

# 전세계 DB산업 현황

## (The State of Databases Today : 1993) ②

### 차 례

I. 개요
II. 데이터베이스 레코드
III. 온라인 검색
IV. 데이터베이스 (표현형태별 분석)
V. 지역 및 국가별 분석
VI. 주제별 분석
VII. 정보제공자별 분석
VIII. 정보유통자별 분석
IX. 데이터베이스 세계의 미래상

### IV. 데이터베이스 (Databases)

#### 정보표현 형태에 따른 분류

DB는 여러가지 방법으로 분류할 수 있다. 그 한가지가 정보표현 형태에 의한 것이다. 정보는 단어, 숫자, 이미지, 혹은 소리등의 형태를 띤다. 따라서 DB도 문자중심, 숫자중심, 이미지(화상)중심, 혹은 오디오중심으로 나눌 수 있다. 이러한 구분은 정보표현형태에 기초한 것인데, 이러한 형태가 파일의 구조와 탐색 혹은 검색S/W에 영향을 미친다 ; 즉 파일구조와 검색기술

이 4가지 분류에 따라 많이 달라진다는 것이다.

DB분류체계중에서, 문자 및 숫자중심의 DB를 분류하는 하부체계에 대한 연구가 오래전부터 구체화되어왔다. 문자중심의 DB들은 다음의 항목들로 분류될 수 있다 ; 서지, 특허/상표권, 디렉토리, 全文등. 처음으로 일반에 상용화된 DB들은 문자중심의 것들이었다. 1960-70년대에 와서는 대부분의 DB들이 문자중심이었고 이러한 문자중심 DB의 대부분이 서지정보 형태였다.

숫자중심의 DB들은 다음의 항목으로 분류 될 수 있다 ; 거래형, 통계형, 시계열형, 자산형, 기타.

일반인이 접근할 수 있는 이미지나 오디오 DB는 1980년대까지는 존재하지 않았다. 최근에 와서 오디오/이미지DB들은 애플사의 매킨토시, IBM과 그 호환기종, 미래의 첨단 소형PC용으로 개발되고 있다. 아울러 越全文化(Hypertext)와 CD-ROM 시스템의 보편화와 함께, 점차 기능이 우수한 멀티미디어 DB시대를 맞게 될 것이다. 멀티미디어 CD-ROM에 오디오, 비디오, 그리고 ASCII텍스트가 혼합되어 동시 실행되고 심지어 다양한 언어에 적합한 여러가지 오디오트랙을 사용할 수 있게 될 것이다.

<表 1>

형태별 분류항목과 연도별 DB의 수, 비율  
(한개의 DB를 하나의 분류항목에 포함시킴)

DB분류항목	1985+		1988		1989		1990		1991		1992	
	수	(%)	수	(%)	수	(%)	수	(%)	수	(%)	수	(%)
문자중심DB*	1926	(64)	2797	(69)	3370	(70)	4080	(72)	4491	(72)	4925	(70)
숫자중심DB	1084	(36)	1136	(28)	1236	(26)	1298	(23)	1370	(22)	1533	(22)
이미지DB			14	((1)	34	((1)	113	(2)	145	(2)	272	(4)
음성DB			1	((1)	2	((1)	16	((1)	27	((1)	83	(2)
전자서비스			90	(2)	134	(3)	170	(3)	172	(3)	146	(2)
S/W DB			4	((1)	10	((1)	12	((1)	55	(1)	39	((1)
총 계	3010	(100)	4042	(100)	4786	(100)	5689	(100)	6261	(100)	6998	(100)

\* 주 : 수는 DB Entry의 수를 의미하며, 1992년에 있어 개별DB의 분류항목별 수는 7907개에다 해당비율을 곱해서 계산할 수 있다.

+ 1985년의 경우, DB분류항목은 단 2개였음.

\* 이 부분의 상세한 내용은 <표3>을 참조바람.

