

引用索引(Citation Index)

柳 京 熙
KORSTIC 調査檢索部長

1. 引用索引이란?

引用索引(Citation Index, CI)이란 在來의 몇 가지 索引方法의 弱點을 많이 補完한 比較的 새로운 方法에 의한 索引이다. 在來式 索引方法에 比하여 그 長點은 첫째, 文獻과 文獻間의 相關關係를 把握할 수 있다는 것이고 둘째, 主題를 分析하여 索引하는 專問家가 必要없이 Man-Machine 索引이 可能하다는 것이다. 그리고 셋째로, 引用文獻을 從來의 主題索引方式에 의하면 科學技術老化의 影響을 많이 받았는데 이 索引法은 그 影響을 받지 아니한다.

CI는 簡單한 概念을 바탕으로 한 것으로서, 어떤 사람(著者)이 어떤 情報를 生産한 경우, 여기에 이미 發生한 다른 情報가 引用되었을 것이다. 論文인 경우, 이 引用된 것을 通常 參考文獻이라고 하고 있다. 이러한 情報를 그 引用情報와 함께 配列하여 만든 目錄이 바로 引用索引이란 것이다.

처음으로 이 方法이 應用되기로는 美國의 Shepard Citation이라는 것이며 이것은 1873年 以後의 法廷記錄을 整理하여 둔 것인데 類似事件의 判例를 簡單히 찾을 수 있도록 한 것이다.

이 시스템을 科學文獻에 適用하기로는 1950年代 初期였다. 그러나 1963年에 처음으로 科學引用索引(Science Citation Index, SCI)이 科學情報研究所(Institute for Scientific Information, ISI)에 의하여 發刊되었다.

2. 主題索引의 難點

오늘날에 이르러 科學技術文獻은 幾何級數적으로 增加되어 必要한 情報를 찾아 내기가 더욱 어렵게 되었다. 推定되기로는 1975년까지는 200萬名이 넘는 科學者가 每年 100萬件以上の 論文을 生産할 것이며, 이미 發行된 論文 1,000萬件을 합치면 그 規模는 어마어마하다. 거기에다가 界面科學에서의 情報交換이 急増하고 있을 뿐 아니라 이미 細分된 各分野間의 情報交換도 해마다 늘어나고 있다.

主題索引, 키워어드索引 및 UDC에 의한 索引 등은 單純히 한 個의 分野 또는 細分分野를 한 個 또는

몇 個의 主題, 單語 또는 記號로서 커버하는 것이 大部分이다. 最近에 있어서도 어떤 個人의 特定分野에의 獨斷적인 執着때문에 他分野에게 有效할 수 있는 情報가 死藏되어 버리는 경우가 많다. 예를 들면, 어떤 化學者가 治療用 藥品이나 人工臟器의 材料를 選擇하기 위하여 化學, 醫學 및 工學分野의 雜誌를 모두 읽어 보아야 한다는 難點이 있다. 그리고 特히 綜合科學적인 性格을 띤 海洋學, 有機化學, 還境科學, 原子力科學 등의 分野도 이제는 1個分野로서의 작은 房에 가두어 들 수가 없게 되어 버렸다.

主題索引 등에서 使用되는 分類語가 曖昧하거나 때로는 다른 解釋에 오히려 알맞은 경우가 자주 있기 때문에 檢索하기가 매우 어렵다. 그리고 主題索引에서는 새로운 概念에 대하여 새로운 表現을 하여야 한다. 事實上 어떤 새로운 概念에 대해서 適合하다고 할 만한 表現은 恒常 나올 수 있는 것이 아니다. 왜냐하면 그 概念을 創案한 著者 自身の 表現이 不適合한 경우가 많기 때문이다. 이런 경우는 急變하는 分野일수록 많은데 生化學에서는 特히 더 많다.

主題索引에서는 반드시 索引專問家가 있어야 하는데 이들의 知的能力和 熟練度가 각각 다르기 때문에 檢索을 위한 統一된 索引語의 抽出이 어렵다. 두사람의 索引者가 同一한 文獻을 分類 또는 索引을 할 때 각각 다른 分類 또는 索引語를 抽出하는 例가 많다. 이러한 問題를 解消하기 위하여서는 索引者의 知識에 相關하지 않고 語義의 困難性에 不拘하고 統一된 索引語를 使用하여야 하며 한편 索引語와 檢索語와의 關聯을 지워야 하는 등 龐大한 作業이 先行되어야 한다.

3. SCI의 作業過程

ISI가 施行하고 있는 SCI는 每年 約 2,200種의 定期刊行物의 內容을 引用索引法에 의하여 索引된 것이다. 이러한 刊行物을 소오스 저어널(Source Journal)이라고 하고 여기에 실린 記事를 소오스 記事(Source Item)라 한다. ISI는 現在 廣告나 書籍評 같은 記事를 소오스記事로 取扱하지 않고 있다고 한다.

