情報検索システムにおける効果的なナビゲーション機能の提案

諸橋 正幸, 堤 泰治郎, 丸山 宏, 野美山 浩
日本アイ・ビー・エム株式会社 東京基礎研究所
〒 242 神奈川県大和市下鶴間 1623-14
Tel: 0462-73-4670, Fax: 0462-73-7413, E-Mail: moro@trlvm.vnet.ibm.com

概要

情報検索システムを利用する際に用いられる代表的な検索方法には、シソーラスのような階層的構造を上から辿っていく方法と、体系化されていないキーワード（フリーターム）とその論理演算で検索する方法の２種類があるが、いずれの場合にも、漏れを生ずることなく適切な数にまで文献を絞り込む過程にかなりのノウハウを必要とするため、素人にはなかなか使いこなせない。

本稿では、この絞り込みの過程でとらえられた文献セットをさまざまな観点から、概観するウインドウを提供することで、本当に欲しい文献セットへと導いていくナビゲーション機能を試作した。

さらに、この機能が、単なる特定の情報をさぐるためのナビゲーションとしての役割ばかりでなく、文献ＤＢの中に埋没している事実を発見する道具としても役立つことを紹介する。

キーワード

情報検索，ナビゲーション，情報の輪郭，データマイニング

Proposal of an effective navigation for information retrieval systems

Masayuki MOROHASHI, Taijiro TSUTSUMI, Hiroshi MARUYAMA, Hiroshi NOMIYAMA
IBM Research, Tokyo Research Laboratory
1623-14, Shimotsuruma, Yamato, Kanagawa 242, Japan
Phone: +81 462-73-4670, Fax: +81 462-73-7413, E-Maile: moro@trlvm.vnet.ibm.com

Abstract

On using information retrieval systems, one of the problems which the users face with is that they have no way of knowing whether the sequence of the conditions, which the users have given to the system, is proper or not before they look through every document of the final results.

We proposed a navigation function that gives information outlining of the selected document set every after inputing a piece of retrieval conditions. The information outling also helps you to find a new aspect of the selected document set, such as the recent topics in a specific research area.

Keywords

Information Retrieval, navigation, Information Outlining, Data Mining

1．はじめに

情報検索システムを利用する際に用いられる代表的な検索方法には、シソーラスのような階層的構造を上から辿っていく方法と、体系化されていないキーワード（フリーターム）とその論理演算で検索する方法の２種類があるが、いずれの場合にも、漏れを生ずることなく適切な数にまで文献を絞り込む過程に問題を含んでいる [1,2]。

たとえば、シソーラスを辿る検索においては、そのシソーラス自体の構築方針や分類に対する考え方を理解していないと、確信を持って、階層を降りていくことができない。さらに、間違えた枝を辿っていたことが分かった時点で、どこまで戻るべきかの判断もつかず、シソーラスの森の中で迷子になるケースも出てくる。学問体系が複雑になり、学際的研究も増えている科学技術文献の情報検索においては、ある程度学問体系を知っている専門家にとっても、シソーラス探索は難しくなっている。

フリータームにより検索を行う場合に見られる典型的な検索手順は、キーワードを思いついた順番に（ＡＮＤの形で）次々に入力していき、個々の文献を斜め読みにできる位の文献数になるまで、その操作を繰り返す手順である。この場合、問題はより深刻で、正しい道筋で検索を行うには、適切なキーワードを早く思いつけるだけの経験や運に恵まれている必要がある。

こうした問題が生ずる原因は、絞り込みの過程で候補にあがってくる文献のセットが、果たして本当に自分の欲しいものを含んでいるのかどうかの確認をとる手段をシステムが提供していないところにある。これに対する解決策の一つとして、ランクつきの結果表示を利用する方法が考えられる。すなわち、検索条件を一つ追加するごとに、該当文献の数を表示するばかりでなく、その中で、今までの条件をもっともよく満たしている上位何件かについて、表題も見ることができるようにする。この表題を眺めながら最適な表題が出てくるまで条件を追加し続ければ、絞り過ぎを防ぐことができる。

