

2 Yahoo! Search Technology (YST) と、検索分野における Yahoo! JAPAN の戦略

井上 俊一 (Yahoo! JAPAN リスティング事業部)
toinoue@mail.yahoo.co.jp

宮崎 光世 (Yahoo! JAPAN リスティング事業部)
miyazaki@mail.yahoo.co.jp

🔍 本稿の目的

Yahoo! JAPAN は 2004 年 5 月 31 日、検索エンジンに Yahoo! Search Technology (YST) を採用した。本稿では採用に至った経緯と、現在の開発体制、検索エンジンの評価方法を紹介し、さらに、Yahoo! JAPAN が考える検索エンジンの未来について現在の考えを述べる。

🔍 なぜ Yahoo! JAPAN は独自の検索技術保有に踏み切ったか

Yahoo! JAPAN はサーファアと呼ばれる編集スタッフが登

録、管理するリンク集からスタートし、表-1 に示すように、Web 検索についてはその時最良と思われる外部の技術を使用してきた。その Yahoo! JAPAN がなぜ Web 検索の自社保有に踏み切ったのか、そこには 3 つの理由がある。まず検索技術の本質的な重要性、次にインターネット全文検索エンジンの技術革新、そしてビジネスモデルの確立である。

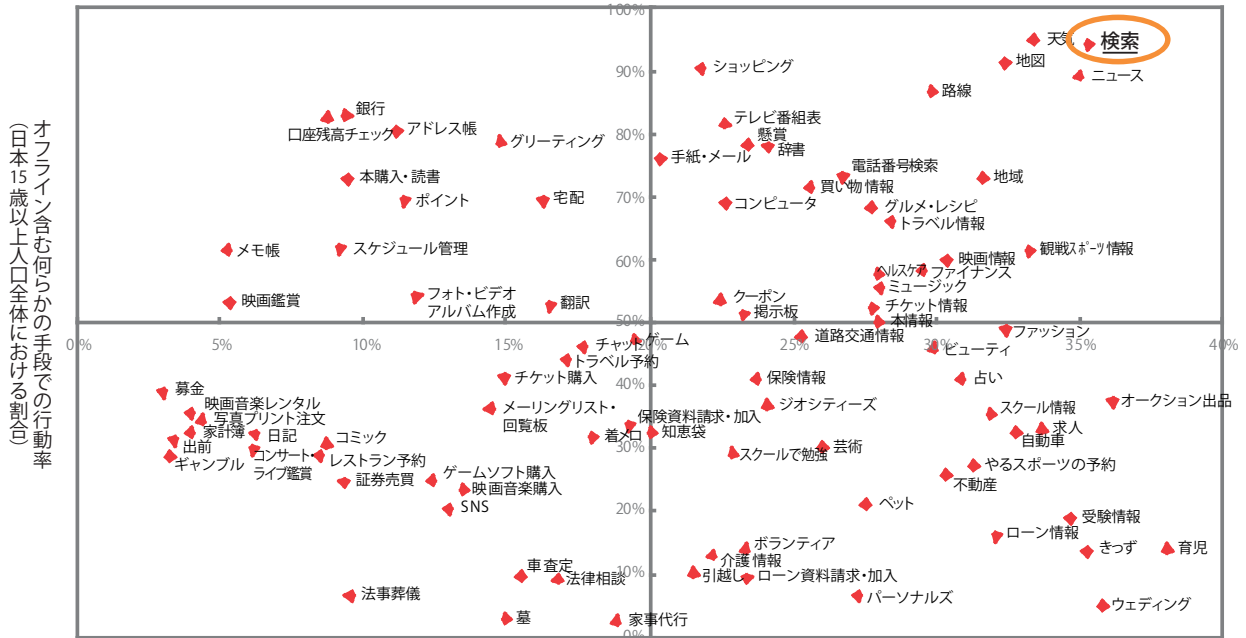
検索はインターネットの本質であり、インフラでありつづける

■インターネットは誰もが情報を発信できるメディアである

インターネットは誰もが情報を発信できるインターパー

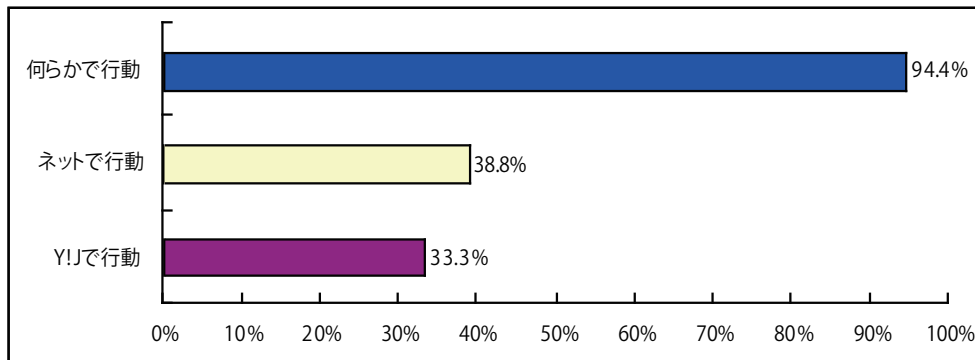
	Yahoo! JAPAN 関連情報	Yahoo! Inc. 関連情報、US の状況
1995 年		Yahoo! Inc. 設立 (4 月) Altavista 全文 Web 検索開始
1996 年	Yahoo! JAPAN 設立 (4 月)	Inktomi 社設立
1997 年		Altavista が多言語対応 広告型検索 GoTo が登場
1998 年		Yahoo! Inc., Inktomi 採用 Google 設立
1999 年	goo をサーチエンジンに採用	GoTo がクリック課金型広告開始 AlltheWeb 開始
2000 年		Yahoo! Inc., Google を採用
