

Google などの検索エンジン

Search engine such as "Google"

～新しいビジネス形態の拡大と新世代型の登場～

森 隆史, 吉田 昌太

Takashi MORI, Shota YOSHIDA

Abstract: When we use internet, we usually use a search engine. The most popular search engine of those exist now is "Google". And, the search engines get to have a better opinion as business recently. This paper describes that the development of the search engines and the new form of business, and in this paper it is considered that the future views of search engines.

1 はじめに

インターネットを利用する上で必ずといって良いほど利用するものに検索エンジンがある。検索エンジンには、現在様々なものがあるが、その中でもトップの人気を誇っているのは、「Google」である²⁾。また最近では、検索エンジンのビジネス形態が新しくなっている。

本発表では、Google 登場以前の検索エンジンと Google、さらには新しい検索エンジンについて述べる。また、検索エンジンの新しいビジネス形態について述べ、今後の検索エンジンについて考える。

2 Google 登場以前の検索エンジン

検索エンジンには、ロボット型とディレクトリ型がある。ロボット型検索エンジンは「ロボット¹⁾」を用いて WWW ページを回収し、それを検索可能な形で提供するサービスである。ディレクトリ型検索エンジンとはたくさんのページを主題別に分類した大規模なリンク集のようなものである。

英語のロボット型検索エンジンにおいては、Google が登場するまで altavista に人気があった。その理由は、集めたウェブサイトの情報量がトップクラスだったこと、豊富な検索オプションが用意されていたことからである。しかし、1998 年ごろから、1 つのキーワードだけで検索すると大量のアダルトサイトが上位を占めてしまう状況が起こり、altavista は使いにくいと言われ始めた。Inktomi や excite といった他のロボット型検索エンジンも同じ状況であった。これは当時の検索エンジンが単純に単語の出現頻度やリンク数だけを比較していたため、それを有料アダルトサイトが逆手にとったために起こった。さらにその副産物として、検索エンジンのポイントアップのためだけに作られたページ²⁾が溢れ返り、ま

もな検索ができなくなっていた。

日本でも 1999 年ごろには同じような状況になり、ロボット型の人気は落ち、多くの日本語ポータルサイトはディレクトリ型に力を注いだ。

3 Google

Google は特許にもなっている「Page Rank」という技術を用い、中身のないページや本来のキーワードと関連の薄いページを振り落とすことに成功した。Page Rank でのページのポイントのつけ方は、そのページに対してリンクしているページのポイントとリンク数で決まる。例えば、あるページがいくつかのリンクを張っているとすると、そのページのポイントを張っているリンク数で割ったポイントが各リンクのポイントになる。またページのポイントは、入ってくるリンクのポイントの合計になる。これを表したものを Fig. 1 に示す。Fig. 1 からわかるように、偽リンクページを作りリンク数だけを増やしても、その偽リンクページ自体に人気がないためにポイントが増えることはない。

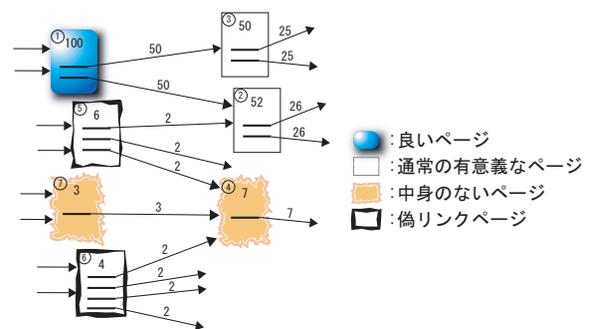


Fig. 1 Page Rank

Google はロボット型でしかも使いやすいため、サービス開始直後から人気が出た。また、多くの特許で技術がガードされているため、他の検索エンジンが同じような技術を開発することは難しい。やがて、Yahoo をはじめ数多くのポータルサイトが Google を採用し始め、ロボット型検索エンジンの世界では Google の独走状態の

¹⁾HTTP クライアント (http サーバに対し送信要求をし、送られてきたデータを受信するソフトウェア) の一種。WWW ページを自動的に回収するソフトウェアのこと。

²⁾中身のないリンクページや同じような中身ばかりのランキングサイトなど。

ような形になっている。

4 新しい検索エンジン

Googleの独走状態は1年以上続いているが、2001年末ごろからようやくWiseNutやTEOMAといった新しい検索エンジンが登場してきた。

4.1 WiseNut

Lycos Japanでは、WiseNutに注目し、2001年12月から日本語版WiseNutをロボット型検索エンジンとして採用している。WiseNutがGoogleと比較しても劣らない理由を以下に挙げる。

