

데이터베이스 관련 기술(Ⅲ)

Technical Approach in Database

목 차

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

5. 데이터베이스 검색기술

정보화 시대를 맞이하여 정보서비스에 대한 요구가 급증할 것으로 보아 많은 양의 정보를 손쉽게 가공하고 검색할 수 있는 기능을 가진 검색시스템이 요구된다. 현재의 검색시스템은 대부분 메뉴 구동방식을 채택하고 있고 또한 정형화된 정보에 대해서는 효율적인 검색이 이루어지지만 비정형화된 정보의 처리에 문제가 있으므로 고도화된 정보서비스 제공에는 적합하지 않다.

단순한 업무의 효율화를 위한 데이터베이스는 대상이 되는 데이터에 표현된 문자나 수치를 입력, 처리하여 여러가지 형식으로 출력할 수 있으면 된다. 그러나 판매전략이나 시장 예측을 수립하기 위해서는 이용자가 요구하는 목적에 맞추어 대상이 되는 데이터의 내용을 평가하거나 분석된 데이터를 입력하여 검색할 필요가 있다. 이용자의 목적과 요구에 맞추어 정보를 수집하고, 정보가 갖는 의미 내용을 분석, 가공, 축적 및 검색할 수 있는 새로운 검색형 데이터베이스 시스템이 필요하다. 우선 환경이 편리하고 사용자의 의도를 파악할 수 있는 자연어 질의 처리기 같은 인터페이스, 인식된질의에 대해 검색결과를 빠르게 찾아낼 수 있는 지능형 색인기법 등의 여러가지

기술이 필요하다.

궁극적으로 데이터베이스검색기술은 대규모 정보검색 시스템을 구축함으로써 방대한 자료의 효율적인 검색이 필요한 정보서비스 분야의 사업화 및 활성화에 도움을 주고, 제조업의 경쟁력 강화에 기여하는 것이다. 데이터베이스 검색기술은 다음 표5-1와 같이 정리해 볼 수 있다.

	내 용	비 고
검색기법	- 지식기반 정보검색 - 문헌검색 시스템	지식기반 검색 모델
인식기법	- 자연어 인식 (한글분석) - 비정형화된 데이터 인식	
색인기법	- 통일어색인기술 - 자연어색인기술	색인을 위한 모델 개발 필요
검색명령어	- 검색명령어 통일 - 자연어 질의 처리기	

표 5-1 데이터베이스 검색기술

가. 검색기술

● 지식기반 정보검색

기존의 데이터베이스 시스템은 단순히 키워드나 불리안 연산자를 이요하여 자료를 찾는 것이다. 이러한 방법은 키워드를 정확하게 입력해야 하기 때문에 유사한 의미의 다른 키워드는 찾을 수 없고, 많은 양이 검색될 경우에 사용자가 요구한 내용을 판단하기 힘들고, 비정형화된 데이터에 대한 검색이 어렵다는 문제점을 가지고 있다. 그러므로 새로 구축할 정보검색시스템은 단순한 키워드나 불리안 연산을 이용하는 것이 아니고 지식베이스(Knowledge base)를 기반으로 정보검색

이어야 한다. 지식베이스를 이용하여 사용자의 질의와 저장된 문서와의 관련성을 판별하여 문서를 검색해 주고 검색된 문서를 사용자의 의도에 적합한 순서로 제시해 줄 수 있어야 한다.

또한 자연어로 입력받아 그 의미를 인식할 수 있어야 한다. 방대한 양의 정보중에서 시스템 사용자가 요구하는 정보만을 선별하여 제공해 줄 수 있어야 하고, 동시에 시스템 초보자도 손쉬운 검색을 할 수 있도록 자연스러운 검색 방법을 제공해야 한다. 지식기반 검색 모델로는 시소러스 기반 검색모델과, 기존의 모델에 퍼지 개념을 도입한 새로운 모델 등을 고려할 수 있다. 시소러스 기반 검색 모델은 사용자의 의도를 파악하여 가장 적합한 순서대로 정보를 검색해 주는 것으로서 사용자의 질의어와 자료의 색인어를 비교하여 가장 가까운 의미를 가진 자료부터 검색해 주는 방법이다.

