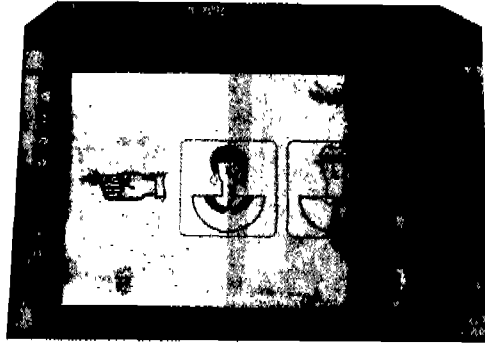


국내외 데이터 베이스 현황과 전망



강 영 채

공학박사 · 과학평론가

※ 머리말

“데이터 베이스란 어느 특정 조직의 응용업무에 공동 사용하기 위하여 운영상 필요한 데이터를 완벽화(完壁化), 비중복화(非重複化), 구조화(構造化)하여 컴퓨터 기억장치에 저장한 데이터의 집합체”라고 정의되어 있다.

정의에서 특정 조직이란 어떠한 목적으로 소규모의 데이터 베이스를 사용하는 개인이나 또는 대규모의 공용 데이터 베이스를 사용하는 기업체 모두를 총칭하며, 운영상의 데이터란 이러한 특정 조직의 응용업무에 필요한 데이터를 의미한다.

또한 이것은 단순히 외부세계로부터 컴퓨터에 입력된 정보나, 시스템에서 출력된 정보 뿐만 아니라 시스템에 저장된 기존의 정보로부터 응용분야에 적합하도록 유도된 새로운 정보도 포함한다.

완벽화는 특정 조직에서 필요로 하는 모든 데이터를 데이터 베이스에 저장하는 것을 의미한다. 예를 들어 고객 데이터 베이스는 모든 고객

과 각각의 고객에 대한 완벽한 정보를 갖고 있어야 한다. 이러한 완벽화는 특정 조직에 대한 모든 정보의 요구를 얻기 위하여 필요하다.

비중복화는 같은 데이터가 두 번 이상 저장되지 않는 것을 의미하며, 이러한 특성은 일치성(一致性)을 보장하고 기억장소를 절약하기 위해서 필요하다. 그러나 비중복화가 항상 쉽게 이루어지는 것은 아니며, 사용자가 빠른 응답시간을 요구할 경우와 데이터의 회복을 보장하기 위해서는 최소한의 중복은 필요 불가결하다.

구조화는 데이터를 효과적으로 처리할 수 있는 방법으로 데이터 베이스를 기억장치에 저장하는 것을 의미한다.

한편 일본 산업구조심의회 정보산업분회 보고서에서 정의한 바에 의하면 데이터 베이스란 데이터를 정리 통합하여 컴퓨터로 처리 가능한 형태로 만든 정보 화일 또는 그 집합체이다. 대량의 정보가 범람하는 오늘날 개개인 모두가 필요한 정보를 독자적으로 찾는 데는 한계가 있지만 데이터 베이스를 활용함으로써 지금까지 단독으

로는 입수가 불가능했던 정보들의 효과적인 활용이 가능해지고 정보의 효용가치를 극대화시키는 동시에 정보 입수를 위한 중복 투자의 손실을 줄일 수 있다. 데이터 베이스는 중요한 사회적 자원의 하나로 인식되어질 수 있다. 이처럼 데이터 베이스란 하드웨어, 소프트웨어, 검색원과 함께 정보화 사회를 지탱하는 지주의 하나로서 다양한 정보를 효율적으로 활용하는 데 있어서 필요 불가결한 데이터 베이스의 정비는 정보화 사회를 향한 전제조건이라고 되어 있다.

데이터 베이스의 어원은 1963년에 시스템 디벨로프사(System Development Corporation)에서 개최한 “Development and Management of Computer-Centered Data Bases” 심포지엄에서 처음 사용하였으며, 데이터 베이스 독립성의 개념을 도입한 데이터 베이스의 모체는 1963년에 제너럴 일렉트릭사에서 개발한 “Integrated Data Store”에서 찾아볼 수 있다. 또한 데이터 베이스 표준화의 작업은 1965년의 CODASYL의 DBTG(Data Base Task Group) 그룹에서 시도되었다.

