

ビットマップディスプレイを使用する引用索引検索システム

稲葉清高

(株) リコーソフトウェア研究所

文献の引用関係 (citation) を使用したデータベースをグラフデータモデルを使用して作成した。このデータモデルでは、引用関係をわかりやすく格納し検索にも有効である。

また、文献間の引用関係をビットマップディスプレイを使用して表示するインタフェースを作成し、引用関係をグラフ状に表示できるようにした。このインタフェースを使用すると、一段の引用関係のみならず複数の段階の引用関係をユーザに理解されやすい形で表現できる。

Citation Index Retrieval System for Bitmap Display

Kiyotaka Inaba

Software Research Center

Ricoh Company Ltd.

1-17, Koishikawa-cho 1-Chome

Bunkyo-ku, Tokyo, 112, Japan

This paper discusses an implementation of citation index on the graph data model and a user interface which uses the benefit of bitmap display.

By using the graph data model, user can retrieve the relationship between documents to documents easier than the relational data model. And the user interface for bitmap display enables user to see the relationship of documents as nodes of network with arcs. This interface also displays multi level references of documents easier than character terminal based interface.

1 はじめに

過去の技術情報を検索し必要な情報をスピーディに入手するために、文献検索システムを使用することが一般的になってきた。文献検索システムには様々なものが存在するが、必要な文献を検索する方法から種別するとキーワードを使用したシステムと引用関係にもとづいたシステムにわけられる。

この二つの検索方法は競合するものでなく、むしろ共用することで必要な情報を正確に入手することが可能になる[1]。キーワード検索システムが主にある主題についてのべられた文献の全数調査をめざすのにたいし、引用索引検索システムはそれぞれの情報の重要度および相関をしめすのに有効である。また、引用索引検索システムは文献の検索のみならずある技術分野の発展度およびその興味の移動を検出することにも有効である[2]。

引用索引検索システムは、文献間の引用関係を記録するデータベース部分(引用文献データベース)と使用者が必要な検索条件を入力し、その条件に適合する文献を表示するユーザーインタフェース部分に分かれる。

引用索引検索システムとしては、Institute for Scientific Information (ISI) の Science Citation Index (SCI) が有名であるが、小規模の文献データベースをパーソナルコンピュータを使用して行う試みもみられる[3]。このような小規模のデータベースを作成する場合には、とくに検索の自由さやその作成に必要な労力が少ないことが要求される。この論文では、リコーソフトウェア研究所で開発されたグラフデータモデルに基づいたデータベース管理システム (DBMS) [4] を使用することで引用文献データベースが容易に構築できることをしめす。

また引用索引検索システムにおいて、複数の文献を原点にしてその文献を引用する文献群にたいし和や積、差などの集合演算をとることで、必要な文献群の品質を向上させるころみがみられる[5]。ところが、現在の検索システムは文字端末を入出力装置としているために、参照関係を視覚的に表示することができない。これは、とくに上でのべた集合演算を行なう場合に問題となる。この問題を解決するために、ビットマップウィンドウ上に参照関係を表示することのできるプログラム[11] をユーザーインタフェースに採用した。この結果ある文献を中心に参照関係を表示し、ユーザーが会話的に参照群への操作をおこなえる可能性を示す。

2 汎用 DBMS を使用した文献データベース

ISI や 文献 [3] 等の過去の主な引用文献データベースは、その中に専用のカーネルを用意して検索機構をサポートしている。これは大型システムの場合効率が重要視され、既存の DBMS を使用するより専用のものを構築したほうが効率向上がはかれるためと思われる。また、文献 [3] のようなパーソナルコンピュータ上のシステムの場合適当な DBMS が存在しないこと、オーバーヘッドが大きいこと、機能として必要なものが DBMS の提供するもののごく一部で十分なことなどによると思われる。

引用索引システムに限定しないと、文献 [6] のように既存のデータベースシステムを使用する可能性を議論したものもみられる。この提案のように関係データモデルを使用することはデータベースの構築が容易になること、一貫性がたもちやすいこと、集合論にのっとりた理論的な検索が可能であることなどの特徴がある。また、関係データモデルを使用した DBMS を使用すると上記の利点のほかに DBMS のユーティリティを使用してプログラム作成にかかる労力を減少させられること、検索機構を高速化できる可能性があることなどの点からも有利である。

ところが引用索引システムにおいては、文書間の関係を容易に記述できるデータモデルでないとデータベースの構築が困難になる。たとえば 実体関連モデル [7] のように、実体間を明示的に結合できるモデルが適切である。これに対し関係データモデルでは、文献間の参照関係を検索するのに結合演算を使用する必要があり、複数段の参照関係の検索は非常に複雑なものとなる。