● 문헌 검색시스템

일반정보의 90%를 차지하는 비정형 문소(텍스트, 그래프, 이미지등의 복합문서)로 표현된 정보에 자산 가치를 부여할 필요가 있다. 일반 데이터베이스와는 달리 필드 전체 내용에서 색인의 가치가 있는 단어(또는 전문에 들어 있는 모든 단어)에 대한 색인을 제공함으로써 비정형 정보에 대한 빠른 검색을 제공해 주는 시스템이 문헌 검색 시스템이다. 일반적인 경우 관계형 DBMS를 사용하고 수동 색인 처리를 하므로 속도 및 사용상의 문제가 발생하고 있다. 일부 자체 개발 검색시스템의 경우 데이터베이스 관리 및 보안 기능이 없고 색인어 추출의 경우도 수작업 경우가 많다. 외국 패키지를 도입하여 사용하는 경우 일반적으로 관계형데이터베이스 관리기능 부족으로 다른 관계형 데이터베이스와 연결하여 사용하며, 한글 색인 추출 등의 한글처리 기능이 미비하다. 수동색인 처리는 판단기준의 가변성(담당자 주관, 사회적 배경 등)이 있고 대량

의 데이터 처리가 미흡하며, 자동 색인의 경우 언어의 모호성이 있다. 문헌검색 시스템과 관계형DBMS의 장점과 단점은 표 5-2와 같다.

	장 점	단 점
문헌검색 시스템	-전문검색 기능(단어, 구 등) -빠른 검색 속도 -대량의 데이터 처리(저장 및 검색) -비정형 데이터 처리 -시소러스 기능	-관리기능 -정형필드 처리 기능 -보안기능 -출력기능
관계형 DBMS	-정형데이터 처리 -보안 및 동기화 -REPORT WRITER, SORTING 등의 출력 -관리기능(MONITORING 등)	-전문검색 기능 -비정형 데이터 처리 -문헌 데이터 출력 및 입력 편집기능 -데이터 크기의 제한 -시소러스 기능

표 5-2 문헌검색시스템과 관계형DBMS의 장단점

나. 인식기법

● 자연어 인식(한글인식)

새로운 검색시스템은 사용자의 편의 및 의도에 맞는 응답을 위해 자연어 질의 처리기가 필요하다. 외국에서 개발된 자연어 질의 처리기는 한글 지원이 미흡하기 때문에 우리에게 그대로 적용하기는 어렵지만 한글처리기술에 대한 연구가 없으면 한글정보 처리 분야에서 기술종속이 우려가 있다. 한글에 대한 분석은 한글로 표현된 질의를 올바르게 인식할 수 있어야 한다. 한글에 대한 인식을 위해 형태소 분석, 구문분석, 의미분석, 담화분석이 필요하다. 형태소 분석은 어절단위로 자연어 질의문 자연어 질의문을 형태소 단위로



분석하는 것이다. 형태소 분석기는 하나의 어절을 이루는 형태소들간의 구성원리(어절 구조 규칙)를 사용하여 어절타입의 예측을 통한 효율적인 분석을 할 수 있어야 한다. 구문분석은 어절 사이의 관계를 밝히는 구문해석 과정이다. 구문 분석기는 여러개의 형태소 분석결과 중에서 구문적으로 올바른 결과만을 선택할 수 있어야 한다. 의미/담화 분석은 한글의 특성인 생략과 대용어의 처리와 정보검색의 질을 높이기 위해서 필요하다. 의미분석은 여러개의 구문분석 결과중 의미적으로 적합한 것만을 선택하기 위해 필요하다. 담화분석은 우리말의 특성인 빈번한 생략과 대용어위 처리를 위하여 필요하다.

● 비정형화된 데이터 인식

비정형화된 데이터는 문자와 같이 어떤 정형화 되지 않은 화상이나 음성정보를 말한다. 화상정보는 주로 책, 신문과 같은 인쇄된 매체에서 얻을 수 있는데 정보의 유용성을 높이기 위해서는 화상정보도 필수적으로 처리해야 한다. 또한 컴퓨터와 사람이 직접대화를 통해 어던 질의를 처리해 주기 위해서는 이러한 비정형화된 데이터에

대해 인식하는 방법이 잘 정립되어 있어야 한다.

