

寄稿

インテリジェント・リサーチのための「リファレション」DBの開発

- 文献末尾にある参考文献をデータベース化すると非常に付加価値の高いデータベースが構築できる。
- 文献間の関係を示す文献属性として新たにリファレションを提案し、それをベースにした情報検索システムをパソコン上で実用化した。
- これを利用すると、特定専門分野の情報構造が明らかになるため、研究者は評価情報や関連性情報をもとに知的で効率的な研究を推進できる。

浅井 勇夫 / 大阪府立大学工学部経営工学科

企業や国家間で知的所有権に対する関心が高まってきた。有体物でない知的なもの、例えば著作物(プログラムや半導体)、営業秘密、ノウハウなどの入手は困難になることが予想される。物的資源の乏しい日本にとって、知的所有権の源泉である質の高い研究開発とデータベースの拡充は緊急の課題である。

新技術や新材料が非常に早いテンポで開発されるため、研究開発の現状を正確に把握することは、永年研究に携わってきた研究者にも難しく、まして新規の参入者には不可能に近い。オンラインによる遡及検索では、読みきれないほどの大量の文献が検索される。文献を収集したあとの処理は研究者の頭脳にゆだねられており、機械化されていないのが現状である。

ここでは、収集した文献のコレクションをデータベース化し、研究に必要な情報を定量的に作成することを試みる。AI的なアプローチではないが、中級の研究者が習得するぐらいの研究に関する専門家情報を得ることが目的である。知的な研究を進める上で大変有用なツールである。

参考文献の新しい利用法

情報現象を計量的にあつかう学問分野に計量情報学(Bibliometrics, Informetrics, Scientometrics)がある¹⁾。引用分析を使った研究動向、分類、評価、系譜などの研究や、ブラッドフォード分布(情報源の特性)、ロトカ分布(科学者の生産性)、情報の寿命分布などの研究である。ここでは、参考文献利用

開発したソフトの特徴は、(1) 文献間の関係をあらわす参考文献を入力、(2) リファレション(Referations)をベースに項目の関連性を分析、(3) メニュー選択方式で多種類の情報を表示——などである。以下、参考文献利用の歴史を見た後、リファレションを提案し、ソフトを紹介する。

の歴史を簡単に見てみる。

文献データベースの現状

現在の文献データベース(文献DB)は抄録誌が母体であり、属性の構成は抄録誌と同じである。コンピュータの導入で処理能力は増大しているが、文献DB(データベース)の構成は変わっ

表1 ●文献データベースの構成

カテゴリ	フィールド
基本項目	番号, 著者名, 標題, 雑誌名, 巻頁, 発行年
補足項目	所属機関名, 使用言語, 発行国, 図表・参照数
付加項目	抄録, キーワード, 分類番号

ていない。一般にDBはデータの蓄積が付加価値をもつため、何年も同じ形をとり、変更が難しい。

文献DBのフィールドは3つの部分から構成される(表1)。どの文献DBも基本項目をもち、大規模な文献DBは補足項目をもつ。付加項目はコストやタイム・ラグの原因であるが、最も付加価値が高く、その質によってDBの価値は決まる。

1文献は10数項目、約1000バイトで構成されるが、大規模な文献DBは数10万~数100万のレコードをもち、ソフトは検索主体である。多数の項目をあつかう管理主体のDBMSと異なり、データの蓄積が価値を生む。

文献DBの分野にも新しい動きが見られる。シソーラスの改良、全文DBの提供、文献内容のファクトDB化、電子出版、CD-ROM化など、新しい情報技術が芽生えている。欧米からは、日本語で書かれた文献をすべて網羅したDBが要求されるだろう。また、現在の文献DBは新しい属性の付加や検索した文献群の統計出力などで付加価値をさらに高めることができる。

図1 ●SCIの参照文献の入力書式

BRADFORD SC(DOCUMENTATION,1949) SMART SC(AM ED RES J,V18,P399,1981) SUMMERS EG(READING RES Q,V19,P102,1983) PRICE DO(KEY PAPERS INFORMATI,P195,1980)

重要な文献属性である参照文献

現在の文献DBに最も追加して欲しい属性は文献末尾の参照文献である。参照文献がDB化されていないのに驚く人がいるかも知れない。参照文献は簡単に入手できるので価値がない、ということではない。その有用性が見いだされたのは最近のことである。情報関係者を含め、それを知っている人は少ない。

参照文献は論文を書く際に参照した論文のリストで、論文末尾に付けることは1つの慣習になっている。参照文献の要点だけを書き、詳細は該当する論文を読みなさいというわけで、歴史的な背景や自説の論理的展開の位置づけなど、その役割は大きい。最近、参考文献や引用文献の表現が参照文献に統一された²⁾。

参照文献は文献と文献を結びつける有用なデータであり、次々にリンクしていけば、文献間の関係はねずみ算式に増えていく。あまり関係がないようなデータが、ある程度蓄積されてくると大きな力を発揮するのがDBの特徴である。現在のコンピュータ技術を使えば、文献DBの属性に参照文献を付加することは可能である。

Citation Indexesの登場

参照文献をDB化する構想は1950年代に提案された。そして61年にISI

社のGarfieldが遺伝学の基金で613種類の雑誌にリストされた約140万の参照文献を処理するシステムを開発し、63年にScience Citation Index (SCI)を冊子体で発行した³⁾。

参照文献の入力は簡単なようで大変難しい。参照文献の書き方が雑誌ごとに異なるため、そのまま入力しただけでは、あとで名寄せしたときに使えない。現在のパソコン程度のコンピュータを使ってGarfieldが考えた入力書式は、第1著者、雑誌の省略形、巻・頁・発行年である(図1)。

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、大変価値の高い関係が得られることである。図2は、2つの参照文献BとCをもつ文献Aを転置した結果を示す。文献BやCを誰が参照しているかすぐ判明する。これを大規模に行ったのが、Garfieldである。著者順に並べたのをCitation Indexesという。

現在、世界の主要な約4100の科学雑誌に使われた参照文献が提供されている。抄録やキーワードの代わりに参照文献を付加した世界で唯一のDBである。著者の書いた論文が他の著者にどれだけ参照されたかが分かるので、米国では科学者の評価基準の1つに使われている。また、文献がどのように利用され発展していくかを示すので、研究を進める際の貴重な情報となる。

