

検索を科学する

塩田 紳二

第4回 MSN サーチツールバー with Windows デスクトップサーチ

今回は、マイクロソフトのデスクトップ検索システム「Windows デスクトップサーチ」について見ていくことにする。これは、MSN サーチツールバーの一部として提供され、ローカルのハードディスク内の検索を行うシステムであり、Google デスクトップサーチと同じ位置づけの製品だ。

マイクロソフトの検索技術

Windows デスクトップサーチは、2004年11月にベータ版が発表され、2005年5月16日に正式版となった。日本語には6月25日に対応し、現在MSNのウェブサイトで配布が行われている¹。

このデスクトップサーチは、マイクロソフトが開発していた「Staff I've Seen(以下、SIS)」という検索システムをベースにしたものだ。これ自体は、2003年のコムデックスで行ったビル・ゲイツ氏の基調講演でも公開されている。実際には、このSISをベースにしたPhlatというプロタイプが作られた。Windows デスクトップサーチは、これを利用して作られた。

SISは、情報検索の研究用に作られたもので、その論文によれば、統一的な情報のリスト表示と時間による並べ替えが情報検索に有効だという。Windows デスクトップサーチには、この成果が盛り込まれている。

登場のタイミングだけを見ると、Windows デスクトップサーチは、Google デスクトップサーチに対抗して急遽作られた

ものようだが、このようにしっかりした出自を持ち、それなりの研究を経て登場した製品である。もっとも、すぐに製品化しなかったものでもあり、Google デスクトップの登場が、Windows デスクトップサーチを製品化するための原動力となったことは否定できないだろう。

※1 : <http://desktop.msn.co.jp/>

動作の仕組み

SISもWindows デスクトップサーチも基本的には、前回(2005年8月号)解説したIndex Serviceをベースにしている。つまり、iFilterを使い、各文書から単語リストを切り出し、これを元にインデックス情報を作成する。検索は、このインデックス情報を使って行われる。

この点では、エクスプローラに組み込まれている検索と性能的な違いはない。しかし、Windows デスクトップサーチには、多くのファイルに対応するためのiFilterが含まれており、Windows XPが標準では検索できなかった情報も検索が可能だ。たとえば、Outlook内の予定や連

絡先なども検索対象となる。標準で対応していないのは、PDFやZIPファイルなどだが、これも、サードパーティーが用意するiFilterを利用することで検索可能となる。Windows デスクトップサーチ(MSN ツールバー)のページからは、主要なiFilterのダウンロードができるページが用意されている。

多くのファイル形式に対応するため、インストール直後には、全ファイルをスキャンしてインデックスを作成する必要がある。この処理は意外と負荷が大きく、大量のファイルがあると1日以上かかることもある。筆者は、Pentium 4/1.5GHzの環境に、16Gバイト、2万4000ファイルほどの資料(PDFやWord、PowerPoint、Excel、テキストファイルなど)を置いてインデックス構築を行ったが、18時間ほどかかった。実際には、プログラムなど他のファイルもあるため、ファイル数は5万程度、ハードディスクの使用量27Gバイトが対象になったと思われる。

なお、インデックス構築は、通常もファイルが変更、追加されるたびに行われ、Windows デスクトップサーチでは、その

制御も可能。システムに余裕がある場合には、インデックス作成を優先することもできるし、ノートPCなどバッテリーで動作しているときの動作も指定が可能だ(図1)。

対象フォルダーなどの指定は細かくできないが、Windows デスクトップサーチでは、プログラムの検索も可能で、Windows ディレクトリーや Program Files ディレクトリーなどを省く必要がないためだと思われる。

ユーザーインターフェイス

検索エンジンとして Index Service を使っているが、それでは Windows XP 標準の検索機能とはどこが違うのか？それは、ユーザーインターフェイス部分である。Windows デスクトップサーチは、2つのユーザーインターフェイスを持つ。1つは、エクスプローラに統合された Windows デスクトップサーチウィンドウである(図2)。もう1つは、タスクバー上に配置されるデスクバーだ(図3)。

デスクトップサーチウィンドウは、検索結果を表示するためのエクスプローラであり、ファイル内容を表示するプレビューウィンドウ(オンオフ可能)を持つ。また、検索結果の多くは、プレビューのほかに、概要表示(大きなアイコン表示)も可能である。

このあたりは Google デスクトップサーチと同じだが、Windows デスクトップサーチでは、さらに検索結果を絞り込むための機能がある。

1つは、ファイルタイプなどによる絞り込みだ。「すべて」、「文書」、「電子メール」、「音楽」、「画像およびビデオ」、「予定」といった情報の種類、ファイル形式を指定して表示することが可能だ。このほか、特定の種類(テキスト、スプレッドシート、プレゼンテーションなど)だけを表示させることもできる(図4)が、すべての種類の文書ファイルを個別に指定できるわけ

