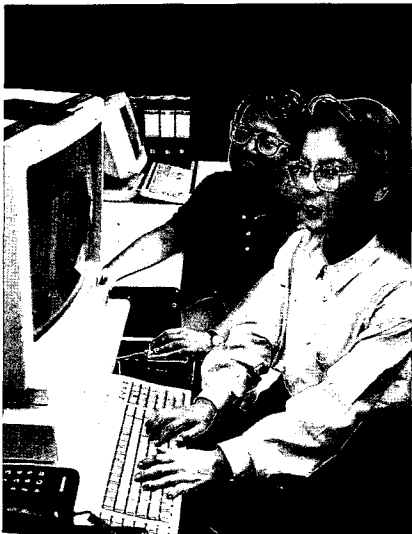


데이터베이스 기술(I)

정리 : 최중태

목 차

1. 개 요
2. DB제작과 생산기술
3. DB유통기술
4. DB이용기술
5. DB검색기술
6. DB시스템기술
7. 결 언



1. 개 요

현대는 정보화 사회라 할 수 있다. 즉 모든 산업활동이나 개인의 생활에서 정보가 중요한 역할을 하는 사회로 되어가고 있다. 정보의 홍수라고 말할 수 있는 정보화 시대를 맞이하여 대용량의 데이터로부터 신속하고 효율적으로 사용자가 원하는 정보를 찾는다는 것은 매우 중요한 문제가 되었다. 선진국에서는 데이터베이스산

업이 널리 확산되어 그 시장도 여타 산업에 비해 보다 큰 비율로 성장하고 있다. 그러나 우리나라의 경우 아직까지는 데이터베이스에 대한 중요성이 기업이나 연구소 같은 특정인에게만 인식되고 있고 일반인에게 까지는 그 인식이 확산되지 못하고 있다. 데이터베이스 초기 단계라 볼 수 있는 우리나라에서도 정부 및 업계에서 데이터베이스의 중요성을 인식하고 많은 노력을 하고 있으나 정보서비스 시장의 국제화가 가속화 되는 시점에서 확고한 예비책이 없다면 해외 의존도가 높은 우리나라의 데이터베이스 산업은 설 땅을 잃을 것이다.

데이터베이스는 수집된 자료를 소정의 형태로 가공, 처리하여 제작되며 제작된 데이터베이스는 온라인(ON-LINE) 또는 오프라인(OFF-LINE) 형태로 이용자에게 제공된다. 다시말해 데이터베이스 산업에는 데이터베이스의 제작과 운영, 유통, 검색에 관련된 기술이 요구 된다. 이러한 과정속에는 여러가지 기술이 관련되어 있는데 정보의 원활한 유통을 위한 표준화와 데이터베이스의 제작, 이용, 유통분야의 기초기술 개발이 필요하며 데이터베이스 조작, 운영 기술 뿐만 아니라 정보수집, 가공, 유통서비스 등 제기술의 발전이 필요하다. 현재 우리나라는 이러한 DB관련기술 분야가 매우 취약한 상태에 있다.

또한 데이터베이스는 기초 기술 뿐만 아니라 첨단기술과 깊이 관련되어 있다. 따라서 데이터베이스 산업의 지원은 데이터베이스 자체만이 아니라 이와 관련된 기술 및 산업에 대해서도 함께 이루어져야 온전한 성과가 이루어질 것이다. 이러한 성과를 이루기 위해 데이터베이스 산업에는 많은 인력을 필요로 한다. 정보통신관련 인력 뿐만 아니라 데이터베이스 설계 및 조작 운영에 관

한 인력과 이들 전체 조직을 효과적으로 관리할 수 있는 인력도 똑같이 필요하다. 그러나 현재까지 국내에서는 정보통신쪽의 전문가 육성에만 치우쳤다고 해도 과언이 아닐 것이다. 앞으로는 보다 균형적으로 데이터베이스 관련 전문가를 양성하고 숫자적으로도 충분한 인력을 확보하는데 관심을 기울여야 할 것이다.

최근 산업기술정보원을 비롯한 여러기관에서 산업체 및 일반인을 대상으로 정보검색 또는 사내 정보화에 대한 교육을 활성화하고 있는 것도 전문인력 확보 및 데이터베이스 저변 확대를 위해서 매우 중요한 기여를 할 것으로 보인다. 또

한 많은 국내 대학에서도 도서관학과를 문헌정보학과로 이름을 바꾸고 학과내용도 바꾸고 있는 것은 이러한 측면에서 매우 바람직한 현상이다.

