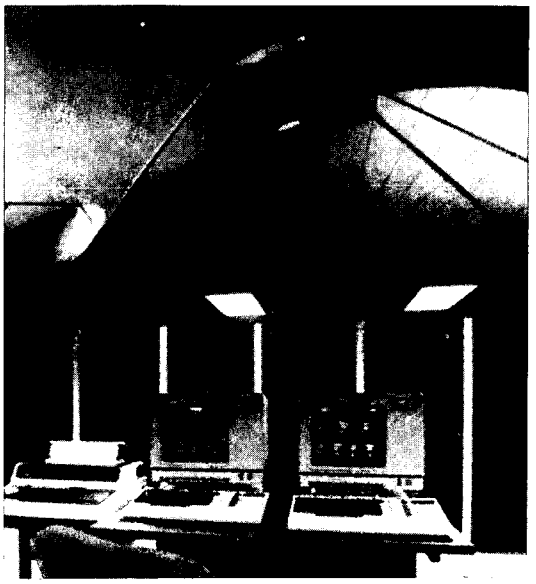
THE STATE OF DATABASE TODAY : 1994(Ⅱ)

목 차

1. 서문
2. 데이터베이스 레코드
3. 온라인 검색수
4. 데이터베이스
   - 데이터베이스 형태에 따른 분류
   - 지역별, 국가별 DB현황
   - 주제별 분류
5. 제작업체
6. 유통업체
7. 결론



데이터베이스

○ 데이터표현 형태에 따른 분류

데이터베이스들은 여러가지 형태로 분류될 수 있다. 그 중 하나가 데이터표현 형태에 따른 분류이다. 데이터는 단어, 기호, 화상, 소리 등의 형태로 나타날 수 있으며 데이터베이스들은 문장지향, 기호지향, 화상지향(정지화상과 동화상) 또는 음향지향(소리)의 형태로 분류 된다. 이러한 형태의 구별은 파일구조와 검색을 위한 소프트웨어에 영향을 주는 데이터 표현형태에 그 기초를 두고 있으며 문장지향 데이터베이스는 서지학, 디렉토리, 전문, 특허/상품 등 처음 외부에서 활용해온 데이터베이스가 문장지향 데이터베이스이다. 1960– 70년대에는 대부분의 데이터베이스들이 문장지향적인 것이었으며 대표적인 것이 서지학이었다.

기호로 구성된 데이터베이스는 거래량, 통계, 타임시리즈, 재산목록 등 외부에서 활용할 수 있었던 화상이나 음향 데이터베이스들은 80년대 중반까지는 존재하지 않았다. 화상이나 음향 데이터베이스는 매킨토시컴퓨터, 쿼트로, IBM컴퓨터, IBM호환기종, NEXT마이크로컴퓨터가 등장하면서 존재하기 시작했다. 매킨토시에서 주로 사용된 하이퍼텍스트와 CD-ROM타이틀의 출현으로 많은 멀티미디어 데이터베이스들을 쉽게볼 수 있게 되었다. 멀티미디어 CD-ROM에 오디오, 비디오, ASCII텍스트가 함께 짜여지면서 다른언어로 다른 오디오트랙에서도 사용될 수 있게 되었다. 모든 데이터베이스는 기본적으로 이와같이 4가지 부류로 분류되고 그들이 담고 있는 내용에 따라 다양한 형태로 분류할 수 있다. 즉, 텍스트와 숫자데이터, 텍스트와 이미지데이터, 이미지와 오디오데이터 등등. <도표1>에서 보듯이 2개의 추가적인 데이터베이스분류(전자서비스와 소프트웨어)가 가능하다. 게시판, 전자메일, 전자회의와 같은 전자정보서비스는 일시적인 데이터나 문서

<도표 1> 데이터표현 형태에 따른 분류
(데이터베이스당 하나의 분류로 구분)

