

# 웹정보콘텐츠 산업의 현황과 동향

웹정보콘텐츠 산업은 정보제공기업이 유·무선 인터넷을 통해 정형화된 데이터베이스 또는 비정형 정보를 제작해 서비스하고 이용자로부터 회원제, 건당 요금 등의 요금체계에 따라 정보이용료를 받아 수익을 올리는 산업이다. 본 고에서는 웹정보콘텐츠산업의 국내외 동향, 제작 유통 환경, 이용환경, 기술 동향 및 향후 전망과 과제에 대해 소개한다.

글 이창한 기자

2004년에 전반적인 시장 하락세에도 불구하고 성장 폭이 다소 감소하기는 했으나, 성장세는 지속되고 있다.

이러한 온라인 매체에 의한 DB의 지속적인 증가의 원인은 기존 오프라인 매체에 의해 유통되는 콘텐츠가 온라인 매체로 전환됐기 때문이다. 네트워크 환경이 광대역화에 따라 이러한 현상은 당분간 지속될 것이다.

## 새로운 웹정보 콘텐츠 기반 서비스 출현

웹정보콘텐츠 기업의 주요 수익 모델은 유료회원 제도를 통한 회원 수입이나 정보 이용료를 통한 종량제 수입이 주종을 이루고 있다. 그러나 정보콘텐츠의 유통 매체가 웹으로 발전함에 따라 인터넷의 포털 사이트를 통한 무료 정보가 폭발적으로 증가해 기존의 정보 콘텐츠 제공자들은 새로운 비즈니스 모델을 찾고 있다. 이러한 비즈니스 모델은 매우 다양화되고 있는데, 콘텐츠 간 융합이나 기존의 온·오프라인 서비스와 융합하는 등의 새로운 수익 모델을 창출하고 있다.

콘텐츠 간 융합의 대표적인 사례로는 법령 콘텐츠에 법률 문헌 콘텐츠를 링크해 제공하거나 전자 지도위에 상점정보를 융합해 이용자에게 필요한 콘텐츠를 한 번에 다양하게 제공하는 웹정보콘텐츠 등이다. 웹정보콘텐츠와 오프라인 서비스가 융합되는 사례는 서적 정보와 오프라인 배송 시스템이 융합된 웹서점, 상품 정보와 전자 결제 서비스가 결합된 인터넷 쇼핑 몰 등으로 이용자는 필요한 서적

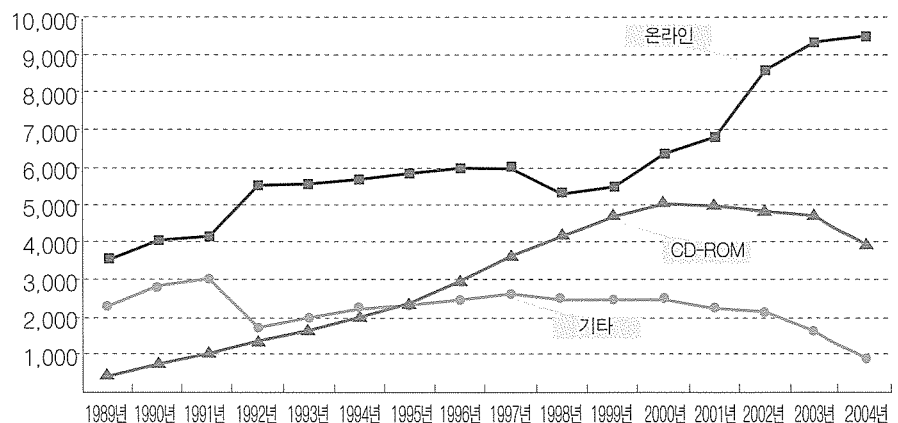
웹 상에서 유통되고 있는 콘텐츠는 e메일·인스턴트 메시징·Surface Web 및 Deep Web 등 네가지 형태로 구분하는 것이 일반적이다. e메일이나 메시징서비스는 웹상에서 개인간 정보를 교환하는 수단으로 이미 정착됐고, Surface Web은 회사나 개인의 홈페이지와 같이 비상업적으로 정보를 제공하는 웹 정보로서 고정형 웹 정보(Fixed Web Pages)라고도 한다. Deep Web은 이용자의 요구에 따라 웹정보를 생성하는 데이터베이스 주도형 웹정보 콘텐츠를 의미한다. 이중에서 정보 유통량이 가장 많은 것이 e메일이고 Deep Web이 그 뒤를 잇고 있다.

본 고에서는 전체 디지털콘텐츠 산업 중 웹정보콘텐츠 부문에 대한 범위를 산업적 측면에서 가장 의미 있는 Deep Web, 즉 DB 주도형 웹정보 콘텐츠(생활정보·특허정보·법률정보 등)에 한하며 게임, 디지털영상물 등을 포함시키지 않았다. 무선인터넷을 통한 정보 콘텐츠에 대해서는 다른 장에서 다루고 있기 때문에 이장에서는 논외로 한다.

## 웹정보 콘텐츠의 증가

웹을 통해 유통되는 DB의 수는 꾸준히 증가하고 있다. <그림 1>과 같이 DB의 유통 매체별 성장률의 추이를 보면, 온라인 매체를 제외한 CD-ROM과 기타 제공매체 등 오프라인 매체의 성장률은 2001년 이후 하락세를 지속하고 있는 반면, 온라인 매체의 성장률은 1999년 이후 꾸준히 상승세를 보이고 있다.

<그림 1> 제공 매체별 데이터베이스 서비스 변동 추이



기타 : 디스켓·마그네틱 테이프·배치파일·DVD 등

이나 상품을 사이버 상에서 주문하고 결제할 수 있다.

정보통신부의 IT839 등과 같은 강력한 정보통신 발전 정책에 따라 2010년 중반에는 유비쿼터스 컴퓨팅 환경이 조성될 것으로 기대되고, 상당 부분의 기존 오프라인 서비스는 온라인 서비스로 전환되고, 현재로서는 상상할 수 없는 많은 서비스들이 창출될 것이다. 이러한 차세대 서비스들은 웹정보콘텐츠와 더욱 쉽게 융합될 것이다.

### 웹정보콘텐츠 산업의 과제

미국 UC 버클리대학의 정보량 계량화 프로젝트에 따르면 전 세계적으로 매년 발생돼 유통되는 정보의 양은 17.9 exabyte(10의 16승 byte)이며, 이중 DB 형태로 유통되는 정보의 양은 0.09 exabyte로서 0.5%에 불과하며, 정보통신 환경의 발전과 함께 발생하는 정보의 양은 연평균 30% 정도 증가되고 있다고 추정하고 있으며, 매년 발생하는 정보의 90% 이상이 디지털 형태로 자기 매체에 저장되고 있다.

이는 웹정보콘텐츠의 대상 소스가 무궁무진하다는 것을 의미한다. 향후 다가올 유비쿼터스 컴퓨팅 환경에서 출현하게 될 유비쿼터스 DB는 제작자가 지급과 같이 인위적으로 정보를 정형화해 축적하지 않아도 유비쿼터스 컴퓨팅 요소들이 매년 수십 엑사바이트 급으로 발생하는 정보를 자동으로 정형화해 축적하고, 이 축적된 정보가 이용자에게는 보이지 않지만 무의식적으로 정보를 이용할 수 있게 하거나 혹은 DB가 스스로 이용자의 여건을 탐색해 여건에 맞는 정보를 제공하는 방향으로 발전할 것이라고 전문가들은 DB의 장기 로드맵을 예상하고 있다. 웹정보콘텐츠는 이러한 로드맵 상에서 유비쿼터스 DB로 지향하는 과정에서 중간 위치에 볼 수 있다.