그후 데이터 베이스에 관한 연구가 꾸준히 진행되었는데, 대표적인 데이터 베이스 시스템으로는 관계 데이터 베이스 시스템, 망구조 데이터 베이스 시스템, 계층구조 데이터 베이스 시스템이 있다. 최근에는 관계 데이터 베이스 시스템이 주류를 이루며, 계층구조 및 망구조 데이터 베이스 시스템보다 폭넓게 사용될 전망이다.

데이터 베이스의 초기과정은 도서관의 자동화 과정과 일치한다. 자료의 축적과 함께 이를 조직화하려는 학문이 바로 도서관학인 까닭에 도서관학 분야에서는 일찍부터 자료의 조직적 정리와 효율적 검색에 많은 연구와 실험을 해 왔었다. 사실상 초기의 데이터 베이스는 바로 도서관의 문헌목록 검색 시스템에서 시작했다고 할 수 있다.

1. 데이터 베이스의 동향

지금 해외에서는 데이터 베이스가 정보산업의 가장 실용적인 분야 중 하나로서 미국을 중심으로 영국, 프랑스, 캐나다 등에서 발달해 왔다. 이는 세계적인 권위를 갖춘 Cuadra/Elsevier사가 출판한 '90년도 데이터 베이스 디렉토리에 등록된 각 국가별 데이터 베이스 사업자수를 보면 이 사실을 확인할 수 있으며 또한 데이터 베이스 산업은 미국과 유럽에서 시작되었지만 전 세계적으로 관심이 집중되고 있음도 알 수 있다. 그러나 우리와 가까운 대만, 홍콩, 인도네시아, 말레이시아 등에도 등록된 데이터 베이스 업체가 있는데 비해 정보화 사회를 부르는 우리나라에는 등록된 데이터 베이스가 하나도 없다는 사실은 실질적인 우리의 정보산업 수준을 말해주는 것이 아닌가 염려된다.

먼저 미국의 경우를 살펴보자. 부가가치 통신망 사업이 미국에서 시작된 것과 마찬가지로 데이터 베이스 역시 미국이 제작 및 이용분야에서 단연 세계의 선두자리를 확고하게 지키고 있다.

미국의 데이터 베이스 산업은 하루가 다르게 발전하고 있어 현재의 상황을 정확하게 파악하기란 쉽지가 않다. 다만 과거 1차 및 2차 정보에 의존하던 데이터 베이스가 오늘날보다 발전하여 부가가치가 월등히 높은 3차 정보, 즉 예측과 판단이 부가된 정보들을 중심으로 하는 데이터 베이스가 활발히 생산되고 있는 실정이다. 프랑스와 같은 조직적인 데이터 베이스 산업구조가 이루어지지는 않았지만 초기에는 정부차원에서, 그리고 NASA의 지원에 의한 민간차원에서 시작된 데이터 베이스 및 온라인 서비스 사업은 점차 확대되어 민간기업 중심으로 발전하게 되었다. 먼저 미국 정부차원의 정보유통체계를 살펴보자.

미국의 데이터 베이스 산업은 정부 주도로 시작되었다. 오늘날 세계에서 가장 널리 사용되고

있는 MEDLINE이나 AGRIS 등이 바로 초창기 미국 정부에서 만들어진 것이며, 이러한 노력이 미국의 정보산업을 오늘날까지 이끌어 온 원동력이 된 것이다. 이는 정부에서 민간기업을 지원하여 성장한 프랑스의 경우와 좋은 비교가 된다고 하겠다.

