

# 인용문헌을 이용한 검색에 관한 연구

## A Study on Document Retrieval Using Bibliographic Citations

김 영 민\*

### 초 록

본 논문에서는 정보이용자가 특정정보를 요구하여 기존의 상업용 데이터베이스를 탐색할 때, 주제색인어를 이용한 전통적인 탐색방법의 결과에 만족하지 않고 더 많은 문헌을 탐색하고자 한다는 가정하에서, 문헌의 내용을 나타내는 주제색인어의 보완적 역할로서 인용문헌을 이용하여 검색효율을 높이는 실험을 함으로써, 현실적으로 이용가능한 새로운 탐색전략기법을 제시하였다.

### ABSTRACT

A user who retrieved relevant documents from the existing commercial databases may be not always satisfied with the results of the traditional bibliographic searches using the subject index terms. On the assumption that the user wants more relevant documents in such instances, this thesis presents an expanded search strategy by carrying out an experiment using bibliographic citations as another content indicator in addition to index terms.

### 1. 서 론

#### 1. 연구의 목적

과학 기술 분야의 발전상은 그 연구결과를 보고하는 문헌의 증가량에서 엿볼 수 있다. 과거

에는, 지수함수적으로 증가하는 문헌과 그 문헌에 포함된 특수한 정보를 요구하는 연구자나 실무자들의 사이를 연결시켜주는 역할을 초록지

\* 한국에너지연구소  
접수일자: 1985. 5.10.

나 색인지가 많아왔다. 오늘날에는, 이와같은 전통적인 서지도구들의 대부분이 온라인 방식을 통해 탐색이 가능하도록 데이터베이스화 되고 있다.

자동화된 문헌검색시스템의 탐색전략은 주로 주제색인어들에 논리관계함수(Boolean Function)를 적용시킨 것인데, 실제로 이러한 논리관계함수로 이루어진 검색방법이 항상 효과적이지 아니라는 연구가 발표된 바 있다.<sup>1)</sup> 온라인 탐색이 가능한 데이터베이스의 이용과 관련된 여러 문제점들중의 하나는, 문헌파일을 효과적으로 탐색하려면 정보요구와 관련된 색인어에 대한 적당한 지식을 가지고 있어야 한다는 것이다. 특히, 복합적이며 여러 주제 분야와 관련이 있는 학문분야의 정보를 요구하는 상황에서는, 전통적인 탐색방법만으로는 좋은 결과를 낼 수 없으므로, 새로운 탐색방법이 요구되고 있다.

샬튼(Salton)<sup>2)</sup>은 검색효율의 향상기법을 벡터에 의한 질문과 문헌의 유사도측정, 용어에 가중치를 부여하는 방법, 파일 클러스터링, 동적 질문조정방식, 인용문헌의 이용등 5가지로 구분하고 있다. 특히 인용문헌의 이용은, 자동화된 검색시스템에서보다는 오히려 연구자들이 관련 주제 분야의 적합문헌들에 포함되어 있는 참고문헌들을 추적함으로써, 인용문헌을 검색에 이용하는 경우가 더 많이 있다.

본 논문에서는 정보이용자가 특정 정보를 요구하여 기존의 상업용 데이터베이스를 탐색할 때, 주제색인어를 이용한 전통적인 탐색방법의 결과에 만족하지 않고 더 많은 문헌을 탐색하고자 한다는 가정하에서, 문헌의 내용을 나타내는 주제색인어의 보완적 역할로서 인용문헌을 이용하여 검색효율을 높이는 실험을 함으로써, 현실

적으로 이용가능한 새로운 탐색전략기법을 제시하고자 한다.

## 2. 연구의 범위 및 방법

본 논문에서는 문헌의 내용을 나타내는 주제색인어와 인용문헌을 함께 이용하여 검색효율, 특히 재현율을 높일 수 있다는 가설하에서 1차 검색은 주제색인어를 이용하여 검색한 후, 2차 검색은 1차검색 결과로부터 나온 적합문헌들의 인용문헌을 이용하는 실험을 하였다.

실험에서는 종합학문적 성격을 띤 생화학과 정보학 분야의 연구자들이 실제 연구에 필요한 문헌을 검색하기 위하여, 연구자들이 제시한 질문으로부터 검색어를 선정하여 8개의 검색문을 작성하였다. 이 검색문들로 각 해당분야의 상업용 데이터베이스인 CA Search와 LISA를 온라인 방식으로 검색하였다. 1차검색의 결과는 질문을 제시한 연구자들로 하여금 적합성 여부를 판별하게 하였고, 그 적합문헌들의 인용문헌집합을 구성하기 위해 인용색인어를 탐색하였다. 이 인용문헌집합에 서지결합법과 동시인용법을 응용시키고, 질문자들로 하여금 다시 적합성 여부를 판별하게 하였다.

- 1) J. Verhoeff, et al., "Inefficiency of the Use of Boolean Functions for Information Retrieval Systems," *Communication of the Association for Computing Machinery*, Vol. 4 (1961), pp. 557-9.
- 2) William Goffman, "A Searching Procedure for Information Retrieval," *Information Storage and Retrieval*, Vol. 2 (1964), pp. 73-8.
- 3) Gerard Salton, *Introduction to Modern Information Retrieval* (New York: McGraw-Hill, 1983), pp. 200-50.

## II. 이론적 배경

### 1. 인용문헌의 유용성

문헌들로부터 본내용문외에 얻어낼 수 있는 데이터가 그 문헌의 내용을 나타내는 중요한 지표로서 역할을 할 수 있는가의 여부는 문헌검색 분야의 관심을 끌어왔다. 특히, 특정저자가 자신의 참고문헌리스트에 소개한 문헌들을 문헌검색에 이용할 수 있는가의 여부를 고려해 볼 수 있다.

인용문헌들이 문헌의 내용을 대표할 때는, 몇가지의 제약이 따르는 데, 첫째는 문헌이 갖는 성격이다. 폭 넓게 주제를 다루는 문헌들, 예를 들어 이론적인 문헌이나 리뷰문헌들은 특정한 내용만을 다루는 논문들보다 더 빈번하게 인용되므로, 같은 주제를 다루더라도 그 문헌이 갖는 성격에 따라, 서로 다른 인용문헌을 포함하게 될 수 있다는 것이다.

둘째는, 인용문헌의 입수 용이도이다. 중요 학술잡지나 도서에 실린 논문들은 이용자가 입수하기 힘든 문헌들 보다 인용이 자주 되며, 출판년도에 따라 인용빈도율도 변화하게 된다.

셋째는, 저자들의 인용습관으로, 저자들이 갖는 다양한 지적 배경에 따라 문헌의 수와 형태가 달라진다. 연구자의 연구분야와의 적합성과는 관계없이 자기인용(Self-Citation)을 선호하는 경향이 있을 수도 있다.

또한 인용문헌을 포함하지 않는 서신, 의견발표, 회보등의 문헌인 경우에는 인용문헌만을 이용한 검색기법에서는 제외된다는 점도 있다.

이러한 문제점들로 인해 인용문헌은 문헌의 내용을 대표하는 역할을 갖지 못하고, 문헌들

간의 일반적인 경향과학과 문헌탐색에 있어서 보조적 역할만을 하여왔다.<sup>4)</sup> 또한 한 문헌이 그 주제 분야에서 차지하는 중요도를 인용색인지에 나타나는 인용빈도수와 동일시함으로써, 중요한 문헌들을 밝히고,<sup>5)</sup> 과학분야에서 비교적 중요한 영향력을 갖는 핵심 잡지군을 파악하는데 이용되어 왔다.<sup>6)</sup>

이와같이 인용문헌이 갖는 제약에도 불구하고 문헌의 내용을 나타내는 주제색인어 대신 인용문헌을 정보검색에 이용한 실험결과들이 계속 발표되었다. 내용지표로서 인용문헌이 갖는 유용성은 다음과 같다.

첫째, 저자들이 인용한 참고문헌리스트들중에는 저자의 연구내용과 관계가 깊은 문헌들이 포함되어 있을 것이 분명하며,

둘째, 많은 분야에서 연구와 개발을 담당하고 있는 전문연구자들이 중요문헌의 인용문헌을 추적하기 위해 인용색인을 이용하고 있는데, 이런 경우의 참고문헌들은 사실상 내용지표로서 이용되고 있다고 볼 수 있다.

셋째, 인용문헌은 저자와 이용자사이의 매개 역할을 하는 색인작성자의 판단보다 오히려 저자의 판단에 의한 것이므로, 인용문헌을 이용함으로써 이용자는 자신이 관심을 가지는 주제 분야의 원자료에 더 가까워진다고 볼 수 있다.

4) Eugene Garfield, "Citation Indexes for Science," *Science*, Vol. 122 (1955), pp. 109-11.

5) S. Dutta and T.S. Rajagopalan, "Literature Citations in Scientific and Technical Periodicals-a Survey," *Journal of Scientific and Industrial Research*, Vol. 17A (1958), pp. 259-61.

6) J.H. Westbrook, "Identifying Significant Research," *Science*, Vol. 132 (1960), pp. 1229-34.

7) M.M. Kessler, "Technical Information Flow Patterns," *Proceedings of Western Joint Computer Conference*, 1961, pp. 247-57.

이외에도 인용문헌을 통하여 문헌검색을 할 때는, 한 문헌내에 인용문헌들이 이미 존재함으로써 색인작성에 드는 부차적인 시간과 비용이 절약되고, 인용문헌들은 거의 자동적으로 쉽게 조작할 수 있으므로, 자동화된 시스템에 비교적 간단하게 입력시킬 수 있다는 유용성을 가지고 있다.

## 2. 문헌들간의 관계 설정 기법

전통적인 문헌검색의 자동화시스템에서는 논리관계 탐색 질문식과 도처파일시스템이 이용되고 있다. 이러한 검색모델에서는 각 문헌과 질문사이의 관계만이 잘 설정되면 효율적인 검색결과를 가져올 수 있으며, 문헌파일내의 문헌들은 서로 독립되어 있다는 가정을 근거로 하고 있다.

고프만(Goffman)<sup>8)</sup>은 문헌파일내의 문헌들은 서로 독립적이 아니며, 실제로 문헌들간의 상호 관련성을 나타내는 탐색전략이 전통적인 검색모델보다 더 우수하다고 하였다. 이와같이, 한 문헌파일내에서 문헌들이 상호 관련성을 갖는다는 개념은 수년동안 거론되어 왔다.

매론(Maron)과 쿤(Khuns)<sup>9)</sup>은 각 문헌마다 통계적인 추론기법을 적용하여 탐색질문을 만족시킬 가능성이 있는 적합성을 수자로 나타내는 확률 색인법(Probabilistic Indexing)을 제시하였고, 도일(Doyle)<sup>10)</sup>은 파일내의 문헌들이 갖는 주제개념들을 연관 도표(Association Map)로 나타냄으로써, 이용자로 하여금 파일내에 나타난 개념들의 도표를 보고 자신의 요구에 맞는 특정 방향을 선택하도록 하였다.

이러한 개념들과는 달리 색인어에 의존하지 않고 별개의 검색문헌 파일을 구성하는 제안들 중의 하나로, 문헌들사이의 관계를 측정하는 데

내용지표로서 저자, 잡지, 저자와 잡지의 혼용을 제시한 것을 들 수 있다.<sup>11)</sup> 이러한 문헌들 사이의 관계 측정방법은 그 주제 분야의 저자와 잡지에 대한 충분한 사전 지식을 필요로 하기 때문에 실제로 응용되기에는 어려운 점이 있다.

다른 제안으로는, 한 주제에 관한 문헌들의 네트워크를 형성하기 위해 문헌내에 저자가 수록한 인용문헌들을 이용하여 문헌들을 서로 연결시켜 주는 것을 들 수 있다.

