

# 정보검색 발전사\*

## The Historical Development of Information Retrieval Systems

사공 철(Chul Sagong)\*\* 서경주(Gyeong-Ju Seo)\*\*\*

### 목 차

- |                          |                                   |
|--------------------------|-----------------------------------|
| 1. 서론                    | 5. 1980년대                         |
| 2. 1950년대까지              | 5.1 전문데이터베이스의 증가                  |
| 3. 1960년대                | 5.2 온라인 인터페이스의 등장                 |
| 3.1 서지정보의 기계가독화          | 5.3 해외 데이터베이스 이용과 국내 정보검색 시스템의 개발 |
| 3.2 오프라인 시스템과 실험적 온라인시스템 | 5.4 CD-ROM의 등장                    |
| 3.3 시소러스의 발전             | 6. 1990년대                         |
| 3.4 후조합색인에 관한 주요 연구      | 6.1 CD-ROM의 발전                    |
| 3.5 학문으로서의 방향제시          | 6.2 멀티미디어 정보검색과 하이퍼텍스트 기술의 개발     |
| 4. 1970년대                | 6.3 지능형 정보검색시스템의 등장               |
| 4.1 온라인 검색시스템의 확대        | 6.4 인터넷의 발전                       |
| 4.2 전문검색 시스템의 등장         | 7. 장래의 전망                         |

### 초 록

1950년대에서 1990년대까지의 정보검색 발전사를 연대별로 기술하였다. 1950년대에서는 후조합색인법과 KWIC방식을, 1960년대에서는 오프라인 시스템과 실험용 온라인시스템, 1970년대에서는 온라인시스템, 전문검색시스템, 1980년대에서는 전문데이터베이스, 온라인 인터페이스, 해외 데이터베이스의 이용 및 국내 온라인시스템, 그리고 1990년대에서는 CD-ROM, 멀티미디어 및 하이퍼텍스트, 인터넷 등의 발전에 관하여 각각 기술하고 장래에 관하여 전망하였다.

### ABSTRACT

The development of information retrieval between 1950s and 1990s is described chronologically. For each decade, the following information retrieval systems are examined : post-coordinate and KWIC indexing methods for the 1950s ; off-line and experimental on-line systems for the 1960s ; on-line and full-text retrieval systems for the 1970s ; full-text databases, on-line interfaces, and overseas and domestic on-line databases for the 1980s ; and finally for the 1990s, CD-ROM, multimedia, hypertext, and Internet. The prospects for the future are also discussed.

\* 1995년도 숙명여자대학교 특별연구비에 의하여 이룩됨

\*\* 숙명여자대학교 문헌정보학과 교수

\*\*\* 숙명여자대학교 문헌정보학과 강사

■ 논문접수일 : 1996년 10월 27일

## 1. 서론

정보검색은 인간의 지적활동에 있어 가장 기본적인 행동 가운데 하나이며, 보다 진보된 정보검색 방법을 위한 노력이 수십년 동안 계속되어 왔다. 도서관의 기본적인 기능은 인류의 지적 유산을 보존하는 것과 이를 제공하는 것, 두 가지로 대별할 수 있는데, 19세기에서 20세기로 넘어오면서 이전의 보존 제일주의에서 제공기능으로 그 중점이 옮겨지게 되었다. 이 변화의 직접적인 요인은, 제2차 세계대전 이후 정보에 대한 개념과 인식이 바뀌어서 정보가 단순히 개인이나 기업의 전유물이 아니라 국가자원이라는 인식이 생긴 데에 있다. 이러한 변화가 실체화된 것이 도큐멘테이션이다.

19세기 말부터 유럽에서 시작한 도큐멘테이션은 1940년대에 들어서 특히 활발하게 전개되었으며 그 후 정보검색이 이 개념을 대신하게 되었다. 정보검색(Information Retrieval)이라는 용어는 1950년 Mooers가 처음 사용함으로써 출현하였으며 Cleverdon과 Thorne이 그들의 보고서에서 이 용어를 사용함으로써 일반화되기 시작하였다. 그러나 정보검색의 개념은 그 이전부터 이미 존재하여 왔으며 Hugo가 성경용어 색인을 편찬하기 위하여 500명의 수도사를 고용했다는 1247년까지 소급할 수 있다.

정보검색의 역사는 컴퓨터의 패러다임을 따라 진전되어 왔다. 본고에서는 검색에 사용된 컴퓨터 기술에 비추어 정보검색의 발전단계를 고찰하여, 그 역사적 흐름과 발전과정을 정리하여 봄으로써 앞으로의 정보검색을 전망하고

자 한다.

## 2. 1950년대까지

1940년대에 반기계화된 시스템이 등장하기 전까지 정보검색은 인쇄된 카드목록이나 책자 형태의 색인이 주류를 이루었다. 즉 분류기호나 주제명으로 단행본을 검색하는 것이 대부분이었다.

그러나 출판물의 양적 증가는 색인과 검색 방식에 변화를 요구하게 되었다. 종래의 분류기호나 주제명 색인을 위주로 하는 전조합색인에서, 후조합색인 방식이 생기게 되었는데, 이는 1950년에 유니텀카드로 발전하였다.

반기계화 방식의 대표적인 시스템은 Batten의 광학일치카드 시스템(optical coincidence card system)과 Mooers의 엣지-notch드카드 시스템(edge-notched card system)을 들 수 있다. 특히 광학일치카드 시스템은 1915년 Taylor가 고안한 것을 Batten이 플라스틱분야의 특허정보 색인에 처음으로 채택한 이래 컴퓨터 이전단계에 상당히 많이 이용된 방식이었다. 또한 1950년대 초에는 마이크로필름 탐색방법이 많이 사용되었는데 그 시스템의 예로 1949년 Engineering Research Associates가 제작하여 후일 예일대를 비롯한 여러 곳에서 개량 사용된 신속선택기(rapid selector)를 들 수 있다.

1945년 최초의 컴퓨터인 ENIAC이 개발되었고 1951년 UNIVAC 1 이후 실용화되어 감에 따라 정보검색도 여명기에 접어들었다.

1950년대는 다양한 색인시스템과 분류시스

템. 정보검색 관련자들 간의 끝없는 논쟁으로 특징지어진다. 이 가운데 가장 획기적인 것은 1950년 Taube가 고안한 유니탐시스템으로서 이는 키워드방식을 채택한 최초의 시스템이다. 통제하지 않은 자연어회에 대하여 관심을 가졌던 Taube는 유니탐이라는 통제하지 않은 단어들(single word)를 사용하여 색인하고 검색시 이 단어들을 조합하여 정보요구를 나타내도록 하였다. 조합시스템이 이전에도 존재하였지만 유니탐시스템은 컴퓨터를 기반으로 하는 후조합시스템이라는 점과 광범위하게 사용된 검색시스템이라는 점에서 최초의 것이라고 할 수 있다(Lancaster 1968).

1953년에는 유니탐과 주제명표목에 의한 색인방식의 검색효율을 비교하기 위하여 대규모 실험이 실시되었다. 즉 기술문헌 15,000건에 대하여 유니탐과 ASTIA(Armed Services Technical Information Agency) 주제명표목을 각각 적용하여 색인한 후 동일한 질문을 가지고 두 시스템에서 검색되는 문헌의 수를 조사한 것이었다. 실험 결과 어느 시스템의 검색효율이 더 높은지 명확히 나타나지 않았지만 도전적이고 새로운 시도라는 이유에서 유니탐시스템이 주목받는 계기가 되었다(Swanson, 1988).

이러한 실험에는, 정보검색시스템은 두 가지 측정방법-‘적합문헌을 얼마나 잘 포착하는가?’와 ‘부적합한 문헌을 얼마나 잘 기각시키는가?’-에 의해 평가되어야 한다는 것이 내재되어 있었다. 그러나 이 생각은 1955년 Kent, Berry, Luehrs, Perry가 이 두 가지 측정방법을 재현율과 적합률로 정의한 후에야 비로소 공식화되었다(Swanson).