따라서 이 장에서는 DB산업의 균형있는 발전을 위하여 개발해야 할 여러 DB관련 기술들을 소개하고 이들 기술과 관련하여 국내외에서 연구되고 개발된 현황을 소개한다. 나아가는 국내 DB산업활성화를 위해 이들 기술을 확보할 방안을 제시하고자 한다. 이러한 기술의 발전을 통해 DB제작 및 유통을 원활히 할 수 있으며 사용자들은 유익한 정보를 편리하고 신속하게 검색할 수 있게 되어 국내의 DB산업을 육성하는 기반을

구 분	기 술 명	비 고
DB제작, 생산기술	<ul style="list-style-type: none"> DB분류법 DB목록 DB자료 부화 기법 DB입출력 기술 정보 축적 및 관리기법 DB설계기법 및 제작도구 개발 DB자원관리 	DB내용, DB제작, DB이용 문자인식, 화상정보, 음성정보 전자사전, 시소러스 PSL/PSA, SUDDS IRDS
DB유통기술	<ul style="list-style-type: none"> 정보통신망 이용기술 비디오텍스 서비스 CD-ROM기술 기타정보 전달기술 	VAN, WAN, LAN 저작도구, 에뮬레이터 RDA, DTP
DB이용기술	<ul style="list-style-type: none"> 맨머신 인터페이스 지능형 단말기 자동번역 자연어 처리 음성인식 및 합성 	소형화, 휴대화 다언어 시소러스
DB검색기술	<ul style="list-style-type: none"> 검색기법 인식기법 색인기법 검색명령어 	지식기반, 문헌검색 자연어(한글), 비정형 데이터 통일어 색인, 자연어 색인 자연어 질의 처리기
DB시스템기술	<ul style="list-style-type: none"> 분산DB 객체지향DB 멀티미디어DB 하이퍼텍스트DB DB머신 데이터 및 저장 관리자 한글정보검색 지원 DBMS 	관계형, 객체지향형 등 하이퍼미디어 저장 시스템

표 1.1 데이터베이스 관련기술

조성하는데 결정적인 역할을 하게 될 것이다.

기술들의 연관관계

데이터베이스 활성화를 위해서는 여러 기술들이 종합적으로 이용되어야 하며 이러한 기술들이 균형있게 개발되고 서로 통합적으로 이용되어야 한다. 이러한 데이터베이스 관련 기술들은 대략 다음과 같은 5가지 분야로 구분할 수 있다. 이들 분야에 대한 세부 기술 요소들은 표 1.1에 요약되어 있다.

- 1) 데이터베이스 제작 및 생산기술 : 원천정보로부터 분류하고 편집하여 유용한 데이터베이스 형태를 제작하고 이를 컴퓨터나 기타 정보서비스 매체에 수록하는 저장 기술을 말한다.
- 2) 데이터베이스 유통기술 : 제작된 데이터베이스를 사용자들에게 서비스를 제공하거나 그 내용을 전달/유통하는 방법과 관련된 기술분야이다.
- 3) 데이터베이스 이용기술 : 사용자들이 데이터베이스를 이용할 때 편리하고 효과적으로 이용할 수 있는 방법을 제공하는 기술이다.
- 4) 데이터베이스 검색기술 : 사용자들이 데이터베이스에 대한 내용을 요청할 때 이를 효율적이고 편리하게 검색할 수 있도록 제공해주는 검색에 관련된 기술이다.
- 5) 데이터베이스 시스템 기술 : 데이터베이스 내용을 컴퓨터에서 효율적이고 편리하게 관리해주는 소프트웨어 시스템과 관련된 기술 분야이다.

2. DB제작과 생산기술

DB분류법

도서관에서 소장된 자료를 열람하기 위해서는 분류된 도서 목록을 참조하게 된다. 이러한 도서 분류와 마찬가지로 데이터베이스에 대해서도 일정한 분류법이 개발되어 사용자 뿐만 아니라 개발자들이 이를 참조하는 것은 매우 중요하다. 현실정에서는 IP(INFORMATION PROVIDER)나 개

발자 마다 분류법이 다르므로 이를 포괄하는 체계적인 분류법이 속히 개발되어야 한다.

