

# 検索を科学する

塩田 紳二

## 最終回 検索のこれから

最近のテレビコマーシャルでは、URLをいわずに「 で検索」といった表現も見られる。たしかに、キーワードで企業などを特定できれば、長いURLは不要である。このように、検索にはいままでなかった利用方法ができつつある。連載の最後となる今回は、検索機能そのものについて考え直してみようと思う。

### あらゆる場面で使われる検索

検索は、コンピュータ処理における基本処理でもあり、さまざまなところで使われている。また、遺伝子研究などにおいても、検索機能は重要な役割を持っているように、「大量のデータがあるところに検索あり」という感じだ。

コンピュータの性能が上がってきたことから、従来は不可能だと考えられていたところでも検索が使われるようになってきている。このことは、たとえば、大量の情報の中から、意味のある関係を見つけるデータマイニングなどという応用分野も出てきている。

また、最近のHDDビデオレコーダーなどに搭載されている「おまかせ録画」や「キーワード録画」といった機能も検索機能を使ったものだ。

確かに検索には便利な面もある。キーワード入力を面倒に思わなければ、URLを覚える必要もない。たとえば、企業名を入れてそのホームページを見つけることも可能だ。ブックマークは、大量に登録すると逆に見つけるのが大変になるし、ここで検索を使うならインターネット

検索を直接使ったほうが便利というわけだ。

調査会社のレポートなどでも、インターネット検索の利用回数はまだ増え続けており、すでにインターネットユーザーの一般的な行動になりつつある。

インターネット内の多くのものが検索できるということで、変わった使い方としては「見つからないものを探す」というものがある。これは聞いた話ではあるが、ある企業の名前を決めるのに「Googleで意味のある検索結果が表示されなかったキーワード」を選んだという。このようにすることで、ある程度ユニーク(一意)な単語を調べることが可能なわけだ。

また、検索を利用する技術も登場しつつある。たとえば、最近の一般ユーザー向けのデジカメ写真修正ツールでは、赤目修正を自動で行うことができる。これは、ソフトウェアで赤目部分を「検索」している。

### 検索キーワードの選定と入力における課題

Googleの成功以来、さまざまな検索機

能が登場し、インターネットだけでなくローカル環境の検索も実現されるようになった。しかし、そこにあるのは相変わらずのキーワード入力である。

キーワードの指定による検索は古くから使われてきたため、慣れているユーザーも多いのだが、一方でいくつかの問題もある。

まず、キーワード自体の「表記ゆれ」などの問題である。人間が日常的に利用している言語では、1つの言葉に複数の表記方法がある。日本語ならば、同じ言葉をひらがなやカタカナで表記したり、外来語では、音引きや促音の表記の違い(サーバとサーバー、ウィンドウとウインドウなど)が出てくる。あるいは、英語であればスペルミスなどもあるだろう。逆に検索対象にこうした表記の違いやスペルミスが含まれる可能性もある。

これらについては、辞書による補正や、正規化(半角、全角の違いやカタカナ、ひらがななどを統一する)などがあるが、検索対象自体に手を付けられないインターネット検索では、利用が困難な場合もある。たとえば、ある言語ではスペルミスとされる綴りが別の言語では正し

いということもあるからだ。Google では、サジェスト機能で、正しいと思われるスペルを提示する機能がある。これは、インデックスに含まれる単語で、キーワードと似たもののうち、出現頻度の高いものを候補として提示しているようだ。こうした機能は、ローカル検索でも必須となるかもしれない。

また、キーワードを文字表記しなければならないことも問題である。人間の会話は「音」で行われるため、そこから得た情報は必ずしも正しい文字表記に変換できるとは限らない。しかし、キーワード検索では少なくとも文字表記されたものが必要である。これは英語圏などでは、母音の省略や文字の置き換えなどで音の似た単語の検索方法などが確立されているからだ。しかし、日本語では、文字情報から音を推測するためには辞書が必要なため、実装しているシステムはあまり多くない。

## キーワードの入力だけに頼らない検索

たとえば、前に述べたハードディスクコーダーでは、録画した番組に対して、テレビ局や時間帯、番組ジャンルといったやり方で整理して表示する機能を持つ機種もある。これなど、検索(データベース)の応用といえるだろう。このようにキーワードを入力させずに検索を行う方法が今後は増えてくるのではないと思われる。

Windows デスクトップサーチでは、ファイルなどの日付による検索が可能だ。こうした検索の場合、たとえば、カレンダーを表示して、これを GUI として検索させることも不可能ではない。Google の検索履歴などでは、一覧表示とともにカレンダーが表示され、クリックするとその日に行った検索の一覧を表示できる(図1)。また、デスクトップ検索では、日付順に結果を並べることも可能だ(図2)。

