

Referation による引用検索

浅井 勇夫

大阪府立大学工学部

1. はじめに

研究者が収集した文献コレクションをデータベース化し、分野の動向分析や技術評価に利用すれば、研究開発の生産性は向上する。ここでは、高性能・低価格なパソコンを使用して、研究者が容易に操作できる文献管理用ソフトを開発することを目的とする。

引用文献データベースは、書誌項目と文献の末尾にある引用文献とから構成する。引用文献のデータベース化によって、Citations (=Citing Papers) よりも広義の Referations (=Citations+Document+References) が作成できる。ここでは、データの作成方法と引用検索(Referation Search) について説明する。

2. 引用文献データベースの作成

データの作成や利用を研究者自身が行なえるように、(i)入力項目の少ない、(ii)操作性の高い、設計を試みた。以下、システムの概要を述べる。

2.1 ハードウェアの構成

- (1)パソコン本体：PC-9801 シリーズ
- (2)ディスプレイ：カラー高解像度CRT
- (3)ディスク：1 MB 用
- (4)増設RAM：256 KB 増設。

2.2 ソフトウェアの構成

プログラミング言語は N88-BASIC(86)を用いた。ソートやマージはマシン語(8086)を使用する。プログラム容量は約 180K バイトである。画面構成やカラー表示に工夫した。また、マニュアルも出力可能である。

初期画面のジョブ選択により、各処理ルーチンを実行する。主なルーチンは、

- (1)メイン：初期画面
- (2)文献ファイルの作成
- (3)引用ファイルの作成
- (4)索引ファイルの作成
- (5)各種リスト
- (6)引用索引。

2.3 文献ファイルの作成

文献は1レコード 256バイトのランダムファイルとして保存する。1レコードのフォーマットの内容を次に示す。

- (1)種別：1桁
- (2)文献番号：4桁
- (3-5)著者：各20桁
- (6)標題：128桁
- (7)出典：35桁
- (8)発行年：4桁
- (9)その他：24桁。

なお、作成時に著者・標題・出典の索引を画面に表示して入力作業を支援する。

2.4 引用ファイルの作成

引用文献の入力は次のような方法で処理した(第1図)。

- (1)引用文献は文献ファイルに登録した文献のみを対象とする、
- (2)引用ファイルは文献番号で構成する、
- (3)著者索引を表示して入力を支援する。

[2引用] Referation Analysis & Informetrics (1160) 【引用文献管理】

文 献	10212 SHALL,H., GRIFFITH,B.C., SOC.STU.SC.4-1,17-40 (1974) THE STRUCTURE OF SCIENTIFIC LITERATURES 1. IDENTIFYING AND GRAPHING 10R SPECIALITIES
5	30152=71 WEINSTOCK,H. CITATION INDEXES (ENCYCLOPEDI MAGRAW-HILL,16-40
引	6 10250=68 MALIN,M.V. THE SCIENCE CITATION INDEX. A LIB.TRE.16-3,374-387
用	7 10071=66 GOFFMAN,W. MATHEMATICAL APPROACH TO THE NATURE. 212,449-452
8	10121=65 PRICE,D.J.DE. NETWORKS OF SCIENTIFIC PAPERS SCIENCE,149,510-515
文	9 00051=64 GARFIELD,E. THE USE OF CITATION DATA IN W ISI 86P
献	10 10083=63 KESSLER,M.M. BIBLIOGRAPHIC COUPLING BETWEEN J.ASIS. 14-1,10-25
1	20263=69 PRICE,D.J.DE. CITATION PRACTICE ASLIB P.21-8,328
番	00237=69 PRICE,D.J.DE. MEASURING THE SIZE OF SCIENCE P.ISR.AC.4- .98-111
2	00123=67 PRICE,D.J.DE. NATIONS CAN PUBLISH OR PERISH SCI.TEC.70- .84-90
3	10122=66 PRICE,D.J.DE. COLLABORATION IN AN INVISIBLE AM.PSYC.,21-11,1011-
4	00422=66 PRICE,D.J.DE. SCIENCE. THE SCIENCE OF SCIEN MED.OPI.RE.1- .98-97
索	5 00753=65 PRICE,D.J.DE. IS TECHNOLOGY HISTORICALLY IN TECH.CUL.1- .553-568
引	

第1図 引用ファイルの作成画面

2.5 索引ファイルの作成

検索や分析を高速化するために、あらかじめ、39種類の索引ファイルと5種類の分布ファイルを作成する。

3. 引用検索

ISI社のSCI,SSCIは、アブストラクトやディスクリプタの代わりに引用文献を入力する。その転置ファイルのCitationsを利用して研究が多数行なわれている。ここでは、Citationよりも広義な概念を導入し、それを利用した新しい文献の検索を試みる。