1950년대에는 자동분류, 자동초록, 자동색인 등에 관한 연구가 성행하였다. Taube의 뒤를 이어 정보검색의 기술적인 발전에 큰 역할을 한 사람은 IBM의 Luhn이다. 그의 관심은 주로 정보의 축적과 검색에 컴퓨터를 사용하는 문제였다. 1957년 Luhn은 컴퓨터를 이용하여 자동으로 원문을 분석하는 방법을 연구하였다. 그는 색인과 탐색에 대한 명확한 이해를 바탕으로 자동초록방법(정확하게는 extracting)을 개발하였는데 이 연구는 색인에 대한 관심을 불러일으키는 자극제가 되었다. 즉 주제를 부여하는 작업은 특별히 훈련을 받았거나 주제영역에 대한 지식을 가진 사람만 할 수 있다고 생각했었는데, 컴퓨터를 사용해서 매칭과 소팅작업뿐 아니라 문헌의 내용분석 등 지적작업도 할 수 있음을 처음으로 제시한 것이다. 그는 문헌에서 단어의 위치와 출현빈도 특성에 기초한 자동색인과 자동 가중치기법을 제시하였고 이어서 자동초록법을 제안하였다. Luhn의 자동색인 방법은 단어의 출현빈도에 근거해서 가장 빈번히 출현하는 단어를 그 문헌의 색인으로 결정하는 것이었다. 또한 그는 특정 단어의 출현양상이 반드시 고려되어야 함을 주장하였는데, 예를 들어 한 단락에 여러번 출현하거나 이어지는 단락에도 계속 출현하는 단어는 더욱 중요한 것으로 간주하는 방법 등이었다. 이러한 Luhn의 연구는 수작업색인시 색인작성자 간에 발생하는 불일치문제, 완벽한 색인을 작성하는 데 드는 비용 문제를 해결하는 하나의 대안으로 제시되었다. 오늘날 출현빈도에 의해서 단어의 중요도를 판단하는 방법은 많은 문제가 있음이 규명되었으나, Luhn의 연구는

원문 처리영역에 있어 가장 기초적인 이론이 되었다. 왜냐하면 자동문헌처리의 중요 영역인 문장론과 의미론에 대한 연구가 여기에서부터 시작되었기 때문이다. 한편 Edmunson은 단어의 상대적 출현빈도가 결정적인 요소라는 것을 제안한 최초의 인물이었다(Cleverdon 1970). 그는 문헌에 나타나는 각 단어의 출현빈도를 조사한 다음 더 자주 출현하는 단어가 더 중요한 단어라고 하였다.

이러한 일련의 자동색인 연구들은 컴퓨터 기능의 발전에서 비롯되었다. 즉 하드웨어 및 소프트웨어의 발달로 말미암아 이전까지는 수치데이터 처리에 주로 사용되었던 컴퓨터가 문헌처리 영역에서도 활용가능해진 것이다. Luhn은 또한 1958년 KWIC 색인법을 개발하였고 이것을 미국화학회가 받아들여 Chemical Titles에 적용되었다. 그는 KWIC을 활용한 SDI(Selective Dissemination of Information)를 제시하기도 하였다(Luhn 1958).

1950년대에는 이처럼 이전에 없었던 새로운 검색기술들이 소개되면서 기존의 내용분석에 사용된 방법들, 문헌과 질문의 주제분석방법에 대한 재평가를 불러일으켰다. 따라서 새로운 통제어휘집들이 간행되었으며 시소러스를 출판하게 되었다.

또한 이 시기에 중요한 인식의 변화가 나타났다. 색인과 내용분석을 위해서 용어통제를 반드시 해야 한다는 생각에서 벗어나 탐색법을 공식화하거나 검색시 부가적인 통제를 함으로써 이를 대처할 수 있다고 인식하게 되었다. 특히 통제어휘집에서 추출한 복잡한 주제어를 문헌에 할당하는 대신 단일어 디스크

립터나 키워드를 할당할 수 있음을 인식하게 되었다.

1956년 캠브리지대학은 자동번역과 정보검색의 유사성을 연구하였는데, 그 연구원 중 한 사람이었던 Spark Johns는 용어 간의 조합관계를 자동적으로 생성해 내는 데 주로 관심을 가졌다. 그녀는 색인 및 검색프로그램의 철저함과 색인언어의 정밀도 사이의 긴밀한 관계를 입증해냄으로써 정보검색에 대한 이해도를 증진시킨 연구결과를 발표하였다(Cleverdon 545). 그리고 1957년에 소련의 인공위성 스푸트닉 발사에 대한 미국의 쇼크는 정보검색분야에도 영향을 주어서 이 분야의 발전에 한 계기가 되었다.

### 3. 1960년대

1960년대는 정보검색 발전사상 여러 가지로 특기할 만하다. 우선 컴퓨터가 본격적으로 상용화되기 시작하면서 이용범위가 확장되었다. 오프라인 배치방식의 시스템들이 개발되어 운용되었고 실험 단계의 온라인 시스템도 등장하였다. 주로 서지정보를 대상으로 하는 후조합색인이 주류를 이루었다. 정보검색이 학문으로서의 방향을 분명히 제시한 것도 이 시기이다.

#### 3.1 서지정보의 기계가독화

1960년대 초반에는 대규모 출판사들이 색인·초록지 생산에 컴퓨터로 통제되는 사진식자장치를 사용하였으며 인쇄물로된 색인·초

록지 출판 후 그 부산물로 자기테이프에 수록된 서지데이터를 얻게 되었다. 이와 같이 생산된 자기테이프는 온라인시스템의 가장 중요한 요소인 데이터베이스의 유래가 되었다. 이 방식으로 생산된 최초의 색인지는 1964년 NLM(National Libraries of Medicine)의 Index Medicus이며 이것은 MEDLARS(Medical Literature Analysis and Retrieval System)에 의하여 생산된 것이다. 1965년 CAS(Cheical Abstracts Services)에서는 Chemical Registry System의 개발에 착수하였고 같은 해 초록이 포함된 기계가독형 POST(Polymer Science and Technology)를, 1967년에 CBAC(Cheical and Biological Activities)와 CA-Condensates를 각각 발행하였다.

이후 여러 색인지와 초록지들이 컴퓨터 식자방식에 의하여 간행되었고 이 과정에서 많은 데이터베이스가 출현하게 된 것이다. 1969년까지 생산된 대부분의 데이터베이스는 자기테이프에 서지데이터를 생산해 내는 방법을 취하였다.

### 3.2 오프라인 시스템과 실험적 온라인시스템

1960년대 들어 컴퓨터는 하드웨어와 소프트웨어 측면에서 획기적인 발전을 하였다. 즉 제2세대 컴퓨터가 출현하면서 부피가 소형화되었고 연산속도 및 기억용량이 대단히 개선되었다. 또한 응용프로그램도 급속한 발전을 하였다.

컴퓨터에 의한 배치방식의 검색이 처음 실

현된 것은 1954년의 일이고 그후 10여년 동안 특수도서관을 중심으로 배치탐색에 관한 기술이 개발되고 축적되어 왔다. 1960년대 전산센터나 정보센터에서는 운영비를 절약하기 위하여 탐색의뢰를 여러 건 모아서 일괄적으로 처리하였기 때문에 탐색의뢰후 수일 혹은 수주일 을 기다려야 결과를 입수할 수 있었다. 1964년 NLM에서 MEDLARS테이프를 이용하여 일괄탐색 서비스를 시작하였는데 이것이 일반대중이 사용할 수 있는 최초의 대규모 소급탐색서비스였다. NLM의 분석기는 탐색질문을 접수한 후 그 질문을 NLM 컴퓨터에서 탐색이 가능한 형태로 코드화하였다. NLM의 평균 응답시간은 약 2주일 정도 걸렸으나 이용자측에 결과가 입수되는 것은 약6주 후가 되었다. 탐색 공식이 적합치 않았거나 결과가 만족스럽지 못한 경우에는 같은 과정이 반복되었다. 일괄탐색은 몇몇 유형의 질문에 대하여는 수작업에 비하여 유용하고 결과도 월등히 좋았으나 즉시성 등 여러 문제가 있었다. 이러한 단점을 보완할 수 있는 대안으로 온라인 탐색이 개발되었다.

