

特集

検索ビジネス 最前線

検索を制する者はネットを制す！

Google、MSN、A9、GETAなど、日米の注目検索サービスを一挙紹介

「ポータル」から「検索」への流れが起きている。そして、その検索対象はウェブコンテンツのみならず、本、ビデオ、店舗、地図などリアル世界のデータへと拡大を始め、そのユーザーインターフェイスもマルチメディアを駆使した使いやすいものに進化してきている。本特集では、従来からの大手検索エンジンのビジネストrendを概観し、それに追い付き追い越そうとする新勢力を、米国勢を中心に紹介する。また、連想検索など、日本独自の研究にも注目した。検索はデジタルパワーがいかに発揮される領域であり、ネットビジネスを行う者には本丸技術となる。世界中で始まったこの次世代検索競争からヒントを読み解いていただきたい。

C O N T E N T S

28

新しいステージに入った検索

30

大手検索ビジネスの動向～過去・現在・未来～

Google、Yahoo!、MSNなど

40

米国・注目検索サービス一挙紹介

50

日本の検索技術

汎用連想計算エンジン「GETA」
行動支援メディアを目指す「gooラボ」

新しいステージに入った検索

佐々木 俊尚 ジャーナリスト

Yahoo!も、最初はディレクトリー型の検索エンジンとして多くのユーザーから人気を得て、現在のポータルNo.1の地位を築いた。

トラフィックの流れはポータルから検索へ

検索エンジンをめぐる企業間競争が激化している。主なプレーヤーはGoogleとYahoo!、そしてマイクロソフトの3社である。

1998年に設立されたGoogleは、ページランクテクノロジーなどの圧倒的な技術力により、それまでの検索エンジンを凌駕し、事実上の標準(デファクトスタンダード)の地位を確立した。そして後にYahoo!に買収されることになるOvertureのビジネスモデルを真似て、キーワード広告のGoogle AdWordsを投入。それまでポータルサイトの「オマケ」にすぎなかった検索エンジンの世界に、収益モデルを作り上げた。そうして多くのトラフィックが検索エンジンに集中するようになり、「検索エンジンこそが、インターネットの集積地である」という考え方がネットの世界に燎原の火のように広がっていったのである。

かつては巨大ポータルサイトがインターネットの集積地の役割を果たし、多くのユーザーはYahoo!やMSN(マイクロソフト)、AOLなどのポータルサイトにまずアクセスし、そこから提示されているサービスラインアップやディレクトリーを経由して、目的のサイトへと誘導されていた。だがこうした中央集権的なネット支配は、世紀をまたぐころから徐々に変容し始めた。Googleの人気の高まるのに連れ、検索エンジンをベースにしてネット上を歩き回るユーザーが増えていったのである。それはたとえばNHKのサイトに行くために、Googleで「NHK」というキーワードを入力するようなやり方だ。Googleが「Google Toolbar」という使いやすいブラウザプラグインをリリースしたことも、そうした傾向に拍車をかけた。

これによって、かつては純然とした情報収集の

ためだけに使われていた検索エンジンの用途は、大きく拡大した。

覇権争いの勃発で拍車がかかった技術開発競争

こうした動きに危機感を抱いたのが、ポータルサイトの王者であるYahoo!とMSNだった。「このままではネットのトラフィックはGoogleにすべて吸収されてしまう」と考えた両社は、相次いでGoogleとの提携を解消し、新たな検索エンジンの独自開発へと踏み切ったのである。

そしてこの3強の戦いは、思わぬ余波を生んだ。競争が激化するのに伴って、検索の対象フィールドがウェブだけでなく、パソコンのデスクトップやオンラインショッピングの商品、映画やビデオなどの動画、世界中のニュース、地図、地域情報などへとどんどん拡大していったのである。このフィールドの拡大は圧倒的で、2004年から2005年にかけては、検索エンジン各社がほとんど毎月のように新たな検索サービスのベータ版をリリースするという驚くべき状況が続いている。

そしてこれに呼応するかのように、3強以外からも検索エンジン戦争に挑戦しようとする企業も続々と現れてきた。たとえばAmazon.comは、A9というテクノロジー企業を新たに設立し、書籍のフルテキスト検索という驚くべき検索サービスを投入して、業界の度肝を抜いた。さらにオーガニックな検索エンジン開発にも乗りだし、秀逸なユーザーインターフェイスを持つ検索エンジンをリリース。検索戦争に本格参入したのである。

またTeomaというエッジな検索テクノロジーを持つAsk Jeevesは、Googleのページランクテクノロジーをさらに進化させ、リンクの「質」を計測する高度な技術を投入し、注目を集めている。

RSS との融合で さらなる変貌を遂げつつある検索

こうした各社の戦いは、さらに大きな動きももたらした。検索エンジンのドラスティックな質的变化

つまり検索エンジンがメディア化していくという動きである。もっと具体的にいえば、それは検索エンジンと情報配信システム(シンディケーション)との融合である。

ここ数年、インターネットのトラフィックは再び大きな変動期を迎えつつある。前出のように、かつてトラフィックはポータルサイトから検索エンジンへと大きく動いた。そして今度は、トラフィックの大きな部分がRSS(Rich Site Summary)へと徐々に動きつつある。

RSSというのは、ウェブやブログのメタデータを構造化してXMLで記述した更新情報で、RSSリーダーなどのアプリケーションを使うことによって、ユーザーは求めるウェブの更新状況を一目で把握することができる。いわば、1990年代に流行したがその後廃れてしまった、プッシュ型情報配信システムの復活ともいえるものだ。

このシンディケーションと検索エンジンとは対立するのではなく、検索エンジンがシンディケーションを包含し、新たなサービスを生み出すようなかたちで進化を始めている。具体的には、このRSSを中心とするシンディケーションと、もともとはプル型のメディアであった検索エンジンが徐々に接近し、ブログのRSSやGoogle Newsのパーソナライゼーション、Amazon.comで人気のあるレコメンデーション(おすすめ)といった新たなサービスが生み出されている。

そしてそんな状況が進むにつれて、検索業界の中からは次のような声も出始めている。

「検索エンジンの再定義が必要なのではないか」

今後、検索エンジンがどのように進化の道をたどり、どこへ向かっていくのか それは、これからの物語である。

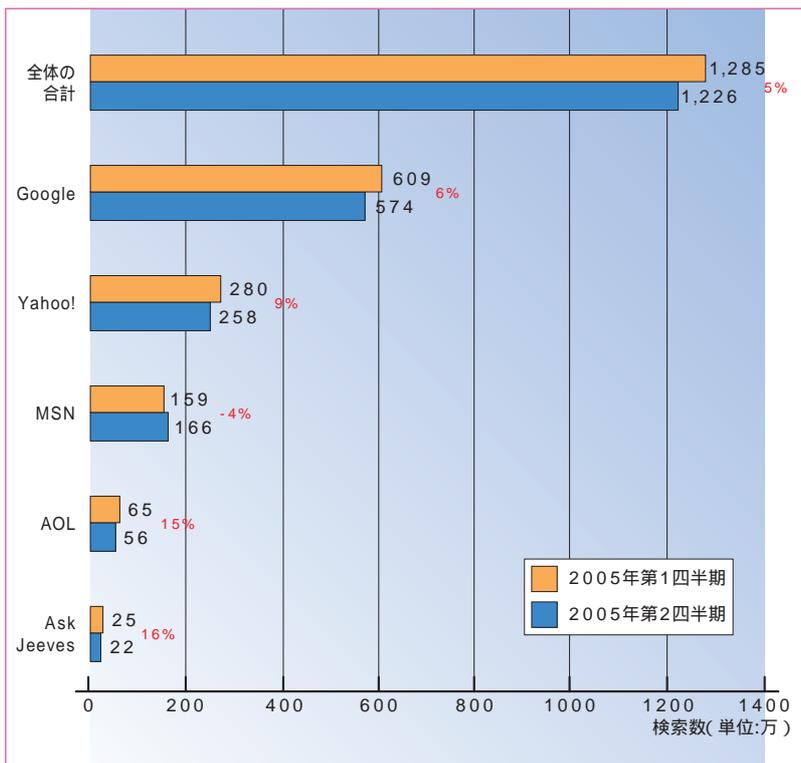


図1 検索エンジン上位5社における2005年上半期の検索数と伸び率。3強と呼ばれるGoogle、Yahoo!、MSNらが1桁台またはマイナスの伸びなのに対し、第2集団となるAOLやAsk Jeevesは大幅な伸びを示している。
(出典: Nielsen//NetRatings MegaView Search, July 2005)

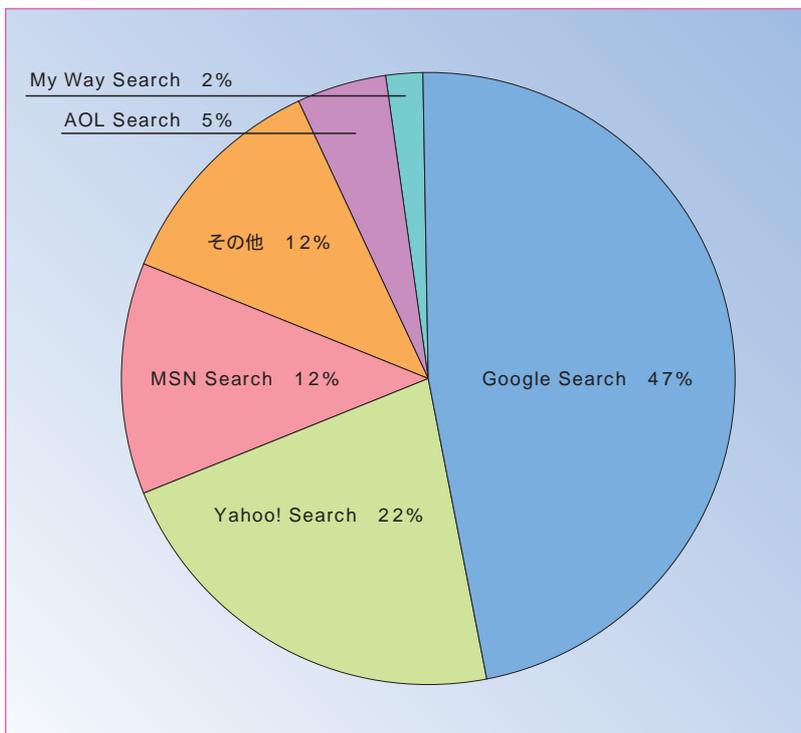


図2 検索エンジンのシェア(2005年7月 米国の家庭と職場)。Googleが圧倒的なシェアを持つが、この1年で本腰を入れ始めたYahoo!とMSNの今後の動向が注目される。ただし日本においては、Yahoo!が約5割でGoogleが約3割と逆転する。
(出典: Nielsen//NetRatings MegaView Search, July 2005)

検索エンジン市場の動向 ～ 過去・現在・未来～

佐々木 俊尚 ジャーナリスト

黎明期には、HotBotという検索エンジンがあった。今のGoogleのように他の検索サイトのエンジンとしてライセンス提供されるなど、人気があった。現在はLycosが運営しており、デスクトップ検索ツールも提供している。

黎明期における 検索エンジンの栄枯盛衰

検索エンジンの歴史をごく端的にいつてしまえば、「Google以前」と「Google以降」に分けることができるだろう。テクノロジー的にもビジネス的にも、それほどまでにGoogleという企業の存在は大きかった。

Googleが登場する以前をざっと眺め渡してみよう。世界で最初のロボット型検索エンジンは1994年1月に有料でスタートしたInfoseekである。有料では顧客はなかなか集まらず、その年の夏には無料サービスに移行している。この同じ年の4月には、後にYahoo!の創設者となるジェリー・ヤンとデビッド・ファイロがディレクトリーの原型となるリンク集を公開している。すべての原点は、この1994年という年から始まったといっていいたいだろう。

しかし黎明期のこの時代、検索エンジンの代名詞のように語られたのは、DECの研究所で開発されたAltaVistaだった。1995年にサービスインしたYahoo!が、公式の検索エンジンとしてAltaVistaを採用したからだ。また1995年10月にサービスを開始したExciteもかなりの人気を集めていた。MagellanやWebcrawlerといった検索エンジンもあった。だがこうした企業の多くは、今ではほとんどが姿を消してしまった。MagellanとWebcrawlerは後にExciteに買収され、そのExciteも2002年に倒産。日本法人だけは生き残って伊藤忠商事の子会社となり、検索エンジンから若者向けのポータルサイトへと変わっている。

またAltaVistaは、Yahoo!がInktomiに乗り換えるのと同時に力を失い、DECがコンパックと合併するとともに切り離され、その後Overtureに買収された。そのInktomiもその後、Yahoo!がGoogleに乗り換えることで破たんの危機に瀕した

