

近傍関係を活用した情報検索

増 井 俊 之[†]

ネットワークや計算機上において様々な情報検索システムが使われているが、多くのシステムにおいて、ユーザがなんらかの形で検索条件を明示的に示してから検索を開始するのが普通である。一方、人間の頭の中では、明示的に検索条件を示すのではなく、漠然とした連想によって情報検索が行なわれていると考えられるが、計算機上の検索においても、これに似た連想的な検索が有効に働く場合がある。本稿では、現在ユーザが着目中の情報に近い情報を自動的にシステムが提示し、ユーザは関連リンクをクリックするだけで連想的な検索を行なうことができる「近傍検索」システムを提案し、その効果について述べる。

Information Navigation by Neighbor Hopping

TOSHIYUKI MASUI[†]

When a user wants to find information on a computer, he usually has to choose the retrieval method and retrieval keyword before he can start the search operation. This means that a user has to know many things beforehand, including the structure of the data set, efficiency of available retrieval methods, effectiveness of using keywords, etc.

In this paper, we propose an information navigation technique based on the closeness between data items. When a user starts viewing a data item, similar items are listed automatically, and the user can select one of them and continue the operation. Since the similarity of information is calculated from various metrics, users can navigate through the wide information space only by selecting an appropriate data item from the list. For example, a user can start by viewing a picture, find a keyword attached to the picture, find another picture which has the same keyword, find a memo written on the day when the picture was taken, find another memo with similar contents, and so on. In this way, he can browse and search the information he needs only by selecting data items, without worrying about what search method should be used for the navigation.

We have implemented an information navigation system *NeighborQuest* based on this idea, and we show how people can use NeighborQuest to find the information they need only by repeatedly selecting related information.

1. はじめに

インターネットや計算機上では様々な検索システムが使われているが、情報の属性や分類のような構造をもとに検索する方法と、情報の内容をもとに検索する方法が主に使用されている。Unix の `ls`, `find` などのコマンドや、インターネットの Yahoo ディレクトリ、書籍の目次などは階層構造をもとに情報を検索するために使われており、Unix の `grep` コマンドやインターネットの Google のような全文検索システム、書籍の索引などは内容をもとに検索を行なうために使われている。これらの使われ方はかなり異なるが、カテゴリやキーワードのような検索条件を明示的にユーザが指定して検索を行なうという意味では類似している。

一方、World Wide Web はリンク関係にもとづいた大規模な検索システムであると考えられる。Web 上でブラウジングを行ないながら情報を捜す場合は、明示的に検索条件を指定するかわりに、リンクによって関連づけられた情報をたどっていくことにより情報検索が行なわれる。

人間の頭の中は計算機データのように構造化されていないし、索引も存在しないため、条件を指定した検索を行なうことはできず、Web のブラウジングにおける検索のように連想的に記憶をたどることによって、必要な情報検索を行なうことができるようになっていると考えられる。たとえば、何かを昔行なった場所を思い出そうとする場合、直接その場所の情報を思い出すことができなくても、そのときの天気/季節/時刻/会話/人物関係/のような関連情報を思い出すことにより、必要な情報にたどりつくことができることが多い。構造化されていないのに情報を思い出すことができるの

[†] 産業技術総合研究所 情報処理研究部門
National Institute of Advanced Industrial Science and Technology
masui@acm.org

は、情報の間に関連が形成されており、連想がうまく働くためだと思われる。

人間の頭の中には膨大な量の雑多な情報が格納されており、全く構造化されていないと思われるにもかかわらず、必要な情報を連想によって思い出すことができるのに対し、コンピュータの場合はファイル名や階層構造のような構造や検索キーワードをきちんと指定しなければ検索を行なうことは難しい。計算機上の情報についても、関連した情報が関連づけられて連想的に検索できるようになっていけば、構造や検索キーワードについて深く考えることなく、人間の頭の中の検索に似た柔軟な検索を行なうことができると考えられる。

本稿では、頭の中の情報を連想的に検索するのと同じような方法で計算機内の情報を検索可能な近傍検索システムを提案し、その実装と評価について述べる。

2. 近傍検索システム

近傍検索システムは、ある情報をもとにしてその近傍の情報を選択するという操作を繰り返していくことにより必要な情報への到達を可能にするシステムである。ある情報に対し、それに内容が近い情報/ 同じ属性をもつ情報/ 作成時刻が近い情報/ 格納場所が近い情報/ などをその情報の近傍情報であると定義し、情報を表示するときにはいつもこのような近傍情報を同時に表示して注目点を移動可能にする。ある情報の近傍情報の中からひとつを選択すると、今度はその情報を中心とした近傍情報が表示される。この操作を繰り返すことにより、関連情報を連想的にたどって目的の情報に到達することができる。