온라인 시스템은 배치시스템이 소개되기 몇 년 전인 1951년 미국 MIT의 Bagley에 의하여 그 기술적인 가능성이 소개되었다. (Bagley 1951) 그러나 컴퓨터 시분할이나 원격터미널 설비, 통신기술과 비용문제로 인해 보급되지 못하였고 본격적인 온라인 서지탐색 시스템은 1960년에 와서 실현되었다.

1960년에 SDC(System Development Corporation)는 ORBIT의 전신인 Proto-synthex를 개발하여 Golden Book Encyclopedia의 전문탐색을 시행함으로써 온라인

탐색을 처음 소개하였다. 이때 선보인 많은 기술이 현재까지도 온라인 탐색에 유용하게 사용되고 있다. 이 탐색에는 언어연구 영역이었던 질의응답프로그램이 응용되었으며 탐색 프로그램은 도치를 사용하였고 대화식 시분할 방식으로 근접탐색과 절단기법을 사용하였다. 이 탐색프로그램은 컴퓨터에 회선으로 연결된 단말기를 이용한 것으로 재호출이나 불 연산 기법 등은 첨가되지 않은 프로그램이었다.

1964년에는 Lockheed Missiles Corporation이 회사 자체 내의 데이터베이스 탐색용인 CONVERSE라는 온라인 시스템을 실험공개하였다. 내부용 시스템의 개발에 이어 1965년 록히드는 NASA(National Aeronautics and Space Administration)가 소장하고 있던 20만 건의 문헌검색을 위한 온라인 검색시스템 개발에 대한 계약을 체결하였다. 이 시스템은 1966년 NASA Ames Laboratory에 설치되었고 1967년에는 NASA 본부로, 1970년에는 전국에 있는 NASA 기관으로 확대되었다. 또한 1969년 독일의 컴퓨터센터에 연결되어 유럽 7개국에서 검색할 수 있게 되었다. 이 시스템이 후일 DIALOG로 발전된 것이다. IBM도 1966년 내부용으로 온라인 서지 검색 시스템을 개발하였으며 이것이 후에 IBM STAIRS시스템이 되었다.

1965년에 SDC는 미 국방성 ARPA(Advanced Research Projects Agency)의 지원하에 전국규모의 온라인 정보검색 네트워크를 구성하였는데, 이는 13개 정부기관 및 민간단체에 약 20만 건의 서지레코드를 온라인 서비스하기 위한 것이었다. 이 시스템에 사용된 소프트웨어는 ORBIT이었고 전국규모의

온라인 시스템으로는 최초의 것이다.

1965년에 MIT의 Technical Information Project는 물리학 관계 문헌 35,000건에 대한 온라인 서지검색 시스템을 개발하였다. 이 시스템은 탐색전략의 평가방법과 학술정보 교환 방법을 제공한 것으로 후속 온라인 시스템의 근간을 제시한 것으로 평가된다.

MDC(Mead Data Central)의 전문검색시스템의 전신인 OBAR(Ohio Bar)가 개발 운영되기 시작한 것은 1960년대 후반이었다. 1967년 OBAR는 Ohio Bar Automated Research의 온라인 전문검색시스템의 개발을 시작하였다. 1968년 Data사가 Mead사에 통합되어 MDC가 되고, 이 회사에서 개발한 온라인 시분할 전문검색시스템은 1969년부터 운영되었다. MDC가 개발한 이 시스템의 소유권을 OBAR는 MDC에게 넘겨 1973년부터 MDC가 LEXIS라는 서비스를 통해 온라인 법률 관련 본문검색서비스를 전국적으로 판매하기 시작하였다. 그리고 1967년에는 펜실베이니아대학, 시라큐스대학, 스탠포드대학과 미국물리학연구소에서 각각 온라인시스템을 개발하였다.

이와 같이 1960년대에는 बै치처리방식이 사용되면서 동시에 온라인 검색이 실험개발되는 등 온라인시스템으로의 길을 개척해 나갔던 시기였다.

1970년대 초까지 이어진 발전기 이후, 온라인 정보검색시스템의 이용은 미국과 유럽에까지 확산되었다. 그럼으로써 बै치탐색방식이 폐기되고 온라인으로 대체되었다.

### 3.3 시소러스의 발전

1960년대는 시소러스 연구와 개발이 매우 활발하게 이루어진 시기이다. 색인작성에 있어서 비통제어회를 사용하는 경우의 문제점들을 극복하기 위한 적극적인 시도의 결과 시소러스가 출현하였는데, 이는 역사적으로 주제명표목표를 발전시킨 것이다. 정보검색 분야에 시소러스라는 용어가 처음 사용된 것은 1957년 영국 도킹(Dorking)에서 개최된 '정보검색 분야에 관한 국제회의'에서 Brownson이 행한 강연에서 비롯되었다. 그는 한 문헌에 나타나는 여러 개념의 관련을 보다 정비된 용어로 변환하는 것이 정보검색의 문제이며, 그 해결법으로서 의미를 상호관련시키는 시소러스를 적용하는 것이라고 하였다. 그 후 이 용어를 널리 소개한 사람은 Luhn, Heald, Herman, Wall 등이다. 도킹회의 2년 후인 1959년에 미국방성은 시소러스가 기계검색에 필요불가결한 것이라고 결론짓고 'Thesaurus of ASTIA Descriptors'를 간행하였다. 이어서 1960년에 NLM이 MeSH (Medical Subject Headings) 제1판을, 1961년 미국화학공학회가 'Chemical Engineering Thesaurus'를 간행하였다. 1962년에는 미국방성이 개정판을 발행하게 되어 시소러스에 의한 색인방법이 인정을 받게 되었고, 1963년 미국대통령 과학기술자문회의에서는 기술계에 대하여 논문을 발표할 때 후일에 있을 검색을 쉽게 하도록 할 책임이 있으며, 이를 위하여 시소러스에 있는 키워드로 된 색인의 부기를 권고하기에 이르렀다. 그 후 시소러스는 전세계적으로 발전하게 되었다.

### 3.4 후조합색인에 관한 주요 연구

1960년에 Swanson은 사람이 작성한 색인과 컴퓨터를 사용하여 원문을 자동 분석하는 방법과의 검색효율을 비교연구하였는데, 이는 직접적인 비교 결과가 나온 최초의 연구였다. 핵물리학 분야의 100여 개의 짧은 논문을 대상으로 50개의 질문을 사용하였다. 그 결과 자동으로 원문을 분석하는 방법에 기반을 둔 시스템의 검색 성능이 사람의 손에 의한 것보다 더 우수한 것으로 나타났다. Swanson은 이에 대해 "그러나 실험결과 중 가장 눈에 띄는 것은 두 가지 방법 모두 관련정보의 비율이 상당히 낮다는 것이고 또한 주목할 것은 컴퓨터에 의한 검색방식이 전통적인 검색방식에 비해 명백히 우월하다는 사실이다"라고 보고서에 적고 있다. (Swanson, 1988) 뒤이어 Swanson은 자연어 질의문을 검색에 사용함으로써 그의 실험영역을 확대하였는데 여기서도 역시 자동으로 원문을 처리하는 시스템이 우수한 결과를 나타냈다.