が、やがてYahoo!とGoogleが決裂するに及んでYahoo!に買収される結末となった。

世界初の検索エンジンだったInfoseekは、親会社であるディズニーによって閉鎖され、日本法人だけがインフォseekの名称を引き継いで生き残った。日本法人は2000年に楽天に買収され、その後ライコスと統合し、現在は楽天のポータル戦略の一環として生き延びている。しかし、独自の検索エンジン開発はもはや行っていない。

このように検索エンジンの歴史を振り返ってみると、その合従連衡の激しさにあらためて驚かされる。まさに死屍累々、弱肉強食の世界である。

話を戻そう。黎明期の検索エンジンには、2つの大きな競争があった。1つはインデックスを他社よりも巨大化させ、ウェブ世界をいかにして覆い尽くすかというデータベース構築競争である。世界中に存在するウェブページの数は爆発的に増加し続け、当時すでに数十億ページにまで達していた。その爆発に検索エンジンは追いつけなくなっており、検索結果に漏れが大量に生じていたのである。

もう1つの競争は、ユーザーからの検索クエリーに対していかに適切に回答を返すかというアルゴリズム競争だった。初期の検索エンジンはきわめて貧弱なアルゴリズムしか持っておらず、SEO(検索エンジン最適化)スパムと呼ばれる悪質な手口に振り回された。どんなキーワードで検索しても、必ず検索結果上位に特定のポルノサイトやカジノサイトなどが表示されるという、信じがたい事態にまで陥ったこともあったのだ。

そうした混乱の状況が続く1998年、彗星のように登場してきたのがGoogleだった。次の章で詳述するが、Googleはこの2つの競争において、圧倒的な技術的勝利を収めたのである。

Googleの登場が火をつけた 検索ビジネス競争

データベース構築競争においては、Googleは分散モデルを導入することで、データベースのスケラビリティを徹底的に推し進めた。1台のスーパーコンピュータで処理するのではなく、大量のマシンを使う分散システムを考え出したのだ。同社のCEOであるエリック・シュミット氏は後に「高性能CPUではなく、より安価で小さなプロセッサを大量に購入して構築する方が、データベース構築には現実的だ」と述べている(「Always On」のインタビュー記事より)。これら数百台のサーバーから、数万のHTML収集クローラーが送り出され、同時並列にウェブページの収集を行い、これを数万台から数十万台ともいわれるマシンに処理させるというものである。

またアルゴリズム競争においても、Googleはページランクテクノロジーと呼ばれる技術を導入することで、検索結果に圧倒的な公平性と正確さを与えることに成功した。ページランクテクノロジーについてはすでに周知の技術なので詳述はしないが、要するに「人気のあるサイトからリンクが張られているサイトは良いサイト」という考え方を具体化したものである。

この2つの技術によって、Googleは従来の検索

エンジンを凌駕した。Googleがその位置を保ち続けた最大の理由は、同社が圧倒的な技術企業であったからにほかならない。

「Googleの成功の秘密は、口コミでその人気を広まったことにある。ファンたちがGoogleの驚異的な機能について熱狂的に騒ぎ立て、その輪がどんどん広がっていったのだ」(「Sunday Herald Online」紙)

そしてGoogleはその性能を認められ、2000年5月にYahoo!の公式検索エンジンとして採用された。圧倒的なリーチ(到達)率を誇る巨大ポータル、Yahoo!の力によってGoogleは一般社会にも認知され、ブランドイメージを上げていくことになる。

ところがGoogleが力をつけていくのに従い、両社の関係は次第にぎくしゃくしたものに変わっていつてしまう。ポータルであるYahoo!よりも、検索エンジンのGoogleの方が注目されるようになってしまったからだ。

実際、検索エンジンのリーチ率では、GoogleはYahoo!を完全に凌駕している。たとえば現在の数字がどうなっているかといえば、市場調査会社Nielsen//NetRatingsの2005年3月の統計によれば、リーチ率は米国でGoogleが43.1%、Yahoo!が26.0%。イギリスではこの比率が57.0%、14.6%とさらに広がり、また非英語圏のフランスでも61.7%、12.0%と圧倒的な差がついている。

Googleのページランクテクノロジーのアイデアは、創業者のサーゲイとラリーが1998年に発表した論文に記されている。

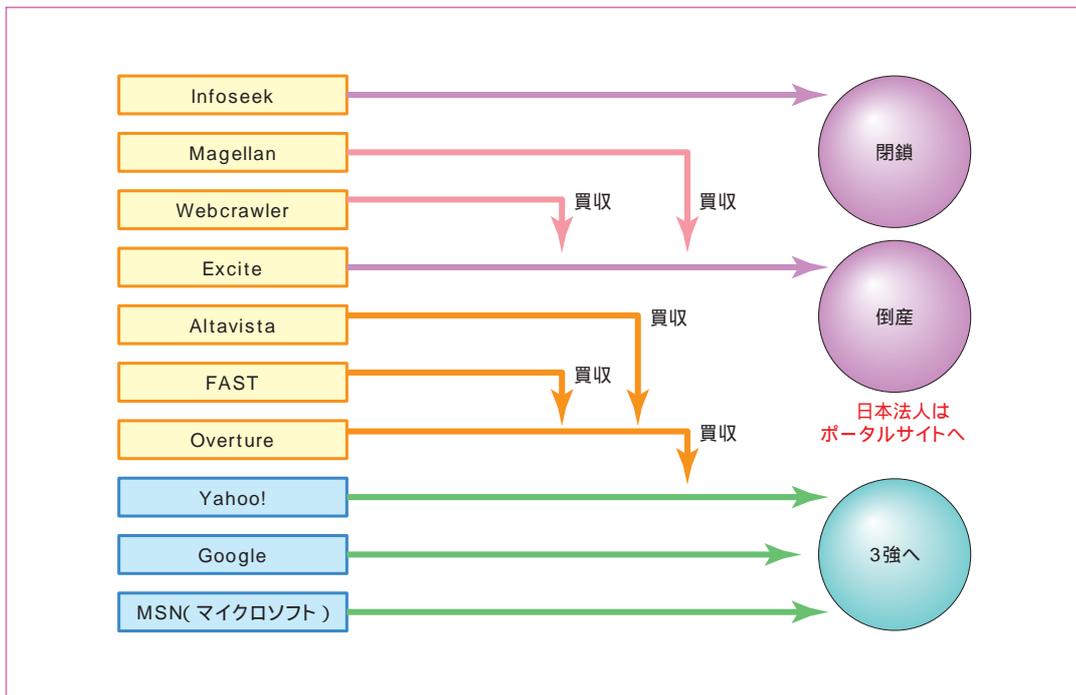


図3 検索エンジン淘汰の過程。他社のユーザーや検索技術を取り込もうと多くの買収・統廃合が行われてきた。

2000年前後は、Yahoo!やMSNに限らず、実に数多くの検索サイトがGoogleをエンジンとして採用していた。

自社のポータルにGoogleを採用したYahoo!としては、この状況は「軒先を貸して母屋を取られる」ことに等しく、看過できる事態ではなかった。

そこでYahoo!は検索エンジンの自社開発の道を選び、その一環としてInktomiを買収したのである。2002年12月のことだった。

さらにYahoo!は、キーワード広告の分野でGoogleと激しい競争を繰り広げているOvertureを、2003年7月に買収した。このOvertureは検索エンジン業界の中ではきわめて重要な立ち位置にある。なぜなら同社は、それまで単なる「ポータルサイトのオマケ」でしかなかった検索エンジンを、立派なビジネスへと仕立て上げたからだ。

Overtureの前身であるGoto.comは、1999年に設立された。このGoto.comが世界で初めて、広告料金を取って検索結果の上位ランキングを販売するという新しいモデルを投入したのだ。このモデルを真似るかたちでGoogleもその後、Google AdWordsというキーワード広告を投入した。さらにGoogle AdSenseというコンテンツターゲット広告も生み出すなど、検索エンジンのビジネス化に邁進していくことになる。

Yahoo!はこのOvertureを買収し、この結果、キーワード広告という魅力のある市場は、Yahoo!とGoogleの2社で覇権を争う構図となった。おまけにOvertureはYahoo!に買収される直前、ルウェーの優秀な検索エンジンであるFAST(Fast Search & Transfer)社の検索部門と、老舗のAltaVistaを買収している。つまりYahoo!はこの買収で、Inktomiを含めて一挙に検索エンジン企業4社を有する業界のメインプレーヤーとなったのである。そしてこれら買収企業のテクノロジーを活用し、2004年2月、ポータルの正式検索エンジンをGoogleから自社開発のYahoo! Search Technology(YST)へと切り替えた。

このGoogleとYahoo!の激しい動きに揺さぶられたのが、マイクロソフトだった。同社のポータルサイト「MSN」は、検索エンジンとしてInktomiとOvertureの2社と契約していた。しかしこの2社がYahoo!の傘下に入ってしまったとあっては、契約続行は「敵に塩を送る」ということにもなりかねなかった。さらにいえば、機を見るに敏なマイクロソフトは、検索エンジンをデスクトップ、ウェブブラウジングに次ぐ第三の波としてとらえていた。



図4 MSN Search。これまで検索ボックスはMSNのトップページに含まれる1コンテンツであったが、Googleの影響なのか新たに単独ページも用意し、非常にシンプルなデザインとなった。この傾向はYahoo! Searchも同様。

ここで逸してしまえば、次世代のメインストリームに乗り遅れかねない。そこで同社は検索エンジンの自社開発へと乗り出し、2005年1月にInktomiから、自社製のMSN Searchに切り替えた(図4)。

ここに、「3強」の激しい戦いが繰り広げられることになったのである。

そしてこの大戦争は、新しい検索エンジンサービスの提供を各社に促し、そうしたトライアルは結果として検索エンジンのカバーする領域を著しく拡大させることになった。検索エンジンの変容が始まったのである。それはパラダイムの転換ともいえるような、ドラスティックな変化だった。

検索領域がウェブからデスクトップへと拡大

Overtureが発明し、GoogleがAdWordsで追隨したキーワード広告の隆盛は、検索エンジンというテクノロジーの世界にビジネスの論理を持ち込んだ。Googleはそれまでのポータルサイトに検索エンジンを提供するOEMビジネスだけでなく、キーワード広告やコンテンツターゲット広告を収益モデルの中核に据えるようになった。

しかし、収益を増やすために単純に広告を増やすだけでは、広告はスパムとなってユーザーからは避けられてしまう。そこでキーワード広告やコンテンツ広告を表示する「場」を増やすことが求められるようになり、その結果として検索のターゲットを拡大させ、「場」の面積を広げていく戦略を検索エンジン各社は採るようになったのであ



図5 Google Desktop Search。ウェブ版と同じインターフェイスにすることで、ユーザーは対象データの違いを意識することなく利用できる。これによって、Googleはユーザー個人のPCをその検索対象として取り込んだ。

る。多くの検索のフィールドが存在し、そこに薄く広く広告を展開していけば、ユーザーにも受け入れられやすいからだ。

従来はウェブのテキストだけを対象にしていた検索エンジンは、動画や商品カタログ、書籍の全文テキスト、電話帳などの地域情報、テレビ番組などさまざまな分野にターゲットを広げていくことになった。そうした面的拡大の象徴ともいえるのは、デスクトップ検索である。

ユーザー個人のパソコンの中身を検索エンジンの対象にすることができれば、検索企業はキーワード広告をこれらのサービス上にも展開できるようになる。もちろん、ハードディスクの中身を検索するという機能は、ウィンドウズXPにもネイティブで組み込まれてはいる。しかし性能としてはきわめて貧弱で、ユーザーの満足できるものではなかった。そもそもファイル名による検索だけでは、多量のデータから目的のデータを探し出すことが難しい(ウィンドウズにも文書内検索の機能はあるが、決して使いやすいものではなかった)。

この分野にいち早く目を付けたのは、やはりGoogleだった。Googleはウェブブラウザベースのウェブ検索とデスクトップ検索を融合させ、情報がインターネット上にあるが、ハードディスク上にあるが関係なしに、ユーザーに対してシームレスに検索結果を返すことのできるユーザーインターフェイスを考え出した。

このツールは「Google Desktop Search」と呼ばれ、2004年10月にベータ版がリリースされた。メールやプレーンテキスト、MS-Officeの文書、



図6 Google 検索アプライアンス。ラックマウントサーバーにGoogleの検索エンジンを載せ、企業内ネットワークの検索環境も提供するGoogle。デスクトップに続き、ファイアウォールの内部も対象に取り込むこととなる。

AOLのチャットのログ、ウェブブラウザの過去の履歴などを検索し、ウェブブラウザの画面上に検索結果を表示するというものだ。Googleはこのアプリケーションの開発に際し、NEC Research Instituteからヘッドハンティングしたスティーブ・ローレンスを投入し、約1年をかけてリリースにこぎ着けた(図5)。

Google Desktop Searchはウェブで検索するのと同じように入力フォームにキーワードを入力して検索ボタンをクリックすれば、それまでに作成されているインデックスを使ってハードディスク内のドキュメントの全文検索を行い、瞬時に検索結果が表示される。きわめて使い勝手の良いインターフェイスを持ち、しかも驚くほど高性能だった。