非キーワード検索でも、日付のように文字で表示できるものは、Windows デスクトップサーチのようにキーワード入力の拡張などで組み込まれてしまうが、あまり便利なものではない。たとえば、先週作った文書を検索する場合には、「日付：先週」などとキーワード入力しなければならない。

その意味では、まだ、Windows に標準で組み込まれている検索機能の日付検索

のほうが簡単でわかりやすい指定方法である(図3)。

また、キーワードを使った検索でも、必ずしもユーザーがキーワードを入力する必要はない。すでにインデックス化が行われているとしたら、インデックスに載っていない、あるいは前半部分も一致しない単語は、どちらにしても探すことができない。原理的には、登録されているキーワードすべてを提示して選択するこ

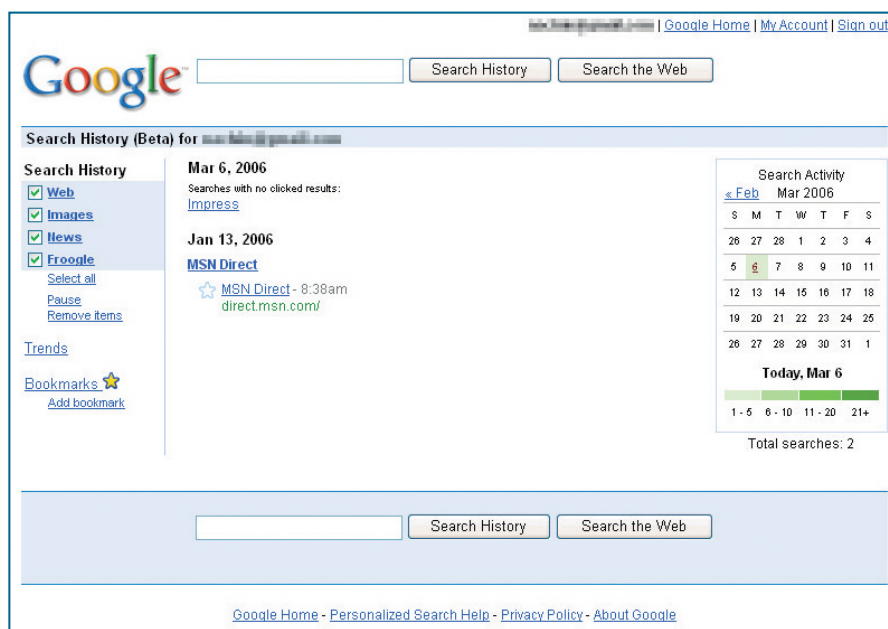


図1 Google のパーソナライズ検索機能には検索履歴機能がある。ページ右側のカレンダーは、検索が行われた日には色が付き、数に応じて色が濃くなる。クリックするとその日に検索されたキーワードとクリックしたサイトのリストが表示される。



図2 デスクトップサーチでは、「タイムライン表示」という形で、検索結果を時間順に並べたものが表示できる。検索履歴と同様にカレンダーから日時が指定できる。

とでキーワード入力は省略が可能である。ただ、あまりに多いので、提示方法を工夫する必要はあるだろう。たとえば、辞典などのように最初の1文字などで分類し、選択範囲を狭めるような方法は必要だろう。

簡単にいうと、ユーザーに文字を入力させるのは、汎用性を高めるという意味もあるが、ユーザーインターフェイスとしては最も手振きのやり方ともいえるだろう。

## 文字以外のさまざまな情報も検索の対象に

画像検索も、そろそろ実用化されそうな技術の1つだ。画像データを前処理していくことで、たとえば、顔のようなものが写っている写真を検索することなどが可能だ(図4)。写真を一枚例示すると、似たような写真や同じ人物の写真を見つけるといったことは、数年ぐらいのうちに一般的なパソコンで可能になりそうである。また、最近のデジカメは、被写体の顔らしき部分を見つけ、そこにフォーカスを合わせるような機能が搭載されている。このような画像認識も広い意味で

検索機能の1つである。

音楽の検索なども、ときどき、企業の研究開発展示などで見ることがある。たとえば、音楽の雰囲気などを自動的に判定して、ユーザーの好みや指定に合わせて自動選択する方法である。ただ、こうした方法は、あまり普及していないところをみると、それほど評判がよくないのかもしれない。

もっとも、こんな複雑なことをしなくても、Amazonがやっているように、「をお買い求めのお客様はほかにこんなものを買っています」という購買行動からの検索のほうが、好みの音楽を的確に検索できるのかもしれない。

MP3のID3タグを使った検索機能では、iTunesのスマートプレイリストのようなものがすでに実用になっている。同じような機能はWindows Media Player 10にもあり、ユーザーが指定した条件で自動的に動的なプレイリストを作成する。

