

## 電子図書館 II —基本設計—

2W-7

諸橋 正幸 堤 泰治郎 丸山 宏 野美山 浩 渡辺 日出男  
日本アイ・ビー・エム株式会社 東京基礎研究所

## 1 はじめに

情報のマルチメディア化とネットワークの広がり、情報検索システムのあるべき姿を急速に変化させている。蓄積情報の面からその変化を考えると、マルチメディアとしての情報表現手段の多様化であり、情報量の増大ということになる。また、利用者の側から考えてみると、ある特定の文献データベースのみを検索する「その分野の専門家」から、より一般のユーザへと広がっていくという変化が予想できる。

電子図書館という名のもとに、そうした新しい環境での情報検索技術のあるべき姿が種々の観点から研究されてきており [1]、我々もそうした技術の統合化の実験プロトタイプとして、電子図書館システム MEL S I N E (Multi-media Electronic Library System with Information Navigating Environment) を提案する。システム実現に向けて考慮したのは、

1. メディアの多様化に対応して、検索・表示技術に対応させること、
2. ネットワークの広がりに対応して、情報の分散化を前提とした検索を可能にすること、
3. 蓄積される情報の巨大化に対応して、蓄積情報の輪郭をいかに検索者に伝えるか、

の3点である。

その中で、現在、中心的なテーマとして取り組んでいるのは3の検索結果（途中経過）のフィードバックの問題で、これに関する我々の提案とそれに基づくシステムの試作について、本稿および [2], [3] で述べる。なお、上記 1, 2 については、当面、オブジェクト指向データベースの利用を考えているが、WWW やパソコンネットまでを蓄積情報空間として考える場合には、より緩やかな構造を採用すべきであろう。

## 2 検索経過フィードバックの必要性

従来の情報検索システムにおける情報（文献）検索の方法は、検索者が思いついたキーワードと、いくつかの書誌情報による属性値（著者、掲載雑誌名、など）を組み合わせで検索するのが典型である。このとき、常に問題となるのが、結果の文献数である。のちの利用を考えると、検索された文献数がせいぜい数十であってほしいが、一度の検索でその数に落ちることはまれで、キーワードを追加したり、属性値を変えたりという何度かの試行錯誤の末、その数に絞り込むといった操作を続けなければならない。「この操作をいかに効率良く行うか」は、対象となる文献DBの内容をどれだけよく知っているか（どういうキーワードを与えるとどの程度の文献数が拾えるかが、ある程度予測できる）に依存するために、専門家以外のユーザには使いにくいという評価につながっているものと思われる。

これに対する解決策としては、

1. シソーラスなどの体系化されたキーワードのセットに対して、個々のキーワードに関連する文献数をあらかじめ計算しておき、それを検索時に表示する
2. ユーザが自発的に数を絞り込まなくても良いように、該当文献をキーワードとの関連度の強い順に表示する（優先順位づけ）

ような工夫が考えられる。

ただし、1 においては、探している情報を的確に表現するキーワードのもとに、適度な数の文献があるかないかの判断ができるだけで、もっとも適切なキーワード（または、その組合せ）が何かを見つけることはできない。また、2 においては、順位づけの基準が問題になるケースがある。優先順位付きの検索の提案は、「与えられたキーワードが議論の中心となっている文献をうまく探す」ことが目的のため、例えば、素人向けの解説記事を読みたい場合と、専門家向けのより深い議論を求める場合の区別はつかないといった欠点がある。電子図書館のように、利用者層の広がりが予想されるような環境で

は、この欠点は無視できなくなるであろう。

### 3 情報検索におけるナビゲーション

以上の議論から、検索の道筋で、よりの確なフィードバックを行い、適切な検索条件（キーワードと属性値の組合せ）を見つける手がかりを与えるには、シソーラスだけでなく、いくつかの観点から文献集合を分類したデータを用意し、それらを組み合わせ（AND 条件）場合の文献数を必要の都度計算することが有効である。ただし、この仕組みが実際のシステム上で効果を発揮するためには、高速な文献数の計算（[2]で述べる）と、文献集合にあった観点（ビュー）の提供（新聞記事DBにおけるビューの設計について[3]で述べる）が不可欠である。

なお、我々の提案は、検索結果の優先順位つき表示という考え方と衝突するものではないので、例えば、結果を表示するウインドウに現れるタイトルなどは、優先順位に従って表示すればよいが、前節で指摘したように、順位付けにも分類と同様、観点を導入することが考えられる。例えば、本屋などで利用されるベストセラー情報のアナロジーから考えれば、電子図書館における参照文献頻度の情報は、キーワードとの関連度とはまったく別の観点からの順位づけのデータになることが予想できる。

## 4 システムの構成要素と機能

### 4.1 検索 GUI

あらかじめ多用されるキーワードを色々な観点でまとめて文献集合を分類したビューを用意し、検索途中の、そのビューにおける該当文献数を表示するウインドウが存在する。ビューとしては、十進分類のようなシソーラス系のビューや、著者あるいは記事内容に関連した土地のビュー、発行年月日のビューなどが有効なビューとして考えられる。これを用いると、単に、欲しい情報をとるのに適切な条件を知るのに役立つばかりでなく、検索をしながら新しい発見を体験することもできる。例えば、「ノンフィクション、ヨーロッパの作家で最近10年以内」という条件を与えて、記事内容に関連した土地のビューを見れば、左記の条件で見つかる文献の地域ごとの件数を知ることができるから、ヨーロッパの作家は、最近、どの地方の出来事に興味を持っているかがわかる。

### 4.2 検索エンジン

検索エンジンとしては、従来のシステムによく見られるようないくつかの代表的検索機能を並存させる。具体的には、フリーターム方式により付加されたキーワードをもとに検索を進めるキーワード検索エンジン（同義語辞書などにより似た意味のキーワードも探す）、一時文献情報から検索を行う全文検索エンジン、書誌情報検索エンジン（RDBなどの検索エンジンを利用。上述の発行年や土地などのビューはこれで検索できる）からなる。

さらに、小分類ごとの文献数を計算するための文献数計算プロセスが検索エンジンとは別に存在し、[2]の方法によりリアルタイムに検索結果のフィードバックを行う。

### 4.3 検索用インデックスの自動生成

情報検索におけるキーワードの自動抽出、抄録の自動生成など、大量の文献から、二次情報を自動的に作り出すための先行研究は、すでに、かなり以前から行われており、すでに実用的な観点からでも利用可能な技術は多い。我々のシステムでも、これらの技術は積極的に利用するが、検索ナビゲーションのために必要なデータ（各ビューにおける分類項目ごとの該当文献数のデータ）にも抽出されたキーワードは利用可能である。

例えば、地名を階層構造で分類した辞書を土地のビューとして使うことを考えた場合、各文献から抽出したキーワードと地名の関連から、分類項目と文献の関連づけが可能である。

## 5 おわりに

素人から専門家まで、広く利用できる電子図書館の構築を目指して、試作システムを開発していく中で、難関の一つである「より適切な情報集合を切り出す」ための検索ナビゲーションに重点をおいて、システムの基本設計を述べた。より実用的なシステムに近づけるには、蓄積情報を的確に表すビューの発見とその表示技術が重要になる。

## 文献

- [1] 藤澤、絹川(1993) 情報検索における自然言語処理, 『情報処理』, vol. 34, no. 10.
- [2] 丸山他(1994) 電子図書館 III, 『情報処理学会第49回全国大会講演論文集』.
- [3] 野美山他(1994) 電子図書館 IV, 『情報処理学会第49回全国大会講演論文集』.