인용문헌을 이용해 문헌들간의 관계를 설정해 주는 기법으로는 서지결합법(Bibliographic Coupling)과 동시인용법(Co-citation)이 있다.

### 2.1 서지결합법

서지결합법은 미국 메사추세츠공과대학의 케슬러(Kessler)<sup>12)</sup>가 소개한 기법으로, "여러 편의 과학문헌들이 한 개 이상의 참고문헌을 공통으로 포함하고 있다면 이 문헌들은 각기 서로에 대하여 의미있는 관계를 가지게 된다. 즉, 서지적으로 결합된다"고 가정하였다. 즉, 다음과 같은 2개의 기준으로 나누어 설명하고 있다.

- 8) William Goffman, "An Indirect Method of Information Retrieval," *Information Storage and Retrieval*, Vol. 4 (1968), pp. 361-73.
- 9) M.E. Maron and J.L. Khuns, "On Relevance, Probabilistic Indexing and Information Retrieval," *Journal of the Association for Computing Machinery*, Vol. 7 (1960), pp. 216-44.
- 10) L.B. Doyle, "Semantic Road Maps for Literature Searches," *Journal of the Association for Computing Machinery*, Vol. 8 (1961), pp. 553-78.
- 11) D.B. Cleveland, "An n-Dimensional Retrieval Model," *JASIS*, Vol. 27 (1976), pp. 342-6.
- 12) M.M. Kessler, "An Experimental Study of Bibliographic Coupling Between Technical Papers," *IEEE Transactions on Information Theory*, Vol. 9 (1963), pp. 49-51.

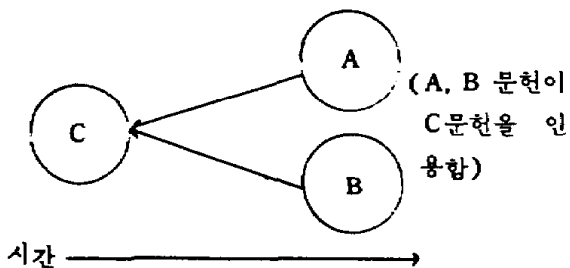
기준 A : 다수의 문헌들이 어떤 특정한 실험 문헌  $P_0$  와 적어도 1개 이상의 결합단위(Coupling Unit)를 가지고 있다면, 이들은 서로 관련있는 그룹  $G_A$  를 이루게 된다.  $P_0$  와  $G_A$  의 구성요소들간의 결합강도는 이들 사이의 결합단위( $n$ )로 측정된다.

기준 B : 적어도 1개 이상의 결합단위를 가지는 문헌들은  $G_B$  라는 관련된 그룹을 이룬다.

이에 대한 조작상의 정의를 클레버슨(Cleverdon)<sup>13)</sup>은 "서지결합법은 특정문헌  $q$  를 인용한 문헌들의 전부 또는 일부를 이용하거나, 또는  $q$  와 어떤 일정 갯수만큼의 인용문헌을 공통으로 갖고 있는 특정 집단내의 모든 문헌을 이용함으로써,  $q$  의 주제 범위인 항목  $X$  를 확장하는 수단"이라 내리고 있다.

서지결합법의 개념은 그림 1에 나타나 있다.

[그림 1] 서지결합법 형태



## 2.2 서지결합법을 이용한 연구

케슬러(Kessler)는 서지결합법을 입증하기 위해 일련의 실험들을 진행하였는데, 첫번째 실험은<sup>14)</sup> 「Proceedings of the Institute of Radio Engineers」에 수록된 40개 논문들을 한 그룹으로 하여 수작업 처리한 결과, 문헌의 성격이 이론적이거나 리뷰적인 문헌들이 서지적으로 강하게 결합되어 있음을 발견하였다.

같은 해에, 컴퓨터 프로그램을 이용하여「Physical Review」 36권들에 수록된 전체 137,000개의 인용문헌을 처리한 실험<sup>15)</sup> 결과, 다음과 같은 결론을 내렸다.

- 1) 서지결합법은 수자라는 측면에서 처리되므로, 용어와 언어에 구애받지 않으며,
- 2) 어떠한 전문적인 지식도 요구하지 않고,
- 3) 한 그룹내의 문헌들을 과거뿐만 아니라 미래에까지 확장시킬 수 있으며,
- 4) 특정한 문헌을 어떤 고정된 분류항목에 넣지 않고, 문헌의 인용 상황에 따라 변화시킬 수 있고,
- 5) 특정한 문헌과 결합단위를 공유하는 문헌들은 그 문헌과 '논리적으로 관련있는 참고 문헌'으로 고려될 수 있다.

또한 계속하여 케슬러(Kessler)<sup>16)</sup>는 「Physical Review」의 35권들에 수록된 8,186개 논문으로 구성된 데이터베이스를 서지결합법을 이용하여 탐색한 결과, 검색될 문헌에 시간상의 제약을 둠으로써 문헌의 수를 통제하는 것이 가능함을 밝혀냈다. 또한, 미국문헌과 서지적으로 강하게 결합되어 있는 중국문헌은 그 분야 과학자들에게 번역하여 제공해야 하므로, 서지결합법이 선택적 번역서비스에 이용될 수 있음을 시사하였다.

13) B.H. Weinberg, "Bibliographic Coupling, a Review," *Information Storage and Retrieval*, Vol. 10 (1974), pp. 189-96.

14) M.M. Kessler, "An Experimental ...," pp. 49-51.

15) M.M. Kessler, "Bibliographic Coupling Between Scientific Paper," *American Documentation*, Vol. 14 (1963), pp. 10-25.

16) M.M. Kessler, "Bibliographic Coupling Extended in Time: Ten Case Histories," *Information Storage and Retrieval*, Vol. 1 (1963), pp. 169-87.

다음 연구는 주제분석 색인법과 서지결합법으로 검색한 결과를 비교한 실험<sup>17)</sup>으로, 검색된 문헌집합은 서로 아주 높은 관련성을 가지고 있음을 증명하고, 서지결합법이 곧 주제검색의 한 방법이 될 수 있다고 제안하였다.

클레버든(Cleverdon)<sup>18)</sup>은 서지결합법을 인용색인의 정확률 향상수단으로 고려하였는데, 공통된 인용문헌의 수가 증가할수록 분류의 범위가 좁아질 수 있기 때문이다. 케슬러(Kessler)가 단순결합만을 제안한 반면, 클레버든(Cleverdon)은 문헌의 전체 참고문헌수와 서지적으로 결합된 참고문헌수의 비율을 계산하는 비례적 결합을 제안하였는데, 이 결합은 각 문헌에 가중치를 부여하는 성격을 띠게 된다.

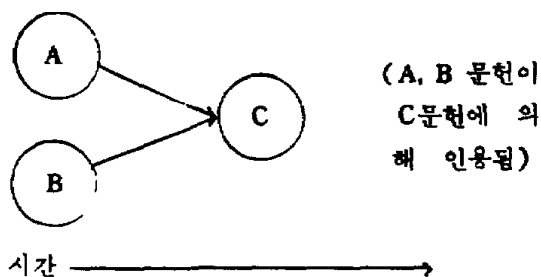
스몰(Small)<sup>19)</sup>은 서지결합법을 이용하여 890개의 잡지를 168개의 잡지군으로 나누고, 이 결과를 수작업으로 잡지군을 분류했을 때와 비교해 보고 잡지의 분류에 이 기법이 실용성을 갖는다고 제안하였다.

이상과 같이 서지결합법은 주로 주제검색의 한 기법으로 발전을 해왔으며, 잡지군의 분류에도 응용이 되어 왔다.

### 2.3 동시인용법

케슬러(Kessler)가 서지결합법을 소개한 지 십년 뒤인 1973년에, 스몰(Small)이 동시인용법을 발전시켰는데, 이 개념<sup>20)</sup>은 두 문헌이

[그림 2] 동시인용법 형태



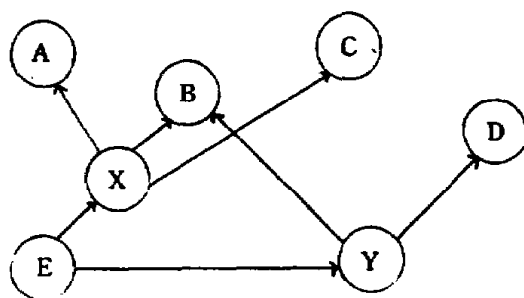
공통으로 함께 제 3자의 문헌에 인용되는 것을 의미하는데, 그림 2와 같이 나타낼 수 있다.

그 결합강도는 역시 두 문헌이 함께 나타나는 횟수에 의해 결정된다.

동시인용법은 두개의 원문헌이 나중 문헌에 의해 인용되는 것이므로, 강하게 동시인용되기 위해서는 많은 수의 저자들이 두개의 원문헌을 인용하여야 한다. 동시인용법은 문헌을 인용한 저자들에 의해 수립되는 관계로써, 동시인용형 태는 서지결합형태와 다르게 나타난다.

서지결합법과 동시인용법에 의한 문헌관계를 그림 3을 통해 살펴볼 수 있다.

[그림 3] 문헌들간의 인용링크



7개 문헌들간의 인용링크를 보면, 문헌 X와 Y는 각기 B, C문헌을 인용하고 있으므로, 두 문헌간의 서지결합단위는 2개이다. 서지결합법

17) M.M. Kessler, "Comparison of the Results of Bibliographic Coupling and Analytic Subject Indexing," *American Documentation*, Vol. 16 (1965), pp. 223-33.

18) B.H. Weinberg, "Bibliographic ...," pp. 189-96.

19) H.G. Small, "Journal Clustering Using a Bibliographic Coupling Method," *Information Processing and Management*, Vol. 13 (1977), pp. 277-88.

20) H. Small, "Co-citation in the Scientific Literature: a New Measure of the Relationship Between Two Documents," *JASIS*, Vol. 24 (1973), pp. 265-9.

에서는 X문헌과 Y문헌이 E문헌에 의해 인용된 것은 측정되지 않는다. X문헌과 Y문헌의 동시인용강도는 1개이다. 하나의 문헌만이 X, Y문헌을 인용하고 있기 때문이다.

결국, 동시인용법은 문헌이 발표된 후, 함께 인용되는 초기문헌들을 연결시키는 것이고, 서지결합법은 초기의 문헌들을 인용하는 나중 문헌들을 연결시키는 것이다. 이 차이는 서지결합법이 문헌의 내부적 관계에 근거한 정적 연결이고, 동시인용법은 문헌들의 외부적 관계에 근거한 동적 연결이라 볼 수 있다. 시간이 지남에 따라 동시인용강도는 계속 증가하나, 서지결합 강도는 일정하다.

#### 2.4 동시인용법을 이용한 연구

동시인용법의 발전은 컴퓨터를 이용해 대규모의 인용문헌을 처리하는 인용색인지의 발전에 힘입은 바 크다. 오늘날, 동시인용법은 동시인용망(Co-citation Network)을 통해 저자간의 상호관련도를 밝히고, 주제구조의 골격 및 개요를 추출함으로써, 특정분야의 파라다임(Paradigm)을 규명하는데까지 사용되고 있다.

실제로 스몰(Small)<sup>21)</sup>은 동시인용망을 개념구조로 바꾸어 개념간의 관계를 파악함으로써, 인용분석을 통해 특정 주제분야의 파라다임을 규명하는 방법을 시도하였다.

화이트(White)와 그리피드(Griffith)<sup>22)</sup>는 동시인용법을 이용해 저자간의 관계를 도식화함으로써 정보학내에서 각 하위주제별 저자군 및 저자의 위치를 규명하였다.

최근에는 렌크(Lenk)<sup>23)</sup>가 동시지명(Co-nomination)을 통한 분석결과와의 유사성을 발견하고 저자의 인용행태가 과학의 사회적 구조와 교도로 상응함을 밝혔다.