ただし、この方法の問題点は、システムは、あくまでも、それまでにユーザが入力した一連のキーワードにもっとも近い文献を探して結果の上位に表示するだけなので、思いついたキーワードに適切なものがないならば（欲しい文献セットを代表する必要最小限のキーワードが含まれていなければ)、いつまでたっても、該当する文献タイトルが表示されない。また、思いついたキーワードが欲しい文献セットの一部のみを示すような語であったときには、望ましいタイトルが表示されるが、すでに絞り過ぎていたことになり、しかも、ユーザはその事実に気づかないことになる。

そこで、望ましい文献セットに早く、しかも的確に到達するためには、検索過程において、候補としてあがってきた文献セットの一部を表示するだけでなく、全体を表示するような工夫が必要となる。

2．情報の輪郭

前節で述べたように、情報絞り込みの過程において、取り出された文献セットの全体像が常に見えていることが、欲しい文献セットに早く辿り着く必須条件である。ここで、任意の文献セットが与えられたとき、その全体像（情報の輪郭）を表現することを Information Outlining と呼ぶことにする。当然のことながら、Information Outline は個々のユーザ（のバックグラウンド、知識の深さなど）により異なるから、それを表す単一の表現手段は存在しない。そこで我々は、文献セットを種々の観点から分類した複数のウインドウ（マルチビュー）を用意し、物事を多角的に捕える人間の認識能力に訴える形で情報の輪郭を捕えてもらうことを考えた。

具体的に文献セットの観点を設定する際に利用できる文献の属性には、

- 文献の書誌情報およびその意味的分類

- キーワードの意味的分類

の2種類がある。

書誌情報からは、著者（著者毎に文献数を数える）、著者の国籍（国別の文献数が分かる）、所属機関（組織別）、発表／掲載年月（年別発表数）、分野（学際的研究テーマに関しては威力を発揮する）などの観点が取り出せる。また、キーワードからは、シソーラス上のカテゴリ（キーワードに対応してカテゴリがわかれば、分類できる）、年代／地域／民族など（社会科学などの場合は対象テーマをこうした観点で分けることができよう）などの観点を用意することができる。

検索の過程で絞り込まれていく文献セットを複数の観点から分類した結果（各観点において各々のカテゴリに属する文献の数を表示したもの）を見ることで、ユーザは情報の輪郭を思い描ける。情報検索において、情報の輪郭を知ることの利点は、

- 検索条件の軌道修正ができる（たとえば、あるキーワードを追加すると、有名な研究機関の発表が0に近づくならば、そのキーワードは文献を絞り込み過ぎたかもしれない）。
- シソーラス上を順番に辿って行かなくても、最適な概念ノードが見つかる（シソーラス上に分布した文献の数を見て、適切な深さのサブ分野が見つかる）。
- あらたな検索条件がわかる。（たとえば、多くの文献がＵＳＡに集中しているならば、他の国の研究は見る必要がないかも知れない。あるいは年別の発表件数を見て、重要な進展のあった年が分かるかもしれない。）

である。以上3点は、いずれも欲しい情報セットへ到達するためのナビゲーションの役割を果たしている。

3．ナビゲーション機能の実現

Information Outlining のためのナビゲーションウインドウには、該当文献数が表示される。この文献数は、検索条件が一つ追加されるたびに再計算される。ただし、検索の途中では該当する文献の数は多いままであるため、ある観点に従って分類し文献数を計算する作業をリアルタイムに行うことは、現実的とはいえない。そこで、文献数を計算する代わりに推定することで、リアルタイムなナビゲーションを実現した[3]。