이러한 발전 과정상에서 웹정보콘텐츠 부문의 해결 과제로는 양질의 대량 데이터 확보·품질 향상·데이터 포맷의 표준화·저작권 보호, 검색 시스템의 개선 등을 들 수 있다. 이러한 과제는 개별적인 문제가 아니고 상호 연관돼 있다. 첫째, 대량 데이터 확보 문제는 대량으로 발생하는 비정형 데이터를 어떻게 효과적으로 정형화시킬 것인가라는 문제로

요약되는데 이를 위해 공공 혹은 민간 부문의 정보 레포지토리와 정보사업자를 법제도적으로 연계할 수 있는 방법이 강구돼야 한다. 둘째, 품질향상 문제는 제공되는 웹정보콘텐츠가 정보로서 그 가치를 갖기 위해서 해당 웹정보콘텐츠에 수록된 정보의 재현율과 정확율을 제고시키는 연구 개발이 필요하다. 셋째, 데이터 포맷의 표준화는 상호 관련이 있는 웹정보콘텐츠의 제작측면이나 이용측면의 융합이라는 차원에서 필수적인 과제이다. 넷째, 저작권 보호는 원시 정보 소유권자, 웹정보콘텐츠 사업자, 최종 정보이용자 사이에 정보유통의 흐름이 원활하게 추진될 수 있는 방향으로 연구돼야 할 과제이다. 다섯째, 검색 시스템의 개선은 자연어 검색, 멀티미디어 콘텐츠 검색, 검색결과 의미론적 해석 등에 대해 해결돼야 할 과제이다.

## 국내 시장 동향

온라인 DB서비스 시장은 2004년 전년대비 매출액 기준으로 113% 증가한 것으로 조사됐다. 아울러 서비스 제공업체 수도 11.1% 증가, 제공 DB 수도 8.1% 증가한 것으로 나타났다. 지난 1992년부터 파악된 온라인 DB서비스 시장 동향은 <표 1>과 같으며, 연평균 증가율을

<표 1> 연도별 온라인 DB 서비스 시장 동향(2000년~2004년)

구분	연도					연평균 증가율 (1993~)
	2000	2001	2002	2003	2004	
DB수(종)	2,440	2,247	2,016	2,478	2,679	19.8%
증가율	10.9%	7.9%	10.3%	22.9%	8.1%	
제작기관수(개)	1,195	1,179	1,206	1,472	1,635	20.0%
증가율	14.1%	1.3%	2.3%	22.1%	11.1%	
매출액(억원)	1,787	1,622	1,171	1,589	3,384	26.1%
증가율	27.4%	-9.2%	-27.8%	35.7%	113%	

고려한 2000년 이후의 동향을 살펴보면 지속적인 성장세에서 2001년~2002년 감소세를 기록한 이후, 다시 2003년 이후 증가세로 돌아선 결과 서비스 제공업체 수와 DB 수의 증가율에 반해 시장 규모의 급성장은 온라인 DB 서비스 시장의 고질적인 문제였던 유료화 기

반이 정착되고 있다고 해석된다.

향후 국내 온라인 DB서비스 시장을 전망해 보면, DB 수는 전체적으로 연평균 9.7%의 상승세를 나타낼 것으로 예측되며, DB 제작기관 수는 7.5%, DB서비스 매출액은 22.8%의 비교적 높은 증가율을 기록해<그림 2>와 같이 2009년에는 약 9,000억원 규모의 시장으로 성장할 것으로 전망된다.

국내에서 제공되는 정보콘텐츠를 주제 분야별로 살펴보면, 비즈니스·경제 분야가 전체의 21.7%를, 레저·스포츠 분야가 11.4%를 차지하고 있다. 2004년 이후의 추이를 살펴보면 비즈니스·경제 분야가 비중이 점점 커지는 가운데, 학술 분야 또한 증가율은 작지만 지속적으로 늘어나고 있다.

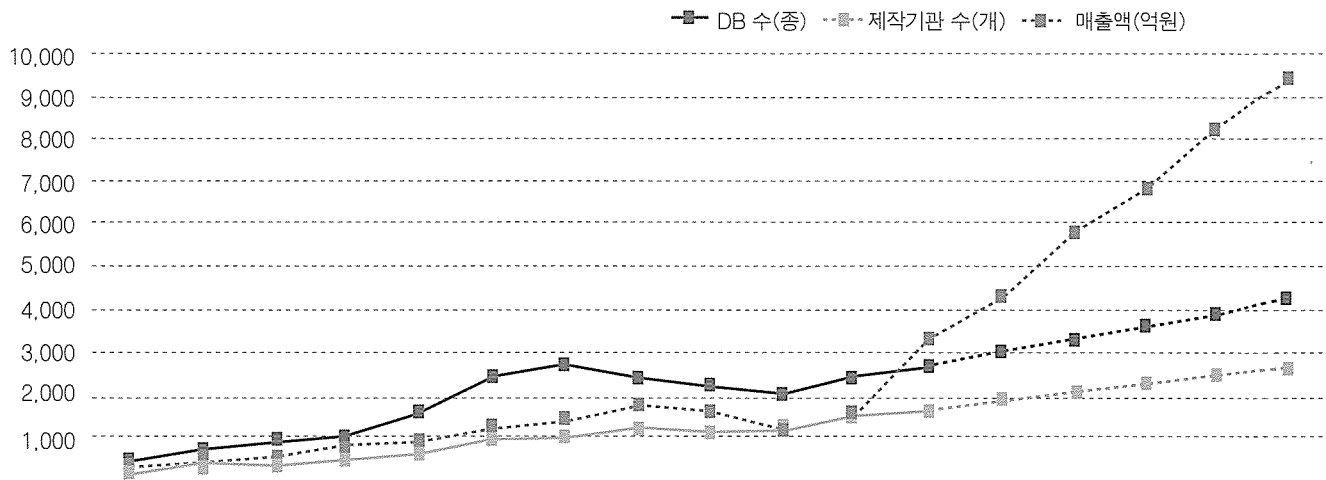
반면 쇼핑·생활 분야의 경우, 대부분의 매매 물품 가격, 상품 사진 등의 정보를 체계적으로 구성해 DB로 제공하고 있으나 2000년대 초 양산된 생활·쇼핑 분야의 온라인 DB서비스는 순수 정보제공으로 인한 매출 발생이 미비해 2003년 이래 급락세를 보이고 있다.

2004년 전체 DB 가운데 유료로 제공된 DB는 전체 56%에서 70%로 14% 증가한 것으로 나타났다. 이러한 변화는 주제 분야별 시장성장에도 변화를 준 것으로 나타났는데, 특히 수익 창출이 용이한 비즈니스 및 문화·예술·엔터테인먼트 분야에서의 약진

이 두드러졌다.