2001 年	Google をサーチエンジンに採用 Yahoo! BB サービスを開始	GoTo がブランドを Overture に
2002 年	検索リニューアル (ダイジェスト表示) Overture/Google のスポンサーサイトを表示	Yahoo! Inc., Inktomi の買収を発表
2003 年	サイト検索結果とページ検索結果を同時表示	Overture が Altavista, FAST 社の Web サーチ部門 (Alltheweb) を買収 Yahoo!, Overture を買収
2004 年	画像・音声・動画検索開始 知恵袋 (知識検索) β 公開 辞書サービス強化 ページ検索を YST に切替 スポンサーサイト広告を Overture のみに変更 ダイレクト検索を開始	Yahoo! Research 設立 Yahoo! Inc., YST を採用 YST が RSS 対応 Yahoo! Inc., FareChase 買収 Yahoo! Inc., Kelkoo 買収
2005 年	商品検索 β 公開	

表-1 YST採用までの経緯



Yahoo! JAPANでの利用率(各何らかでの行動者ベース)

図-1 生活における行動とYahoo! JAPANの利用率
 出典:「生活に関する情報源調査」 調査時期:2005年2月17~21日 調査対象:15歳以上のYahoo!リサーチ・モニタ男女 1,140人
 日本人口/インターネット人口にあわせて質問結果を集計修正



結果:質問したさまざまな日常行動89項目の中で、最も高い「インターネットでの行動」率(38.8%)

- しかも38.8%中, Yahoo! JAPANでの行動率は33.3%
- つまりネット利用者を100%とすると85.8%がYahoo!JAPANを利用

図-2 質問「過去12カ月間に知りたい・興味がある情報を探したり調べたりしましたか？」
 出典:「生活に関する情報源調査」

ソナなメディアである。これを Yahoo! JAPAN では「My Media」と呼んでいるが、インターネットの最も重要な本質であり、他のメディアとまったく違う点である。Web という仕組みを使えば誰もが容易に情報発信できるがゆえに、非常に多様な情報が存在しており、世界のどこからでもその情報にアクセスできるようになっている。そこで必然として起こるのが、世界のどこかにいる誰かが、まったく別のところにいる何の面識もない誰かが発信している情報を探すとという行動である。情報が提供され、知