다. 색인기법

정보홍수 시대를 맞이하여 많은 양의 정보가 쏟아져 나오고 있는데 이러한 정보를 정보이용자의 요구에 맞추어 제공하기 위해서는 색인기법이 필요하다. 좋은 색인 기법이 되기 위해서는 문서의 내용을 명확히 표현하는 표현력과 대량의 문서 집합속에서 특정 문서를 정확히 구분할 수 있는 식별력을 갖추고 있어야 한다.

데이터베이스에서는 저용내용의 핵심을 색인언어로 표현하므로 색인방법의 결정은 시스템 구축상 매우 중요한 부분이다. 색인언어의 종류로는 통제어 방식과 자연어 방식이 있는데 각각 장단점이 있으므로 자사의 업종분야 규모에 따라 결정해야 한다. 표5-3에서 보는 바와 같이 통제어 방식은 분류, 시소러스 및 용어집 등을 이용할 수 있다. 자연어 방식에는 자유 키워드 부여와 자동 키워드 추출방법이 있다. 이들 여러가지 색인 언어중에서 채택할 수 있는 기준의 예를 들면 자사의 사업분야가 명확하게 한정되어 있을 경우에는 통제어 방식의 시소러스를 활용할 수 있으며, 사업대상이 자주 변화할 경우에는 자연어 방식인 자유 키워드로 색인하는 것이 효율적이다. 예산과 인력이 충분하다면 통제어와 자연어를 병용하는 것도 고려할 수 있으며 선진국의 주요 데이터베이스는 검색 효율을 극대화 하기 위하여 통제어와 자연어의 장점을 모두 수용한 병용 방식을 대부분 채택하고 있다.

이와 같은 색인기법에 이외에 자동 색인기법, 지능형 색인기법 등을 고려할 수 있는데 이러한 색인기법은 기업체 단위로 보유하는 것보다 데이터베이스진흥센터 같은 기관에서 구축하여 무료로 공급함으로써 데이터베이스산업의 활성화에 기여할 수 있다.

● 통제어 색인기술

시소러스 방식을 채택할 경우에는 기존 시소러

구 분		통 제 어			자 연 어	
		분류코드	시소러스	용 어 집	자유 키워드	차동 추출어
통제어와 자연어의	장 점	검색호율(재현율, 적합율) 제고			색인부하가 적음 전문용어, 신조어 색인 기능	
	단 점	검색시 참고용 자료 필요, 색인 부하가 높음, 개념 변화 추가기능 저조			검색효율 저조	
검색어수	수십-백	수천-만	수백-만	무 한	무 한	
색인부하	소	대	중	소	무	
기억용량	소				대	
적용대상	큰 개념을 어느 정도 한정할때	세부개념, 구조 적인 검색이 필 요할때	병렬개념, 구조 적인 검색이 불 필요할때	보조 키워드로서 유효	긴 문장항목이 포함되어 있을 때	

표 5-6 통제어 색인기술과 자연어 색인 기술의 비교

스를 그대로 이용하는 방법, 기존 시소러스의 해당 분야를 선택하여 사내용으로 재편상하는 방법, 자사의 고유 시소러스를 작성하는 방법등을 고려할 수 있다. 현재 국내에서 발간된 시소러스는 1992년 시스템공학연구소가 제작한 40,000여 용어 규모의 과학기술 용어 시소러스와 1993년 한국언론연구원에서 제작한 10,000여 용어 규모의 신문기사 종합 시소러스가 있다. 통제용어집으로는 산업기술정보원과 시스템공학연구소가 공동으로 제작한 40,000여 용어 규모의 과학기술 한글일 용어번역집이 있다.

자사 고유의 시소러스 제작은 극히 소규모가 아니면 많은 인력과 예산이 필요하므로 사전에 충분한 검토가 필요하다. 이외에 통제어 방식으로는 과학기술 용어번역집 혹은 핸드북 등에서 필요한 용어를 추출하여 사내의 색인 작업용 키워드 집으로 제작하여 활용하는 방식이 있다.

● 자연어 색인 기술

필요한 자료를 효율적으로 찾기 위해서는 정보를 저장할때 자료를 색인해 놓아야 한다. 색인을 하기 위해서는 단순한 키 값만으로는 효율적이지 못하고 통계 및 의미를 기반으로한 색인기법이 개발되어야 한다. 자동색인 기능을 제공하는 시스템 개발이 필요하며 이는 모든 분야에서 색인되어야 하므로 각 분야별로 색인 작업을 도와주기 위한 용어 사전이나 색인지침 등의 개발이 먼저 이루어져야 할 것이다. 자연어 처리 소프트웨어를 활용하면 더욱 효율적이지만 한글 데이터에 대해서는 아직 완벽하게 적용할 수 있는 단계는 아니라고 판단된다. 자연어 색인기술에는 컴퓨터에 의한 자동색인을 의미하는 자동 키워드 추출 방법과 임의의 키워드를 부여하는 자유 키워드 방법이 있다.

