

文献関連情報をもつ研究室用 文献データベースの構築

二村祥一 大澤伊作 岡本則秀

大分大学 工学部 知能情報システム工学科

我々は、研究室単位で手軽に利用できる文献データベースシステムの構築を行っている。このシステムでは、書誌情報とは別に文献分類情報、参考文献情報などの文献の関連情報を追加することにより、通常の文献検索機能に加えて、ある文献と他の文献との関連を調べたり、特定研究分野における文献の位置づけなどを見つけるなどのLAのための諸機能を実現した。また、複数の文献データベースを統合することにより、複数研究領域の中で効率的に文献を検索、評価できるようにした。システムのユーザインターフェースはXウインドウシステムを用いたGUIであり、ユーザの情報検索と検索結果把握がより容易になるようにした。

Construction of a Laboratory Document Database System
with Document Relevant Information

Shouichi FUTAMURA Isaku OSAWA Norihide OKAMOTO

Department of Computer Science and Intelligent Systems
Oita University
700 Dannoharu, Oita 870-11, Japan

We are now developing a document database system which can be used easily at laboratories in universities and companies. In our document databases, we utilized document reference and document classification information as well as the bibliographic information. Adding to the ordinary document retrieval functions, this enabled us to find relationships between documents, and to estimate the location of a document in a specific study field, and so forth. These functions lead the further development of the laboratory automation. Moreover, we implemented the integration facility of several document databases by which we can efficiently and effectively retrieve desired documents and locate them in several study fields. We used X-window system as the graphic user-interface which made us to retrieve documents and understand the results of the retrieval easily.

1. はじめに

文献データベースは大学や企業の研究者の間で研究活動を支援するものとして重要な役割を果たしている。文献データベースは、不特定多数のユーザのための大規模文献データベース[2,5]と、研究室単位で手軽に利用できる文献整理のための小規模文献データベース[3]の2種類が必要とされる。我々は、後者に属する研究室自動化(LA)のための文献データベースシステムの構築を行っている。

このシステムは文献の書誌情報に加えて、文献の分類情報と参考文献情報を合わせ持つことにより、通常の文献検索機能に加えて、文献と他の文献との関連を調べたり、特定研究分野における文献の位置づけなどを見いだすことができる[4,8]ようにした。このシステムは一研究室の文献データベースに止まらず、他研究室で構築した文献データベースをも同時利用できるようにして、複数研究領域の中で文献を探しそれを位置づける手助けとなるようにも配慮した。システムのユーザインターフェースはUNIXワークステーション上でX-Win d o wを用いて作成し、文献情報をハイパーテキスト[7]として表現することにより、ユーザの情報検索と結果情報把握とがより容易になるようにした[1]。

以下、まず第2章で文献データベースの構成とシステム構成について述べる。第3章で文献データベースシステムの機能について、第4章でユーザインターフェースとシステム支援機能について、第5章で複数の文献データベースの統合について述べる。第6章では、作成したシステムについての考察を行う。

2. 文献データベースの構成

文献データは図1に示すように文献内容表、文献分類表、文献関連表の三つの関係表として構成した。文献内容表は文献の書誌情報を含む。項目のうち、文献コードは文献を一意に識別するため用意した項目で、文献の発行年・著者の姓・文

<文献内容表>

文献コード：1987conklin17
著者名：E.Jeffry Conklin
標題：Hypertext: An Introduction ...
雑誌名：IEEE Computer
発行年：1987
ページ数：25
参照ページ：17-41
キーワード：hypertext, database, ...
参考文献数：24

<文献分類表>

1987conklin17 | MM | 0.5
1987conklin17 | HT | 0.5

<文献関連表>

1987conklin17 | 1945bush101
1987conklin17 | 1967nelson
...