ではない。

Google デスクトップサーチでは、検索語として「 filetype:拡張子」として特定のファイルタイプのみを対象とすることができる。しかし、Windows デスクトップ

サーチでは、このあたりを GUI で操作でき、クリックするだけで再検索できる点、および単純な拡張子指定ではなく、画像やビデオ、文書といったカテゴリー指定が細かく可能という点が違う。

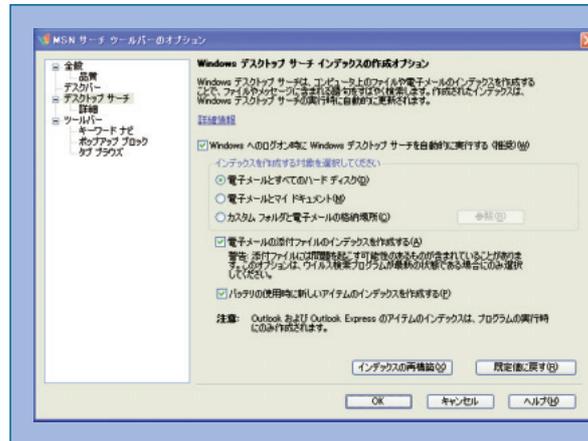


図1 Windows デスクトップサーチでは、インデックス処理や対象などの指定ができる。また、ノートPCなどではバッテリー動作時にインデックス化を行うかどうかの指定が可能。

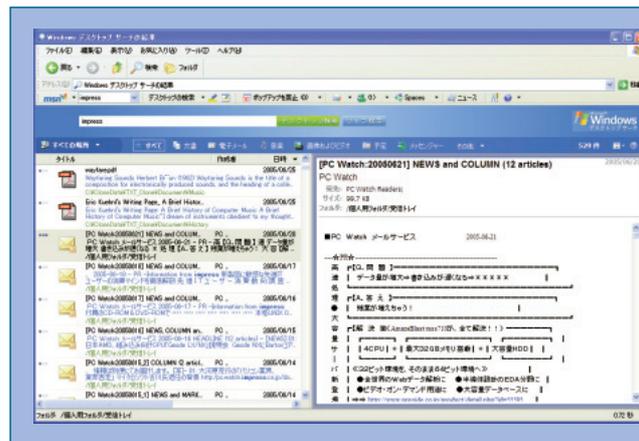


図2 Windows デスクトップサーチのメインウィンドウ。これはエクスプローラ内部に表示されている場合。場所や対象を指定しての絞り込み検索が簡単に行える。

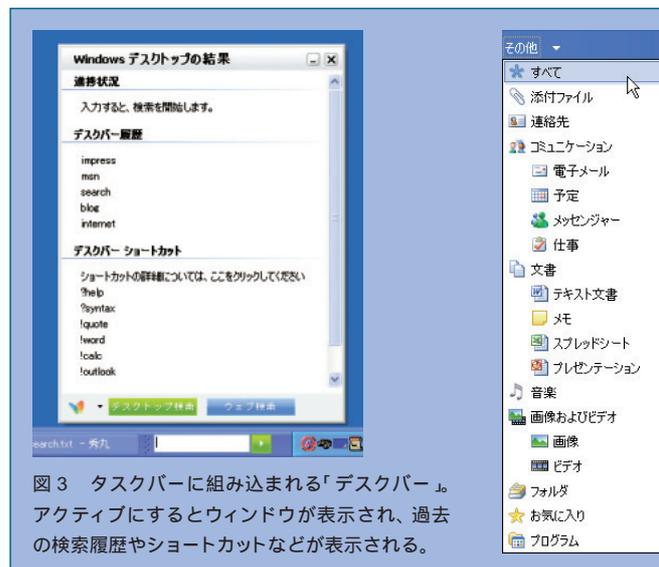


図4 Windows デスクトップサーチウィンドウでは、[その他]メニューから検索対象を限定できる。

図3 タスクバーに組み込まれる「デスクバー」。アクティブにするとウィンドウが表示され、過去の検索履歴やショートカットなどが表示される。

もう一つ、Windows デスクトップサーチでは、検索対象が存在する場所を指定しての検索ができる(図5)。細かくフォルダーなどが指定できるのではなく、ファイル、Outlook、Outlook Expressのどれかを指定できる。メールなどが大量にあり、それらを除外したいときなどは便利だが、Google デスクトップサーチのファイル、メール、ウェブ履歴といった選択とほとんど違いはない。なお、Windows デスクトップサーチでは、ウェブ履歴の検索はできない。このあたりはぜひとも対応してほしい部分だ(さらにいえば、ウェブ検索と同時にすることもできない)。