入手한 소오스 저어널에 있는 有用한 데이터를 편집

하거나 標識해 두고, 여기서 外國語로 된 題目은 英文으로 翻譯한다. 이것을 보고 IBM카드에 穿孔을 한다. 여기에는 소오스記事와 이 소오스記事안에 있는 脚註나 參考文獻에 있는 著者, 題目, 誌名 등의 데이터가 穿孔이 된다. 穿孔된 카야드는 다시 소오스記事와 直接 比較하면서 檢孔을 하게 된다. 檢孔이 끝난 카야드의 데이터를 바로 磁氣테이프에 收錄시킨다. 이 過程에서 著者名이나 題目에 틀린 스펠링이 있는지를 確認하는 過程이 電子計算機에 의하여 行하여진다.

最終적으로 電子計算機에 의한 Photo-composition machine을 使用하여 편집결과를 印刷하도록 되어 있다. 그리고 各種 索引集은 Photo-offset 印刷로서 作成되며 여러 가지 目的에 의한 統計表도 適切한 書式에 따라서 찍혀 나오게 하였다.

4. SCI의 形式

SCI는 다음 세 가지로 構成되어 있다.

- ① 引用索引(Citation Index, CI)
- ② 소오스索引(Source Index, SI)
- ③ 퍼어뮤터엄主題索引(Permuterm Subject Index,

PSI)

이 세 가지 索引은 서로가 關聯되어 있으며 每年 季刊으로 發行되는데 年末에 나오는 것은 當該年度의 累加索引이다. 1969年度 SCI의 年間 累加索引은 約 1400面, 11卷으로 되어 있는데 이것은 341,000個의 소오스記事에서 抽出한 4百萬個의 引用文獻이 索引되어 있다.

A. CI

CI는 引用된 著者의 알파벳順으로 되어 있고 引用文獻의 엔트리는 첫번째 著者名, 發行年度, 誌名, 卷, 面의 順으로 되어 있다. 그 예는 그림 1에 나타나 있다.

어떤 著者에 대한 引用記事가 둘 이상인 경우는 그 引用年度의 順序대로 配列되어 있다. 그리고 어떤 參考文獻을 引用한 소오스記事는 그 著者의 알파벳順으로 그 行 바로 아래에 配列하였다. 그래서 CI를 通하여서는 特定人의 論文이 누구누구의 論文에 引用되었다는 事實을 쉽게 찾아 볼 수 있도록 편집되었는데 여기에 각각의 書誌데이터가 있기 때문에 關聯文獻을 찾는 데 아주 便利하다.

Citing Author	Reference	Source	Year	Publication	Volume	Page	Citing Source Year
ABRAHAM BM	PESHKOV VM	MOTA AC	65	PHYS REV	98	551	
		HEINICKE W	65	PHYS REV	11	209	
		ANDERSON AC	65	PHYS REV LETTERS	133	133	
		MESS RW	65	PHYS REV LETTERS	177	266	
		MOTA AC	65	PHYS REV LETTERS	20	649	
		ABRAHAM CT	65	PHYS REV LETTERS	39	5878	
		TANG DT	65	PHYSICA	41	260	
		ABRAHAM DJ	65	PHYSICA	177	266	
		ABST 109604	65	PHYS REV	12	143	
		ABRAHAM E	65	PHYS REV	8	48	
		DERMARDE A	65	PHYS REV	1	189	
		FARKAS P	65	PHYS REV	279		
		ABRAHAM EE	65	PHYS REV	14	473	
		APEL JR	65	PHYS REV	58	1	
		ABRAHAM EP	65	PHYS REV	24	74	
		MORIN RB	65	PHYS REV	2	337	
		RITZERFE W	65	PHYS REV	12	291	
		CLAESSEN M	65	PHYS REV	91	1396	
		BODANSZK M	65	PHYS REV	140	24	
		KATZ E	65	PHYS REV	152	477	
		RIEBER M	65	PHYS REV	58	266	
		KASBERAR MG	65	PHYS REV	77	579	
		MORIN RB	65	PHYS REV	68	77	
		DYMOCK IW	65	PHYS REV	57	352	
		KATZ E	65	PHYS REV	163	352	
		BODANSZK M	65	PHYS REV	31	364	
		STOODLEY RJ	65	PHYS REV	50	377	
		TAN HS	65	PHYS REV	35	155	
		ABRAHAM F	65	PHYS REV	79	377	
		HOLYNSKI R	65	PHYS REV	46	183	
		ABRAHAM FF	65	PHYS REV	91	1396	
		ABRAHAM FF	65	PHYS REV	81	1401	
		ABRAHAM FF	65	PHYS REV	42	412	
		ABRAHAM FF	65	PHYS REV	21	468	
		ABRAHAM FF	65	PHYS REV	21	189	
		ABRAHAM FF	65	PHYS REV	21	189	
		ABRAHAM FF	65	PHYS REV	21	231	
		ABRAHAM FF	65	PHYS REV	64	375	
		ABRAHAM FF	65	PHYS REV	88	209	
		ABRAHAM FF	65	PHYS REV	159	1110	
		ABRAHAM FF	65	PHYS REV	176	1661	
		ABRAHAM FF	65	PHYS REV	13	208	
		ABRAHAM FF	65	PHYS REV	25	47	
		ABRAHAM FF	65	PHYS REV	39	3287	
		ABRAHAM FF	65	PHYS REV	39	4791	
		ABRAHAM FF	65	PHYS REV	50	1303	
		ABRAHAM FF	65	PHYS REV	39	4799	
		ABRAHAM FF	65	PHYS REV	40	446	
		ABRAHAM FF	65	PHYS REV	34	5811	
		ABRAHAM FF	65	PHYS REV	40	446	
		ABRAHAM FF	65	PHYS REV	50	1303	
		ABRAHAM FF	65	PHYS REV	48	732	
		ABRAHAM FF	65	PHYS REV	50	1303	
		ABRAHAM G	65	PHYS REV	50	1303	
		PAUPE J	65	PHYS REV	125	160	
		KRAPPE HJ	65	PHYS REV	17	191	
		LARUSO HA	65	PHYS REV	8	69	
		ABRAHAM GJS	65	PHYS REV	1124	441	
		BOND RF	65	PHYS REV	69	268	
		ABRAHAM MC	65	PHYS REV	208	459	
		BERKWER I	65	PHYS REV	216	193	
			65	PHYS REV	125	1067	

