同じように有用と考えている人がいるかもしれません。ここでは、私自身の研究に必要なコンピュータプログラムを探してみました。この情報は海外にある可能性が高いため、ロボット型で有名な海外サイト AltaVista を使うことにしました。表題中のキーワードを並べただけでは候補が多すぎて有用情報に辿りつけません。このような場合には、様々な追加キーワードを試してみることになります。結局、このプログラムを適用する最適化問題の名称「TSP」を加えることで絞り込みました。



AltaVista では、「全て」のキーワードを含むペー

ジを探す場合、キーワードの前に半角のプラス記号を付け半角スペースで区切って並べます。単にスペースで区切ると、与えたキーワードの「いずれか」を含むページを探します。

該当ページが長文の場合、キーワードがそのページのどこに含まれているか分からないことがあります。この時には、Webブラウザのメニューにある「ページ内検索」の機能を使うと便利です。



今回も、可能性の高い候補ページが長文であったため「Kernighan」をページ内検索し、周辺を拾い読みして見つけました。

4 検索のヒント

複数のキーワードを組み合わせ指定する場合、それらキーワードの全てを含むとする AND 型の指定と、いずれか少なくとも一つ以上のキーワードを含むとする OR 型の指定とが良く使われます。要は、多すぎれば AND 型でいくつもキーワードを並べて絞り込み、少なすぎれば OR 型で範囲を広げればよいことになります。また、あるキーワードを含まないとする NOT 型もありますが余り使いません。絞り込み途中で目障りな項目が目立って来た場合などに利用します。

表 2 にこれらの指定方式を例示しています。多くのサーチエンジンでは、AND/OR のどちらの指定かを、キーワード入力窓の傍にある「全ての語/いずれかの語」のようなメニュー選択で指示します。

メニュー選択がなく、キーワードの前にプラス記号(半角)が付けば、そのキーワードは必ず含まなければならない、単にスペースではいずれか一つ以上を含めば良いことを表す場合もあります。マイナス記号(半角)の付いたキーワードを含むページは一覧から除かれます。他にも様々な指定方式がありま

表 2: 複数キーワードの組み合わせ例 (詳しくは各サーチエンジンの HELP を参照)

サーチエンジンの例	キーワード「boy」、「girl」をともに含むと指定	キーワード「boy」、「girl」のいずれかを含むと指定	キーワード「boy」を含み「girl」を含まずと指定
goo Infoseek Lycos 他多数	boy girl 全ての語を含む (AND)	boy girl いずれかの語を含む (OR)	いくつかのエンジンは 拡張機能等に対応
AltaVista Yahoo! 他	+boy +girl	boy girl	+boy -girl
Lycos の パワーサーチ 等の拡張機能	boy AND girl	boy OR girl	boy NOT girl
Yahoo!Japan の エキスパート 検索	boy*girl	boy+girl	boy#girl

すが、これら指定方式はサーチエンジンによって少しずつ異なりますので、それぞれのサーチエンジン上で使用方法を確認してください。

さらに、サーチエンジンによっては、括弧を使った優先順位付けや、キーワードが互いに近くにあることの指定、キーワードが表題やリンク名など特定の場所にあることの指定、画像など特別の形式であることの指定なども可能です。

検索の成否は、キーワードの選択に大きく依存します。目的の情報にベストマッチするキーワードを思いつけば成功です。最初の検索で提示された候補ページをいくつか見ていくと追加すべきキーワードが鮮明になってきます。また関連キーワードを提示してくれるサーチエンジンもあります。

該当するページが多すぎたり少なすぎたりした場合には以下のような手段があります。

● 多すぎる場合

- ディレクトリ型 (または併用型) サーチエンジンを使う。
- ロボット型では、たくさんのキーワードを AND 接続で並べてみる。
- 「リンク集」などをキーワードに追加して、関連情報の要約ページを探す。
- サーチエンジンを変えてみる。

● 見つからない場合

- ロボット型 (または併用型) サーチエンジンを利用する。
- たくさんのキーワードを OR 接続で並べてみる。
- キーワードを短くしてみる。または、分割してみる。
- 同義語、類義語、略語、広義概念の用語などに変えてみる。
- サーチエンジンを変えてみる。
- 検索対象を日本から全世界に広げる。

各サーチエンジンで検索方法が微妙に異なりますので、あまり浮気せずに少数のサーチエンジンを使い慣れることが良い様に思います。

5 Web 上の様々な情報

例えば以下のような情報が Web から手に入ります。サーチエンジンで探してみてください。

地図、郵便番号、電話番号、航空機・宿などの予約状況・受付、文献検索・取り寄せ、英和・和英辞書検索、本の検索・購入、特許情報、乗換案内、時刻表、審議会報告・

答申、統計データ、図書館蔵書検索、オンラインジャーナルなど

また、以下のようなサイトも情報検索には便利でしょう。自分に合うものを探してみてください。

- リンク集を集めたページ。例えば「リンクのリンク」などで、探してください。サーチエンジンとは、また違った便利さがあります。
- サーチエンジンを集めたページ。「メタサーチ」などで探してください。複数のサーチエンジンを利用する検索ページなどもあります。

6 おわりに

Web 上には人間の興味あることはなんでもあると言える状況になってきました。例えば、あるページ(アリアドネ)は、心理学、言語学、教育学、政治学、経済学等、学術分野毎に分類整理したリンク集であり、調査研究に役立つリサーチツールとなっています。また、あるページ(Uncover)では17000もの学術雑誌の目次を無料で検索することが可能です。多くの図書館からリンクがあります。

Web 上の情報は日々更新されており、今日便利なページが明日もそうであるとは限りません。必要となった時点で便利なページを探すことが肝要であると考えて、この記事では具体的 URL(アドレス)はサーチエンジンのみに限りました。上のページへは、文章中のキーワードを組み合わせれば辿りつけると思います。サーチエンジンで探して見ませんか。