りたい人がそれを見つけ、情報が活かされ、それがさらなる情報発信を生む。その相乗効果が Web を歴史上存在したことの無い巨大な情報の塊に育てた。繰り返しになるが、インターネットは誰もが情報を発信・受信できる最初のメディアであり、検索はそのインフラである。それが検索技術を保有する最大の理由になっている。

図-1, 2は「検索」が現代人の生活において誰もが行う日常の行動であり、そのインフラとして「検索」(ここでは Yahoo! JAPAN) が機能していることを示す調査結果である。

インターネット全文検索エンジンにおける技術革新

■ディレクトリサービスからはじまった Yahoo! JAPAN がなぜ?

Yahoo! JAPAN はこれまでディレクトリサービスとインターネット全文検索を対立してとらえたことはなく、本の目次と索引のように両方必要なものとして考えてきた。ただ、インターネット全文検索が未熟だったころには、100%人間が作成したディレクトリの方が、たとえば公式サイトを確実にヒットさせるには精度が高いという現実もあった。そのため、ディレクトリサービスを強化しながら、インターネット全文検索エンジンについてはその時々ベストと考えるサービスと提携してきた。

ディレクトリサービスとしての Yahoo! JAPAN は Web サイトを探す最初の場所として定着し、先に述べた相乗効果の中で成長を遂げることとなる（その間にニュースや天気予報の提供、オークションの開始等、いわゆるポータルとしての展開もあるが、その点については割愛）。が、規模の拡大とともにスケーラビリティの問題に突き当たるようになる。Web にある情報量があまりにも膨大になったために起こる質と量の二律背反である。Yahoo! JAPAN は検索する人にとって価値があるならば、すべての情報を検索できるようにしたいと考えており、この問題を解決する必要があった。

一方、インターネット全文検索の世界では、精度と網羅性を飛躍的に上げたいいわゆる「第2世代」の検索エンジンが登場する。これらの検索エンジンは、コモディティ（日用品）と化した PC サーバやネットワーク帯域をふんだんに使い、オープンソースソフトウェアを駆使して、それまでは考えられなかった膨大なデータ処理を行って検索性能を飛躍的に上げた。第2世代の検索エンジンには多くの特徴があるが、機能としてポピュラーなのはリンク分析やアンカーテキストを用いること、結果として最も明らかなのはナビゲーションクエリ（探している情報のありかが誰にも明らかであるクエリ、たとえば会社名や有名ブランド名等）について非常に高い精度で結果を出す点等が挙げられる。つまり、キーワード検索においてディレクトリサービスが担っていた部分がある程度カバーしながら、それまで考えられなかった網羅性を両立する性能が実現してきたのである。そして、百億以上ものドキュメントを1秒以下で検索する巨大なシステムが人々のコモディティとして定着した。

今後のディレクトリの位置付けだが、大きく2つの面で高い価値を持ちつづけるだろう。1つ目はクリックだけで検索ができるということ。リテラシーの観点から、あるいはテレビやゲームのリモコンのようにキーボード

がない環境等を想定すると、キーボードを必要としない検索は今後ますます重要になる。もう1つは、良い Web 検索を作るときに良質なディレクトリは欠かせないということだ。たとえば企業間取引を行う企業群を取り出したり、「都市銀行」のように社会的に規定されたサイト群をリストする時にディレクトリのメタ情報が役立つ。地域情報等もディレクトリが生きる分野だ。たとえば Yahoo! 商品検索ではディレクトリ登録サイトから無料の Feed を受け付けるサービスを行っているが、これができるのは、ディレクトリが、質が担保されたオンラインショップが分類登録されたサイト集だからである。