구체적으로 증권, 경제통계자료, 입찰, 특허 및 규격정보와 같은 비즈니스 분야는 전년대비 제공 DB수는 1.5배, 매출규모는 3배 증가를 보였다. 그리고 영화, 만화, 운세 등과 같은 문화·예술·엔터테인먼트 분야의 경우

〈그림 2〉 국내 온라인 데이터베이스 서비스 시장 전망



	1992	1993	1994	1995	1996	1997	1998	2,000	2001	2002	2003	2004	2005	2006	2007	2008	2009
DB 수(종)	460	714	907	1,061	2,488	2,789	2,440	2,247	2,016	2,478	2,679	3,054	3,313	3,349	3,984	4,256	
제작기관 수(개)	254	389	397	400	551	919	1,047	1,195	1,179	1,206	1,472	1,695	1,867	2,058	2,269	2,465	2,675
매출액(억원)	300	464	501	754	917	1,240	1,408	1,787	1,622	1,171	1,589	3,384	4,261	5,750	6,801	8,184	9,856

〈표 2〉 주제 분야별 정보콘텐츠 서비스 비율

분야	2001년	2002년	2003년	2004년	증감율
언론/미디어	6.7%	8.5%	5.3%	10.2%	4.9%
교육/취업	11.8%	12.1%	22.6%	9.8%	12.8%
비즈니스/경제	13.5%	15.1%	15.2%	21.7%	6.5%
문화/예술/엔터테인먼트	15.5%	12.6%	14.9%	8.5%	6.4%
레저/스포츠	5.5%	4.8%	5.1%	11.4%	6.3%
정치/사회	5.2%	4.4%	2.7%	10.4%	7.7%
쇼핑/생활	23.2%	25.1%	16.4%	9.2%	7.2%
컴퓨터/인터넷	10.6%	9.9%	8.0%	2.1%	5.9%
학술	8%	7.5%	9.8%	11.3%	1.5%
기타				5.4%	
합계	100	100	100	100	

자료 : 한국데이터베이스진흥센터

전년대비 제공 DB 수는 38% 감소했으나 매출은 2.7배 증가한 것으로 조사됐다.

이는 소규모로 제공되던 DB들이 점차 대규모 DB로 흡수돼 DB의 수는 감소했으나, 인터넷망의 속도개선과 더불어 제공되는 정보의 형태가 다양해지면서 2배 이상의 고성장을 이룬 것으로 풀이된다. 또한 문화·예술·엔터테인먼트 분야의 경우, 양질의 고가 DB서비스 증가로 인해 성장률이 높게 나타난 것으로 분석된다.

## 해외시장 동향

정보통신 기술의 급속한 발전과 함께 지식 정보 사회에 대한 요구가 본격화되면서 웹정보콘텐츠 산업도 성장·발전하고 있다. 웹정보콘텐츠 산업의 발전과 밀접한 관계를 가지고 있는 인터넷 보급률도 꾸준히 증가하고 있으며, 이와 더불어 각국의 전산화가 가속화되면서 세계 웹정보콘텐츠 산업도 점차 그 범위를 확대하고 있다.

Miniwatts International에서 발표한 Internet World Stats에 따르면 2004년 12월을 기준으로 세계 인구의 12.7%인 8억 1,000만명이 인터넷을 사용하고 있는 것으로 나타났다. 인터넷 보급률 상위 국가는 스웨덴, 홍콩, 미국, 네덜란드 등이며, 우리나라는 7위에 랭크됐다. 그러나 인터넷 이용자수 상위 국가는 1억9,000만명인 미국을 비롯해, 9,400만명인 중국, 6,700만명인 일본 등이며, 우리나라는 3,100만명, 브라질이 1,800만명으로 나타났다.

## 세계 시장

게일 디렉토리에 따르면 2004년 기준 세계에서 제작·유통되고 있는 DB 수는 약 1만 3,959개, DB 제작기관수 약 3,416개인으로 파악됐다. 이는 1993년(DB 수 : 8,261개, DB 제작기관수 : 2,744여개)과 비교할 때 DB 수는 약 1.7배, DB 제작기관수는 약 1.2배 증가했다. 1983년 대비 1993년 DB수가 2.7배, 제작기관 수가 6.5배 성장한 것과 비교하면, 다소 성장세가 둔화되고 있는 것을 알 수 있다. 이것은 1970년대 중반부터 1990년대 초반까지는 미국을 중심으로 전 세계의 DB 산업이 활성화된 시기로서 급격히 증가했음을 확인

할 수 있다. 그러나 1995년 이후 DB 수는 물론 제작·유통기관 수 전반에 걸쳐 증가 속도가 둔화되는 경향을 보이고 있다. 이는 정보기술 산업의 발전과 더불어 DB의 제공기반이 변화돼 오프라인 중심에서 온라인으로 점차 이동하는 현상이 두드러졌고, 이에 따라 시장의 확대보다는 제공기반별 비중의 변화가 이루어지고 있는 것으로 해석된다.

여기서 미국, 일본, 한국 그리고 세계 DB 산업을 비교·분석함에 있어, DB 산업 규모 비교 지표로 DB 수, 제작업체 수 그리고 매출액을 활용했다. 각 국가별 시장 비교 결과는 <표 4>와 같다.

미국은 전체 시장 규모면에서 유럽의 6.6배, 일본의 8.0배, 한국의 175배에 이르고 있으며, 생산성 측면에서는 일본의 2.2배, 한국의 54.3배에 이르고 있다. 한편, 한국과 일본을 비교해 보면, DB 수는 각각 2,191종과 2,478종으로 큰 차이를 보이지 않으나, 제작업체 수는 한국이 일본의 10.7배(2002년 8배), 매출액은 일본이 한국의 21.9배(2002년 28배)에 이르는 것으로 나타났다.

이와 같이 두 아시아 국가의 DB 종당 매출액을 비교해 보면 일본이 15억8,000만원으로 한국의 24배에 이르는 것으로 나타났다 일본은 DB의 품질과 전문성을 바탕으로 높은 생산성을 보이고 있으며, 산업의 경쟁력에서도 한국을 앞서 있는 것으로 보인다.

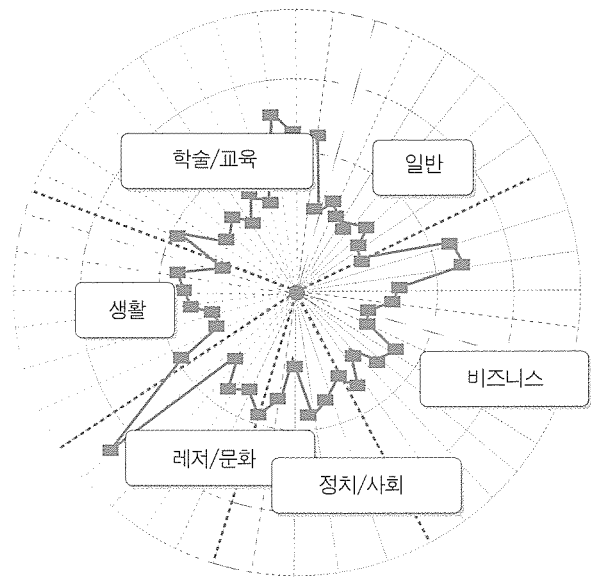
**미국시장**

<표 4> 국가별 DB 서비스 시장 규모('03)

구분	세계 <sup>1)</sup>	미국 <sup>2)</sup>	유럽 <sup>3)</sup>	일본 <sup>4)</sup>	한국 <sup>5)</sup>
DB 수(종)	13,959	8,000*	-	2,19	2,478
제작업체 수(개)	3,416	-	-	137	1,472
매출액	47조9,425억원 <sup>6)</sup> (402억\$)	27조8,723억원 <sup>6)</sup> (233억7,100만\$)	4조2,019억원 <sup>6)</sup> (27억9,900만\$)	3조4,757억원 <sup>6)</sup> (3,118억\$)	1,589억원
DB 종당 매출액	34억3,000만원	34억8,000만원	-	15억8,000만원	64억1,000만원