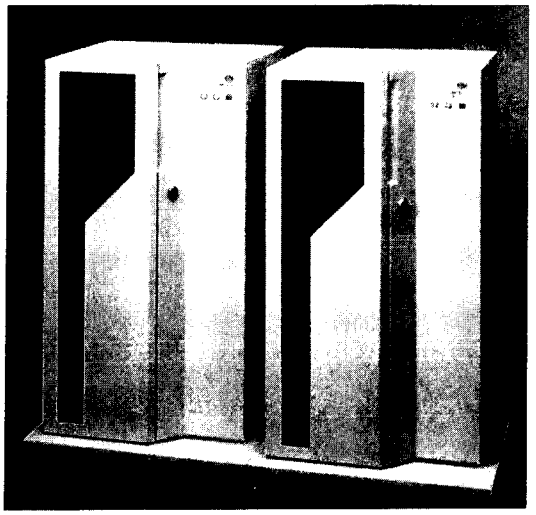
DB분류 (one/db)	1985 No.	(%)	1988 No.	(%)	1989 No.	(%)	1990 No.	(%)	1991 No.	(%)	1992 No.	(%)	1993 No.	(%)
문장지향	1728	(64)	2797	(69)	3370	(70)	4080	(72)	4491	(72)	4925	(70)	5421	(72)
기호지향	972	(36)	1136	(28)	1236	(26)	1298	(23)	1370	(22)	1533	(22)	1437	(19)
화상			14	(<1)	34	(<1)	113	(2)	145	(2)	272	(4)	340	(4)
음향			1	(<1)	2	(<1)	16	(<1)	28	(<1)	83	(2)	106	(1)
전자서비스			90	(2)	134	(3)	170	(3)	172	(3)	146	(2)	203	(3)
소프트웨어			4	(<1)	10	(<1)	12	(<1)	55	(1)	39	(<1)	31	(<1)
총계	2700(100)		4042(100)		4786(100)		5689(100)		6261(100)		6998(100)		7538(100)	

* DB수는 디렉토리에 있는 DB 엔트리 수임
* 1985년에는 2가지 종류의 DB만 있음 (자세한 것은 도표3 참조)

<도표 2> DB엔트리 특성에 따른 분류(년도별 수와 퍼센트)

DB분류 (multiple/db)	1985 No.	(%)	1988 No.	(%)	1989 No.	(%)	1990 No.	(%)	1991 No.	(%)	1992 No.	(%)	1993 No.	(%)
문장지향	1728	(64)	3147	(69)	3409	(70)	4213	(72)	4661	(72)	6497	(71)	6652	(72)
기호지향	972	(36)	1278	(28)	1250	(26)	1360	(23)	1422	(22)	1974	(21)	1763	(19)
화상			16	(<1)	34	(<1)	98	(2)	151	(2)	358	(4)	417	(4)
음향			1	(<1)	2	(<1)	16	(<1)	28	(<1)	109	(1)	130	(1)
전자서비스			101	(2)	136	(3)	178	(3)	179	(3)	193	(2)	249	(3)
소프트웨어			5	(<1)	10	(<1)	12	(<1)	57	(1)	52	(1)	38	(<1)
총계	2700(100)		4548(100)		4841(100)		5877(100)		6498(100)		9183(100)		9249(100)	

* DB수는 디렉토리에 있는 DB 엔트리 수임
* 1985년에는 2가지 종류의 DB만 있음 (자세한 것은 도표3 참조)



가 아닌 형태의 데이터를 포함한다. 그러나 많은 주요 벤더들이 복합 데이터를 팔고 있지만 그들의 데이터는 한 분류로 구분했다. 전자메일과 전자회의 서비스들은 개별적인 데이터베이스들로 계산되어지지 않지만 이들 데이터베이스들은 복합적인 데이터베이스로 집계하였다.

현재 상용적인 ON-LINE서비스들은 이와같은 데이터베이스들을 ON-LINE으로 서비스하고 있다. 예를 들면 AMERICA ONLINE社는 다양한 소비자 중심의 소프트웨어와 전자게시판 외에 애플컴퓨터와 매킨토시II에 사용되어지는 공용소프트웨어를 제공하고 있다. 또한 BYTE INFORMATION EXCHANGE社는 전자회의 서비스와 다양한 공개 소프트웨어를 제공하며, COMPUHELP社는 각종고장을 치유할 수 있는 소프트웨어를 다운로딩으로 처리하고 있다. 이와 같은 소프트웨어 데이터베이스들은 소프트웨어 패키지형태를 담고 있어 소프트웨어 데이터베이스 그 자체를 분류 하였다. 그들은 디렉토리 데이터베이스와는 그 내용이 전혀 다르다.