### 3. 인용문헌의 이용

#### 3.1 인용문헌에 의한 문헌검색

인용문헌을 이용하여 파일내의 문헌들간의 상호관계를 결정한다는 것은 논리관계 함수와 도치파일에 근거한 전통적인 검색방법에 영향을 미치게 됨을 의미하므로, 문헌의 내용을 나타내는 주제색인어와 인용문헌사이에 어떠한 관계가 있는가를 밝히고자 하는 연구들이 진행되었다.

케슬러(Kessler)<sup>24)</sup>는 주제분석색인어와 서지결합법으로 자기 검색한 결과, 서지적으로 결합된 인용문헌들이 색인어에 의해 검색된 문헌들을 대신할 수 있다고 제시하였다.

셀튼(Salton)<sup>25)</sup>은 인용문헌의 제목에서 얻어낸 새로운 단어들로 연관문헌검색기법(Associative Document Retrieval Technique)을 확장시키는 실험에서, 인용문헌들이 특정문헌 집단에 들어있지 않는 적합한 색인어를 제공하게 되므로 보통의 색인어만을 이용한 검색결과보다 더 융통성있는 검색결과를 나타낸다고 밝혔다.

21) H. Small, "Co-citation Context Analysis and the Structure of Paradigms," *Journal of Documentation*, Vol. 36 (1980), pp. 183-96.

22) H.D. White and B.C. Griffith, "Author Co-citation: a Literature Measure of Intellectual Structure," *JASIS*, Vol. 32 (1981), pp. 163-71.

23) Peter Lenk, "Mapping of Fields Based on Nominations," *JASIS*, Vol. 34 (1983), pp. 115-22.

24) M.M. Kessler, "Comparison ...," pp. 223-33.

25) Gerard Salton, "Associative Document Retrieval Techniques Using Bibliographic Information," *Journal of the Association for Computing Machinery*, Vol. 10 (1963), pp. 440-57.

그레이(Gray)와 할리(Harley)<sup>26)</sup>는 한 저자가 인용한 참고문헌들의 내용을 색인어로 나타낸 결과, 저자가 집필한 문헌의 주제와 관계가 있음을 밝혀냈다. 또한, 인용문헌의 내용을 나타낸 색인어에 컴퓨터를 이용하여 가중치를 부여하는 방법을 제시하고, 새로운 문헌에 색인어를 부여할 때 인용문헌이 색인작성자에게 도움이 될 것이라고 제안하였다.

마이클슨(Michelson)<sup>27)</sup>의 연구는 인용문헌을 초기검색에서보다 오히려 탐색과정중의 피이드백 정보로서 이용한 실험이었다. 원문헌 벡터를 그 문헌의 저자명, 그 문헌을 인용한 문헌들, 그 문헌이 인용한 문헌들로 구성하여 실험한 결과, 검색효율중 정확률과 재현율이 향상되었음을 밝히고, 인용문헌이 검색초기의 탐색에도 효과가 있을 것이라 하였다.

살톤(Salton)<sup>28)</sup>은 SMART 시스템을 통해서 문헌과 질문벡터를 확장하기 위하여 인용문헌을 이용하여 실험한 결과, 주제를 나타내는 색인어들만으로 구성된 표준벡터보다 검색효과가 더 좋음을 밝히고, 문헌의 내용을 인용문헌들로 표현하였을 때, 낮은 재현율과 높은 정확률의 결과를 가져온다고 하였다.

오커너(O'Connor)는 두 번의 실험<sup>29)30)</sup>으로 본문중 인용문헌이 있음을 나타내 주는 문장에서 추출해 낸 단어들을 이용하여 검색했을 때, 20% 정도로 재현율을 더 높일 수 있고 인용문헌이 있음을 나타내 주는 문장에서 추출한 단어와 인용문헌의 서명에서 추출한 단어를 이용하여 검색한 결과와 주제색인어와 초록에 나타난 단어를 이용하여 검색한 결과를 비교한 결과 모두 50%의 재현율을 나타내었다고 하였다.

이상의 실험들의 결과로 보아 인용문헌이 문

헌의 내용을 나타내는 주제색인어를 대신할 수 있으며, 적어도 주제색인어의 보조역할을 할 수 있음을 알 수 있다.

### 3.2 인용문헌에 의한 문헌자동분류

인용문헌은 용어와 언어에 관계없이 파일내에 들어있는 유사한 문헌들을 찾아내는 문헌의 자동분류실험에 주로 이용되어 왔다.

아이비(Ivie)<sup>31)</sup>는 문헌들 사이에서 공통으로 인용된 문헌들에 기초를 두고 데이터베이스를 여러 개의 문헌클러스터로 나누었다. 이용자와 시스템간의 상호작용을 통해 질문을 대표하는 문헌들과 데이터베이스내의 문헌들간의 유사도를 측정하여, 적합한 클러스터를 선정하여 검색하도록 하였다.

가장 성공적으로 인용문헌을 이용한 실험은

- 
- 26) W.A. Gray and A.J. Harley, "Computer Assisted Indexing," *Information Storage and Retrieval*, Vol. 7 (1971), pp. 167-74.
  - 27) D. Michelson, "An Experiment in the Use of Bibliographic Data as a Source of Relevance Feedback in Information Retrieval," in G. Salton, *The SMART Retrieval System: Experiments in Automatic Document Processing* (New York: Prentice-Hall, 1971), pp. 430-43.
  - 28) Gerard Salton, "Automatic Indexing Using Bibliographic Citations," *Journal of Documentation*, Vol. 27 (1971), pp. 98-110.
  - 29) J. O'Connor, "Citing Statement: Computer Recognition and Use to Improve Retrieval," *Information Processing and Management*, Vol. 18 (1982), pp. 125-31.
  - 30) J. O'Connor, "Biomedical Citing Statement: Computer Recognition and Use to Aid Full-Text Retrieval," *Information Processing and Management*, Vol. 19 (1983), pp. 361-8.
  - 31) J. Bichteler and R.G. Parsons, "Document Retrieval by Means of an Automatic Classification Algorithm for Citations," *Information Storage and Retrieval*, Vol. 10 (1974), pp. 267-78.



프라이스(Price)<sup>32)</sup>와 쉬미노비치(Schiminovich)<sup>33)</sup>의 연구들로서, 인용문헌을 이용해 자동분류표를 만들어 낸 결과, 프라이스(Price)는 수작업보다 더 좋은 분류시스템을 만들어 내었고, 쉬미노비치(Schiminovich)는 94%의 재현율과 93%의 정확률을 얻어내었다고 하였다.

코크(Kwok)<sup>34)</sup>는 문헌의 서명과 그 문헌이 인용한 문헌들의 서명에서 빈도수에 의해 단어들을 추출하여 분류한 결과, 문헌의 내용을 나타내는 직설적인 단어들을 집중시켜 모을 수 있고, 인용문헌을 이용한 자동분류와 유사한 결과를 나타내며 유사한 문헌그룹들이 있을 때, 이 그룹들을 구분해 내기가 용이하다고 하였다.

비히텔러(Bichteler)와 파슨즈(Parsons)<sup>35)</sup>는 연구자들이 그들의 정보요구에 적합하다고 제시한 문헌과 그 문헌의 인용문헌들로 탐색질문을 나타내는 실험을 한 결과, 한 가지 주제 분류체계에 기반을 둔 검색절차보다 비교적 더 낮은 재현율과 높은 정확률을 나타낸다고 하였다.

후에 비히텔러(Bichteler)와 이이튼(Eaton)<sup>36)</sup>은 앞의 실험에서 이용했던 데이터를 이용하고, 물리학자들이 제공한 문헌리스트를 준문헌의 인용리스트로 사용하여 실험집단내의 각 문헌들과 준문헌 사이의 유사도를 계산해내는 데 서지결합법을 이용한 간단한 알고리즘으로 처리하였다. 계속하여 이들은<sup>37)</sup> 서지결합법과 동시인용법을 함께 이용한 유사도 측정방법이 서지결합법만을 이용한 문헌검색보다 더 향상된 결과를 나타낸다고 하였다.

인용문헌을 이용한 문헌의 자동분류의 유사도 측정방법은 2가지로 볼 수 있는데, 하나는 암슬러(Amsler)<sup>38)</sup>가 제안한 것으로, 서지결합법과 동시인용법을 모두 이용하였다. 즉, 문헌  $i$

는 두개의 집합으로 나타낼 수 있는데,

$A_i$  :  $i$ 가 인용한 문헌들의 집합

$D_i$  :  $i$ 를 인용한 문헌들의 집합

으로, 문헌  $i$ 의 전체 문헌집합은  $|A_i \cup D_i| = L_i$ 이고, 문헌  $i$ 와  $j$ 의 주제유사도는  $(L_i \cap L_j) / (L_i \cup L_j)$ 로 측정된다.

다른 하나는 쉬미노비치(Schiminovich)<sup>39)</sup>가 자동분류표를 만들기 위해 사용한 인용문헌간의 서지적 패턴 발견공식(Bibliographic Pattern Discovery Algorithm)으로, 서지결합법을 이용한 것이다.

### 3.3 이동상의 문제점

인용문헌을 이용한 검색시스템이 몇 가지 제

32) N. Price and S. Schiminovich, "A Clustering Experiment: First Step Towards a Computer-generated Classification Scheme," *Information Storage and Retrieval*, Vol. 4 (1968), pp. 271-80.

33) S. Schiminovich, "Automatic Classification and Retrieval of Document by Means of a Bibliographic Pattern Discovery Algorithm," *Information Storage and Retrieval*, Vol. 6 (1971), pp. 417-35.

34) K.L. Kwok, "The Use of Title and Cited Titles as Document Representation for Automatic Classification," *Information Processing and Management*, Vol. 11 (1975), pp. 201-6.

35) J. Bichteler and R.G. Parsons, "Document ...," pp. 267-78.

36) J. Bichteler and E. Eaton, "Comparing Two Algorithms for Document Retrieval Using Citation Links," *JASIS*, Vol. 28 (1977), pp. 192-5.

37) J. Bichteler and E. Eaton, "The Combined Use of Bibliographic Coupling and Cocitation for Document Retrieval," *JASIS*, Vol. 31 (1981), pp. 178-82.

38) Robert Amsler, Application<sup>2</sup> of Citation Based Automatic Classification, *International Technical Report 72-12* (Austin: The University of Texas, 1972), p. 42.

39) S. Schiminovich, "Automatic ...," pp. 417-35.

시가 되고 있는데, 여미쉬(Yermish)<sup>40)</sup>는 서지레코드에 들어있는 인용문헌들 간의 관계를 효과적으로 조작할 수 있도록 하고, 시스템과 이용자간의 상호작용이 가능한 정보검색시스템을 설계하였다. 각각의 문헌레코드에는 REFLIST(특정문헌이 인용한 문헌리스트)와 CITELIST(특정문헌이 인용된 문헌리스트)가 포함되어 있고, 전통적인 색인어 탐색과 인용문헌들을 결합시킨 탐색방식을 이용할 수 있다.

최근에는 사용가능한 상업적 검색시스템과 인용색인 데이터베이스를 이용하여, 특정주제에 관한 적합문헌을 탐색하기 위한 전략으로서 동시인용법을 이용할 수 있다고 제시되고 있다.<sup>41)</sup>

가필드(Garfield)<sup>42)</sup>는 동시인용법을 이용한 클러스터링을 통해 설정된 연구전선의 특정성에 기반을 둔 검색시스템인 BIOMED SEARCH의 시험테스트를 보고하고 있다.

인용문헌을 이용한 검색시스템이 제시되고 있지만, 실제로 상업적인 온라인 방식에 의한 검색시스템에서 인용문헌을 광범위하게 이용할 수 없는 이유는 다음과 같다.