모든 DB는 이러한 4가지 기본적인 항목에 따라 분류될 수 있고, 그중 많은 DB들이 全文과 통계정보, 全文과 이미지정보, 이미지와 오디오정보, 혹은 이러한 것들의 혼합과 같은 복합적 형태의 정보를 포함하고 있기 때문에 여러가지 다양한 방법에 의해 분류될 수 있다.

그외에 두가지의 DB분류항목(전자서비스와 S/W)이 <표1>에 제시된 자료에 추가되어 있는데, 이것들 둘다 문자와 숫자를 쓰고 있기 때문에 표현형태에 따라 분류될 수도 있다. 하지만, 두가지가 사용되는 방식이 다르고 독특한 특성을 가지고 있기 때문에 추가항목으로 제시되어 있는 것이다. 전자게시판, E-Mail, 그리고 전자회의와 같은 전자정보서비스도 정보를 포함하고 있지만, 그것들은 일시적이고 기록

보존적인 형태가 아니다. 그럼에도 불구하고 몇몇 주요한 DB유통업체들이 DB검색 서비스와 같은 방식으로 이러한 서비스를 제공하고 있기 때문에 그것들도 DB안에 포함시켜야 한다. E-Mail과 전자회의서비스는 개별DB의 숫자과목에 포함시키지 않았지만, 그러한 전자서비스도 하나의 DB로 포함시킬 때는, DB와 연관성이 있는 것으로 간주하여 하나의 DB로 분류되는 것이다. 최근에는 몇몇 온라인 공용서비스를 통하여 S/W DB도 온라인으로 이용할 수 있게 되었다.(예를들면, America Online은 여러가지 소비자지향적인 DB와 전자게시판뿐만아니라 Apple Macintosh용의 공용 S/W와 Shareware를 제공하는 등 다면적 서비스를 제공하고 있고, Byte Information Exchange는 전자회의 서비스를 제공하면

<表 2>

형태별 분류항목과 연도별 DB의 수, 비율  
(하나의 DB Entry가 여러 분류항목에 동시에 포함되는 경우)

DB분류항목	1985+		1988		1989		1990		1991		1992	
	수	(%)	수	(%)	수	(%)	수	(%)	수	(%)	수	(%)
문자중심DB*	1926	(64)	3147	(69)	3409	(70)	4213	(72)	4661	(72)	6497	(71)
숫자중심DB	1084	(36)	1278	(28)	1250	(26)	1360	(23)	1422	(22)	1974	(21)
이미지/화상			16	(1)	34	(1)	98	(2)	151	(2)	358	(4)
음성DB			1	(1)	2	(1)	16	(1)	28	(1)	109	(1)
전자서비스			101	(2)	136	(3)	178	(3)	179	(3)	193	(2)
S/W DB			5	(1)	10	(1)	12	(1)	57	(1)	52	(1)
총 계	3010	(100)	4548	(100)	4741	(100)	5877	(100)	6498	(100)	9183	(100)

\* 주 : 수는 DB Entry의 수를 말함

+ 1985년의 경우 DB당 하나의 분류항목에 포함시켰음.

\* 자세한 사항은 표3을 참조바람.

서 공용S/W를 온라인 서비스하고 있다. CompHelp는 시각 장애자용 PC서비스와 함께 S/W 다운로드 서비스도 하고 있다.) 몇몇 정보제공자들은 S/W를 내장한 CD-ROM DB를 제공하고 있다. 그러한 DB들은 코드변환과 S/W 사용에 대한 설명을 포함하고 있기 때문에 S/W DB로 분류된다. 그것들은 Directory DB로 분류되어 있는 S/W의 Directory와는 다른 것이다.

여기서 다루어진 모든 DB는 위에 제시된 6가지 항목에 따라서 코드화되어 있다. 표1은 1985년과 1988-92년까지를 대상으로 6가지 분류항목과 연관시켜 DB의 수와 비율을 제시하고 있다.

표1은 여러 분류항목에 중복분류될 수 있는 DB들을 하나의 항목으로 나눈 후 이 조사연구에서 대상으로 삼은 6,998개의 DB Entry를 분류항목에 따른 수와 비율로

나누어 나타내고 있다.