1) 세계 시장은 '게일 디렉토리 2005', 'Simba information, 2004.12', 'Electronic Publishing Service, 2004. 8', 'Global Entertainment and Media Outlook, PwC, 2004' 참조.  
 2) 미국 시장은 '게일 디렉토리 2005'에 발표된 세계 DB 중 미국에서 제작된 DB entry 수를 의미하며 매출액은 미국통계국의 온라인 정보 서비스 시장관련 자료 중 2003년 데이터베이스 서비스 매출액을 활용.  
 3) 유럽 DB산업은 IRN Research의 'European Online Information Market 2004' 자료를 참조했으며 제시데이터는 2003년 자료임.  
 4) 일본 DB산업은 '일본데이터베이스 백서 2004'를 참조했으며 제시데이터는 2003년 자료임.  
 5) 한국 DB산업은 '2004 국내 데이터베이스 산업 현황 및 전망보고서' 참조.  
 6) 환율은 100엔 = 1114.74원, 1달러 = 1192.60원, 1유로 = 1,501.24원 적용 (2003년 12월 31일 적용)  
 \*) 추정치

<그림 3> 온라인 정보콘텐츠 서비스 주제 분야별 분포



자료 : 한국데이터베이스진흥센터

<표 3> 세계 DB 서비스 시장

구분	2000년	2001년	2002년	2003년	2004년
데이터베이스 수(종)	12,417	12,959	16,417	18,214	13,959
제작업체 수(개)	4,017	3,879	4,042	4,094	3,416

자료 : 게일디렉토리 2005

미국 인구통계국(U.S. Census Bureau)에서 발표한 서비스업 연간 조사(Service Annual Survey) 결과에서 나타난 미국의 2003년 정보 산업의 변화를 살펴보면 <표 5>와 같다. 보는 바와 같이 미국 전체 정보 산업

에서 정보서비스 시장이 차지하는 비율은 10% 미만으로 방송, 출판 등의 타부문의 시장 규모가 높은 편이다.

그러나 1998년 이후 전체 정보 산업의 세부 분야별 비중은 <표 5>와 같이 영화·음반 산

업(NAICS 512)과 방송·통신 산업(NAICS 513)은 크게 변화가 없는 가운데 2000년대에 접어들면서 출판 산업의 비중이 조금씩 줄어드는 반면 정보서비스의 비중은 점차 늘어나고 있다.

이것은 출판을 통해 오프라인으로 유통되던 정보가 점차 온라인에서 제공되는 형태로 변화하고 있음을 나타낸다. 이러한 정보서비스 산업에서 온라인 DB 서비스에 해당하는 시장 부분을 살펴보면 <표 6>과 같다.

미국의 온라인 DB 서비스 산업은 2003년 233억7,100만달러로 전년대비 6.3% 증가했으며, 성장률도 2002년 4.5%에 비해 1.8%p 증가했다. 인구 1만명당 매출액도 2002년 대비 5.4%증가한 82만6,100달러로 조사됐으며, 성장률은 2002년의 3.2%보다 2.2%p 증가한 것으로 조사됐다. 즉, 매출시장 규모 및 인구 1만명당 매출액 모두 2003년도 전체 정보서비스의 증가율 0.8%에 비해 높은 것으로 조사돼, 향후 온라인 정보서비스 시장의 안정적인 성장을 기대할 수 있으며, 오프라인 출판시장이 점차 온라인으로 제공기반을 전환하면서 온라인 DB 서비스 시장의 성장은 지속될 것으로 전망된다.

<표 5> 미국 정보 산업 시장 규모

(단위 : 백만달러)

구분	1998년	1999년	2000년	2001년	2002년	2003년	
정보 산업(NAICS51)	687,991	767,379	845,687	868,917	875,323	889,276	
증감률	-	11.54%	10.20%	2.75%	0.74%	1.59%	
세부분야	출판 산업	200,576	218,124	232,069	231,714	230,916	232,427
	영화 및 음반 산업	59,404	64,023	67,745	69,152	74,020	77,901
	방송 및 통신	381,017	425,127	469,349	485,410	482,496	490,390
	정보서비스 및 데이터 처리 서비스	46,994	60,105	76,524	82,641	87,891	88,558

### 유럽 시장

유럽은 Bonnier, Thomson, Elsevier, Datamonitor 등 세계 유수의 정보서비스 업체가 활동하는 세계의 주요 정보 시장 가운데 하나이다. 이러한 유럽의 정보시장을 살펴보기 위해 기초 자료로 IRN 연구소에서 발간하는 '유럽 온라인 정보 시장(European Online Information Market)' 보고서를 활용했다. IRN은 영국 시장조사협회(BMRA)의 회원사

로서 영국을 포함한 유럽 전역의 관련 시장을 분석하는 전문 기관이다.

유럽의 온라인 정보 시장은 크게 과학·기술·의료 정보(STM) 시장과 법률·세금·규정(LTR) 등을 포함한 비즈니스 정보 시장으로

<표 6> 미국 DB 서비스 시장 규모

연도	2001년	2002년	2003년
DB 수	6,889	7,049	8,000***
성장률	-	2.3%	13%
매출액*(백만\$)	21,047	21,993	23,371
성장률	-	4.49%	6.27%
인구**	277,017,622	280,540,330	282,909,885
성장률	-	1.3%	0.8%
만명당 매출액/전체 인구 대비	75.98만\$	78.40만\$	82.61만\$
성장률	-	3.18%	5.38%
정보서비스 시장 성장률****	8.0%	6.4%	0.8%

\*) 자료 : Gale Directory of Database(2005), U.S. Census Bureau, Service Annual Survey(2004)  
 온라인 정보서비스 분야 가운데, 인터넷 접근 서비스, 정보검색서비스, 디렉토리 및 DB, 인터넷, 비디오·오디오 스트리밍, 증시사진정보 및 기타 서비스 포함  
 \*\*) 매년 1월 인구 : U.S. Census Bureau 참조  
 \*\*\*) 2003년 자료는 추정치, 2004년 DB 엔트리 수는 9,709개로 나타남  
 \*\*\*\*) 정보서비스 및 데이터프로세싱 서비스 시장 성장률

구분할 수 있다. 2003년 온라인 정보 시장은 2002년 대비 14% 성장한 27억9,900만(약 4조2,019만원)유로 규모로 나타났다. 이것은

는 전년대비 24%의 두 자리대 성장률을 보였다. 이러한 STM 정보 시장의 성장은 정보서비스 전체 시장 규모 가운데 온라인 정보서비스 시장의 비중이 높아진 것 때문으로 보인다.

이러한 온라인 정보서비스 시장을 국가별

로 비교해 보면, 영국, 프랑스, 독일의 3국이 전체 온라인 정보시장의 78%를 차지하고 있어, 실질적으로 유럽 시장을 주도한다고 볼 수 있다. 그 중에서도 특히 영국은 유럽 전체 STM 시장의 36%, 비즈니스 시장의 37%를 차지하고 있어 유럽의 정보서비스 시장의 핵심을 이룬다. 영국의 온라인 서비스 시장은 <그림 4>와 같이 1999년~2001년까지 급속한 성장을 보이다가 2001년~2002년 성장이 둔화되기 시작했다. 특히 비즈니스 정보의 경우에는 2003년에 마이너스 성장을 기록했다.

