

---

## インターフェイスの街角 (59) — 近傍検索

増井俊之

---

---

### 検索システムの問題点

通常の計算機操作のほとんどは、広義の“検索”というよいでしょう。UNIX でもっともよく利用されているのは、ls や grep などのファイルを探すためのコマンドだと思います。また、メーラー (MUA) を使ってメールの返事を書くといったごくありふれた作業でも、

- デスクトップやメニューからメーラーを検索
- 返信アイコンを探してクリック
- リストから返事すべきメッセージを検索
- 返信の文章を作成
- 宛先を検索してセット
- 送信ボタンをみつけて送信

という手順になります。文章を書くときに利用する“かな漢字変換”も辞書からの漢字検索と考えれば、計算機における操作の大半は検索ということになります。したがって、検索システムは計算機の操作性を向上させるうえできわめて重要な要素といえます。

計算機上での一般的な検索では、情報の属性や分類などの“構造”をもとにする方法と、情報の内容から文字列を探す方法がもっとも一般的でしょう。ls や find などはディレクトリの階層構造にもとづいて検索するツールであり、書籍にたとえれば目次から本文の内容を探す方法に相当します。また、grep や Namazu などの全文検索は内容から検索する手段であり、書籍でいえば索引で本文の内容を探す方法に相当するでしょう。

従来は検索といえばこれら 2 つの手法が主流でしたが、最近では“超整理法”のように時間を基準に検索する方法も

ポピュラーになってきています。これは、目次や索引以外のレベルでの検索が可能ということを示した点で大きな意義があったといえます。

これらの方法は、どちらかといえば独立して使われることが多いようですが、データベースに対して検索処理を明示的に指定するという意味では似ています。

一方、人間の頭のなかには階層構造になっていませんし、索引もなく、時間順に並んでもいません。しかし、それでも必要十分な情報検索能力をもっています。たとえば、かつて何かをした場所を思い出そうとした場合、場所そのものをすぐに思い出せなくても、そのときの天候や季節、時刻、会話、誰と一緒にいたかといった情報をたぐっていけば、たいして必要とする情報にたどり着けるはずですが、構造化されていないのに情報をきちんと思い出せるのは、情報のあいだに連想が形成されており、連想がうまく働いたためだと思います。

人間の頭のなかにはいろいろな情報が適当に放り込まれていて、まったく構造化されていないようにもみえますが、必要な情報は連想によつて的確に探しあてられる仕組みになっているようです。これに対して計算機の場合には、ファイル名やディレクトリ階層といった構造やキーワードを正しく指定しなければ検索をおこなうことすらできません。計算機上でも、各データのあいだで関連性を見つける連想が働き、構造やキーワードなどを意識せずに検索できるようになれば、頭のなかでの検索と同じような感覚で情報を探し出せるかもしれません。

2000年5月号で、Q-Pocket という情報管理システムを紹介しました。Q-Pocket では、ファイル名やディレクトリ構造などを考慮せずにデータをどんどん蓄積していても、時間順に情報を管理して曖昧検索をおこない、目指

ず情報を的確に検索できるようにしています。

Q-Pocket はテキスト検索は得意ですが、キーワードが曖昧だと検索自体が難しくなります。さらに、写真などのテキストを含まない情報は検索対象にできないという問題がありました。また、各データが関連づけられているわけではないので、データ間の関係を用いた検索も不可能でした。

そこで、キーワード検索の代わりに連想的なリンクを重視し、頭のなかの情報を連想によって探しあてるのと同様な方法で計算機内の情報を検索する「近傍検索システム」を試作してみました。

---

## 近傍検索とは何か

近傍検索とは、1つの情報をもとに、その近傍の情報をたどる操作を繰り返していくことによって目的とする情報を探し出す手法です。

まず、ある情報と内容や作成日時、置き場所などが近いものを“近傍情報”と定義します。そして、1つの情報を表示するときは、つねにその近傍情報も表示することにします。近傍情報のなかから1つを選択すると、さらにその情報に関する近傍情報も表示されるため、関連する情報を連想的にたどって最終的に目的とする情報に到達できる可能性が高くなります。

まったく単純なコンセプトですが、これだけでも次のようにけっこう強力な連想検索が可能になります。

例 1：ある製品をどの展示会で見たのか思い出したい

- 会場で、誰かとその製品について話したことを思い出す。
- 展示会と一緒にいった人の情報を Web で調べる。
- その人に関連する写真リストのなかから展示会で撮ったものを選ぶ。
- 写真を撮った日時から展示会を特定する。

例 2：“なぞなぞドア”の写真を探す

- UNIX MAGAZINE の記事でとりあげたのだから、その原稿を書いたころに写真を撮ったはずだと思う。
- UNIX Magazine ディレクトリを参照する。
- なぞなぞ認証の原稿を参照する。