図2 ●参照文献の転置

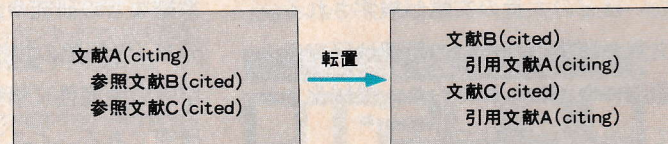


図3 ●78年発行の文献601の参照文献と引用文献

番号	年	名前
1205	86	引用文献 Citations
1129	84	
1115	83	
824	80	
601	78	文献
608	78	参照文献 References
626	77	
579	76	
670	75	
278	73	
56	72	

このDBはDIALOGでオンライン検索でき、2種類のファイルがある。

- 34: SCISEARCH (他に, 87, 97, 186)
- 7: SOCIAL SCISEARCH
著者Aの論文を誰が参照したかは、
S₁CR=著者A
で検索できる。なお、著者名は図1のような略記法を用いる。

文献間の関連性の測定

1つの文献に直接関係する文献群は2種類あることが明らかになった。

- 参照文献
References, Cited Papers
 - 引用文献
Citations, Citing Papers
- である(図3)。

前者の参照文献は簡単に知ることができ、年月が経過しても増えることはなく、むしろ退化していくように感じられる。前節のCitation Indexesから得られる後者の引用文献は、文献の利用に応じて増えていき、あたかも生命をもっているように思われる。

文献と文献がどれくらい関連しているかを客観的に測定できれば、文献を取りまく現象が分析可能になる。従来使っていた方法は、

- Reference Analysis
参照文献だけを用いる分析
- Citation Analysis
引用文献だけを用いる分析

の2種類である。

63年にMITのKesslerは書誌結合(Bibliographic Coupling)という測度を提案した⁴⁾。これは、2つの文献の参照文献に共通な文献を数え、文献間の関連性の測度とした(図4)。文献のす

べてのペアについて、書誌結合を計数し、それが大きいほど関係が深いと見なした。

73年にISI社のSmallは共引用(Co-Citation)を提案し⁵⁾、参照文献の代わりに引用文献を用いて測度を求めた(図5)。この共引用のカウントにはSCIを用いる。Citation Indexes以外の利用法を見つけるのに、約10年かかったことになる。

引用分析を用いた科学の科学

SCIが現れるまでは、手作業で参照文献から頻度統

計をとり、研究情報の特性や構造などを分析していた。この分野ではPriceの業績が顕著である⁶⁾。引用文献の概念が普及するにつれて、参照文献の分析から引用文献の分析へと移行した。

最近の計量情報学の研究はSCIを用いて引用分析するのが多い。Garfieldの研究が代表的である⁷⁾。SCIの中の膨大な量の引用文献から関連性の高い文献群を見つけ、クラスタリングやマッピングする。その汎用ソフトも開発されている⁸⁾。その結果、分野、著者、雑誌間の関係が、マクロ的ではあるが、定量的・視覚的に提供され、利

図4 ●文献1075と1048の書誌結合

Doc.	1075	84	1048	82	Doc.
Ref.	962	81	527	76	Ref.
	1151	78	654	75	
	527	76	435	75	
	653	74	684	75	
	93	67	653	74	
	705	55	652	70	
	9	48	6	67	
	226	35	705	55	
	537	26	537	26	

図5 ●文献671と631の共引用

Cit.	1104	83	1104	83	Cit.
	998	82	905	83	
	981	81	608	78	
	985	81	607	78	
	980	81	619	77	
	608	78	638	77	
	619	77	639	77	
	639	77	634	76	
	630	77	583	76	
	620	77	636	76	
583	76	631	76	Doc.	
Doc.	671	75	631	76	Doc.

用できるようになった。

SCIを用いた研究は引用分析の研究者でさえも簡単に行えない。なぜなら、SCIは大変価値があり、高価だからである。文献の発行後、少なくとも5年以上経過しないと引用文献は蓄積されない。そのため、引用文献だけを用いる分析は、現在よりも少し前の情報現象の分析になる。

参照文献だけでも価値はあるが、そ

れをDB化すると、図2に示したような参照文献、引用文献の転置が可能となる。これによって付加価値の高い引用文献が作成でき、また文献間の関連性が測定できる。

このような新たな情報分析を可能とするため、以下に、筆者が提案した新たな文献属性、「リファレーション(Refererations)」について述べてみたい。

新たな文献属性リファレーションの提案

日本では、ドクメンテーションや図書館情報学などの情報を管理する技術は、あまり重要視されていない。単一民族で終身雇用性の社会では情報を記録する必要はなく、データや情報は誰かの頭のなかにあり、いつでも無料で聞き出せたからである。しかし、国際化の波と「新人類」の登場で、独自のDBを作成する必要がでてきた。

筆者は引用文献の有用性に着目し研究を行ってきた⁹⁾。しかし、SCIのDBを利用できなかったため、研究分野の計量情報学に関する文献DBを作成し、参照文献の入力方法や引用分析の研究をしてきた。しかし、情報量の少ない

文献群から有用な情報を抽出することは非常に困難であった。そこで、わずかなデータを有効に活用する方法を見いだす必要にせまられた。

リファレーションの定義

特定分野の文献を収集すると、最も新しい文献と最も古い文献の発行年の差は数十年に及ぶ。そのような時に、文献の参照文献だけや引用文献だけを属性とみなすと、文献の発行年により不都合なことが発生する。

新しい文献は、参照される機会がないため引用文献はゼロである。しかしすべての文献を参照する機会をもつ

で参照文献は多い。古い文献の場合は、逆に引用文献は多いが参照文献はゼロである。発行年が途中の場合は、ある程度の引用文献や参照文献をもつ。

発行年の異なる文献同士をマッチ

ングさせるには、参照文献だけとか引用文献だけでなく、その両者の和を使うと都合がよい。これに気が付いたのは、大規模なSCIデータではなく、独自に開発したDBを利用していたためである。

筆者は引用文献と参照文献の和に文献自身を加えたものを、「リファレーション(Refererations)」と名付け発表してきた¹⁰⁻¹²⁾(図6)。文献は参照文献でなくリファレーションを持つと見なすほうが合理的である。現在のSCIデータからはリファレーションの作成は難しい。なお、リファレーションはReferencesの前半とCitationsの後半から造った新造語である。

リファレーションを用いた関連性の測定

書誌結合や共引用と同様に文献間の関連性は、新しいリファレーションを使って定義することができる。図7は2つの文献860と137のリファレーション間に共通な文献を実線で示したものである。