図1. 文献データ

献の開始ページ（本の場合は発行年と著者の姓のみとした）から構成した。ここでは、同一人物が同年に同じ開始ページで文献を出すことは少ないと仮定している。キーワード項目は文献主題を表わす語句10～20程度を、文献の付加キーワードあるいは抄録や序文から手動抽出したキーワードにより設定した。文献分類表は文献を研究分野（ここでは文献分類と呼ぶ）で重み付けをして分類したものである。我々の研究室の文献分類と該当文献数を表1に示す。文献関連表は各文献とその参考文献とをそれぞれ文献コードにより示したものである。これら三つの表に加えて、文献内容表の標題およびキーワード項目から抽出したキーワードによりキーワード表を、別途作成している。

これらの表は文献コードをキーとしてデータベースシステムSybase[6]により関係データベースとして構築している。文献データベースシステムの構成を図2に示す。検索システムは通常の情報検索機能であるパターンマッチ検索と、文献間の関連を重視した構造検索に大別した。Sy

表 1. 研究室の文献分類

文献分類	文献数
DB(database)	69
IR(information retrieval)	50
OBJ(object oriented)	31
MM(multi media)	19
TX(text processing)	19
DM(data model)	16
HT(hyper text)	13
...	
DB OBJ	14
DB MM	6
IR TX	6
HT IR TX	5
...	