図1のようにイベント/写真/メモ/メールなどの雑多な情報がいろいろな属性によって関連づけられている場合、時刻が同じであるとか内容が近いとかいったリンクをたどっていくことによって、連想的に情報を検索していくことができる。

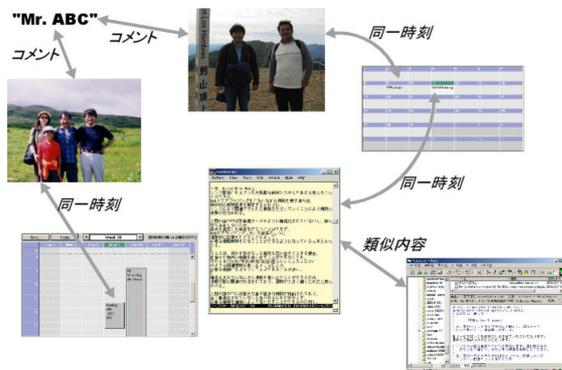


図1 近傍関係の例

近傍検索システムを使った情報検索のシナリオの例を以下に示す。

例1: ある製品を見た展示会を思い出したい

- 会場で誰かとその話を議論したことを思い出す
- 展示会と一緒にいった人の情報を参照する
- その人に関連する写真リストの中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: 東大寺南大門の写真を捜す

- 奈良案内システムを昔作ったことを思い出す
- 奈良案内システムのディレクトリをさがす
- プログラムの作成日付を調べる
- プログラムを書いたころの写真を格納したディレクトリを参照し、その中から南大門の写真を捜す

例3: 「なぞなぞドア」⁶⁾の写真を捜す

- 雑誌に記事を書いたのだから、原稿を作成したころに写真を撮ったはずだと思う
- 雑誌記事の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
- 原稿を作成した周辺の時期の写真から、なぞなぞドアをさがす

例4: 以前どこかの学会で発表された論文を捜す

- その学会の直後に山登りに行ったことを思い出す
- 山登りの写真を捜す
- その写真を撮った日の前後のイベントを調べる
- その中から学会を選び、Webページに行って原稿を取得する

例5: 類似システムのソースを捜す

- 携帯端末用のシステムのプログラムソースを参照する
- 同じディレクトリにある Makefile を参照する
- その Makefile と中身が似た Makefile を参照し、類似システムのソースをみつける

このような連想的な検索は、誰もが頭の中では常に実行していると思われるが、計算機上では簡単には実行できないことが多い。たとえば例3の場合、まず山登りに行ったことを思い出して、ファイルブラウザや写真管理ソフトでその写真を捜し、ファイルの属性を調べることによって撮影日時を調べ、予定表を参照してその日の周囲の予定を参照して学会を捜し、学会のホームページを検索するといった手間が必要になる。それぞれの検索ステップにおいて必要な操作は異なっており、やりたいことと実際の操作とのマッピングは難しいため、計算機の操作に詳しいユーザでないとなかなか連想的な検索を手際よく実行することは難しい。これらのステップがすべてクリックひとつで実行

できるようになっていけば、このような連想的な検索を誰でも簡単に実行できるようになる可能性がある。

構造を利用する検索でもキーワードによる検索でも、現在一般的な検索手法では、検索条件を人間が明示的に指定しなければならないのが普通であるし、構造を用いて検索するのか/キーワード検索をするのかなどといった検索手法を指定するのは面倒なことである。フォルダウィンドウを開いたり、Unixのlsコマンドを発行したりしてファイルの存在を確認するような作業は、計算機ユーザが毎日のように行なっている作業ではあるが、常に計算機内のデータの構造を把握しつつ適切な操作を行なわなければならないため、心理的負担が大きな作業である。また、GoogleやNamazuのような全文検索システムを用いて内容検索を行なう場合は、全文検索で情報がみつかる可能性が高いことを理解したり検索の操作方法を覚えたりしなければならないため、やはり負荷が高い作業であるといえる。この方法を使ってどのように検索を行なうかを常に即座に判断しながらうまく情報にアクセスできるようになるためには習熟が必要であり、計算機の利用をさまたげるひとつの原因になっていると考えられる。近傍検索システムでは、見えているリストの中から必要そうなものを選択するというブラウジング操作を繰り返すだけで結果的に目的の情報に到達できることを目標としている。