参照文献だけを用いる書誌結合は6編、引用文献だけを用いる共引用は1編であるが、リファレーションを使うと11編になる。書誌結合と共引用の和よりも4編多くなる。直接のリンクや参照文献と引用文献のリンクがあるため、それだけ情報量が増えたことになる。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計数するためであった。

関連性測度として3種類を考えた。それは関連度数、関連係数、関連指数である(図8)。リファレーションの数が多いとリンクする度数は多くなる。

図6 ●リファレーションの定義

1199	85	Citations	Refererations
1193	85		
1110	83		
1097	83		
857	81	Document	=Citations
863	78	References	+Document
547	77		
104	74		
121	65		
120	63		
			+References

図7 ● 文献860と137との関連度数

Cit.	1173	85			
	1078	84			
	1110	83			
	1081	83	1081	83	
	978	81			
Doc.	860	81	860	81	
Ref.	597	78	808	78	
	607	78			Cit.
	605	78			
	592	77	567	77	
	586	77			
	527	76			
	525	76	525	76	
	119	75	119	75	
	90	73			
	137	73	137	73	Doc.
	156	73			
	285	71	136	71	
	70	70	70	70	
	11	69	11	69	
	69	69	69	69	
166	69				
260	69	260	69		
		10	68	Ref.	
93	67	93	67		
319	67	296	67		
35	62	66	66		
81	60	81	60		
705	55	261	49		
8	34	9	48		

それを調整するのが関連係数で、よく使う測度である。また、度数と係数の中間的な測度として関連指数を求めた。図7の場合、文献860と137間の関連度数は11、関連係数は $11 / (29 + 19 - 11) = 0.297$ である。

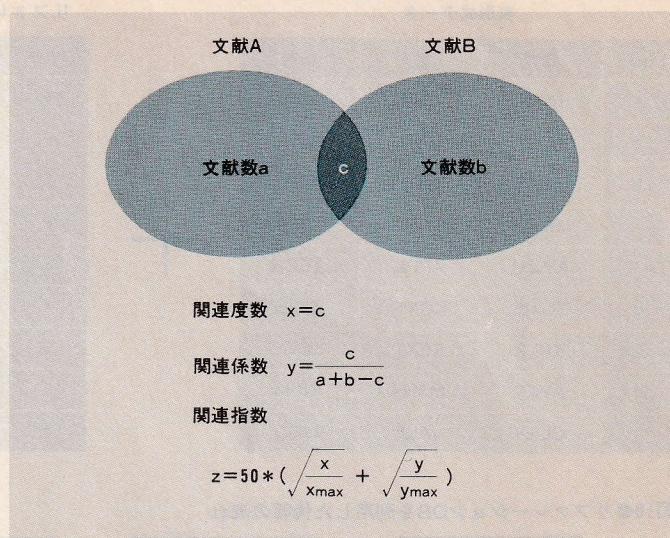
文献Aに關係の深い文献群を求めするには、他のすべての文献との関連性測度を求め、それを大きい順に並べればよい。文献Aと他のすべての文献と

の関連度数を求めるとは膨大な計算量になる。しかし、非常に簡単なアルゴリズムを開発し利用している。また、文献群と文献群との関連性は関係する文献の累積度数を用いた。

リファレションDBの構成

文献DBはフィールドに文献属性、各レコードに文献を対応させた表形式のデータ構造をもつ。リファレションは属性なので文献DBの中に入れるべきものであるが、ここでは分離して文献DBに付随するファイルとみなし、それをリファレション・ファイルと呼ぶ。
リファレション・ファイルの各レコードには、レコード間の関係を示すデータ、例えばレコード番号が可変長

図8 ● 関連度数， 関連係数， 関連指数の定義



で入る。このようなリファレション・ファイルをもつ表形式のDBをリファレションDBと名づける(図9)。

DBのデータ・モデルは、複雑な階層モデルや網モデルから簡単な関係モデルへと発展した。エキスパート・システムの知識ベースの構築にルール型やフレーム型が使われているが、もう少しシンプルなアプローチがあってもよい。そうすれば、現在までに蓄積したDBやソフト資源を有効に活用できる。

ここで提案したリファレションDBは、表形式データにレコード間の関係を表すリファレション・ファイルを追加しただけであり、汎用性が高く簡単な構造をしている。現存する情報を単に検索するのではなく、関連性測度を使って、現存しない高度な情報をつくりだす。知識ベースにおけるリファレション型アプローチとみなせる。

インテリジェント・リサーチのすすめ

研究の現状や方向の把握、あるいは重複研究の回避のために、研究者は文

図9 ● リファレションDBの定義

表形式データ				リファレション・ファイル	
No.	著者	標題	出典	No.	リファレション
1	ROBI	ANAL	SIMU	1	1, 5
2	SLAM	ON T	ACLM	2	2, 3, 4, 7
3	BASS	INDE	PROC	3	2, 3, 5, 8
4	BLAI	ORGA	JASI	4	2, 4, 6, 7
5	MAZU	APPL	TECH	5	1, 3, 5, 7, 9
6	TOME	COMP	SCIE	6	4, 6, 8
7	KATZ	TEXT	JDOC	7	2, 4, 5, 7, 9
8	AVER	MANA	INFM	8	3, 6, 8
9	CLEV	JALU	LIBQ	9	5, 7, 9

図11 ● 書誌情報の作成ルート

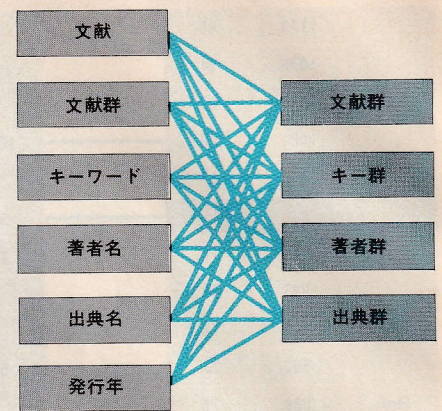
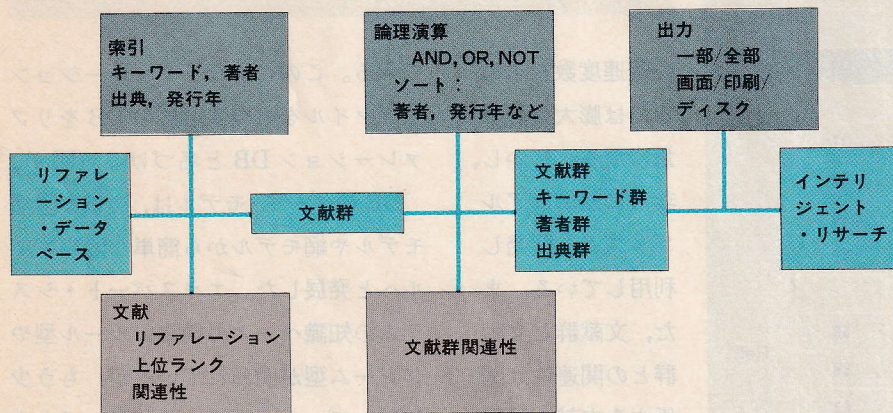