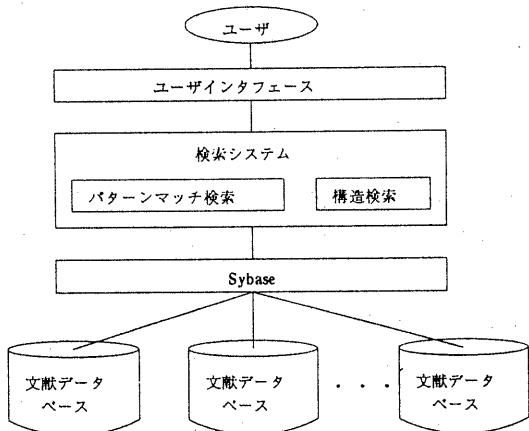


図 2. 文献データベースシステムの構成

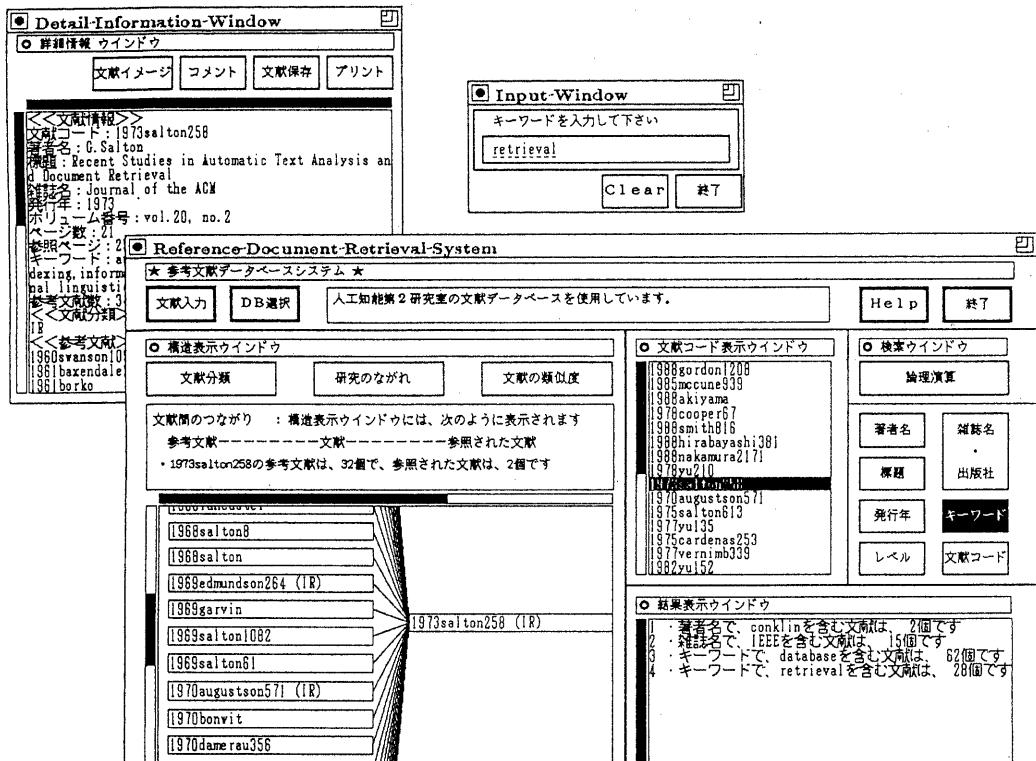


図 3. 文献データベースシステムのユーザインターフェース

baseシステムの利用により複数の文献データベースの統合が可能である。

3 文献データベースシステムの機能

3. 1 パターンマッチ検索

通常の検索は文字列のパターンマッチによる検索である。これは文献内容表の書誌情報について、指定項目値に検索文字列と一致する文字列があるかどうかを調べる。

図3を用いてパターンマッチ検索の例を示す。この例ではキーワード項目に "retrieval" を含む文献を検索している。検索結果が文献コード表示ウィンドウに文献コードで一覧表示される。検索結果ウィンドウには今回の検索により集合番号4の検索集合が表示されている。検索結果に対しては集合番号により論理演算を行うことができる。文献コード表示ウィンドウで文献コードを選択することにより、文献間のつながりが構造表示ウィンドウに表示される。選択文献について、その参考文献を左に、また選択文献を参照している文献を右に表示する。

3. 2 構造検索

構造検索は、特定の研究分野（文献分類）について関連のあるキーワードや関連深い高参照頻度参考文献を調べたり、特定文献についてその類似文献を調べたりする検索である。構造検索では文献分類表を仲介役として使用する。ここでは、文献分類表を、表2に示すように、文献分類単位だけでなく、文献分類の組（およびその任意の部分組）をも単位にして処理する。

構造検索のための前準備として、文献分類表とキーワード表から、表3に示す文献分類とキーワードとの関連表を作成した（表3は後述の”文献分類”ボタンでの処理の中で表示される）。これにより文献分類に特有のキーワードが分かる。また、文献分類表と文献関連表から、表4に示す文献分類と参考文献との関連表を作成した（表4は後述の”研究のながれ”ボタンで表示される）。これにより文献分類毎の高参照頻度参考文献が分か

表2. 文献分類表

文献コード	文献分類	重み
(組を考慮しないとき)		
1987conklin17	MM	0.5
1987conklin17	HT	0.5
1983salton	IR	1.0
1988pollock232	AI	0.25
1988pollock232	IR	0.25
1988pollock232	MAIL	0.25
1988pollock232	MMI	0.25
...		
(組を考慮したとき)		
1987conklin17	MM HT	1.0
1983salton	IR	1.0
1988pollock232	AI IR MAIL MMI	1.0
...		

る。これら二つの表は適時自動作成し、頻度回数のしきい値を決めて頻度回数が少ないものは削除できるようにした。

表3、表4を使って、似通ったキーワードや参考文献をもつ文献分類間に関連を見つけることができる。キーワードを媒介にして、文献分類間の関連を求めた例を図4に示す（図4は後述の”文献分類”ボタンにより表示される）。図中の実線、波線は文献分類間に関連があることを示す。特に、実線は指定文献分類に対する関連を示している。

特定文献に対する類似文献を求めるためには文献の類似度を用いる。類似度は、文献分類表とキーワード表を用いて調べ、同じ文献分類に属し同じキーワードを数多くもつ文献を類似度の高い文献とみなす。類似文献は後述の”文献の類似度”ボタンにより表示される。類似文献の文献コードに文献分類が付加されて、指定文献とともに構造表示される。

表3. 文献分類とキーワードとの関連

文献分類	キーワード	参照回数
DB	data	14
DB	database	106
DB	dbms	3
DB	deductive	8
DB	design	11
DB	distributed	18
DB	document	4
...		
OBJ	data	5
OBJ	database	28
OBJ	design	5
OBJ	model	8
OBJ	object	30
OBJ	oriented	27
...		
DB OBJ	database	25
DB OBJ	model	5
DB OBJ	object	26
DB OBJ	oriented	24
...		