그림 1. CI의 한 예

B. SI

CI에는 引用된 記事나 引用記事의 첫번째 著者名만이 明示되어 있다. 그러나 SI에는 각 引用記事의 全著者의 이름이 明示되어 있다. SI는 소오스記事 著者名의 알파베트順으로 配列되어 있는데 여기에는 共著者名, 記事題目, 誌名, 卷, 號, 面, 年度, 記事區分, 參

考文獻數 그리고 雜誌整理番號 등이 있는데 KORSTIC에서 發刊하는 文獻速報와 形式이 비슷하다. 다만 KORSTIC의 速報記事는 UDC分類順에 의하여 配列한 것이지만 SI는 著者名의 알파베트順으로 配列된 것이 다를 뿐이다. SI의 한 例를 그림 2에 表示하였다.

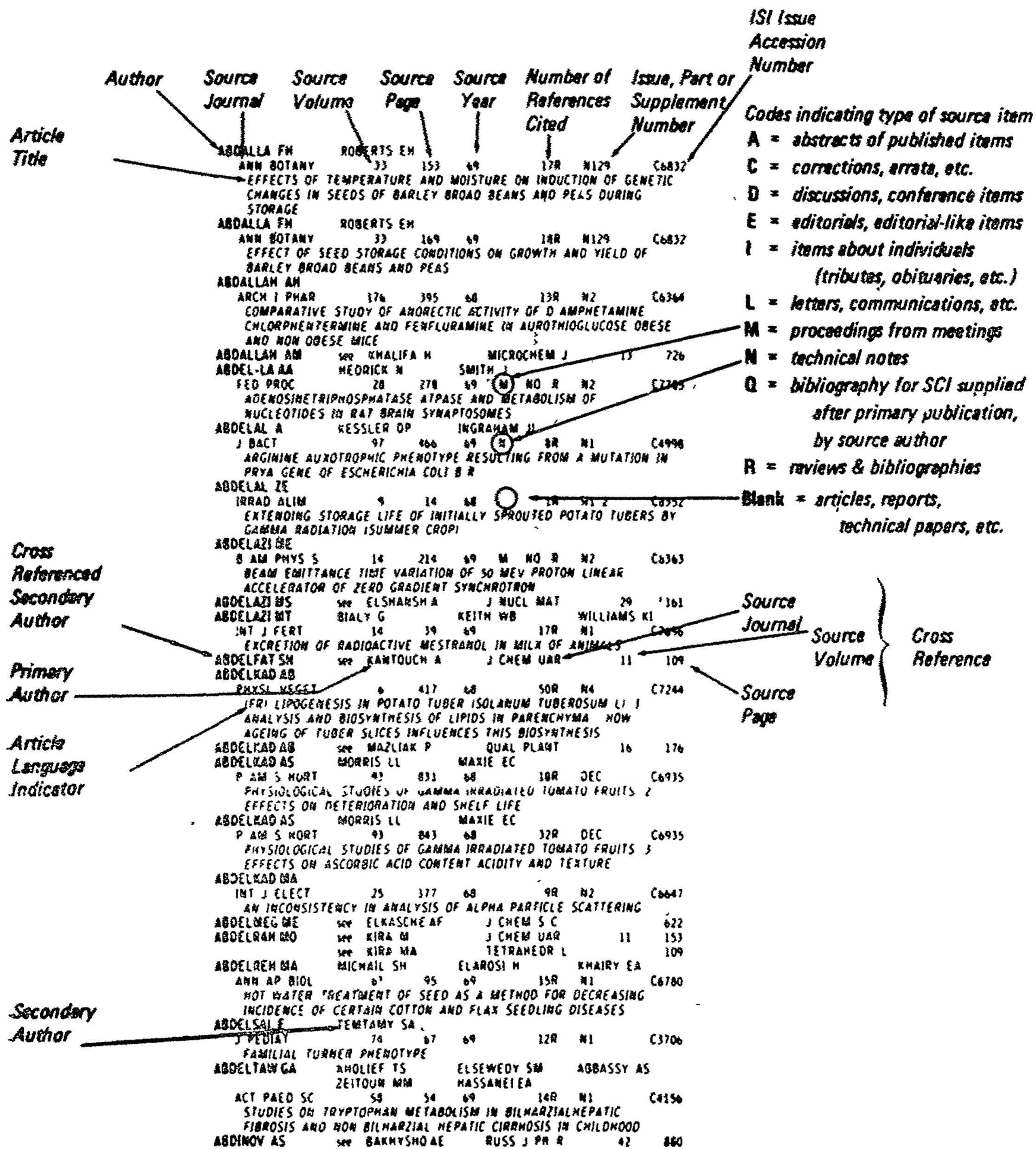


그림 2. SI의 한 例

C. PSI

퍼어뮤터엄主題索引의 퍼어뮤터엄이란 말은 Per-muted Term을 줄여서 만든 것이다. 記事題目에서 冠詞나 前置詞와 같은 單語를 除外한 낱말이 n個가 있다고 하면 각각의 낱말을 見出用으로 順次的으로 내어 보는데, 見出用으로 내어 보내는 낱말을 1次터엄이라고 하고 나머지 낱말들을 모두 2次터엄이라고 한다. 그림

3에서 보는 바와 같이 2次터엄은 바로 1次터엄 아래에 알파베트順으로 配列되어 있다. 그리고 每2次터엄 뒤 에 그와 關聯된 記事의 著者名이 있다. 1次터엄은 알 파베트順으로 配列되어 있어서 찾고자 하는 特定主題 에 關聯된 著者를 쉽게 찾을 수 있도록 되어 있다.

이 技法은 KWIC索引과 類似한 것이지만 그보다 識 別하기 쉽게 되어 있다.