자동 키워드 추출 부문은 단어간의 통계적 상

호의존관계를 이용한 색인방법 및 단어의 역문헌 출현빈도에 따른 색인방법을 구현하여 비교분석하는 것이다. 자유키워드 부여방법은 기사에서 사용하는 용어를 그대로 사용하거나 색인자 임의로 키워드를 부여하는 것인데 이때에는 여러명의 색인자가 가능한 일정한 형태의 색인어를 추출할 수 있는 색인 매뉴얼의 제작이 필수적이다.

라. 검색명령어와 자연어 질의 처리기

사용자들의 데이터베이스를 편리하게 사용하며, 모든 데이터베이스를 같은 명령어로 검색이 가능하도록 통일된(표준화된)명령어 개발이 필요하다. 데이터베이스 유통업체들이 제공하는 데이터베이스 서비스는 상이한 검색 명령어와 검색 화면을 제공하기 때문에 사용자가 데이터베이스 정보를 검색하는데 큰 어려움을 느끼므로 정보검색의 혼란과 데이터베이스 유통산업 활성화에 장애요인이 되고 있다. 데이터베이스 이용자에게 용이한 시스템 사용환경 제공과 데이터베이스 이용활성화 및 데이터베이스 유통산업의 발전을 위해 데이터베이스 검색명령어, 검색화면 및 데이터 입력방식의 표준 마련이 시급하다.

표준안이 마련되면 데이터베이스를 이용하는 사용자 측면에서는 각기 다른 시스템의 데이터베이스를 검색색할때 새로운 검색명령어를 익힐 필요가 없게 된다. 데이터베이스 검색화면을* 통일함으로써 새로운 데이터베이스 검색에 적용할 수 있다. 초보자라도 쉽게 접근할 수 있는 표준검색 명령어 및 검색 화면을 제공함으로써 데이터베이스 활용인구의 저변확대 및 정보산업 활성화에 기여할 수 있다. 서로 다른 데이터베이스간의 정보 전달시 공통검색 명령어 및 검색화면을 이용하여 데이터베이스를 공유할 수 있으므로 데이터베이스 제작의 중복 투자를 줄일 수 있다.

● 검색명령어 통일

사용자 인터페이스가 잘 정립되었다고 하더라도

도 기본적으로 사용해야 할 검색명령어 있다. 검색 시스템마다 고유의 명령어를 사용한다면 사용자는 매우 불편을 겪게 될것이다. 사용자들이 데이터베이스를 편리하게 사용하도록 모든 데이터베이스들과 검색시스템들이 같은 명령어로 검색되도록 통일된(표준화된) 명령어를 개발한다면 이러한 불편이 해결된다.

● 자연어 질의 처리기

PC의 보급으로 데이터베이스에 대한 수요자는 기업이나 연구소만이 아닌 개인으로 더욱 확대되고 있어 좀 더 다양한 수요자와 수준의 차이를 극복할 수 있는 노력이 필요하다. 사용자에게 편리한 사용환경을 제공하며 사용자의 의도를 파악하여 원하는 문서를 쉽게 찾을 수 있도록 도와주는 자연어 질의 처리기가 필요하다. 자연어 질의 처리기는 사용자가 시스템에 정의된 명령어를 사용하지 않고 자연어로 질의할 수 있는 질의 처리 시스템으로서, 사용자가 시스템의 명령을 외우지 않아도 시스템을 조작하여 검색을 할 수 있으며 불리안 연산자로서는 부정확하게 표현되기 쉬운 사용자의 의도를 정확하게 포착할 수 있는 것을 의미한다. 이것을 구현하기 위해서는 앞에서 언급한 한글에 대한 분석이 잘 정립되어야 한다. 한글을 기반으로 하는 검색 서비스를 제공하기 위해 한글에 대한 정보처리 기술도 발전해야 한다.