表4. 文献分類と参考文献との関連

文献分類	文献コード	参照頻度
HT	1945bush101	3
HT	1963engelbart	2
HT	1968engelbart395	2
HT	1976chen9	2
HT	1981nelson	3
HT	1984mccracken293	3
HT	1987conklin17	7
...		
MM	1980chang	2
MM	1982haskin207	2
MM	1984copeland316	3
MM	1984mccracken293	3
MM	1986delisle132	2
MM	1986larson62	2
...		
HT MM	1984mccracken293	3
HT MM	1986delisle132	3
HT MM	1986shneiderman	2
HT MM	1987conklin17	2
HT MM	1987shneiderman	2
...		

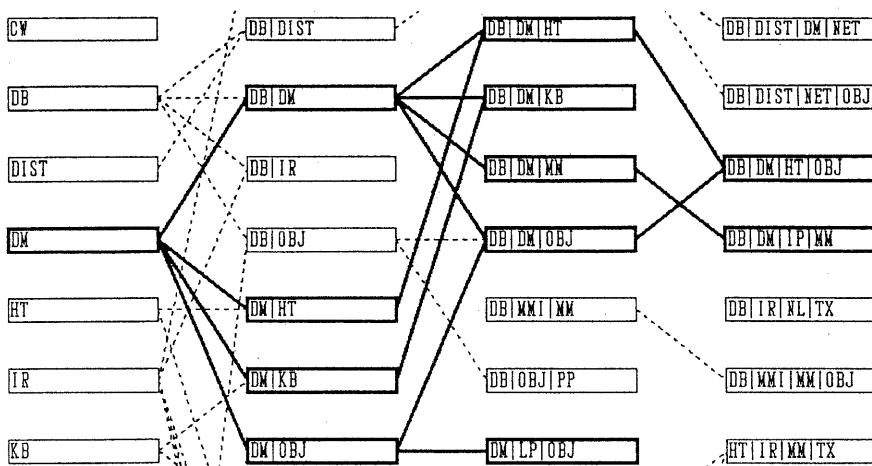


図4. 文献分類間の関連

4. ユーザインターフェースと支援機能

4. 1 ユーザインターフェース

システムの利用画面は既に図3に示した。このインターフェースはUNIX分散ワークステーション上でX-Windowを用いて作成した。

システムのベースとなるウインドウは、検索ウインドウ、結果表示ウインドウ、文献コード表示ウインドウ、構造表示ウインドウからなっており、それぞれ、パターンマッチ検索を行うときの項目選択や論理演算、検索履歴表示、検索結果の文献コード表示、文献の関連表示をするのに使われる。構造表示ウインドウは、パターンマッチ検索により検索された文献を中心とした文献間のつながりを示す(図3を参照のこと)。

このほか構造表示ウインドウには3個のボタンを持たせ、それぞれの機能は以下のようにした。“文献分類”ボタンは登録されている文献分類を表示し、また指定文献分類と関連の深いキーワードを表示する。“研究の流れ”ボタンは文献分類またはキーワードを入力して特定分野の研究の流れを表示する(図5)。“文献の類似度”ボタンは入力された文献と類似度の高い文献を表示する。また、ベースウインドウには“DB選択”ボタンがあり対象データベースを選択するためのメニューを表示する。

詳細情報ウインドウは、文献情報を見るためのものであり、文献コードの示す文献データを表示する。文献データを保存したい場合もしくは保存している文献データを見たい場合は“文献保存”ボタンをクリックすればよい。

4. 2 支援機能

文献データベースシステムを支援する機能として次の4個を用意した。

(1) 文献データ入力(文献入力ボタン)：ウインドウシステムを利用し、ユーザが文献データを会話形式で入力できる。これにより、入力データの定形化がなされ、文献を読後すぐに各ユーザがデータ入力可能である。

(2) 文献のイメージ表示(文献イメージボタン)：文献の書誌情報だけの情報では文献自体を捉えにくい。文献の抄録や参考文献情報、図表などをイメージスキャナで取り込み、検索結果についてイメージ表示する。

(3) 文献に対するコメント付け(コメントボタン)：各文献にユーザがコメントを付けることができるようとした。コメント付けにより同じ文献に対するユーザによる文献評価の違いに対応できるようにした。これにより管理者が文献の文献分類情報やその重みを変更することが可能になる。