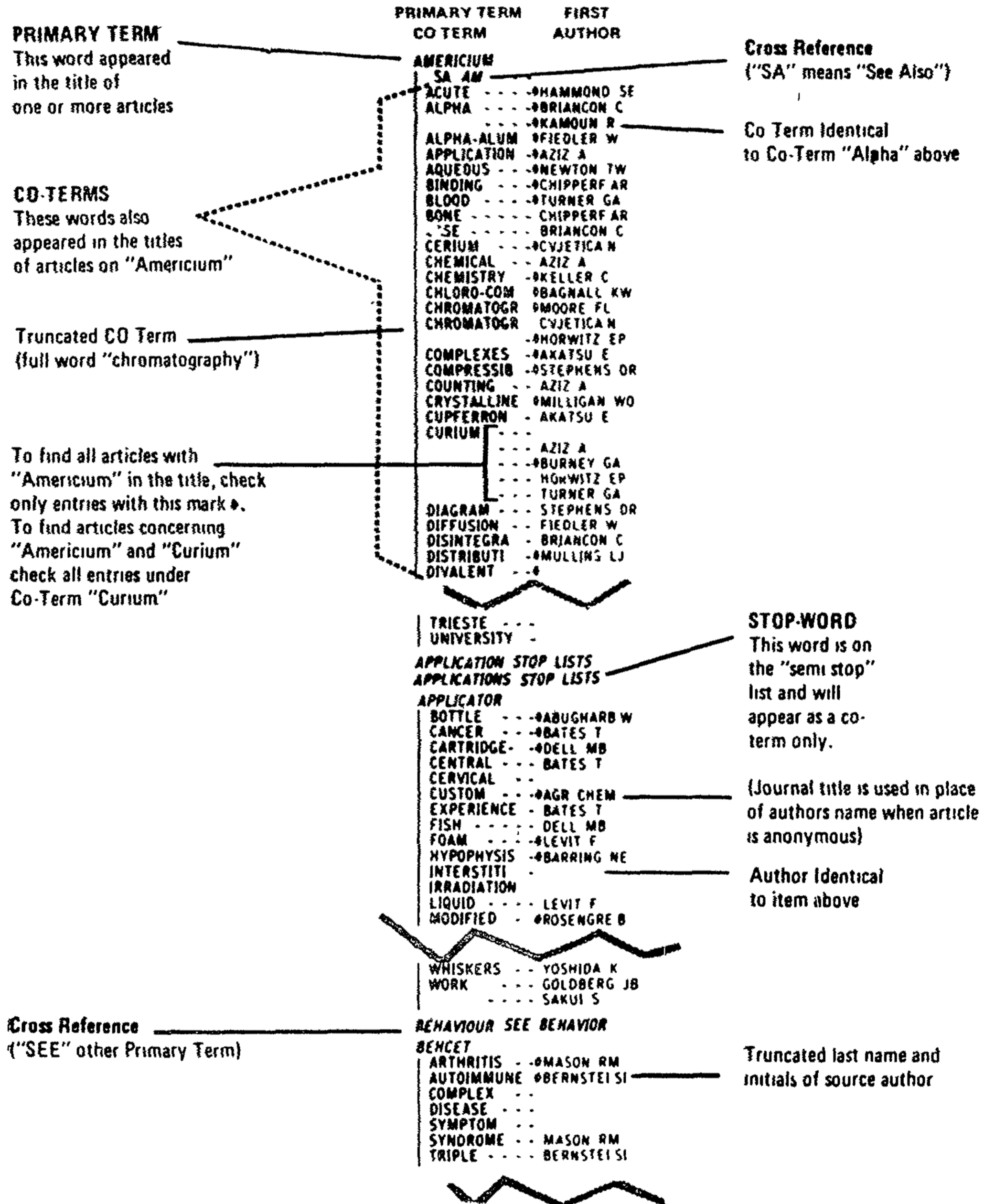


그림 3. PSI의 한 예

5. SCI의 使用法

SCI를 使用하여 要求하는 文獻을 檢索하는 方法을 簡單히 說明하기로 한다. 먼저 檢索者는 찾고자 하는 主題에 關聯된 記事에 적혀 있는 著者名으로부터 始作한다. CI의 著者名이 알파벳順으로 配列되어 있으므로 여기서 그 著者名을 찾는다. 그 著者名 아래에 그 著者의 記事를 引用한 다른 著者名과 書誌事項을 읽어 낼 수가 있다. 그 다음, 引用한 著者名에 따라 SI에서 그 記事의 題目을 읽어 낼 수가 있다. 共著者인 경우에도 SI에서 찾을 수 있다. 여기에서 檢索者가 찾고 있는 主題와 關聯이 있는지 없는지를 判斷하게 된다. 이 過程에서 要求하고 있는 文獻과 關聯이 있는 다른 著者를 多數發見할 수 있으며, 이들 著者名에 따라서 繼續적으로 反復하여 찾으면 關聯文獻을 많이 찾을 수 있다. 이렇게 反復하는 過程을 사이클링 (Cycling)이라고 한다.

萬一, 檢索者가 찾고자 하는 主題에 關聯된 著者名을 알 수 없을 때에는 PSI를 使用하면 된다. 即 CI나 SI를 볼 수 있도록 著者名을 알아내는 道具라고 생각하면 된다. 檢索者가 찾고자 하는 主題에 대한 몇 가지의 檢索語를 羅列하여 두고 이것을 PSI에서 該當되는 著者를 읽어 낼 수가 있다. 著者가 確認만 되면 앞에 말한 사이클링 過程을 繼續 反復함으로써 目的하는 多數의 關聯文獻의 書誌的 事項을 알아 낼 수 있다.

예를 들어 보면, 隕石에 있는 生命體에 관하여 興味를 갖고 있는 檢索者가 있다고 하자. 그는 62年度 Science誌 第137卷 (623~628)面에 실린 Harold C. Urey의 論文만 가지고 있다면 먼저 CI에서 Urey의 '62 Science를 確認한다. 그 아래쪽에 Mueller, G.란 사람이 65年度 Nature誌 第205卷 1200面에 論文을 실었는데 여기에 Urey의 論文을 引用하였다. Mueller, G.란 이름으로 SI를 찾으니 바로 그 論文의 題目이 “炭素質 隕石에서의 마이크로構造 解釋”이란 것이었

