

The State of Database Today:1994

세계 데이터베이스 산업을 잘 이해하기 위하여 1970년대부터 1992년까지 유통된 컴퓨터가독형 데이터베이스를 수집하였으며, 1993년에는 Gale Directory of Database(이하 GDD)에 포함된 여러 통계를 분석해 보았다. 이 글은 컴퓨터가독형 DB뿐만 아니라 GDD에 나타난 통계를 비교분석하여 DB산업의 '94년 동향을 살펴보고 그 특징을 알려주려는 데 목적이 있다.

이런 데이터베이스의 통계를 계속 활용하고 전체적인 데이터베이스 비교통계를 만들기 위해 데이터베이스 엔트리들을 특별히 등록하였다.

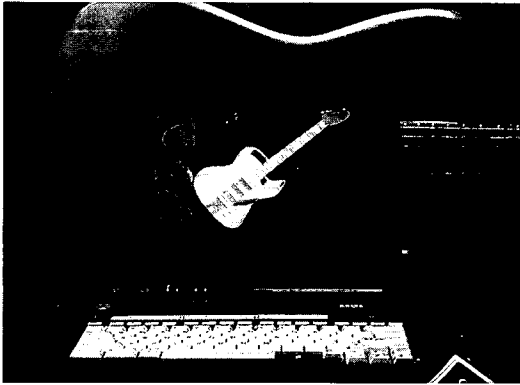
이 글은 과거의 데이터베이스산업이 지역별, 산업특성에 따라 어떻게 변화되어 왔는가를 통계별로 설명하고 있다. — 편집자 주 —



서문

1 994년에 발간된 디렉토리를 보면 8,261개의 데이터베이스와 7,538개의 서브파일들이 있다.(여기서 엔트리는 단일 데이터베이스가 아니라 서브파일의 집합과 데이터베이스군을 의미한다. 데이터베이스는 때로는 파일이라고도 하며 데

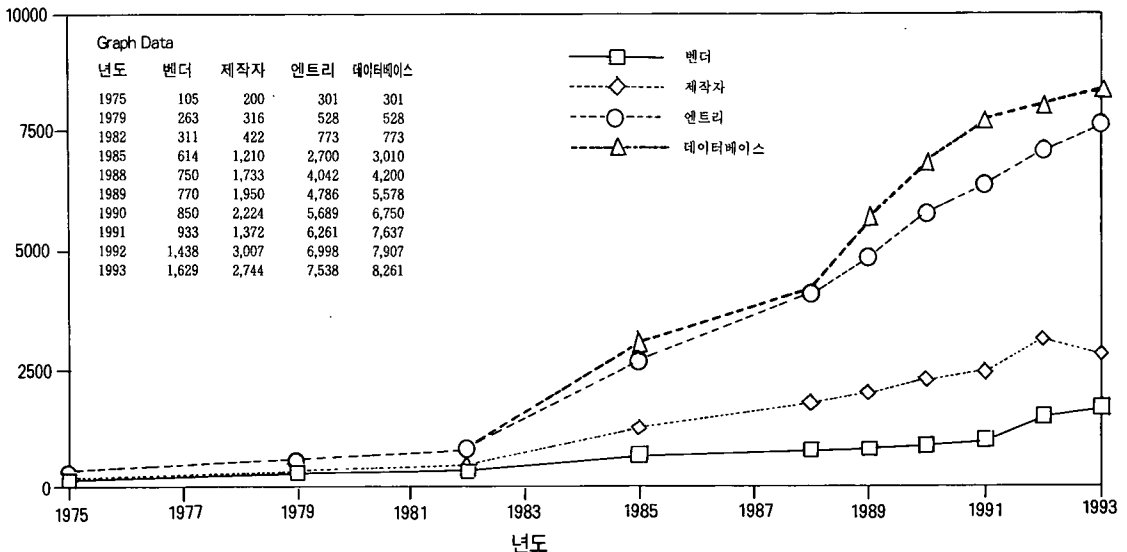
이터베이스군은 서브파일이라고도 한다) 몇개의 데이터베이스 엔트리들은 온라인서비스나 CD-ROM제품에 저장되어 중복되어 나타나기도 하는데 그러한 숫자는 9천1백3십6개에 이른다. 9천1백3십6개의 미디어에 실린 엔트리들중 4백1십1개는 없어진 DB인데 없어진 DB는 제작자들이 더