6. 데이터베이스 시스템기술

가. 분산데이터베이스

● 필요성

현재의 데이터베이스 서비스는 대형 컴퓨터를 호스트로 두고 데이터베이스를 집중관리 하는 것이지만 앞으로 데이터베이스 산업이 활성화 되면 그 종류도 많아질 것이며 정보 생산업자나 지역에도 제한이 없게 될것이다. 따라서 데이터베이스는 여러지역에 분산 유지될것이며 이러한 데이

터베이스를 사용하는 지역에 대한 제약이나 사전 지식어 없어 자유롭게 사용할 수 있어야 할 것이다. 또한 각 지역에서의 데이터베이스 서비스는 그 지역국에서 서비스하게 함으로써 중앙 집중형으로 서비스할 경우 발생할 수 있는 사용자 과다로 인한 시간 지연문제나 데이터 충돌 현상을 해결할 수 있으며 그 지역 실정을 잘 반영하는 데이터베이스를 수집, 관리할 수 있다. 따라서 데이터들이 여러 장소에 분산 유지되는 경우 사용자에게 일관성 있는 서비스를 제공하고, 데이터의 과다한 중복없이 데이터베이스를 일관성 있게 유지하는 기술이 필요하다.

● 분산형DBMS 형태

이미 단일 위치에서 가동하고 있는 DBMS를 작업량이나 위험을 분산시킬 목적으로 하는 하향식 구축방법과 이미 여러 위치에서 독립적으로 가동하고 있는 DBMS를 정보 자원의 유효 이용이라는 관점에서 통합하는 상향식 전개방법이 있다. 또한 각 구성 DBMS의 성격에 따라 동종(HOMOGENEOUS)구성과 이종(HETEROGENEOUS)구성으로 나눌 수 있으며 컴퓨터 네트워크로서 WAN을 이용할 것인지 LAN을 이용할 것인지에 따라서도 그 구성형태상의 차이점이 있다. 단일 DBMS에서 문제가 되는 트랜잭션관리, 동시성제어, 회복에 관한 모든 사항들이 분산 환경에서는 더욱 복잡하기 때문에 이에 대한 효율적인 관리가 필요하다. 즉, 데이터가 여러지역에 분산되어 유지되고 있기 때문에 발생하는 데이터의 중복을 얼마나 최소화할 것이며, 데이터에 대한 갱신작업시에 어떻게 각 중복데이터를 갱신할 것인가, 다수 사용자가 이용하고 있는 데이터에 대한 일관성은 어떻게 유지할 것이며, 한 지역 DBMS나 선로의 고장시 이용자에게 어떻게 투명한 서비스를 제공하며 고장에 대한 복구는 어떻게 할 것인가에 대한 연구가 이루어져야 한다.



● 분산DBMS의 연구개발 상황

분산 DBMS의 효시는 미국 CCA(COMPUTER CORPORATION OF AMERICA)사의 SDD-1이다. 그 이후 수 많은 연구개발이 진행되어 왔다. 표6-1에 주된 시스템연구 개발 상황을 기술하였다. 아직 우리나라에서는 분산DBMS가 개발되지 못한 상황에 있어 이의 개발이 시급하다.

나. 객체지향 데이터베이스

● 필요성

객체지향 데이터베이스 시스템은 객체지향 데이터베이스를 정의하고 조작할 수 있는 데이터베이스 시스템을 의미한다. 객체지향 개념은 CAD/CAE/CASE/CAM시스템, 지식기반 시스템, 다중매체 정보시스템과 첨단 사용자 인터페이스 시스템과 같은 차세대 데이터베이스 응용을 위한

