オフィス文書自動検索の一実現方法

7Q－2

落合尚良　池田崇博*1　野村直之*1
NEC情報システムズ　*1 NEC情報メディア研究所

1 はじめに

GUI上でワープロ、メール執筆等の本業を行っている時に、バックグラウンドで関連文書を自動的に検索して提示する、オフィス文書自動検索の一実現方法を提案する。近年のオフィス環境では、１人１台のＰＣが普及し、ウィンドウシステム上で、ユーザーが手動で気軽にマルチタスクを扱えるようになっている。そのおかげで、本業に関連する、検索などの別作業を割り込ませながら仕事を進める様態が一般的となっている。この従来無かった環境では、マルチタスクのおかげで業務の効率が上がる一方、本業への集中度が落ちて、次のような問題点が現れるようになった。

問題点1) *割り込み作業の内容により気が散る、本業から脱線、何をしようとしていたか忘れ、重要でない作業に走る。*

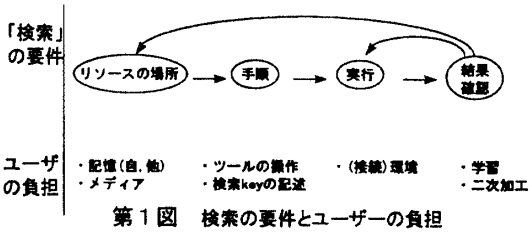
問題点2) *割り込み作業の遂行のための手順の想起、関連情報の所在の想起、手順の実行、結果の確認方法の想起などに思考のコストおよび時間コストがかかる。*

これらの新たな問題点の解決につながる技術の中で、ユーザによる曖昧な指示を解釈し、必要に応じて時間／空間の制約を越えて「良きにはからって」くれるソフトウェアはエージェントと呼ばれる。本稿では、エージェントによる代行に問題をシフトさせず、解決の対象を上記の問題点2)に絞って、「自動検索」の新しいユーザインタフェースを提案する。

従来の関連研究では、検索の条件指定を自動化する試み、すなわち、明示的にキーワードを指定せずに類似文書を検索する文書アクセスを行うための手法として、たとえば文献[1]がある。但し、主眼は文字列成分分布の類似度比較という要素技術の側にあり、具体的にどのような状況でユーザの検索行為を省力化できるかが明示されてはいない。

2 従来の検索ユーザインタフェースの問題点

「検索」の成立には、「情報の所在の確認」、「情報へのアクセス手順の確認」、「アクセス・入手の手順の実行」、そして「検索結果の確認のための操作」の４つの要件が成立しなければならない。最後の「検索結果の確認のための操作」には、検索結果から必要なものを短時間で選びやすくしたり、場合によっては発想支援や深い理解のために検索結果の「眺め」＝ビューを複数の軸で切り替える操作なども含まれる。複数の軸で検索結果をランキングして表示するビューは、LotusNotesをはじめとするグループウェア・クライアントの標準機能となっている。



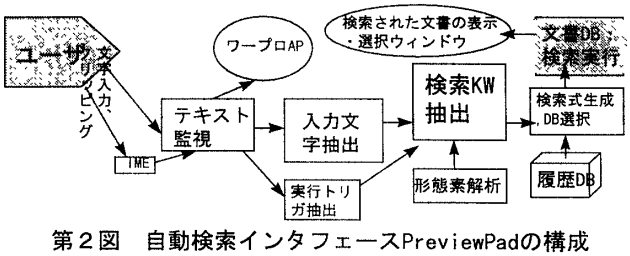
第１図　検索の要件とユーザーの負担

第１図に上述の、検索のための４つの要件をまとめた。右向きに、隣り合う順に並ぶ矢印は、時間軸上の順序を表わしている。左向きに戻る２本の線は、１つの検索の結果の確認後、繰り返し再検索をする際の典型的な操作の流れを示す。

前記問題点2)とは、４つの要件のいずれが欠けても検索行為が成立しない、ということから発生するユーザの負担である。「リソースの場所」を記憶の断片やファイルシステムに乱雑に分散するメモ・ファイルから手繰り、操作法を思い出す。また、検索式記述のシンタックス規約を参照し、システム側の都合による応答時間にしばられ、さらに検索結果の読み方や再利用の方法も学習・想起することになる。

3 自動検索インタフェースPreviewPadの設計

前記の検索の要件の全てが欠けた場合でも、現在のユーザの本来作業に関連する文書が自動的に検索されるようなインタフェースPreviewPadを設計した。第２図にその構成を示す。



第２図　自動検索インタフェースPreviewPadの構成

An Architecture for Automatic Office Document Retrieval, Takayoshi Ochiai, Takahiro Ikeda and Naoyuki Nomura NEC Corporation, 4-1-1 Miyazaki, Miyamaeku, Kawasaki 216