### 일본 시장

일본DB진흥센터(JDPC)의 발표에 따르면 <표 7>과 같이 2003년 일본 DB 산업의 매출액은 3,118억엔으로 전년대비 14.2% 증가한 것으로 나타났다. 이는 DB 수와 제공기업 수가 모두 하락한데 반해 두자리 대의 성장률을 보인 것으로 눈여겨 볼만하다.

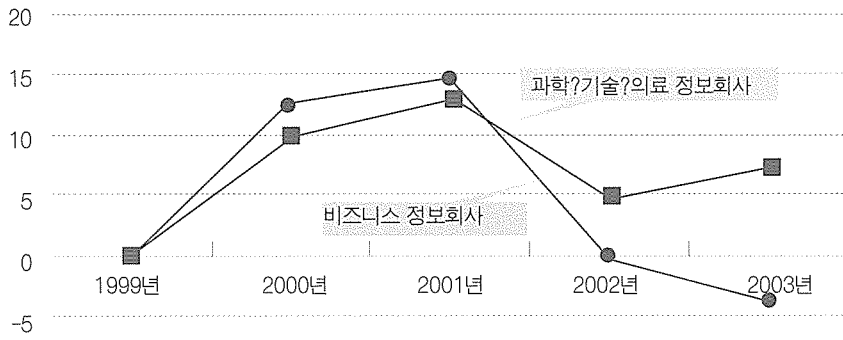
일본의 DB 서비스의 주제 분류별 제공 현황을 살펴보면, 전체 DB의 32%를 차지한 비즈니스 분야가 가장 많은 것으로 조사됐고 다음으로 일반(29%), 학술(28%) 순으로 나타났다.

한국의 26배, 일본의 1.2배, 미국의 0.15배에 이르는 시장 규모이다. 정보서비스 시장의 전체 성장률을 비교해 보면, 온라인 정보시장의 68%를 차지하고 있는 비즈니스 정보 시장은 3%의 성장을 보였으며, STM 정보는 6%의 성장률을 보였다.

온라인 정보 시장의 주요 분야별 시장 성장률은 비즈니스 정보는 전년대비 9%의 성장을 보이며 한 자리수에 머무른 반면, STM 정보

<그림 4> 영국 정보서비스 업체의 매출 성장률

단위: Sales(Change year-on-year)



## 제작 · 유통환경

인터넷의 발전으로 정보 콘텐츠의 제작 · 유통 환경은 과거 PC통신 시절보다 대량의 디지털 정보를 양산시켜 정보 콘텐츠 제작에 있어서 보다 용이한 환경을 제공하고, 정보 콘텐츠 유통 구조 제공을 위한 다양한 채널이 형성되고 있다.

정형화된 웹정보콘텐츠로 구축되기 위해서는 웹정보 콘텐츠의 기획 · 설계, 원시정보 수집, 정보의 가공 및 저장의 단계를 거치게 된다.

웹정보콘텐츠의 기획 · 설계 단계에서는 콘텐츠 구축의 목표를 명확히 해 마켓 포지셔닝을 하고, 구축된 콘텐츠의 이용 볼륨의 확대, 이용 대상의 극대화 및 콘텐츠 생산성의 확대에 대해 고려해야 한다. 이러한 사항을 고려해 데이터에 대한 논리적 · 물리적 모델링을 수

행하고 타겟 이용자의 특성에 맞추어 인터페이스를 설계한다. 특히 이 단계에서는 네트워크 환경의 변화에 탄력적으로 대응해야 하는데, 제작하고자 하는 웹정보가 유통될 네트워크의 종류에 따라 기획 · 설계 방향이 매우 달라지기 때문이다.

원시정보 수집 단계에서는 웹 에이전트의 사용 가능성, 디지털 원시정보 수집의 가능성, 원시정보의 수집의 비용, 공공부문의 정보의 상업적 이용 가능성 등을 고려해 프로세스 체계를 수립한다. 최근 사회 전반적으로 정보통신 기술의 적용이 보편화되면서 원시정보 자체가 디지털 형태가 많기 때문에 제작자 입장에서는 비용 절감 차원에서 적극적으로 검토 하되 저작권 문제에 대해 고려해야 한다.

정보의 가공 및 저장 단계에서는 우선 메타 데이터의 가공 정도 목표를 설정해야 한다. 예를 들어 사용할 분류 테이블과 통계 용어집의

활용 여부, XML 혹은 PDF와 같은 원문 정보의 형식 등을 사업성, 인력의 수준 및 소요 예산 등을 고려해야 한다.

최근 정보통신 기술의 발달에 따라 웹정보의 제작과 관련된 소프트웨어와 하드웨어의 기술이 동반해 발전하고 있다. 특히 인하우스 DB를 구축해 기업 활동에 전략적으로 활용하는 기업의 지식 관리 활동이 대표적이다. 기업 활동에 필요한 정보를 지식화해 전략적으로 활용하기 위한 BI(Business Intelligence), 고객의 충성도에 대한 특성을 DB로 구축해, 고객에게 맞춤형 서비스를 제공하기 위한 CRM (Customer Relationship Management), 기업내부의 정보자원을 전략적으로 활용하기 위한 ERP(Enterprise Resource Planning), 영업 직원의 업무를 지원하기 위한 SFA(Sales Force Automation) 등이 대표적인 사례이다.

1990년대 하이텔 · 천리안 · 나우누리 등의 PC 통신망을 출현으로 정보 콘텐츠 유통의 개념이 일반 국민에게 보급되기 시작했으며, 2000년대 초 인터넷이 보급되면서 웹정보 콘텐츠 제공 환경이 본격적으로 자리 잡기 시작했다.

정보 콘텐츠 시장은 현황 조사가 처음으로 실시됐던 1992년에는 388억원 규모를 이루었으며, 2004년에는 3,384억원으로 8배 이상 성장했다. 그러나 정보 콘텐츠 제공 환경이 PC통신에서 인터넷으로 바뀐 전환기(2000~2002년)에는 시장 규모가 일시적으로 축소되기도 했으나 웹정보 콘텐츠가 정착하기 시작한 2003년부터는 다시 성장세를 기록하고 있다.

인터넷으로의 전환기에 시장 규모가 축소된 가장 큰 원인으로는 PC통신기반에서는 PC통신사(한국통신(현 KT) · 데이콤 · 나우콤 등)에서 이용요금을 일괄적으로 징수하는 과금 형태를 유지해 웹정보콘텐츠 제공업체가 안정적으로 수입을 확보할 수 있는 채널을 유지했으나, 개별 과금체계를 요구하는 인터넷기반에서는 유료화 방안이 웹정보콘텐츠 제공업체의 숙제로 부담됐으며 무료정보의 난립으로 인해 상용 웹정보콘텐츠 이용에 대한 필요성이 희석되며 결국 유료화 마인드가 퇴보하는 결과를 낳았다. 그러나 안정적인 온

<표 7> 일본 DB 산업 규모

구분	2000년	2001년	2002년	2003년
DB 수(개)	2,458	2,552	2,379	2,191
성장률		3.82%	6.78%	7.90%
매출액(억엔)	2,916	2,979	2,730	3,118
성장률		2.16%	8.36%	14.21%
제공기업 수(개)	176	186	179	137
성장률		5.68%	3.76%	23.46%

자료 : 일본DB진흥센터, 데이터베이스백서

<표 8> 일본 DB 서비스 주제 분야별 비중

주제	비즈니스	일반	학술	사회 · 인문	기타	합계
DB수(개)	70	637	615	190	49	2,191
비율(%)	32	29	28	9	2	100

자료 : 일본DB진흥센터, 데이터베이스백서

라인 결제 시스템이 도입되고, 이용자의 유료 정보사용에 대한 거부감이 해소돼 웹정보 콘텐츠의 매출이 제고되고 있다.