3 グラフデータモデルを使用した引用文献データベース

グラフデータモデル [8] は関係データモデルを拡張して、表中のオカレンス間にリンクとよばれる実体レベルの参照関係を直接表現できる構造を付加したデータモデルであり、実体関連モデルによって記述されたスキーマを容易に実現できることが文献[9]で述べられている。

このデータモデルを使用して、引用索引システムのようなオカレンス間の関係を表現した例としては、文献[10]に地下鉄の駅をつないだネットワーク型の表現に使用したものがある。グラフデータモデルを使用することで、前節でのべた関係データモデルの特徴がそのまま使用できるほかに、実体間の多対多の関係をリンクをたどると

いうプリミティブな操作で記述できる。また現在のイン
ブリーメンテーションではサポートされていないが、ある
リンクを何回かたどる操作 (recursive closure) も可能で
ある。

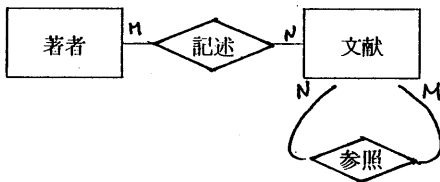
recursive closure の例としては、人間をタプルとし、
「親」「子」というリンクがそれぞれある人の親、子をへ
のリンクとした場合、ある人を中心に二回「親」リンク
をたどって到達するノードの集合は祖父母をあらわし、
逆に「子」リンクを二回たどって到達できるノードの集
合が孫をあらわす。また「親」リンクを一回以上たどっ
て行って到達できる人間の集合はある人の先祖になる。
これらの操作を関係データモデルで記述すると複雑なも
のになることであろう。

上の例の人間を書誌情報に「親」「子」のリンクを「参
照」「被参照」のリンクにおきかえたものはまさしく引
川文献データベースのスキーマとなる。

引用文献データベースにおいては、それぞれの文献の
名前、発行日等を含んだ実体と著者 (複数のこともある)
の実体が必要である。今回は簡単化のため、文献を複数
含む論文集のようなものおよびその編集者は情報として
収録しないことにした。

以上二つの実体にはある文献を書いた著者という関係
と(ある文献に複数著者がいる可能性もあるし、ある著
者が複数の文献を書いた可能性もあるので、多対多の関
係となる) および文献間の参照/被参照の関係 (これも多
対多の関係) が存在する。

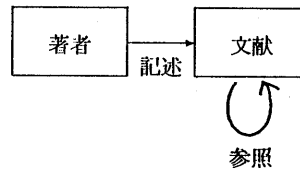
以上を実体関連モデルを使用して図示すると以下のよ
うになる。



図の中で「著者」および「文献」の二つの箱は実体を
あらわし、それぞれある文献の著者と書誌情報を格納し
ている。また、「記述」は著者と文献をつなぐ関連をあら
わし、これによってだれがある文献を書いたかという情
報を記述している。さきほどのべたように著者と文献は
多対多の関係にあることを著者と文献を結ぶ線上の M
と N という文字であらわしている。また「参照」とい
う関連は「文献」どうしをつなぐもので、この関連を使
用してある文献を参照している文献はなにかという情報
を記述している。

この構造をグラフデータモデルで記述すると以下の様

になる。



グラフデータモデルにおいては、「著者」と「文献」と
いう二つの箱がレコード型をあらわし(レコード型は関
係データベースの用語でいう表と同じもの)、「記述」と
「参照」のふたつの矢印がリンク型をあらわす。リンク
型は、始点と終点の二つのレコード型(同一のレコード
型でもよい)をもち、始点のレコード型のなかのあるオ
カレンス(これをレコードという)と終点のレコード型の
なかのオカレンスを結ぶポインタ(これをリンクという)
の集合である。

ある、レコードからリンクをたどる操作はむしろネッ
トワーク型のデータベースの基本操作に似るが、グラフ
データモデルを特徴づけるのはこの操作をレコードの集
合からリンクの集合をたどりレコードの集合にいたるも
のに拡張したところにある。すなわち始点レコード型に
対して選択操作をほどこし、その操作の結果できるレコ
ードの部分集合から出発するリンクの集合すべてに対して
先にのべたたどる操作をほどこし、到達する終点レコ
ードの集合をもとめるという操作を記述することがデー
タ操作言語をもちいて容易に書けることである。実際には
選択操作は始点レコード型に制限されているわけではな
くリンク型でつながれた任意のレコード型に対してほ
どこすことができる(くわしくは文献 [8],[4] を参照してほ
しい)。

データ操作言語の中でこのリンク型によってつながれ
たレコード型を記述するための記法を「パス式」と呼ぶ。
このスキーマ上でパス式表現を用いた検索例をいくつか
紹介しよう。