1962년 Cleverdon은 미국과학재단의 연구비를 받아 색인언어의 성능을 비교하는 Cranfield Project를 수행하였다(Cleverdon 1970). 그 결과는 UDC, 주제명표목표, 파시트 분류, 유니텀 등 4가지 색인언어 사이에 중요한 차이점이 거의 없음을 보여주었다. 특히 통제되지 않은 단일어가 조절된 주제명표목만큼 효과적이라는 것이다. 또한 Cranfield Project는 조합(coordinate)이 정확률을 향상시킨다는 것과 사람의 손에 의해 조절되는 그 어떤 색인어보다 단일어가 더 효율적이라는 점을 증명해 냈다.

1960년부터 Salton은 SMART시스템을 연구하였다. 이는 자연어로 된 문헌을 탐색요구에 적합한 순서대로 목록해 내는 자동원문처리시스템이다. 그는 자연어로 된 초록이 색인을 만드는 것과 비슷한 성능을 얻을 수 있다는 것을 알아냈다.

### 3.5 학문으로서의 방향제시

1961-1963년에는 정보검색에 관한 주요 저서들이 간행되었다. Vickery의 'On Retrieval System Theory', Kent의 'Textbook on Mechanized Information Retrieval', Becker와 Hayes의 'Information Storage and Retrieval' 등이 이 시기에 간행되었는데 이들 자료는 정보검색에 관한 최초의 교과서로 미국은 물론 전세계의 대학에서 오랫동안 교과서 또는 지침서로 채택되어 왔다. 또한 이들 저서에 의하여 당시까지의 각종 기법이 정리되고 통일되었으며 정보검색이라는 용어도 마침내 정착되었다(사공철 1983).

## 4. 1970년대

이 시기의 특징은 온라인 검색의 확대, 데이터통신의 발전, 전문검색시스템의 등장 등을 들 수 있다.

### 4.1 온라인 검색시스템의 확대

1960년대 말과 1970년대 초기의 온라인 서비스는 실험용으로 개발된 시스템들이 한 기관이나 한정된 이용자를 대상으로 서비스하면서 기술 축적을 해나가던 시기였다. 그리고 1970년대는 여러 정보검색시스템을 통해 대규모 데이터베이스에 접근하는 것이 정보검색의 주류를 이루었다.

온라인 정보검색시스템의 운영 주체가 정부에서 상업적인 기관으로 전환된 것은 1960년대 말과 1970년대 초기 데이터베이스 벤더들에 의해 이루어졌다. 그들은 데이터베이스를 구입하고 재포장하여 온라인 서지검색서비스를 제공하였다. 1977년 Williams는 상업적 기관들의 온라인검색서비스의 성공을 발표한다 있다.

1971~1972년에 록히드사는 일반 대중에게 DIALOG시스템을 상용서비스하기 시작하였는데 대상 데이터베이스는 ERIC(Educational Resources Information Center), NTIS(National Technical Information Service), NAL(National Agricultural Library) 등 미국 정부기관이 제공하는 것 이외에 Engineering Index, Psychological Abstracts 등 민간기관에서 간행되는 것도 포함되었다. DIALOG는 오늘날 전 주제분야에 걸쳐 약 350개의 데이터베이스를 보유하고 100여국의 이용자에게 서비스를 제공할 정도로 성장하였다.

SDC의 ORBIT도 1972년 상업적인 온라인 서비스를 시작하였으며 1992년 이후 InfoPro Technologies사에 의해 운영되다가 1994년 프랑스의 QUESTEL사에 합병되어 기존의 QUESTEL 서비스와 병행하여 운영되고 있



다. 현재 QUESTEL/ORBIT는 약 150개의 데이터베이스를 제공하고 있으며 특히 과학기술과 특허정보 제공에 주력하고 있다.

1971년 MEDLINE(MEDLARS Online)은 미국 전체에 서비스를 시작하였으며 Index Medicus에 색인되는 모든 정보가 포함된 데이터베이스도 제공하였다. 이 당시 90여 개 이용기관에서 1975년에는 200여 기관으로 가입기관이 늘어났고 캐나다, 영국, 프랑스 등 해외로도 확장되어 국제적 시스템으로 발전하였다. 같은 시기에 Battelle Memorial Institute와 SDC도 온라인 검색서비스의 판촉 활동을 개시하였으나 Battelle은 후에 SDC에 통합되었다.

BRS는 1976년에 서비스를 시작하였는데 초기에는 미국 내에서만 서비스를 하였으나 후에 세계적인 규모로 성장하였다. 그러나 수차례에 걸쳐 운영권이 변경됨에 따라 그 명칭도 변화하게 되었다. 1989년에 Maxwell Online사에 매각되었다가 1992년 이후에는 ORBIT과 함께 InfoPro Technologies사에 의해 운영되었다. 1994년 3월에 다시 CD Plus Technologies사에 의해 운영되면서 CDP Online으로 명명되었다가 1995년에 이 회사가 OVID Technologies로 상호를 변경하면서 OVID Online이 되었다. 현재 약 80개의 데이터베이스를 온라인으로 제공하며 서지데이터베이스와 비서지데이터베이스를 모두 제공하고 있어 2차정보뿐 아니라 1차정보의 직접탐색에도 유용하다.

NLM에서는 AIM/TWX를 MEDLARS로 확대하고 1971년 그것을 MEDLINE(MEDLARS Online)이라 하였으며 그의 검

색시스템은 ELHILL이다. 1972년 NLM은 종래에 사용하던 텔렉스에 의한 서비스를 중지하고 Tymshare사의 새로운 공중회선인 Tymnet에 의한 서비스를 실시함에 따라 이용자 수가 폭증하게 되었다.

법률분야에서도 이 시기에 여러 상업시스템이 등장하였다. 1960년대 말 OBAR 시스템을 인수받은 Mead Data Central은 1973년 LEXIS를 통해 법률 및 판례의 전문탐색서비스를 전국으로 판매하기 시작하였다. 이어서 1975년 WESTLAW도 법률시스템으로 판매를 시작하였으며 1975년부터는 전문데이터베이스와 전문검색시스템을 추가함으로써 LEXIS와 경쟁하였다.

이 시기에는 OCLC(Online Computer Library Center)의 배치처리 방식이 시분할 방식으로 바뀌어 목록자가 직접 레코드를 생성할 수 있게 되는 등 온라인 목록이 대규모 학술도서관에 등장하기 시작하였다.

1972년 동양권에서는 처음으로 JICST(Japan Information Center of Science and Technology)가 '科學技術文獻速報'의 데이터베이스를 제공하기 시작하였고 1977년에는 JICST가 개발한 JOIS(JICST Online Information System)가 가동되었다. 처음에는 일본 내에서만 운영되다가 1985년 해외에서도 이용할 수 있게 되었다. JOIS에 의해서 '科學技術文獻速報'를 영어초록문으로 자동번역한 것이 JICST-E인데 해외에 있는 이용자들이도 탐색이 가능하게 된 것이다.

1975년 드디어 한국에서도 컴퓨터에 의한 정보검색이 실시되었다. 즉 한국과학기술정보센터(현재는 산업기술정보원)에서 CA-Con-

densates를 배치방식으로 SDI서비스하기 시작한 것이다. 이는 후에 온라인탐색으로 확대되었다. 1970년대 후반에는 온라인시스템인 TECHNOLINE이 개발되었고 1977년 영국에서 개발된 온라인시스템인 BLAISE에 대하여 서비스하게 되었다. 그리고 한국통신기술연구소(현재는 한국전자통신연구소), 한국전력주식회사, 한국원자력연구소(현재는 한국에너지연구소) 등이 INSPEC, INIS 등 외국의 데이터베이스를 입수하여 검색서비스를 시작함으로써 연구소를 중심으로 한 온라인 정보 검색 시스템이 본격적으로 가동되기 시작하였다.