新しいビジネスモデルの成功

■スポンサーサイトのビジネスモデルとは

上記の状況を受け、Yahoo! JAPAN は常に次世代の検索サービスについて検討を重ねていたが、それを大きく後押ししたのがスポンサーサイトの存在である。検索サービスの成長を阻害してきたのは、コストばかりで儲からないという状況だった。それが、この新たなビジネスモデルによって一変した。

スポンサーサイトは、現在の Yahoo! Inc. の Search Marketing 部門（旧 Overture Services 社）が、かつて「GoTo.com」という名称だった時代に考案したビジネスモデルで、具体的にはクエリに連動したテキスト広告であるが、CPC（Cost Per Click：クリックあたり課金）であること、さらにそのクリック単価がオークションモデルによって決まるという点に大きな特長がある。ユーザには、広告ではあるが、自分が探している情報に関する内容であり、広告主としては支払うコストに対して確実にアクセスが得られるという、両者に利がある優れたモデルである。IAB（Internet Advertising Bureau）の発表によれば、米国で年間1兆円規模に成長したインターネット広告市場で4割のシェアを占めるまでに成長、日本を含めた世界各国でも急成長をとげている。

今も昔も Web 検索の最大の課題は技術・ビジネス両面のスケーラビリティであり、検索サービスがペイしていくためには、このビジネスモデルは欠かせない。Yahoo! Inc. はスポンサーサイトに関する特許を Google にライセンスしている。

🔍 Yahoo! Search Technology とは 検索技術を自社開発に

Yahoo! Inc. と Yahoo! JAPAN は親子関係の会社ではないため、検索エンジン選択についての検討は両社別に行っていた。そして、Yahoo! Inc. ではさまざまな検索エンジンを比較検討した結果、結局検索技術を自社技術

化することを決断し、Inktomi 社と Overture 社（当時の社名）を買収した。Overture 社は著名な検索サービスである Alltheweb、Altavista を買収しており、結果その時点では Inktomi、Alltheweb、Altavista と 3 つの検索エンジンを保有することになった。Yahoo! Inc. はこれら 3 つの検索エンジンと Yahoo! の技術を融合して、新しい検索エンジンを開発した。これが Yahoo! Search Technology (YST) である。

一方、Yahoo! JAPAN はこの新しい検索エンジンが第 2 世代のクオリティと、Web の拡大に適應できるスケールビリティを持った、深い協業が可能な検索エンジンであると評価して採用に踏み切った。自社開発の重要性は Yahoo! JAPAN も同様に認識しており、単純に優れた技術として採用するだけでなく、奥の奥まで共同で開発できるよう体制面も同時に整備した。

商用検索エンジンにおける検索結果の評価方法

「RCFPT」について

Yahoo! JAPAN で検索エンジン进行评估する際は、社内で「RCFPT」と呼ぶ、下記のような性能を重視する。この RCFPT それぞれに対して、性能指標が存在する。これらの中で特に重要なのが RCF であり、最も重要なのは R: Relevancy である。これは検索ワードに対して、どれだけ適切な検索結果を表示しているか、つまり検索結果の質を表す項目である。各項目に対し、複数の評価指標があるが、この中で Relevancy を調べるために用いている方法の 1 つを紹介する。

1. R: Relevancy (クエリと検索結果の関連性)
2. C: Comprehensiveness (網羅性)
3. F: Freshness (新鮮さ)
4. P: Presentation (検索結果の提示方法)
5. T: Trust (信頼性)

Relevancy 評価の一例

紹介する方法は特殊なものではなく、大学や研究室などで行われている方法と同じだと思われるが、より組織的に実施している点が違うと想像する。

1. クエリセットを準備する（実際のクエリからランダムに抽出）。
2. クエリセットに対して検索エンジンから X 件までの結果を抽出する（X の数値はテストの目的によって変更する）。
3. 抽出した検索結果をバラバラに並び替える。
4. 各結果に関して、クエリに対する関連性の評価を行う。