오늘날 미국에는 세계적으로 알려진 데이터 베이스업자만도 2000여 업체에 달하며, 이중 소규모의 데이터 베이스를 서비스하는 업자들까지 모두 합하면 1만여 업체가 될 것으로 예상된다. 하지만 보편적으로 데이터 베이스가 황금알을 낳는 거위를 아님을 점차로 확인하고 있는 상황이다. 즉 초기의 많은 데이터 베이스업자들은 지속적인 투자와 관리의 어려움, 그리고 사업성 부진 등을 이유로 거의 도태되었으며, 그들이 만든 데이터 베이스는 몇몇 대규모 업체들에 흡수되어 오늘날 데이터 베이스 업체들은 대기업체 또는 세계적인 언론매체 등의 계열회사로 바뀌어 가고 있다. 그러나 고무적인 점도 있는데 데이터 베이스 회사가 대형화됨에 따라 많은 자본이 투입됨으로써 점차로 고급화되어 가고 있다는 것이다. 그 한 예로 미국의 세계적인 언론, 출판사인 맥그루 힐(McGraw Hill)사가 서비스하는 DRI의 경우, 광범위한 예측정보를 제공함으로써 전 세계의 정치, 경제 관련자들의 지침이 되고 있다. 그러나 여전히 데이터 베이스 자체의 경제적 자립도는 미약한 것으로 인식되어 있다.

우리나라 기업들이 열심히 추종하는 일본의 경우 데이터 베이스 산업은 상대적으로 부진한 실정이다. 미국이나 영국, 프랑스, 캐나다 등 경제적으로 비슷한 국가들에 비해 일본의 데이터 베이스 서비스 산업은 부진한 편이다. 미국이 정부 주도로 민간 참여를 유도한 반면, 일본은 영국형을 상당히 가미했다고 볼 수 있다. 즉, 영국은 미국과는 반대로 정보사업을 철저한

공익사업으로 보고 민간기업의 참여를 봉쇄했다. 또한 데이터 베이스를 국가적인 거대사업으로 보고, 장기계획과 함께 중앙집권 형식의 대표적인 사례로 인식하고 있다.

미국의 형태에 이러한 영국식을 가미한 일본은 일본식 공익성을 강조하며, 정보유통체계를 위한 투자를 해왔지만 사실상 선진국 수준에는 못 미치고 있다. 그러나 자체 데이터 베이스 개발이 미비한 반면, 외국의 데이터 베이스를 통한 정보검색은 세계에서 가장 조직적으로 발달하여 정보의 가공 및 서비스보다 정보입수 및 분석측면에서 매우 발달한 형태를 보이고 있다.

물론 어느 것이 이상적이라고 결론을 내릴 수는 없지만 결과적으로 미니텔을 중심으로 하는 프랑스의 정보통신이 가장 모범적으로 운영되고 있다는 의견이 지배적이고 보면, 정부 주도로 발전한 프랑스 형태를 따르려는 우리나라의 정보화 노력에 어느 정도 공감할 수 있다. 다만 프랑스가 세계적인 부러움을 사고 있는 데도 불구하고 제2의 미니텔이 나오지 않고 있음을 볼 때, 단순한 IP(Information Provider) 지원과 단말기 보급만으로 데이터 베이스가 간단하게 이루어질 수 있으리라는 우리나라 관계자들의 기대는 분명 무리인 것 같다.

이와 같이 국내 데이터 베이스 수준은 아직 걸음마 단계에도 접어들지 못한 상태이다. 다만 정보화 사회 혹은 정보산업 등 핑크빛 꿈에 젖은 일부 업자들의 사전지식도 없는 무분별한 수입으로 해외 데이터 베이스에 대한 이름만 알려져 있을 뿐이다.

세계에서 가장 규모가 큰 다이알로그를 비롯해서 가장 많은 사용자를 확보하고 있는 컴퓨서브, 가장 신뢰도가 높은 정보를 제공하는 DRI 등 웬만한 것들은 거의 수입되어 있다는 사실에 독자 여러분은 놀랄 것이다. 불과 1~2년만에 이렇게 다양한 데이터 베이스가 수입될 수 있었

던 이유를 살펴 보도록 하자.

대부분의 기업간 계약은 물품인도 및 기술이전 등이 사전에 합의되고, 또 상표도입의 경우에는 그 상품의 질과 유사한 품질의 제작능력을 갖추고 있어야 한다. 물론 물건을 수입할 경우에도 시장이 어느 정도 확보되어야 하며, 그 상품에 대한 사후 소비자 관리능력 역시 갖추고 있어야 한다.