표 6-1 분산형데이터베이스 시스템의 연구개발 현황

구분 시스템명	연 대	기계(OS)	통신망	설치 언어와 크기	사용자 언어	유 형	수행자
SDD-1	'76-'79	DEC-10과 DEC-20으로 합계 4대	ARPANET	BCPL 25,000행	Datalanguage (관계 언어)	동종 구성 institutional	CCA(Computer Corporation of America)
분산형 INGRES	'76-	DEC VAX 11/780과 11/750 (UNIX) 각각 5대	ETHERNET (3MB)	Language C 비분산형 INGRES의 1.7배	QUEL EQUER		캘리포니아 대학 버클리분교
Sytem R*	'79-'84	IBM 4341 3대	SNA Virtual Circuit	PLS/III	SQL PLI/SQL		IBM San Jose 연구소
ENCOM- PASS	'81	Tandem Non Stop comput ers	Wide area net(56 KB)	TAL	ENFORM	동종 구성 Commercial ('81)	Tandem Computer사
DDTS	'79-	Honeywell level-6 mini-compu ters	Bus 결합 (Coax-cable)	Language C 20,000행	GORDAS(on ECR)	이종 구성 institutional	Honeywell사
특별명 없음	'80(?) -	An Prime computers	Primeret	Implemeation 대부분원료	CODASYL형 언어	institutional	Prime Computer사
Multibase	'80-'84	DEC VAX 11/780(V MS)	ARPANET	Ada	DAPLEX (on FDM)	이종 구성 institutional	CCA
DDM	'80-'84	DEC VAX 11/780(V MS)	ARPANET	Ada	DAPLEX (on FDM)	동종 구성 institutional	CCA
SHARD	'82-'86	DEC VAX(UNIX) 외 는 SUN(UNX)	DARPA Internet	Language C	QUEL		CCA IBM사
DB2 버전2, 릴리스 2	'88	IBM MVS	VTAM (프로토콜 LU6.2)	공표되어 있지 않다	SQL (ANSI 표준)	이종 구성 commercial	IBM사

시스템명	구분	연 대	기계(OS)	통신망	설치 언어와 크기	사용자 언어	유 형	수행자
SIRIUS DELAT		'79-'81	Intertechnique IN2000 mini computers 3대	DANUBE (LAN)		Relational 언어(SQL에 유사)	이종 구성 institutional	INRIA
POLYPHENE		'76-'81		CYCLADES		Relational 언어		Grenobel 대학(프)
POREL		'77-'82	mini computers	FRAGNET	PASCAL		동종 구성 institutional	Stuttgart 대학(독)
VDN		'77-	Intel 8086 microprocessor와 UNIX based기계	wide area 및 local area net	PASCAL	레코드 based언어와 SQL	동종 구성 commercial (1985)	Nixdorf Computer 사(독)
UDS-D '84-'86	Siemens (B/S2000)		Siemens net	SPL	CODASYL형 언어	동종 구성 commercial	Siemens 사(독)	
JDDBS		'78-'83	FACOM M170F (AIM) 과 NEC Acos-700(ADA BAS)	local area net	PL1	RQL (QUEL에 유사)	이종 구성 institutional	일본정보처리개발 협회
DEIMS		'80-'82	NTT의 DIPS 기계	wide area net 와 채널 결합	SYSL	COBOL과 SYSL	동종 구성 commercial (1983)	NTT(개발담당은 일본전신전화공사)
RDB/DV		'82-'86	FACOM M380 2대와 M180 2대	FNA Virtual Circuit	사외비밀	RDB/QL	동종 구성 institutional	후지통신연구소 (주)
분산형 FRED		'85-'88	mini와 micro computer	OKINET 2000 모델 20 (10MB)	Language C	SQL (ISO표준)	동종 구성 institutional	오끼전기공업(주)

풍부한 기능을 지원하는 데이터의 모델의 적합한 기반을 형성한다. 왜냐하면 객체지향 개념은 기존의 프로그래밍 언어에서 보았던 객체를 클래스로 그룹화하는 개념(레코드를 레코드형으로, 튜플을 릴레이션으로 그룹핑하는 것에 상응), 객체와 이를 구성하는 객체들간의 집단체화 관계(중첩 객체)와 객체의 클래스와 이로부터 세분화된 객체들의 클래스 사이의 일반화 관계와 같은 데이터 모더링 개념을 이미 포함하고 있기 때문이다.

현재 정보시스템을 구축할때 관계형 데이터베이스는 사용의 단순성, 이론적 배경이 확실한 점 등의 특징을 인정받아 급속히 사용이 증가되고 있지만 다음과 같은 불충분한 점이 있다.

- 실세계를 모델화하는 능력의 부족
 - 질의어와 프로그래밍 언어와의 불용합
 - 소프트웨어 엔지니어링 기술침투의 어려움
- 반면 객체지향 데이터베이스는 다음과 같은 특징을 가지고 있다.
- 객체개념에 의해 내부표현이나 내부조작이 은폐(encapsulation)되므로 다중매체의 데이터 표현을 자연스럽게 취급할 수 있다.
 - 계승(inheritance)등을 사용한 추상화 지원에 의거한 데이터 정의 능력이 있으므로 사용자가 새로운 형의 객체를 용이하게 정의할 수 있다.
 - 메세지에 의거한 객체 조작이 가능하므로 데이터베이스의 동적 상태 기술능력이 뛰어나다.