첫째, 앞에서 기술한 바와 같이 인용문헌이 갖는 제약들로 인해, 주제색인어를 대신하여 검색을 할만큼 내용지표로서 완벽하지 못하므로 색인어의 보조역할을 목적으로 인용문헌을 이용하려면, 색인작성시 드는 비용외에 인용문헌을 입력시키는데 상당한 비용이 추가되며,

둘째, 검색시스템에서 다루고 있는 주제분야의 문헌들의 인용문헌을 모두 입력시켜야만 검색효과가 나아지게 되므로, 기억용량이 큰 대규모 컴퓨터에서만 이용이 가능하며,<sup>43)</sup>

셋째, 인용문헌들간의 관계를 설정하는 기법들이 충분히 실험되어 오지 않았으며, 각 학문분야마다 결합강도의 임계적 기준치(Critical

Threshold Value)가 결정되어야 한다.<sup>44)</sup>

그러나 무엇보다도 인용문헌을 처리할 수 있도록 특별히 고안된 시스템이 부족하다는 점이 큰 장애가 되고 있다.

### Ⅲ. 인용문헌을 이용한 검색실험

#### 1. 실험의 목적

온라인 방식에 의한 서지정보의 탐색은 주제별 서지의 작성, 소급탐색(RS:Retrospective Searching), 최신정보주지탐색(SDI Search: Selective Dissemination of Information) 등의 여러가지 목적으로 이용될 수 있다.

이와같이 다양하게 이용될 수 있는 온라인 방식에 의한 서지정보의 탐색에 있어서, 검색결과가 만족스럽게 나타나지 않을 때가 있다. 특히, 여러 학문분야와 관련이 있는 주제나, 새롭게 부상하고 있는 주제의 검색에 있어서는, 관련된 학문분야가 검색시스템내에 포함되어 있지 않을 수도 있고, 아직 학문이 성숙되지 않아서 용어의 설정이 완전하지 못해 적당한 색인어가 준

40) L.C. Smith, "Citation Analysis," *Library Trends*, Vol. 29 (1981), pp. 83-106.

41) Howard D. White, "Cocited Author Retrieval Online: An Experiment with the Social Indicators Literatures," *JASIS*, Vol. 32 (1982), pp. 16-21.

42) E. Garfield, "ISI's On-line System Makes Searching So Easy Even a Scientist Can Do It," *Current Contents*, Vol. 13 (1981), pp. 5-8.

43) L.N. Rolling, "A Computer-aided Information Service for Nuclear Science and Technology," *Journal of Documentation*, Vol. 22 (1966), pp. 93-115.

44) J.A. Virgo, "The Review Article: Its Characteristic Problems," *Library Quarterly*, Vol. 4 (1971), pp. 275-91.

재하지 않음으로 인해, 검색결과가 아주 좋지 않을 수가 있다. 그러나 이와같은 분야가 온라인 방식에 의한 서지정보의 탐색을 적극적으로 이용하고자 하며, 또한 활발히 연구가 진행되고 있는 분야라는데 문제가 있다.

본 실험의 목적은 특히 이러한 다학문적인 주제분야에 인용문헌을 이용한 검색을 행하였을 때, 검색효율이 향상되는 지를 알아보고자 하는 것이다. 또한 이 실험을 위한 실험집단을 따로 만들지 않고, 기존의 상업용 온라인 데이터베이스를 이용하여 검색함으로써, 현실적으로 이용가능한 탐색전략을 제시하고자 하였다.

## 2. 실험의 설계

### 2.1 실험대상 분야의 선정

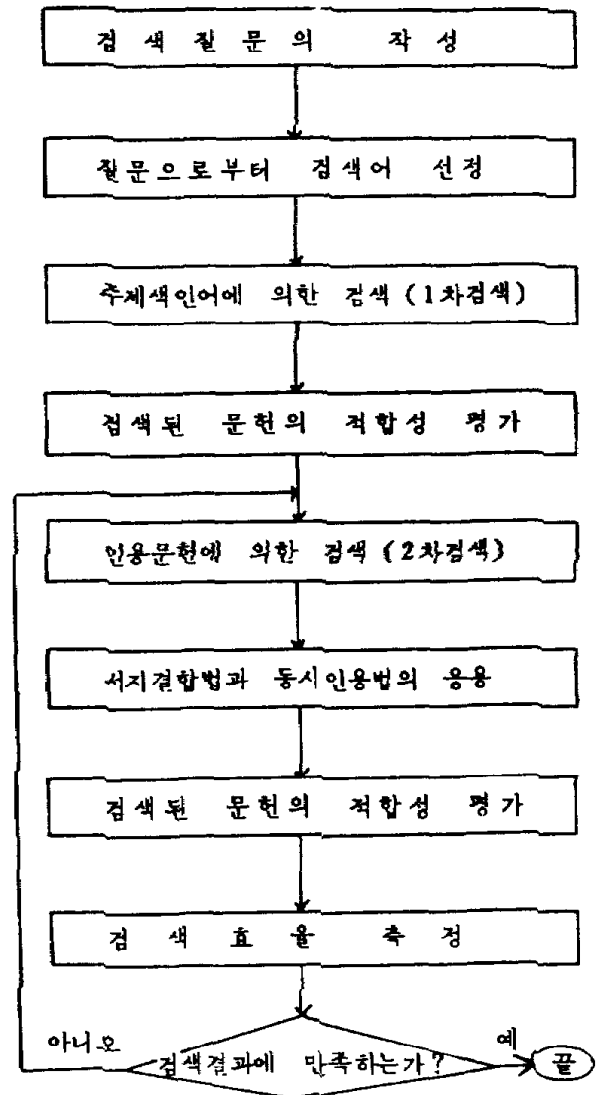
본 실험에서는 인용문헌을 이용한 검색결과, 재현율이 높아질 것으로 예상되는 종합학문적 성격을 띤 생화학과 정보학분야를 검색질문의 실험대상 분야로 선정하였으며, 선정이유는 다음과 같다. 즉, 최근 생화학 분야의 발전 속도는 눈부신데, 특히 DNA의 배열순서 결정과 단백질 합성의 상호관계, 서로 다른 종끼리의 유전자 융합을 이용한 유전공학의 발달, 세포내에서의 대사조절, 세포막에 대한 연구가 가장 활발한 연구대상이 되고있다. 학문의 연구결과를 수록하는 속도가 비교적 늦은 출판물중의 하나로 교과서를 들 수 있는데, 거의 매 4년마다 개정되어 나오는 외국의 생화학 교과서의 내용이 전판에 비해 약 절반이 새로와 진다는 사실<sup>45)</sup>만으로도 그 발전추세를 알 수 있다.

정보학 분야도 새롭게 부상하는 학문의 하나로, 전산학, 통계학, 수학등의 학문과 깊은 관계를 가지고 있는 종합학문적 성격을 띤 분야라 할 수 있기 때문이다.

### 2.2 실험방법

전반적인 실험과정을 순서도로 나타내면 그림 4와 같다.

[그림 4] 실험과정의 순서도



#### 1) 검색질문의 작성

새롭게 부상하고 있고, 복합적 학문성격을 띤 생화학과 정보학 분야의 연구자들이 그들이 관

45) 송정석, 「생화학」(서울:연세대학교 생화학교실, 1982), p. ii.

심을 가지고 있는 분야중 비교적 적합문헌을 찾기 힘들었던 주제로 검색질문을 만들도록 하였다.

생화학 분야의 질문자들은 생화학 박사학위를 취득한 교수 2명과 석사과정의 대학원생 1명으로 구성하였으며, 정보학 분야의 질문은 본 연구자가 직접 작성하였다. 질문의 수는 생화학 분야에서 6개, 정보학 분야에서 2개로 모두 8개이다.

질문을 구성하는 검색어의 선정은 질문자들 자신이 직접 행하게 하였고, 질문자 4명중 3명은 자동화된 문헌검색시스템을 사용해본 경험이 있었다.

2) 주제색인에 의한 검색(1차검색)

질문자들이 질문으로부터 작성한 검색문들로 기존의 상업용 데이터베이스를 온라인으로 탐색하였다.

생화학 분야의 데이터베이스로는 Chemical Abstracts Service 사의 CA Search를 탐색하였는데, CA Search는 과학, 기술분야의 데이터베이스중 가장 많은 축적정보건수를 가지고 있으며, 광범위한 주제를 망라하고 있다.

정보학 분야의 데이터베이스로는 Learned Information Ltd. 의 LISA(Library and Information Science Abstracts)를 탐색하였는데, 이 데이터베이스의 축적정보건수는 '84년 10월 현재 43,000건에 달하고 있다.<sup>46)</sup>

생화학 분야의 검색은 질문자중의 1명과 약학분야를 전공한 탐색전문가와의 간단한 면담을 행한 후, 실시하였다.

3) 인용문헌에 의한 검색(2차검색)

상업용 데이터베이스를 이용한 1차검색 결과로 나온 문헌리스트를 질문자에게 제공하여, 그들의 질문에 적합한 문헌들을 가려내게 하였

다. 여기에서 적합문헌으로 평가된 문헌들을 이용해 2차검색을 실시하였다.

적합문헌들은 자기 2차검색을 위한 하나의 도입문헌(entry document)의 역할을 하게 되는데, 이 도입문헌을  $X_0$ 라고 정의하면,  $X_0$ 의 인용문헌집합은  $X_0$ 에 의해 인용된 문헌(Cited document)집합과  $X_0$ 를 인용한 문헌(Citing document) 집합으로 구성된다.  $X_0$ 의 인용문헌 집합은,

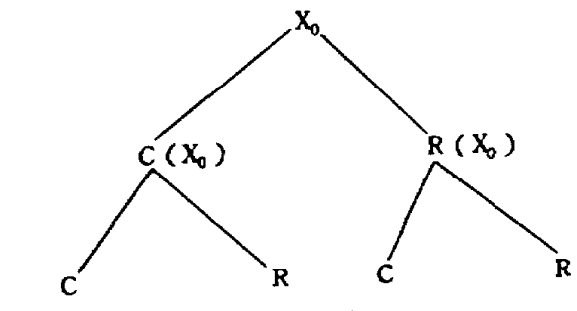
$$\{ R_i, C_i \} \quad \left( \begin{array}{l} R : \text{인용된 문헌} \\ C : \text{인용한 문헌} \\ i = 1, 2, 3, \dots, n \end{array} \right)$$

으로 나타낼 수 있다.

인용된 문헌집합은 실제 적합문헌으로부터 직접 검색하였고, 인용한 문헌집합은 인용색인지인 Science Citation Index와 Social Science Index를 통해 검색하였다.

이 인용문헌집합을 질문자에게 제공함으로써, 적합성 여부를 판별하게 한 후, 적합문헌으로 판별된 문헌들을 다시 도입문헌으로 하여 인용문헌집합을 구성할 수 있다. 그림 5는 도입문헌  $X_0$ 의 인용된 문헌과 인용한 문헌의 네트워크를 나타내고 있다.

(그림 5) 도입문헌( $X_0$ )을 이용한 인용문헌 네트워크



46) 「해외데이터뱅크 이용안내」, (서울 : 한국데이터통신주식회사, 1984), p.32.

이와같은 반복 검색은 수차례 되풀이 할 수 있는데, 되풀이함에 따라 새로운 문헌들이 나타나게 된다. 반복 검색은 추가할 새로운 문헌이 나타나지 않을 때까지 계속할 수 있다. 즉, 새로운 문헌들의 변화율이 0이되면, 이 반복 검색은 중단하게 된다. 실제로 이와같이 검색작업을 반복할 경우, 생물의학 분야에서는 하나의 도입문헌으로 1,000~1,500개의 새로운 문헌들을 찾아낼 수 있다고 보고되었다.<sup>47)</sup>

이와같은 반복 검색을 실시하여 새로운 문헌이 검색되는 결과는 주제의 범위, 그 주제분야의 연구량, 문헌을 인용하는 저자들의 인용패턴에 따라 달라진다.