하지만 데이터 베이스는 어떤가? 이 점을 많은 국내 도입업자들은 간과하고 있는 것이다.

데이터 베이스란 정보의 창고로서 간단하게 생각하면 가장 손쉬운 수입 상품이다. 즉, 실질적인 물건의 도입도 필요없으며, 수입업자가 갖추어야 할 제작능력이나 공정 등도 일체 필요가 없다. 다만 돈을 갖고 계약만 하면 이미 물건을 판매할 수 있는 것이다.

하지만 이는 크나큰 오판이 아닐 수 없다. 데이터 베이스란 정보를 모은 창고로서 정보를 필요로 하는 불특정 다수가 모두 소비자이다. 그러나 우리에게도 아직 이러한 정보를 돈으로 환산하는 습관이 전혀 되어 있지 않다. 또, 대량의 정보들을 모아둔 데이터 베이스에서 필요한 정보를 찾아내는 작업은 상당한 기술과 경험을 요하는 전문작업인 것이다.

한편 국내 데이터 베이스의 생산실정을 살펴 보도록 하자. 데이터 베이스를 생산하는 업자들 DP(Database Producer) 혹은 IP(Information Provider)라고 한다. 그러나 아쉽게도 아직 국내에서는 데이터 베이스 서비스가 전무한 실정이다. 물론 최근 정보통신진흥협회에서 발간한 국내 데이터 베이스 목록에는 약 200여개의 국내 데이터 베이스를 소개하고 있지만 사실상 대부분의 데이터 베이스가 자사 경영을 위한 내부인사 및 회계 데이터 베이스의 형태를 띤 것으로 사실상 정보로서의 가치가 거의 없는 것들이 대부분이다. 그렇다고 우리나라에는 데이터 베

이스가 없다는 의미는 아니다.

특히 관공서나 국가기관에서 상당히 가치있고 부가가치가 높은 데이터 베이스를 구축하고 있으나, 정보 공유에 대한 의식구조, 그리고 사용자의 요구가 아직은 적은 까닭에 이들 데이터 베이스 서비스가 원활히 제공될 수 있는 여건이 되지 못한다는 것이 가장 큰 문제점이다.

2. 데이터 베이스에서 제공받는 정보들

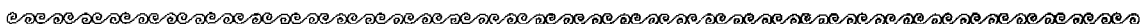
대부분의 시사성 데이터 베이스는 언론사에서 만드는 것으로, 그들은 데이터 베이스를 위해서가 아니라 그들이 발행하는 신문이나 잡지 등을 통해 이미 정보들이 수집, 분석되며 다만 이를 입력하여 컴퓨터에 옮기기만 하면 값진 데이터 베이스가 되는 것이다.

이외에 과학기술정보 역시 오랜 기간의 축적에 의해 상당한 가치를 인정받고 있다. 대부분의 첨단분야가 그렇듯이 데이터 베이스 역시 학교나 학회 등 학술분야에서 시작되었기 때문에 이를 연구하는 학자들의 관심분야가 먼저 데이터 베이스화 되기 시작한 것이다.

이는 오늘날 과학기술정보 중 컴퓨터와 정보통신 관련분야가 가장 많은 비중을 차지하는 사실만으로도 잘 알 수 있다. 또한 과학기술정보는 데이터 베이스화 되기 이전에도 세계의 많은 학자들이 문헌을 통해서 광범위하게 수집 활용하고 있었으며, 많은 기업체에서도 제품 개발을 위한 자료수집에 상당한 투자를 해왔던 분야이다.

그런 까닭에 수집과 검색이 손쉬운 데이터 베이스 형태로의 요구가 가장 컸던 것이 사실이다. 이외에도 각종 수치정보 등 다양한 종류의 데이터 베이스가 세계적으로 서비스되고 있다.

다양한 종류의 데이터 베이스들은 온라인 서비스업자의 호스트 컴퓨터를 통해서 일반 사용자에게 제공된다. 즉 온라인 서비스업자들은 여러 개의 데이터 베이스들을 모아 일반 대중에게



것은 기본적인 사항일 뿐 이 정도로 당장 정보를 검색할 수 있는 것은 아니다.