図10 ● リファレションDBを利用した情報の流れ



使用すると、さらに図10の下半分の機能が加わる。すなわち、任意の文献のリファレションの検索、任意の文献に関連の深い文献群の検索、検索した文献群に関連の深い文献群、キーワード群、著者群、出典群の検索が可能になる。

図11に示すように、書誌情報は全方向から得られる。そして、従来の検索方式と異なり関連性の高い順番に出力される。これは一種の専門家情報とみなすことができる。このようなDBシステムを利用すれば、特定分野の書誌的な構造が理解でき、研究効率は高くなる。

次に、筆者が実現したリファレションDBを具体的に述べる。

文献の収集活動を行っている。しかし、研究に関連する文献はオンライン検索の普及などで増大する一方である。もしコンピュータに支援された研究活動が可能になれば、専門家・非専門家を問わず研究開発は促進する。

今まで見てきたように、参照文献をDB化すると、

- リファレションの作成
- リファレションによる関連性の測定

が可能になる。特に、引用文献数を含む文献リストは古い文献の評価基準になる。

これらを使用した新しい文献情報シ

ステムは図10のようになる。

従来の文献システムは、キーワード、著者、出典、発行年などのキーで検索した文献群を論理演算したり並べ替えて出力していた。リファレションを

リファレションDBの開発

高性能なパソコンが出現したため、今までの研究成果をソフトの形にまとめることができた。パソコンのカラー・グラフィックス、画面制御などの機能を使ったユーザー・フレンドリなソフトである。約1000編の文献を処理できるが、主記憶が大きくなれば処理量は増大する。

まずシステムの操作概要を述べ、文献ファイルの作成、参照文献の簡単な入力方法の開発、各種の索引ファイルの作成、そしてサブDBの作成など、データの作成に関することを述べる。

初期画面と操作概略

操作性を高めるために階層的なメニ

ユー方式を採用した。プログラムを立ち上げると、初期画面があらわれる(図12)。そこには、データの内容や更新日付が示される。使用するデータは計量情報学関係の1230文献である。表2は各ジョブのFキーの一部を示す。

画面は、次のように構成した。

- 1行目 (水色)
ジョブ名とFキー名の表示
- 2行目 (黄色)
入力用コメントの表示
- 3行目以降：情報表示。
ジョブは1~2行目に表示されるメニューとの対話で進行する。1行目のFキーは常に割り込み状態になっており、選択すると紫色に変わる。2行目のコマンド選択はテン・キーで行う。入力時のリターン・キー操作は省略でき、必要な時は[CR]で指示する。ゼロ・キーを押すと1つ前のジョブに戻る。また、操作途中でヘルプ・キーを押すと、マニュアルを見ることができる。

文献ファイルの作成

研究分野に関係した文献の存在が明

らかになれば、入手・未入手にかかわらず文献ファイルに登録し利用する。

初期画面で[6文献]を選択すると文献ファイルの作成画面になる(図13)。画面は上下2つに分かれ、上部は文献ファイルの作成画面で、カード・イメージで文献を入力する。

なお、文献ファイルは、1レコード256バイトの基本項目だけで構成した(カッコ内はバイト数)。

- 種別 (1)
- 文献番号 (4)
- 著者1 (20)
- 著者2 (20)
- 著者3 (20)
- 標題 (128)
- 出典 (35)
- 発行年 (4)
- その他 (24)

下部は5種類の索引(文献番号、著者、標題、出典、発行年)の表示画面である。表示する索引は作成画面のカーソル位置で自動的に決まる。索引の探索とスクロールは非常に速

い。特に、省略形の出典を入力する際に有用である。

- 作成時のメニューを示すと、
- [英数字] データの入力
 - [←→] カーソルを左右に移動
 - [CR] カーソルを項目ごとに移動
 - [INS] 空白文字を挿入
 - [DEL] カーソル位置の文字を削除
 - [ESC 索引 CR] 索引の探索
 - [↑↓] 索引を頁単位にスクロールなどの機能がある。

参照ファイルの作成

参照文献を入力する方法が開発できれば、リファレション概念を適用し

表2 ● ファンクション・キーの一部

初期画面	
1. リスト	28種類の情報を出力
2. 検索	13種類の文献検索
3. 関連性	検索文献群の書誌情報
6. 文献	文献ファイルの作成・編集
7. 参照	参照ファイルの作成・編集
8. 索引	約50種の索引ファイル作成
9. サブ	特定文献群に限定
リスト	
1. 引用	文献のリファレションの表示
2. KWIC	キーワードのKWIC表示
3. 索引	番号、著者、出典、発行年、種別、標題索引の表示
4. ランク	引用、キーワード、著者、出典ランクの表示
5. 名前	キーワード、著者、出典
6. 度数	引用、発行年、種別
7. 分布	ブラッドフォード分布、ジップ分布、ロトカ分布、特性分布、成長・寿命分布
8. 書式	1行、圧縮、定型
検索	
1. 引用	関連性順、引用、引用ランク、ランク順をもとに検索
2. 索引	番号、キー、著者、出典、発行年、種別をもとに検索
3. ランク	キー、著者、出典ランクをもとに検索
4. 操作	検索式、ソート、削除
5. 表示	検索文献群の表示

図12 ● 初期画面

キー	総数	延数	作成年月
文 献	1230		文献ファイル 86/09/1
著 者	932	1689	引用ファイル 86/09/2
出 典	284	1230	
年 別	47	1230	文 献 索 引 86/09/1
種 別	5	1230	キ ー 索 引 86/09/2
標 題	596	1230	引 用 索 引 86/09/2
キーワード	1648	7183	分布の推定 86/09/2
REFERENCES	1230	5510	本日の日付 87/01/1
CITATIONS	1230	5510	
REFERATIONS	1230	12250	[HELP] マニュアル

