

데이터베이스 당면과제와 대책

박홍식
산업기술정보원 원장

지금 세계는 산업 사회에서 정보화 사회로 급속히 이행되고 있다. 정보화 사회에서는 정보 통신과 데이터베이스가 차지하는 역할이 매우 큰데, 그 중에서도 정보의 효율적인 수집과 활용이 경제활동의 원동력이 될 뿐만 아니라 인간의 생활 환경에 큰 변화를 주게 된다. 이처럼 데이터베이스가 정보화 사회에서 담당하는 역할이 매우 중요하다는 것은 아무리 강조해도 지나침이 없을 것이다.

전 세계적으로 데이터베이스산업의 규모를 보면 1975년부터 1995년의 20년간 데이터베이스의 수는 301개에서 9,207개로 31배나 증가하였고, 레코드의 수도 5,200만개에서 81억 6,000만개로 157배나 증가하고 있다. 한편 우리 나라에서 현재 제작되고 있는 데이터베이스의 수는 1,136개로서, 용도별로 보면 상용 데이터베이스가 621개, 인하우스 데이터베이스가 515개로 1991년도의 424개에서 2.6배 증가하고 있다. 그러나 양적으로는 어느 정도 증가되고 있으나 질적인 면에서 많은 문제점도 드러나고 있다. 게다가 최근에는 데이터베이스의 세계에서 일고 있는 새로운 양상에 대처하지 않으면 안되게 되었다. 즉 인터넷의 데이터베이스화, 행정의 정보화에 의한 디지털 정보 유통의 촉진에 의한 정보의 무료화 등 상용 데이터베이스를 둘러싼 환경은 점점 어려워지고 있다.

지금부터 10년은 과거 10년간 일어난 변화의 속도 보다 훨씬 빠른 속도로 변화할 것이 예상된다. 지금까지 당연하다고 생각하던 것이 전혀 다

른 양상으로 나타날 수 있다는 것을 데이터베이스 종사자들은 유념해야 할 것이다. 여기서는 1997년 벽두를 맞이하여 데이터베이스 산업의 발전을 위해 우리나라 데이터베이스 업계가 주시해야 할 몇가지 현안문제에 대하여 살펴본다.

인터넷 데이터베이스의 확대

최근 인터넷이 급속히 보급됨에 따라 국제 데이터베이스 시장에서 전자 정보의 유통시스템이 혁명적으로 변모됨으로써 정보의 생산에서 엔드 유저의 이용에 이르기까지 정보 환경의 패러다임이 변화되고 있다. 20여년의 역사를 가진 미국의 기존 상용 온라인 서비스 산업이 생존에 위협을 받을 정도로 변혁이 임박해 있다.

현재 많은 기업들이 이 분야에 집중하여 세계적으로 매수, 합병, 제휴가 이루어지고 있으며, 인터넷과 World Wide Web에 참가하려는 전략을 세우고 있다. 인터넷의 등장이 가져온 데이터베이스산업의 변혁 양상을 보면 제일 큰 것으로는 정보 요금의 가격파괴로서 인터넷상에서는 대부분의 정보가 무료로 제공되고 있다. 예를 들면 기업의 개요를 알고싶을 때 기업이 제공하는 홈페이지를 보면 대략의 내용은 알 수 있다. 신문사나 TV방송국의 뉴스, 국가의 통계와 같은 1차적인 정보도 데이터베이스업자의 손을 거치지않고도 입수할 수 있는 시대로 변하고 있다. 무료 정보는 신뢰할 수 없는 것이 많다고 하는 것이 사실이지만 그래도 어느 정도 수요를 충족시키고 있는것 또한 사실이다.

최근에는 점점 많은 사람들이 정보 수집 수단 하나로서 인터넷을 활용하고 있다. 그 결과 정보 제공 분야에서 데이터베이스업자가 활약할 수 있는 장이 좁아지고 있다. 따라서 인터넷 시대에 데이터베이스업이 존립하기 위한 대응책 마련이 시급히 요구되고 있다.

행정 정보의 무료 공개

행정의 정보화가 추진됨으로써 민간 데이터베이스업계에는 중대한 문제가 발생하고 있다. 행정 측면에서 전자화된 정보가 계속적으로 제공된다. 여기에 포함되는 것은 관보나 백서 등의 인쇄물, 각종 위원회에서 사용하는 보고서나 자료 외에 새로운 법률, 국회에서의 수상의 연설문, 심의 내용 등 여러 가지가 있다. 각종 통계도 물론 포함된다. 이들 정보는 원칙적으로 무료로 제공된다. 중대한 문제는 바로 여기에 있다. 현재 통계 정보나 유가 증권 보고서 등의 정보는 데이터베이스업자가 유료로 제공하고 있는데, 이것이 행정 서비스로서 무료로 입수할 수 있게 되면, 누구도 민간 데이터베이스를 이용하지 않을 것이다. 이러한 예는 이미 미국에서 일어나고 있다. SEC(Securities and Exchanges Commission)의 EDGAR(Electronic Data Gathering Analysis and Retrieval)와 행정 서비스로서 정보를 제공하고 있는 FedWorld 2가지가 바로 그것이다.