このGoogleの新しい検索に激しい焦燥感を抱いたのが、マイクロソフトだった。ウィンドウズとOfficeを擁し、デスクトップが自社の牙城であると自負する会社にとって、デスクトップ分野にGoogleの進出を許すというのは、とうてい容認できることではなかった。

そこでマイクロソフトはGoogleに対抗してデスクトップ検索を開発し、ベータ版を2004年12月にリリースした(図7)。ウィンドウズの次期バージョンであるウィンドウズビスタにOSネイティブな検索として、MSN Searchを使うオプションを標準で加えることや、インターネットエクスプローラ(IE)のツールバーに標準でMSN Searchの検索ボックスを表示させることなどを表明したのである。

ウィンドウズのAPIは完全には公開されていないから、マイクロソフトが本気で高性能なデスク

アップルのMac OS X 10.4には、OSに統合されたSpotlightと呼ばれるデスクトップ検索機能が搭載されている。OS標準の検索機能としては、ウィンドウズをリードしたと同社はアピールしている。

日本のジャストシステムも、同社の日本語処理技術を活かしたConcept Searchというデスクトップ検索ツールを提供している。

Windows Desktop Searchは、さすがにOSやブラウザとの親和性も高く、MS-Office 文書の検索オプションも充実している。

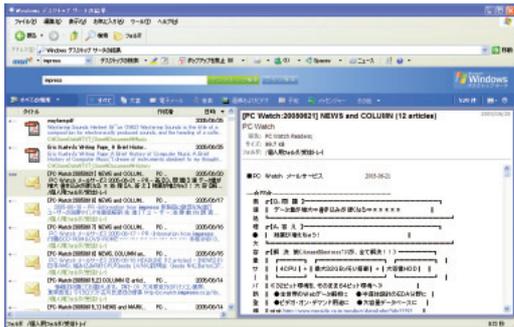


図7 Windows Desktop Search. ウェブブラウザやOS、メジャーアプリケーションを持つ強みを活かし、高度な連携機能を実現している。次期ウィンドウズでの戦略が逆転できるかどうかのカギを握る。

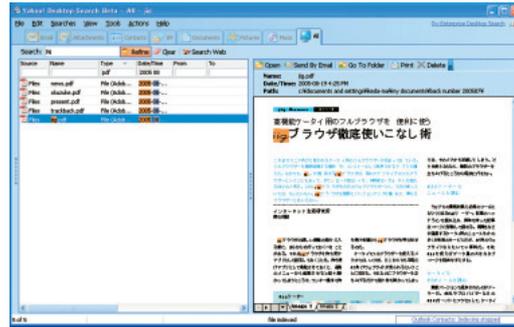


図8 Yahoo! Desktop Search. Google、マイクロソフトに続いて投入してきたが、いまだベータ版で先行する2社に遅れをとっている。他社からのライセンス製品を持ってきた点に、Yahoo!の焦りが見て取れる。

トップ検索の開発に邁進すれば、Googleよりも有利に展開できる可能性がある。しかも米国内での「Search Engine Guide」などの報道によれば、マイクロソフトはMSN Searchを同社のインターネットサーバーソフトウェアであるIIS(Internet Information Services)と連動させることで、通常の検索エンジンがインデックス化しにくいような動的なコンテンツを含んだウェブページも検索できるようにする計画も持っているという。マイクロソフトが市場を押さえている標準的な製品と検索エンジンサービスを連動させていくことで、事態を有利に運ぼうとしているのは明らかだった。

振り返ればこの戦術は、かつてのウェブブラウザ戦争で同社が採用したのとまったく同じといえる。1990年代後半、ブラウザ市場を独占していたのはネットスケープナビゲーターだった。この市場にマイクロソフトは、ウィンドウズにIEをバンドルするという戦術を採ったのである。この手法は業界の強い批判を浴び、独占禁止法違反の裁判にまで持ち込まれた。だが判決が出るころには、IEの攻勢でネットスケープはすっかり弱体化し、市場の大半を失った挙げ句に最終的にAOLに吸収合併される結果となった。

この戦術が果たしてデスクトップ検索でも有効に作用するのかわかりませんが、まだわからない。だがこの分野で出遅れていたYahoo!も2005年1月に「Yahoo! Desktop Search」のベータ版をリリースした(図8)。このツールはデスクトップ検索専門のX1 Technologiesからのライセンスを受けたものである。MS-OfficeやPDF、Photoshopなど200

種類以上ものファイルをインデックス化できるなどきわめて高機能で、先行するGoogleやマイクロソフトを凌駕しかねない。この分野でも3強の三つ巴の競争が続いており、今後も検索エンジンの主戦場の1つとなっていそな雲行きだ。

より検索とビジネスとを結びつける商品検索

検索エンジンの対象を拡大させる動きとして、もう1つ注目されるべき大きな分野がある。それは商品検索だ。ウェブ全体を検索するのが通常の検索エンジンだが、商品検索ではオンラインショッピングサイトで販売されている商品だけに絞って検索する。それはただ単に検索エンジンのサブセットということではない。検索という行動がダイレクトに購買へと結びつくため、アフィリエイト的な手法と融合させれば、検索エンジンのビジネス化の1つとして大きなマーケットを生み出す可能性がある。この分野で先陣を切ったのも、やはりGoogleだった。

Googleは2002年のクリスマス商戦というきわめて早い時期に、Froogleと呼ばれる商品検索エンジンのベータ版をリリースしている(図9)。Froogleはオンラインショッピングサイトの商品情報を収集し、商品の名前を検索すると、商品の写真と価格を並べて表示してくれる。リンクをクリックすれば、その商品を販売しているサイトに誘導されて、そこで商品を購入することができる。価格で検索を絞り込むことも可能だ。



図9 Froogle。Googleを商品を探すために使っているユーザーは多いが、それを踏まえて商品検索と価格比較に特化した検索エンジンとして提供している。当然、より広告ビジネスとしての可能性が高い領域だ。

興味深いのは、このサービスが特定のオンラインショッピングサイトと提携して行われているのではないことだ。検索して表示される商品は、Googleが収集している他のウェブサイトと同様、Froogleが世界中のオンラインショッピングサイトからクローラーを使って自動的に集めてきた商品データなのである。特定のサイトと連携して垂直統合を行うのではなく、オープンなシステムを流通業界に対して提供するというモデルをGoogleは選んだのだ。

GoogleはこのFroogleと同じ時期に、宅配便追跡サービスもリリースしている。宅配便の荷物以外にも、飛行機の登録番号や特許番号の検索、さらには空港の離発着状況を表示する機能も備えられており、このサービスとFroogleを統合すれば、Eコマースに対するトータルソリューションをGoogleは提供できることになる。つまりは商品検索の巨大なオープンポータルを構築し、膨大な数に上るGoogleのユーザー層を取り込んでいくという新たなビジネスを作り上げることが可能になるということだ。

そしてこの分野でも、Yahoo!はGoogleに追随した。2003年9月に、商品の価格比較を行える「Yahoo! Product Search」のベータ版をリリースしたのである。そしてマイクロソフトもかなり遅れて、2005年3月にショッピング検索の「MSN Shopping」ベータ版を公開した。

だがこうした3強の動向よりも、もっと驚くべき動きを見せたのは、突如として現れた第四の勢力オンライン書店Amazon.comの試みだった。

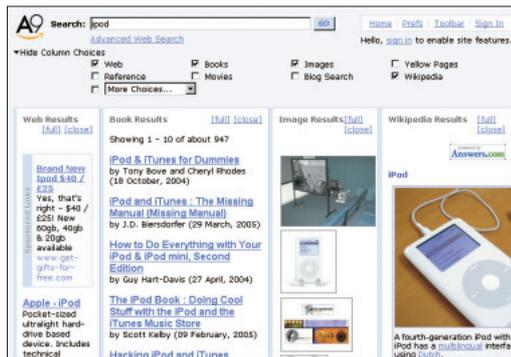


図10 A9。Amazon.comが商品検索サイトとして投入した。Amazon.comに誘導するだけではなく、あくまでもショッピング検索として、販売店舗情報や地域情報なども提供し、通常の検索エンジンとしても高度な機能を備える。

Amazon.comは2003年10月、A9という子会社を設立し、独自の商品検索テクノロジーのリリースを行うと発表したのである(図10)。そして設立直後の10月下旬には、Amazon.comで販売されている書籍のフルテキスト検索をスタートさせ、業界をあっといわせた。

A9のフルテキスト検索は「Search Inside Book」という名称で、スタート時には出版社190社の12万冊、計3300万ページという膨大な量が検索対象になっていた。これをすべてスキャンし、テキスト化させたのである。

さらにA9は、2004年9月に本格的なオーガニック検索エンジンもリリースした。検索結果をウェブやイメージ、書籍、映画などに分類し、それぞれを伸縮可能なコラム形式で表示するというユーザーインターフェイスを持つ。さらにユーザーのウェブブラウジング履歴を記録し、その履歴の中から検索できるといったパーソナライズ機能も持つなど、秀逸なトライアルが満載の検索エンジンである。

もともとAmazon.comのマーケティングテクノロジーには定評があった。たとえば特定の本をサイトで表示させると、「この本を買った人はこんな本も買っています」とおすすが表示されるBook Matcherというフィルタリング技術は、同社の特許にもなっているほどである。そしてA9設立の目的について、Amazon.comの広報担当者は米メディアの取材に、「単なるオンラインショッピングサイトから、ITサービスを包括的に提供する企業に進化するためだ」とコメントしている。つまりは

「Froogle」というのは耳慣れない単語だが、「frugal(質素、儉約)」と「Google」をかけたあわせた造語だという。

A9は、その機能やAmazonの戦略という面で注目されるが、開発者の間ではAjaxによって構築されているという点でも話題となった。Google Maps同様、軽快な表示はAjaxによるもの。

Google Print で使われている書籍ページの画像は、キルタス社が開発した自動ブックスキャナを使ってデジタル化されている。1ページずつページをめくり撮影するという気の遠くなる作業が行われている。

Google と同じように、E コマースのトータルソリューション提供企業へと脱皮を図ろうとしているのである。

初めてテクノロジー分野で他社に先を越された Google も、この年の暮れからフルテキスト検索の開発を開始した。そして Amazon.com に遅れること約1年後の2004年10月、Google Print というサービスをスタートさせたのである。検索キーワードとマッチする言葉を含んだ書籍を検索し、検索結果表示ページにその書籍へのリンクが表示されるというものである。リンクをクリックすると、該当キーワードが含まれているページが表示され、同時にその本がどこで購入できるかという情報も表示される。

検索ビジネスは3強を中心にさらに激しく展開

検索対象の拡大という局面においては、Google がまず先に新たなサービスを投入し、それに他社が追従するという構図だった。テクノロジードリブンで成長してきた Google は、インターネットユーザーが驚喜するさまざまなサービスをベータ版として提供してきた。これまで紹介したデスクトップ検索や商品検索だけではない。ニュース検索 (Google News) や地域検索 (Google Local)、地図検索 (Google Maps) など、数え上げればきりがなく (図 11)、この6月には Google が昨年買収した Keyhole の 3D 地図ソフトの新版 Google Earth、7月には月面の探索ができる Google Moon を発表し、その斬新さに人びとを驚かせ、そして楽しませた。Google のこうしたトライアルが結果的に、検索エンジンがカバーする領域を次々と拡大させてきたのである。

そうした Google の先鋭的な戦略は、同社の独特の経営スタイルによって保たれてきたといってもいいかもしれない。Google は通常の企業のように役員会が率いる構成を持たず、技術者が経営を主導した。「戦略や新製品を展開する主導権を握っているのは、2人の Google 創設者か、もしくはごく少数のメンバーで構成している技術チームだ。従来の企業のような決定プロセスは採用していない。おかげで Google はきわめてスピーディーに改革を進めることができるし、これがわが社の



図 11 Google Maps。地図検索はこれまで国内でも存在したが、衛星写真を使った表示によって多くの注目が集まり話題となった。API を公開することで派生サービスが数多く生まれている点も見逃せない。

最大の強みになっている (E・シュミット氏の同インタビュー記事より)

だが同社には2004年、大きな転機が訪れた。同年4月に、新規株式公開 (IPO) を米証券取引委員会 (SEC) に申請したのである。幹事会社はモルガン・スタンレーとクレディスイス・ファーストボストンという名門証券2社である。そして調達する予定の資金は、約27億ドル。米国のインターネットの会社として、過去最大のIPOだった。公開時の時価総額は250億ドル (約3兆円) にも上り、新たな巨大企業の誕生を見せつけた。

だが Google はこのIPOによって、新たな難問も抱え込むことになった。上場した公開企業となると、常に四半期ごとの収益を開示し、株主の期待に応えて株価を上げ続けなければならない。株主からは「収益の上がるビジネスをすぐに実行すべきだ」という圧力がかかり続ける。つまりは前出のようなベータ版サービスを、すぐに収益の上がるビジネスへと作り直すように求められることになるのである。