例 1

まず、「E.F.Codd の書いた文献」は

著者 [名前 = "E.F.Codd"] ../ 記述 ../ 文献

と書ける。この例では「著者」という始点レコード型
の名前というフィールド(関係データモデルにおける属
性と同じ)に関しての選択操作をほどこし、そこから「記

述」リンク型をたどって「文献」レコード型へいたるパスの集合を求めている。

例 2

次に「Garfield の書いた Citation Index を参照している文献」は「文献 1」という「文献」レコード型に対するレコード異称をもちいて

```
著者 [名前 = "E.Garfield"] ../ 記述 ../  
文献 1[名前 = "Citation Index"]  
../ 参照 ../文献
```

と書ける。ここでレコード異称が必要なのは、「参照」リンク型が同じレコード型を始点と終点にもつために、始点のレコードと終点のレコードを区別するためである。もし、レコード異称を使用しないとパス式の意味は「自分自身を参照している文献」という意味になる。

また、パス式が作り出すパスの集合(パス集合)どうしの集合演算をほどこすことで、ある文献が参照している文献のうちで別の文献が参照していないものというような操作も記述できる。

このようにグラフデータモデルおよびパス式を使用した引用文献データベースでは参照関係を含む検索を直感的に理解しやすい形で表現できる。

4 グラフ表示による引用関係の表示

現在提供されている引用索引検索システムは、入出力装置として文字端末装置を仮定しているものが多く、引用関係をわかりやすく表示することは不得意である。とくにある文献を参照している文献をさがすといった一段の検索ならばよいが、ある文献を参照している文献を参照しているものという検索や、ある文献を中心に参照/被参照の関係にある文献を表示するといった複雑な検索になると出力結果を解析することが非常に困難になる。また小規模なデータベースを個人もしくは小さな集団のために構築するとすると、大型計算機を使用するよりワークステーションもしくはパーソナルコンピューターを使用するほうが効率がよい。このさいには、ワークステーションやパーソナルコンピューターの画面表示を十分に使用したあらたな表示方法が考えられる。

グラフデータモデルを実現した市販 DBMS「G-BASE」で使用するユーティリティプログラムとして、画面上にネット状のデータを表示するプログラムが試作された[11]。

このプログラムは入力としてノードおよびアークの情報を与えると最適な配置をさがしだし、発見的な手法でノードをレイアウトするものである。原理的には、VLSI 用 CAD に使用されているフロアプランニングと似た方法を使い、いくつかの基本的なルール(アークの多いノードは中心部におかないと後でレイアウトへの制約が大きくなること等)、ad hoc なルールおよびレイアウトを失敗した場合のバックトラックからなる。

このプログラムを使用することで、引用文献データベースのようなネット状になるデータを画面上に視覚的に表示することが可能ではないかと考えた。このシステムはデータベースから与えられたサブグラフを取り出す部分とそれを画面上に配置する部分の二つから成る。データベースからサブグラフを取り出す部分は、ある文献を中心に参照/被参照の関係にある文献を前節で議論した方法でとりだし、各文献と文献間の接続情報を一時ファイルに書きだす。又後者の表示部は先ほど説明したグラフ表示プログラムをほぼそのまま利用している。

出力された検索の一例を次のページに示す。この検索は P.P.Chen の“Entity-relationship Model”の論文を参照している論文のある特定の論文誌に関して検索したもので、図中に“Chen76”と記されているのが“Entity-relationship Model”の論文をあらわし、そこから伸びている線はこの論文を参照しているものを現している。この検索の場合この論文を参照している論文が三つあり、それぞれの論文が参照している論文が 3 編、6 編および 17 編あることが容易にわかる。

5 今後の課題

本節では今回試作したシステムの今後の課題についてのべる。

- 収録されるデータ件数を多くすることでこのデータモデルが実用的な検索にたえることを証明する。また、収録する情報も今回省略したものを含むようにする必要がある。
- データベース自身の問題ではないが、過去の引用文献データベースでは怪視されてきた第二著者以外の情報を以下に利用するかということも情報検索の点からは興味がある。
- ユーザーインタフェース部に関しては、現在のディスプレイプログラムでは、時間的に過去のものを例えば左側に表示するなどの取り扱いをしていないので配置に時刻のような制約をつけることで、表

示したときにどの文献がどの文献を参照しているのか容易にわかりやすくすることが必要である。

- 今回は単に表示プログラムとしてユーザーインタフェースを設計したが画面上でマウスを使用して会話的操作をゆるすことで、さまざまな改良が考えられる。たとえば、今回のシステムでは画面上には第一著者名とその発表年の下二桁をつないだものを表示しているが、この長方型のなかをマウスで指示すると他の情報をその近くに表示することが考えられる。