본 실험에서는 반복 검색을 2회로 한정하였으며, 2차의 1회검색 결과에 만족하지 않는 경우에만 2회검색을 실시하였다.

4) 서지결합법과 동시인용법의 응용

2차검색 결과로 구성된 인용문헌집합에 서지결합법과 동시인용법을 응용하였다.

서지결합법에서는, A,B문헌이 C문헌을 공통으로 인용하고 있을 때 A,B문헌은 서지적으로 결합되었다고 보며, 결합단위는 1이 된다. 동시인용법에서는, A,B문헌이 C라는 문헌에 의해 동시에 인용되었을 때, A,B 두 문헌은 동시인용되었다고 보며 결합단위는 1이 된다.

그런데 본 실험에서는 A,B에 해당되는 문헌들이 이미 적합문헌들로 구성되어 있으므로 2차검색의 결과로 구성된 인용문헌집합내의 문헌들의 적합성 여부를 가리는데, 서지결합법과 동시인용법을 이용하였다. 결합단위는 인용문헌집합에서 공통으로 인용되고 있거나, 인용하고 있는 문헌들의 횟수로써 나타내었다.

5) 검색결과의 평가

1차, 2차검색의 결과가 나오면 수정재현율<sup>48)</sup>

과 정확률의 평가기준을 이용하여 검색효율을 측정하였다.

실험결과의 평가 및 분석은 본 연구자가 행하였다.

2.3 실험내용

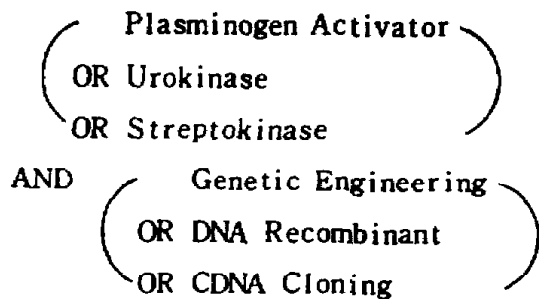
2.3.1 검색질문의 작성

질문자들에게 비교적 최근에 연구가 활발히 진행되면서, 검색문헌수가 많지 않을 것으로 예상되는 분야에서 질문을 작성하도록 하였다.

8개의 질문으로부터 적당한 검색어를 선정하였고, 검색어들은 불리안(Boolean)논리관계로 조합하였다.

질문 1~6은 생화학 분야이며, 질문 7~8은 정보학 분야이다. 8개 질문의 내용과 검색을 위해 작성된 검색문은 다음과 같다.

질문 1 : 플라즈미노젠 활성제에 유전공학적 방법을 적용시킨 문헌들.



질문 2 : 특별한 기능을 갖는 프린그랜드(Pren Gland)의 원형질내에 축적되는 미토콘드리아 효소의 전구물질에 관한 문헌들

(Precursor) AND (Mitochondrial)

질문 3 : 아세테이트 키베이즈의 분포, 반응메카니즘, 분석방법, 유전공학적 방법을 적용한 연구 문헌들.

47) T.R. Kochtanek, "Document Clustering, Using Macro Retrieval Techniques," *JASIS*, Vol. 34 (1983), pp. 356-9.

48) 후에 설명할 것임.

(Acetate Kinase) AND

Review

OR Assay

OR Distribution

OR Mechanism of Reaction

OR Kinetics

OR Genetic Engineering

질문 4 : 유미를 배설하는 환자의 카이로마이 크론내에 나타나는 에이포지질단백질 B에 관한 문헌들.

(Apolipoprotein B) AND (Chylomicrons) AND (Chyluria)

질문 5 : 모노클로날 항체를 가진 환자에게서 볼 수 있는 저밀도 지질단백질에 관한 문헌들.

(Low Density Lipoprotein) AND (Monoclonal Antibodies)

질문 6 : 글리코실레이션이나 당뇨병 환자에게서 볼 수 있는 저밀도지질단백질이나 카이로마이 크론과, 에이포지질단백질 B, B-100, B-48에 관한 문헌들.

(Low Density Lipoprotein)  
OR Chylomicrons

AND

(Apolipoprotein B  
OR Apolipoprotein B-48  
OR Apolipoprotein B-100)

질문 7 : 계량서지학적 측면에서 전염이론과 이를 응용시킨 문헌들.

Epidemic Theory

질문 8 : 서지결합법과 동시인용법을 응용한 검색에 관한 문헌들.

(Bibliographic Coupling) AND (Cocitation) AND (Retrieval)

### 2.3.2 주제색인어에 의한 검색(1차검색)

질문자들이 작성한 검색문으로 각기 CA Search와 LISA데이터베이스를 검색한 결과, 검색된 문헌수가 전혀 없거나 극히 적은 질문의 경우 검색되는 문헌수를 늘이기 위해서, 질문 3, 4의 검색문에 대폭 수정을 가하였고, 질문 1, 5, 6, 7, 8의 경우에도 약간의 수정을 하였다. 수정된 검색문은 다음과 같다.

질문 1 :

(Plasminogen Activator  
OR Urokinase  
OR Streptokinase)

AND

(Genetic  
OR DNA  
OR Cloning)

질문 2 :

(Precursor) AND (Mitochondrial)

질문 3 :

Acetate Kinase

질문 4 :

(Apolipoprotein) AND (Chylomicrons)

질문 5 :

(Low Density Lipoprotein) AND (Monoclonal)

질문 6 :

(Low Density Lipoprotein)  
OR Chylomicrons

AND (Apolipoprotein) AND

(Glycosylation)  
Diabetes

질문 7 :

Epidemic

질문 8 :

( Cocitation  
OR Bibliographic Coupling )  
AND ( Retrieval )

이상의 수정된 검색문으로 검색한 결과, 각 질문에 대한 검색된 문헌수와 검색된 적합문헌수는 표1과 같다.

< 표 1 > 1차검색의 결과

	검색된 문헌수	검색된 적합문헌수
질문 1	36	10
2	5	3
3	43	13
4	38	11
5	12	7
6	2	1
7	8	4
8	9	2

질문 3 과 4 는 검색어를 더욱 확장한 결과, 표 1에서와 같이 검색된 문헌수가 많음을 알 수 있다. 특히, 질문 3은 Acetate Kinase 라는 주제검색을 행한 것이라 볼 수 있다. 질문 2와

6 은 질문의 내용만을 완전하게 나타낼 수 있는 검색어만을 선정함으로써, 검색된 문헌수가 매우 적게 나타나고 있으며, 이 질문과 같은 주제의 연구결과는 많이 발표되지 않았음을 알 수 있다.

2.3.3 인용문헌에 의한 검색(2차검색)

표1에 나타난 각 질문에 대한 적합문헌들을 도입문헌으로 하여, 각 문헌의 인용한 문헌집합과 인용된 문헌집합으로 구성된 인용문헌집합을 검색하였다. 인용된 문헌집합은 각 문헌의 참고문헌에서, 인용한 문헌집합은 인용색인지인 SCI 와 SSCI 에서 탐색하였다.

인용색인지의 기본적인 탐색절차는 그림 6 과 같이 나타낼 수 있다.

[ 그림 6 ] 인용색인지의 탐색절차

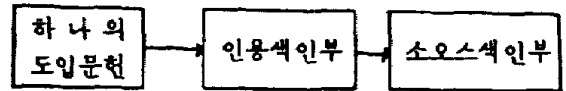
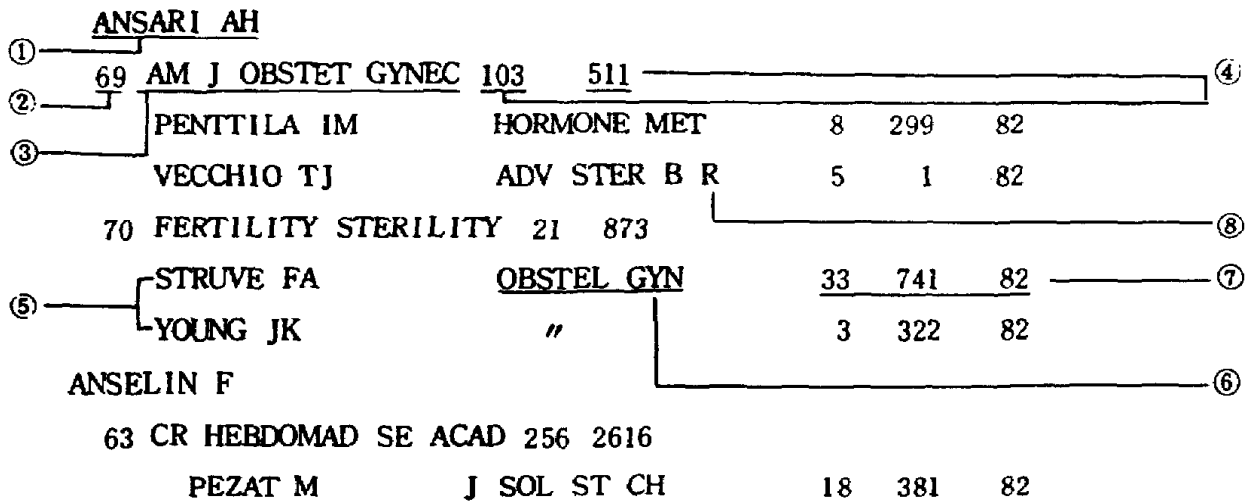


그림 6에서 도입문헌은 주제색인어에 의해 검색한 1차검색결과로부터 나온 적합문헌들로 구성되며, 각 도입문헌의 저자명으로 인용색인



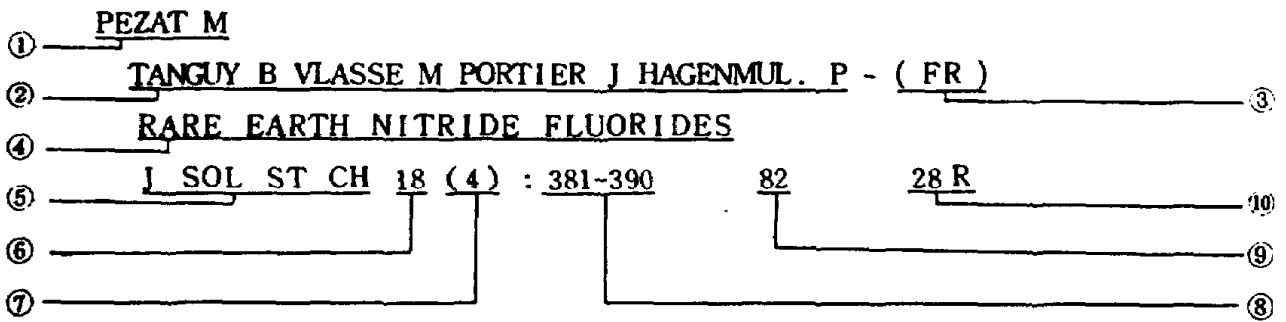
- ① 인용된 문헌의 저자명
- ② " 출판년도
- ③ " 잡지명
- ④ " 권수, 페이지
- ⑤ 인용한 문헌의 저자명
- ⑥ " 잡지명
- ⑦ " 권, 페이지, 년도
- ⑧ 문헌의 성격을 나타내는 코드  
(예 : M→Meeting, R→Review)

부"를 탐색하게 되는데, 저자명 밑에는 저자의 저작들이 수록되어 있고, 저작들 아래에는 각

저작을 인용한 문헌들이 수록되어 있다. 도입문헌을 저작들 중에서 찾아내어, 인용문헌들을 찾아낸다. 저작을 찾을 때, 필요한 서지사항은 문헌의 잡지명, 권호, 페이지, 출판년도이다.

인용색인부의 엔트리구성은 앞면과 같다.

인용색인부에서 알 수 있는 인용문헌의 서지사항은 인용한 문헌의 저자, 잡지명, 권호, 페이지, 출판년도이므로, 인용문헌의 서명을 찾기 위해서 인용한 문헌의 저자명으로 소오스 색인부를 다시 탐색하게 된다. 소오스 색인부의 엔트리구성은 아래와 같다.