A

CITATION INDEX SECTION

CITED AUTHOR	CITING AUTHOR	CITED PUBLICATION	CITING YR	VOL	PAGE
UREY HC		31-PHYS REV		38	1969
	ANN T	J CHEM PHYS	65	42	932
	PILLAI MGK	AUST J CHEM	65	18	261
	TADOKORO H	J CHEM PHYS	65	42	1432
	DOWN ML	GEOL J AM B	65	76	287
		62-SCIENCE		137	623
	MUELLER G	NATURE	L 65	205	1200
	ANDERS E	SPACE SCI R	R 64	3	583

B

SOURCE INDEX WITH TITLES

MUELLER CW	RAD CORP AM	3166694 US	65	P	7R	JAN 19	CL317/235 SYMMETRICAL POWER TRANSISTOR
MUELLER DC	INT BUS MACH	3165719 US	65	P	2R	JAN 12	CL340/166 MATRIX OF COINCIDENCE GATES HAVING COLUMN AND ROW SELECTION
MUELLER EG	SEE SMITH HL	3166263 US					
MUELLER FH	MUELL CO	3168280 US	65	P	1R	FEB 2	
MUELLER G	CLAUS G	NATURE	205	L	10R	N4977	62558
		INTERPRETATION OF MICRO-STRUCTURES IN CARBONACEOUS METEORITES					
MUELLER H	SINGER GD	Z PHYS	182	366	65	22R	N4 61967
		THERMOLUMINESCENCE AND EXO-ELECTRON EMISSION FROM ALKALI AZIDES					

그림 4.