また、Windows デスクトップサーチは、エクスプローラと同じく、検索結果の並べ替えが可能だ。これは、ファイルリスト上部のカラムをクリックするだけでよく、日付だけでなく、ファイル種別や Windows デスクトップサーチが付けたランクやフォルダー名などで並べ替えることができる。前述の SIS の研究によれば、ユーザーが行う検索対象は、作成されてからあまり時間がたっていないことが多いという。多くの場合、ユーザーは、ランク付けよりも日付順のほうを利用し、ヒット率も高かったそうだ。なお、この Windows デスクトップサーチウィンドウは、MSN ツールバーや Outlook に組み込まれる MSN ツールバー for Outlook から起動し、デスクトップ検索を行うことができる。

もう一つのデスクパーでは、キーワードを入れると小さなウィンドウが表示され、インクリメンタル(差分)検索が行われ、検索履歴へのアクセス(図6)が可能になる。また、ここでは、ショートカットが利用でき、“!calc”などと入力することで電卓プログラムが起動するといったことが可能。ユーザーがショートカットを定義することもできる。

インクリメンタル検索が行われると、検索結果がカテゴリー毎(電子メールやファイルなど)に表示される。ただし、すべての検索結果ではなく、日付順で上位

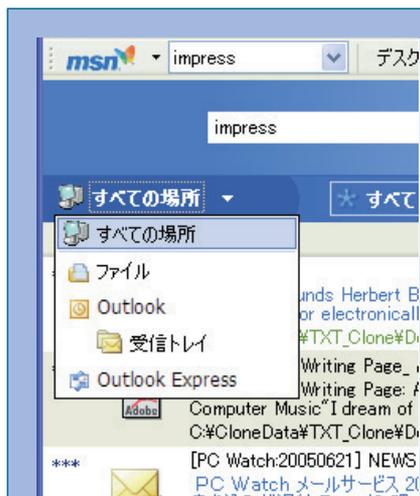


図5 検索対象となる場所としては、ファイル、Outlook、Outlook Expressが指定できる。



図6 デスクパーに文字を入れていくとインクリメンタル検索が行われる。このときの表示は、日付順で新しいものから表示されていくようだ。

のものだけが表示される。簡易なインターフェイスだが、これだけで間に合うことも少なくない。「すべてを表示」を選べば、Windows デスクトップサーチウィンドウへ移動することもできる。

Google デスクトップサーチと比べると、検索後の並べ替えやフィルタなどの機能が充実していて、しかも、それらはエクスプローラなどとうまくつながっている感じである。そういう意味では、マイクロソフトが自身の土俵で勝負しているという感じがした。

この Windows デスクトップサーチは、エクスプローラだけでなく、インターネットエクスプローラ(IE)内部でも動作する。不思議なことに、エクスプローラとIEでは、その動作が微妙に違う(図7)。まず、プレビュー領域でのプラグインを使った文書データの表示では、エクスプローラ内部で動作しているときにセキュリティ機能により、ActiveX プラグインの動作がブロックされ、警告が行われる(図8)。Word や Excel といったファイルの表示では問題はないが、PDF などの表示では必ず表示前に警告が出る。しかし、IE の内部で動作しているときには警告が出なくなる(ただしセキュリティ設

定が必要)この警告の表示はかなりうっとうしい。

キーワード入力

Windows デスクトップサーチは、入力したキーワードそのままを検索するのではなく、場合によっては、単語に分解して検索を行う。たとえば、「健康診断」という検索語の場合、「健康」、「診断」のどちらか、あるいは両方の語が含まれているファイルもヒットする。このため、見つかった候補に必ずしも入力した検索語通り(この場合には「健康診断」)の語が含まれているとは限らない。したがって、検索結果に違和感を感じる場合がある。

このような場合には、全体をダブルクオート(")でくくり、たとえば「"健康診断"」などとして検索を行う必要がある。

このほか、入力時には、AND や OR、NOT といった論理演算子や日付や日付範囲指定、メールの差出人などの情報項目の指定を行うことができる(表1)。

特に、検索対象を限定した場合には、たとえば、連絡先の会社名などの指定ができる。この点では、非常に細かい検索条件の指定が可能といえるが、細かい指

定をすればするほど、入力は複雑になり、使い勝手は落ちてしまう。実際には、あまり使われない機能になるのかもしれない。ユーザーとしては、単に検索語を入れて結果が出るほうが便利だからだ。

Windows デスクトップサーチでは、属性の指定は、検索候補を減らすことはできても、新たな候補を増やす働きはない。すでに単純なキーワード検索で可能性のある候補はすべて列挙されているはずだからだ。候補があまりに多かった場合、検索内容によっては有効だが、前述のようにキーワードの分解が行われるため、まずは、フレーズ検索(検索単語をダブルクォートで囲む)を行ってみるほうが効果的だろう。