GUI環境でユーザが前面ウィンドウに入力したり、コピー・バッファへのクリッピングを行ったテキストをシステムが監視し、本来の送出先へ送るのと同時にテキスト内容を取得し、解析対象とする入力文字列を抽出する。同時に、ユーザのタイピングが一定以上途絶えた等、GUI上の適当なイベントの直後のタイミングを実行トリガとして抽出し、入力文字列に対して形態素解析を実行する。ここで、固有名詞やサ変名詞の取捨選択等、予めユーザが指定したパラメータに従って検索キーワードを抽出する。

続いて、過去に発行した検索式やその選別・評価の結果、さらにバックエンドの文書データベースの所在や格納形式、それらのアクセス・コストを含む履歴データベースを参照しながら、検索キーワードを組み合わせた検索式を生成する。この際に、過去に類似の検索式の送り先となった文書データベース名を履歴データベースから参照しながら、高ヒット率が見込まれ且つ検索結果が高速に低コストで得られることが期待される文書データベースを選択する（文献[2]）。

こうして選択された文書データベース（一般には複数）に対して発行済みの検索式を実行し、その結果を、本来作業のウィンドウの近傍に配置したウィンドウPreviewPadに表示する。PreviewPadには、ヒットした文書群の一覧とともに、最高ゆう度の文書内容を表示する。デフォルトモードでは、以上が全て自動的に連動する。検索式生成のための各種条件、履歴データベースの参照／非参照、そして検索された文書の表示・選択ウィンドウを広げておくか、あるいは平常時はアイコン化しておくか、等はユーザ指定のパラメータによりモードを変更できる。

4 PreviewPadの実装と評価

クライアント機上で、形態素解析や検索式の発行、送出などのプロセスを実行するため、強力なマルチタスク機能を備えたWindows NTをターゲットOSとした。開発環境としては、Delphi2.0Jを用いた。バックエンドの文書データベースには、不定形マルチメディア文書を扱うことができ、且つ添付文書の内部にまで全文検索の実行可能なLotusNotesを用いた。API関数によってデータベース内容を直接読み書きできる文書アクセス機能モジュールを作り、その検索結果をクライアント機上で実行中のPreviewPadアプリケーションに引き渡す、という実装を行った。

評価の際に組み合わせたワープロソフトは、一太郎７R1およびMSWord95、日本語入力IMはATOK10およびMSIME95である。対象ウィンドウごとに自動検索結果の履歴を管理し、内容表示、一覧表示を行う。その表示内容からのクリップボード経由による再帰的な検索も実行可能とした。

約１万文書、２０ＭＢの、主としてテキストのみからなる文書データベースに接続した状態で、一太郎７R1およびMSWord95の４文書間でウィンドウを切り替えながら入力を続け、数日間使用した。主記憶48MB、周波数166MHzのPentiumマシンで、特にストレス無く執筆作業を続けることができた。これは、第２図に示した各機能モジュールが各々ほぼ線形時間で実行されることから、妥当な結果である。つまり、実装の具体的な出来やコンパイラの性能によって体感速度が決まる。

肝心の自動検索に要する時間であるが、これは文書データベースの所在、サイズ、ネットワークトラフィック等に依存するため、一概には評価できない。但し、全文検索のパフォーマンスに関して言えば、その性能をほぼそのまま享受できる。これは、先述のグループウェア・クライアントに比べて、大量の文書が存在した時には勝る点である。なぜなら、自動検索機能をもたないグループウェア・クライアントでは、存在する全文書に対して一旦線形時間のコストをかけてビューを構築しない限り、中身の文書を閲覧できないからである。１万文書の規模のデータベースでは数分間を要することがある。全文検索では文書データベースのサイズに比例するような時間コストはかからないため、大量の文書が溜まれば溜まるほど、本PreviewPadの比較優位性が高まることになる。

5 おわりに

従来検索の成立に不可欠であった４つの要件全てを欠いても自動的に検索を実行するインタフェース、PreviewPadを設計・試作し、利用可能性を評価した。本インタフェースでは、本業の執筆を継続しつつ、行き詰まった場合や参考文書の必要性を感じた場合にのみ、 PreviewPadを開いて既に検索済みの文書群をブラウズするのを典型的な使用形態とする。ここには第１図に示した、検索自体に要するユーザの操作はなく、且つ、検索結果を必要としない限り、従来の本業で行っていた以外の一切のキー入力、マウス操作が不要である。

最近入力した、あるいはクリップしたテキストに含まれるキーワードを含む、ネットワーク上でアクセス可能な関連文書が既に検索された状態を実現するため、従来のホワイトカラーのワーク・スタイルを保ったまま、発想支援効果の拡大を期待することができる。今後は、これらのメリットの定量評価を試みるとともに、キーワードの選別手法に様々な要素技術を適用し、取捨選択・評価の試行を繰り返すことにより、広域ネットからの大量文書からの関連文書の絞り込み性能を上げるなどの改良を施していきたい。

参考文献

[1]湯浅他「Document Query by Example文書の例示検索」，情報処理学会第52回全国大会
[2]池田他「広域ネットを含むバックエンドデータベース仮想化の一手法」，情報処理学会第 52 回全国大会