

## 전세계 DB산업 현황

### (The State of Databases Today : 1993) ①

이 글은 美 Gale Research Inc. 에서 전세계적으로 유명한 3개 DB디렉토리를 종합하여 발간한 『Gale Directory of Databases』 ('93년 1월판) 중에서 세계적으로 우수한 DB 7,907개를 분석한 일리노이대 마르타 E 윌리엄스의 글을 번역, 요약한 것으로 앞으로 3회에 걸쳐 게재할 예정이다.  
 <편집자주>

#### 개요 : 데이터베이스 레코드, 온라인 검색, 데이터베이스, 정보제공자 그리고 유통자의 숫적 변화 추이

이 글에서는 전세계적으로 7,097개의 DB와 아울러 그중에서 고유의 부속화일을 가진 6,998개의 DB(Entry)를 분석대상으로 삼았다. (여러개의 DB와 부속화일이 하나의 DB로 취급되기도하기 때문에 이 글에서는 이들을 'Entry'라고 칭함) 그리고 어떤 DB들은 CD-ROM이나 다른 전자미디어 형태의 DB로 중복취급되어 이같은 오프라인(off-line) DB를 합하면 8,495개가 된다. 이들중 375개는 실제로 이용가능한 화일형태를 유지하고 있지 못하기 때문에 제외시켜야 한다. 따라서 진정한 유통DB는

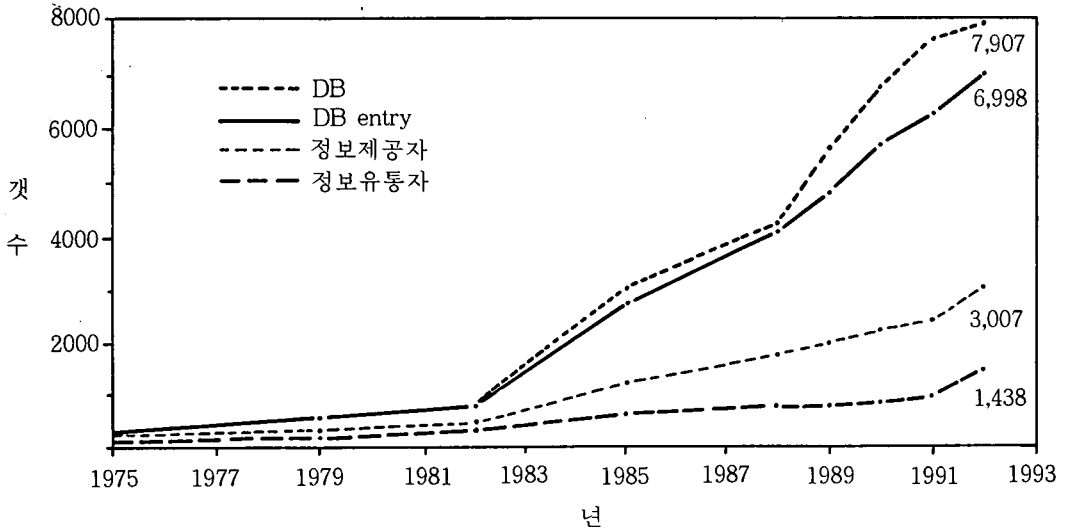
8,120개라고 볼 수 있고 이 글에서는 온라인 형태의 Entry DB인 6,998개를 주 분석 대상으로 하였다.

DB산업의 규모는 DB 레코드, 온라인 검색 횟수, DB, DB entry, 정보제공자 그리고 유통업체의 양적 규모에 의해 설명될 수 있다. (<표 1,2,3,4>참조)

표1은 DB entry, DB, 정보제공자 그리고 유통업체수의 증가동향을 보여준다. 표2에 나타난 개별 DB의 수는 DB entry의 수에 다 각 entry내에 포함된 DB들을 합쳐놓은 것으로 DB entry의 수와 함께 막대그래프로 나타나 있다. 정보제공자의 수는 개별 DB entry의 수와 일치한다. 정보제공자들 중에는 CD-ROM, 디스켓등의 다른 미디어 형태로 정보를 제공하는 업체들이 많아서 표1의 정보제공자에서는 중복된 부분은 제외했다. 정보제공자/유통업자를 포함한 유통자의 총수도 표1에 함께 나타나 있다.