- |                       |          |
|-----------------------|----------|
| ① 인용한 저자명             | ⑥ 잡지의 권수 |
| ② 공저자명                | ⑦ 잡지의 호수 |
| ③ 언어코드(예 : FR→French) | ⑧ 페이지    |
| ④ 서명                  | ⑨ 출판년도   |
| ⑤ 잡지명                 | ⑩ 참고문헌수  |

본 실험에서는 인용한 문헌의 검색을 수작업으로 행하였으나, 온라인 데이터베이스인 SCI Search와 Social SCI Search의 검색도 가능하다.

이렇게 검색된 인용문헌집합은 다시 질문자에게 제공하여, 적합성 여부를 판별하도록 하였다. 적합성 평가후 만족할 만한 결과를 가져오지 않은 질문은, 2차검색의 1회검색에서 나온 적합문헌을 도입문헌으로 하여 2회검색을 실시하였다. 이에 해당되는 질문은 7과 8이었다.

각 질문에 대한 2차검색의 1회, 2회검색결과는 다음과 같다. 단, 인용문헌이 단행본인 경우는 계산에서 제외하였다. 적합문헌수의 비율은 전체인용문헌집합의 적합문헌수를 100%로 하여 계산한 것이다.

49) SCI 나 SSCI 와 같은 인용색인지는 인용색인부 (Citation Index), 소오스색인부(Source Index), 순열주제 색인부(Permutated Index)의 3부분으로 구성된다.



질문 1 :

인용된 문헌 총수	254
중복되지 않는 인용된 문헌 총수	213
적합문헌수	26(43 %)
인용한 문헌 총수	87
중복되지 않는 인용한 문헌 총수	75
적합문헌수	34(57 %)

질문 2 :

인용된 문헌 총수	80
중복되지 않는 인용된 문헌 총수	71
적합문헌수	8(19 %)
인용한 문헌 총수	83
중복되지 않는 인용한 문헌 총수	63
적합문헌수	34(81 %)

질문 3 :

인용된 문헌 총수	297
중복되지 않는 인용된 문헌 총수	206
적합문헌수	33(89 %)
인용한 문헌 총수	55
중복되지 않는 인용한 문헌 총수	49
적합문헌수	4(11 %)

질문 4 :

인용된 문헌 총수	317
중복되지 않는 인용된 문헌 총수	193
적합문헌수	10(23 %)
인용한 문헌 총수	323
중복되지 않는 인용한 문헌 총수	202
적합문헌수	34(77 %)

질문 5 :

인용된 문헌 총수	120
중복되지 않는 인용된 문헌 총수	99
적합문헌수	12(50 %)
인용한 문헌 총수	21
중복되지 않는 인용한 문헌 총수	19

적합문헌수 | 12(50 %)

질문 6 :

인용된 문헌 총수	19
중복되지 않는 인용된 문헌 총수	19
적합문헌수	3(13 %)
인용한 문헌 총수	26
중복되지 않는 인용한 문헌 총수	26
적합문헌수	20(87 %)

질문 7 :

	1 회검색	2 회검색
인용된 문헌 총수	53	24
중복되지 않는 인용된 문헌 총수	47	18
적합문헌수	9	8(44 %)
인용한 문헌 총수	3	74
중복되지 않는 인용한 문헌 총수	3	46
적합문헌수	1	21(56 %)

질문 8 :

	1 회검색	2 회검색
인용된 문헌 총수	20	50
중복되지 않는 인용된 문헌 총수	18	35
적합문헌수	7	14(48 %)
인용한 문헌 총수	4	85
중복되지 않는 인용한 문헌 총수	4	61
적합문헌수	1	22(52 %)

각 질문에 대한 인용된 문헌집합과 인용한 문헌집합 즉, 인용문헌 집합내에서 중복되지 않는 적합문헌수와, 1 차검색 결과와 중복되지 않는 적합문헌수는 표 2 와 같다.

< 표 2 > 2차검색의 적합문헌수

	인용문헌집합내의 적합문헌수	1차검색과 중복되지 않는 적합 문헌수
질문 1	59	54
2	38	37
3	37	34
4	43	40
5	22	21
6	23	23
7	29	27
8	34	33

각 질문에 대한 1차검색 결과의 적합문헌수와, 1차검색의 적합문헌과 중복되지 않고 2차검색에서 추가로 검색된 적합문헌수는 표 3과 같다.

< 표 3 > 1, 2차검색의 적합문헌수 비교

	1차검색의 적합문헌수	2차검색의 적합문헌수	합계
질문 1	10	54	64
2	3	37	40
3	13	34	47
4	11	40	51
5	7	21	28
6	1	23	24
7	4	27	31
8	2	33	35

2차검색의 1회, 2회검색을 행한 질문 7, 8의 1차검색과 2차검색에서 추가로 검색된 적합문헌수는 표 4와 같다.

< 표 4 > 질문 7, 8의 1차, 2차검색 (1회, 2회 검색)의 적합문헌수

	1차검색의 적합문헌수	2차 1회검색의 적합문헌수	2차 2회검색의 적합문헌수	합계
질문 7	4 (13%)	9 (29%)	8 (58%)	31
8	2 (6%)	8 (23%)	25 (71%)	35

2.3.4 고전문헌을 이용한 검색

질문 7,8의 경우 각 질문의 내용을 대표할 수 있는 고전문헌들을 질문자가 이미 알고 있으므로, 이 문헌들을 도입문헌으로하여 인용문헌집합을 찾고자하는 검색을 실시하였다.

질문 7의 도입문헌은 계량서지학에 전념이론을 적용시킨 고프만(Goffman)의 문헌이며, 질문 8의 도입문헌은 스몰(Small)과 케슬러(Kessler)의 문헌으로 각기 동시인용법과 서지결합법을 소개하였다. 이 도입문헌들은 다음과 같다.

Goffman, W. *Nature*, Vol.202 (1966), pp. 449-52.

Small, H. *JASIS*, Vol.24 (1973), pp.265-9.

Kessler, M.M. *American Documentation*, Vol.14 (1963), pp.10-25.

질문 7,8의 고전문헌을 이용한 검색결과는 표 5,6과 같다.

< 표 5 > 질문 7의 고전적 문헌을 이용한 검색결과

	전체 인용문헌수	적합문헌수	인용 문헌집합내에서 중복되지 않는 적합 문헌수
인용된 문헌집합	4	2	14
인용한 문헌집합	23	12	

< 표 6 > 질문 8 의 고전적 문헌을 이용한 검색결과

		전체인용 문헌수	적 합 문헌수	인용문헌 집합 내에서 중복되 지 않는 적합 문헌수
인용된 문헌집합	케슬러 스 물	1 7	1 4	23
인용한 문헌집합	케슬러 스 물	39 32	14 11	

고전적인 문헌을 이용한 검색과, 1차, 2차 검색결과에의 적합문헌수의 비교는 표 7 과 같다.

< 표 7 > 고전적인 문헌을 이용한 검색과 1차, 2차 검색결과에의 비교

	1차, 2차검색의 적합문헌수	고전적인 문헌을 이용 한 검색의 적합문헌수
질문 7	31 (100%)	14 (45%)
8	35 (100%)	23 (66%)

2.3.5 서지결합법과 동시인용법의 응용

2차검색 결과로 만들어진 인용문헌집합에서 서지결합법과 동시인용법을 응용하였다. 결합단위 (n)는 각각 인용된 문헌집합과 인용한 문헌집합에서 공통으로 인용되고(인용하고) 있는 빈도수로 나타내었다.

질문 6의 경우는 1차검색에서 검색된 적합문헌수가 1개이었으므로, 인용문헌집합내에서는 결합단위를 이루는 문헌이 존재하지 않게 되어 결합단위는 나타나지 않았다.

질문 6을 제외한 각 질문마다 결합단위를 이루는 문헌들의 수와 적합문헌수는 표 8에서 표 14와 같다.

< 표 8 > 질문 1 의 결합단위를 이루는 문헌수와 적합문헌수

서지결합법 결합단위	문헌수	적 합 문헌수	동시인용법 결합단위	문헌수	적 합 문헌수
n = 1	10	4	n = 1	8	5
2	6	3	2	2	2
3	5	3	.	.	.
4	1	1			

< 표 9 > 질문 2 의 결합단위를 이루는 문헌수와 적합문헌수

서지결합법 결합단위	문헌수	적 합 문헌수	동시인용법 결합단위	문헌수	적 합 문헌수
n = 1	7	3	n = 1	20	15
2	1	1	.	.	.

< 표 10 > 질문 3 의 결합단위를 이루는 문헌수와 적합문헌수

서지결합법 결합단위	문헌수	적 합 문헌수	동시인용법 결합단위	문헌수	적 합 문헌수
n = 1	22	6	n = 1	4	0
2	6	4	2	1	0
3	5	4	.	.	.
4	2	2			
5	3	3			
6	2	2			
7	1	1			

< 표 11 > 질문 4 의 결합단위를 이루는 문헌수와 적합문헌수

서지결합법 결합단위	문헌수	적 합 문헌수	동시인용법 결합단위	문헌수	적 합 문헌수
n = 1	35	3	n = 1	45	10
2	17	3	2	16	8
3	8	1	3	8	4
4	1	0	4	2	1
5	2	0	5	1	1
6	1	0	6	-	-
7	-	-	7	1	1
8	1	0	.	.	.

< 표 12 > 질문 5의 결합단위를 이루는 문헌수와 적합문헌수

서지결합법 결합단위	문헌수	적합 문헌수	동시인용법 결합단위	문헌수	적합 문헌수
n=1	15	2	n=1	4	2
2	3	2	2	3	3
3	1	1	.	.	.

< 표 13 > 질문 7의 결합단위를 이루는 문헌수와 적합문헌수

서지결합법 결합단위	문헌수	적합 문헌수	동시인용법 결합단위	문헌수	적합 문헌수
n=1	4	3	n=1	4	2
2	4	4	2	3	3
.	.	.	3	1	0
			4	2	2
			5	-	-
			6	-	-
			7	-	-
			8	1	1

< 표 14 > 질문 8의 결합단위를 이루는 문헌수와 적합문헌수

서지결합법 결합단위	문헌수	적합 문헌수	동시인용법 결합단위	문헌수	적합 문헌수
n=1	6	4	n=1	18	11
2	1	1	2	3	3
3	1	1	-	-	-
4	1	1			

2.3.6 검색효율의 측정

검색에 의하여 얻어진 정보가 질문자의 요구를 어느 정도로 만족시켜 주었는가를, 즉 검색효율을 측정하였다.

검색효율의 측정에는 보통 재현율(recall ra-

tio)과 정확률(precision ratio)로 나타내게 된다. 재현율과 정확률의 산출공식은 다음과 같다.

$$\text{재현율} = \frac{\text{검색된 적합 문헌수}}{\text{적합 문헌 총수}} \times 100$$

$$\text{정확률} = \frac{\text{검색된 적합 문헌수}}{\text{검색된 문헌 총수}} \times 100$$

위의 재현율 산출공식을 이용하려면, 한 시스템내에 포함된 적합문헌 총수를 알아야 하므로, 소규모 실험집단에서나 가능하다고 볼 수 있다. 본 실험에서는 축적문헌건수가 방대한 상업용 데이터베이스를 이용하여 검색하였으므로, 이 데이터베이스내에 포함되어 있는 각 질문에 대한 적합문헌총수를 알아낸다는 것은 거의 불가능하였다. 그러므로 본 연구에서는 수정재현율로 측정을 하였는데, 수정재현율은 각 질문에 대한 1,2차검색에서 검색된 적합문헌수를 백분율로 나타내었다. 즉,

< 표 15 > 각 질문에 대한 1, 2차 검색의 수정재현율

	1차검색의 수정 재현율	2차검색의 수정 재현율
질문 1	16 %	92 %
2	8 %	95 %
3	28 %	79 %
4	22 %	84 %
5	25 %	79 %
6	4 %	96 %
7	13 %	94 %
8	6 %	97 %
계	16 %	89 %

$$1차(2차)검색의 수정재현율 = \frac{1차(2차)검색의 적합문헌수}{1,2차검색의 적합문헌수} \times 100$$

즉 각 질문에 대한 1,2차검색의 수정재현율은 표15와 같다.