데이터베이스 분류법은 데이터베이스의 소유자 또는 제공업체, 내용, 장소(지역), 작성일자, 크기, 이용자 수 등 여러가지 사항을 고려해야 한다. 이러한 분류는 데이터베이스 구축 이전에 먼저 실시 되어야 하는 과제로서 DBMS, IP, 사용자는 물론 데이터베이스 산업이 필수적으로 준수해야 하는 기본분류 방법이 되어야 한다.

데이터베이스 산업의 진흥정책은 무엇보다도 데이터베이스의 체계적 분류를 바탕으로 집중적인 투자가 이루어져야 그 효율성을 기대할 수 있으며 더 나아가 국제교류 및 국제 경쟁력의 강화도 가능하다. 따라서 데이터베이스 표준 분류안과 관련된 연구의 기본방향은 첫째, 데이터베이스 산업 분야를 체계적으로 분류하여 잠재적 시장 기반을 활성화시키는 정책수행의 기본 자료로 삼을 수 있도록 하는 것이며 둘째, 관련기관에서는 데이터베이스의 분류 및 용어상의 혼돈에 따른 오류와 업무상의 중복을 피할 수 있게 하기 위한 것이다.

데이터베이스는 DB내용측면, DB제작측면, DB이용측면으로 분류할 수 있다. 이것을 세분하면 표1.2와 같이 9개 항목으로 구성할 수 있다.

DB내용측면	<ul style="list-style-type: none"> • 제작기관별 • 주제(분야)별
DB제작측면	<ul style="list-style-type: none"> • 가공형태별 • 표현형태별 • 언어별 • 가공완성도 및 갱신주기별
DB이용측면	<ul style="list-style-type: none"> • 검색방식별 • 제공매체별 • 용도별

표 1.2 데이터베이스 분류

DB목록

데이터베이스에 대한 분류가 이루어지면 데이

대분류	중분류	소분류
일반	<도서·논문>	단행본 / 논문 잡지 / 목 사전 / 목 록
	<인물·인력>	일반 기 과 학 기 술
	<신문·뉴스>	일 반 경 제, 산 업, 기 술 * 국 내 * 국 외 기 타
생활	<보 건>	공 연 / 전 시 비 디 오 / TV 스 포 츠 / 바 둑 / 경 마
	<가 정·이 사>	
	<민 원 안 내>	
	<취 업>	
	<기 상>	
	<레 저>	
	<부 동 산>	관 광 안 내 예 약 교 통
<은 행·보 험>		
	<주 민·배 달>	
	<자 동 차>	
	<기 타>	
교육	<진 학>	
	<학 습>	
	<어 학>	
일반법률	<일반 법률>	

터베이스 목록을 구축하여 이를 관리하는 것이 필요하다. 데이터베이스 목록은 기존의 정보 서비스를 제공하는 하이텔, 천리안 등에도 잘 정리가 되어 있으며 한국데이터베이스진흥센터(DPC)에서도 93년 하반기에 DB표준화분과위원회 내의 DB표준분류실무작업반에서 표준목록을 위한 기초 연구가 있었다.

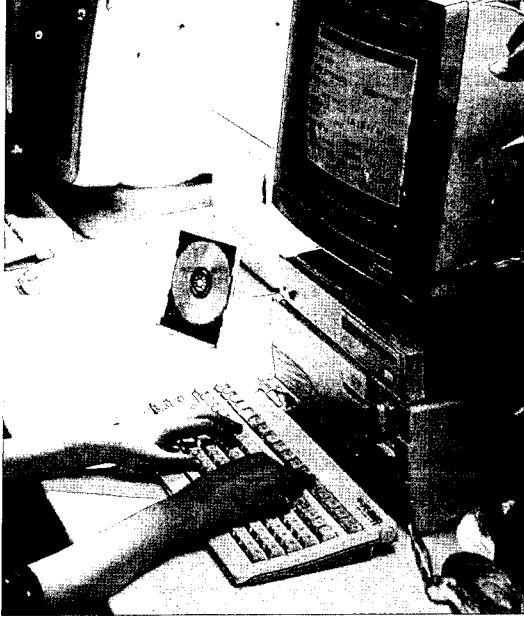
이러한 목록이 표준화된다면 데이터베이스 제작업체의 중복 투자를 막을 수 있고 이용자들에게 사용 가능한 데이터베이스의 소재를 한 눈에 알아볼 수 있도록 하여 데이터베이스 이용을 촉진할 수 있다. 예로 DPC에서 발간한 알기쉬운 한

국의 데이터베이스 편람 1994에 실려 있는 주제(분야)별 데이터베이스 목록은 표 1.3에 있다.