DB관련 여러부문의 숫적변화를 통해서 볼 수 있듯이, DB산업의 성공은 '종이중심의 서비스'에서 'PC중심의 서비스'로 정보산업이 변동된 추세변화의 결과이며 컴퓨터로 검색가능한 DB의 이용도와 그러한 이용에 따른 검색횟수를 통해 측정될 수 있는 것이다. (<표4> 참조)

<표1> DB유통업체, 정보제공자, DB entry, DB의 숫적 변화



그래프자료 (단위 : 개)

년	유통업체	정보제공자	DB entry*	DB*
1975	105	200	301	301
1979	263	316	528	528
1982	311	422	773	773
1985	614	1,210	2,700	3,010
1988	750	1,733	4,042	4,200
1989	770	1,950	4,786	5,478
1990	850	2,224	5,689	6,750
1991	933	2,372	6,261	7,637
1992	1,438	3,007	6,998	7,907

\* 1개이상의 매체로 서비스되는 DB entry는 한번만 계산됨.

이 분석을 통해 알게 된 것은 DB 레코드가 최소한의 단위이고 그 다음으로 DB entry, 개별 DB, 온라인 검색, 정보제공자 그리고 유통자순으로 단위 혹은 실체가 커지는데, 작은 단위일수록 증가속도가 빨라

서 DB 레코드의 숫적 증가속도가 가장 빨랐다는 것이다.

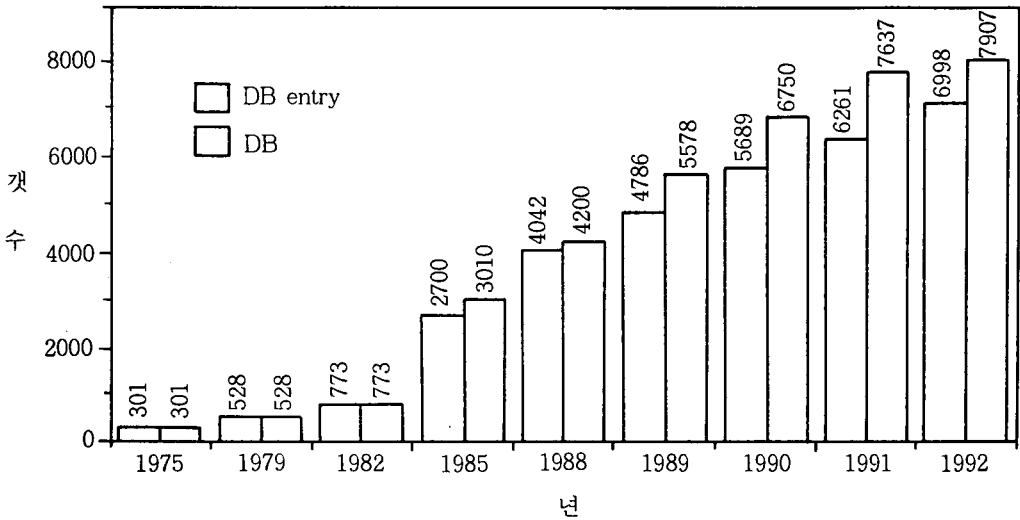
### 데이터베이스 레코드 (Database Records)

1975년부터 1992년까지 17년동안, DB 레코드수는 5천2백만개에서 45억2천7백만개로 87배가 증가했다. 반면, DB는 301개에서 7,907개로 26배가 증가했으며, DB entry수는 301개에서 6,998개로 23배 증가했다.<sup>1)</sup> 정보제공자의 수는 개별정보제공자가 다수의 DB를 개발하고 있어 증가속도가 다소 느렸다. 정보제공업자는 200개에서 3,007개로 15배 증가했고, 업체당 평균 2.6개의 DB를 제공하고 있는 셈이다. 유통

1) DB entry과약에서 국내DB로는 한국경제신문의 '경제뉴스' 1개만 포함되었음.