1979년 말에는 전세계적으로 59개의 정보 검색시스템이 운용되고 있었고 이들 시스템에서 온라인으로 탐색이 가능한 데이터베이스도 400여 개나 되었다.

이와 같이 온라인 정보검색이 발전할 수 있었던 가장 중요한 요소 중의 하나는 데이터베이스 작성기관의 대형컴퓨터와 원거리에 있는 이용자의 단말기를 연결하는 저렴한 데이터통신 네트워크의 발전이라 하겠다. 제한된 이용자에게 제공되던 정보검색 시스템이 1970년대에 전국적인 규모로 발전한 것은 통신네트워크의 발달에 기인한다. 이는 패킷교환방식과 부가가치망의 보급에 의한 것이다. 1970년 미국의 Tymshare사에 의하여 개발된 부가가치망인 Tymnet를 비롯하여 패킷교환방식인 Telenet, 캐나다의 Datapac 그리고 1980년 유럽공동체 가맹 9개국이 공동개발한 EU-RONET(European Online Information Network)이 서비스를 개시하였다. 이 네트워크를 이용하는 각국의 온라인 서비스기관이

공동으로 행하는 정보서비스를 DIANE(Direct Information Access Network for Europe)이라 한다. 그후 각국에 많은 네트워크가 개설되었다.

한편 1970년대에 구미선진국에서는 정보네트워크가 일반화되어 도서관 소장정보의 상호 검색이 가능해졌다. 이미 1932년 미의회도서관에 협동목록과(Cooperative Cataloging Division)가 만들어지면서 목록에 관한 정보 교환이 미국 및 캐나다의 약 400개 기관으로 확장되었으며 컴퓨터와 통신기술의 향상에 힘입어 정보네트워크의 발전이 가속화되었다. 1966년 미의회도서관은 LC MARC(Machine Readable Cataloging)을 개발하였으며 이 LC MARC 포맷은 서지정보의 입력과 전송의 표준양식으로 자리잡았다. 또한 1967년 OCLC(Ohio College Library Center, 후에 Online Computer Library Center로 개칭)가 설립됨으로써 1970년대에는 본격적인 컴퓨터 네트워크시대가 시작되었다. OCLC는 LC MARC와 회원 도서관으로부터 입력된 목록정보를 기초로 목록 데이터베이스를 구축하여서 여러 도서관이 온라인 네트워크를 통하여 정보를 공유하도록 한 것이며 이에 자극받아 연구도서관 중심의 RLIN(Research Libraries Information Network), 워싱턴주의 WLN(Washington Library Network) 등과 같은 온라인 네트워크 들이 등장하였다.

#### 4.2 전문검색 시스템의 등장

이 시대의 또다른 특징은 과거 서지정보검색 일변도에서 벗어나 명감이나 통계, 전문데이터베이스로 수록정보의 형태와 검색시스템이 다양화된 점이다. 특히 전문검색시스템은 특징적인 것으로서, 참조검색이 서지사항을 확인한 후 원문을 요구하는 2단계를 거치는 번거로움이 있는 데 비해 전문검색은 한 과정으로 전문을 입수할 수 있는 편리한 방법이다.

법률분야를 비롯해서 신문이나 학술연구분야 등에서 전문데이터베이스와 전문검색서비스가 등장하였다. 즉 MDC는 1979년부터 LEXIS와는 별도로 온라인 본문검색시스템 NEXIS를 운영하여, 신문과 잡지의 전문검색을 제공하였다. 신문기사에 대한 대표적인 초기 온라인 시스템으로는 New York Times Information Bank(NYTIB)를 들 수 있다. 이것은 뉴욕타임스와 68종의 잡지 기사를 대상으로 초록 색인하여 탐색을 제공하는 목적으로 1971년에 시작된 시스템으로 1972년 신문사 외부의 유료이용자에게도 온라인으로 정보를 판매하였으나, 1983년부터는 NEXIS를 통해 뉴욕타임스의 전문탐색을 제공하고 있다. 1977년 캐나다의 신문 Globe & Mail지에 대한 전문검색서비스인 InfoGlobe도 등장하였다. 그리고 1978년 Harvard Business Review 온라인 데이터베이스를 시작으로 학술분야에 전문데이터베이스가 등장하였다(유재욱, 노정순 1996).

## 5. 1980년대

1980년대에는 온라인 정보검색시스템이 전 세계적으로 광범위하게 이용되고 국내에서도 해외 데이터베이스의 이용이 본격화된 시기이며 몇개의 국내 온라인시스템이 자체 개발되기 시작하였다. 또한 CD-ROM이 새로운 정보저장 매체로 등장하고 이를 통한 데이터베이스의 탐색이 보편화되면서 정보검색사의 또 다른 한 장이 시작된 시기이다.

### 5.1 전문데이터베이스의 증가

1970년대에 시작된 전문데이터베이스와 전문데이터베이스 검색시스템의 급격한 증가는 1980년대에 더욱 뚜렷한 특징으로 나타났다.

전술한 바와 같이 인쇄물 출판의 부산물로 쉽게 기계가독형 전문을 얻을 수 있다는 점은 데이터베이스의 구축에 결정적인 요인으로 작용하였다. 즉 1980년대 들어 PC나 워드프로세서, 컴퓨터를 사용한 인쇄법 등의 보급으로 많은 기계가독형 전문데이터베이스를 부산물로 얻을 수 있게 되었는데 그 수가 급격히 증가하였다. 즉 1979년에 4.2%(17개)였던 전문데이터베이스가 10년 만인 1989년에는 34%(1,381개)로 증가하였다. 이는 서지정보 뿐만 아니라 원정보의 직접검색이 가능해진 것을 의미한다.

DIALOG와 BRS는 1983년 이후 본격적으로 단행본과 학술잡지 및 일반잡지에 대한 전문검색서비스를 시작하였다. 대표적인 학술잡지 데이터베이스는 HBR(Harvard Business Review)로서, BRS는 1982년 이후, DIALOG는 1983년 이후 이에 대한 서비스를 개시하였다. 이외에도 40개 이상의 일반잡지와

기사 전문이 Magazine ASAP에, 80개 이상의 상업잡지 기사전문이 Trade & Industry ASAP에 수록되어 1984년 이후 DIALOG를 통해 검색되고 있다. 특히 주목할 것은 미국 화학회가 간행하는 학술잡지 18종에 대한 전문검색 서비스가 1983년 6월부터 BRS를 통해 ACS Journals Online이라는 이름으로 제공되고 있다(정영미 1993).

이와 같이 전문검색의 대상은 판례, 신문, 잡지 등의 1차 정보원이 주종을 이루나, 명감이나 백과사전과 같은 참고자료도 탐색이 가능하게 되었다. 1980년 11월 테스트된 OCLC의 채널 2000이라는 뷰데이터 시스템에 Academic American Encyclopedia가 등장한 이래 백과사전 역시 여러 네트워크시스템을 통해 검색할 수 있게 되었다.

## 5.2 온라인 인터페이스의 등장

최종이용자의 직접탐색을 돕기 위한 인터페이스의 등장도 1980년대의 특징이다. 온라인 인터페이스는 검색시스템에 대한 전문적인 지식과 경험을 가진 탐색대행자가 탐색을 수행하는 것이 아니라 정보를 필요로 하는 정보요구자가 직접 온라인 탐색을 수행하도록 시스템과 이용자를 연결시켜 주는 탐색보조 시스템이다. 1980년대에 실험용 온라인 인터페이스에 대한 연구가 활발하게 이루어지기 시작하였는데, 그 기능과 성능이 매우 다양하며 명칭도 중계자시스템(intermediary), 게이트웨이(gateways), 전단(front-ends)시스템, 투명(transparent)시스템 등 다양하게 사용된다.

