

# リファレション型データベースによる 知的検索の可能性

浅井 勇夫 (大阪府立大学工学部)

keywords document database, referation database, referation search,  
intelligent information retrieval

## 1. はじめに

データベース (DB) に関する情報処理技術は、1) 入力技術: データ構造やドキュメンテーション、2) 処理技術: コンピュータのハードやソフト、3) 出力技術: ユーザインタフェースに分けることができる。また、技術の状態は、a) ある程度成熟、b) 現在進行中、c) 近未来のものに分けることができる (第1表)。この20年位の間、DB技術の何が変化し、何が変化しなかったかのであろうか。

一般に、文献DBはテーブル型のデータ構造で構成される。属性の数は比較的少ないが、レコード数は非常に多い。速報性や網羅性を高めるためにDBは絶えず更新され、レコード数は着実に増加し、それにつれてDBの価値は高くなる。レコード数の増加に比例して、キーワード数や分類項目数は増加しない。そのため、大規模なDBほど情報検索は難しいという現象が生じる。

ここでは、テーブル型DBにレコード間の関係情報を追加したリファレション型DBを提案し、新しい情報検索の世界を示す。本論にはいる前に、最近開発されているデータモデルの特徴について考察する。

第1表 データベースに関する情報処理技術

技術	ある程度成熟	現在進行中	近未来
入力	テーブル型DB ファクトDB	知識ベース ハイパーテキスト	マルチメディア
処理	大型コンピュータ 大容量ディスク リレーショナルDB	パソコン CD-ROM	並列処理 ニューロ
出力	プリンタ出力 オンライン化	CRT表示 分散化	マルチウィンドウ

ASAI, Isao

University of Osaka Prefecture

591 堺市百舌鳥梅町4-804 Tel. 0722-52-1161

591 4-804 Mozu, Sakai, Osaka, JAPAN

## 2. 最近のデータモデルの特徴

テーブル型DBから派生したものにリレーショナルDBがある。これはデータ部分から属性の定義部分を独立させ、選択、射影、結合などの関係演算（ファイル処理）を可能にしたDBである。文献DBのオンライン検索システムには、検索（選択）、検索結果のフォーマット出力（射影）はあるが、検索結果の再利用（結合）は考慮されていない。

データや情報を直接DB化し利用するDBに、ファクトDB、経済統計DB、知識ベース、そしてハイパーテキストがある。前2者は、テーブル型のデータ構造をもつ。ファクトDBは属性の定義が難しく、経済統計DBは時間軸を含む3次元テーブルである。検索の基本概念はいずれも文献DBに類似している。

エキスパートシステムにおける知識ベースは、従来あまり使用しない階層型やネットワーク型のデータ構造をあつかう。知識ベースの構築には、知識とその関係の入力が必要である。IF～THEN～ルールで蓄積したデータへのアクセスを推論部と呼ぶが、条件に応じて関係するリンク先を探索する。関係が複雑になると探索が困難になると、見落とした関係先には到達しないなど、知識ベースはその作成に非常に負荷がかかるシステムである。

最近、ハイパーテキストが開発されている。これは情報管理用の簡易言語の一種で、情報それ自体をDB化する。レコードは見出しとその説明の2つの属性から構成する。百科辞典における見出しとその解説は代表的な例である。従来のDBで行うような情報内容を各属性に配置するという面倒な作業はない。その代わりに情報の中に現れる語句のリンク先を定義する必要がある。われわれはリンクをランダムにたどりながら、必要な情報を探索する。

第2表は各種のデータモデルの特徴を簡単にまとめたものである。新しいタイプのDBと従来のテーブル型DBの大きく異なる点は、レコード間の関係を持つか持たないかである。レコードをバラバラに蓄積するのではなく、レコード間の関係情報を含むようなDBの開発が新しい課題である。

第2表 各種のデータモデルの特徴

	テーブル型 DB	知識ベース	ハイパー テキスト	リファレー ション型DB
内容 性質 規模	書誌・抄録 客観的 大規模	知識 主観的 小規模	情報 主観／客観的 小規模	文献・情報 主観／客観的 小規模
タイプ 関係 単位	テーブル型 なし 属性（複数）	階層型 条件付き 1センテンス	ネットワーク型 あり（複数） 1パラグラフ	テーブル型 あり（複数） 属性（複数）
計算機 処理 検索	大型 手続き マッチ	ミニコン 推論 連続リンク	パソコン 手続き 離散リンク	パソコン 手続き 関連／類似性