- 原稿の執筆前後に撮った写真のなかから、なぞなぞドアをみつける。

例 3：学会で発表された論文を探す

- その学会の直後に山登りに行ったことを思い出す。
- 山に登ったときの写真を探す。
- その写真を参照し、その前後に開催されたイベントを調べる。
- そのなかから学会を選び、Web ページで公開されている論文を入手する。

通常、人間はこのような作業を自然に頭のなかでおこなっているはずですが、計算機上ではそれほど簡単に実行できません。上記の 3 番目の例でいえば、

1. 山登りに行ったことを思い出す
2. ファイルブラウザや写真管理ツールでそのときの写真を探して撮影日時を調べ
3. 予定表を開いてその日時に移動して学会の予定を参照し
4. 学会のホームページに移動して検索する

といった一連の作業が必要になります。近傍検索システムを利用して、これらの作業をすべてクリック 1 つで実行できるようにしておけば、このような連想的な検索もかなり楽になるはずです。

構造を利用するものにしても、キーワードにもとづくものにしても、現時点で利用可能な計算機上の検索システムでは、人間がなんらかの指定を明示的におこなう必要があります。しかし、構造を使うのか、それともキーワードで検索するのかなどと、いちいち方法を指定すること自体が面倒です。毎日のおこなっている作業とはいえ、ファイルを探そうとするたびに `cd` や `ls` でディレクトリをたどるのは、あまり効率的とはいえません。

Google などの全文検索システムが身近になってきたとはいえ、キーワードの指定による検索は、ブラウザ上でクリックの繰り返しによって情報を探す方法にくらべると、それなりに手間のかかる作業だと思います。

近傍検索システムは、目の前に見えているリストのなかからなんとなく必要そうなものを選択するブラウジング操作を繰り返していけば、最終的には目指す情報にたどり着けるというアイデアにもとづいています。

図 1 ezplus ディレクトリを参照したところ



図 2 写真を選択



そもそも、Web のようなハイパーテキスト・システムは誰でも使える検索機構として考案されたものではないでしょうか。たとえば、Yahoo!に代表される“ディレクトリ検索サービス”は、情報の構造化によってハイパーテキスト検索を可能にしたシステムといえます。

これらのシステムのように情報が適切に構造化されていれば、リンクをたどることで情報を効率的に探し出せるかもしれません。しかし、個人が作成したファイルなどは、そもそもデータが構造化されていないので、ファイルシステムのディレクトリをたどる程度の検索が精一杯です。

近傍検索システムを利用すれば、自分の計算機に蓄積されたファイルのように構造化が不十分な場合でも、近傍情報にもとづいて自動的に構造を作り、Yahoo!と同様な感覚で連想的に検索していくことが可能になります。

## 近傍検索システムの例

以下では、ブラウザから使えるように CGI で実装した近傍検索システムの例を紹介します。

図 1 は、近傍検索システムで ezplus/index.html を参照したところです。ページの一番上の入力枠に /user/masui/ezplus/index.html と指定して近傍検索システム (neighbour.cgi) を呼び出すと、このディレクトリに置かれているファイルが一覧表示されます。同時に、このディ

レクトリの更新日時に近い時刻のファイルアクセスやメモ作成などもすべてリストされます。

8月号で au (KDDI) の GPS 機能付き携帯電話のプログラム開発について解説しましたが、ezplus ディレクトリにはそのときの資料が入っています。通常のファイルブラウザでディレクトリをアクセスしても、参照できるのはそのディレクトリに置かれているファイルの内容だけです。一方、近傍検索システムでは、index.html とほぼ同じところに作成したメモやファイル、撮った写真なども同時に一覧できるようになっています。

私が常用しているマシンでは、重要なファイルを書き出したときはその時刻を記録するようにしてあります。また、アイデアなどのメモ類は日付と時刻からなるファイル名で保存しているので、ある日時近辺のアクティビティを簡単に一覧表示することができます。

作成日時が 2002/6/24 となっている ezplus ディレクトリには、記事で紹介したシステムに関連する資料が置かれているので、すぐ近くに関連する情報やファイルが散見されます。その直後には“慶應講義”というイベントも記録されています。ここには、所用で慶應義塾大学湘南藤沢キャンパス (SFC) に行った際、研究室を見学させていただいたときの写真もリストされています。つまり、携帯電話の Java 開発キットである ezplus ディレクトリを表示することにより、それを紹介した原稿ファイルや関連する

図 3 怪しげな電球の写真を選択



アイデア、近い日付のイベントや写真などをすべて1つのページ内に表示することができるわけです。

写真のなかから1枚をクリックすると、画面はその写真を中心として図2のように変化します。これらの写真には、キーワードも付けられます。その場合には、注目しているものと同じキーワードをもつ写真の一覧が表示されます。この日に撮った写真には「20020625 慶應 SFC」というキーワードを付けておいたため、このキーワードをもつすべての写真がリストされています。