이상 파일을 갱신하지 않거나 공개적으로 활용할 수 없는 파일을 말한다. 411개의 없어진 DB들을 빼고나면 8천7백2십5개의 데이터베이스엔트리가 남는다. 1번이상 수록된 데이터베이스들은 모두 1개의 수로 계산했다. 도표(1,2,3,4)에서 보듯이 데이터베이스산업의 규모는 데이터베이스레코드, 온라인검색, 데이터베이스, 데이터베이스엔트리, 데이터베이스 프로듀서, 벤더들의 크기로 나타내어질 수 있다. 데이터베이스엔트리, 데이터베이스 프로듀서, 벤더들의 수의 크기는 <도표 1>에 잘 나타나 있다.

개개의 데이터베이스의 갯수는 한개의 엔트리에 포함된 데이터베이스집합(Sets)과 데이터베이스 엔트리수를 합한 것이다. 데이터베이스엔트리 수와 개개의 데이터베이스수는 <도표 2>인 막대그래프에 잘 나타나 있다. 데이터베이스프로듀서의 수는 데이터베이스프로듀서와 엔트리프로듀서에 중복표기되어 있다. 많은 온라인 데이터베이스프로듀서들이 역시 CD-ROM이나 디스켓으로 이용하는 데이터베이스들을 제작함에 따라 이중으로 계수하게 된다. <도표 1>은 그러한 중복계산을 피한 숫자들이다. 또한 그들이 만든 데이터베이스를 직접 판매하는 프로듀서들은 역시 프로듀서와 벤더수에 중복으로 나타난다. 이렇게 중복된 전체 벤더수가 <도표 1>에 보여주고 있다. 이렇게 숫적으로 증가한 데이터베이스산업의 성공의 열쇠는 무엇보다도 컴퓨터의 출현으로 인한 정보산업의 도래로 인한 것이다. 이러한 수치들은 컴퓨터가 읽을 수 있는 데이터베이스들의 사용수로 그 성공을 측정할 수 있으며 그러한 수치는 <도표 4>에 잘 나타나고 있다. <도표 4>를 분석해보면 데이터베이스레코드가

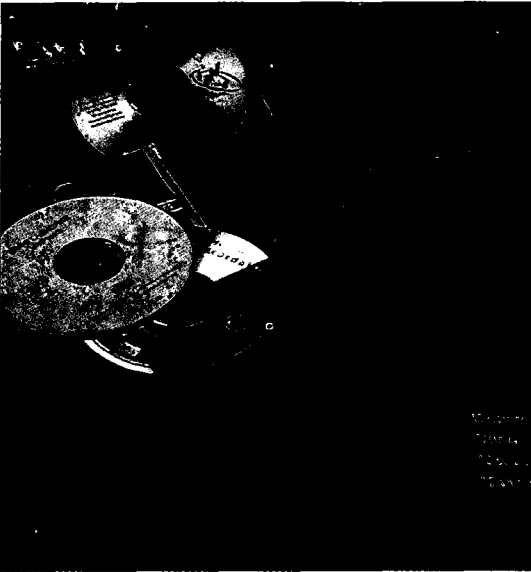
<도표 1> 벤더, 제작자, 엔트리, DB수의 증가



주 : 한개이상의 벤더에 의해 제공되어지는 DB는 한번만 계산

개별의 데이터베이스, 데이터베이스엔트리, 온라인검색, 프로듀서, 벤더 다음에 오는 가장 작은 최소단위임을 알 수 있다.

엔터티의 증가는 벤더들의 입장에서 가장 적은 규모로 엔터티규모가 감소하였고, 데이터베이스 레코드 수가 가장 빠른 속도로 증가했기 때문에 가능했다.



데이터베이스 레코드

<도표 3>에서 알 수 있듯이 18년간(1975-1993)이라는 기간 동안에 레코드수는 107배 만큼 성장해 왔다. - 레코드수는 5천2백만개에서 55억7천2백만개 늘어났는데 이는 데이터베이스수가 <도표 1>에서 알 수 있듯이 3백1개에서 8천2백6십8개로 27배 늘어난 것과 대조적이다. - 데이터베이스엔트리는 3백1개에서 7천5백3십8개 25배 늘어났다.