図 13 ●文献ファイルの作成画面

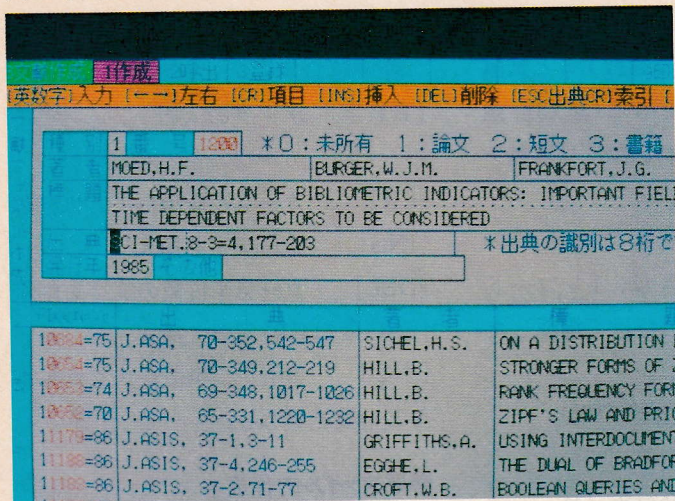
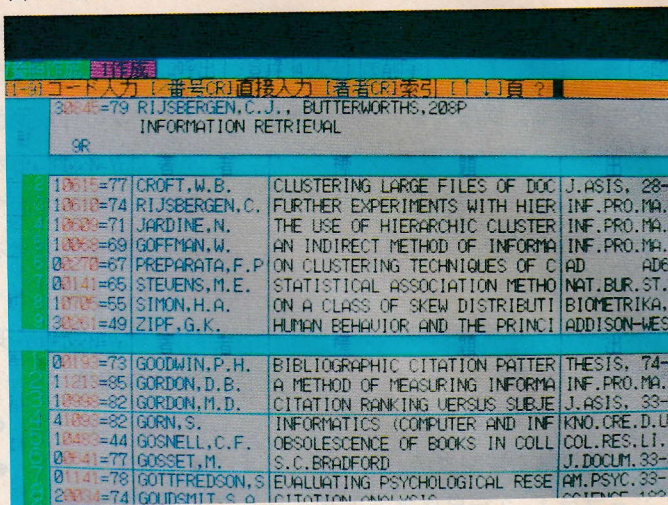


図 14 ●参照ファイルの作成画面



た分析が可能になる。先駆者の Garfield は遺伝学分野を対象とした参照文献の入力を試みた。しかし、参照文献の全部を入力したため科学技術全般に拡大し、大規模でオープンな DB になった。このため、マクロ分析ではあるが、科学全体の構造や分野間の特性が明らかになる。

ここでは、Garfield の方法とは異なり、次のような条件を設けた。すなわち、リストされた参照文献の全部ではなく、特定分野に関係あるものだけを対象とする。従って、参照文献は文献ファイルに登録した文献だけを対象とし、参照ファイルは文献番号で構成する。クローズドな DB であるが小規模で大変あつかいやすい。検索スピードも速く、分野内の特性などマイクロ分析が可能になる。

図 14 は参照ファイルの作成画面の例である。画面は上中下の 3 つに分かれる。上部は文献、中央は参照文献、下部は著者索引の表示画面である。作成時のメニューを示すと、

[1-9] 索引コードの入力

[番号 CR] 文献番号の直接入力

[著者 CR] 著者索引の探索

[↑↓] 索引を頁単位でスクロールなどがある。

参照文献の入力方法は 2 種類あり、1 つは画面下部の索引コードの入力、もう 1 つは前もって調べた文献番号の直接入力である。入力するのはコードや文献番号だけであるが、文献の基本項目全部を入力したのと同じ効果をもつ。

なお、参照文献は入力ごとに、

- 発行年が越えていないか
- 同じ参照文献はないか

のエラー・チェックを行うようになっている。

索引ファイルの作成

リスト、検索、関連性などの出力処理を高速に行うために、あらかじめ索引ファイルや分布データを作成する。この索引の更新をしなければ、新たに入力した文献や参照文献は利用できない。

索引の作成はソート処理が中心であ

る。16 ビットのパソコンは 1 秒間で約 1500 のデータをソートする。次に、各ジョブの内容を示す。

1. 文献：各種索引の作成

- 著者 (20 桁) + 発行年 (3 桁)
- 出典 (8) + 発行年 (3)
- 発行年 (3) + 著者 (8)
- 種別 (1) + 発行年 (3) + 著者 (5)
- 標題 (8) + 発行年 (3)

2. キー：標題キーの作成

- 標題キー (12) + 発行年 (3)

3. 引用：引用索引などの作成

- Reference (4) + 発行年 (3) + 著者 (1)
- Citation (4) + 発行年 (3) + 著者 (1)

● Referations の作成

4. 分布：パラメータの推定

- ブラッドフォード分布、ジップ分布、ロトカ分布、成長・寿命分布。

RAM ディスクを使用して 1230 文献を処理する時間は、文献 (65 秒)、キーワード (339 秒)、引用 (241 秒)、分布 (59 秒) の各ジョブ・トータル約 12 分である。

また、英文標題からの単語の切り出し方は、

- アルファベット 2 文字以下
- AND, FOR, FROM, NOT, SOME, THE, WITH を除く文字列とした。

サブ DB の作成

分析や出力は文献全部を対象とするが、検索した文献群だけに限定する場合は、このサブ DB を実行する。分布データを除く全索引ファイルを書き換える。入力関係の [6 文献], [7 参照], [8 索引] は使用できなくなるが、

いつでも文献全部の状態へ戻ることができる。

例えば、標題に CITATION をもつ文献は 1230 編の中で約 330 編ある。そのサブ DB を作成すると、以降の分析や出力は 330 編だけに限定したものになる。これは DB の利用価値を非常に高くする。なお、330 編の場合、RAM-DISK を使用した処理時間は約 75 秒である。

以上見てきたように、単なる文献 DB の作成とは異なった全く新しいタイプの DB が誕生した。次に、その利用法について述べる。

の所から自由に表示できることである。

参照文献の入力によってもたらされる第 1 の利点は、文献のリファレションが年代順にリストできることである (図 15)。これにより、研究者は文献の発展過程を容易に理解できるようになり、分野全体の中での文献や理論の位置づけが可能になる。