5. 4. における採点結果を集計して指標化する。たとえば下記のようなことを調べる。

- (ア) トップ 10 件の全体的なクオリティはどうか？
（1 位から 10 位まで傾斜配点した総合得点）
- (イ) トップ 1 のクオリティは完璧といえるか？
- (ウ) 逆に悪い結果はどうか？

このような分析により、全体性能を上げるチューニングも可能になるし、検索結果の性能を下げている個別の問題も発見できる。YST では常にインデックスが更新されつづけているが、これらのスコアを通して、Relevancy の変化も把握できる。

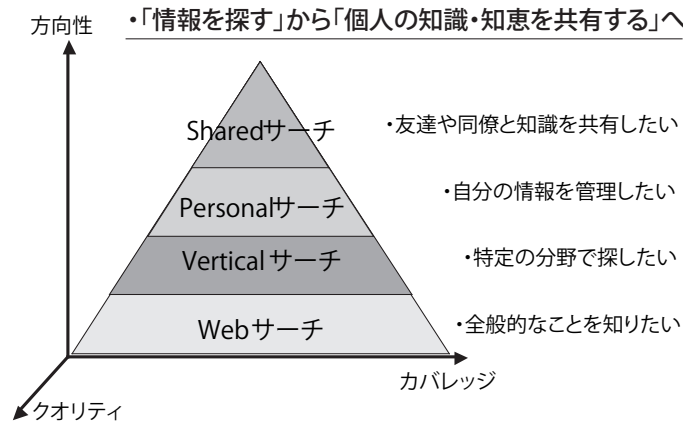
なお、これらの非常に重要なデータを作成するのは専門のエディトリアルスタッフである。クエリのバリエーション、Web 上の情報のバリエーションがあまりにも広いので、Web を知り尽くした専門スタッフの方が、質の高いデータを作成できている。

ログの分析とライブテスト

Yahoo! JAPAN において検索サービスの改善はすべてログデータの解析からはじまる。ページのどこがどれだけクリックされているか、たとえば検索結果からあるページに行ったがすぐ戻って来てしまった、といったデータは問題点の抽出に非常に役立つ。また、社内で「ライブテスト」と呼んでいるが、ブラウザクッキーごとに関連したセッティングやデザインの検索結果を表示し、クリック動向等を分析するといったことも行っている。そこでの経験則だが、検索結果デザインを変更しても画面全体のクリック総数が変わることはめったになく、ここが増えれば別の場所が減るといったようにトレードオフとなることが多い。しかしデザインの改善が成功すると、全体のクリック数が底上げされる。

Web 検索に答えはない

Web は Structured なデータではない。文章があり、表があり、会話があり、機能があり、画像や音声、動画といったメディアがあり、経済があり、社会があり、法があり、そして巨大である。Web 検索の Relevancy 評価を有効に行い、恒常的に向上させるためには、スケールする理論・技術と、それを実際に市場のニーズにつなげていく質の高いデータが必要である。顧客のニーズをつかむためログを分析し、リサーチをかけ、プロダクトにしっかりと方向性を出す。スケールする技術的枠組み（マシンラーニングが多用される）を設定し、エディタ、エンジニア、サイエンティストらが一体となって開発を行っている。



サービス例:	
Web サーチ	Yahoo! Search Technology
Vertical サーチ	ダイレクト検索 <search.yahoo.co.jp/promo/direct.html> Yahoo! 知恵袋 <knowledge.yahoo.co.jp/> Yahoo! 商品検索 <psearch.yahoo.co.jp/> Yahoo! 画像・音声・動画検索 <beta.search.yahoo.co.jp 参照 >
Personal/Shared サーチ	後述