公開企業に対するこうしたプレッシャーが、先進的なテクノロジー企業である Google の文化にどう影響するのかは、まだ分からない。

マイクロソフトのMSN担当副社長であるユスフ・メディー氏は2004年6月、ゴールドマンサックス主催のカンファレンスで「今やMSNの全精力は、新しいMSN Searchに傾けられている」と宣言している。またYahoo!も傘下のOvertureをYahoo! Search Marketing Solution (YSMS) に改称させるなど、統一ブランドの確立に邁進し、

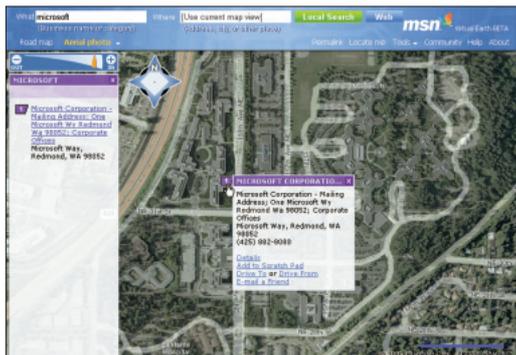


図12 MSN Virtual Earth。Google Mapsにわずかに遅れて公開された、マイクロソフトの地図/地域情報検索。こちらも衛星写真が使われているほか、Google Mapsにはない機能も備える。



図13 Ask Jeeves(Ask.jp)。自然言語検索をウリに登場し、その後Teomaを買収したりIACに買収されたりしながら第二集団の位置につける。2005年に日本市場に本格参入し、日米ともに高いユーザーの伸び率を見せている。

Googleを激しく追い上げている。

これら技術競争を繰り広げているライバル社の三つ巴の戦いは、今後も激化の一途をたどりそうな勢いである。

独自の手法で3強の追撃を開始した Ask Jeeves

Google、Yahoo!、MSN(マイクロソフト)という検索エンジン3強。これに対し、Amazon.comやAsk Jeevesのようなエッジなテクノロジーを持つ新興パワーも出現しつつある。

Ask Jeevesの誇るTeomaテクノロジーは、純粹な情報検索というきわめてオーガニックな検索エンジンを志向している(図13)。

Teomaの考え方は、「インターネットの世界は、同じテーマや話題を共有するウェブページの集合体によって構成されている」というものだ。だからTeomaの検索結果は、単にURLの一覧だけでなく、求める検索結果のウェブページがどのような集合体に属している、その集合体がどのようなエキスパートやオーソリティーによって主導されているのかを表示させる仕組みになっている。これはGoogleのページランクテクノロジーを一步推し進めたコンセプトであり、情報のリサーチという分野に特化した検索エンジンといえるだろう。

ウェブ世界全体を広範囲に収集し、だれもが満足できる検索結果の提供というコンセプトにおいては、現状ではGoogleやYahoo!が抜きんでており、新興勢力の参入する余地は少ない。

しかし検索エンジンは利用人口が増え、そのフィールドが爆発的に拡大するのに従って、セグメントも分かれつつある。たとえばYahoo!のリサーチラボは今春、Mindsetというサービスをリリースしている。検索結果について、広告的な要素と情報収集的な要素のバランスをユーザーが選べるといったものだ。開発者のブログには、「検索エンジンを使って物を買いたいときもあれば、情報を集めたいときもある。しかし一般的な検索エンジンの検索結果では、物を購入するためのサイトのデータと、情報収集に役立つようなサイトのデータは渾然と表示され、自分の求める情報がどこにあるのかを探すのが面倒だ。Mindsetなら、そうした課題を解決することができる」と書かれている。

Ask Jeevesの試みは、この「情報収集」というセグメントに集中的に切り込もうとするものだ。そして、すでに紹介したAmazon.comやFroogleなどの商品検索は、「物を購入する」というセグメントに特化した検索エンジンなのである。

ブログによってさらなる進化を遂げる検索

さらにここに来て、検索エンジンをドラスティックに変容させる大きな動きとして、インターネットにおける情報のシンディケーション(配信システム)との融合がある。具体的にいえば、旧来のプッシュ型の情報配信システムと、もともとはプル型のメディアであった検索エンジンが徐々に接近し、ブログのRSSやGoogleニュースのパーソナライ

日本では、テレビCMで話題になった「一発検索」のAsk.jpだが、そのエンジンとして使われているTeomaの技術は、Googleのページランクに匹敵するものとして評価されている。

日本では、2004年にブログサービスが次々と登場したが、そのおかげでブログ検索の利用価値も高まってきたといえる。商用化されたものには、ライブドアの「未来検索」やドリコム「News & Blog Search」などがある。

ゼーション、Amazon.comのレコメンデーションといった新たなサービスを生み出しつつあるのだ。

この背景には、ブログの隆盛という大きな波があるのは間違いない。ブログと、そのコンテンツを配信するRSSという仕組みが登場したことで、インターネットのトラフィックは従来の「ポータルサイトからのプル型閲覧」「検索エンジンを使ったプル型閲覧」という2つの経路に加え、「RSSリーダーを使ったプッシュ型閲覧」という経路が新たに加わったのだ(図14)。そして検索エンジン業界の新興パワーの中からは、このブログとRSSを積極的に検索に取り込み、融合させようとする動きが続々と生まれている。

たとえば、ブログのリアルタイム検索エンジンとして知られる Technorati もその1つ(図15)である。ブログに書かれたエントリーを、ほぼリアルタイムで検索することができる。検索結果には「リンクを見る」というボタンもあり、その検索結果がどこからリンクが張られているのかを調べることも可能で、それぞれのサイトへのつながりや信頼性を無限に追いかけられる仕組みになっている。またURLを検索フォームに入力すれば、そのURLにどこからリンクが張られているのかを一覧表示することもできる。インターネットのコミュニティ性に注目し、コミュニティに特化した検索エンジンとなっているのだ。

Technoratiは日本でもデジタルガレージとの提携によって、7月1日から日本語サービスを開始しているが、同社のブログ&モバイル事業戦略室副室長の竹内崇也氏は「ブログというのはカンパセーションであるというコンセプトが中心となっている。あるできごとやニュース、話が水滴が落ち



図15 Technorati。ブログ検索エンジンとして先陣を切って登場し、その特性を活かしたリアルタイム性を武器に、独自のニーズを掘り起こしている。旬な話題や商品評価の調査などマーケティング分野への応用が期待されている。

るように始まり、そこから波紋のように広がっていく。それがどのようにカンパセーションされていくのかを、Technoratiは見せている」と説明する。

このテクノロジーを使って、Technoratiは検索結果を1つのコンテンツとしてOEM供給すると同時に、ブログデータをもとにしたリサーチソリューションを提供するビジネスを実現している。企業が新商品を発売したり、キャンペーンをスタートさせたりするなどのアクションを起こした際、そのアクションがどのようにしてブログの世界で波紋となって広がっていくのかということモニターして、マーケティングの見地から分析することが可能になるというわけだ。従来のGoogle型検索では、インデックスの更新にそれなりの時間がかかるため、その「波紋」を時間軸で分析していくのは難しかった。だがTechnoratiはリアルタイムにブログを検索することで、波紋の波及速度までも計測することが可能になっている。いわばバイラ

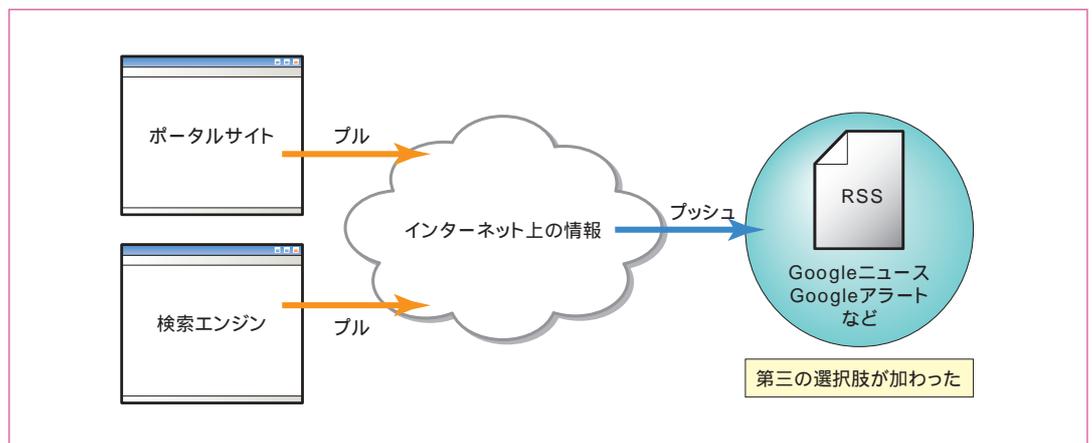


図14 これまでのプル型閲覧にプッシュ型閲覧が加わることで、インターネットの新しい機能やサービスが生まれてきた。

ル(クチコミ)マーケティングを検索エンジンと組み合わせさせたようなサービスとなっているのである。

東京工業大学精密工学研究所の奥村研究室で開発されているブログ検索エンジン blogWatcher も、アプローチは異なるものの Technorati と同様のコンセプトを持っている(図16)。

blogWatcher はブログを、Movable Type などのブログツールで作成されたものだけでなく、日本に1990年代からあったウェブ日記なども含んだ「個人による時系列による記述」としてとらえている。このため他のブログ検索のようにRSSフィードを検索するのではなく、Googleと同じようにクローラーを放ってHTMLを収集し、そこから独自のアルゴリズムで「ブログ的なページ」を抽出するという方法を採用している。

検索エンジン自体は当初はGETA(50ページ参照)、現在はLuceneを使っている。奥村学助教授は言語処理やデータマイニングの専門家で、検索エンジンによってインデックス化されたブログデータをさまざまにマイニングしている。

バースト度(注目度)は、特定のキーワードがブログスフィア全体でどのような出現頻度になっているのかを調べ、グラフとして表示する。ウェブ上でその話題が「いつ」「どのくらい」注目されたかを調べることができる。また評判検索は、特定の製品などについてブログでどのように言及されているかを自動的に解析し、製品の評価に関する部分のみを抜き出すというものだ。ポジティブな評価を赤色、ネガティブな評価は青色でハイライト表示し、これらを集計して評判の推移をグラフ化してビジュアル表示もできるようになっている。このあたりは言語処理のテクノロジーを駆使しており、



図16 blogWatcher。これまでは検索用のインデックスとして収集されてきたウェブ上の情報は、分析されることでさらに新しい価値をもたらそうとしている。

検索エンジンの可能性を実感させるサービスとなっている。

blogWatcher のテクノロジーが実現しているサービスは Technorati と同様、マーケティング的にも非常に興味深いものだ。実際、奥村研究室はソフト開発会社のホットリンクと提携し、同社によってblogWatcher は製品化されている。

奥村助教授は「ウェブ全体を対象にする検索エンジンでは、Googleの1人勝ちが続いているが、ブログなどのニッチな市場では、用途に特化した検索がさまざまに出現してきている。今後の検索エンジンはそうした分野に可能性が秘められていると思う」と話す。

実際、検索した結果をさらにマイニングするという分野はGoogleのウェブサービスなどによっても注目されている新分野で、このフィールドではまだまだ新たなテクノロジー、新たなビジネスが登場してくる可能性は高いといえるだろう。

世界中のウェブを対象にした検索では大分があるものの、特定分野や日本語という言葉に限ると、まだまだ国内の検索技術が活躍する場はあるといえる。

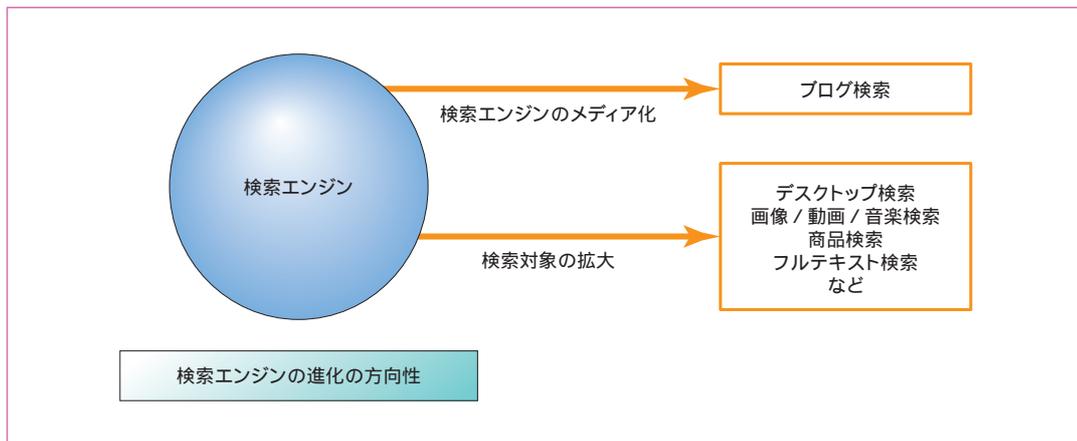


図17 検索の進化は、「メディア化」と「対象領域の拡大」という2つの大きな流れに向かっていく。

米国・注目検索サービス一挙紹介

瀧口 範子 / トーマス・スパーク (Kanabo Consulting Inc.) / 金松 洋子 / 編集部

Google <http://www.google.com/>

検索においてトップに立つ Google。ここ最近の動きとしては、インタラクティブ地図開発の新興企業 Keyhole を買収するなど地図技術の統合を進め、地域情報と地図の合体による新しい情報プラットフォームの確立を狙っている。Google Maps(地図検索)、Google Desktop Search(デスクトップ検索)、Image Search(画像検索)、Google