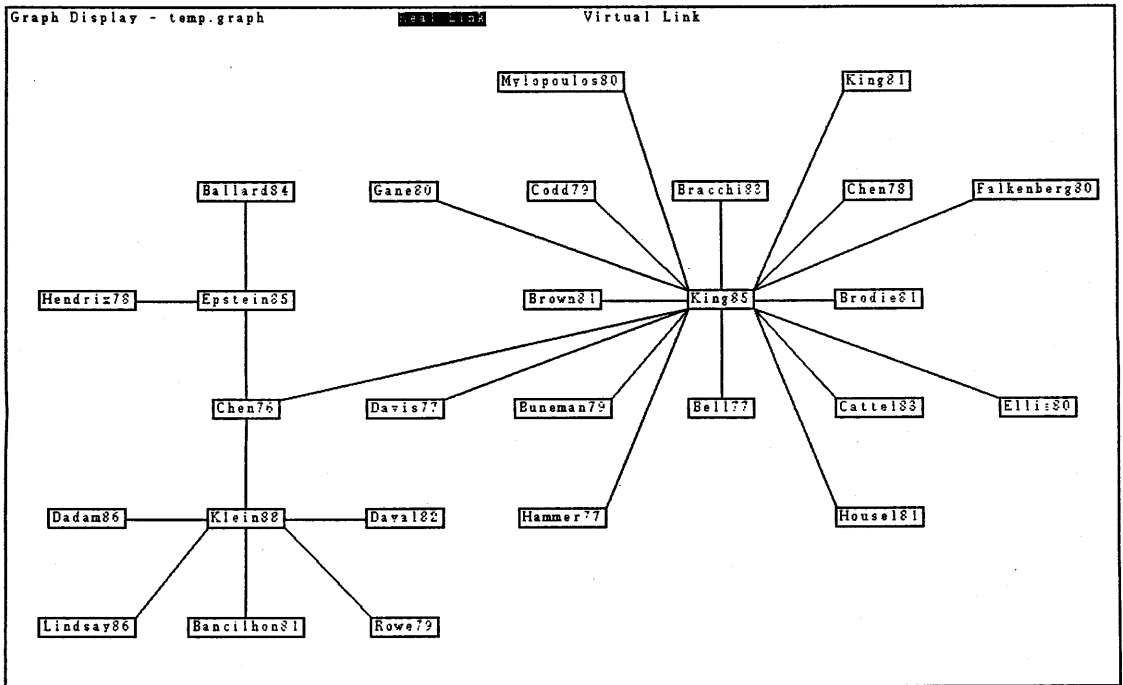
6 ま と め

本稿ではグラフデータモデルを使用した引用文献データベースとそのデータベースを使用するビットマップディスプレイを活用したユーザーインタフェースについて述べた。グラフデータモデルのもつパス式という表現を使用することで、引用文献データベースのようなタプル間のつながりが重要なシステムにおいても関係データモデルのもつ集合論にのっとった検索が可能であることが示せた。また効率のみを重んじた従来のシステムと違い、ワークステーションクラスのハードウェアを活用することで使用者にわかりやすい引用関係をしめすことのできるインタフェースを作成できた。

高解像度のビットマップディスプレイがより安価になり、CDROMのような大容量の記憶装置が一般的になるとここで述べたような大量のデータを整理して表示するためのユーザーインタフェースの重要度もより一層増すものと思われる。

References

- [1] M.L.Pao, T.T.W.Fu. *Titles retrieved from Medline & from Citation Relations*. Proc. ASIS Annu Meet 1985, Vol. 22, pp.122-123, 1985.
- [2] E.Garfield. *Mapping the World of Biomedical Engineering: Alza lecture(1985)*. Annals of Biomedical Engineering, Vol.14, pp.97-108, 1986.
- [3] 浅井勇夫. パソコンによる引用文献データベースの開発. 情報科学技術研究集会発表論文集, Vol.21, pp.21-31, 1985.
- [4] (株) リコー. G-BASE システムマニュアル, 1988.
- [5] E.Garfield. *Citation Indexing - Its Teory and Application in Science, Technology, and Humanities*. John Wiley & Sons, New York, 1979.
- [6] R. G. Crawford. *The Relational Model in Information Retrieval*. J. ASIS, pp.51-64, Jan. 1981.
- [7] P.P.Chen. *The Entity-Relationship Model: Toward a Unified View of Data* ACM Trans. Database Systems, Vol.1, No.1, pp.9-36, 1976.
- [8] H.S.Kunii. *Graph Data Language : A Highe Level Access-Path Oriented Language*. Ph. D. Dissertation of The University of Texas at Austin, May 1983.
- [9] H.S.Kunii. *DBMS With Graph Data Model for Knowledge Handling* Proc. FJCC '87, pp. 138-142, 1987.
- [10] 青木保一, 品川嘉久ほか. メトロネットワークにおける最適経路問題. コンピュータロール, No.17, pp124-130, 1987.
- [11] Y.Watanabe. *Huristic Graph Displayer for G-BASE*. Ricoh Software Research Center Technical Report, SRC-TR-87-05, 1987.



グラフ表示による出力例.