よく知られた KWIC, 文献, 著者, 出典, 発行年, 種別, 標題など 7 種類の索引が簡単な操作で画面に表示できる。そのメニューを以下に示す。

- 索引 (nn) 表示可能な索引と文献数
- [/索引 CR] 表示する索引の探索
- [↑↓] 索引を前後にスクロール
- [CR] 文献の表示。

図 16 は KWIC 索引の例で、書式が圧縮や定型の場合は KWOC 索引になる。

インテリジェント・リサーチの実例

今まで見過ごされていた参照文献がコンピュータの支援で簡単に入力できるようになった。その結果、研究に役立つ高度な情報が得られ、研究者はそれをもとに研究を進めることができる。ここでは、文献情報のリスト、分野全体の情報、索引ベースの検索、新しい情報検索の世界、そして高度な書誌情報の利用などを紹介する。

リファレション・索引のリスト

初期画面で [1 リスト] を選択すると文献の内容がいろいろな角度から出力できる。文献は新しい順に表示され、その書式には 1 行、圧縮、定型の 3 種類がある。特徴として、引用・参照文献数が表示されるために重要文献かどうかの判断ができること、そして任意

分野全体の情報

引用、キーワード、著者、出典に関する 4 種類のランク、キーワード、著者、出典に関する 3 種類の名前、引用、発行年、種別に関する 3 種類の度数、そして 10 種類の分布グラフが F キー

図 15 ● 文献 860 のリファレション・リスト

文献番号	発行年	種別	論文	5C	23R
ASAI, I.					
J. ASIS, 32-2, 113-119			その他		
A GENERAL FORMULATION OF BRADFORD'S DISTRIBUTION. THE GRAPHIC APPROACH					
11173=85	12	0	11	NELSON, M. J.	SPLIT SIZE-RANK MODELS FOR
11878=84	10	1	8	ITO, H.	BRADFORD LAW IN R&D EXPAN
41110=83	32	0	31	SCHUBERT, A.	QUANTITATIVE STUDIES OF S
11891=83	15	1	13	SUMMERS, E. G.	BRADFORD'S LAW AND THE RE
12872=81	7	1	5	KANTOR, P. B.	RIGOROUS SOLUTION OF THE
12802=81	24	5	29	ASAI, I.	A GENERAL FORMULATION OF
12807=78	9	5	3	PRAUNLICH, P.	BRADFORD'S DISTRIBUTION.
12807=78	9	7	1	BROOKES, B. C.	FREQUENCY-RANK DISTRIBUTI
12807=78	25	8	16	DROTT, M. C.	AN EMPIRICAL EXAMINATION
12802=77	13	7	5	BOOKSTEIN, A.	PATTERNS OF SCIENTIFIC PR
12802=77	14	3	10	LEIMKUHLER, F. F.	OPERATIONAL ANALYSIS OF L
12807=76	58	36	21	PRICE, D. J. DE. S.	A GENERAL THEORY OF BIBLI
12807=76	12	1	10	HASPERS, J. H.	THE YIELD FORMULA AND BR
12818=75	32	10	21	POPE, A.	BRADFORD'S LAW AND THE PE
12802=73	13	7	5	LAURENT, S. M.	BRADFORD'S LAW AND THE

図 16 ● KWIC 索引の例

索引	文献数	索引	文献数
11857=83	0	5	+RNS AMONG BIBLIOG = A
01146=80	2	0	+CATION = A MODEL OF
02955=80	5	0	+YSIS OF A CO-CITATION
12848=78	0	0	+NCE JOURNAL CITATIO =
12802=73	0	0	+UAL EXPERIMENTS USING
11218=72	0	0	ON BY THE ADDITION OF
12807=81	0	2	+UTOMATIC INDEXI = THE
12809=80	0	0	+ OBJECTIVE AND SUBJ =
12807=80	0	3	F SECONDARY SERVICES.
42820=79	0	16	OF SCIENTIFIC GROWTH.
42802=76	1	1	ROBLEMS IN ESTIMATING
42802=76	1	1	ROBLEMS IN ESTIMATING
12802=73	3	2	MASTER'S THESES. ONE
02802=65	2	0	+IFIC RESEARCH = THE
32217=63	0	4	+INFORMATION (DE = THE
02802=59	1	0	AN APPROACH TO THE

表3 ●引用文献数のランク

No.	番号	年	引用	参照	著者名
1	120	63	110	10	PRICE, D. J.
2	121	65	96	9	PRICE, D. J.
3	9	48	78	0	BRADFORD, S. C.
4	56	72	77	30	GARFIELD, E.
5	140	73	50	8	SMALL, H.

表4 ●参照文献数のランク

No.	番号	年	引用	参照	著者名
1	94	74	25	99	LINE, M. B.
2	567	77	9	94	NARIN, F.
3	104	74	25	87	MEADOWS, A. J.
4	750	79	36	87	GARFIELD, E.
5	981	81	6	80	SMITH, L. C.

とメニュー選択で簡単に出力できる。

表3と表4は引用文献ランクや参照文献ランクの例である。1963年に発行のPriceの著書⁶⁾が引用文献数110でトップである。これは1230文献のなかの110文献がPriceの文献を参照していたことを示し、かなり高い数字である。

上位ランクの文献は、この分野の研究を始める人の必読文献リストとみなせる。研究に関する専門情報が客観的に得られることの意義は大きい。

名前リストや度数リストが用意されている。また、分野全体の特性を調べるために、ブラッドフォード分布やロトカ分布など計量情報学であつかう10種類のグラフをプロットする。図17は情報源の分布を表すブラッドフォード分布の例で、画面の左側は曲線の数学モデルとその推定値を示す¹³⁾。索引の作成時に前処理しているため、グラフは数秒で描かれる。

索引ベースの検索

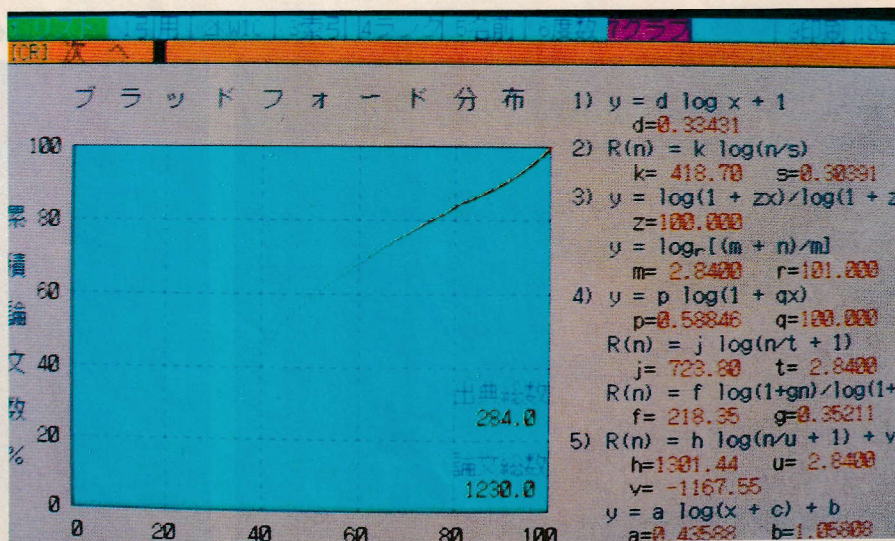
初期画面で[2 検索]を選択する。必要な文献群の検索が容易に行えるように引用や索引の内容はすべて画面に表示される。9種類の表示をもとに6種類の索引検索が可能である。画面は上下の2つに分かれており、上部は検索結果、下部は索引などを表示する。