다. 檢索者가 이 程度의 情報만으로 滿足할 수 없을 때에는 Urey나 Mueller의 論文에 引用된 다른 文獻의 著者名으로 CI에서 이와 같은 過程을 反復하면 된다. 이 例를 그림 4와 같이 圖示하였다.

6. 引用索引技法의 應用

引用文獻分析(Citation Analysis)에 의하여 圖書館이나 情報센터에서의 入手資料의 選擇基準을 作成한다면 情報資料의 有效壽命을 決定한다면, 더욱 나아가서 各國別 引用狀況을 把握하여 그 나라의 科學水準을 測定하는 것이 可能하다고 言及된 바가 있다.⁽¹⁾

SCI는 引用文獻分析을 위한 完全한 데이터를 含有하고 있으므로 目的에 따라서 自由롭게 必要로 하는 統計를 作成할 수 있기 때문에 手動的인 統計方式은 이에 比하여 無色하게 되었다.

1955年, 알렌(G. Allen)博士는 歷史研究에 이 引用索引技法을 利用하도록 勸告한 바가 있다. 그는 書誌引用네트워크를 構成하였으며 1964년에는 實用的인 方法을 만들어 내었다. 이 技法은 社會學的 및 歷史學的인 事件, 發生年度, 各 事件間의 相關關係 및 相對的인 重要度 등을 알아 내는데 便利하기 때문이다. 뿐만 아니라 科學技術의 歷史的 背景을 究明하는 데도 아주 有用하다. 그림 5는 核酸에 관한 15個 論文의 引用網의 한 例인데 이를 通하여 그 歷史的 및 社會的 背景을 觀測할 수가 있다. 그리고 그림 6은 멘델(G. Mendel)의 遺傳法則에 관한 論文의 引用네트워크인데 여기에 關하여 簡單히 살펴보기로 하겠다. 멘델의 革新的인 論文은 비록 1865年에 發表되기는 하였어도 當時 科學界가 無識하여 35年이 지난 1900年에 와서 이것의 價値를 “再發見”하였다고 一般的으로 알고 있다. 그러나