3. 近傍検索システムの実装

以上のような考えにもとづき、Webブラウザで使用するこのできる近傍検索システム NeighborQuest を試作した。NeighborQuest は CGI プログラムとして動作する。

NeighborQuest では以下の基準にもとづいて近傍度を判定し、注目しているファイルの近傍にあるファイルを周囲にリストする。

- 注目しているファイルの含まれるディレクトリ構造
- ファイルの更新日時
- ファイルの内容

ファイルの内容の類似度計算には文書間の類似度計算ライブラリ GETA (Generic Engine for Transposable Association) ⁷⁾ を使用している。検索対象となる全てのファイルについてあらかじめインデックスを作成しておき、tf・idf法にもとづいた類似度が注目ファイルに近いものをリストしている。

図2は、NeighborQuestで“ezplus/index.html”とい



図2 “ezplus”ディレクトリを参照したところ

うファイル参照したところである。ページの一番上のファイル名入力枠で“ezplus/index.html”を指定することにより NeighborQuest を呼び出すと、このディレクトリの中のファイルがリストされ、またこのファイルを更新した時刻近辺で作成されたファイルやメモなどがすべてリストされる。

このディレクトリには au の GPS 携帯電話のプログラム開発に関する情報が格納されている。通常のファイルブラウザでディレクトリをアクセスした場合はそのディレクトリの内容しか参照することができないが、NeighborQuest では、index.html を作成したのと同じ頃に作成したメモやファイル、そのころ撮影した写真なども同時に一覧できるようになっている。

著者が使用している計算機では、ファイルをエディタで書き出したときはその時刻を記録するようになっており、またアイデアなどのメモは日付/時刻をもとにしたファイル名に格納しているため、ある時刻近辺のアクティビティを簡単にリストすることができる。

ezplus というディレクトリの作成時刻は 2002/6/24 であるが、このディレクトリは雑誌記事で解説したシステムに関連したものであり、すぐ近くに関連する情報やファイルが散見されている。また、その直後に「慶應講義」というイベントも記録されており、その頃に慶應大学 SFC を訪問し、研究室を見学したときの写真もリストされている。つまり、携帯の 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、それに関連した雑誌原稿のファイルや関連したアイデア、同時期のイベントや写真などをすべてひとつ

<http://geta.ex.nii.ac.jp/>

のページに表示することができたことになる。

写真のひとつを選択すると、画面はその写真を中心として図3のように変化する。写真にはキーワードをつけることができるようになっており、注目している写真と同じキーワードをもつ写真がリストされるようになっていく。この日撮った写真には「20020625 慶應SFC」というキーワードをつけてあったため、同じキーワードをもつ写真がリストされている。



図3 写真を選択

訪問した研究室には変わった電球が置いてあったので、この写真に「電球」というキーワードを付加し、その後でこの電球の写真を参照すると、今度は図4のように「電球」というキーワードも表示されるようになる。

これだけであれば、一般的な写真整理システムと同じであるが、NeighborQuestでは、キーワードを使って連想的に検索を行っていきることができる。今この写真に「電球」というキーワードをつけたので、「電球」というキーワードから図4のようなページを得ることができるが、普通の写真整理システムとは異なり、写真と同時にこの写真を撮ったとき日時付近のイベントやその頃作成したファイルなどもリストされているので、「変わった電球を見たのと同じ頃に作ったプログラム」がすぐに捜せることになる。また、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できるなど、関連する情報をどんどん連想的にリンクして検索することができるようになる。



図4 変わった電球の写真を選択

「変な電球をどこでいつ見たか思い出したい」「慶應大学に行ったときに何を見たか思い出したい」「ある人とどこかを見学したはずだが、どこか思い出したい」といた漠然とした検索をすべてクリックを繰り返すだけで実行できることになる。

図5は、カレンダーを計算するPerlプログラム“calendar”をNeighborQuestで参照した結果である。このプログラムを作成した頃の日記が下側に表示されているのに加え、このプログラムと内容が近いファイルが右上に表示されているが、ここでは全く別の場所にあるPerlプログラム“expand”がリストされている。