図-3 Yahoo! JAPANの戦略

🔍 日本市場に向けた開発体制

先に述べたように、Yahoo! JAPAN は YST を共同開発できる検索エンジンとして採用した。現在の YST の検索結果には Yahoo! JAPAN による成果も含まれている。

日本語処理技術の重要性

日本市場を見たときに、まっ先に問題になるのは日本語の特殊性である。日本語は単語と単語の間に切れ目がなく、漢字・ひらがな・カタカナ・英語などがまざりあって使われる。さらに送りかなの送り方、漢字の当て方、旧字・新字等、文字列を処理する検索エンジン泣かせの言語である。たとえば表記ゆれは Web 上に存在するテキスト情報、たとえばアンカーテキスト（リンクに使われる言葉）を使えばある程度吸収できるのだが、それだけでは不十分である。Yahoo! JAPAN はこの問題を重視し、日本語処理専任のエンジニア・サイエンティストチームを設けて開発を行っている。

日本市場に向けた最適化

日本市場の特殊性は言語だけではない。たとえば、日本語 Web の規模は、英語の Web に比べれば小さい。これはリンク分析に影響があると考えられる。巨大な匿名掲示板、日記文化…等、Web コミュニティのあり方も違う。グローバルな検索エンジンの日本語版ではダメで、日本人、日本市場のための検索サービスとならなければならない。そのために顧客から寄せられるフィードバックを今後のサービスに活かしていきたいと考えている。

各国 Yahoo! の検索担当者が集まると、各国の状況がよく分かるのだが、非常に興味深い。たとえば中国では Web 検索に対するスパム行為が非常に大きな問題である。韓国ではブロードバンドの進展を反映して動画や画像が満載の非常にリッチな検索結果が好まれている。

🔍 Yahoo! JAPAN 検索サービスの今後

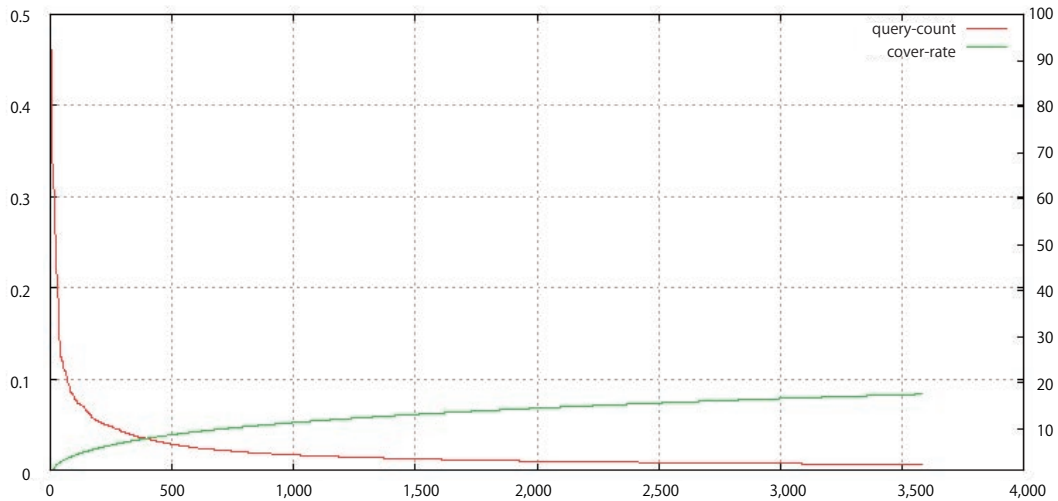
戦略

図-3 に示すように、戦略の中核にあるのは Personalization と Shared Search であり、その詳細は後述する。同時に、我々が「Vertical Search」と呼ぶ分野を絞った専門的な検索、特に Web 上のあらゆる商店から物を買うことができる商品検索と、Yahoo! JAPAN で「Yahoo! 知恵袋」と呼んでいる Q&A サービス (Knowledge Search) による知識の蓄積、これら 2 サービスは特に重要である。さらに、画像・音声・動画といったメディアデータを含めた展開、デスクトップ (PC) 内データの検索、モバイル、それらを統合するダイレクト検索等、検索サービスの強化に向け、非常に広範な取り組みが行われており、それをシンプルな検索窓の裏側に結実させようとしている。