### 3. リファレーション型データベース

DBの大規模化と大衆化で知的検索への要求は強いが、テーブル型DBに依存している限り簡単に解決できない。その解決策の一つが、前節で見えてきたレコード間の関係情報の追加である。筆者は、文献の参照文献をDB化し、小規模ではあるがパソコン上で処理するソフトを開発し、内外で発表してきた[1-7]。そして、テーブル形式のDBを生かし知的検索を可能にするDBとして、リファレーション型DBの研究を行ってきた。

文献DBの場合におけるリファレーション型DBの簡単な例を第1図に示す。図中のリファレーションファイルは、各レコード間の関係を表わすファイルである。例えば、文献3は文献1、文献3、そして文献6と関係が深いことを示す。

文献ファイル (テーブル型)					リファレーションF	
No	著者	表題	出典	発行	No	Ref.
1	Luhn, H. T.	Keyword-In-Content I	Am. Doc. 11-	1960	1	1, 3
2	Small, H. G.	The Structure of Sci	Sci. Stu. 4-	1974	2	2, 5
3	Vickery, A.	A Reference and Refe	J. Doc. 43-1	1987	3	1, 3, 6
4	Lancaster,	Information Retrieva	John Wiley	1968	4	1, 4
5	Garvey, W. D	Research Studies in	Inf. Stor. 8	1977	5	2, 5
6	Chubin, D. E	Citation Classics An	J. ASIS, 35-	1984	6	3, 6

第1図 リファレーション型データベースの簡単な例

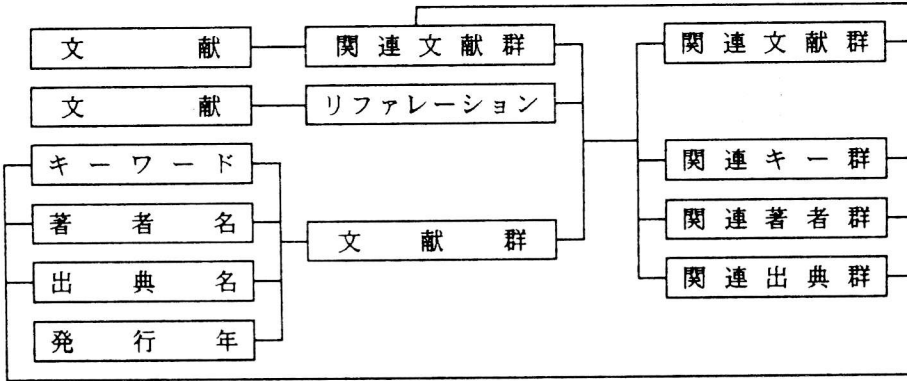
このようなシンプルなデータ構造は目新しいものではない。例えば、見出しのキーワードとそれに関連したキーワードからなるDBを考えてみる。見出しキーワードを属性が1つしかないテーブル型DBとみなせば、それに対応する関連キーワードはリファレーションファイルとみなすことができる。このようなDBは、キーワード間の関係を示すだけでなく、そのリンク先を求めたり、あるいはクラスタリングなどが可能になる。

文献DBの場合、文献と文献を結びつける客観的なデータとして、文献末尾の参照文献がある。現在のソフト技術を使えば、参照文献の入力は可能である。参照文献を使う利点の1つは、文献(citing paper)と参照文献(cited paper)を転置したとき、被引用文献が得られることである。筆者は参照文献と被引用文献の和に文献自身を加えたものを、リファレーションと名付けた[1]。入力は1回であるが、その逆も使うので、データ量は倍増する。

これからのDBにはレコード間の関係情報が必要であると理解しても、関係データを実際に収集するのは難しい。対象とするDBの領域ごとに、独特の方法を考えなければならないかも知れない。ハイパーテキストは、パーソナルなレベルでリンクを自由に挿入・削除できる。時間尺度を含むDBには、前に説明した文献DBの方法が応用可能である。すなわち、あるレコードの関係を求めるには、それ以前のレコードとの関係だけを考慮すればよいという命題が成り立つ。

#### 4. リファレーション検索

文献DBのリファレーション検索の流れ図を第2図に示した。従来の検索は、図の左下にあるキーワード、著者名、出典名、そして発行年などを手がかりに文献群を求めた。新しい検索では、文献間の関係データが入っているため、あるキー項目に関連性の高い集合を取り出すことが可能になる。ある文献に似た文献群、ある文献群に関連の高い文献群、あるいは、文献群を構成する属性の特徴などが検索できる。