従来の検索方法と同様に、文献番号、キー、著者、出典、発行年、そして種別をもとにした検索が可能である。図18はキーワードをベースにした検索画面の例で、画面下部はキーワードとその文献数である。語尾変化のあるキーワードが簡単に見つかるので検索の際に役立つ。

索引検索時のメニューを示すと、[m1,m2CR]検索最初、最後コード入力[/索引CR]索引の探索[↑↓] 索引を頁単位にスクロールである。

メニューは大体同じパターンに設計してある。その他に、キー、著者、出典の各ランクをもとにした検索もできるようにになっている。

図17 ●ブラッドフォード分布



新しい文献検索の世界

リファレションを用いた引用検索には3種類のタイプがある。それは、

- 文献のリファレションの検索
- 引用ランクの上位文献の検索
- 文献に関連性の高い文献群の検索である。

図19は文献のリファレションの検索例である。画面の下部には各文献がもつリファレション数がリストされる。文献コードの527を入力するとリファレションの58文献が検索される。次に、引用ランクの上位文献の検索はリファレション、引用文献そして参照文献の3種類のランクが表示されるので、その中から1つを選

択し、必要な文献数を入力する。

最後の関連性の高い文献群の検索は、ある1つの文献に対して、前に述べたリファレーションを用いた3種類の関連性測度を計算し、その結果を高い順に並べ替えて表示する。従って、測度の1つを選択し、必要な文献数を入力するだけで検索できる。この部分はマシン語を使うため、複雑な計算やソートを数秒で実行する。

検索した文献群に対して操作や表示ができる。その主な機能は、

- 演算 検索式の AND, OR, NOT
- ソート 発行年順, 第1著者順, 出典順, 番号順
- 削除 検索式の一部・全部削除
- 表示 検索式番号, 出力文献数, 書式(1行, 圧縮, 定型, 番号)の指定

である。

高度な書誌情報の利用

検索した文献群の特性は初期画面の[3 関連性]を選択することで得られる。リファレーションをベースにした関連性の測度を用いて、検索した文献群に関連の高い文献群、キー群、著者群、そして出典群が検索できる。

従来の検索では文献だけしか検索できなかったが、ここでは特定の文献群に対して、類似の文献群、よく使われるキーワード、研究者グループ、そして発表する情報源などが客観的に得られる。これはまったく新しい情報検索である。

まず検索した文献検索式の中から1つを選択する。次に検索する書誌項目のFキーを押すと、3種類の関連性測

図18 ●キーワード索引支援による検索

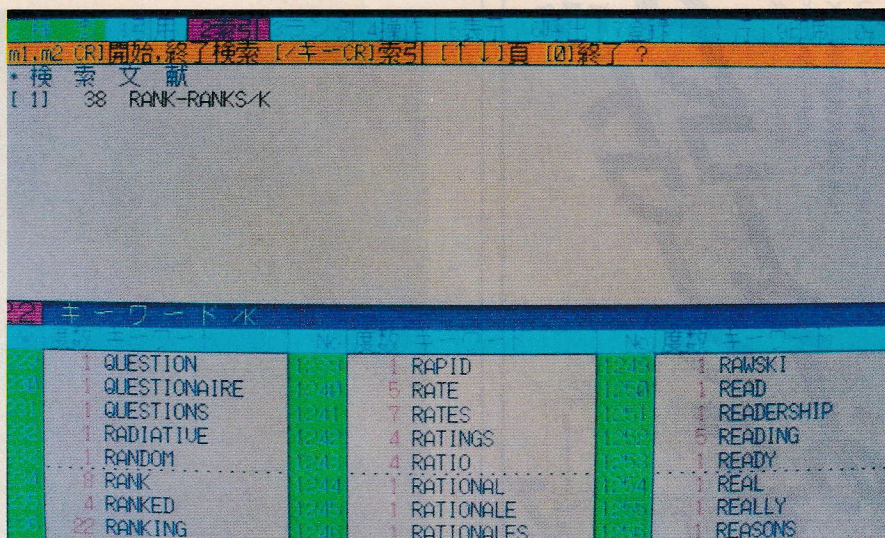
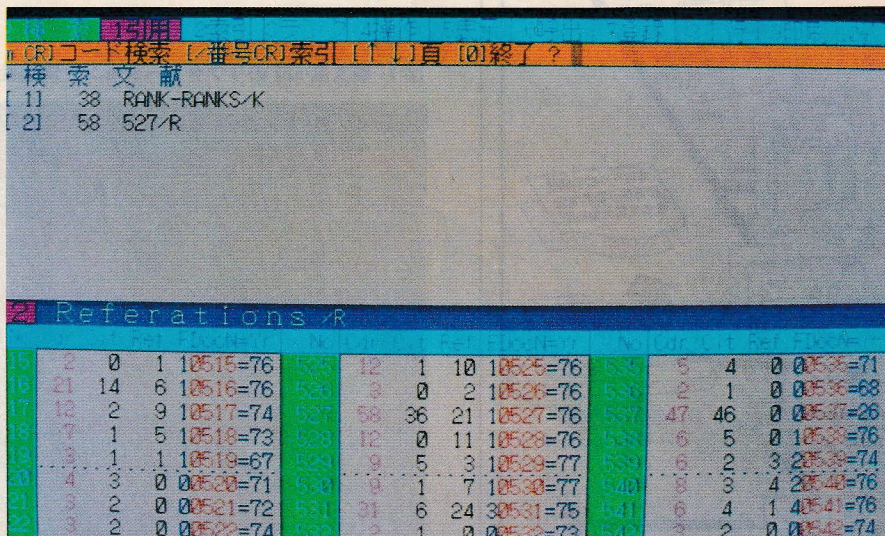


図19 ●文献527のリファレーションを検索



度が表示される。そのメニューは、
[1-3, 1-総数 CR] 測度の選択と上位からの検索数

[↑↓] 測度の頁単位のスクロール
[P] 関連性測度の印刷
である。論理演算, ソート, 削除, 表示などができる。

図20は文献527のリファレーションの58編の文献に関連した文献群を検索した例で、図21は同じ文献群から

著者群を検索した例である。小規模で簡単なDBであるが、そこから得られるものは大きい。対象とする文献群はどんなものでもよい。例えば、最近発行の文献群を使えば動向分析になる。単にマッチングさせるものと異なり、直接に関係がないものも見つけだせる。

なお、使用したパソコンはNECのPC-9801 VM2で、2メガバイトのRAMを増設している。また、ソフトは

図 20 ●文献 527 のリファレーションと関連性の高い文献群の検索例

2文献

1-3,1-1002 (R) 選択, 迄検索 (↑ ↓) 自 (P) 印刷 (Q) 終了?