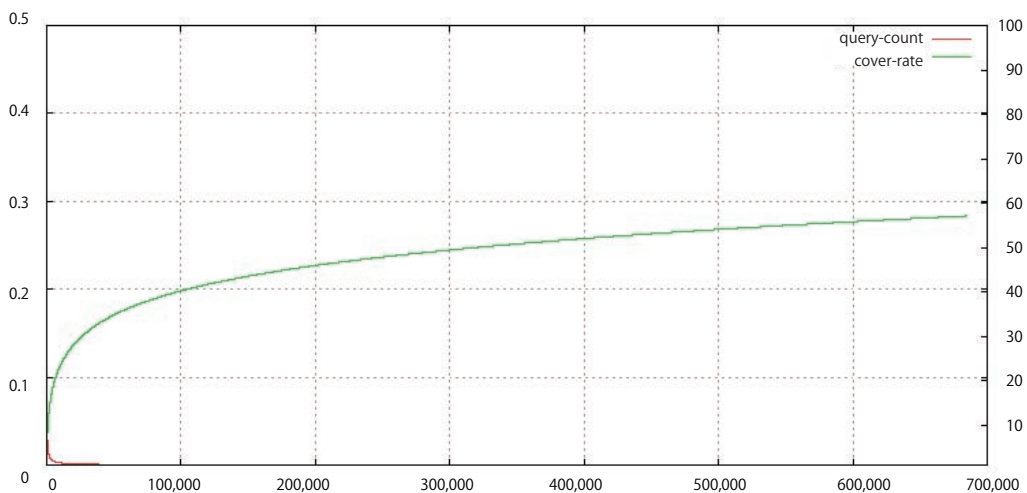
My Media としての検索サービス “Personal Search”

検索という行動はおおむね個人的である。それを端的に表す図として図-4 を示す。これは 1 日の検索ログを

クエリディストリビューション：検索回数1,000回までの分布



クエリディストリビューション：検索回数10回までの分布



左右Y軸：PV占有率(%)
X軸：検索ワードランキング

query-count (赤線)：当該順位検索ワードのPVが全体のPVに占める割合 (縦軸左)
cover-rate (赤線)：当該順位までの全検索ワードのPV合計が全体のPVに占める割合 (縦軸右)

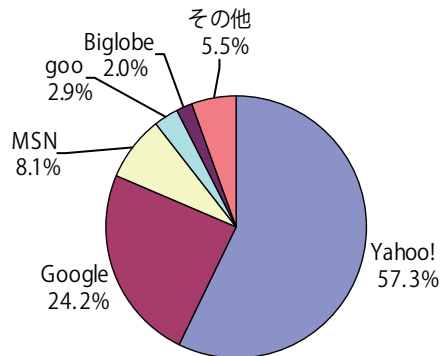
図-4 「クエリカーブ」1日の検索ログを分析. 検索ワードごとに検索数を合計し, 上位から並べたもの.

検索回数順に並べたものだが、皆が共通して検索するワードはごく一部で、後は緩やかに限りなく1に近づきながら伸びていく。その1日に1回しか検索されないクエリは、1日の全クエリの60%以上を占める。つまり、現状でもWeb検索サービスは、一人一人の検索要求に対するために、何千万というバリエーションの検索結果を表示しているのである。インターネットが「My Media」として強力に機能し、検索がそのインフラとして個人と情報、個人と個人との出会いを作り出している