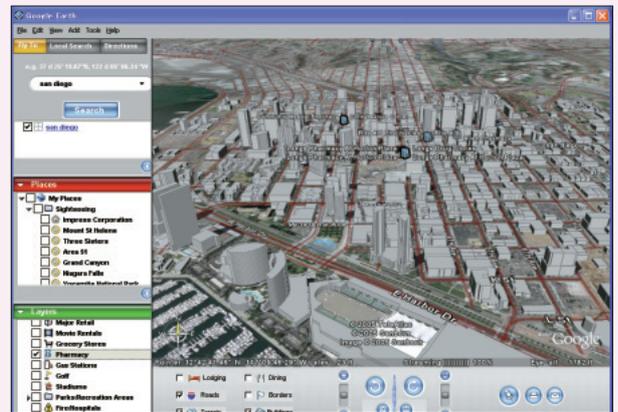
操作が軽快にできる衛星写真の Google Maps

Video(動画検索)、Google Print(ページイメージ付きの書籍検索)など、常に他社に先駆けて新しいサービスを提供することでリードし続けている。1998年の創設時から検索キーワード入力用のボックスを中心としたシンプルなサイト作りを行っていたが、数か月前にサイトをパーソナライズできるツールを発表。気に

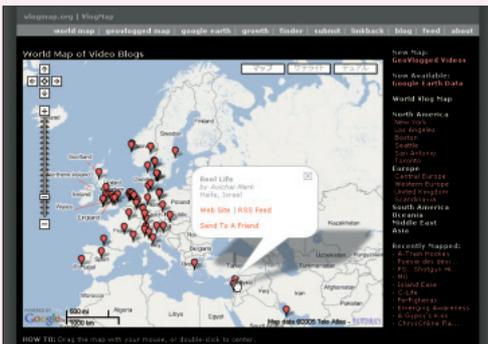
入ったリンクをカテゴリー化し、さらにメールや金融情報などよく使うリンク、ニュースなどを1つのページにまとめられるようにした。また個人の検索履歴を参照して検索順位を決定する技術も統合した。このパーソナライズツールによって、Yahoo!などのポータルにも対抗するポジションが明らかになってきた。



Google Maps。衛星写真による表示が話題になったが、シンプルなインターフェイスと軽快な動作という「Googleらしさ」も人気の理由である。
<http://maps.google.co.jp/>



Google Earth。買収した Keyhole のソフトを Google が改良して公開した。Google Maps よりも多機能で、3D 表示などが可能。
<http://earth.google.com/>



Google Maps で注目すべきは、API を公開することで、誰でも地図データの上に独自の情報を重ねて表示できる新しいサービスを生み出せるようにした点。Video Blogger の居場所を示した「vlogger Map」(左)とユーザーが写真などを地図上にマッピングできる「はてなマップ」(右)。

Yahoo! Search <http://search.yahoo.com/>

My Web のパーソナライズ機能に注目

ポータルとして圧倒的なシェアを誇り、日本では検索エンジンのシェアでも Google に差をつけている Yahoo!。2004 年 2 月に検索サービスを Google から自社開発技術による Yahoo! Search Technology (YST) に切り替え、総合的なメディアサイトとしての強みを生かし、そのコンテンツやサービスを流用したソリューション型の

検索結果を提供する。現在ベータ版としてアップされている開発中のサービスに「My Web」がある。これは、これまでのブックマークに代わって特定のウェブページとリンクを Yahoo! のサーバー上に保存できるもので、その保存と検索が「My Web」からワンクリックで行えるうえ、そのファイルの共有もできる。

また、「context search」というツールは、自然言語での検索を可能にしたもので、ページごと、あるいは数単語ごとに、その内容に関連した情報をワンクリックで検索して小さなスクロールボックスにリストアップする。今後究極のターゲット広告に利用される可能性もある。



context search で、Yahoo! News から関連情報を検索した結果。「Search Related Info」をクリックするとスクロールボックスが表示される。



Yahoo! Audio Search で検索した結果。音声検索は日本語でもベータテスト中。画像や動画検索はすでに対応している。

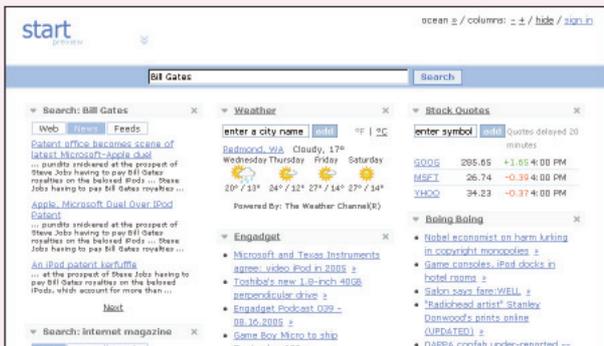
MSN Search <http://search.msn.com/>

Virtual Earth 衛星写真の解像度の高さ

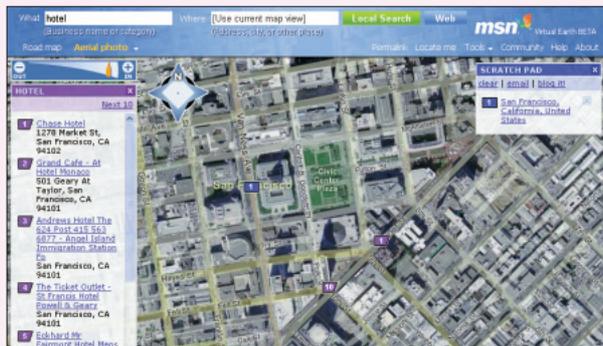
MSN のインデックスページ数は 50 億を超える、この数は Google を超えると同社は語る。検索結果には、エンカルタの 150 万項目の情報に加え 4 万項目のその他の事実情報を統合して的確性を加えている。Google に追隨してデスクトップ検索ツールと地図・地域検索サービスを投入した。

また、自社内のインキュベーション企業である Start.com では、Google 風のシンプルなインターフェイスの入力ページを実験的に開発した。また、地域情報の検索にも力を入れており、現在ベータ版で提供されている Virtual Earth では、検索結果のリストとともに、その店舗や企業の地図が呼び出

される仕組み。衛星地図の統合にも積極的に取り組んでいる。MSN は検索エンジン市場への進出では明らかに乗り遅れたが、ウィンドウズ OS による高い市場占有率、メディア、ニュース、メールなどを含めた全方位的な MSN のリソースをどう検索に統合させていくかに注目が集まっている。



マイクロソフトが次世代検索技術の研究として位置づける Start.com。検索ページをカスタマイズすることで、個人的な情報がポータルとして構成できる。
<http://www.start.com/myw3b/>

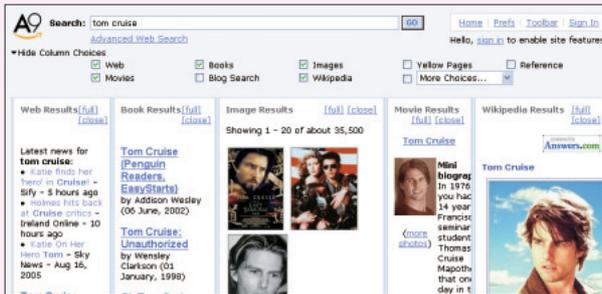


Virtual Earth でサンフランシスコのホテルを検索したところ。地図をドラッグしてスクロールすると、周辺情報も追隨して更新される。
<http://virtualearth.msn.com/>

ショッピングサイトのAmazon.comが開発した検索エンジンで、2004年4月にスタートした。検索技術にはGoogleを導入している。キーワード入力ページはシンプルだが、検索結果のリストは縦に分割された欄に上記の検索エリア別の結果が並列される。検索エリアは、上記以外にも200種類以上のカテゴリーから選ぶことができ、

自分用にパーソナライズした検索結果の組み合わせを呼び出すことができる。イエローページ欄には、特に力を入れており、検索結果に関連した店舗や会社の住所、電話番号のリストとともに、写真と地図も呼び出され、その1つをクリックすると同じ通りにある他のビジネスもリストアップされる。写真はクリックでスクロールし、そ

の通りを歩くように店先の風景を眺めることができる。A9では現在、このイエローページのために全米の主要都市の通りからの風景を動画に収めており、ローカル情報と店舗写真を組み合わせるデータベースはますます増える予定だ。店舗や会社自体、一般消費者も写真やコメントを寄せられる仕組みを作っている。



ウェブ、書籍、画像、映画、Wikipediaを対象に「tom cruise(トム・クルーズ)」で検索した結果。すべての結果が並列表示される。



テスト中の地図サービス。かなり詳細な街並みの写真が表示されるが、個人を特定できるレベルであるためプライバシー問題も懸念される。

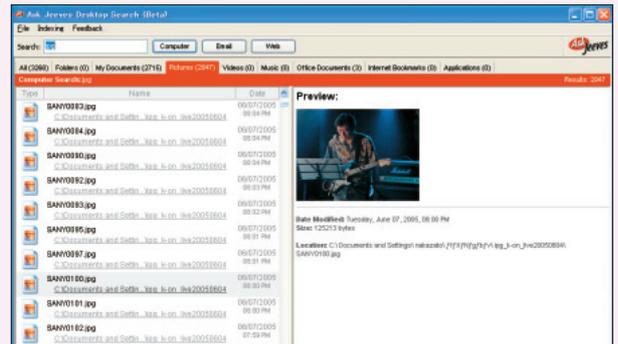
1990年末に自然言語で検索ができるサイトとして登場し、注目を集めた。その後、2001年にTeomaを買収し、クローラーによるインデックスから検索結果をリストアップする方法に転換。対象となるウェブの領域を限定し、適合性によって検索領域を特定する技術で、適切な回答を出す確率を高めているとされる。先頃、「Web

Answer」という、自然言語で入力された検索のクエリーに対して、リアルタイムでウェブを検索して検索結果を出す機能を統合したが、これはAsk Jeevesの強みである自然言語検索の対象を利用したもの。このほかにも検索結果を絞り込んだり、逆に関連性のある情報に拡張する「Zoom」といった機能(一種の連想検索)もある。

Ask Jeevesは、2005年3月にIAC(InterActiveCorp)に買収された。IACは傘下に旅行サイト、地域情報、コミュニティーなど多様なインターネット企業を抱えており、そうした企業が持つリソースとの相乗作用は大きい。最近も、世界屈指のブログRSSリーダーであるBloglinesも買収し、検索サービスへの統合を進めている。



「beatles」と検索すると、右側には「Rolling Stones」や「John Lennon」など、キーワードの拡大や絞り込みのリンクが表示される。



Google、Yahoo!、マイクロソフト同様に、Ask Jeevesもまたデスクトップ検索ツールを提供している。

Blinkx <http://www.blinkx.com/>

動画・画像・音声データ検索の精度の高さ

デスクトップに常駐したソフトウェブがウェブとローカルのPCを自動検索するため、必ずしもキーワードを入力する必要がない。同ソフトが画面上で閲覧している文書を把握しておき、ユーザーの文書や受信メールのフォルダー、ウェブの中から検索事項に関連した文書を提示する。検索結果は、対象の背景にあるコンセプト別に類別し、そこから絞り込まれたコンテンツのみを関連度の高い順に表示する。ウェブページの単語を入力したりハイライトにしたりするのとは

ほぼ同時に検索結果が提示される。また、同社の検索クローラーでは、インターネット上の音楽やビデオファイルに加え、電子メールの添付ファイルも検索の対象とする。音声・ビデオ画像の検索では、正確度を高めるため、高性能な音声認識技術を利用している。現在、同社製ソフトおよびサービスはいずれも無償提供されており、正式なビジネスモデルは公表されていない。



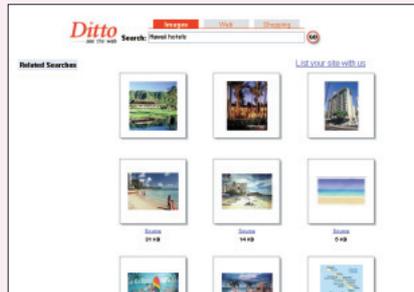
エリック・クラプトンに関するウェブページを開いているときに動画ボタンを押すと、内容を解析してE・クラプトンに関するクリップが表示される。