(4) 履歴情報をとる：参照された文献の履歴をとっていき、関心の高い文献を調べる。履歴情報をを利用して管理者は文献にコメントを加えることができる。

5. 文献データベースの統合

研究者が自分の研究に他の研究分野の内容を取り入れようとするとき、該当研究分野の文献情報を見つけるのは困難である。そういう時に、該当研究分野の研究室の文献情報を見たり、自分のもつ文献情報と他の研究室の文献情報を合わせて見ることができれば、文献を探しそれを評価する手間が省け、必要な文献を効率的に見つけることができる。このことは、特に境界領域研究において有益であろう。

文献に対して付与される文献分類やその重みは、データ作成者それぞれの価値観によるものなので、研究室毎、あるいは個人毎に違ったものとなる。複数の研究室の文献データベースを統合して検索することで、同じ文献に対して研究室毎の異なった評価や位置づけを見ることができる。そのため本システムでは、同じ文献でもデータベースが違えば別の文献として扱っている。また、統合により総文献数が増加することから、パターンマッチ検索や文献間のつながりを調べるときに幅広い検索ができるほか、新しい文献分類を含めて文献分類間の関連を見いだすことができる。

各文献データベースはSybaseで管理している。それぞれのデータベース構築は単一のシステムを介して行うようにして、データベース名を

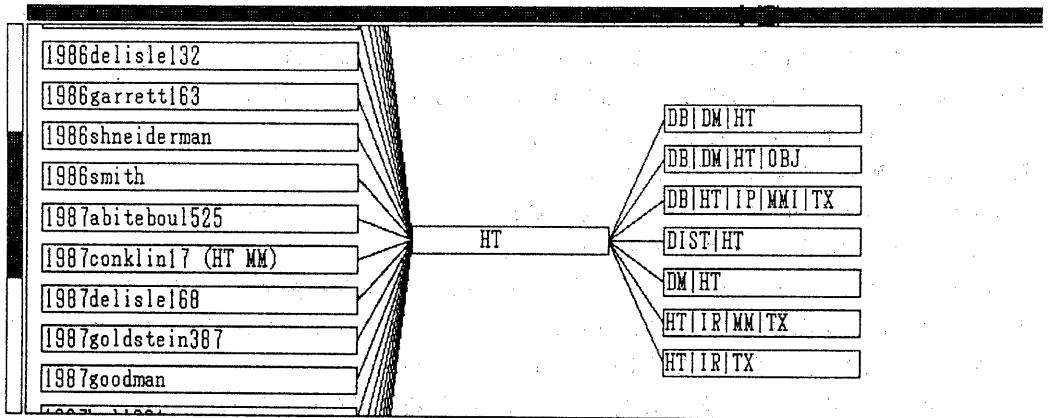


図5. 文献分類HTにおける研究の流れ

除いた、表名や属性名は全て同一に設定している。べることで研究の流れを見いだすことができ、その文献データのユーザ入力はデータ入力用ユーザインターフェースを用いて行う。このインターフェースによって入力データは同一のフォーマットになる。入力データは、データベース管理者に送られ、管理者によるチェックの後にデータベースに追加される。統合して検索する場合は、パターンマッチ検索においては同じ検索コマンドを複数のデータベースに送りそれぞれの結果をまとめたものを検索結果とする。この検索結果には、どのデータベースから検索されたものか分かるように各文献に對して研究室名を付けている。また、構造検索においては、文献分類とキーワードとの関連表および文献分類と参考文献との関連表はデータベースが選択されたときに自動作成するようにした。

6. 考察

同一研究分野の研究者同士は、研究過程で利用する文献は似通っており、文献に対する評価や関心は近いものであるといえる。そこで本システムでは、それらの文献データを手軽に蓄積できるようにし、通常の文献検索機能に加えて、文献のもつ意味と文献間のつながりを示し、研究室単位のデータを統合することを可能にした。