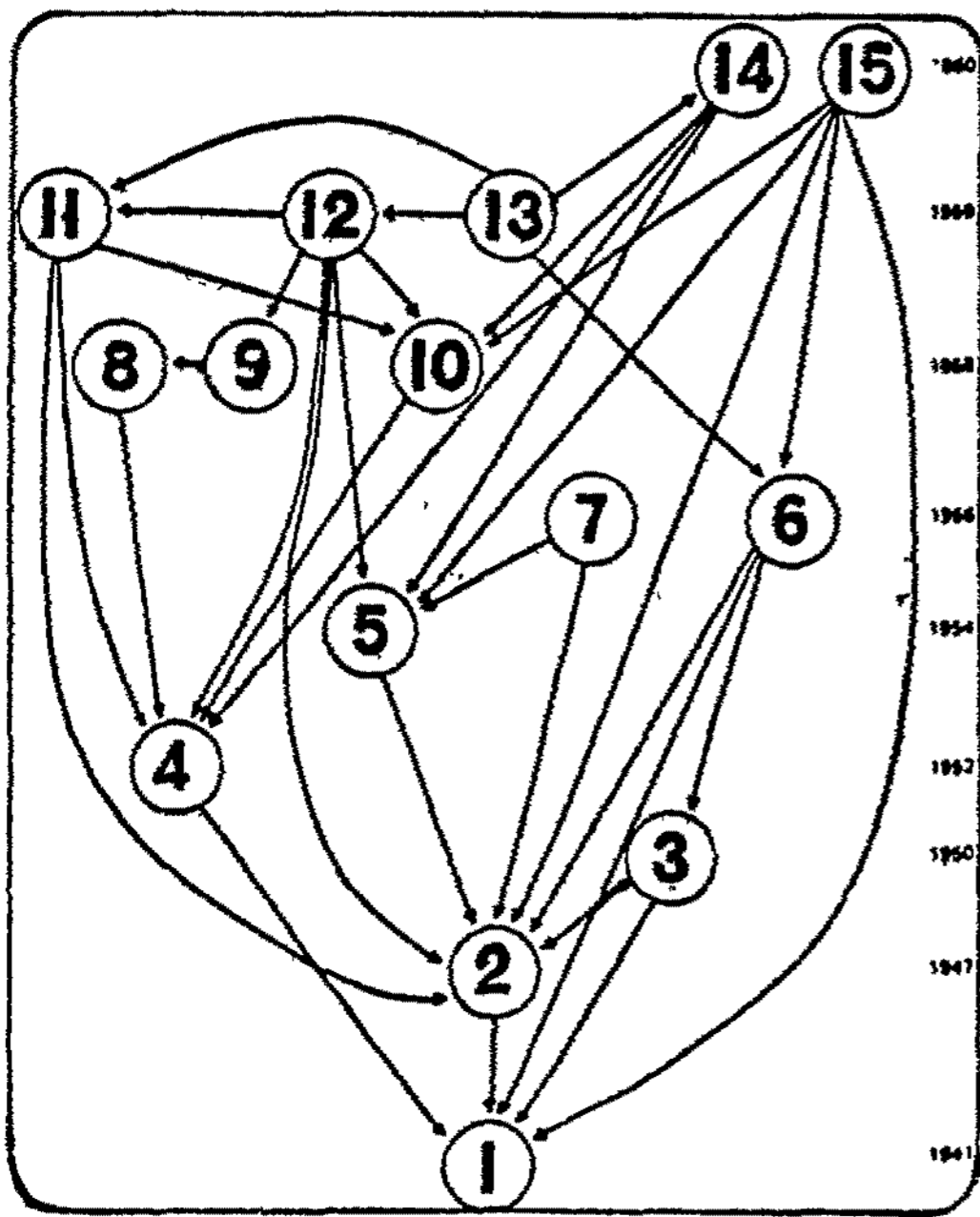


그림 5.

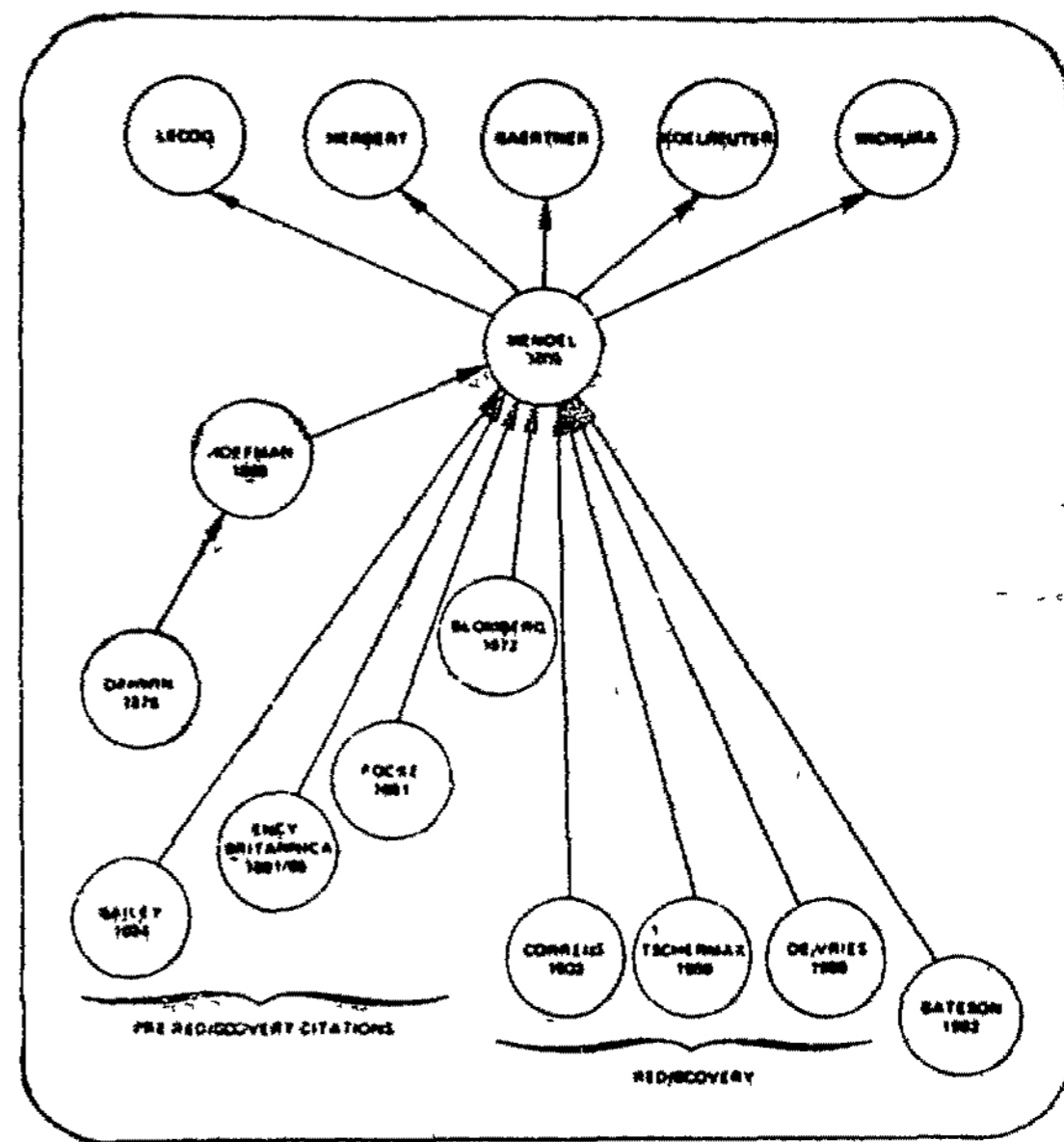


그림 6.