Ditto <http://www.ditto.com/>

サムネイル画像で表示される検索結果一覧

画像とそれを含むウェブサイトを検索し、画像のサムネイルによって結果として表示する。インターネット上に画像の数が増えるに従って、テキストだけの検索結果リストに加えて画像が果たせる役割が大きいくというアプローチのもとに開発された検索エンジンである。サイトのクローリングによって画像やイラストなどを選んでランク付けしたインデックス化を行い、ユーザーの検索に対しては人間の手による編集も合わせたフィルタリングを行っている。ただし、現在

のところは、検索結果表示に画像以外の情報がほとんどなく、クリックするまでサイトの内容がわからないという短所もある。収益源は、検索結果に合わせた広告表示、ショッピング情報におけるクリックスルー、他のサイトへの画像検索技術供与など。Dittoの前身となる Arriba は、無断で写真を使用したとして写真家から訴えられた過去を持つ。サムネイルの使用は無罪とされたが、フルサイズでの使用は著作権侵害にあたるとして罰金を払い渡された。



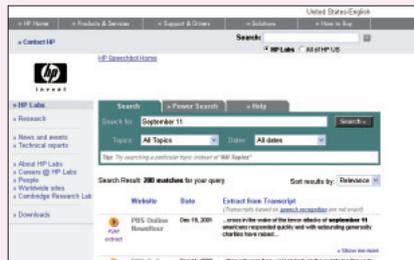
「Hawaii hotels」で検索すると、関連した画像がサムネイル表示される。画像以外の情報はクリックして見る。

Speechbot <http://speechbot.research.compaq.com/>

音声認識技術による動画・音声の文脈検索

音声認識技術によって番組をテキスト化し、それをインデックス化している。米国の人気テレビやラジオ番組のウェブサイトから、音声とビデオが検索できる。現在二十数番組の約15000時間分のコンテンツを持つ。検索はキーワード入力によって行う。音声認識技術はまだ完全には正確にテレビやラジオで語られた文章を読み取れないが、

強調される言葉や繰り返し出てくる言葉などを認識することによって、求めるキーワードが位置する文脈を検索できるという。もともとは、ヒューレット・パッカードに買収されたコンパック・コンピュータの研究所が始めた研究。現在も研究プロジェクトとして存続している。



「September 11」を検索すると、この言葉に言及された番組一覧が表示される。リンクをクリックすると音声ファイルがダウンロードできる。

Singingfish <http://search.singingfish.com/>

豊富なメタ情報で検索できるテレビ番組検索

ファイルのメタデータをもとに、テレビやラジオのニュース、ウェブサイトから、ビデオやオーディオファイルを検索する。メタデータにはテキストで出演者、タイトル、内容、ファイルのサイズなど70種類の記述が行われている。検索できるファイル形式は、MP3のほかに、QuickTime、Real、Win-

dows Mediaなど。ビジネスモデルは、検索結果に合わせた広告掲載、ペイドコンテンツ、メディア企業向けの社内検索エンジン統合サービス(月々の検索ボリュームとカスタマイゼーションの度合いによって課金)など。2003年にAOLによって買収されている。



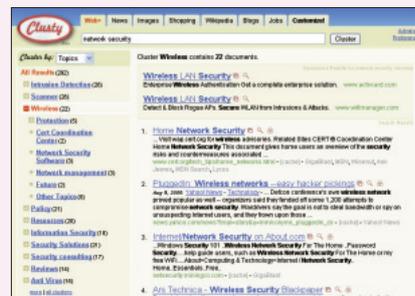
「Paris Hilton」を検索すると、香水のプロモーションビデオやテレビ番組などへのリンクが表示される。

Clusty <http://clusty.com/>

クラスタリングエンジンによるカテゴリー分け

複数の検索エンジンの技術を用いるメタサーチ(Looksmart、Lycos、MSN、Open Directory、Teoma(Ask Jeeves)、GigaBlast、Wisenetを利用)。メインの検索結果は、複数の検索結果を統合したうえで、Clusty自身のヒューリスティックなランキングによって表示される。その際、どの検索エンジンによる結果なのか出所も明らかにされる。Clustyの大きな特徴は、Vivisimoが開発したクラスタリングエンジンを利用して、検索結果を分野ごとにファイルに分けて表示すること。これは検索結果

ページの左側に配置されている。クラスタリングエンジンは、テキストの内容や言語的な類似性によってウェブページをまとめており、自動分析に人間の手を加えた分類手法を採る。検索結果を見るいくつかの便利なツールがあるのも注目される。検索結果リンクを別ウィンドウにクリックして見せるもの、プレビュー、そして左側のクラスターファイルのどこに分類されているかをハイライトするツールなどがある。さらに、サーチエリアのタブやサーチする対象のカスタマイズも可能。



「network security」と検索すると、左欄に検索結果がカテゴリー別に区分されて表示される。

Dogpile <http://www.dogpile.com/>

役立つ情報にアクセスできるインテリファインド

MSN、Google、Yahoo!、Ask Jeeves、About、Looksmartなどの検索結果を統合するメタ検索。検索結果は、適合性に合わせて1本化したものと、複数の検索サイトの結果を相互に見比べる画面の両方を表示できる(ユーザーによるカスタマイズが可能)。検索のキーワードによって、ショッピングサイトや商業的なリンクが検索結果の上位にリストアップされるなどの読み分けをする。例えば、「デジタルカメラの値段」と入力す

るとショッピングサイトが、そして「デジタルカメラの技術」と入力すると技術情報誌の記事などが上位に来る。また、検索結果とともに関連した検索キーワードのリストを表示し、希望の結果が得られない場合はワンクリックで他の関連キーワードでの検索が可能になる。検索結果リストの上部には、同社が開発したインテリファインド技術によって商品情報、上映映画館情報、音楽ファイルなど、役に立つ情報がすぐに呼び出せる仕組みもある。



「Digital camera technology」と検索すると、複数の検索エンジンの結果を統合して表示する。ワンクリックで各エンジンごとにも表示できる。

Mamma <http://www.mamma.com/>

複数の検索エンジンの総合 / 個別結果を比較

MSN、Teomaなど複数の検索エンジン、ディレクトリー、ペーパークリック広告検索サイトの検索結果を統合するメタ検索。結果リストの決定には独自のアルゴリズムを用いるが、これはそれぞれの検索エンジンの結果の重複を票と数えて順位決定に利用するものである。希望する検索エンジンだ

けをえるパワー検索の設定も可能。検索結果リストの右側には、検索をさらにリファインするためのサブカテゴリや関連ワードが表示される。

ツールバーをダウンロードすれば、検索結果を保存して後で閲覧することもできる。



「New York hotels」と検索すると、広告、スポンサー、検索結果、関連リンクが同時に呼び出される。

Ujiko <http://www.ujiko.com/>

ユニークな検索画面とユーザーのレベル制度

検索結果を円環状にビジュアライズし、それをさらに内容によって色分けする。ユーザーの利用状況をユーザー自身のコンピュータに記憶させ、利用することに使い勝手が進化する仕組みを取り入れており、新しいサイトを10か所訪れるごとに1レベルずつ機能性が上がっていき、最終的には気に入ったサイトにハートマークを付けて分類保存したり、不要なサイトをその後の検索から取り除いたりするフィルター機能が付加される。

検索エンジンにはYahoo!を採用しており、

4億ページへのアクセスが可能。レベル10まで進むと、Kartooのメタ検索エンジンを利用できるようになる。

ユーザーの利用状況やカスタマイズ情報は、Ujikoのサーバーではなくユーザー自身のコンピュータに保存される(Ujiko側には検索キーワードのみが送信される)。そのため、個人情報の保護という面で、他の検索エンジンに対してアドバンテージがあるとUjikoは強調している。

Ujikoは、Kartooが開発した2つ目の検索エンジンである。



「London subway」の検索結果は、地下鉄駅やローカルニュースなどに色別に分類され、円環状に表示される。カーソルを重ねるとリンク先のページイメージが表示される。

Kartoo <http://www.kartoo.com/>

奇妙なデザインのカテゴリー表示に注目

セレブリティ、テクノロジー、エンターテインメント、金融情報など、カテゴリー分けされたウェブの情報を検索できる。検索結果をマップ状にビジュアライズ表示し、その1つにカーソルをあてると、関連する他の検索結果がラインで結ばれる。また同時に、左側の欄にはそのリンクの内容がテキストで表示され、右側の欄には検索結果に関連した他のサイトへのリンク、印刷、保存、特定の検索結果をマップから消去する、別のサイトを追加するといった操作がで

きる。マップ部分とテキストを分けることによって、テキストリストを読む苦痛を和らげている。Yahoo!、AllTheWeb、MSN、Altavista、Teomaなど十数個の検索エンジンの結果を統合するメタ検索でもあり、100億ページにアクセスができる。フランスで開発され、現在は英語や欧州言語を中心に運営している。もともとはデータマイニングやナレッジマネジメントの技術開発を行っており、企業向けのサービス業務も数多く行っている。



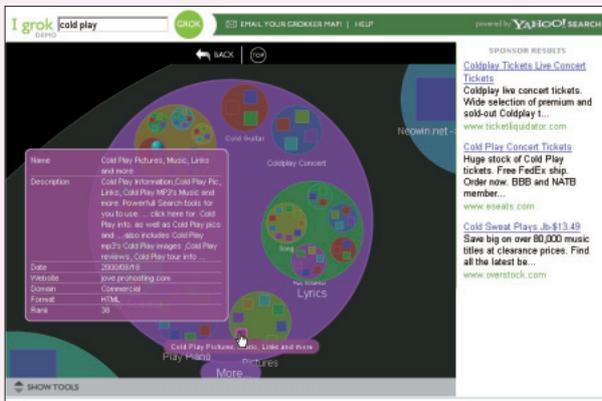
「Stanford University」を検索。関連する検索結果がマップ状に表示される。カーソルを重ねると、関連する結果同士が線で結ばれる。

Grokkerは、2001年にサンフランシスコで創業したベンチャー企業のGroxis(グロキス)による検索サービスだ。検索結果を種類分け(クラスター化)し、可視化(イメージ表示)する。これまでの文字による検索結果表示から、把握しやすいビジュアル表示への変化といえる。技術的には、自然言語処理と、HCI(Human Computer Interaction)の研究者によって、Javaを使って開発されている。「Grok」という言葉は、ロバート・ハイラインの『異星の客』(原題: A Stranger in a Strange Land)の中に出てくる言葉で、「何かを深く理解して、消化する」という意味だ。これまでGroxisは、スタンフォード大学の図書館情報部門との共同研究によってシス

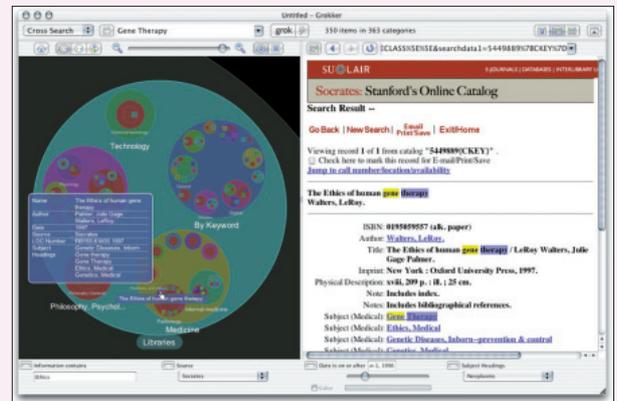
テムの開発を行っている。2004年9月からは、同大学内のネットワークであるSUNetにつながるコンピュータで、大学にかかわる人たちにサービスを提供している。大学サイドのプロジェクトマネージャーであるクリス・ボーク氏によると、Grokkerの特徴は、同時にいくつかのデータベースからの検索を可能にするインターフェイスにもあるという。例えば、スタンフォード大学を例にとってみると、「ソクラテス」と呼ばれる、1980年から大学の図書館で使ってきた図書検索だけでなく、ライセンス契約を結んでいる図書出版情報サービスのEBSCO、米国国会図書館、IEEE、Google、Yahoo!など、15ほどのデータベースから選択して、その内のいくつかのデータベー

スから同時に検索することができる。また、それらの結果をカテゴリーに分けてグラフィカルに表示するので、特に、学部の学生たちには、選択の範囲を広げることができ、ビジュアルに情報が見られることで人気を呼んでいるようだ。

Groxisとスタンフォード大学は、今後、アクセスできる情報データベースの種類を増やすべく努力を続けていくという。また、Groxisの営業担当ディレクターのアドレアナ・ラドニッチ氏によると、GrokkerはYahoo!やGoogleなどの検索サービスと競合するものではなく、世にある検索エンジンと提携して補充し合うものだという。