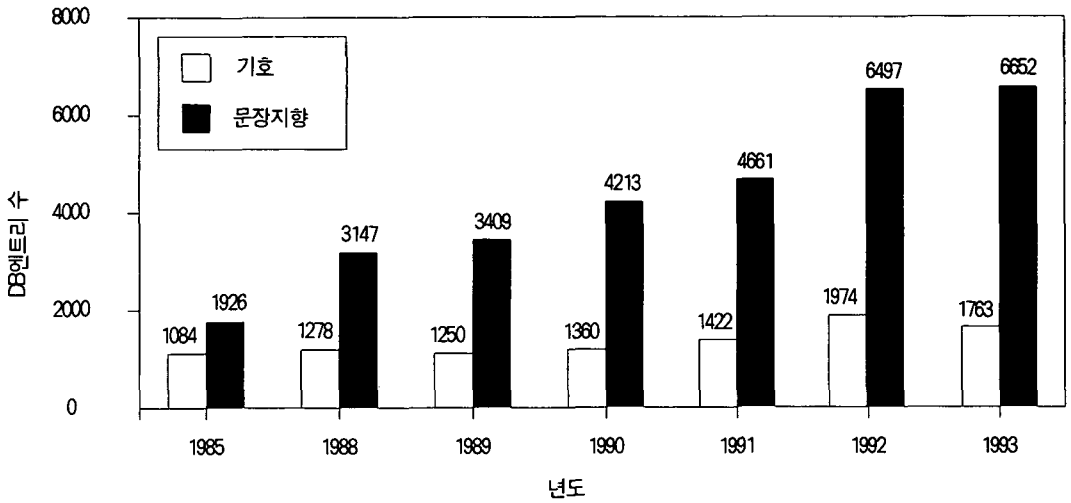
디렉토리에 나타난 데이터베이스는 위에 나타났듯이 6개로 분류될 수 있다. <도표1>은 85년과 88년부터 93년까지의 분류된 데이터베이스의 수와 퍼센트를 나타내고 있다. <도표1>은 게일 디렉토리에 나타난 중복되는 7,538개의 데이터베이스 엔트리를 표준화하여 분류했으며 <도표2>는 그렇지 않은 수치이다. 전반적으로 볼때, 문장지향 데이터베이스들은 전화번호 데이터베이스와 신문기사 데이터베이스, 화학구조나 특허, 기업정

<도표 3> 문장지향 DB의 수와 퍼센트(년도별)

문장지향 DB의 재분류	1985		1988		1989		1990		1991		1992		1993	
	No.	(%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)	No.	(%)
서지학	1094	(57)	1162	(37)	1223	(36)	1367	(32)	1425	(31)	1715	(26)	1739	(26)
특허/상표	#	#	55	(2)	58	(2)	80	(2)	85	(2)	47	(<1)	91	(1)
전분	535	(28)	1285	(41)	1412	(42)	1786	(42)	2040	(44)	3077	(47)	3155	(48)
디렉토리	287	(15)	613	(19)	9	(23)	952	(23)	1074	(23)	1611	(25)	1600	(24)
사전	10	(<1)	32	#	#	(1)	23	(1)	32	(<1)	47	(<1)	67	(1)
기타	#		#			(<1)	4	(<1)	5	(<1)	0	(0)	0	(0)
총계	1926(100)		3147(100)		3409(100)		4212(100)		4661(100)		6497(100)		6652(100)	

수치는 디렉토리에 있는 DB엔트리의 수이다.
85년 특허/상품정보는 서지정보의 일부가 포함되어 있음.
85년 DB는 오직 한가지 방법으로 분류되었으나 그 이후 다양하게 분류되었음.

<그림 1> 년도별 DB엔트리 수



보 데이터베이스 그리고 CD-ROM데이터베이스의 증가로 꾸준한 성장을 보이고 있다.

이와같은 수치는 <도표 3>에 잘 나타나 있다. 문장지향 데이터베이스중 서지학데이터베이스가 가장 많으며 디렉토리 데이터베이스가 그 뒤를 잇고 있다. 화상/영상 데이터베이스는 1990년 이래로 4배이상 성장했고 88년 이래로 26배 성장했다. 오디오 데이터베이스는 88년 1개에서 93년 130개 늘어나 강세를 나타냈다. 소프트웨어 데이터베이스는 88년 5개에서 92년 57개로 늘어났으며 93년 38개로 감소 했다.