이 시기에 Pro-search나 Wilsearch와 같이 이용자의 마이크로컴퓨터에 장착시켜 온라인 탐색에 사용하는 인터페이스가 상품화되었으며 EasyNet, Business Computer Network, Paper Chase, Einstein, Informaster 등과 같이 일단 이용자가 시스템에 접속하면 종료하지 않은 상태에서 다른 이용자를 연결시켜 주는 게이트웨이시스템도 등장하였다. EasyNet은 가입자에게 DIALOG, BRS, ORBIT, WILSONLINE 등 10여 개 시스템의 탐색을 연결시켜 준다. 각 시스템의 탐색 방법을 모르고 EasyNet 시스템언어로 탐색문이 입력되어도 각각의 검색시스템 언어로 자동변환되어 탐색된다.

DIALOG와 BRS도 이 시기에 온라인 탐색보조 인터페이스를 개발하여 제공하기 시작하였다. 즉 BRS는 1983년 After Dark, 1985년 Saunders Colleague, 1986년 Breakthru를, DIALOG는 1985년 Knowledge Index, 1987년 Business Connection과 Medical Connection을 제공하기 시작하였는데, 이들은 탐색서비스기관의 컴퓨터에 전단 소프트웨어를 탑재시킨 것이었다.

1980년대 초에 시작된 온라인 인터페이스에 대한 연구는 그 후 인공지능형 컴퓨터와 하이퍼텍스트 등의 기술개발로 꾸준히 성장하고 있다(유재욱, 노정순).

## 5.3 해외 데이터베이스 이용과 국내 정보 검색 시스템의 개발

1980년대에는 전문화된 데이터통신망을 통하여 해외 데이터뱅크의 이용이 본격화되었

다. 국내에서는 1970년대에 이미 해외 데이터뱅크의 이용이 시작되었으며 1980년대에는 한국데이터통신의 통신망인 DACOM-NET를 통해 미국의 DIALOG, LEXIS/NEXIS, 프랑스의 GSI-ECO와 QUESTEL, 일본의 JOIS 등이 연결되게 되었다. (사공 철 등 1990) 또한 1980년대에 우리 나라에서 온라인시스템을 자체 개발하였다. 1984년 산업연구원은 KIETLINE이라는 온라인시스템을 개발, 정보서비스를 실시하였으며 그후 KININTI-IR로 개칭, 운용하고 있다(현재는 산업기술정보원이 운영). 이 시스템에 포함된 데이터베이스는 CA Search, INSPEC, WPI, NTIS 등 해외 유명 데이터베이스와 자체제작한 국내의 산업, 경제, 과학, 기술 관련 정보 등 그 내용이 다양하다. 1985년 데이콤이 비디오텍스 시스템인 천리안을 개발한 이래, 한국PC통신의 HITEL, 에이텔의 POS-Serve, 코리아네트의 Info-Serve 등이 등장하였다. 그 밖에 중앙일보, 연합통신, 대한무역진흥공사 등에서 자체개발하여 운영되던 시스템이 일반인에게 온라인으로 탐색가능하게 되었다. 1986년에는 한국경제신문사가 KETEL시스템을 개발하여 온라인으로 서비스하기 시작했는데, 이 시스템은 자동색인 기법을 적용한 신문기사 전문 데이터베이스를 비롯하여 기업, 증권, 부동산, 세무, 유통, 문화오락 등의 정보를 제공하였다. 1992년부터 HITEL과 통합되어 현재까지 운영되고 있다.

#### 5.4 CD-ROM의 등장

기록매체로서의 획기적인 장을 연 CD-

ROM이 등장한 것도 이 시대의 특징이다. CD-ROM은 1983년 소니사와 필립스사가 개발한 것으로서 1986년에는 관련 12개 회사가 표준구조인 하이시에라 포맷(high sierra format)을 제안하여 그 이듬해 국제표준기구에서 국제표준규격(ISO 9660)으로 채택하였다. CD-ROM은 기억용량이 크고 신속한 액세스가 가능하며 경제적인 매체라는 장점으로 인하여 등장 이후 상당히 빠른 속도로 보급되기 시작하였다. 1985년 최초의 상업용 CD-ROM인 Biblio File이 등장한 이래 서지, 색인, 목록, 초록 등과 같은 2차자료와 백과사전, 사전류 등 참고도서, 방대한 기억용량을 필요로 하는 전문정보를 저장하는 데 중요한 매체로 부각되었다.

## 6. 1990년대

### 6.1 CD-ROM의 발전

전술한 바와 같이 1980년 중반부터 등장한 CD-ROM은 짧은 역사에도 불구하고 가장 광범위한 저장매체로 위치를 확고히 하고 있다. CD-ROMS IN PRINT 1995년판에 의하면, 1990년대 초 300여 종이었던 것이 1995년에는 3,800여 출판사에서 8,000여 종의 CD-ROM을 생산하고 있다. 미국의 경우 1991년에 이를 이용하는 정보서비스는 5%를 점유했으나, 1992년에는 약 9%로 증가하였다. 국내에서도 역시 봄을 이루어 유통물량이 1994년 80만 장, 1995년 5백만 장에 이어 1996년에는 8백20만 장에 달할 것으로 예상된다.

이와 같이 CD-ROM이 확산되는 이유는 온라인 연결에 따르는 비용, 통신료, 인쇄료를 지불하지 않고도 이용자 특히 최종이용자가 데이터베이스를 탐색할 수 있다는 점과, 멀티미디어 PC의 보급증가에 기인한다고 볼 수 있다. CD-ROM의 이용은 온라인 탐색을 감소시키는 것으로 알려지고 있다. 특히 ERIC, NTIS, MEDLINE 등의 데이터베이스가 CD-ROM으로 생산됨에 따라 이에 대한 온라인 탐색이 줄어드는 결과를 가져왔다(사공철 1994). 현재는 DIALOG와 같은 온라인서비스기관까지 CD-ROM 제품의 생산과 판매에 뛰어들고 있으며 앞으로 40% 정도의 데이터베이스가 CD-ROM으로 출판될 것으로 예측되고 있는데 이렇게 될 경우 온라인 서비스에 변화가 있을 것으로 예상된다.

1990년대에 들어와 국내에서도 CD-ROM이 등장하였다. 1991년 큐닉스컴퓨터가 CD-ROM 한 장에 영문성경 8종과 한글성경 1종, 관련서적 14권, 관련사진 1,100여 개를 담은 '성경라이브러리'를 최초로 내놓은 이후, 최근에는 여러 기관에서 다양한 자료를 CD-ROM으로 제작하고 있다. 그러나 많은 수의 초록이나 전문 데이터베이스를 CD-ROM화하는 외국과는 달리 국내에서는 게임이나 교육관련 타이틀이 다수를 차지하고 있다.

CD-ROM은 기존의 온라인 서비스를 통하여 정보를 이용하던 많은 이용자를 흡수하여 온라인 서비스와 더불어 정보서비스의 또다른 축을 형성할 것으로 예측된다. 그러나 한편으로는 인터넷의 등장으로 CD-ROM이 더 이상 성장하지 않을 것이라고 예측하는 전문가들도 있다.

## 6.2 멀티미디어 정보검색과 하이퍼텍스트 기술의 개발

1990년대에는 기존의 데이터베이스에서 처리되었던 숫자나 문자, 원문뿐 아니라 화상, 음성, 비디오 등 멀티미디어 데이터를 처리할 수 있는 멀티미디어 데이터베이스가 출현하게 되었다.