현재 국내의 웹정보콘텐츠의 유통 구조는 소비 지향 중심으로 발달돼 있다. 생활, 오락 등 일상생활 분야의 웹정보콘텐츠는 다음·네이버·야후 등과 같은 인터넷 포털 사이트를 중심으로 형성돼 있다. 반면, 전문 지식형 웹정보콘텐츠는 일부 민간 사업자가 제공하고 있으나 규모가 영세하고, 대부분 공공 기관이 제공하는 상황이다. 다시 말하면 전문 지식형 웹정보콘텐츠는 미국이나 유럽의 Dialog나 Factiva와 같은 대형 정보 유통 벤더가 국내에는 부재하므로 전략적 육성 정책이 필요하다.

한편 최근 웹정보콘텐츠 유통 시장에서 새롭게 등장한 신디케이터(Syndicator)는 제작업체들에게 정보콘텐츠를 위탁·양도 받아 이들을 통합해 새롭게 상품화된 정보콘텐츠를 구축해 개인이용자나 기업체를 대상으로 서비스를 제공하고 있다. 그러나 이들 신디케이터 사업자의 규모가 영세해 본연의 웹정보 콘텐츠 신디케이션 업무를 원활히 수행하고 있다고 보기는 어렵다.

### 제작·유통 환경 전망

향후 정보콘텐츠 제작 환경은 전문화, 고도화돼 첨단 입력 인터페이스(음성인식 입력, 영상인식 입력)기술, 인력중심의 입력환경, 전자문서관리시스템(Electronic Document Management System : EDMS)의 발달 등을 통해 데이터 입력 시간을 더욱 단축되고 더욱 쉽고 편리하게 정보콘텐츠 제작이 가능해질 것으로 기대되고 있다. 또한 대용량 데이터 처리, 고속 데이터 추출, 자가 관리형(Self-Managed) 데이터 모델링 기술의 발전을 통해 e러닝, GIS와 같은 전문분야를 겨냥한 특화된 정보콘텐츠 제작 시스템이 대두될 것으로 전망된다.

정보콘텐츠 유통 환경은 인터넷과 함께 최근 확산되고 있는 모바일 환경이 자리를 잡아 가며, 유통 시장을 더욱 확대해 나가고 있다. 또한 유비쿼터스 기술의 도입, 광대역 모바일 기기의 확산, 체계화된 과금 체계 등은 웹정보

콘텐츠 서비스 환경을 더욱 개선시킬 것으로 기대되고 있다.

2010년 이후 유비쿼터스 네트워크 환경이 성숙되면, 정보콘텐츠의 유통 환경은 더욱 고도화될 것이다. 이러한 환경에서는 정보 콘텐츠 간 융합과 다른 서비스와 융합 현상이 더욱 용이하고 새로운 비즈니스 모델에 의해 유통될 것이다.

웹정보콘텐츠의 제작·유통 시장은 정보의 생산·수집·가공·축적·검색·유통·제공, 이용을 위한 모든 경제 주체의 활동이 집약된 정보통신기술의 집합체로, 정보의 교류를 필요로 하는 모든 산업 부문에 적용이 가능해 그 전망이 매우 밝은 편이다. 특히 최근 사회적으로 고급 전문 인력의 과잉 공급 현상은 웹정보콘텐츠 부문의 전문 분야에서 수용할 수 있는 호기이고 전문 인력의 고급 지식과 웹정보 콘텐츠를 활용해 신규 지식 서비스를 창출할 수 있는 기회를 제공할 것이다.

그러나 전술한 바와 같이 국내 웹정보콘텐츠

(그림 5) 웹정보콘텐츠 제작·유통 환경 발전 과정

구분	1990년	2000년	2010년	
	정보	지식	지능화	
웹정보콘텐츠 제작 환경	인력 중심	입력장치(OCR 등) 발달	음성/영상 인식 입력 특화된 DB 제작 시스템 자가관리형 데이터 모델링 기술	
데이터베이스 유통 환경	PC 통신	인터넷	유무선 통합	유비쿼터스 네트워크

자료 : 일본데이터베이스진흥센터, 데이터베이스백서

츠의 소비 지향 중심의 유통 구조, 공공 기관의 무료 웹정보 콘텐츠 공급, 기업 이용자의 정보 가치에 대한 인식부족 등 유통 구조 측면에서 해결해야 할 과제가 산적해 있다.

### 기술동향

공개 소프트웨어의 개발 보급은 국가의 정보 산업 육성의 기본 정책 측면에서 매우 중요한 과제이다. 웹정보콘텐츠를 인터넷에서 보다 용이하고 확실하게 획득할 수 있게 하는 DOI(Digital Object Identifier)와 학술 논문 등의 검색에서 유용성이 기대되는 OpenURL 등에 대해 소개한다.

### 공개 DBMS와 상용 DBMS 벤더의 대응

공개 DBMS로는 여러 종류가 있으나 미국을 중심으로 일반화되고 있는 것은 MySQL, PostgreSQL, InterBase가 대표적이다. MySQL은 현재 웹을 중심으로 한 정보네트워크에 적용되고 있으며, 매월 50만에서 100만 모듈이 복제돼 보급되고 있다. PostgreSQL은 비교적 인지도가 높고, 현재 저렴한 가격에 고기능 DBMS로서 기능 강화가 진행 중이다. InterBase는 상용 버전과 공개 버전 간에 어느 정도의 기능 차이가 있으나 확실하게 알려져 있지 않다. 이들 공개 DBMS의 품질 평가 결과가 공표되고 있는데, MySQL의 경우 소스 코드 1000행 당 버그가 0.09개로서 상용 DBMS 평균 버그가 0.57개에 비하면 품질이 우수하다. 실제 적용 사례로는 YahooUSA가 MySQL을 폭넓게 이용하고 있다고 발표했다.

한편 RDBMS 시장을 리드해 온 오라클과 같은 상용 벤더들의 대응도 치열하다. 오라클

10g에서 실현한 128노드의 클러스터링 기술은 공개 DBMS에서는 실현 불가능한 기술로 보여진다. 대규모 그리드 컴퓨팅을 실현하기 위해 개발된 최초의 플랫폼으로서 종래 학술 연구 영역에서 이용돼 온 그리드 컴퓨팅 기술을 대기업 정보시스템에 적용할 수 있도록 지원한다. 그리드 컴퓨팅이란 광범위하게 분산된 하드웨어를 하나의 가상 컴퓨터로 간주해 집중적으로 관리하는 기술이다. 현재 기업의 정보시스템은 대개 메모리 50%, CPU 20% 수준에서 사용되는 비효율적인 환경에 놓여 있다. 이를 하나의 그리드로서 배치를 최적화함으로써 효율적인 데이터 처리가 실현될 수 있다. 아울러 XML 기능을 강화했고, 저장

DB의 용량의 상한선을 128 테라바이트까지 확대했다. 그리드 연결을 구현하기 위해 DB의 이행, 물리적인 재배치 기능과 고속 기능이 구현됐다. 운영 관리 비용을 절감하기 위해 많은 물리 스토리지, 논리적 DB, 개별 애플리케이션을 집중적으로 관리하고 시스템을 다운시킬 수 있는 부하 증가에 대비한 기능을 증설했으며, 장애 대비 자동 복구 기능을 구현했다.