정확률 측정공식은 수정하지 않고 이용할 수 있으므로, 각 질문에 대한 1차 2차검색의 정확률 측정은 표16과 같다.

각 질문마다 서지결합법과 동시인용법을 응용시켜 나온 결과의 검색효율을 결합단위별로 계산한 결과는 표17,18과 같다.

<표 16> 각 질문에 대한 1, 2차 검색의 정확률

	1 차 검색			2 차 검색			
	검색된 문헌수	적합문헌수	정확률		검색된 문헌수	적합문헌수	정확률
질문 1	36	10	28%	R	213	26	12%
				C	75	34	45%
2	5	3	60%	R	71	8	11%
				C	63	34	54%
3	43	13	30%	R	206	33	16%
				C	49	4	8%
4	45	11	24%	R	193	10	5%
				C	202	34	17%
5	12	7	58%	R	99	12	12%
				C	19	12	63%
6	2	1	50%	R	19	3	16%
				C	26	20	77%
7	8	4	50%	R	65	17	26%
				C	49	22	45%
8	9	2	22%	R	53	21	40%
				C	65	26	40%
계	160	51	31%	R	919	130	14%
				C	548	183	33%
					1,467	313	21%

(R : 인용된 문헌, C : 인용한 문헌)

< 표 17 >

n ≥ 1 일 때의 검색효율

		총 적 합 문헌수	검색된 문헌수	검색된 적 합 문헌수	수 정 재현율	정확률		총 적 합 문헌수	검색된 문헌수	검색된 적 합 문헌수	수 정 재현율	정확률		
질문 1		26	22	11	42 %	50 %		34	10	7	21 %	71 %		
2	인 용 된 문 헌	8	8	4	50 %	50 %	인 용 한 문 헌	34	20	15	44 %	74 %		
3		33	40	22	67 %	55 %		4	5	0	0	0		
4		10	65	7	70 %	11 %		34	73	25	74 %	34 %		
5		12	19	5	42 %	26 %		12	7	5	42 %	71 %		
7		8	4	3	38 %	75 %		21	11	8	38 %	73 %		
8		14	9	7	50 %	78 %		22	21	14	64 %	67 %		
계			111	167	59	53 %		35 %		161	147	74	46 %	50 %

< 표 18 >

n ≥ 2 일 때의 검색효율

		총 적 합 문헌수	검색된 문헌수	검색된 적 합 문헌수	수 정 재현율	정확률		총 적 합 문헌수	검색된 문헌수	검색된 적 합 문헌수	수 정 재현율	정확률		
질문 1		26	12	7	27 %	58 %		34	2	2	6 %	100 %		
2	인 용 된 문 헌	8	1	1	13 %	100 %	인 용 한 문 헌	34	-	-	-	-		
3		33	19	16	49 %	84 %		4	1	0	25 %	0 %		
4		10	30	4	40 %	13 %		34	28	15	44 %	54 %		
5		12	4	3	25 %	75 %		12	3	3	25 %	100 %		
7		8	2	2	25 %	100 %		21	7	6	29 %	86 %		
8		14	3	3	21 %	100 %		22	3	3	14 %	100 %		
계			111	71	36	32 %		51 %		161	44	29	18 %	66 %

결합단위에 따른 검색효율은 표 19와 같다.

< 표 19 > 검색단위별 검색효율

	검색효율	n ≥ 1	n ≥ 2
인용된 문헌	수정재현율	53 %	32 %
	정확률	35 %	51 %
인용한 문헌	수정재현율	46 %	18 %
	정확률	50 %	66 %
전체인용문헌	수정재현율	49 %	24 %
	정확률	39 %	57 %

2.4 실험결과 분석

본 절에서는 앞절 2.2~2.4에서 작성한 표를 이용해서, 실험결과를 분석해 보았다.

인용문헌을 이용하여 재현율을 높이고자한 가설을 세웠는데, 각 주제에 대한 데이터베이스내의 적합문헌 총수를 알 수가 없었으므로, 각 1, 2차검색에서 얻어진 적합문헌수를 100%로 보았을 때의 수정재현율추정이 표 15에 나와있다. 각 질문마다 2차검색의 수정재현율이 모두 월등히 높은 것을 알 수 있으며, 1차검색의 평균수정재현율은 16%, 2차검색의 평균수정재현율은 89%에 달하고 있다. 즉, 주제색인어를 이용하여 1차검색을 실시한 결과, 적합한 문헌을 만족스러울 정도로 얻지못했을 경우, 2차검색을 인용문헌을 이용하여 실시함으로써, 약 72%의 수정재현율이 증가하였음을 알 수 있으므로, 2차검색으로 더욱 많은 적합문헌들을 찾아낼 수 있음이 입증되었다.

주제의 범위가 상당히 한정되어 있고, 연구결과가 많이 발표되지 못한 질문 6에서는, 1차검색으로 얻어진 1개의 적합문헌을 이용하

여 2차검색에서는 23개의 적합문헌을 찾아낸 것을 볼 수 있다. 특히 질문 3과 4는 1차검색시 주제를 확장하여 검색함으로써, 검색된 문헌수가 각기 43, 38개나 되었는데도, 2차검색에서 각기 79%, 84%의 수정재현율을 나타낸 것을 알 수 있다. 질문 7과 8은 1, 2차 검색결과에 만족하지 않고, 2차검색의 2회검색을 실시한 결과, 2차검색의 1회검색에서는 27%, 2차검색에서는 68%의 수정재현율을 나타내었으므로, 필요한 경우 인용문헌을 통해 계속적으로 검색을 반복하면 일정 횟수까지는 더 많은 적합문헌들이 검색될 수 있음을 알 수 있다.

인용된 문헌집합과 인용한 문헌집합에서의 검색결과는 표 16으로 비교해 볼 수 있는데, 질문 2, 4, 6에서는 인용한 문헌집합에서 더 많은 적합문헌을 검색해 내었고, 질문 1, 5, 7, 8의 경우는 서로 거의 비슷하였고, 질문 3에서는 인용된 문헌집합에서 더 많은 적합문헌을 검색해 내었다. 두 문헌집합의 평균 수정재현율은 인용된 문헌집합이 41%, 인용한 문헌집합이 57%로 나타났는데, 인용한 문헌집합을 이용한 검색결과가 더 우수하다고 볼 수 있으나 인용문헌을 통하여 재현율을 높이고자하는 검색에서는 인용된 문헌을 이용한 검색이나 인용한 문헌을 이용한 검색중 하나만을 선택하는 것보다는 두가지 검색을 모두 행하는 것이 바람직하다고 볼 수 있다.

또한, 2차검색에서 검색된 적합문헌들이 1차검색시 사용했던 검색문들을 완전히 만족시키기 때문에 적합문헌으로 평가되었다면, 주제색인작업이 제대로 행하여지지 않고있다고 볼 수 있다. 그러나 질문 1~6의 질문자들이 적합하다고 평가한 문헌들에는, 연구자들이 선택한 점

색문과 일치하지 않더라도 질문의 이해와 연구에 필요한 문헌들의 상당수가 포함되어 있음을 알 수 있다. 그러므로, 인용문헌을 이용한 검색은 질문자가 특정 주제에 대한 전반적인 연구 문헌을 필요로 할 때, 더욱 효과적이라는 결론을 내릴 수 있다.

질문 7의 2차검색의 1회, 2회검색에서 나온 적합문헌들은, 전염이론을 적용시킨 문헌들이나 전염이론을 다루고 있는 문헌들로써 구성되는데, 1차검색에서 검색되지 못한 이유는 다음과 같다. 2차검색결과로 나온 적합문헌들을 수록하고 있는 총 18개 잡지 중 LISA에 포함되어 있는 잡지는 JASIS, Journal of Chemical Information and Computer Science, A-slib Proceedings 인 3개 뿐으로 17%에 불과하다. 전염이론 자체가 본래 의학분야의 전염병학에서 비롯되어, 제량서지학에 응용시킨 것으로, 이 전염이론과 관계된 문헌들은 의학분야 잡지와 과학분야 잡지에 많이 포함되어 있다. 결과적으로, 여러 학문과 관련있는 주제검색에서는 인용문헌을 통해 관련된 문헌을 찾아내는 것이 효과적임을 알 수 있다.

질문 8의 경우에는 서지결합법과 동시인용법을 이용한 정보검색에 관한 문헌들을 검색하고자 한 것으로, 2차검색에서 검색된 적합문헌들의 서명을 살펴보면 'Bibliographic Coupling'과 'Co-citation'이란 용어대신, 'Cited', 'Citing' 또는 'Citation'이란 용어들을 많이 사용하고 있음을 알 수 있었다. 그러므로 인용문헌을 이용한 2차검색결과는, 질문 8로부터 검색문을 작성할 때 'Bibliographic Coupling', 'Cited', 'Co-citation', 'Citing', 'Citation'에 'Retrieval'을 AND로 결합시켜 검색한 결과와 같은 효과를 갖는 것을 알 수 있

다. 즉, 1차검색에서 질문의 내용을 나타내는 검색어들 중 일부만으로 검색을 한 후, 인용문을 이용하여 검색한 결과는, 검색어 전부를 이용하여 1차검색을 한 결과와 같이 효율적이라는 것을 알 수 있다.

한 주제의 고전적인 문헌을 알고 있는 질문의 경우, 이 문헌들을 도입문헌으로 시작하여 검색을 하였다. 이러한 검색을 행한 질문 7, 8의 검색결과가 표 7에 나와 있다. 질문 7에서는 1개의 고전적 문헌으로 검색한 결과 1차, 2차의 1회, 2회검색에서 찾아낸 적합문헌들 중 45%를, 질문 8에서는 2개의 고전문헌으로 검색한 결과 1, 2차의 1회, 2회검색에서 찾아낸 적합문헌들의 66%를 각각 찾아내었다. 이와같이 한 주제의 고전적인 문헌을 알고 있다면, 이 문헌을 이용한 한 차례의 검색으로 1차, 2차의 1회, 2회의 3차례 검색을 행하여 얻어낸 적합문헌수의 56% 이상을 찾아내었음을 알 수 있다.

이상과 같이 인용문헌을 이용했을 때, 수정재현율이 높아졌다는 사실외에, 8개 질문에 대한 1차, 2차 검색결과의 정확률을 표 16에서 살펴볼 수 있다. 8개 질문에 대한 1차검색의 평균 정확률은 31%가 되며, 2차검색의 평균 정확률은 20%로서, 주제색인어를 이용한 검색결과의 정확률보다 인용문헌을 이용한 검색의 정확률이 약 10%정도 떨어진 것으로 나타났다. 재현율과 정확률은 서로 반비례 관계에 있으므로 당연한 결과라 할 수 있는데, 이 실험의 목적은 적합문헌을 검색하기 어려운 주제를 검색 질문으로 선정하여 재현율을 높이는 데 있으므로, 이 정도로 떨어진 정확률은 그리 큰 문제가 되지 않는다고 볼 수 있다.

2차검색 중 인용된 문헌집합에서의 정확률은



12%, 인용한 문헌집합에서의 정확률은 32%로서, 인용한 문헌집합의 정확률이 더 높은 것으로 나타났다.