프로듀서수는 개인이 많은 데이터베이스를 만들기 때문에 저성장을 계속해왔다. 1993년에 프로듀서들은 2백개에서 2천7백4십4개인 14배 성장했으며 평균 3건의 데이터베이스를 이용하게 했다. 1993년 데이터베이스프로듀서수가 1992년에

비해 감소했는데 이는 개인프로듀서들에 의해 제작된 데이터베이스들이 감소했기 때문이다. 벤더의 수는 더딘 속도로 성장해 왔는데 한 벤더가 여러개의 데이터베이스서비스를 해왔기 때문이다. 사실 DIALOG Information사와 같은 벤더들은 수백개의 데이터베이스를 제공하고 있다. 데이터베이스벤더의 수는 1991년부터 1993년까지 CD-ROM제품을 이용한 벤더수의 증가로 급격히 성장했다. 18년간동안 벤더의 수들은 1백5개에서 1천6백2십9개로 15.5배 성장했다.

데이터베이스레코드의 수는 데이터베이스수만큼 비례적으로 성장하지 못했다. 데이터베이스레코드수의 증가는 3억1천만개의 레코드수를 기록하는 1983년까지는 성장세가 둔했다. 1983년부터 84년까지 3배이상(약 10억만개 레코드수)성장했고, 1984년부터 1987년까지 2배(20억만개 이상), 1987년부터 1993년까지 약 2.7배인 55억7천2백만개의 레코드수를 기록했다. 1975년에는 데이터베이스들이 평균적으로 17만3천개의 레코드를 가지고 있었으며 1985년에는 약 50만개 레코드를 갖고 있다. 1993년에 데이터베이스엔트리들은 평균적으로 73만9천개의 레코드를 갖고 있거나 개별 6만4천개의 레코드를 갖고 있다.

평균 데이터베이스의 크기가 이렇게 다른 것은 3백4십4개의 데이터베이스가 1백만개 이상의 레코드를 - 그중 9개는 1억개이상의 레코드를, 43개는 1천만개에서 1억개의 레코드를, 292개는 적어도 백만개에서 1천만개 미만의 레코드를 갖고 있기 때문이다.

344개의 대규모 데이터베이스를 제외하면 1993년 데이터베이스엔트리들은 평균적으로 10만6천개의 레코드를 갖고 있고 데이터베이스들은 평균적으로 9만7천개의 레코드를 갖고 있다. 이와같이 대규모 데이터베이스들은 어떻게 포함시키느냐에 따라 데이터베이스의 평균레코드수는 매우 다르게 나타난다.

데이터베이스레코드를 숫자화한 엔터티들은 매

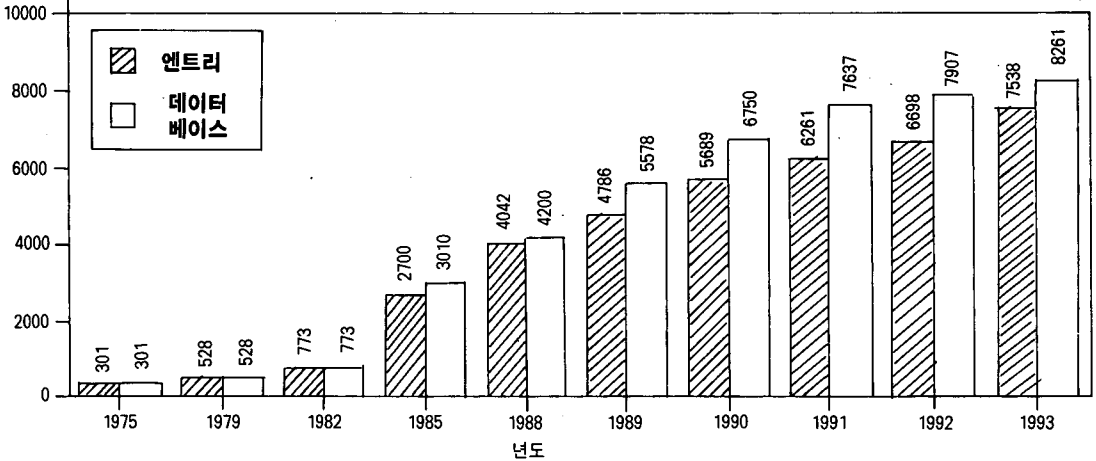
우 다양하지만 일반적으로 2백개에서 2천개 단어들로 구성된다(비단어로 구성된 레코드들은 저장을 위해 사용되는 바이트수로 나타냄). 레코드들은 인용구, 초록, 뉴스스토리, 자서전, 요리법, 타임시리즈, 소프트웨어프로그램, 화학명, 화학구조, 재산목록, 가상적인 것들에 대한 서술이나 목록을 의미한다. 아마 데이터베이스시장규모에 가장 영향을 준 한가지 중요한 요인은 80년대 중반부터 시작된 상업용 비지네스데이터베이스들의 증