<표2>

DB entry 및 DB의 숫적증가 동향



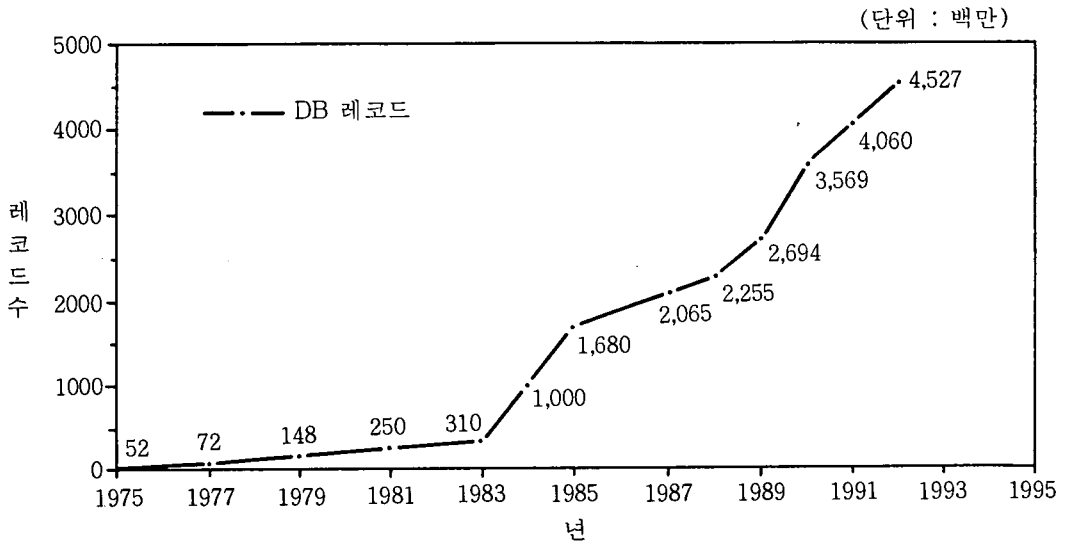
업자는 수많은 DB서비스를 제공하고 있기 때문에 비교적 느린 속도로 증가하고 있다. 실제로, DIALOG와 같은 유통업체는 수백개의 DB를 제공하고 있다. DB유통업체의 수는 CD-ROM제품 유통자의 빠른 숫적증가에 힘입어, 1991년과 1992년 사이에 급격하게 증가했다. 그리고 많은 경우 CD-ROM생산업체는 한가지 제품만을 유통시키고 있다. 지난 17년동안 유통업체의 수는 104개에서 1438개로 13.7배가 증가했다.

DB 레코드 수의 증가율은 DB수의 그것과 일치하지 않았다. 왜냐하면, 한 DB의 평균크기가 일정하게 유지되지 않았기 때문이다. DB레코드의 증가는 그 숫자가 3억 1천만개가 된 1983년까지는 비교적 느렸다. 그런데 1983년과 1984년의 1년 사이에, 3배 이상에 해당하는 거의 10억개로 늘었고, 1984년부터 1987년사이에는 2배인 20

억개로 늘었다. 그리고 1987년부터 1992년 사이에는 2.19배가 늘어 40억5천3백만개에 도달했다. 한 DB당 평균레코드수를 보면, 1975년에 173,000개, 1985년에는 거의 500,000개에 달했다. 1992년 DB entry당 평균레코드 수는 647,000개이고 개별 DB(discreet DB)당 평균레코드수는 573,000개였다. DB 평균크기에 대한 이와같은 도표들은 상당부분 왜곡되어 있는데, 왜냐하면, 전체에서 불과 288개의 DB가 100만개이상의 레코드를 차지하고 있고, 그 288개중 10개는 1억개이상의 레코드를, 앞선 10개를 제외한 26개는 1천만개와 1억개사이의 레코드를 차지하고 있기 때문이다. 그리고 252개는 1천만개 보다는 적지만 적어도 100만개를 차지하고 있다. 288개의 매우 규모가 큰 DB를 제외하면, 1992년 DB Entry당 평균레코드수는 125,000개이고, DB당 평균은 111,000개이다. 것처럼 규모