EDGAR는 SEC가 미국의 모든 상장기업에 대하여 재무, 기업 정보를 전자적으로 제출하는 것을 의무화함에 따라 생긴 컴퓨터에 의한 정보 수집, 해석, 검색 서비스이다. SEC는 1995년 가을부터 인터넷상에서 무료로 이들 기업 정보를 제공하게 되었다.

단, 같은 정보를 서비스하고 있는 민간 벤더에의 영향을 고려하여 제공되는 정보는 각 기업이 보고한 그대로의 상태로 색인도 없고 정리도 되지않은 상태로 제공된다. 또한 보고시점에서 24시간 지연된 상태에서 공개되고 있다. 한편 민간 벤더는 타사 데이터와 비교가능한 형태나 보기 쉽게 정리된 형태로 정보를 가공하여 유료로 제공함으로써 상용 벤더로서의 시장을 유지해 나가고 있다.



미국의 정부 기관이 인터넷상에 갖고있는 사이트의 수는 700개 가까이 되며, 또한 매주 새롭게 사이트가 생기고 있다. 이러한 모든 정부 기관을 커버하고 있는 것이 FedWorld이다. 상무성의 산하기관인 NTIS(National Technical Information Service)가 1992년 11월부터 FedWorld로 인터넷을 통해 정부 정보의 온라인 제공을 개시하고 있다. 여기에서는 정부의 대부분의 정보를 입수할 수 있다. FedWorld에 들어가서 U.S. Government Information Service를 클릭하면 40항목으로 나누어진 Subject Categories의 화면이 나타난다.

여기에서 희망하는 정보가 포함되어 있는 카테고리 클릭하면 사이트가 출현한다. 여기에 예시된 Administration and Management에서는 FEMA나 CIA 등 46개의 사이트가 가능하다.

이와같이 EDGAR나 FedWorld를 통하여 정보를 무료로 입수할 수 있게 되면, 정부 정보를 데이터베이스화 하여 유료로 제공하고 있는 민간업자는 생존의 위기를 느낄 수밖에 없다. 미국의 행정 정보화는 상당히 진전되어 있다. 우리나라가 이러한 사태를 맞이하기까지는 아직 시간이 걸릴것으로 예상되지만 우리나라도 조만간 행정 정보화의 진전에 의해 이와 같은 문제가 발생할 것이 틀림없다.

양질의 데이터베이스 구축만이 살길

인터넷의 대규모 데이터베이스화, 디지털화한 행정 정보의 유통 등 상용 데이터베이스의 존립 기반을 흔드는 사태에 대처하기 위해서 데이터베이스 업자들은 1차적인 정보를 정리, 분석하는 등 부가가치를 가진 상품을 만들지 않으면 안될 것이다.

단순히 정보를 유통시키는 데이터베이스만으

로는 경쟁력을 가질 수 없는 시대에 돌입하고 있다. 인터넷에서 무료로 제공되고 있는 정보, 행정이 제공하는 정보는 정보 서비스업에 있어서는 원재료와 같은 것이다. 원재료를 어떻게 가공하여 제품으로 처리하는가 하는것이 정보 벤더, 상용 데이터베이스 업자에 주어진 사명이다. 앞으로 10년간은 이러한 부가 가치 부분에서 경쟁하는 시대가 될 것이다. 이용자의 입장에서는 사용하기 쉬운 데이터베이스를 요구하고 있다. 이를 위해 첫번째로 생각할 수 있는 것은 자연어에 의한 검색이다. A사의 데이터베이스의 검색에 익숙한 사람이 B사의 데이터베이스를 검색하려고 할 때 시간이 더 걸린다던가 검색 자체가 불가능한 경우도 있다.

이것은 A사와 B사간에 검색 명령어가 틀리거나 키워드 부여방법이 다르기 때문이다. 이용자는 양방의 데이터베이스를 사용하려면 A, B 양사의 검색 방법을 모두 숙지해야 하기 때문에 큰 부담이 된다. 이를 피하기 위해서는 A, B 양사의 검색 방법을 같게 하는 것을 생각할 수 있지만 그다지 현실적이지 못하다. 그보다도 소프트웨어를 개발하여 A사의 검색 방법이 B사의 검색 방법으로 자동변환 되게끔 하면 이용자는 1가지 검색 방법만 알면 된다.