가일 것이다. 즉 타임시리즈데이터베이스나, 전화번호 디렉토리데이터베이스나 신문상의 데이터베이스들이 바로 데이터베이스시장규모를 크게 증가시키는데 일조를 했다.

이와같이 대규모데이터베이스들은 서지정보(참고문헌, 초록, 전문)가 있고, 기업정보, 타임시리즈, 특허정보, 전화번호부나 주소록, 인명록, 자서전, 고용기록, 화학정보, 지역정보, 뉴스, 단어목록(전문용어, 단어, 시소러스), 상표목록, 조달과

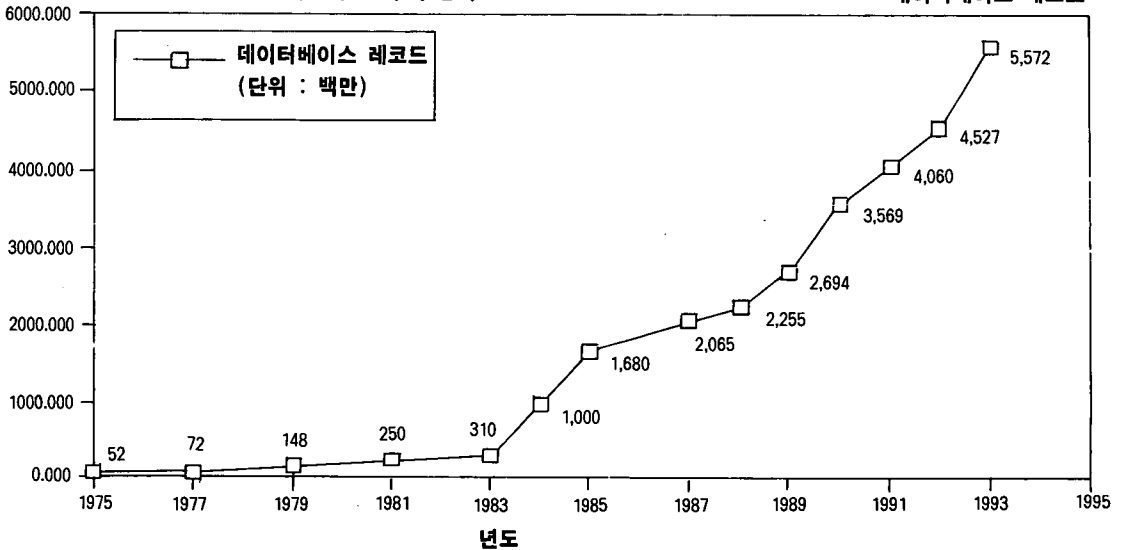
<도표 2> 데이터베이스 엔트리와 데이터베이스 수의 증가

엔트리의 증가



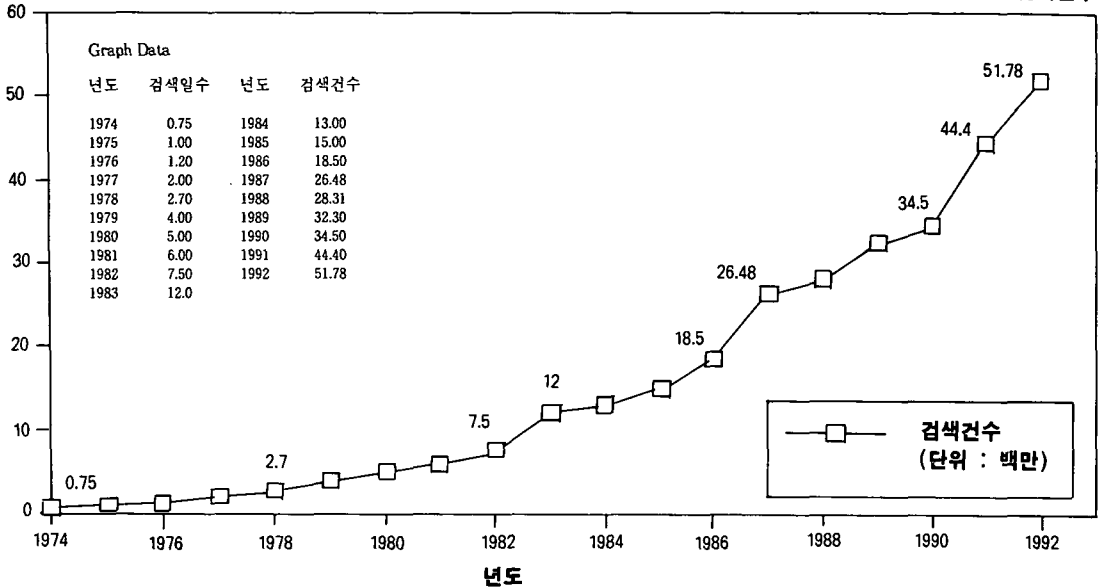
<도표 3> 데이터베이스레코드수의 증가

데이터베이스 레코드



<도표 4> 주요 미국벤처의 온라인 데이터베이스 검색건수의 성장

검색건수



계약집 등이 있다.

온라인 검색건수(Online Searches)

많은 데이터들은 미국에서 문자로 된 데이터베이스를 주로 제공하고 있고 Information Market Indicators(IMI)에 의해 제작되어지며 이들 데이터베이스는 문자로 된 데이터베이스를 제공하고 있고, Mead Data Central, West Publishing Company, DIALOG Information Services, ORBIT온라인과 BRS온라인서비스를 갖고있는 Info Pro Technologies같은 유통업체에서 제공하고 있다.