ここで、研究室にあった「怪しげな電球」を撮った写真に「電球」というキーワードを付けてみます。そして、この電球の写真を参照すると、今度は図3のように「電球」というキーワードも表示されます。

これだけなら、よくある写真整理システムと同じですが、近傍検索システムではキーワードを用いた連想的な検索が可能になっています。写真には「電球」というキーワードが付けられているので、このキーワードをもとに検索すると図3のようなページが得られます。ただし、普通の写真整理システムとは異なり、写真そのものだけでなく、撮影日時前後にあったイベントや、そのころに作成したファイルなどもリストされているため、「怪しげな電球を見たのと同じころに作ったプログラム」もすぐに探し出せます。また、そのとき一緒にいた人をキーワードとして登録しておけば、その人物から検索を進めていくことも可能

です。このように、関連する情報をどんどん連想的にリンクして検索できるようになるわけです。

このシステムでは、「あの怪しげな電球をいつどこで見たのか思い出したい」「SFCに行ったとき、何を見たかを思い出したい」「ある人とどこかを見学したはずだが、どこだったかを思い出したい」といった漠然とした検索が、すべてクリックを繰り返していだけでおこなえるようになります。

## 近傍検索がもたらすもの

近傍検索が当たり前になれば、ファイル管理の手法も従来とはかなり変わってくるかもしれません。

通常、UNIXのファイルはディレクトリ構造をもとに管理するので、関連性のあるファイルは同じディレクトリか、近くのディレクトリに置いておく人が多いと思います。たとえば、文章と密接に関連する写真や画像は、たいていその文章と同じか、近くのディレクトリに保存しておくでしょう。さもないと、あちこちに散らばって、どこに何をしまったかが分からなくなるからです。

しかし、作成日時などをもとに近傍情報に簡単かつ確実にアクセスできるのであれば、無理をしてまで同じディレクトリに置く必要はなくなります。近傍検索システムの場合は「時間的に近いもの」を基準としてアクセスできるので、写真ファイルが文章とはまったく異なるディレクトリに置かれていても問題はありません。

このように、いろいろな属性ごとに「近傍情報」をもとにした検索ができると、属性を利用して関連する情報のあいたを縦横無尽にジャンプすることもできます。時間だけでなく、内容や作成者など、多種多様な属性ごとに近傍に存在する情報を簡単に参照できれば、従来とまったく異なる手法によるファイル管理も可能になります。

## キーワードと連想的検索

デジタルカメラの普及とともに、デジタル画像を管理するソフトウェアも一般的になりつつあります。たしかに、それぞれ特色のある検索手法が用いられていますが、いずれもいまひとつ有用性に欠けるような気がします。

写真の検索だけが目的であれば、キーワードを付けてもあまり嬉しくありません。けっきょくのところ、キーワー

ドや色彩などの情報にもとづいた写真の検索が最終目標になっていて、その写真からさらに別のデータを検索することはできないのが、“いまひとつ”と感じる理由の1つのような気がします。

たとえば、これらのソフトウェアでは、文字列や撮影日時などをキーワードにした検索は可能でも、撮影日時に関連する日記やメモなどのデータにはアクセスできないのが普通です。しかし、さきほど挙げた例のように、まずキーワードから写真を検索し、続いてその写真に関連するイベントやメモを検索し、……といった連想的な検索が可能になれば、写真以外のさまざまなファイルとも有機的につながり、あらゆる関連情報が容易に検索できるのではないのでしょうか。

おそらく、人間の頭のなかでもそのような方式で検索がおこなわれているのでしょう。上記のような有機的な検索が可能になるのなら、写真にキーワードを付ける手間は十分にもとがとれるように思います。

---

## おわりに

最近の電子百科事典のなかには、近傍検索システムに似た連想的検索が可能なものもあります。このような事典では、たとえば“ジンギスカン”をキーワードとして検索すると、ジンギスカンという人物についての情報だけでなく、時代背景や地理、同時代のほかの地域の様子なども提示されます。これによって、ジンギスカンが生きた時代についての総合的なイメージが得られるように工夫されているわけです。こういったシステムを利用すれば、1つの情報を単純に暗記するのではなく、複数の情報を関連づけて記憶にとどめられるので、効率的に学習できるだけでなく、いったん憶えた情報を忘れにくくなるといった効果もあるようです。

さまざまな分野の学問や個人情報管理、そして検索などの世界において、関連づけを有効に利用した有機的な検索手法がさらに重視されるようになってほしいものです。

今回紹介したシステムは、私自身の環境に強く依存したものになっているため、まだ公開できる段階にありません。今後、より汎用的にしたうえで配布しようと考えています。

(ますい・としゆき ソニー CSL)