국내에서도 1995년 7월부터 데이콤이 고속 전용망을 기반으로 문자나 그림, 정치화상뿐만 아니라 움직이는 동화상과 음향까지 포함하는 멀티미디어형 서비스 '매직콜'을 시작함으로써 온라인 멀티미디어 정보서비스라는 새로운 장을 열었다. 최근에는 하드웨어 기술의 급격한 발전과 멀티미디어 기술의 발전에 힘입어 하이퍼미디어 시스템에 대한 연구가 활발히 진행되고 있다. 특히 인터넷의 급속한 발전은 하이퍼미디어에 대한 필요성을 더욱 고조시켰다.

하이퍼미디어 시스템은 노드와 노드를 연결해주는 링크들의 네트워크로 구성되며, 시스템 사용자는 노드와 링크조직을 따라 자유롭게 탐색항해(navigation) 함으로써 원하는 정보를 검색할 수 있게 된다. 따라서 하이퍼미디어 시스템은 매우 융통성있는 구조를 제공하고 인간의 인지모델과도 흡사하다는 장점을 갖는다.

Bush는 1945년 Atlantic Monthly에서 'As We May Think'라는 논문을 통해 하이퍼미디어 시스템의 고전이라 할 수 있는 Memex에 관한 아이디어를 제시하였다. Bush는 Memex를 책, 기록물, 정보 등을 마이크로필름에 저장하고 빠른 속도로 융통성있

게 사용하도록 해주는 기기라고 설명하고 있다. 그는 이 시스템에서 마이크로필름 간의 관련 관계를 묶어 주는 개념을 제시하였다. 이 개념은 하이퍼미디어의 핵심을 이루는 것으로서, 결국 Memex를 최초의 하이퍼미디어 시스템으로 자리잡게 하였다. Bush는 현대 정보검색의 시작과 하이퍼텍스트를 위한 아이디어의 기원이 되었고 후에 Nelson이 이를 발전시켰다(Lancaster, Warner 1993). 인터넷 열풍을 몰고온 WWW(World Wide Web)은 실질적인 최초의 글로벌 하이퍼미디어 시스템이다.

하이퍼미디어 공간에서 문서는 색인을 통하지 않고도 직접 브라우징만으로 접근할 수 있도록 연결되어 있다. 이러한 차이는 전형적인 정보검색 시스템에서는 색인을 통해 연결된 문서로 접근하는 반면에, 하이퍼미디어 시스템에서는 한 문서에서 다른 문서로 이동을 하면서 원하는 정보에 접근하게 된다. 직접 문서에 접근하여 정보를 볼 수 있다는 특징은 새로운 형태의 정보검색 기법으로, 이러한 기법의 장점은 접근한 문헌을 통해 좀더 적합한 문헌을 찾는 데 필요한 관련개념을 획득할 수 있다는 데 있다. 그러나 방대한 정보공간 내에서 단순한 브라우징만을 통해 접근한다는 것은 매우 어려운 일이므로 기존 정보검색 기법과의 결합을 통해 좀더 효과적인 접근을 시도할 수 있다.

### 6.3 지능형 정보검색시스템의 등장

1980년대 후반부터 인간의 사고과정 및 의사결정과정을 컴퓨터로 구현시키려는 지능형

정보시스템, 지능형 컴퓨터 등이 본격적으로 등장하였다. 문헌정보학에서는 인공지능의 한 영역인 전문가시스템 기법을 도입하여 지능형 정보검색을 구현하려는 연구가 활발히 진행되었다. 즉 참고사서, 색인 전문가, 분류 전문가, 온라인 탐색 전문가 등 특정한 정보처리 업무를 수행하는 전문인을 대신하거나 보조할 목적으로 전문가시스템들이 개발되었는데, 이 시스템들의 특징은 전문적인 지식을 담고 있는 지식베이스와 추론기능을 제공하는 지식기반 시스템이라는 것이다.

문헌정보학 영역에서의 지능형 정보검색시스템이란 용어는 Spark Johns가 처음 사용하였는데, 그는 지능형 정보검색시스템을 이용자가 잘못 기술한 질의어와 이용자가 실제로 원하는 적합문헌의 후보 사이를 추론기능을 통하여 검색하는 시스템이라 하였다(Spark Johns 1983). Brooks의 정의는 좀더 구체적이며 특히 이용자에 관한 지식의 활용을 강조하고 있다. 즉 지능형 정보검색시스템은 주제지식, 문헌에 관한 지식(색인), 이용자에 관한 지식을 소장하며, 개별적인 이용자에 관한 정보와 이용자가 해결하고자 하는 정보가 입력되면 위의 지식을 이용하여 문제해결에 필요한 문헌들을 추론하여 검색하는 컴퓨터시스템이라고 하였다(Brooks 1987).

현재 지능형 정보검색에 관한 연구는 크게 두 분야로 나뉘어서 수행되고 있는데, 하나는 정보검색을 응용대상으로 한 인공지능의 각 연구영역(자연언어처리, 지식표현, 추론 등)에서의 기본적인 연구이고, 다른 하나는 전통적인 정보검색기법과 인공지능 분야의 기법을 접목한 정보검색시스템에 관한 연구이다. 범

용성을 가진 지능형 정보검색 시스템은 아직 개발되지 못한 상태이지만 멀지 않은 장래에 실현될 것으로 전망된다.

#### 6.4 인터넷의 발전

1990년대 온라인 정보검색의 가장 대표적인 특성은 인터넷이라고 할 수 있다. 이것은 1969년 미국방성이 4개의 호스트컴퓨터를 연결하여 사용하던 ARPANET이라는 네트워크에서 유래되었으며 그동안 확장과 대체를 거듭하여 오늘날 우리가 사용하는 세계적인 규모의 기간통신망이 된 것이다. 인터넷은 TCP/IP 통신프로토콜을 사용하는 컴퓨터는 기종에 관계없이 서로 통신할 수 있는 전산망으로, 전세계에 산재한 많은 데이터베이스를 몇초 안에 온라인으로 접근할 수 있다. 인터넷은 그 자체가 통신망이면서 또한 Gopher, Wais, Web 등과 같은 정보검색 서비스도 제공한다. 이런 것들은 이미지, 오디오, 비디오에 대한 검색을 제공함으로써 온라인 정보검색의 새로운 장을 열고 있다. 특히 문자 위주였던 기존의 2차원적 서비스에서 벗어나 멀티미디어, 다중접속(multi-session), 하이퍼미디어(hypermedia)로 입체화된 멀티미디어 서비스를 표현하는 데 있어서 Web방식은 이미 대중화의 길에 접어들었다. 1990년대에는 새로운 경쟁자로 등장한 인터넷의 급성장이 전통적 온라인 시스템에 많은 영향을 미칠 것이다. 현재 전세계 2백여 개 국가에서 약 6~7천만 명의 PC 이용자들이 접속하는 것으로 추정되고 있다. 국내의 경우도 1994년 6월 한국통신, 1994년 11월 나우콤을 시작으로 현재

20여 곳이 접속서비스를 제공하고 있으며 연결 호스트도 7만 대에 이른다.

## 7. 장래의 전망

이상으로 지난 50년간 정보검색분야의 연구와 활용, 변화와 성장을 살펴보았다. 검색의 기본 원리에서부터 기술적인 측면에 이르기까지 큰 발전을 이루어왔으며 그것은 지금도 계속되고 있다. 더욱이 정보검색은 컴퓨터 하드웨어 및 소프트웨어 기술 그리고 통신 기술의 발전에 영향을 받는 점을 고려할 때에 21세기에는 새로운 역사가 시작된다고 할 만큼 비약적인 발전을 하리라고 예상된다. 정보검색의 장래를 몇 가지로 정리해 보면 다음과 같다.

첫째, 지난 10년간 꾸준히 증가해 온 데이터베이스의 수가 앞으로는 성장률이 서서히 감소할 것으로 예측된다. 물론 데이터베이스의 수는 많아지겠지만 이용자의 정보요구를 충족시킬 만큼 전 주제영역에 걸쳐 양질의 데이터베이스가 생산되는 성숙기에 이르면 데이터베이스 품질에 대한 중요성이 커지고 아울러 비경제적이고 비효율적인 데이터베이스는 소멸되리라고 예상된다. 또한 특정 이용자, 특정 주제를 대상으로 하는 전문 데이터베이스는 매우 많아질 것이다.