표2는 일대일 대응과 관계없이上記의 DB들을 분류한 것이다.

1개의 DB entry가 1개이상의 분류항목에 포함될 수 있어서, DB entry수보다 표2에 분류된 DB수가 더 많게 나타나고 있다. 게다가, 전화번호DB, 신문DB, 화학·특허 그리고 기업관련DB 그리고 CD-ROM DB가 점점 증가함에 따라 문자중심DB의 비율이 더 커지고 있다. 그리고 단어중심DB내에서 하부분류항목에 대한 분석이 표3에서 제시되고 있는데 여기에 보면 全文DB가 거의 서지정보DB만큼 많음을 알 수 있다. 또한 문자와 숫자중심의 DB들이 수효에 있어서 계속 증가하고 있음을 알 수 있다. 이미지/비디오 DB는 1990년 이래로 거의 3배이상 늘었고, 1989년 이후로는 10배이상 증가했다.(표2) 또한 음성

정보DB의 수도 1988년의 1개에서 1992년에는 109개로 빠르게 증가했음을 알 수 있

다. 그리고 S/W DB는 1990년의 12개에서 1992년에 와서는 57개로 증가했다.

<表 3> 세분화한 문자중심 DB의 연도별 갯수와 비율

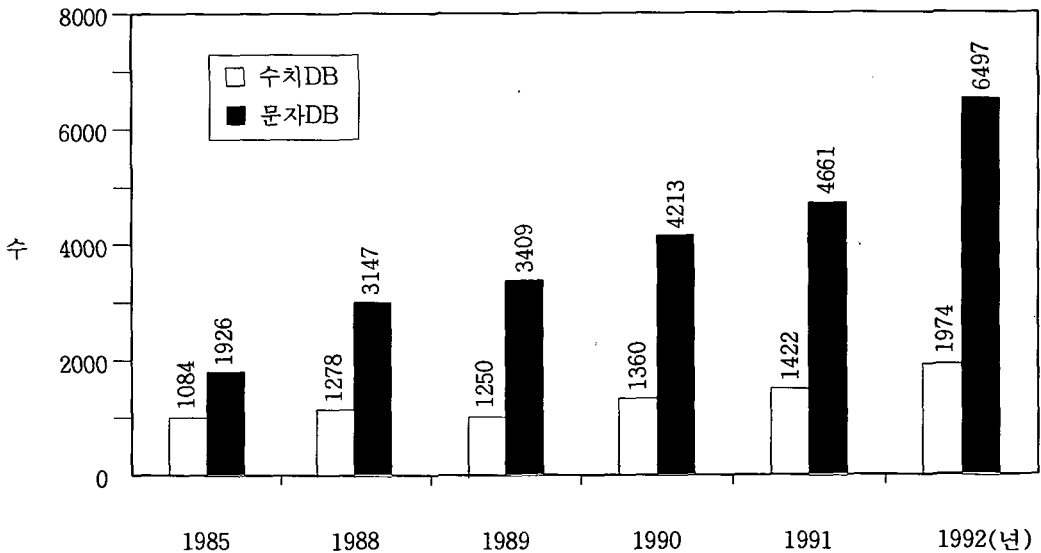
문자중심DB 의 세부항목	1985+		1988		1989		1990		1991		1992	
	수	(%)	수	(%)	수	(%)	수	(%)	수	(%)	수	(%)
서지정보	1094	(57)	1162	(37)	1223	(36)	1367	(32)	1425	(31)	1715	(26)
특허/상표	#	#	55	(2)	58	(2)	80	(2)	85	(2)	47	(〈1)
全文	535	(28)	1285	(41)	1412	(41)	1786	(42)	2040	(44)	3077	(47)
디렉토리	287	(15)	613	(19)	707	(21)	952	(23)	1074	(23)	1611	(25)
사전류	10	(〈1)	32	(〈1)	9	(〈1)	23	(1)	32	(〈1)	47	(〈1)
기타	#	#	#	#	#	#	4	(〈1)	5	(〈1)	0	(0)
총 계	1926	(100)	3147	(100)	3409	(100)	4212	(100)	4661	(100)	6497	(100)