인문·사회 과학 전반	<인문·사회과학 전반>
과학기술	<전 반> <물리·화학> <전자·통신·컴퓨터> <생명·식품> <해양·선박> <기계·금속> <원 자 력> <기 타>
의학	<의 학>
규 격 재산권	<규 격> <저 작 권> <특 허> <상 표>
	<경 제> ┌ 일반(동향/정책) ├ 통 계 ├ 물 가 <산 업> ┌ 일반(동향/정책) ├ 시 장 ├ 통 계 <무 역> ┌ 법 규 / 관 세 ├ 가 격 / 거 래 └ 업 체 / 해 외 지 사
	<금 융> ┌ 일반(동향/정책) ├ 자 금 ├ 외 환 / 환 율 ├ 채 권 <증 권> ┌ 신 용 ├ 일 반 / 투 자 └ 시 황 / 동 향
	<기 업> ┌ 기 업 동 정 ├ 기 업 정 보 ├ * 일 반 프 로 필 ├ * 재 무 보 └ 흥 보 <입찰·경매> <시장·가격> <세무·노무>

표 1.3 주제(분야)별 데이터베이스 목록

<자료: '94 알기쉬운 한국데이터베이스 편람>



DB입출력 기술

● 문자인식

문자인식을 위하여 문자를 입력하는 방법으로는 크게 온라인 입력방법과 오프라인 입력방법으로 구분된다. 오프라인 문서 입력시스템에 필요한 영상 입력장치로는 스캐너, 카메라, 팩시밀리 등이 사용될 수 있으며 온라인 인식시스템에 필요한 펜 입력장치는 전자펜, 마우스, 타블렛(TABLET), 터치스크린 등이 사용될 수 있다. 이러한 영상 입력장치나 펜 입력장치에 의해서 입력된 문자 정보는 컴퓨터내에서 구동되는 문자 인식시스템에 넘겨지며 문자인식시스템에서 인식되어 모든 문자들이 코드 형태로 변환된다. 이렇게 코드 형태로 변환된 문자들은 컴퓨터에서 구동되는 문서 편집기에 읽어 들여 원하는 형태로 삽입, 삭제, 재구성 등 편집할 수 있게 되며 기존의 문서편집기에서 작성한 문서들과 같이 별도의 화일 형태로 보관할 수 있을 뿐만 아니라 프린터로 출력할 수도 있다.

이런 문자 인식기술은 방대한 원시정보를 사람이 직접 입력하지 않고도 편리하게 입력할 수 있어 방대한 DB구축시 필수적으로 사용될 수 있을 것이다. 그러한 현재의 문자 인식기술로는 인식이 낮아 실용화되기 어려운 실정에 있다. 문자

의 인식이 99.99%이상이 될때 실용화가 가능할 것이다. 따라서 문자 인식을 높이는 인식기술이 개발되어야 하며 이러한 인식과 관련된 여러 주변 장치들도 개발되어야 할 것이다.

● 화상정보

문자, 도형, 화상데이터를 자동적으로 입력하는 장치에는 비교적 값이 저렴하고 가장 널리 사용되는 ITV카메라, ITV카메라 보다 높은 해상도로 입력하는 FSS(FLYING SPOT SCANNER), 기계식 주사로 화상을 입력하는 원통 주사기, 복사기와 같이 원고를 그대로 취급하는 평면 주사방식의 입력장치인 화상주사기(IMAGE SCANNER), 화상을 전기신호로 변환하는 소자로서 소형화, 고신뢰화, 저소비 전력화가 실현되는 전하결합장치(CCD : CHARGE COUPLED DEVICE) 등이 있다.

이러한 입력 장치를 통하여 들어오는 데이터의 형식은 다양하며 그 양 또한 방대하기 때문에 이러한 데이터에 대한 통일된 코드방식과 압축 및 복원기술이 개발되어야 한다.

● 음성정보

음성데이터는 전화, 마이크로폰, 종합정보통신망(ISDN)등을 통하여 입력될 수 있으며 스피커나 데이터화일의 형태로 출력시킬 수 있다. 화상 데이터보다 작은 양이지만 문자보다는 많은 양의 저장공간을 차지하므로 음성데이터에 대해서도 압축과 복원이 필요하다.