2차검색에서 만들어진 인용문헌집합에 서지결합법과 동시인용법을 응용하여, 결합단위별 재현율과 정확률을 산출한 결과가 표 19에 나와있다. 표 19에서 알 수 있는 바와 같이 재현율과 정확률은 반비례 관계에 놓여 있으므로, 정확률을 높이려고 하면 결합단위를 높여서 검색하면 된다. 반면에 관련된 모든 문헌을 찾고자 한다면, 결합단위를 낮추어서 검색하면 재현율이 높아지게 된다. 이러한 결정은 질문자가 검색하고자 하는 범위에 따라 정하면 된다. 2차검색에서 나온 인용문헌집단이 방대한 경우, 서지결합법과 동시인용법을 응용하여 재현율과 정확률을 조절할 수 있다.

결론적으로, 질문자가 1차검색결과에 만족하지 않는다는 가정하에서 정확률은 다소 떨어지더라도 더 많은 적합문헌을 찾고자 한다면 인용문헌을 이용하여 2차검색을 행하는 탐색전략이 생화학과 정보학분야와 같이 종합학문적 성격을 띤 검색질문에 효과적임이 실험결과 입증되었다.

#### IV. 결 론

초록지나 색인지와 같은 전통적인 서지도구를 이용하여 문헌검색을 하면 것이, 1970년대 초부터 온라인 방식에 의한 정보검색시스템으로 전파되기 시작하였다. 이러한 정보검색시스템은 일반적으로 주제색인에 의한 검색을 하게 되는데, 이용자의 정보요구를 만족시키지 못하는 경우가 늘어나게됨에 따라 주제색인에만

의존하지 않는 새로운 탐색전략이 필요하게 되었다.

본 논문에서는 생화학과 정보학 분야를 대상으로 8개의 질문을 작성하여, 1차검색은 주제색인에 의해 기존의 상업용 데이터베이스를 탐색한 후, 그 결과로 나온 적합문헌을 이용하여, 인용문헌을 검색하는 2차검색을 실시하였다.

이 논문의 실험결과, 다음과 같은 사실이 밝혀졌다.

1. 높은 재현율을 원하는 경우 주제색인을 이용한 1차검색에 만족하지 않고 인용문헌을 이용한 2차검색을 수행하도록 하는 검색모형이 효과적임을 입증하였다. 실험에서 종합학문적 성격을 띤 생화학과 정보학분야의 검색질문에서 주제색인에 의한 검색을 보완하기 위해 인용문헌을 이용한 검색결과, 주제색인에 의한 1차검색보다 정확률은 10%정도 떨어졌지만 수정재현율은 70%정도 증가하였다.

2. 질문의 내용을 최적으로 나타내지 못하는 검색어를 선정하여 검색하더라도 인용문헌을 이용한 검색방법으로 적합문헌들을 거의 모두 찾아낼 수 있음이 입증되었다. 실제로 질문 8의 경우 검색을 실시한 'Bibliographic Coupling'과 'Co-citation' 검색어외에 'Cited', 'Citing'과 'Citation'이란 검색어를 이용한 것과 같은 효과를 얻었다.

3. 하나의 도입문헌으로 인용된 문헌집합과 인용한 문헌집합을 찾아내고, 이 인용문헌집합에서 나온 적합문헌을 다시 도입문헌으로 이용하는 반복검색작업을 행한 2차검색의 1, 2회 수정재현율은 각기 27%, 68%로 나타났는데, 반복검색작업을 계속 행하면 재현율은

향상되거나 정확률은 떨어지게 됨으로 질문자와의 피이드백에 의해 반복횟수를 정하는 것이 필요하다.

4. 인용된 문헌집합과 인용한 문헌집합에서의 수정재현율은 자기 41%, 57%로써, 인용한 문헌집합의 수정재현율이 더 높으나 그 차이가 크지 않으므로 한가지 인용문헌집합을 이용한 검색방법보다는 두가지 인용문헌집합을 이용한 방법이 바람직하다.

5. 2차검색결과에 의한 인용문헌집합에 서지결합법과 동시인용법을 응용시킨 결과, 결합단위별 수정재현율과 정확율은  $n \geq 1$  일때 각각 49%, 42%,  $n \geq 2$  일때 각각 24%, 57%로 나타났다. 이것으로 서지결합법과 동시인용법에 의한 주제검색이 하나의 탐색기법이 될 수 있음이 증명되었다.

본 연구는 소규모의 실험집단을 만들어 실험을 행한 것이 아니고, 기존의 상업용 서지데이터베이스를 이용하여 현실적으로 가능한 범위에서 실험을 함으로써, 검색효율중 재현율을 높이는 탐색전략을 제시하였다.

본 연구에서는 종합학문적인 성격을 띤 생화학과 정보학 분야를 대상으로 검색질문을 선정하였으므로 종합학문적 성격을 띤 분야에 대한 실험이 수행되어야 할 것이며, 기존의 정보검색시스템에서 검색결과로 문헌의 서지사항과 인용문헌을 함께 제공할 수 있도록 하는 방법이 검토되어야 할 것이다.

### 참 고 문 헌

1. 司空哲, 「情報檢索論」 서울: 아세아문화사, 1977.
2. 송정석, 「생화학」 서울: 연세대학교생화학교실, 1982.
3. 「해외데이터뱅크이용안내」 서울: 한국데이터통신주식회사, 1984.
4. Alt, F.L., "Citation Searching and Bibliographic Coupling with Remote On-Line Computer Access," *Journal of Research of the National Bureau of Standards-B. Mathematical Sciences*, Vol. 72B (1968), pp. 61-76.
5. Bichteler, J. and Eaton, E., "Comparing Two Algorithms for Document Retrieval Using Citation Links," *JASIS*, Vol. 28 (1977), pp. 192-5.
6. \_\_\_\_\_, "The Combined Use of Bibliographic Coupling and Co-citation for Document Retrieval," *JASIS*, Vol. 31 (1980), pp. 178-82.
7. Bichteler, J. and Parsons, R.G., "Document Retrieval by Means of an Automatic Classification Algorithm for Citations," *Information Storage and Retrieval*, Vol. 10 (1974), pp. 267-78.
8. Cleveland, D.B., "An n-Dimensional Retrieval Model," *JASIS*, Vol. 27(1976), pp. 342-6.
9. Doyle, L.B., "Semantic Road Maps for Literature Searches," *Journal of the Association for Computing Machinery*, Vol. 8 (1961), pp. 553-78.
10. Dutta, S. and Rajagopalan, T.S., "Literature Citations in Scientific and Technical Periodicals-A Survey," *Journal of Scientific and Industrial Research*, Vol. 17A (1958), p. 259-61.
11. Garfield, E., "Citation Indexes for Science," *Science*, Vol. 122 (1955), pp. 109-11.

12. Garfield, E., "ISI's On-line System Makes Searching So Easy Even a Scientist Can Do It," *Current Contents*, Vol. 13 (1981), pp. 5-8.
13. Goffman, W., "A Searching Procedure for Information Retrieval," *Information Storage and Retrieval*, Vol. 2 (1964), pp. 73-8.
14. \_\_\_\_\_, "An Indirect Method of Information Retrieval," *Information Storage and Retrieval*, Vol. 4 (1968), pp. 361-73.
15. Gray, W.A. and Harley, A.J., "Computer Assisted Indexing," *Information Storage and Retrieval*, Vol. 7 (1971), pp. 167-74.
16. Kessler, M.M., "An Experimental Study of Bibliographic Coupling Between Technical Papers," *IEEE Transactions on Information Theory*, Vol. 9 (1963), pp. 49-51.
17. \_\_\_\_\_ "Bibliographic Coupling Between Scientific Paper," *American Documentation*, Vol. 14 (1963), pp. 10-25.
18. \_\_\_\_\_ "Bibliographic Coupling Extended in Time: Ten Case Histories," *Information Storage and Retrieval*, Vol. 1 (1963), pp. 169-87.
19. \_\_\_\_\_, "Comparison of the Results of Bibliographic Coupling and Analytic subject Indexing," *American Documentation*, Vol. 16 (1965), pp. 223-33.
20. \_\_\_\_\_, "Technical Information Flow Patterns," *Proceedings of Western Joint Computer Conference*, 1961, pp. 247-57.
21. Kochtanek, T.R., "Document Clustering, Using Macro Retrieval Techniques," *JASIS*, Vol. 34 (1983), pp. 356-9.
22. Kwok, K.L., "The Use of Title and Cited Titles as Document Representation for Automatic Classification," *Information Processing and Management*, Vol. 11 (1975), pp. 201-6.
23. Lenk, P., "Mapping of Fields Based on Nominations," *JASIS*, Vol. 34 (1983), pp. 115-22.
24. Maron, M.E. and Khuns, J.L., "On Relevance, Probabilistic Indexing and Information Retrieval," *Journal of the Association for Computing Machinery*, Vol. 7 (1960), pp. 216-44.
25. O'Connor, J., "Biomedical Citing Statement: Computer Recognition and Use to Aid Full-Text Retrieval," *Information Processing and Management*, Vol. 19 (1983), pp. 361-8.
26. \_\_\_\_\_, "Citing Statement: Computer Recognition and Use to Improve Retrieval," *Information Processing and Management*, Vol. 18 (1982), pp. 125-31.
27. Price, N. and Schiminovich, S., "A Clustering Experiment: First Step Towards a Computer-generated Classification Scheme," *Information Storage and Retrieval*, Vol. 4 (1968), pp. 271-80.
28. Rolling, L.N., "A Computer-aided Information Service for Nuclear Science and Technology," *Journal of Documentation*, Vol. 22 (1966), pp. 93-115.
29. Salton, Gerard, "Associative Document Retrieval Techniques Using Bibliographic Information," *Journal of the Association for Computing Machinery*, Vol. 10 (1963), pp. 440-57.
30. \_\_\_\_\_, "Automatic Indexing Using Bibliographic Citations," *Journal of Documentation*, Vol. 27 (1971), pp. 98-110.
31. \_\_\_\_\_, *Introduction to Modern Information Retrieval*. New York: McGraw-Hill, 1983.
32. \_\_\_\_\_ *The SMART Retrieval System: Experiments in Automatic Document Processing*. New York: Prentice-Hall, 1971.

33. Schiminovich, S., "Automatic Classification and Retrieval of Document by Means of a Bibliographic Pattern Discovery Algorithm," *Information Storage and Retrieval*, Vol. 6 (1971), pp. 417-35.
34. Small, H., "Co-citation Context Analysis and the Structure of Paradigms," *Journal of Documentation*, Vol. 36 (1980), pp. 183-96.
35. \_\_\_\_\_, "Co-citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents," *JASIS*, Vol. 24 (1973), pp. 265-9.
36. Small, H.G., "Journal Clustering Using a Bibliographic Coupling Method," *Information Processing and Management*, Vol. 13 (1977), pp. 277-88.
37. Smith, L.C., "Citation Analysis," *Library Trends*, Vol. 29 (1981), pp. 83-106.
38. Verhoeff, J., et al., "Inefficiency of the Use of Boolean Functions for Information Retrieval Systems," *Communication of the Association for Computing Machinery*, Vol. 4 (1961), pp. 557-9.
39. Virgo, J.A., "The Review Article: Its Characteristics and Problems," *Library Quarterly*, Vol. 4 (1971), pp. 275-91.
40. Weinberg, B.H., "Bibliographic Coupling: A Review," *Information Storage and Retrieval*, Vol. 10 (1974), pp. 189-96.
41. Westbrook, J.H., "Identifying Significant Research," *Science*, Vol. 132 (1960), pp. 1229-34.
42. White, H.D., "Cocited Author Retrieval Online: An Experiment with the Social Indicators Literatures," *JASIS*, Vol. 32 (1981), pp. 16-21.
43. White, H.D. and Griffith, B.C., "Author Cocitation: A Literature Measure of Intellectual Structure," *JASIS*, Vol. 32 (1981), pp. 163-71.