3.1 Referations

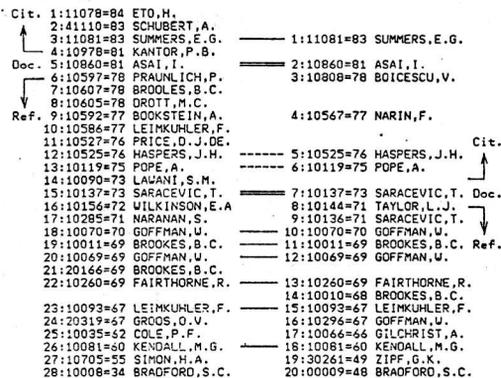
SCI,SSCIを用いた分析は、(1)文献末尾の引用文献を全部入力する、(2)第1著者、出典省略形、巻・号・頁・発行年を入力する、(3) Citations のみを使用する。従って、(1)から分野間の関係が把握できるという長所をもつが、(2)から引用カウントによる評価は第2著者以降を含まず不正確になる、標題キーワード分析は標題を含まず不可能である、そして(3)から比較的旧い文献しか分析できない、という短所をもっている。

ここで行なう分析の特徴は、

- (1) 収集した文献群に関する引用文献のみを使った分野内分析を行ない、分野間分析は行なはない、
- (2) 文献番号を入力するだけであるが、すべての書誌項目を入力したことになるため、標題キーワード分析などが可能になる、
- (3) 入力した References(=Cited Papers)から転置した Citations(=Citing Papers)を作成し、その両者から Referations(=Citations+Document+References)を作成する。これを用いた分析は、ある分野内の新しい文献と旧い文献を同じレベルで扱うことができる。

3.2 Referations による文献の関連性

1963年に、M.M.Kesslerは References 間のカウントをもとに書誌結合を、1973年に、H.Small は Citations間のカウントをもとに共引用を提案した。ここでは、Referations 間のカウントをもとにした関連性の測度を提案する。第2図は書誌結合や共引用を含むと共に、Reference と Citation とのカウントを含む例を示す。



第2図 文献 860と 137の関連度数

ここで用いる関連性の測度を定義すると、

- (1) 関連度数 = z
 - (2) 関連性 1 = z / (x + y - z)
 - (3) 関連性 2 = SQR (z^3 / x / y)
- ここで、文献 a の Referations A の文献数を x、文献 b の Referations B の文献数を y、そして A と B に共通な文献数を z (=A U B) とする。関連性 2 は関連度数と関連性 1 との中間的な新しい測度である。

3.3 Referation Search

著者、出典、キーワードによる文献の検索は、ランクや索引の支援により実行できる。さらに、任意の文献に関する Referation をもとに次の4種類の文献群が検索できる。

- (1) Referations からなる文献群、
- (2) 関連度数の高い文献群、
- (3) 関連性 1 の高い文献群、
- (4) 関連性 2 の高い文献群。

また、検索結果の操作(AND, OR, NOT)と出力は任意に行なうことができる。

第3図は、Referation Search の出力画面の1例を示す。文献番号 860を入力すると、数秒以内に検索・表示される。

[5] 28 引用 (0860)

[6] 22 関連度数 (0860(8))

[7] 23 関連性 2 (0860(0.80))

[8] 14 (引用 (0860) AND 関連性 2 (0860(0.80)))

[9] 37 (引用 (0860) OR 関連性 2 (0860(0.80)))

●関連度数(0860)									
1:10860=81	28M	28T	6:10567=77	12M	103T	11:10156=72	11M	31T	
2:00009=48	17M	76T	7:10093=67	12M	45T	12:10069=69	10M	31T	
3:10011=69	15M	53T	8:10982=81	11M	27T	13:10119=75	10M	32T	
4:10260=69	15M	53T	9:10137=73	11M	20T	14:10525=76	9M	12T	
5:40736=77	14M	63T	10:41031=82	11M	32T	15:10148=48	9M	33T	

●関連性1(0860)									
1:10860=81	----	28M	6:10260=69	.238	15M	11:10090=73	.206	7M	
2:10137=73	.297	11M	7:10156=72	.229	11M	12:10069=69	.204	10M	
3:10525=76	.290	9M	8:41031=82	.224	11M	13:10170=70	.200	9M	
4:10982=81	.250	11M	9:10605=78	.209	9M	14:10119=75	.200	10M	
5:10011=69	.244	16M	10:10978=81	.207	6M	15:10093=67	.197	12M	

●関連性2(0860)									
1:10860=81	----	28M	6:10525=76	1.47	9M	11:10093=67	1.17	12M	
2:10011=69	1.46	16M	7:10982=81	1.33	11M	12:10069=69	1.07	10M	
3:10260=69	1.55	15M	8:40736=77	1.25	14M	13:10119=75	1.06	10M	
4:10137=73	1.54	11M	9:10156=72	1.24	11M	14:10978=81	1.05	6M	
5:00009=48	1.52	17M	10:41031=82	1.22	11M	15:10605=78	1.04	9M	

第3図 Referation Search の例

4. おわりに

引用ファイルから得られる Referation を利用した引用検索の可能性を明らかにした。今後、Citation の代わりに Referation を用いた分析が有用となろう。

【参考文献】

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回情報科学技術研究会、1984年