<표3>

DB레코드수의 증가 동향



가 큰 DB를 통계에 포함시키느냐 아니냐에 따라, DB 즉 '일반적인' DB당 평균레코드수는 달라질 수 밖에 없다.

DB레코드로 간주되는 대상은 범위가 넓지만, 대체로 200-2,000단어사이에 해당된다. 레코드는 인용문, 초록, 소식, 일대기, 고유화학물명, 화학구조식, 재무정보, 요리법, 시계열표, S/W 프로그램, 이미지, 혹은 실제적인 어떤 것의 설명이나 목록등의 어느 한가지 형태를 가진다.

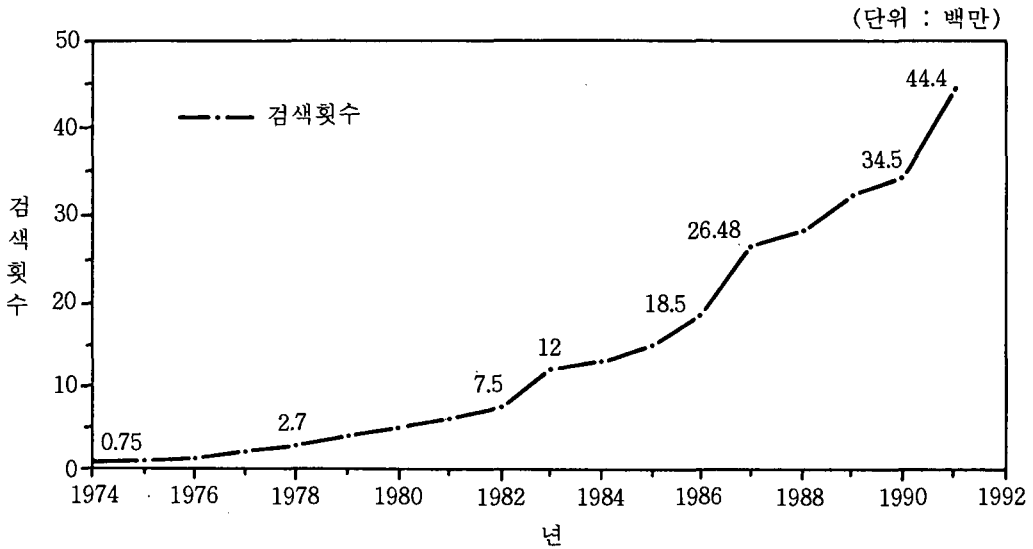
DB의 평균크기 증가에 영향을 미치는 가장 큰 요인은 아마도 1980년대 중반부터 시작된 상용 비즈니스정보의 증가에 있다고 볼 수 있다. DB평균크기의 증가는 1980년대 중반부터 시작된 규모가 큰 시계열 DB의 등장과 전화 Directory DB, 신문 DB의 출현에도 원인이 있었다. 규모가 큰 DB중에는 차지하는 빈도수가 줄어드는 DB들이 있는데, 그것은 서지정보(참고, 초

록, 혹은 全文); 기업정보; 시계열정보; 특허정보; 전화와 주소목록; 인사, 인명록 그리고 고용관계 목록; 화학자료; 인구자료; 뉴스; 어휘정보; 상표; 그리고 조달과 계약정보등이 그것이다.