現在のGrokker(ウェブ版)は、検索エンジンとしてはYahoo! Searchを利用し、その検索結果を解析してクラスター分けする処理を行う。



Grokkerによるスタンフォード大学図書検索「ソクラテス」は、現在はインストール型のアプリケーションだが、今後はウェブ版の提供も予定している。

検索エンジンのアルゴリズムに加え、インターネットユーザーの集成的な利用状況(どれだけのユーザーがそのサイトを利用し、どれだけの時間を過ごしたかなど)を統合して結果の表示順位を決定する。検索キーワードの入力は、インクリメンタル検索が可能で、1文字入力するたびに最もポピュラー

なキーワードをリストにして表示する。検索結果は、写真やロゴも表示されるため、テキストだけのリストに比べて見やすいのが利点。商品検索をすると、人気のオンラインショッピングサイトの商品ページに直結するなど、ユーザーのアクションをサポートする構造となっている。



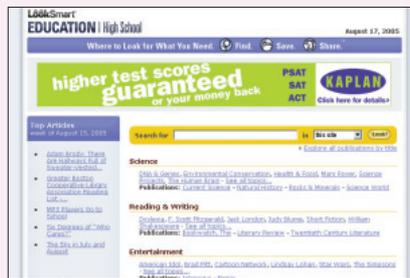
トップページの「Top Searches」をクリックすると、検索利用、人気検索、毎日の広告収入、広告主別の売上げの統計を見ることができる。

Looksmart <http://www.looksmart.com/>

人の手で集められたユニークな検索カテゴリー

基本的には、人間の手によって集められたウェブサイトのディレクトリーで、コンピュータ、エンターテインメント、趣味、家族、宗教、社会、政治、仕事、スポーツ、旅行などの垂直カテゴリーが用意されている。検索は、エリア別などではなく、大学生用(24HourScholar)、ティーンエイジャー用(Teenja)、中高生用(GradeWinner)

子供を持つ親用(GoBelle)、記事検索用(findArticles)など、ユーザーに合わせた検索のための別サイトも複数運営する。目的に合わせた検索結果をすばやく出すために、こうしたユーザー別サイトの利点を説く。ビジネスモデルは広告掲載だが、検索結果の順位は商業/非商業サイトの区別は影響しないという。



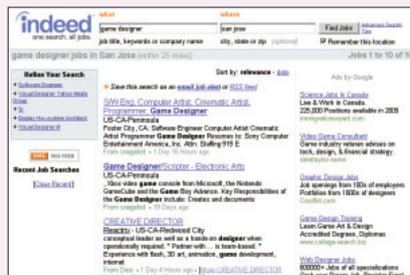
高校生用のパーティカル検索で「human genome」と検索すると、科学ニュースや教材会社などの結果が並ぶ。

Indeed <http://www.indeed.com/>

職種と求人情報に限定した検索とRSS配信機能

新聞や企業の求人情報、また求人情報ネットワークやウェブ上にある求人情報など1000以上のソースから情報をアグリゲートして、職種と都市によって検索する。検索結果リストの左側には、検索をさらに絞り込むための関連ワードとサブピックが並ぶ。ホットジョブなど先行する人気求人情報サイトは、有料で情報をリストの上位に掲載

し、その後にウェブ上にある求人情報を掲載するが、インディードは適合性によって結果の順位を決定する。求める情報をメールやRSSフィードで受け取ることも可能。社名、職種、都市だけで検索することもできる。ビジネスモデルは検索結果に合った求人広告で、検索結果ページの右側の欄にリストで表示される。



「San Jose」で「game designer」と検索すると、複数の求人情報を統合したものが表示される。

About <http://www.about.com/>

機械検索ではなく専門家による回答で差別化

政治、ビジネス、技術、教育、スポーツ、エンターテインメントなど、生活のあらゆる分野について質問(自然言語、あるいはキーワード入力)をすると、専門家による回答が得られるもの。専門家はアバウト・ドットコムが選考し、トレーニングをした500人近い「ガイド」と呼ばれるグループからなる。ウェブのリンクを検索結果として与えるのではなく、あるレベル以上の知識を回答として与えるという方法。ガイドは、1つの分野について、いくつにも枝分かれしたサブテーマを設けて自身の分野を確立している。ガイドは、コンテンツの作成に対して最初の月に500ドルの報酬を得るが、その後は

ページビューが増えれば月々500ドルとページビューの数に合わせた報酬が加えられる。ビジネスモデルは、バナー広告と検索結果に合わせた広告掲載。コンテンツを作成するウェブサイトはおおむね失敗に終わる傾向があったが、Aboutは検索と専門分野別に細分化されたコンテンツを組み合わせることで、広告収入を得るモデルとして成り立っている。設立は1997年で、2005年にニューヨーク・タイムズによって買収された。現在のユーザー数は2000万人とされている。



「Yoga」のページ。ガイドがヨガについて説明し、スポンサーや関連リンクも同時に表示される。

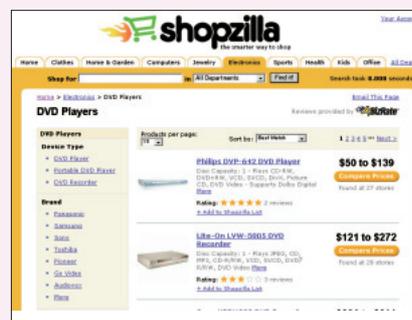
オンラインショッピング・サイトに掲載されている商品(家庭用品、コンピュータ、ファッション、ジュエリー、旅行チケットなど)の価格比較、4万5000以上の店舗で売る2500万点以上の商品が検索できる。

商品の検索とともにそのオンラインショップの対応やウェブサイトを購買者が評価したものを統合して消費者に提供するという、レーティングシステムによって商品の検索とショップの信頼性を組み合わせるようにしている。

検索結果のランキングには、キーワードの

マッチだけでなく、価格、ショップの評判、商品の人気度、在庫などを考慮して順位を決める「ショップランク」という独自のアルゴリズムを利用する。

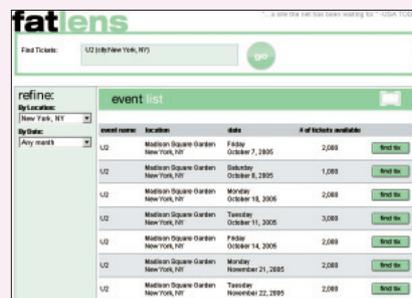
ビジネスモデルは、検索結果に合わせた広告掲載とペイドリスティング。ただし、後者の場合でも料金を支払っているかどうかは、複数のショップがショップランクでまったく同等の評価がなされない限り、検索結果のランキングには影響を及ぼさないと同社では述べている。



「DVD Players」を検索すると、オンラインショップの評価と価格が比較できる。

本サービスでは、提携先の小売情報を利用した従来のショッピング比較サイトと異なり、対象に関連した情報を独自に検索する。TicketMasterをはじめYahoo!、eBay、Craigslist、TicketsNowなどのEコマースおよび求人サイトを対象に検索する。コンサートやイベントの名称のほか、オプションとして開催地も入力できる。サービスを提供するIM2社では、オンライ

ンショッピングの簡略化や高速化に焦点を当てた、特許申請中の技術を活用している。現在のところ、収益源は販売業者のサイトへトラフィックを転送する際の提携料に依存している。Yahoo!およびGoogleのキーワード広告を購入しているが、このビジネスモデルでは、最初にトラフィックを獲得する時点でコストがかかってしまう問題は認識している。



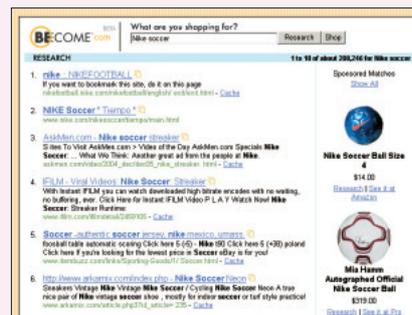
「U2」と検索して、場所を「New York」で設定した画面。コンサートのスケジュールとチケットの価格一覧が表示される。

家庭用品、コンピュータ、ファッション、ジュエリー、ゲームソフトなどの商品検索と価格比較、レビュー検索ができる。

第三者のショッピングネットワークやスポンサー企業のフィードとペイドリスティングなどに頼らず、ショッピング情報に関連したサイトだけをクロールする技術を利用している。ランキングのために独自のアルゴリズムを用い、特定分野で重要なウェブサイト

が他の重要なウェブサイトとどの程度結びついてきたかを判断する。関連度による人気だけではなく、検索のコンテキストを理解することによって不要なサイトが表示されるのを防ぐと同社では説明する。

そのほか、ショッピングに関連したニュースの見出しや、動的な検索単語の提案ツール、消費製品に特化した(単語の)スベルチェク機能も備える。



「Nike soccer」で検索すると、ウェブのリサーチ情報と右欄のショッピング情報を分類されて表示される。

Feedster <http://www.feedster.com/>

さっきの出来事も検索対象になるフィードの即時性

HTMLの代わりにRSSフォーマットによるウェブコンテンツを対象とするため、適時性を特徴とする。

100万種の情報源から更新されたインデックスを1時間に数回供給している。

既存の検索エンジンでもこれに近いペースで更新情報を入手可能だが、HTML文書ではRSSに比べて情報の具体性が低いうえ、既存のコンテンツに対し更新されたポス

ティングを発見できない。

このほか、検索中に発見したフィードを登録できるRSSリーダー機能や、RSSでフィードされているMP3ファイル、画像、動画、Bit Torrentのコンテンツを指定して検索する機能もある。

収益源は広告を主体としており、課金対象の検索結果やフィードに対するスポンサーシップなどで広告戦略を展開する。



「大黒」(日本語でも検索できる)で検索すると、数時間前に行われたサッカーの試合に関する結果が表示された。「大黒」はこの試合で得点した選手の名前だ。

検索エンジンの未来

検索エンジンはインターネットにおいて強い影響力を持っている。ユーザーにとって単に便利なツールというだけでなく、ネットビジネスにおいても重要な存在となった。そして、この検索エンジンの未来を考えると、インターネットの未来を考えることに少なからず重なってくる。

検索エンジンの未来を予想するにあたって重要となるキーワードは、「パーソナライゼーション」「セマンティックウェブ」「ウェブ2.0」だろう。

パーソナライゼーションは、すでにいくつかの検索エンジンも取り組んでいるもので、例えていうなら「あなたのためだけに作られたコンテキスト検索」だ。検索するユーザーの特性や趣向を結果に反映させることで、目的とする情報を的確に探し当てることができる。おそらく将来の検索エンジンは、自己学習機能を持ち、ユーザーが使うたびに最適化されていって、さらに使いやすくなるだろう。

これはユーザーにとっては便利になる一方

で、提供する企業にとっても大きなメリットがある。パーソナライズによって、広告のターゲティング精度がより高まり、ROIが向上するからだ。ただし、あまりにも個々の人に最適化しようとして情報を集めると、プライバシーの問題も生じてくる。

また、セマンティックウェブも検索エンジンの未来に大きくかかわってくる。現在の多くのウェブページは単なるドキュメントだが、セマンティックウェブはそれらに意味の判定を与えるものだ。検索対象であるウェブのデータに「意味」を持たせることで、より高度な検索処理が可能になる。

セマンティックウェブを考えたティム・バーナーズ・リーによれば、現在の検索エンジンは、まだまだ貧弱だという。

将来、セマンティックウェブの仕組みが完成すれば、検索エンジンは文脈までも理解できるようになるかもしれない。

最後に、ウェブ2.0というのは次世代のプラットフォームとしてのウェブを指す。この言葉が示そうとしているのは特定のテクノ



ジェフ・ルルト
ECジャパン株式会社のサーチマーケティング担当バイスプレジデント。国内外の検索ビジネスおよびSEOを専門としている。7月に佐々木俊尚氏との共著『検索エンジン戦争』(アスペクト刊)を刊行。インプレスINTERNET Watchに「そこが知りたい! 検索エンジンの裏側」を連載している。
<http://www.ecjapan.jp/>

ロジーではないが、例えばブロードバンドや安価なストレージ、モバイル機器、RSS、Ajax、ウェブサービスなどが関連してくる。検索エンジンは現在、その収益を広告から得ているところがほとんどだ。その結果、検索結果が商業主義の影響を受け、信頼性や中立性が損なわれる危険性を持っている。そうすると、利益主義に陥らない検索エンジンの存在も必要になってくるかもしれない。例えば、赤十字のような非営利団体が運営する検索エンジンだ。商業主義の影響を受けずにインターネット上の情報にアクセスできるというのは、人々の基本的人権でもあると思う。

日本の検索技術 汎用連想計算エンジン「GETA」

塩田 紳二 フリーライター

GETA ロゴ
国立情報学研究所以外のサイトでもこのGETAは使われており、そこにはGETAのロゴがある。実際、googleで「連想検索」を検索すると、上位に現れるのは、GETA自身やGETAを使って構築されたページの情報だ。

書籍検索に活用される 日本生まれの検索エンジン

国立情報学研究所が運営するサイトの1つに「Webcat Plus」というものがある*1(図18)。書籍に関しては日本最大級のデータベースの1つであり、同研究所が構築中のGeNii 学術コンテンツポータルの一部だ。