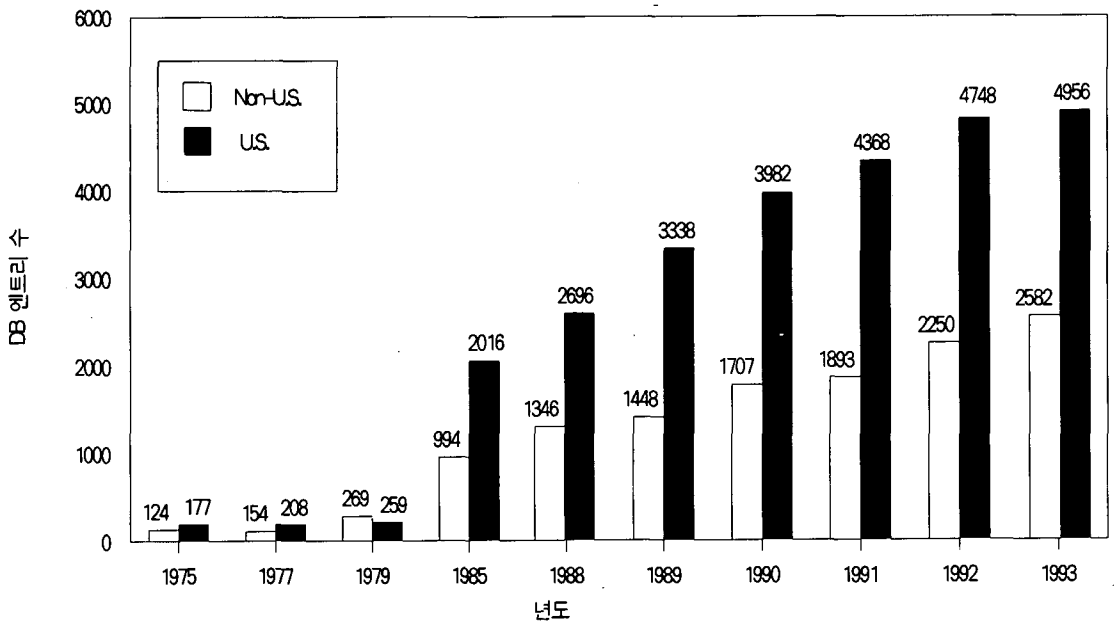
○ 지역별, 국가별 DB현황

게일디렉토리에 나타난 7,538개의 데이터베이스엔트리는 아프리카, 아시아, 중동, 호주, 동아시아, 동유럽, 서유럽(영국포함), 남/북아메리카 등 8개지역에서 대부분 제작된 것들이다. 지역별로 유통되고 있는 데이터베이스 엔트리와 데이터베이스 수치가 <도표 4>에 잘나타나 있다. 아시아와 중동지역의 데이터베이스는 중국(대만포함)과 이스라엘에서 나온 것들이다. 극동지역의 데이터베이스들은 홍콩, 일본, 싱가폴에서 유통되는 것들이다.

<도표 4> 년도별 지역에 따른 DB엔트리

지역	1991 DB 엔트리	(%)	1992 DB 엔트리	(%)	1993 DB 엔트리	(%)
아프리카	7	(<1)	7	(<1)	10	(<1)
아시아	28	(<1)	25	(<1)	36	(<1)
호주	119	(2)	161	(2)	189	(3)
동유럽	11	(<1)	12	(<1)	82	(1)
동아시아	155	(3)	171	(2)	164	(2)
북미	4424	(71)	4768	(68)	5094	(68)
남미	44	(<1)	16	(<1)	25	(<1)
서유럽	1473	(24)	1838	(26)	1938	(26)
총계	6261	(100)	6998	(100)	7538	(100)

<그림 6> 미주지역과 그밖의 지역의 년도별 DB엔트리



북아메리카지역은 5천개이상의 데이터베이스 엔트리가 등록되어 있고, 서유럽(영국포함)은 1,900이상의 데이터베이스 엔트리들이 있다. 지역별로 볼때 미국만이 오직 수천개의 데이터베이스 엔트리를 보유하고 있다. 100개 이상의 데이터베이스 엔트리를 보유하고 있는 국가들은 다음과 같다. 호주 182개, 캐나다 480개, 영국 641개, 프랑스 288개, 독일 342개, 일본 153개를 각각보유하고 있다. <그림 6>은 미주지역과 그밖의 지역으로 구분하여 데이터베이스 엔트리의 제작수를 설명하고 있다