第2図 文献データベースのリファレーション検索

リファレーション検索は次のような特徴をもつ。

- (1) 多様な検索：従来は文献群のみ検索したが、リファレーション検索では関連文献群、関連キーワード群、関連著者群、そして関連出典群が検索できる。
- (2) ランク検索：従来は検索される文献群は同じウェイトであったが、リファレーション検索では関連した順番（ランク順）に出力される。
- (3) 関連した検索：従来は検索キーにマッチしたものだけしか検索できなかったが、リファレーション検索では、検索キーを含まないものも検索する場合がある。DBに入っている情報が加工されて出力される。これは、まさに知的検索とみなすことができる。

これを可能にするのは、文献間の関連性である。それはリファレーションを使って定義する。第3図は文献227と文献434のリファレーションとその文献間に共通な文献を示したものである。合計8編の文献がリンクされている。リファレーションの定義に文献自身を含めた理由は直接のリンクを数えるためである。検索要求に応じて、文献と文献、あるいは文献と文献群の関連性を計算し、その高い順に出力する。

従来の情報検索は多数のレコードの中から検索式にマッチしたものだけを検索する。知識ベースの推論は条件に合うリンクを1つ1つたどりながらリンク先の内容を検索する。これに対して、リファレーション検索は検索式に関連したものを検索する。知識ベースやハイパーテキストではリンクを直接使うため、関係データの収集には慎重さが要求される。しかし、リファレーション検索ではリンクを間接に使うため、データ収集は前者より柔軟である。

	番号	年		番号	年	
被引用	959	81	--- 共引用	959	81	被引用
	781	78	--- 共引用	781	78	
	104	74		674	77	
	434	73	=== 直接	434	73	
文献				505	73	文献
	32	72	--- 引-参	32	72	
	405	72				
	31	71	--- 引-参	31	71	
参照	88	71		764	68	参照
	227	68	=== 直接	227	68	
	406	68				
	37	67	--- 書誌結合	37	67	
参照				96	67	参照
	228	67	--- 書誌結合	228	67	
	440	65		122	66	

第3図 文献227と文献434の関連度数のカウント

## 5. おわりに

情報検索の研究は、検索の精度を高めるために情報を如何に表現するかというレコード内部の研究が主であったが、今後はレコード間の関係を如何に表現し活用するかというレコード外部の研究が必要になってくる。ここで示すリファレーション型DBやリファレーション検索は、レコード外部の研究の一つの試みである。

文献DBの場合、参照文献を利用したリファレーション型DBを構築すると、新しい知的検索の世界が出現することを示した。多数のテーブル型DBが莫大な経費をかけて作成されているが、それをマッチングだけにしか使わないのは非常に残念なことである。今後、情報を個別にとらえるだけでなく、情報と情報との結びつきをDB化する方向に進む。それを処理する汎用ソフトの開発が望まれる。

## 参照文献

1. 浅井勇夫、パソコンによる引用文献データベースの開発、JICST論文集, 21, 21-31 (1984)
2. 浅井勇夫、Referationを用いた特定専門分野情報の分析、JICST論文集, 22, 135-142 (1985)
3. 浅井勇夫、引用分析、オペレーションズ・リサーチ, 31(3)18-23(1986)
4. Asai, I., The Structure and Usage of A "Referation" Database On Micro-computers For Personal Document Management, The Application of Micro-computers in Information and Libraries, K.-D. Lehmann and H. Strohl-Goebel (eds.), Elsevier Science Publishers, 510-517(1987)
5. 浅井勇夫、インテリジェント・リサーチのためのリファレーションDBの開発、日経コンピュータ, 1987-2-16号, 137-149(1987)
6. Asai, I., "Referation" Database for Document Information Analysis, Proc. the ASIS Annual Meeting, 24, 1-5(1987)
7. Asai, I., Referation Search: A New Method of Intelligent Information Retrieval, ASIS Mid-Year Meeting, Abstracts, 33(1988)

第19回

# ドキュメンテーション・シンポジウム

## The 19th Documentation Symposium

期日 1989年6月13日(火)・14日(水)

会場 機械振興会館 (東京都港区芝公園3-5-8)

- 6月13日(火) A会場 ホール 地下2階  
B会場 研修2号室 地下3階
- 6月14日(水) B会場 研修2号室 地下3階  
C会場 研修1号室 地下3階

主催 社団法人 情報科学技術協会  
INFOSTA-NIPDOK

後援 (特)日本科学技術情報センター・(社)日本図書館協会  
専門図書館協議会・(社)日本マイクロ写真協会