## あらゆるアプリケーションに搭載される検索機能

Google デスクトップサーチや Windows

デスクトップサーチでは、SDK(ソフトウェア開発キット)が提供されており、検索機能をほかのアプリケーションからも利用できるようになっている。これを使えば、自身のアプリケーションに簡単にデスクトップ検索機能を搭載できる。いままでは、ユーザーにファイルを指定させるファイル指定のダイアログボックスが普通であったが、今後は、対応ファイルをすべてリストアップしたようなダイアログが普通になってくるのではないと思われる。

特に、今年後半に登場するといわれている Windows Vista では、デスクトップ検索機能が標準搭載されるため、対応アプリケーションは、必要なモジュールを添付することなくインストールが可能になる。現状では、Windows デスクトップサーチも、Google のものも、アプリケーション自体とは別にインストールが必要なため、サードパーティーのアプリケーションからは利用しにくい面がある。しかし Vista からは、少なくとも Windows デスクトップサーチに関しては、こうした問題はなくなるだろう。

検索機能が標準になると、ファイル名



図4 顔や青いシャツなどを認識して画像を検索するシステム。インテルがIDF (Intel Developer Forum)で公開したDiamond Snap Findというソフトウェア。

図3 Windows XPの検索機能では、ファイルの変更日時を簡単に指定できる。時間に関係する検索なら、こうしたユーザーインターフェイスのほうが自然だろう。



なども必ずしもわかりやすいものを付ける必要がなくなり、極論すれば、ユーザーが指定せずに、機械的に番号などを割り振ってファイル名にすることも可能になる。

いまの Windows などでは、ユーザーが作成した文書なり情報なりは、ファイル名を手がかりに指定しているため、ユーザー自身がファイル名を付ける必要がある。しかし、ファイル名を手がかりにする必要がなければ、ユーザーが名前を付ける必要はない。

また、Windows Vista への搭載は見送られたが、マイクロソフトは、WinFS と呼ばれるデータベース構造を持ったファイルシステムを開発中だ。簡単にいえば、これは、アプリケーションが自身の情報を管理するためのデータベースで、そのスキーマ定義などを共有することで、アプリケーション間の情報共有を促すことができる。具体的には、電子メールやアドレス帳などがシステム定義されていて、ほかのアプリケーションからこれを読み込むことも、独自の情報を付加して使うことも可能になるものだ。

これまで、パソコンの中では、さまざまなアプリケーションが独自の情報を管理してきた。それらを相互に利用することを阻んでいたのは、独自のファイル形式である。しかし、XML を使ったり、この WinFS のようなシステム提供のデータベースを使ったりすることで、その壁は崩れつつある。こうなると、検索機能が必要となる場面が増え、必然的に、アプリケーションは検索機能を強化しなければなくなるだろう。

マイクロソフトが WinFS の搭載を断念する代わりに搭載したのが、実は Windows デスクトップサーチなのである。共通データベースではないため、アプリケーション間で直接データを共有することはできないが、検索機能があるために、ユーザーレベルでは、分散している情報をまとめて扱えるようになるわけだ。

## 検索技術の向上に伴う ダークサイドの出現

ただ、検索技術が向上していくにつれ、その副作用あるいは弊害なども起こりつつある。たとえば、検索機能は容易に「検閲」機能に成り得る。米国政府は、通信や通話などを傍受して特定のキーワードを検索しているといわれているし、一部の国では、特定の情報がインターネット経由で国内に入らないように規制が行われている。

コンピュータの性能が向上すればするほど、こうした機能も向上することになる。

インターネットはさまざまな形で使われているため、自分に関係する情報がどう利用されているのかはもうわからない状態だ。たとえば、電子メールサービスと

検索機能を組み合わせれば、その運営者は、ある特定のニュースレターがどの程度のユーザーに利用されているかを判定することが可能である。あるいは、頻繁に話題になる商品名を調べることもできるだろうし、特定の商品名に対してネガティブな表現がどの程度使われているかを調べることもできるだろう。こうした調査結果で個人を特定することはできないが、プライバシーが侵害されていると感じる人も少なくないと思われる。

卑近な例では、Google デスクトップサーチのサイドバーにある写真表示やクイックアクセス(最近アクセスしたサイトの一覧)に好ましくない内容が表示されてしまったなどということもある(図5)。

こうした点で、検索には危険な一面もあることは常に意識しておく必要がある。



図5 Google デスクトップサーチのサイドバー。最近アクセスしたファイルやウェブページ、保存してある写真などを表示する機能がデフォルトで存在するため、望まない情報まで表示されてしまう可能性がある。

## [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的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](mailto:im-info@impress.co.jp)