또한 일반적으로 사용하는 일상어로 검색할 수 있게끔 하면 복잡한 명령어 등을 생각할 필요도 없어진다. 미국의 데이터베이스는 이미 영어의 질문문으로 검색할 수 있는 기술이 등장하고 있다. 두번째로 생각할 수 있는 것은 복수의 다른 분야의 데이터베이스를 마치 하나의 데이터베이스와 같이 사용하는 방법이다. 이것은 현재 인터넷에서 브라우저를 사용하여 정보를 찾아내는 것과 비슷하다. 보통 정보를 찾을 때에는 하나의 소스가 아니라 몇 개의 소스로부터 다면적으로 찾는 경우가 많다.

예컨데 기업의 정보를 찾을 때 신문기사, 재

무 정보, 경영자의 인물 정보, 업계 정보 등 여러 가지 정보를 수집한다. 이 때 각각의 데이터베이스에서 하나 하나 조사하는 것은 익숙치 않은 사람에게는 대단한 작업이 된다. 이것을 어느 하나의 데이터베이스에 들어가면 그 다음은 필요에 따라 계속적으로 조사할 수 있는 기능이 있다면 편리할 것이다. 복수의 데이터베이스를 종합적으로 제공하는 기능이라 할 수 있다. 세 번째는 정보의 비주얼화 즉 이미지로서 정보를 제공하는 방법이다.

예를 들면 지도의 활용 등 정보를 시각화하여 제공하는 것이다. 그림이나 도표로 표현된 정보의 장점은 한눈에 전체의 상황을 잘 알 수 있는데 있다. 예를 들면 경제적인 지역의 통계 데이터를 지도상에 플롯하거나 3차원적으로 표시하거나 복수의 정보를 중첩하여 보기쉬운 형태로 표시하는 것이다.

끝으로 이용자는 항상 데이터베이스의 이용요금이 비싸다는 불만을 첫째로 들고 있다. 이용자의 입장에서는 이용요금이 높기 때문에 이용이 제한되고, 제공자 입장에서 보면 이용량이 늘어나지 않아 매출액이 감소된다. 그렇다고 이용료를 내리는 것도 쉬운 일은 아니다. 이용자도 제공자도 납득할 수 있는 적절한 정보 요금을 결정하기란 어려운 일이다. 같은 정보라도 받는 측에 따라 그 가치가 다른 것이 정보라고 하는 상품이 가진 특이한 성격이다. 데이터베이스의 위력은 과거부터 현재까지의 방대한 데이터로부터 필요한 데이터를 망라적으로 조사하여 순간적으로 찾아내는 것이다. 사람이 직접 수행함으로써 소요되는 시간을 대폭 단축해 주는 것이 데이터베이스가 가진 효용중의 하나이다. 제공자의 입장에서 보면 시간은 돈으로, 데이터베이스의 편리성을 크게 평가하고 싶고, 이러한 의미에서 보면 데이터베이스의 이용료는 그다지 높지않다고 주장하고 싶을 것이다. 양자의 불만을 해결하는 하

나의 방법은 이용자를 늘리는 방법이다. 이용자가 증가하면 반대로 요금을 인하해도 문제는 없다.

여기에서 이용자를 늘리기 위해서는 무엇보다도 사용하기 쉬운 데이터베이스를 많이 제공하는 것일 것이다. 그러나 데이터베이스 이용자를 증가시키기 위해서는 제공자의 노력만으로는 한계가 있다. 데이터베이스 이용의 활성화를 위한 이용자 교육의 역할이 중요하다. 사회 전체의 정보마인드를 높이기 위해 정보화 교육에 지금까지 투입한 힘의 몇배의 힘을 투입하지 않으면 안된다.

제도적 측면

또한 정보를 누구나 자유롭게 사용하고 유통할 수 있게끔 하기 위해 지적소유권 문제를 해결하지 않으면 안된다. 프라이버시의 보호, 시큐리티의 확보에도 충분한 배려가 있어야 할 것이다. 또한 정보의 유통을 저해하는 규제의 철폐는 물론이고, 정보 공개의 원칙을 결정하는 정보 공개법도 제정하지 않으면 안된다. 데이터베이스의 이용, 제공을 촉진하기 위해서는 부가 가치를 가진 사용하기 쉬운 데이터베이스를 제공함과 동시에 이들 과제의 해결을 위해 사회 전체의 노력도 요구되고 있다. **DC**