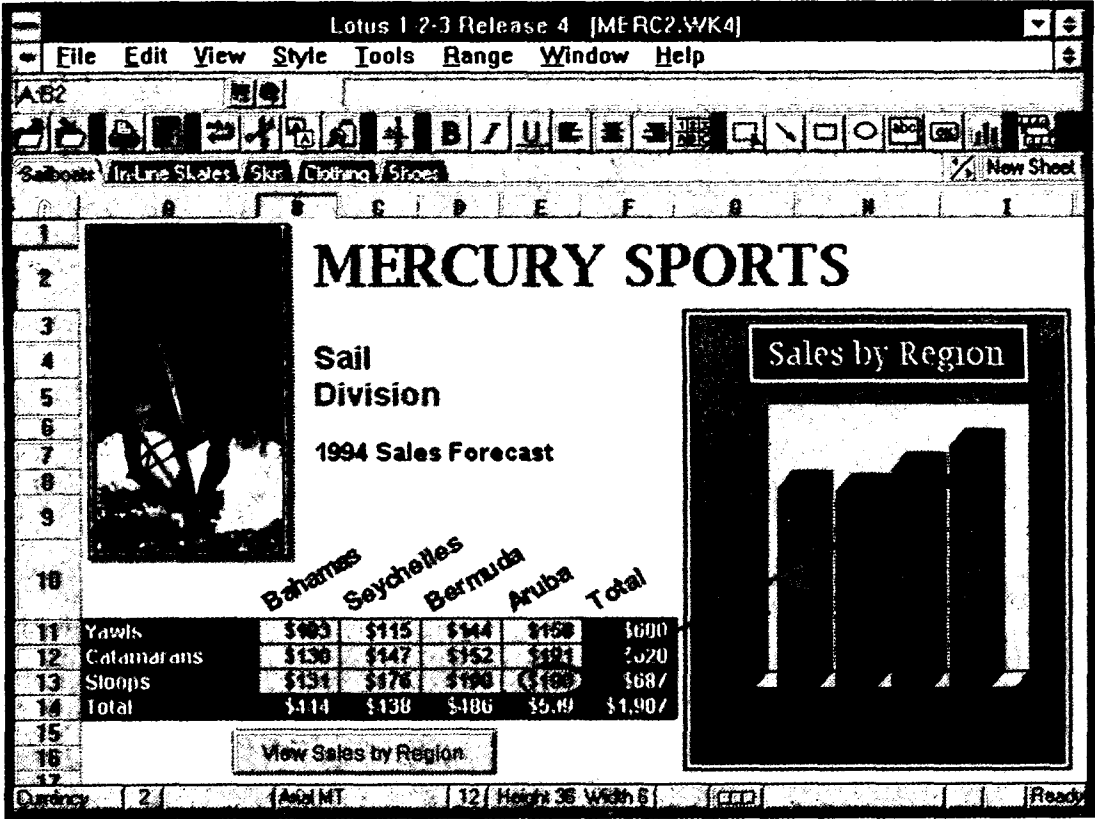
\* 갯수는 DB Entry를 말함.