その推定法の大筋は以下の通りである。

- すべての検索条件ｘについて、そのｘを含む文献数ｆ（ｘ）をあらかじめ計算しておく。これは、文献ＤＢの作成時に行うが、新着文献の登録時にも変更の必要がある。この変更は、ペア（ｘ，ｙ）についても同様である。このｆ（ｘ）を記憶するのに要する大きさは、検索条件集合の大きさＮに比例する。
- 異なる観点同士のすべての検索条件のペア（ｘ，ｙ）について、ｘとｙを同時に含む文献数ｆ（ｘｙ）をあらかじめ計算しておく。これには、Ｎの自乗の領域が必要である。
- ｍ個の観点から選んだ任意の検索条件の組合せｘ１，ｘ２，…，ｘｍについて、それらすべてを同時に含む文書数ｆ（ｘ１ｘ２…ｘｍ）を、ｆ（ｘｉ）およびｆ（ｘｉｘｊ）（ただし、$1 \leqq i \leqq m$，$1 \leqq j \leqq m$，$i \neq j$）を用いて推定する。

検索時には、上記１，２項は計算しないこと、３項の計算は文書の数に依存しないことから、検索時における再分類の計算はリアルタイムに行える。

4.「情報散策」におけるナビゲーションの効果

従来の情報検索システムの目的は、ある明確な目的を持ったユーザに、いかに早く適切な情報を提供するか、という点にあったが、最近の data mining の研究などに見られるような、新たな事実の発見という目的を情報検索システムに持たせることも、重要になりつつあると思われる。

こうしたシステムの捕え方に基づいて、ナビゲーション機能の効果を実験した例を紹介する。以下の例は、日本経済新聞社が提供している日経NEEDSの1年分（1992.12～1993.11）の新聞記事を対象にしている。

4．1 地域に関する観点の利用

キーワード（フリータ－ム）を入力することで限定された新聞記事のセットに対して、地名に関するキーワードを手がかりに県別の記事本数を表示する。

[検索ステップ1]

キーワード「殺人」を入力し、殺人に関する記事1,095件の地域分布を見る。地域別件数の計算は、記事中のキーワードに因っているから、地名と記事の関連は、殺人の行われた場所、犯人の捕まった場所、被害者の出身地、等々が考えられるが、地図上では区別されない。

4．2 期間に関する観点の利用

期間（月または日）別に件数表示する。この観点の表示にあたっても、先に入力された検索条件「殺人」は有効であるから、グラフに示されるのは、月別または日別の殺人記事件数である。

[検索ステップ2]

ここで、地図のウインドウから山梨を選ぶ。これにより、検索条件は「山梨に関連した殺人」（山梨で起きた殺人事件、山梨県人が犯した殺人事件、山梨県人が殺された殺人事件、等々）ということになる。この条件のもとに、地図ウインドウと期間ウインドウの件数が再計算される。このとき、期間ウインドウの報道件数が、8月以前は0で、かつ、8月のみが以上に突起していることがわかる。このことから、8月に山梨に関連した大きな殺人事件があり、その関連記事が集中したという推測がなされる。

4．3 関連キーワード

記事中に引用されたキーワードの回数を数え、頻度順に表示するウインドウである。このウインドウにより、多用されたキーワード「信金ＯＬ誘拐殺人」により、該当記事のタイトルを見なくても、検索された記事の多くが対象としている事件が分かる。

さらに、期間に関する観点のグラフを日別でみると、事件の進展があるたびに記事の本数が増えるところから、記事の拾い読みが可能になる。

5．おわりに

コンピュータシステムの発展の過程で、システムの動作の可視化の重要さが指摘されている。情報検索システムにおいても、この可視化は重要なテーマであると考えられる。ここで提案したナビゲーションもそうした発展の方向で試作した機能であるが、当初の、目的を持った検索の効率化という出発点から構築したプロトタイプが、「情報散策」という新しい使い道のあることを示唆してくれた。今後、ネットワークの高度利用が進み、ｗｗｗのような、検索サービスに特別の専門家をおかないような情報ソースからの検索が頻繁に起こるようになると、情報散策のような検索形態はますます増えていくと思われる。

参考文献

[1] 三輪,「情報検索システムにおけるユーザインタフェースの条件」情報処理学会情報メディア研究会,3-3,1991

[2] 下山他,「サーチャーのノウハウに見る検索インタフェース」情報処理学会ヒューマンインターフェース研究会,41-2,1992

[3] 丸山他,「電子図書館 III － Information Outlining」情報処理学会第 49 回全国大会講演論文集 (分冊 4)