- 検索結果の数がGoogleと同じである。
- 検索結果の現れ方がGoogleと大きく違う。
- 「コンテキストランキング³」を使用している。

ただし、コンテキストランキングが効き過ぎてしまうと、キーワードの持つメインの意味のサイトばかりが上位にランクされてしまう。そこで、検索結果の上位に現れるサイトについては、よく使われる意味別にフォルダ分類して表示する機能も搭載されている。

4.2 TEOMA

TEOMAは現時点では実用段階の検索エンジンではないが、その検索方法や表示方法においては、Googleを凌ぐ可能性がある。TEOMAとGoogleの比較をTable 1に示す。TEOMAの表示方法は、もっとも信頼できるサイトを並べる通常のランク付けによる表示と、トピックごとの表示、検索されたテーマの専門家によって作られた「Expert's link」の表示の3つがある。Table 1からわかるように、検索方法、表示方法ともにTEOMAはGoogleよりも優れているといえる。しかし、登録ページ数がまだ少ない上に、アメリカでもベータテストが続いている段階であり、本格的なサービスはまだまだ先になりそうである。

Table 1 TEOMAとGoogleの比較

	TEOMA	Google
評価方法	関連の高いページからのリンク数	良いページからのリンク数
表示方法	検索結果を3つの方法で表示	Page Rankによる通常のランク付けによる表示
登録数 ⁵⁾	1億ページ未満	10億ページ

³ ウェブサイトの中に登場する単語同士の並び方(文脈)をデータベース化し、どういった意味で使われることが多いかというランキングを作る機能。

5 検索エンジンの新しいビジネス形態

最近多くのポータルサイトが検索サイトの機能に取り入れているのが、アメリカのOvertureが始めた「Pay Per Performance」というサービスである。これは、人気になりそうなキーワードの検索で、提携したポータルサイトの上位に表示させる代わりに料金を取る仕組みである。Overtureが人気のある理由は、キーワードごとの細かい料金設定にある。広告主が1クリックごとに支払う金額は非常に高い⁴が、活発なユーザを呼び込めることを考えるとそれだけ広告価値が高いといえる。また、キーワードごとに細かいクリック情報が公開されているため、広告を出す企業から見て、対費用効果がわかりやすいのもOvertureのセールスポイントといえる。検索エンジンポータル側から見ても、広告営業を代行してくれるので好意的である。ただしOvertureを使う際には、ロボット型の検索結果とはっきり区別をつけることがユーザの信頼を失わないために重要なことである。

6 おわりに

今後、前節で挙げたようなWiseNut、TEOMAといった検索エンジンやPay Per Performanceに参入するポータルサイトが増えるだろう。Googleは、2月2日からPay Per Performanceを導入、Yahooもいずれは独自にサービスを開始することを宣言している。そのことで、一時は儲からないといわれた検索エンジンビジネスが見直され始めている。

こうした流れの中から、これまで挙げたGoogleやWiseNut、TEOMAのようなロボット型検索エンジンをさらに上回る検索技術が登場することもありうる。次に登場しそうな技術は、機械的に単語や文字の位置をデータベース化するだけでなく、言語学的にきっちりと構文解析を行ったデータベースを使う検索エンジンである。現在のコンピュータでは、処理能力が不足しているため無理だが、将来的には実現される可能性が高いといえる。

参考文献

- 1) 戸村賢一、月刊アスキー 4月号、pp.137-144、2002。
- 2) 色々な検索エンジン!! [海外の検索エンジン]、<http://viva.wisnet.ne.jp/search/en/index.html>
- 3) Googleの秘密-PageRank徹底解説、<http://www.kusastro.kyoto-u.ac.jp/~baba/wais/pagerank.html>
- 4) WWW情報検索システムの概説、<http://www.etl.go.jp/~yamana/JEIDA/chapter1/chapter1.html>
- 5) グーグルに続く検索エンジンは?、<http://www.hotwired.co.jp/news/news/technology/story/20010822302.html>
- 6) オプトインメール/オプトアウトメール、<http://www.sw.nec.co.jp/word/aa/aa025.html>

⁴ オプトインメール(ダイレクト電子メールの一種)1通が1~4円なのに対し、Pay Per Performanceでは1クリック30セント程度。