### 온라인 검색(Online Searches)

DB, 정보제공자, 그리고 유통자의 수적 증가를 나타내는 통계를 보면, DB산업이 어느정보 자리매김하고 있음을 보여주지만, 컴퓨터를 통한 접근이 가능한 정보의 이용과 그러한 이용을 위한 기술의 발전에 의한 진정한 성공은 DB의 이용활성화를 통해 나타난다. 또한 DB의 이용도는 검색횟수(연결시간)와 검색비용에 의해서 측정될 수 있다. 그런데 아쉽게도 전체DB산업에 대한 관련자료는 얻기 어렵고 단지 대강의 추측만이 가능할 뿐이다.

<표4> 미국의 주요한 문자중심 DB (Word-oriented DB) 유통 업체에 있어서의 온라인DB 검색횟수 증가동향



그래프자료

(단위 : 백만)

연도	검색횟수	연도	검색횟수
1974	0.75	1983	12.00
1975	1.00	1984	13.00
1976	1.20	1985	15.00
1977	2.00	1986	18.00
1978	2.70	1987	26.48
1979	4.00	1988	28.31
1980	5.00	1989	32.30
1981	6.00	1990	34.50
1982	7.50	1991	44.40

그와 관련해서 기존에는 Information Market Indicators, Inc. (IMI)를 통하여 주요한 문자 중심의 DB유통업체, (예를 들면, Mead Data Central, West Publishing

Company, DIALOG Information Service, ORBIT Search Service, BRS Information Technologie 그리고 U.S. National Library of Medicine)에서 제공하는 DB이용에 관한 자료가 만들어져 왔다. 1982년부터 현재까지의 자료들은 IMI를 통하여 만들어져 왔고, 그 이전의 자료는 이미 발표된 여러 자료를 혼합하여 유통업체들이 제공하여 왔다. <표4>는 1974년의 75만번 검색에서 1992년에는 년당 4억4천4백만번 검색으로 까지 증가해온 미국 온라인네트워크를 통한 DB이용의 성장 추세를 나타내고 있다. 만약 경제거래DB(주식, 전자주문, 신용조회, 항공기예약등)와 소비자 서비스체계를 통한 검색을 여기에 추가한다면, 검색횟수는 훨씬 늘어날 것이다. 그런데 불행하게도 그러한 자료는 존재하지 않는다.

유통업체, 정보제공자, DB entry, 그리고



DB는 표1에 함께 나타나 있는데, 여기에서 몇가지 다른 매체형태로 존재하는 DB수는 제외시켰다. 검색횟수를 보여주는 <표4>와 레코드수에 관한 <표3>은 서로 매우 다르기 때문에 유통업체, 정보제공자, 그리고 DB와 함께 같이 도표에 나타낼 수 없다. <표3>은 DB레코드수의 증가동향을 나타낸다.

1989년 자료부터 시작해서, DB레코드의 총갯수(알려진 레코드수+미확인된 DB에 대한 추정 레코드수)는 미확인 자료에 대한 추정방법에 따라 많이 달라진다. 추정이라는 방법은 백만개 이하의 레코드를 가진 DB들의 평균갯수에 따라 행해진다. 여러해동안 기존 및 신규DB를 막론하고 DB 크기에 대한 공개자료를 기초로 하면, 추

정의 방법이 좀더 정확한 수치를 반영할 수 있을 것이다. <표1>과 <표3>은 전 세계적으로 유통되는 모든 형태의 공용 DB에 대한 자료를 나타내는 반면, DB검색 및 이용의 성장현황을 나타내는 <표4>는 <표1>에 나타난 DB의 한부분, 예를 들면 정보센터/도서관에 있는 문자중심의 DB에 대한 미국의 이용현황과 같은 일부분에 한정된 현황을 나타낸다. 이것은 전체DB 이용현황의 한부분에 한정된 반면, DB<표1>, DB레코드<표3>, DB이용현황<표4>의 곡선모양은 아주 유사하다. 다른지역에서 사용하거나, 거래DB같은 다른 유형의 DB들이 추가된다면, 곡선의 모양은 그대로 이겠지만, 숫자는 비례적으로 더 커질 것이다.