なお、こうした条件指定などの機能は、おそらく、Index service が持っているプロパティキャッシュやそれを対象とした検索機能を利用しているのだと思われる。

かなり意欲的だが、もう少し

Windows デスクトップサーチは、Google デスクトップサーチの置き換えになるか？ この間に関しては「いいえ」と答えざるを得ない。1つは、Windows デスクトップサーチは、ウェブの閲覧履歴の検索をサポートしていないからだ。

一応、IE の履歴機能があり、これを起動すれば、たしかに検索自体は可能ではある。しかし、1つのユーザーインターフェイスですべて検索できる Google デスクトップでの検索に比べると使い勝手が落ちるのは確か。

一方、予定や連絡先が検索可能なのは、現時点では、Windows デスクトップサーチの明らかなアドバンテージである。Outlook が自社製品という強みを活かした方法だ。これが示すように、ユーザーの検索に対する要望をすべて1か所でまかなうのがデスクトップ検索のあるべき姿だと思う。それゆえ、Windows デ

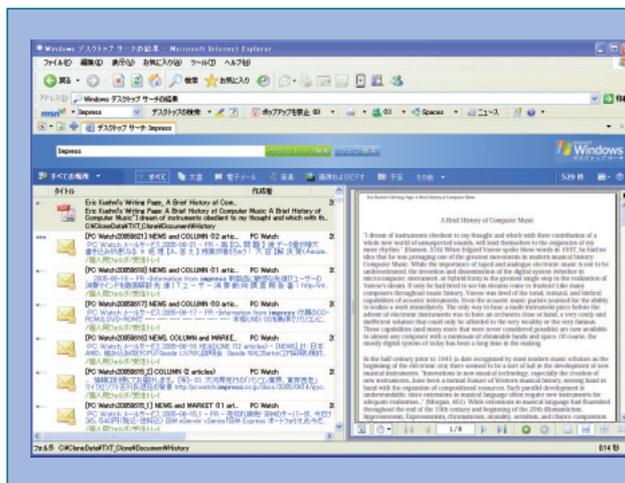


図7 IEのツールバーにあるMSNツールバーからデスクトップ検索を行うと、Windows デスクトップサーチウィンドウはIEの中に表示され、このときにはタブブラウジングが可能になる。

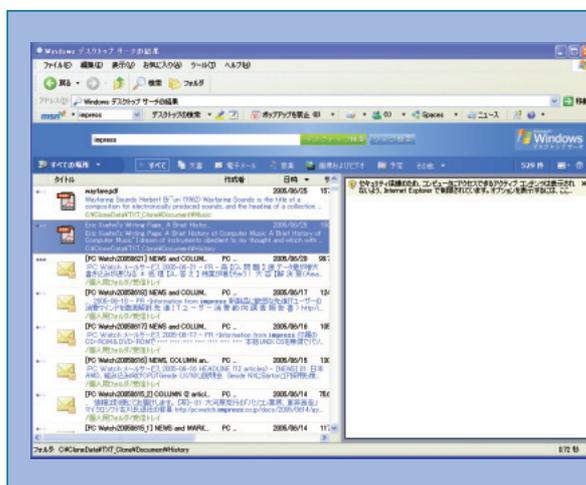


図8 エクスプローラの中でWindows デスクトップサーチを実行している場合には、一部のファイル形式のプレビューがブロックされてしまい、すし使いづらい。

キーワード	意味
NOT key	key を含まない文書の検索
-key	NOT と同じ
key OR key	検索語のどちらかが含まれている文書を検索
"key"	「key」そのものが含まれる文書を検索
(key key...)	語順に関係なく指定されたすべてのkeyを含む文書を検索
propaty >/< data	「日付:」や「サイズ:」と組み合わせると大小比較を行う
前/後:date	日付の前後を指定
サイズ:num	対象データのサイズ指定
日時:date	対象データの日時指定
場所:place	対象データの場所(ファイル/Outlookなど)を指定

表1 クエリーの形式。このほか、対象ごとに「差出人:」などのプロパティ値があり、指定が可能。

スクトップサーチで、ウェブ履歴やウェブとデスクトップ検索を同時に行えないのは残念だ。

このようにGoogle デスクトップサーチとWindows デスクトップサーチは、主要な機能は一致するものの、それぞれ違った

メリットを持つ。どちらもインデックスのための処理時間とディスクスペースを消費するため、両方インストールしておくのもあまり効率的ではない。どちらもPlugin (Google)やiFilterで拡張が可能であるため、今後の改良に期待したいところだ。



[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