當時 科學界에서 無識하여 이를 理解하지 못하였던 것은 分明히 아니다. 적어도 4名 以上の 當時의 科學者가 “再發見”되었다고 한 1900年 以前에 멘델의 研究結果를 引用한 事實이 있다. 다윈(Darwin)은 1876年에 發表한 論文에 호프만(Hoffman)의 論文(1869)을 引用하였는데 멘델의 論文(1865)은 引用하지 않았다. 호프만은 멘델의 것을 다섯차례나 引用하였는데 다윈은 왜 멘델의 것을 引用하지 않았을까? 이것은 아

주 異常한 일이다. 이러한 異常한 點에서 歷史的으로 重要한 事實의 그 社會的 背景을 究明하는 데 큰 功용이 있을 것이라는 것은 確實하다.

引用索引의 다른 한 가지의 興味로운 應用例를 紹介 하겠다. 노벨賞의 受賞者를 이 技法으로서 豫測하여 주는 試圖로서 SCI의 1967年度 데이터를 바탕으로 引用計數가 가장 많은 著者의 順序대로 랭킹을 定하였다. 랭킹 50位까지의 統計를 보면 그림 7과 같다.

	RANK	AUTHOR	TOTAL TIMES CITED	RANK	AUTHOR	TOTAL TIMES CITED
	1	LOWRY OH	2921	26	ELIEL EL	721
	2	CHANCE B	1374	27	STREITWIESER A	717
	3	LANDAU LD	1174	28	MULLIKEN RS	712
	4	BROWN HC	1150	29	JACOB F	711
	5	PAULING L	1063	30	BORN M	710
Nobel Prize in Physics 1969	6	GELLMANN M	942	31	BRACHET J	706
	7	COTTON FA	940	32	WINSTEIN S	702
	8	POPLE JA	933	33	ALBERT A	687
	9	BELLAMY LJ	906	34	LUFT JH	674
	10	SNEDECOR GW	904	35	OEDUVE C	673
	11	BOYER PO	893	36	VONEULER US.	668
	12	BAKER BR	876	37	FIESER LF	666
	13	KOLTHOFF IM	853	38	HUISGEN R	661
	14	HERZBERG G	842	39	NOVIKOFF AB	655
	15	FISCHER F	826	40	GOODWIN TW	643
	16	SEITZ F	822	41	BARTON DHR	632
	17	OJERASSI C	801	42	FISHER RA	631
	18	BERGMEYER HU	754	43	BATES DR	627
	19	WEBER G	750	44	FLORY PJ	626
	20	REYNOLDS ES	748	45	STAHL E	626
	21	MOTT NF	741	46	DEWAR MJS	619
	22	ECCLES JC	737	47	GILMAN H	618
	23	FEIGL F	729	48	FOLCH J	618
	24	FREUD S	727	49	DISCHE Z	614
	25	PEARSE AGE	726	50	GLICK O	609

그림 7. 引用計數랭킹

이 그림에서 랭킹 6位の Gellmann, M.가 1969年度 노벨 物理學賞의 受賞者이고 랭킹 41位の Barton, D. H.R.이 亦是 同年度 노벨 化學賞의 受賞者였다. 이 데이터가 비록 斷片的인 것이었지만 受賞豫想者 50名 가운데서 2名이나 適中하였다는 것은 아주 特記할 만한 일이며, 나아가서 이러한 技法으로 다른 計量社會學的인 各種 豫測의 可能性을 보여 주고 있다. 그리고 學界, 產業界 및 政府機關에 從事하고 있는 많은 研究管理者에게 眞情한 科學技術的 背景을 가진 人材의 評價 手段의 道具가 될 수가 있다.

아직까지 引用네트워크는 手動的으로 만들고 있는데 별지 않아서 이 네트워크의 다이어그램도 電子計算機와 디스플레이를 通하여 自動적으로 얻어질 수가 있을 것이다. 그러면 歷史學者나 計量社會學者가 電子計算機 앞에 앉아서 人名, 單語, 引用文獻 또는 場所 등만의 몇 가지의 單語로서 質問하게 될 수 있을 것이며 電

子計算機는 論文이나 冊의 目錄을 알려 줄 뿐만 아니라 그 歷史的인 흐름을 圖示하여 주게 될 것이다.

이상과 같은 여러 가지의 可能性은 從來의 索引方法으로서는 不可能한 것이며, 多目的으로 利用이 可能한 引用索引技法이 널리 普及되기를 바란다.

參 考 文 獻

- (1) 柳京熙. 科學情報의 壽命測定에 관한 研究. 서울. 韓國 科學技術情報센터, 1971.
- (2) Garfield, E. Citation Indexing for Studying Science Nature 227 669~671, Aug., 1970.
- (3) Malin, M.V. The Science Citation Index: A New Concept in Indexing. Library Trends 17 374~387, Jan., 1968.
- (4) Price, D.J.S. Networks of Scientific Papers Science 149 510~515, Jul., 1965.