둘째, 탐색측면을 보면, 미래의 정보검색은 이용자 지향적이고 지능을 갖춘 지능형 정보검색이 될 것이다. 현재 전세계적으로 이용가능한 상용 온라인 정보검색시스템은 약 800개에 이르는데 이 시스템들은 대부분 각각의 이용자-시스템 인터페이스를 제공하고 있다. 이



용자들은 어떤 정보를, 어떤 방법으로, 또 어떤 정보망을 통해 찾아야 효율적인지를 알아야 하는데, 이러한 문제에 대한 해결방법으로 사용자 위주의 인터페이스가 상당히 발전하리라고 예상된다. 따라서 정보를 필요로 하는 최종 이용자가 편리한 인터페이스 환경에서 직접 탐색을 수행하는 이용자 직접탐색이 증가하게 될 것이며 이용자는 자기가 알고 있는 하나의 명령어를 가지고 어떤 호스트이든 접근할 수 있게 될 것이다. 또한 앞으로 지능형 컴퓨터가 발달하게 되면 자연어 및 멀티미디어 처리가 가능해지고 연상과 브라우징에 의한 검색 및 기존정보를 이용할 수 있는 추론 기능, 학습기능 등이 가능해져서 지능형 정보 검색을 수행하게 될 것으로 전망된다.

셋째, CD-ROM과 온라인 데이터베이스의 상호보완관계를 예측할 수 있다. CD-ROM의 급속한 증가와 아울러 온라인 데이터베이스에 대한 CD-ROM 버전의 생산도 계속 증가할 것이다. 그러나 CD-ROM이 갖는 몇 가지 단점 - 수록정보의 최신성 결여, 수록범위의 부족 등과, 반대로 경제적인 축적매체라는 장점으로 인하여 온라인 데이터베이스와 CD-ROM 데이터베이스를 상호보완적으로 이용할

것이다.

넷째, 최근 폭발적으로 증가하고 있는 인터넷 이용은 정보검색 분야에 많은 영향을 미칠 것으로 예상된다. 즉 대부분의 상용 정보서비스에 원격 접속이 가능하게 됨에 따라 온라인 정보검색 시장에 변화를 가져올 것이다. 인터넷을 통해 뉴스나 잡지 기사를 무료로 이용할 수 있으며 DIALOG, ORBIT, OCLC 등 온라인 정보검색시스템들이 www를 통해 데이터베이스 목록, 상품 및 서비스, 명령어 등에 관한 정보를 제공하고 있다. 이러한 정보들은 앞으로 더욱 늘어날 전망이어서 유사한 정보를 유료로 제공하고 있는 온라인 정보시장에 타격을 줄 것으로 보인다.

이렇게 볼 때 장래의 정보검색은 멀티미디어 데이터베이스에 대한 온라인검색, 인터넷, 통합적 네트워크 정보시스템, 다양한 정보저장매체를 대상으로 한 정보검색 등 여러 측면이 혼재하게 되리라고 전망된다. 과학기술이 발달을 거듭하는 한 기록된 지식의 세계를 탐험하고 그 가운데 필요한 정보를 획득하고자 하는 인간의 노력은 계속될 것이며 무엇보다도 인간의 정보검색 방식과 유사한 시스템을 찾으려는 연구가 앞으로도 계속될 것이다.

## 참 고 문 헌

- 김경창 (1996). "멀티미디어 데이터베이스에서의 정보검색". 데이터베이스월드. 96. 7, p.16-23.
- 김성혁 (1992). "문헌정보학 영역의 지능형 정보시스템에 관한 고찰". 정보관리학회지. 9(1), p.165-180.
- 김윤실 (1996). "원문단체자료로서 뉴미디어 도입의 효과분석." 정보관리연구. 27(2), p.1-32.
- 김현희, 배금표 (1993). "퍼지정보검색시스템

- 의 검색효율에 관한 연구”. 정보관리학회지. 10(1), p.31-52.
- 김휘출 (1996). “www을 이용한 데이터베이스 검색”. 데이터베이스월드. 96. 10, p.20-24.
- 사공철 (1983). 정보검색론. 서울 : 아세아문화사.
- (1994). “온라인 시스템의 역사적 발전”. 정보관리학회지. 11(2), p.111-130.
- 사공철 등 (1990). 최신정보검색론. 서울 : 한국도서관협회.
- 신용태 (1996). “인터넷”. 국회도서관보. 33(2), p.3-16.
- 유사라 (1996). 정보화사회와 도서관 정보네트워크. 서울 : 나남출판.
- 유재욱, 노옥순 (1996). 온라인 탐색. 서울 : 한국도서관협회.
- 이만재 (1996). “데이터베이스 유통매체로서의 CD-ROM”. 데이터베이스월드. 통권35 호, p.22-27.
- 정영미 (1993). 정보검색론. 서울 : 구미무역.
- 정영미, 유재복, 김태환 (1996). 온라인 정보 검색. 서울 : 교보문고.
- 최윤철 (1996). “하이퍼미디어와 정보검색”. 데이터베이스월드. 96. 7, p.8-15.
- 한국정보관리학회 편 (1995). 정보학의 실제. 서울 : 한국도서관협회.
- 도서관 정보관리편람 (1994). 서울 : 한국도서관협회.
- Bourne, C. P. (1963). Methods of Information Handling. New York : Wiley.
- (1980). “On-line systems : history, technology, and economics”. Journal of the American Society for Information Science. May, 1980, p.155-160.
- Brooks, H. M. (1987). “Experts Systems and Intelligent Information Retrieval.” IPM, 23(4), p.367-382.
- Cleverdon, C. W. (1970) “Review of the Origins and Development of Research : 2. Information and its Retrieval”. the 44th Aslib Annual Conference, University of Aberdeen, 20th-23rd September, p.540.
- Convey, John (1989). Online information retrieval. London : Library Association Publishing Ltd.
- Enser, P. G. B. (1995). “Progress in documentation pictorial information retrieval”. Journal of Documentation. 51(2), p.126-170.
- Gilchrist, Alan (1996). “Blissful spiders : forty years of information retrieval”. Managing Information. 96. 5, p.35-37.
- Jones, Spark (1983) “Intelligent Retrieval.” Intelligent Information Retrieval : Informatics 7. London : Aslib.
- Khoshafian, Setrag and Baker, A. Brad (1996). Multimedia and

- imaging databases. San Francisco : Morgan Kaufmann Publishers.
- Lancaster, F. W. (1968). *Information Retrieval Systems : Characteristics, Testing and Evaluation*. New York : Wiley.
- (1979). *Information retrieval systems*. New York : John Wiley & Sons.
- Lancaster, F. W. and Warner, A. J. (1993). *Information retrieval today*. Arlington : Information Resources Press.
- Lilley, Dorothy B. and Trice, Ronald W. (1989). *A history of information science 1945-1985*. San Diego : Academic Press.
- Luhn, H. P. (1958). "A Business Intelligence System", *IBM Journal of Research and Development*, 2 (4), p.314-319.
- (1959). *The Keyword-in Context Index for Technical Literature*. York Town Heights : IBM Corporation.
- Meadow, C. T. (1992). *Text information retrieval systems*. San Diego : Academic Press.
- Salton, Gerard (1987). "Historical note : the past thirty years in information retrieval". *Journal of the American Society for Information Science*. 38(5), p.375-380.
- Swanson, Don R. (1988). "Historical note : information retrieval and the future of an illusion". *Journal of the American Society for Information Science*. 39(2), p.92-98.