一研究室では、少ない文献データでも参考文献の多くは共通しており、高参照頻度参考文献を調

べることで研究の流れを見いだすことができ、その分野を研究する人の必読文献がわかった。また高参照頻度参考文献の時期と文献分類を調べることで各分野の研究の原点がわかり、文献の発行年を調べることでどの時期にどのように研究が盛んであったかわかった。

文献分類とキーワードの関連、文献分類と参考文献の関連において、文献分類の中には同じ文献分類に属する文献が少くとも、共通なキーワードや参考文献が数多く見つかる文献分類や、逆に同じ文献分類に属する文献が多いにも関わらず共通なキーワードや参考文献が少ない文献分類があることがわかった。

文献の位置づけの評価方法として、多くの文献から長期間参照されている文献を知ることは重要である。“文献間のつながり”表示では、どの様な文献を参照し、どの様な文献から参照してきたかがわかる。“文献間のつながり”表示で、同じ文献分類に属する文献を数多く参照し、同じ文献分類に属する文献から数多く参照されている文献は、その研究分野における一連の流れの中にある文献といえる。逆に、自分と異なる文献分類に属する文献を数多く参照している文献は、注目すべき文献で新たな研究分野の始まりとも考えられ、異なる文献分類に属する文献から数多く参照されている文献は、該当研究分野のことを、他の研究分野

の人が参照できるようにまとめた展望論文であるといえる。さらに、各研究室での文献分類毎の文献数を調べることでその研究室の研究の特徴がわかり、また一研究室で多くの文献分類に属する文献は、その研究室での関心の高い文献といえる。

文献データベースの統合において、文献分類とキーワードとの関連表に、複数のデータベースで同じ文献分類に属するキーワードがあれば、そのキーワードは該当文献分類に関連の深いものと考えられる。同様に、文献分類と参考文献との関連表に、複数のデータベースで同じ文献分類に属する参考文献があれば、その参考文献は該当文献分類において評価や関心の高い文献と考えられる。またこのときに、統合により参照回数がしきい値を超えた文献は幅広く利用される文献であるといえる。

7. おわりに

書誌情報に文献分類情報、参考文献情報などの文献の関連情報を付加することにより、通常の文献検索機能に加えて、研究室運営に必要な文献整理機能などを含む構造検索を実現した文献データベースシステムを構築した。

このシステムでは特定研究分野における研究の流れや、研究分野間の類似性を捉えることができた。また、複数の文献データベースを統合することにより、必要な文献を効率的に検索、評価できるようになった。システムのユーザインタフェースはウインドウシステムを用いたG U Iであり、その元でユーザが、容易に文献データを入力でき、また必要に応じて文献にコメント付けできるようにした。また、利用履歴を取ることにより、文献データベースの改善を図れるようにした。

現在は、文献分類の重みは表示するだけであり、積極的には使用してはいない。重みを考慮した検索を考えていきたい。また、ユーザの使用目的の違いで、検索結果を変えるなど、よりユーザ親和性のあるシステムにしていきたい。

最後に、文献データベース統合に際して、文献データ提供に御協力いただいた大分大学工学部の藤田米春教授に感謝致します。

参考文献

- [1] 認知科学研究会訳：知的インターフェース
－人とマシンとの知的相互作用－，海文堂，1991.
- [2] 根岸正光：学術情報センターのデータベース、情報処理，Vol.33, No.10, pp.1144-1153, 1992.
- [3] 松田孝子：研究活動におけるデータベース技術の応用と展望：情報処理, Vol.23, No.10, pp.1015-1018, 1982.
- [4] 三輪眞木子：データベースサーチャの視点：情報処理，Vol.33, No.10, pp.1162-1170, 1992.
- [5] 全国共同利用大型計算機センター，学術センター：オンラインデータベース利用ガイド 第13版，1993.
- [6] 計算機マニュアル：S y b a s e 日本語マニュアル，東芝エンジニアリング，1990.
- [7] Conklin,E.J.: Hypertext, IEEE Computer, Vol.20, No.9, pp.17-41, 1987.
- [8] Salton,G. & McGill,M.J.: Introduction to Modern Information Retrieval, McGraw Hill, 1983.