# 1985년의 특허/상표정보는 서지정보에 포함되었음.

+ 1985년에는 각 DB가 한가지 방법으로 분류되었지만, 그이후에는 여러가지로 분류되기 시작

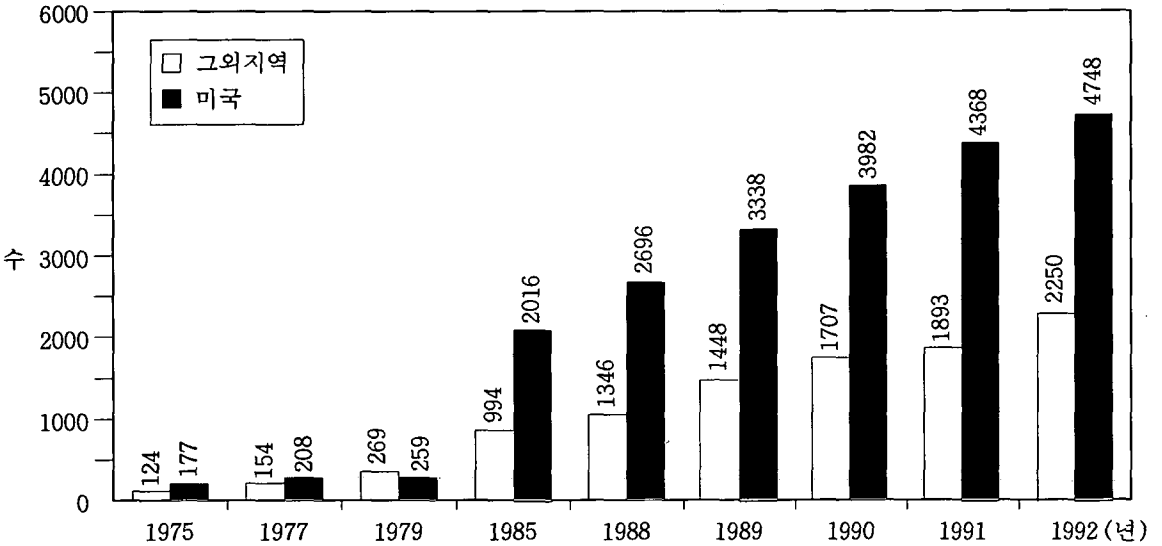
<도표 5> 연도별 DB Entry의 수





<도표 6>

미국과 그외 지역의 연도별 DB Entry 성장현황



<표 4>

주요 지역별 DB현황(1992)

지 역	1991		1992	
	DB Entries	(%)	DB Entries	(%)
아프리카	7	(〈1)	7	(〈1)
아시아	28	(〈1)	25	(〈1)
오스트레일리아	119	(2)	161	(2)
극동아시아	155	(3)	171	(2)
동부 유럽	11	(〈1)	12	(〈1)
서부 유럽	1473	(24)	1838	(26)
북미	4424	(71)	4768	(68)
남미	44	(〈1)	16	(〈1)
총 계	6261	(100)	6998	(100)

\* 주 : 한 DB가 여러 주제분류에 포함될 수 있음.

### V. 지역과 국가별 분석(Region and Country of Origin)

이 글에서 다루고 있는 6,998개의 DB entry는 8개의 주요한 지역에서 만들어진 다 ; 아프리카, 아시아 그리고 중동, 호주, 극동아시아, 동부유럽, 영국을 포함한 서부 유럽, 북미, 그리고 남미이다. 남극대륙의 DB는 하나도 포함되지 않았다. 이러한 지역별 DB entry와 DB의 분포상황은 표4에 나타나 있다. 아시아와 중동의 DB는 중국과 대만, 인도, 그리고 이스라엘의 것이고 극동지방의 DB는 홍콩, 필리핀, 일본, 싱가포르, 그리고 태국의 것이다. 1,000개 이상의 DB를 보유한 지역은 4,700개 이상을 가진

북미와 1,800개 이상을 가진 영국을 포함한 서유럽 등 2군데이다.

국가별로 보면, 미국이 1,000단위 이상의 DB를 보유한 유일한 나라이며, 100개 이상의 DB를 보유한 나라는 호주(168), 캐나다(455), 영국(652), 프랑스(315), 독일(280), 그리고 일본(161)이다. DB는 18개 언어로 만들어져 있다(중국어, 영어, 덴마크어, 독일어, 핀란드어, 폴란드어, 불어, 네덜란드어, 그리스어, 히브리어, 이탈리아어, 일본어, 노르웨이어, 폴란드어, 포르투갈어, 러시아어, 스페인어, 그리고 스웨덴어) <도표6>은 전세계 DB를 미국과 그외 지역으로 나누어 나타낸 것이다. [B]

(번역 · 정리 - 이석기)