### XML 웹서비스의 보급

이용자의 애플리케이션 소프트웨어를 웹상에서 연결시키는 기술인 XML 웹서비스는 XML을 이용해 분산 관리되는 DB의 콘텐츠의 표현 규칙이 일반화된다는 것을 전제로 한다. 기존에는 EDI 등 기업간 혹은 시스템간 데이터 교환에 있어서 항상 정보 제공측과 수신측의 교환 형식 즉 규칙에 대한 개별 확인이 필요했다. XML 웹서비스에서는 이러한 파일을 DB 제공측에서 WSDL(Web Services Description Language)로 규정하고 어떠한 서비스가 존재하고 있는가를 고지하기 위해 UDDI(Universal Description, Discovery and Integration)도 준비하고 있다. XML로 구성된 데이터를 이용 목적, 애플리케이션 소프트웨어를 특정하기 위해서는 최적의 유통 환경을 제공한다. 기존 시스템간 정보 연계에 있어서 인터페이스 설계와 구현이라는 공정이 대폭적으로 축소될 수 있기 때문에 소량 데이터의 교환에 대해서도 충분한 효과를 발휘한다. 기존에는 키워드, 기업명 등으로 검색하고 검색 결과를 Cut and Paste 방법으로 필요한 항목에 입력 혹은 프린트 출력해 판매관리 시스템에 재입력 하는 반복적 업무를 피할 수 있다. 이미 광범위하게 활용할 수 있는 개인 정보는 인터넷에 유·무료에 관계없이 공개되고 이러한 정보를 독자적으로 관리하는 어플리케이션도 다양하다. 구체적인 예를 들면, 주소와 함께 지도를 관리하는 학교시스템, 거래처의 기업개황을 입력하는 판매관리시스템, 제조사 제품 부품 데이터를 사용하는 CAD 시스템 등에서 현재는 개별적으로 필요한 데이터의 입력표를 작성해 개별 입력을 수행하거나 특정 제조사와 계약에 의해 데이터 수집을 위한

인터페이스 소프트웨어를 이용하고 있으나, XML 웹서비스는 이러한 과제를 용이하게 해결할 수 있다. 정보의 XML화가 발전하고 있으므로 어플리케이션 측의 XML 웹서비스 대응이 기대된다.

### 정보검색기술 동향

필요한 정보에 가능한 짧은 시간과 노력으로 접속하고자 하는 정보 검색의 요구에 대해 인터넷 기술은 집중 관리가 불가능할 정도로 엄청난 양의 정보가 전 세계에 산재된 웹 서버에 축적돼 있고, 이들 정보를 링크해 대응할 수 있다. 하이퍼링크의 개념은 URL에 의한 상호 연계로 어느 정도 실현됐으나 실제 URL에 의한 링크만으로 원하는 정보를 검색하는 것은 한계가 있고, 보다 효과적인 방법이 요구되고 있다. 이러한 정보 연계, 링크 과제에 대한 두 가지 접근 방법이 주목되고 있다. 소재 장소를 나타내는 URL이 아니라 대상이 되는 메타 정보를 주고받는 OpenURL과 대상이 되는 정보를 특정하는 식별자를 이용하는 DIO(Digital Object Identifier)가 대표적이다.

### OpenURL

인터넷에서 검색은 검색 엔진을 이용한 프리키워드 검색이나 디렉토리 검색을 하는 것이 일반적이다. OpenURL은 검색하고자 하는 논문 등의 내용을 나타내는 메타데이터를 직접 URL로 기술하고 서버에 송신하는 기술 방법을 정한 것이다. 이 방법을 사용해 정보 원인 오리지널 데이터에 링크시키는 것이 가능하다. 이러한 링킹 서버와 링크 타겟·정보 원으로부터 구성된 서비스 패키지가 제공되고 있다. 조회에 사용되는 OpenURL의 일례로서 최초에 링킹 서버의 URL이 기술되고, 다음에 정보원에 관한 검색 조건이 메타 데이터로서 [데이터 명칭 : genre 등]=[값, article 등]의 구성으로 되어 있다. 링킹 서버는 이 메타데이터를 이 메타데이터를 기본으로 대상이 되는 정보원을 포함한 링크 타겟에 연결돼 필요한 데이터를 제공한다. 링크 타겟은 초록 DB, 전자저널 온라인 서점 등 전체 연결 시스템이 대상이며, 개별 인증에 의해 구독권의

유무에 의한 접근 제어, 대체 문헌 가이드 등의 기능을 제공한다. 링크 서버의 서비스 상품으로서 SFX, LinkFinder Plus, 1 Cate 등이 있다.

### DOI

DOI는 책의 ISBN, 정기간행물의 ISSN과 같이 텍스트·화상·동영상 등 전체 관리 대상이 되는 전자 객체에 식별 번호를 부여하는 관리 코드 체계 DOI 코드와 그 하나로 결정되는 코드로부터 저작물이 존재하는 URL로 변환시키는 시스템인 DOI 디렉토리, 저작물이 관리되는 서버로 구성된다. 전체 저작물의 식별과 소재 관리를 일원화해 디지털 저작물의 유통과 권리 관계의 제어를 목표로 한다.

DOI 코드는 DOI 관리기관에서 국제DOI기관(IDF)가 발행하는 출판 코드인 접두 코드와 출판사가 독자적으로 부여하는 개별 인식 코드인 접미 코드로 구성된다. 자체는 어떤 의미도 부여되지 않는 영숫자 문자열이지만 출판사 코드는 지역과 업종을 식별하는 디렉토리를 포함하고, 접미 코드는 종래 식별 코드를 사용하는 경우, 코드의 종류를 선택사항으로 사용할 수 있다. DOI 디렉토리는 국제 DOI재단에 의해 관리되고 출판사가 DOI와 그 실체가 존재하는 URL을 항상 갱신·등록하고 있다. 이용자로부터 전달된 DOI에 대응하는 출판사에서 변환한 URL을 송신하는 기능을 갖고 있다. 저작물 서버는 출판사가 독자적으로 관리하는 서버이고, 문헌 등의 실체가 이동할 때에는 DOI 디렉토리에 대한 갱신은 출판사의 책임으로 수행된다. DOI 디렉토리로부터 전송된 변경후의 URL에 의해 실체 데이터를 이용자에게 전송하는 기능을 갖는다.

DIO는 국제적인 상업 출판사에 의해 생산자 즉, 출판사가 직접 소비자에게 출판물을 판매하고 요금을 징수하는 것을 목적으로 개발됐다. 이러한 전자적 유통 방식에 의해 정보의 중개업자인 서점, 대리점 도서관의 역할 재편이 예상된다. 물론 현재는 디지털화에 의한 저작권의 취급, 이용허락 관리 등 해결과제가 잔존한다.