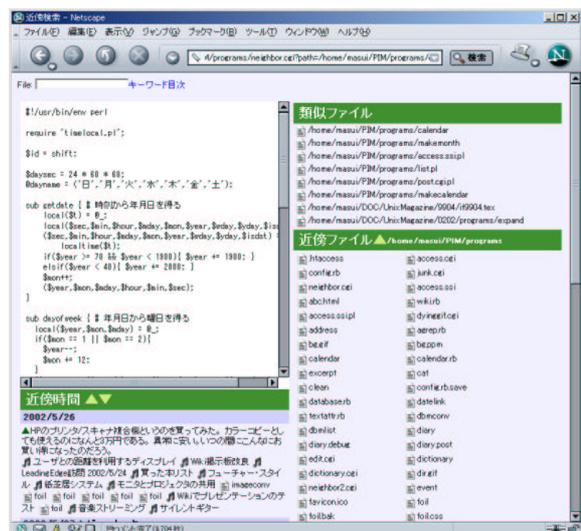


図5 プログラムの参照

“expand”プログラムを選択すると、これを中心として画面は図6のように変化する。このプログラムが作成された2001年12月ごろの情報や、ディレクトリ内のファイルなどが表示されている。

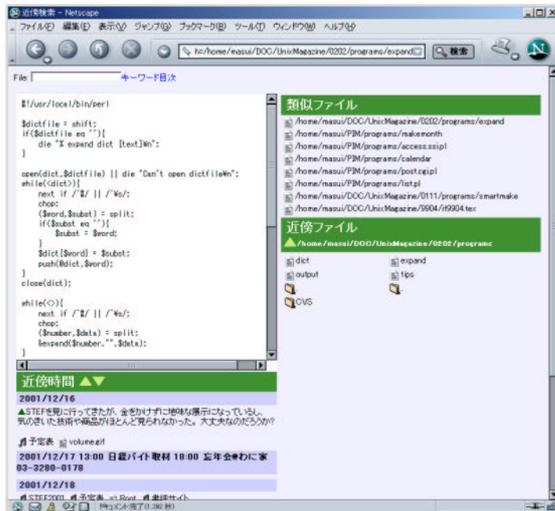


図6 似たプログラムの参照

4. 議 論

4.1 近傍を積極的に利用したファイル管理

近傍検索がうまく使えれば、従来とはかなり異なる方法でファイルを管理することができるようになる。

最近広く利用されている計算機では、ファイルはディレクトリの木構造を利用して管理するものが多く、関連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のが普通である。たとえば、前述のように、文章に関連した写真は、整理のために文章と同じか近くのディレクトリに置いておくのが一般的であるし、そうしておかなければ参照したり整理したりするのに不便である。

しかし、作成時刻などをもとに近傍情報に簡単にアクセスできることが確実であれば、同じディレクトリに置く必要はなくなってしまう。近傍検索システムの場合、時間的に近いファイルは近傍としてアクセスできるため、写真ファイルが全く別の写真ディレクトリに置いてあっても問題ない。

このように、いろいろな属性における近傍で検索ができると、属性を使って縦横無尽に関連情報間をジャンプすることができる。内容的な近さ、作者の近さなど、いろいろな属性において近傍に存在する情報を簡単に参照できれば、従来と全く異なるファイル管理が可能になるだろう。

4.2 予測入力への利用

あるテキストの近傍にあるテキストは内容や話題が近いものであることが多いため、そのテキストを編集する場合には近傍テキストの情報が参考になることが多いと考えられる。小松らは、日本語の動的補完システム Nanashiki⁴⁾ を利用することによって関連テキストを利用した効率的な予測型入力システムを提案している。小松の予測入力システムでは、予測型テキスト入力システム POBox²⁾ と同様に、ユーザの入力に応じて入力単語を予測して候補として提示するが、関連するテキストに含まれる単語を優先して候補として採用するようになっている。「議論」「原稿」のような単語をよく使用するユーザが「g」を入力した場合、通常はこれらの単語が候補になるが、「岐阜での学会」に関するメールに返事する場合は「岐阜」や「学会」のような単語が優先的に候補となる。小松のシステムでは、同じ編集バッファ内にある単語のみが優先的に候補として使用されるようになっているが、近傍テキストに含まれる単語や文章を予測に利用することにより、予測入力の効率が向上すると考えられる。

4.3 キーワードによるリンクの有効利用

デジカメの流行によって計算機上の画像の数は激増しており、様々な画像検索手法を利用した写真管理システムが普及しつつある。撮った写真に常にキーワードをつけるようにしておけば通常のテキスト検索手法を利用して写真を検索することができるので便利であるが、写真を検索するだけのためにいちいちキーワードを付加する作業は面倒であるにもかかわらず見返りが少ないため、実際に運用するのは困難と思われる。Henry Lieberman の Aria¹⁾ システムでは、メールに写真が添付されている場合のようにメールテキストと画像に明らかな関連がある場合に画像に自動的にキーワードを付加したり、簡単なユーザ操作によって画像とテキストを関連づけたりしやすいような工夫がされているが、画像をみつけることが最終目標である画像検索システムを利用している限り、キーワード付加のメリットが手間に見合うことはなさそうに思われる。