또한 음성과 화면이 동시에 출력될 경우에는 각 정보에 대한 동기화가 필요하다. 동기화란 화면의 출력속도와 음성의 출력속도를 적절히 조정하여 화면과 음성이 일치하게 하는 것이다. 음성의 저장 형태와 화면의 저장 형태가 다르고 각 정보의 복원에 대한 알고리즘을 다르게 사용할 경우에는 종종 화면 정보를 설명하는 음성정보의 출력속도가 맞지 않아 정상적인 정보를 제공할

수 없다.

정보 축적 및 관리기법

데이터의 축적에 있어서 변화하는 정보와 이용자의 요구에 대하여 데이터의 통일성과 일관성을 유지하는 것이 중요하다. 색인과 검색의 성공적인 모델은 사용하는 정보에 대한 정확한 구조와 사용기술이 높은 실용성과 효율성을 가졌을 때 가능하다. 정보의 구조는 대량으로 구축될 수 있는 것이어야 하고 실시간 처리가 요구되는 경우에 빠른 계산방법이 만들어져야 한다.

● 전자사전

대규모 문자자료의 신속하고 정확한 입력, 입력된 자료 내용을 일목요연하게 알 수 있게 하는 도구의 개발(요약색인), 대규모 자료의 저장, 저장된 자료의 검색이라는 일련의 과정을 전문가의 도움없이 효율적으로 수행하기 위한 시스템에는 전자사전이 필요하다.

전자사전은 정보통신서비스 하위영역 기초어휘에서 시작하여 모든 분야를 나타내야 한다. 어떤 질의가 들어오면 전자사전에 의해 동의어, 반의어 등의 다양한 서비스 제공 받아 그 질의를 올바르게 인식하고 처리할 수 있어야 한다.

● 시소러스

시소러스는 일종의 지식베이스로서 개념을 표현하는 노드와 개념들 사이의 관계를 나타내는 링크로 구성된다. 용어들의 동의어, 반의어 뿐만 아니라 축약어, 외국어, 협의어, 광의어를 표현한다. 용어간의 상하위 관계 또는 동의어 관계(SYNONYM)등을 나타내고 있으며 이들과 같이 밀접한 관계는 아닐지라도 관련이 있는 용어를 나타내는 관련어 관계 등으로 구성된다.

시소러스를 기반으로 하는 정보검색 시스템의 장점은 첫째, 문서의 색인을 위한 색인어가 시소

러스로 부터 선택되기 때문에 문서에서 사용된 특정용어에 관계없이 같은 내용을 갖는 문서는 같은 색인어에 의해 검색될 수 있다는 것이다. 둘째, 시소러스의 링크는 단어들 사이의 연관성을 정확하게 표현하며 단어들 사이의 연관성 정보는 정보검색 시스템의 검색효율 향상에 사용될 수 있다는 것이다.

DB설계 기법 및 제작도구 개발

● 개요

데이터베이스 설계는 관련된 응용분야에 대한 데이터베이스의 스키마를 개발하는 과정이다. 응용 프로그램의 수행 속도가 데이터베이스의 구조와 액세스 경로의 효율성에 의존하기 때문에 데이터베이스 설계는 아주 중요하다. 데이터베이스 응용분야가 급속히 증가하고 크기와 복잡도가 한층 커짐에 따라 데이터베이스 설계 과정 그 자체는 여러단계들로 구성되며 각 단계에는 서로 다른 분석 및 결정 작업이 포함된다. 설계 및 설계 결정 사항들은 문서나 다이어그램 형태로 정확히 표현된다. 흔히 설계자들은 널리 인정받기에 충분한 논리적 단계들을 규정하는 설계 방법론들을 택하려 한다. 그러한 방법론은 일반적으로 현실 세계를 분석하는 것에서 부터 모형화하고 특정 DBMS에 적합한 데이터베이스 스키마를 만드는 데까지 두루 적용된다.

● 데이터베이스 설계단계

데이터베이스 설계의 목표는 여러가지가 있다. 특정사용자와 응용의 정보 내용 요구를 만족시키는 것, 자연스럽게 쉽게 이해할 수 있는 정보구조를 제공 하는것,그리고 처리의 요구 조건과 응답시간, 처리시간, 저장공간 등과 같은 성능 목적을 지원하는 것 등이다.