* 文献検索式

[D01] 38 RANK-RANKS/K
 [D02] 58 527/R
 [D03] 25 527/R<X
 [D04] 25 527/R<Y
 [D05] 12 (3*4)
 [D06] 16 (2*3)

(1) 関連度数 (2) 関連係数 (3) 関連指数

1	33120=63	523	7018	41031=82	0.1011	315	41031=82	89.88	315
2	10567=77	519	6032	10527=81	0.0968	270	10527=76	87.63	398
3	38750=79	419	7192	10527=76	0.0968	398	10567=77	86.32	519
4	10121=65	410	6148	10268=69	0.0972	359	10268=69	84.57	359
5	10527=76	398	3364	41032=82	0.0834	331	10527=81	83.42	270
6	40580=81	340	3712	10527=81	0.0811	176	40580=81	81.04	390
7	10268=69	359	2958	40980=81	0.0806	390	33120=63	82.29	523
8	30104=74	342	6554	41023=82	0.0803	212	41032=82	81.01	331
9	41032=82	331	2784	10527=81	0.0803	270	10527=81	80.93	331

図 21 ●文献 527 のリファレーションと関連性の高い著者群の検索例

4著者

1-3,1-728 (R) 選択, 迄検索 (↑ ↓) 自 (P) 印刷 (Q) 終了?

* 著者検索式

[A01] 15 527/R<X
 [A02] 15 527/R<Y
 [A03] 15 527/R<Z
 [A04] 4 (1*2)
 [A05] 9 (2*3)
 [A06] 10 (3*1)

(1) 関連度数 (2) 関連係数 (3) 関連指数

1	GARFIELD, E.	2750	53708	AURAMESCU, A	0.1362	519	PRICE, D. J. D	77.56	231
2	PRICE, D. J. D	2313	27782	HAITUN, S. D.	0.1271	786	HAITUN, S. D.	79.76	78
3	BROOKES, B. C	1592	19372	DROTT, M. C.	0.0870	411	AURAMESCU, A	71.72	51
4	NARIN, F.	1152	19082	NARANAN, S.	0.0861	362	GARFIELD, E.	69.39	275
5	SMALL, H.	1144	23548	FAIRTHORNE, P.	0.0849	423	BROOKES, B. C	68.66	159
6	GRIFFITH, B.	862	14500	HUBERT, J. J.	0.0810	712	HUBERT, J. J.	69.71	71
7	COLE, S.	832	15892	ZLINDE, P.	0.0810	417	DROTT, M. C.	69.21	41
8	COLE, J. R.	661	15370	KANTOR, P. B.	0.0875	178	FAIRTHORNE, P.	64.71	423
9	HAITUN, S. D.	786	5452	UNOS, H.	0.0830	302	NARIN, F.	62.29	1152

MS-DOS 上の N 88-日本語 BASIC (86) コンパイラと 8086 マシン語で開発した。

おわりに

表形式の簡単な文献ファイルに、各文献間の関係を表すリファレーション

・ファイルを追加したリファレーション DB を用いて、研究を進めるのに必要な一種の専門家情報を得るソフトを開発した。今まで見捨てられていた参照文献も、コンピュータを利用して分析すれば価値が高くなることが明らかになった。

参考文献

- 1) 仲本秀四郎「計量情報学」, 『ドクメンテーション研究』, 31-4, p.p. 161-168, 1981.
- 2) 『参照文献の書き方: 科学技術情報流通技術基準』, SIST 02-1984, JICST.
- 3) Weinstock, M., "Citation Indexes", *Encyclopedia of Library and Information Science*, vol. 5, 1971.
- 4) Kessler, M. M., "Bibliographic Coupling Between Scientific Papers", *Am. Documentation*, 14-1, p.p. 10-25, 1963.
- 5) Small, H., "Co-citation in the Scientific Literature; A New Measure of the Relationship Between Two Documents", *J. ASIS*, 27-4, p.p. 265-269, 1973.
- 6) デレック・プライス (島尾永康訳) 『リトル・サイエンス ビッグ・サイエンス』, 創元社, 1970.
- 7) Garfield, E., *Citation Indexing: Its Theory and Application in Science, Technology and Humanities*, John-Wiley & Sons, 1979.
- 8) 宮本定明『計量書誌統計処理パッケージの作成』, 1984.
- 9) Asai, I., "Adjusted Age Distribution and Its Application to Impact Factor and Immediacy Index", *J. ASIS*, 32-3, p.p. 172-174, 1981.
- 10) 浅井勇夫「パソコンによる引用文献データベースの開発」, 『第21回 JICST 論文集』, 21-31, 1984.
- 11) 浅井勇夫「Referation を用いた特定専門分野情報の分析」, 『第22回 JICST 論文集』, p.p. 135-142, 1985.
- 12) Asai, I., "The Structure and Usage of A "Referation" Database On Microcomputers For Personal Document Management", *Appl. Micro-computers in Inf., Doc. & Libr.*, p.p. 510-517, North-Holland, 1986.
- 13) Asai, I., "A General Formulation of Bradford's Distribution: The Graph-Oriented Approach", *J. ASIS*, 32-2, p.p. 113-119, 1981.

日本独自の基礎研究を発展させるためには、研究者の頭脳に頼った手探りの研究からコンピュータを利用した知的な研究へ脱却することが必要である。研究者が独自のリファレション DB をもとに研究を進めることのできる研究社会の到来が望まれる。 **NC**

煩わしい手順や規約よさらば
TORNADO/VATで
今日から軽快アクセス

富士通ソーシャルサイエンスラボ