## 향후 전망 및 과제

전세계적으로 생성되는 정보 중에서 웹을 통해 유통되는 정보는 0.34% 정도이며, 이중에서도 정형화돼 이용자가 편리하게 이용할 수 있는 웹정보콘텐츠의 비중은 더욱 낮다. 이러한 현상은 기본적으로 인터넷이 안고 있는 한계, 즉 발생하는 정보를 웹으로 유통시키기 위해서는 인위적으로 입력, 가공하는 과정이 필수적으로 따라야 하며, 실세계 공간과 웹이라는 가상공간이 결합될 수 없다는데 근본 원인이 있다. 기술의 발전에 따라 현재의 웹 환경은 무선 네트워크 환경을 거쳐 궁극적으로 유비쿼터스 컴퓨팅 환경 아래 정보 콘텐츠가 생성되고 유통될 것이다. 즉 이용자에게는 보이지 않지만 무의식적으로 정보를 이용할 수 있게 하거나 혹은 DB가 스스로 이용자의 여건을 탐색해 여건에 맞는 정보를 제공하는 유비쿼터스 DB로 발전하게 될 것이다. 이러한 정보 콘텐츠 제작 유통의 발전 과정상에서 해결돼야 할 과제는 대략 대량의 정형 데이터 확보, 품질 향상, 데이터 포맷의 표준화, 검색 시스템의 개선 등으로 요약된다.

**데이터 정형화**= 현재 대부분 정보의 1차 생성은 비정형 형태의 데이터이다. 실세계에서 말을 하거나 글을 쓰거나 어떤 행위를 할 때 발생하는 정보원의 형태는 비록 전자적 매체에 수록돼 유통된다고 할지라도 대부분이 비정형 형태에 의존한다. 예를 들어 웹상에서 문서 정보를 검색할 경우 PDF 파일 형태를 자주 접하게 되는데 이는 디지털 형태의 정보이지만 정형화된 데이터로 볼 수 없다. 정보로서 그 가치를 극대화시키기 위해서는 문자 형태의 이미지 데이터를 문자 데이터로 변환시키는 기술이 있어야 하는데, 대표적인 것이 OCR 기술이다. PDF 파일에 포함된 모든 문자 정보를 디지털 환경에서 자유자재로 활용할 수 있는 정형 데이터로 전환시켜야 한다. 이러한 과정은 음성 정보나 영상 정보의 경우에도 동일하게 적용된다. 즉 음성 인식·패턴 인식 등의 기초 기술이 필요하게 된다. 대량의 정형 데이터 확보를 위해서는 비정형 정보를 정형화된 정보로 변환시키는 기술이 더욱 세련되고 고

도화돼야 한다.

**데이터 품질 관리**= 발생 정보의 정형화 문제와 연장선상에서 고려해야 할 것이 데이터 품질 문제이다. 전세계적으로 가장 큰 규모의 웹 포털 사이트인 구글의 최대 과제가 웹상에서 대량으로 제공되는 정보를 수집해 고도의 품질을 유지하는 DB로 구축해 효율적인 검색을 가능하게 하기 위해 다양한 웹정보를 자동적으로 인식, 분류하고, 태깅 작업을 자동화하고자 하는 것이다. 구글의 사례를 보듯이 웹상에서 유통되는 관련 정보를 DB로 구축하기 위해 특정 목적을 위한 범위 한정, 개별화, 정보 필터링 혹은 선택 등의 기능이 필요하다. 정보의 범위를 한정할 때 필요 정보를 누락시키지 않고 정보 범위를 설정하며, 정보 콘텐츠에 카테고리나 분류를 부여하는 것이 웹정보콘텐츠의 품질 향상에 있어서 매우 중요한 과제이다. 정보량의 증가에 따라 범위의 한정, 선택, 분류 및 태깅 작업을 정확하게 수행할 수 있는 자동화 기술이 더욱 요구된다.

**데이터 포맷 표준화**= 정형화된 웹정보콘텐츠가 효과적으로 공유 및 교환되기 위해서는 데이터 포맷이 표준화돼야 하는데, 이와 관련해 가장 관심 있게 보아야 할 것이 메타데이터 레지스트리(Metadata Registry)와 XML 기반 웹 DB이다. 메타데이터 레지스트리는 메타데이터의 명세를 표준화하고, 중앙등록기관을 통해 등록·인증하도록 해 데이터의 표준화를 포괄적으로 지원하는 시스템이다. 'ISO/IEC JTC1/WG2(데이터 관리 및 교환)'에서 국제표준 'ISO/IEC 11179 - Metadata Registry'가 국제표준으로 제정됐고, 미국 환경청의 EDR(Environmental Data Registry), 미국 교통부의 ITS(Intelligent Transportation System) 데이터 레지스트리 등 여러 공공 기관이 국제표준을 채택하고 있다. 메타데이터 레지스트리는 전자정부·전자상거래·사이버교육·데이터웨어하우스·ERP 등 웹 기반에서 발생하는 데이터의 이질성 문제를 해결할 수 있는 중요한 개념으로 인정받고 있다. 한편 XML 기반 웹 DB는 차세대 웹 DB로서 현실적인 활용 방법론에 대

한 연구가 진행 중이다. 마이크로소프트의 Access 2002와 SQL Server 2000, 오라클의 10g 등 DBMS가 XML 도큐먼트를 지원하고 있다. XML은 웹상의 다양한 데이터의 교환 수단으로서 널리 활용될 것이며, 머지않아 대량으로 유통될 XML 기반의 문서를 효율적으로 관리할 수 있는 XML 기반의 웹 DB가 요구되고 있다. 현재는 XML 기반의 웹 DB와 관련한 기본적인 표준들이 제안되고 다양한 저장 모델, 검색 기법들과 함께 상업용 제품들이 제시돼 기술, 성능 및 실용적 측면에서 평가를 받고 있다.

**정보검색 시스템 개선**= 정보량의 증대와 정보 내용의 다양화에 따라 보다 효율적인 정보 검색시스템이 요구된다. 종래의 정보검색 서비스와는 달리 웹정보 포털 검색 서비스인 구글 혹은 웹정보의 망라적(Exhaustive)인 보존을 위한 Internet Archive의 검색 시스템인 Wayback Machine 등에서는 대량의 퍼스널 컴퓨터를 이용해 대용량 정보처리를 하고 있는데 이러한 신기술은 유비쿼터스 컴퓨팅 환경에 있어서 정보검색 시스템의 프로토타입에 대해 시사하고 있다. 정보검색 시스템의 검색 효율을 향상시키기 위해서는 콘텐츠 작성 시 더블린 코어와 같은 메타데이터의 채택, 콘텐츠 식별 번호의 부여에 의한 관련 정보와 링크 혹은 색인, 분류에 의한 콘텐츠의 체계화 등이 전제 조건이다. 이러한 전제하에 자연어 검색, 이미지 콘텐츠 검색, 검색 인터페이스 개선, 이중의 정보원 링크에 의한 조직화, 특정 목적을 위한 범위 한정 등의 검색 기능 개선, 검색 인터페이스의 커스터마이징, 콘텐츠의 자동 필터링 및 선택 기능 등 정보검색 시스템의 개량에 다양한 시도가 기대된다. 특히 최근 대학이나 연구소를 중심으로 차세대 웹으로서 검색 결과의 의미론적 해석을 컴퓨터가 자동적으로 수행할 수 있도록 하는 시멘틱 웹(Semantic Web)이 국내외에서 연구되고 있다. 향후, 정보검색 시스템이 검색 결과의 의미론적 해석, 검색 질문식의 자동 전개, 인물·장소·사물 등의 특정화, 검색결과의 링크 기능 등을 어디까지 자동화할 수 있는가 등이 과제이다. 🌐