<도표 4>를 보면 온라인데이터베이스의 성장은 1974년 연간 7십5만건 검색수에서 1992년 연간 5천1백7십만건으로 증가했다. 이러한 수치는 특정 유통업체들의 부분 통합수치인데 여기에 거래된 데이터베이스중, 예를들면 주식, 컴퓨터 주문, 카드 조회, 항공기 예약등을 합치면 검색수는 엄청나게 늘어난다. 그러나 불행히도 그 수치는 알 수가 없다. <도표 4>에 검색 수를 나타내는 도표는 데이터베이스레코드 수와 너무 차이가 나

서 <도표 1>과 같이 유통수, 제작수, 데이터베이스와 함께 한 도표로 나타낼 수가 없다. 한편 <도표 3>은 데이터베이스 레코드 수의 성장을 보여주고 있다.(1989년 데이터부터 시작된 전체 데이터베이스레코드수 산정방법은 중간에 사라진 데이터수를 더하여 계산하게 되었다.) 이렇게 사라진 데이터를 계산하기 위해 7백만건 이하의 데이터베이스레코드를 평균하여 가산했다.

수년간의 기존 데이터베이스와 새로운 데이터베이스의 유통을 토대로 하여 데이터베이스의 합산법은 점차 정확성을 기하게 되었다. <도표 1과 3>은 공식적으로 우리가 사용할 수 있는 세계적인 모든 데이터베이스의 종류수를 나타내고 있다.

<도표 3>과 <도표 4>에서 알수 있듯이 데이터베이스 레코드수와 데이터베이스 검색수는 비례적인 성장을 해왔음을 보여준다. 만약 다른 나라의 검색수나 다른 종류의 데이터베이스수를 가산하더라도 <도표 3, 4>에 나타난 곡선의 형태로 나타날 것이며 다만 수적인 변화만이 있을 것이다.