○ 주제별 분류

많은 유저들이 원하는 데이터베이스를 찾기위하여 주제분류를 먼저 검색한다. 예를들어 만일 어떤 엔드유저나 서쳐가 화학정보를 찾고자한다면 그는 먼저 화학 주제 데이터베이스를 뒤질것이다. 만일 선택한 주제필드에 그것이 존재하지 않다면 주제를 나타내는 서지 데이터베이스가 그 주제를 찾는데 도움을 줄것이다. 많은 데이터베이스는 <도표 5>에서 보듯이 몇개의 주제로 분류될 수 있다. 예를들면 자연건강사업 데이터베이스는 인체과학과 비지니스 분야에 모두 속한다. <도표 5>는 주제별로 분류될 수 있는 데이터베이스의 숫자가 실제 데이터베이스 숫자보다 많다는 것을 보여주고 있다. 하나의 주제의 내용을 충실히 담고 있는 데이터베이스는 하나의 데이터베이스로 묶었다. 예를 들어 어떤사람이 1993년의 비지니스에 관련된 데이터베이스의 퍼센트를 알고 싶다면, 8,103개의 관련 데이터베이스중 2,644개(약33%)의 비지니스 관련 데이터베이스를 쉽게 가려 볼 수 있을 것이다. 그러나 만일 관련 주제가 데이터베이스 가운데 적은 부분을 차지 한다면 주제 분류가 불가능하게 된다.

데이터베이스 주제 분류중 비지니스 관련 주제가 제일 많으며 다음은 과학/기술/엔지니어링 분야이며 89년 이래로 2배이상씩 성장해온 법률관

<도표 5> 주제별 DB의 수와 퍼센트

주제별 분류	1988 No.	(%)	1989 No.	(%)	1990 No.	(%)	1991 No.	(%)	1992 No.	(%)	1993 No.	(%)
비지니스	1815	(34)	1,687	(33)	1,956	(33)	2,101	(33)	2,624	(33)	2644	(33)
소비자관련	186	(3)										
일반	301	(6)	327	(6)	416	(7)	450	(7)	700	(9)	837	(10)
건강/생명과학	433	(8)	576	(11)	651	(11)	690	(11)	728	(9)	742	(9)
인문주의	84	(2)	184	(3)	216	(4)	248	(4)	314	(4)	330	(4)
법률	441	(8)	447	(9)	531	(9)	574	(9)	885	(11)	949	(12)
교육(학문적인)	29	(<1)	335	(7)	368	(6)	366	(6)	296	(4)	230	(3)
뉴스	428	(8)	186	(4)	233	(4)	291	(4)	385	(5)	369	(4)
사회과학	460	(9)	393	(8)	418	(7)	453	(7)	447	(6)	473	(6)
과학/기술/엔지니어링	1,184	(22)	996	(19)	1,154	(19)	1,210	(19)	1,492	(19)	1529	(19)
총계	5,361	(100)	5,131	(100)	5,943	(100)	6,383	(100)	7,871	(100)	8,103	(100)

주) 1. 중복주제가 있는 DB는 하나의 주제 DB로 분류했다.

<도표 6> 유통별 DB의 분류

유통매체별 분류	1989 다중사용번호		1990 다중사용번호		1991 다중사용번호		1992 다중사용번호		1993 다중사용번호	
	No.	엔트리	No.	엔트리	No.	엔트리	No.	엔트리	No.	엔트리
온라인	3524	2711	4018	3041	4170	3200	5486	4519	5564	5569
배치	999	769	1252	948	1321	1014	389	320	481	396
CD-ROM	433	333	715	541	1019	782	1321	1088	1648	1360
디스켓	478	368	626	474	695	533	676	557	781	645
마그네틱테잎	787	605	906	686	954	732	584	481	600	495
휴대가능용							39	33	57	47
총계	6221	4786	7517	5689	8159	6261	8495	6998	9136	7538

주) 하나의 DB는 여러가지 유형의 유통이 있을 수 있으며 여기에 나타난 수치는 발생할 유통의 모든 수를 포함한다.

련 주제가 그뒤를 잇고 있다. 또한 많은 비지니스 데이터베이스들은 주로 시장조사 및 인구통계 관련된 주제가 많아 지금은 비지니스 분야 또는 사회과학 분야로 분류된다. 따라서 89년 부터 소비자관련 주제는 없게 되었다. <도표 6>은 유통되는 방법에 따른 데이터베이스의 수와 퍼센트를 나타낸 것이다. 93년을 보면 총 7,538개의 데이터베이스 엔트리를 유통방식에 따라 분류하여 본 결과 9,136개의 사례가 나오게 되었다.

온라인으로 유통되는 데이터베이스는 데이터베이스수 65%보다 낮은 61%를 기록했으며, CD-ROM은 15%보다 많은 18%를, 디스켓은 9%, 마그네틱테이프는 7%, 배치파일은 5% 그리고 휴대할 수 있는 데이터베이스의 유통수는 1%미만으로 나타났다. <다음호에 The State of DATABASE TODAY 3가 연재됩니다> DB