3. 정보검색방법

정보검색은 일반적으로는, 실사용자가 직접 검색하는 것과 전문 정보검색원을 활용하는 두 가지 방법이 있다.

물론 실사용자가 검색원이 되는 경우도 예상할 수 있지만 여기서 자기검색이란 회사 등의 일반사원들이 정보검색에 대한 비전문적인 지식으로 직접 검색하는 것을 말한다.

자기검색이란 검색을 필요로 하는 일반 개인이 비정기적으로 검색을 실시하는 경우를 말하는데, 이는 오늘날 우리나라 대부분의 기업들이 필요한 정보를 획득하는 방법이기도 하다. 그러나 비전문 검색원이라 하더라도 자기검색을 실시하는데 필요한 기계적 장비는 전문 검색원이 사용하는 장비와 동일하거나 유사한 수준의 것이어야 하며, 또 검색원이 가진 조직적인 자료는 없다 하더라도 당장 필요한 검색을 위한 충분한 자료가 준비되어 있어야 효과적인 검색이 가능하다.

외국의 정보검색 교육을 보면 일본의 경우 정보검색 교육이 다음 세 가지 형태로 이루어지고 있다. 즉, 대학에서의 정규과정, 학회, 협회 및 전문단체, 정보제공업자 및 벤더 등이다.

미국의 형태와 큰 차이가 없으나 미국은 대학 교육을 중심으로 활발하게 실시되고 있는 반면, 일본은 학회 중심의 교육이 활성화되어 있고, 특히 기업체 내부에서의 사원교육으로 활발히 실시되고 있다.

일본 대학의 경우 도서관학 교육은 다분히 사서 양성적 색채가 짙고 교육 역시 4년제 대학보다 2년제 초급대학 중심으로 이루어지고 있다. 때문에 정보검색을 위한 교과과정이 설치되어 있는 학교는 그리 많지 않으며, 교육 또한 그나

마 형식적인 데 그치고 있다. 일본에서 가장 먼저 시작된 곳은 게이오 의숙대학으로서, 현재는 조직적인 교육이 실시되고 있는 대학중 하나이다.

한편 초급대학에서 정보검색 기술자 양성을 적극적으로 시행하고 있는 산업능력 대학에는 정보검색을 필수과목으로 취급하고 있으며, 이외에도 오늘날 이 분야를 교과목에 도입하는 학교의 수는 점점 증가하고 있는 추세이다.

정보제공업자 및 벤더에서 시행되고 있는 교육은 역시 특정 데이터 베이스 검색 테크닉에 한정되고 있는데, 일본 과학기술정보센터(JICST), 일본 경제신문사 등이 자사의 데이터 베이스 검색을 위한 교육을 실시하고 있다. 한편 벤더의 경우 마루젠 MASIS 센터, 키이쿠니아 서점, 유사코, 일본 SDC 등에서 정기교육을 시행하고 있다.

일본의 검색원 교육은 실질적으로 미국과는 큰 차이가 있다. 미국의 검색원은 대체로 도서관학 또는 정보처리학과에서 정규교육을 받은 사람들로 이루어져 있는 반면 일본의 경우 다양한 전공분야를 공부한 사람들이 취업후 단기교육으로 기본적인 검색원의 자질을 익히고, 이 이상은 소속회사의 내부교육으로 특정분야의 전문성을 갖춘 검색원이 되는 경우가 대부분이다.

우리나라의 경우 아직 검색원 교육에 대해서는 언급할 단계가 못되는 실정이지만 최근 그 중요성을 인식하고 있는 일부 대학의 도서관학에서 이러한 과목을 교과과정에 추가하고 있으며, 미국과 같은 별도의 학과를 개설할 움직임을 보이고 있는 학교도 있는 것으로 알려지고 있다.

국내에서는 최근 에임스 컨설팅에서 컴퓨터 입문에서부터 PC 통신과 그 활용 데이터 베이스(DB) 검색법을 가르치고 있다.