画像の検索を行なうために画像にテキストを付加するのではなく、情報間のリンクを増強するという考えにもとづいて画像にリンクを付加するのであれば状況は変わってくるかもしれない。近傍検索システムでは、ある写真にキーワードを付加するという作業は、その写真の属性を定義したというだけにとどまらず、そのキーワードに対して関連のある写真/日時/場所などを定義したのと同じことになるため、適用範囲が非常に広がる。前述の「電球」の例では、キーワードをひ

とつ追加したことにより、キーワードから写真を検索し、それに付加されたキーワードから別の写真を検索し、その写真からイベントやメモを検索し...といった連想的検索が可能になっている。単純なキーワードの付加により写真やテキストが有機的につながりあうことになり、あらゆる情報が容易に検索できるようになることが認識されれば、多少の手間をかけて写真や画像にキーワードを付加する作業は一般化するかもしれない。

4.4 自動提示にもとづくインタフェース

一般的な検索システムでは、検索条件や検索手法をユーザが指定してはじめて検索結果を得ることができ、NeighborQuestの場合はユーザは提示されている候補から必要なものを選択するだけであり、条件や手法を指定する必要がない。与えられた選択枝の中から必要なものを選ぶという行為は、何が必要なのか考えたり、必要なものを得る手段を考えたりすることに比べると負担は小さいと思われる。

水口らは、情報探索を行なうための提示型インタフェースを提案している⁵⁾。提示型インタフェースとは、ユーザが何もしない状態でも新しい情報が次々と提示されるようになっており、自分が欲しい情報が出現したときだけユーザがアクションを起こすことにより必要な情報を取得できるようにするという手法である。ユーザがあまり考えなくても情報探索を行なうことができるという点で近傍検索と提示型インタフェースは似ているが、近傍検索のユーザは能動的な選択操作の繰り返しでシステムを制御することができる点で提示型インタフェースよりは自由度が高い。また、RhodesらのRemembrance Agent³⁾は、ユーザが作成中のテキストに内容に近いファイルをシステムがリアルタイムに検索してユーザに提示することによって関連情報を思い出したり入力作業を効率化したりできるようにしたシステムであるが、これらのシステムのように、検索条件指定や検索開始/終了処理など、ユーザの試行錯誤的な操作を極力少なくし、ユーザはただ選択操作のみ行なうにしたシステムは有望であると考えられる。

5. 結 論

リンクのクリックを繰り返すことにより情報探索を行なうというスタイルはWebブラウザの普及により非常にポピュラーなものとなっている。本稿では、近傍にある情報へのリンクを自動生成することによりWebブラウジングと同じ使い勝手に連想的に情報探索を行なうことができる近傍検索システムを提案し、その特

長について述べた。本手法をさらに使いやすく改善することにより、既存の計算機操作の多くを置き換えることを目指していきたい。

参 考 文 献

- 1) Lieberman, H., Rosenzweig, E. and Singh, P.: Aria: An Agent For Annotating And Retrieving Images, *IEEE Computer*, pp. 57-61 (2001).
- 2) Masui, T.: POBox: An Efficient Text Input Method for Handheld and Ubiquitous Computers, *Proceedings of the International Symposium on Handheld and Ubiquitous Computing (HUC'99)*, pp. 289-300 (1999).
- 3) Rhodes, B.J. and Starner, T.: Remembrance Agent: A continuously running automated information retrieval system, *Proceedings of The First International Conference on The Practical Application Of Intelligent Agents and Multi Agent Technology (PAAM '96)*, pp. 487-495 (1996).
- 4) 小松弘幸, 高林哲, 増井俊之: 日本語動的単語補完方式 Nanashiki を活用した予測入力, インタラクティブシステムとソフトウェア IX: 日本ソフトウェア科学会 WISS2001, 近代科学社, pp. 67-74 (2001).
- 5) 水口充, 浦野直樹: 提示型インタフェースによる情報探索システム, インタラクティブシステムとソフトウェア VI: 日本ソフトウェア科学会 WISS'98, 近代科学社, pp. 159-164 (1998).
- 6) 増井俊之: なぞなぞドア, *Unix Magazine*, No. 8 (2001).
- 7) 西岡真吾, 今一修: 汎用連想計算エンジン GETA とそれに基づく連想検索システム, 情報処理学会自然言語処理研究会 研究報告, No. 137 (2000).