ことが現れている。

このような状況がありながら、現在の検索サービスは、誰が検索してもキーワードが同じであれば誰にも同じインデックスから同じアルゴリズムで、同じ情報を返すことしかできていない。これは情報検索において非常に大きな限界である。検索に関して市場の声を聞けば、誰もが「キーワードに工夫を凝らさなくとも」、「自分の好み」を理解してくれる「気の利いた」「インタラクティブな」サービス、といった将来像が浮かんでくる。

サーチマーケットシェア (PV)



ビデオリサーチインタラクティブWeb Reportデータをもとに作成。 2005年2月/自宅からのアクセス
 ※Yahoo!/Google/MSN/goo/Biglobe/excite/Nifty/infoseek/fresheye/Navar
 計10サイトにおける検索結果ページの推定PVデータをもとに作成

図-5 日本における検索マーケットPVシェア

均一で押し付けがましい巨大なシステムではなく、使えば使うほど手になじむ検索エンジンである。これは単純に“Apple”が果物かコンピュータであるか、といった同音異義語の問題ではない。Web上・身の回りに「情報」が溢れるようになり、さらに、自分の興味ある分野だけに絞ったとしても、情報が過剰になる時代が到来し、人々がその情報を適切にハンドリングできるように助けなくてはならないということである。

Shared Search

My Mediaとしての検索サービスが「第3世代」と呼べる領域に達するかどうかは、最終的には「Shared Search」が握っている。検索エンジンによって発見された有用な情報は、現在でもメールやインスタントメッセージで関係者にシェアされることで広まっていく、つまり最も有用な情報は口コミにあるが、それは検索サービスにフィードバックされていない。その人の関係者からシェアされる情報を総合することで、My Mediaとしての検索サービスが、ずっと早く完成することになる。Yahoo! JAPANは図-5に示すように日本のサーチシェアの過半数を握り、すでに月間1,300万ID(2005年3月中にログインしたYahoo! JAPAN ID数)がログインするサービスである。新世代のPersonal & Shared Searchの実現に向け、我々のポテンシャルを最大限に活かしていきたい。

人々の生活を豊かにするライフエンジン

Yahoo! JAPANのミッションは人々の生活にあらゆるシーンで貢献する「ライフエンジン」になることである。つまり人々の生活に役立ち、生活を豊かにしたいと考えており、そのために必要な技術はどんどん取り入れていきたい。現場においては、ビジネスと情報処理技術、ビジネスと研究開発の、今までになく広く深い協業体制が求められている。

今後のYahoo! JAPAN 検索サービスにご期待いただくとともに、志を同じにする方にはぜひ我々のチームに参加いただきたい。

参考 URL

- 1) Yahoo! SEARCH (ベータ版), beta.search.yahoo.co.jp
- 2) ダイレクト検索, search.yahoo.co.jp/promo/direct.html
- 3) Yahoo! 知恵袋, knowledge.yahoo.co.jp/
- 4) Yahoo! 商品検索, psearch.yahoo.co.jp/
- 5) Yahoo! 画像・音声・動画検索, beta.search.yahoo.co.jp 参照
- 6) Yahoo! NEXT, next.yahoo.com
 Yahoo! Inc. での新しい検索プロダクトが、実験的なものも含めここで紹介されている。
- 7) Yahoo! Research Labs, research.yahoo.com/
 Yahoo! Inc. にて基礎研究を行っている部門。マシンラーニング、Webデータマイニング等。
- 8) Yahoo! Search “My Web”, myweb.search.yahoo.com/myresults/benefits
 Yahoo! Inc. における検索パーソナライゼーションの第一歩。
- 9) Yahoo! Search “My Web2”, myweb2.search.yahoo.com/
 “Shared Search” の1つの形。SNSをコミュニティプラットフォームとして使い、Taggingにより情報を整理共有する。
- 10) YSDN:Yahoo! Search Developers' Network, developer.yahoo.net/
 Yahoo! Inc. の検索技術をAPIの形で、無料で提供している。
 (平成17年7月14日受付)