このWebcat Plusでは、検索条件を文章やキーワードの羅列で指定する。文字列にさえなっていれば、どんなものでもよい。何かキーワードを入れて検索を行わせると書籍のリストが表示され、さらに関連するキーワードも表示される。単に「検索」と入れると「エンジン」「情報」「Google」といった関連ワードが提示される(図19)。リストされた書籍のうち、自分が探しているものに近い書籍を選んでクリックすれば、内容や目次などの情報が表示される。そこから文字列を拾って、検

索条件として指定すれば、今度は、その本と似た内容を持つ本がリストアップされる。このサイトの検索機能は、単なるキーワード検索ではなく、入力されたキーワードに関連していると思われる単語を含む書籍が検索される。

一般に検索において、入力された単語そのものではなく、それに関連する単語を検索することを「連想検索」という。Webcat Plusではこの連想検索を実現するため、同研究所で作られた「GETA」と呼ばれる連想検索のためのエンジンが使われている*2。国立情報学研究所以外のサイトでもこのGETAは使われており、そこには、GETAのロゴがある。実際、Googleで「連想検索」を検索すると、上位に現れるのは、GETAやGETAを使って構築されたページの情報である。

言語依存のない 汎用ライブラリーとして無料公開

GETA(Generic Engine for Transposable Association)とは、「汎用連想計算エンジン」という名称のプログラム(ライブラリー)のことだ。

このGETAを使うと、頻度付きの出現単語リスト同士の類似性を高速で計算することができる。Webcat Plusの例でいえば、あらかじめ書籍の概要文から頻度付きの出現単語リストを作っておき、検索文字列から抽出した単語リストとの類似性を求め、最も類似性の高い文書を候補として挙げるようになっている。また、その過程で、検索単語と同時に使われる他の単語を求めることができる。これが、関連ワードとして提示されるわけだ。

これを開発したのは、同研究所のソフトウェア研究系プログラミング言語研究部門の高野明彦教授らのチーム(写真1)。

単語リスト同士の類似性を求めるためには、多



図19 「検索」という単語を入れただけで、関連すると思われる書籍や、関連語としてエンジンやデータベースといった単語が表示される。

図18 国立情報学研究所が提供する「Webcat Plus」。国内最大級のデータベースを持つ書籍の検索サイトだ。



数の単語をデータとして持たねばならない。GETAでは、ChaSen(茶筌)などの形態素解析ツールを使って対象文書から頻度付きの単語リストを作り出す。この部分はGETAではなく、利用者が行う処理である。

GETAは、こうしてできた単語リスト同士の類似度を求める計算処理を高速化するとともに、類似度を求めるための式を利用者が指定できるようになっている。そのために「汎用」なのだ。またGETA内部では、言語的な処理を行っていないので対象言語にも依存しない。

このGETAは無料で公開されており、誰でもこのライブラリーを利用することが可能だ。

連想検索は 検索スキルの低い初心者にも有効

GETAによる連想検索では、検索用に指定された単語リスト(検索用文字列から抽出されたもの)と各文書の単語リストをつき合わせて類似性の高い文書を見つけるが、この時点で見つけた文書の単語リストが得られており、その中で頻度の高い単語を取り出していくことで、関連するキーワードが抽出できる。今度は、これをユーザーに提示して選択させることで、より正確な単語リストを作り出せるわけだ。逆に、候補として挙げた文書があれば、その文書と関連性の高い文書を検索させることもできる。このようにGETAの連想検索は、単語から文書、文書から単語も同じように扱えるため、単なるキーワード検索のようにも扱えるし、関連のある文書を検索する場合にも利用できる。

広く使われている検索では、検索対象に含まれているであろうキーワードを指定する。これを一致検索というが、問題は、我々がいつでも正しいキーワードを知っているとは限らないということだ。あるいは、対象に含まれていることは確実にあっても、検索結果が大量になってしまい、有効なキーワードではないことがありえる。極端な話、英語の文書であれば、「the」が必ず含まれているが、これで対象文書を検索しても必要な結果に到達することはできないだろう。こうした場面では連想検索が有効で、特に検索に慣れていないユーザーには有効な検索方法といえる。

「どれだけ探せるか」から 「どうやって探せるか」へ

同じくGETAを使ったサイトに「新書マップ」がある*3(図20)。これも高野教授らが中心になって作ったサイトの1つである。

ここでは、国内で販売されている新書をグラフィカルなインターフェイスで探すことができる。最初は、単語や文章を入れる必要があるが、あとは、ウィンドウ内に並んだ単語やテーマを選択していくだけでいい(図21)。検索というと、Googleのようなインターフェイスばかりが目につくが、情報収集能力や検索機能がある程度強力になってくると、今度は、ユーザーインターフェイスの勝負ということになるだろう。GETAを使うことで、検索機能を簡単に向上させることができる。「ゲタ」を履かせてもらったら、検索エンジンは、別の面で勝負しなければならないということだ。

- * 1 : <http://webcatplus.nii.ac.jp/>
- * 2 : <http://geta.ex.nii.ac.jp/>
- * 3 : <http://shinshomap.info/>

GETA = ゲタ = 下駄?
GETAという名称は、情報の類似度などを判定する式などを研究する場合に、その基盤となる環境を提供することで、研究者に「ゲタを履かせる」という意味もあるのだという。



写真1 国立情報学研究所の高野明彦教授(左)とGETAのプログラミングを担当した西岡真吾特任教授(右)。

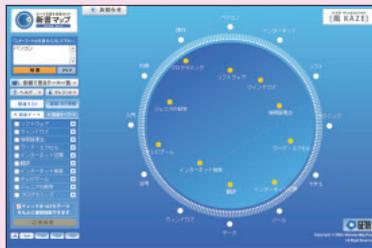
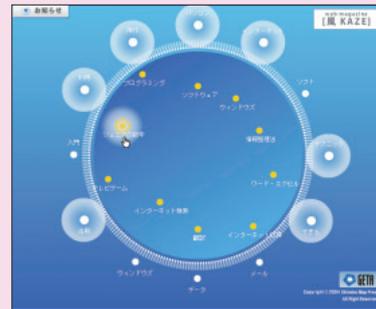


図20 「新書マップ」では、キーワードや文章を入れると関連するテーマが円内に、関連キーワードが円周上に表示される。

図21 テーマをポイントすると、それに関連する単語がハイライト表示される。



日本の検索技術 行動支援メディアを目指す「goo ラボ」

野本 幹彦 フリーライター

独自の技術や考え方を盛り込んだ 地図サービス

「goo」では、次世代の検索機能を研究した結果を「goo ラボ」*1と呼ばれるサイトで公表している(図22)。これらの技術の中には、将来gooの正式サービスとして提供されるものもある。

欧米諸国の言語に比べ、日本語は解析するのが難しいといわれている。唯一の国産ポータルサイトで検索エンジンも独自に開発しているgooでは、どのような技術開発が行われているのだろうか。

gooの独自調査によれば、ユーザーが予想しない日本語特有の表記ゆれなどによって、8回に1回は求める検索結果が得られないという。一般的な重要度からランキングを付けて検索結果を表示しても、人によって求めるものはさまざまである。最初の結果ページに出てくるものが目的のものとは限らないのだ。

gooでは、このような問題を極力回避して求める検索結果を得るために、「検索対象範囲」を絞り込み、「意味内容」や「分類」による検索を可能にする研究が進められている。また、単なるキーワードでの検索ではなく、実際の行動につながる

検索を可能とする「生活行動支援メディア」を目指しているという。

これらの研究の1つの答えが、goo ラボから生まれ、7月26日にリニューアルされた「goo 地図」だ。この新たな地図サービスでは、「goo 地域」という地域情報サービスと連動して店などの情報が表示されているようになっている。他の地図サービスにはない特徴として、スクロールやダブルクリックで地図を移動すると、表示される地域情報もそれに追隨してリアルタイムで更新されるという機能がある(図23)。これは、日本では電車で移動する人が多いため、「目的地の最寄り駅を探し、そこから移動しながら店を探す」という行動が考慮されたものだ。

ユーザーの行動まで考えたインターフェイスの工夫は、通常のテキスト検索ページでも活かされている。たとえば、現在公開中の映画タイトルを入力すると、公開映画館の情報が表示されるが、過去に特定の地域の映画館の情報をクリックしたことがあればその地域の映画館に絞り込んでくれる。また、「～駅」と検索すると、それを駅名の検索と判断し、その周りの地域情報や時刻表などのリンクが用意される。



図22 goo ラボでは、さまざまな次世代技術に触れることができる。実際に自分で使ってみて、将来の検索を体験してみたい。



図23 地図を移動させると、地図内とその右側の「検索結果」に表示される情報も刻々と変化する。これはGoogle Mapsなどにはない特長。

goo ラボで誕生する 次世代の検索技術

「関心事アンテナ」は概念検索を応用したもので、検索キーワードに類似した概念の言葉を含むページも検索結果に含めるといったものだ。キーワードは単語だけではなく文章でもかまわない。

この検索の利点は、知りたいことが曖昧でも望んだ結果を得やすいということだろう。たとえば、中田英寿という名前のサッカー選手がチームを移籍したことを知らなければ「中田英寿 移籍」というキーワードは出てこない。しかし関心事アンテナでは、「サッカー選手の移籍情報が知りたい」と考えて「サッカー選手 移籍」と入れれば、「中田がボルトンに移籍した」というニュースを知ることができるのだ(図24)。そのページに「サッカー選手」という単語が含まれていなくても、「サッカー選手」の類似・関連語として「中田英寿」も対象となる。同じ意味でも多様な表現がある事象を検索する場合、自分のボキャブラリーを超えた結果が得られるという点で、ユーザーを支援しているといえるだろう。

「ニュース記事分類・検索実験」は、ニュース記事検索結果をトピックカテゴリー別に分類し、絞り込みを可能とする検索だ。このようなカテゴリー別の分類にはさまざまな方法が使われてきたが、goo では検索対象をニュース記事に限定し、人物、組織、場所などの属性に着目することで分類しやすくし、対象となる単語が人名なのか地名なのかを認識できるようにしているといふ(図25)。



写真2 goo ラボで次世代の検索技術に取り組む(右から)NTTサイバーソリューション研究所の奥雅博氏、NTTレゾナント・ポータル事業本部の浜野輝夫氏、小澤英昭氏。

行動支援メディアを目指すうえで 必要なもの

goo では、今後も日本特有の生活行動をベースに、検索エンジン側の仕様をユーザーに強制するのではなく、普通の人がわかりやすいインターフェイスとサービスを目指している。これを実現させるために、以下のポイントが重要だという。

まず、意味内容を特定しにくい日本語という言語の特徴をよく踏まえ、「ニュースのみ」や「ブログのみ」などの検索対象範囲を絞った検索でより使いやすいサービスの提供を目指す。また、今後は個人情報などの出すべきではない情報をフィルタリングする技術を開発していく必要もある。こういった特定の分野に絞って分類していく検索技術が成熟していった後に、将来的には分野を絞らずにウェブ全体で適切な検索結果を返せるような技術に発展させていきたいとのことだ。

* 1 : <http://labs.goo.ne.jp/>

フィルタリング技術

8月18日に公開された画像・動画検索の「マルチメディア マイスター」では、独自の類似画像検索・分類技術により、精度の高い結果が得られる。また、猥褻画像や動画などの有害コンテンツをフィルタリングによって排除する機能も備える。



図24 検索語句の類似語までを検索対象とする関心事アンテナ。「サッカー選手に関連する情報」として中田英寿が結果に出てきた。



図25 人物、組織、場所などの属性で分類し、左側の「トピックツリー」に意味内容で絞り込むためのヒント(言葉)が表示される。



[インターネットマガジン バックナンバーアーカイブ] ご利用上の注意

このPDFファイルは、株式会社インプレスR&D(株式会社インプレスから分割)が1994年～2006年まで発行した月刊誌『インターネットマガジン』の誌面をPDF化し、「インターネットマガジン バックナンバーアーカイブ」として以下のウェブサイト「All-in-One INTERNET magazine 2.0」で公開しているものです。

<http://i.impressRD.jp/bn>

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容(技術解説、URL、団体・企業名、商品名、価格、プレゼント募集、アンケートなど)は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者(執筆者、写真の撮影者、イラストの作成者、編集部など)が保持しています。
- 著作者から許諾が得られなかった著作物は収録されていない場合があります。
- このファイルやその内容を改変したり、商用を目的として再利用することはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用する際は、出典として媒体名および月号、該当ページ番号、発行元(株式会社インプレス R&D)、コピーライトなどの情報をご明記ください。
- オリジナルの雑誌の発行時点では、株式会社インプレス R&D(当時は株式会社インプレス)と著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接のおよび間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

このファイルに関するお問い合わせ先

株式会社インプレスR&D

All-in-One INTERNET magazine 編集部

im-info@impress.co.jp