데이터베이스 설계는 요구조건과 구현을 포함해서 그림1과 같이 5단계로 나눌 수 있다. 데이터

베이스 설계는 통상적으로 두개의 활동이 병행해서 진행된다. 하나는 데이터베이스의 내용과 구조의 설계에 관한 활동이고 또 다른 하나는 데이터의 처리와 응용소프트웨어 설계에 관한 활동이다. 이 두 활동은 서로 밀접하게 관련되어 있다. 그렇다고 데이터베이스 설계가 반드시 그림 1과 같이 진행되는 것은 아니다.

그러나 컴퓨터 기술의 발달로 위와 같은 각 단계를 컴퓨터를 사용하여 자동화하는 도구가 절실히 요구되고 있으며 앞으로 방대한 데이터베이스를 구축하기 위해서는 이러한 데이터베이스 설계 전단계를 통합하여 자동화하는 도구가 필요하다. 특히 데이터베이스 생산업자에게는 데이터베이스 설계를 위한 통합제작도구를 개발하는 것이 필수불가결하다.

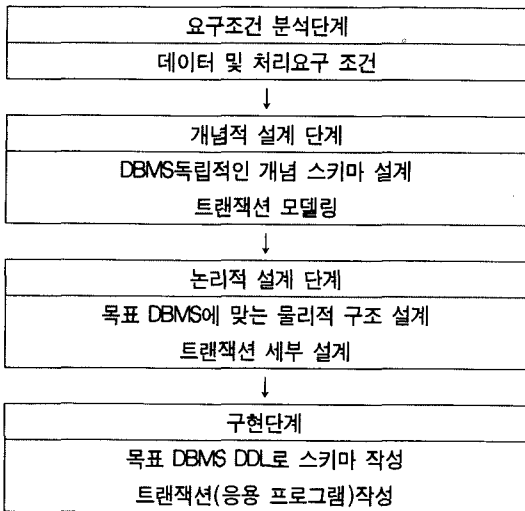


그림 1 데이터베이스 설계 단계

● 외국현황

데이터베이스 설계의 첫단계라고 할 수 있는 요구조건 분석을 자동화하는 도구로 PSL/PSA (PROBLEM STATEMENT LANGUAGE/PROBLEM STATEMENT ANALYZER)가 있으며, 논리적 모델링과 물리적 모델링을 지원하고 논리적 모델을 정규화 시키는 등의 기능을 지원

하는 SUDDS(SEATTLE UNIVERSITY DATA-BASE DEVELOPMENT SYSTEM)가 있다.

데이터베이스 자원 관리

데이터베이스 구축하는 영역이 다양해지고 내용이 점점 방대해지고 있기 때문에 데이터베이스를 체계적으로 관리하지 않으면 데이터베이스들 간에 비슷한 내용이 중복되어 노력과 자원을 낭비할 수 있다. 또한 데이터베이스를 제공하는 업체가 자신들이 제작할 데이터베이스의 원시정보를 모두 직접 입수하고 편집, 저작할 경우 많은 시간과 노력이 들게 된다. 따라서 데이터베이스 제작에 이용될 수 있는 가장 기본적인 정보들을 미리 개발하여 제공한다면 데이터베이스 제작업체들은 이 기반정보를 이용하여 자신의 데이터베이스를 더욱 풍부하고 편리하게 제작할 수 있을 것이다. 이를 위해 전자지도, 전자사전, 정부의 기본통계 등에 대한 정보를 종합적으로 구축하여 데이터베이스 제작을 위한 종합자원을 구축하고 관리하는 시스템의 개발이 필요하며, 이러한 종합자원을 관리할 기구가 필요하다.

데이터베이스 자원을 종합적으로 관리하고자 할 경우 이 자원을 보다 체계적으로 분류하고 구축하는 기술이 필요하다. 예를들면 이들 정보의 효과적인 관리를 위해 정보자사전시스템(IRDS) 기술이 사용될 수 있을 것이다.

이런 데이터베이스 자원에 대한 정보는 일반적인 데이터베이스 제공자가 손쉽게도 빠짐없이 정보를 제공하는데 도움이 된다. 제공하고자 하는 정보가 많을지라도 일정한 형식이 없으면 누락되기 쉽고 결국에는 정보로써 제구실을 하지 못할 수 있다. 또한 제공된 정보가 특정 응용시스템의 형식만 따른다면 정보의 가치는 반감할 것이다. <이 글은 체신부 통신진흥협의회 내 DB산업육성분과회에서 연구 수행한 "DB산업활성화방안 연구"의 하나인 DB관련기술(수행자:강원대 김진호 교수, 서울시립대 홍의경 교수)을 실은 글이다.> DB