다. 멀티미디어 데이터베이스

● 필요성

지금까지 구축된 데이터베이스의 데이터 유형은 대부분이 텍스트 위주였다. 아직까지는 데이터베이스에 대한 사용이 미미하고 서비스 제공 수준도 바람직하지 못한 단계이기 때문에 큰 불만족은 느끼지 않고 있다. 그러나 데이터베이스

구축과 이에 대한 이용이 활성화 된다면 문자 데이터 뿐만 아니라 도형, 영상, 화상, 음성 등 모든 형태의 데이터를 처리할 수 있는 기술이 요구된다. 이러한 멀티미디어를 다룰 수 있는 DBMS는 일부 상용되고 있으나 이를 어떻게 온라인 서비스 할것인가에 대한 문제점은 남아있다. 즉 기존의 텍스트 위주의 데이터베이스에서 수행되었던 각종 기술을 어떻게 적용할 것인가에 대한 연구가 진행되어야 한다. 어찌면 이러한 멀티미디어 데이터에 대한 효율적인 저장과 검색을 제공할 수 있는 새로운 데이터베이스 엔진이 요구될지도 모른다.

라. 하이퍼텍스트 시스템

종래의 책과 같은 정보매체 방식에서는 순차적 정보 접근만이 허용되므로 비순차적이고 연관적인 인간의 사고로 정보를 분석하고 이해하는데는 상당한 어려움이 따른다. 따라서 사람의 연상작용과 유사하게 비순차적으로 필요한 정보를 상호 연결하여 나아감으로써 정보를 얻는 방식이 등장하게 되었는데, 이를 하이퍼텍스트(HYPertext)라 한다. 최근에는 멀티미디어 기술의 발달로 문자 정보뿐만 아니라 이미지, 그래픽, 오디오, 비디오, 애니메이션 등의 정보를 다룰 수도 있는데 이를 강조하여 하이퍼미디어(HYPERMEDIA)라고 표현하기도 한다.

하이퍼미디어 시스템은 정보를 노드와 노드를 연결해 주는 링크들의 네트워크로 구성하며, 시스템 사용자는 노드와 링크조작을 따라 자유롭게 탐색함으로써 원하는 정보를 검색할 수 있다. 그러나 하이퍼미디어가 모든 응용 분야에 적합하지는 않지만 전자사전, 전자백과사전, 교육분야, 의학정보분야, 온라인 분석, 소프트웨어 공학(CASE), 전자출판, 프로젝트 관리, 외국어 교육 등의 응용분야에 적합하다. 이처럼 하이퍼미디어 시스템은 융통성 있는 구조와 자유로운 탐색이 제공하는 잇점 때문에 많은 분야에 이용이 되고 있지만

이같은 자유로움과 융통성은 몇가지 중대한 문제점을 일으킨다. 그중 대표적인것이 하이퍼미디어 네트워크상에서의 사용자 방향 상실과 링크를 생성, 관리하고 선택하는 데 따른 인식의 부담이다. 또한 이러한 멀티미디어 정보에 대한 서술정보를 어떻게 저장 관리하며 이에 대한 질의와 탐색은 어떻게 할 것인지에 대한 연구가 필요하다.

마. 데이터베이스머신

● 필요성

요즘들어 날로 대중화되어 가고 있는 관계데이터베이스 시스템에서 복잡한 질의 또는 대용량 데이터베이스의 신속한 처리를 위하여는 고성능 데이터베이스 시스템이 절실히 요구되며, 이러한 필요성은 최근 정보화 사회 추세에 의하여 점차 증대되고 있다. 고성능 관계 데이터베이스 시스템의 실현을 위해서는 관계DBMS에서의 효율적인 질의처리방법이 필수적으로 요구된다. 그러나 기존의 단일 프로세서 컴퓨터상에서 운용되는 DBMS는 그 처리 용량의 한계로 인하여 충분한 성능을 내기 어려우며 이러한 한계는 다중 프로세서를 사용한 질의의 병렬처리를 통해 극복할 수 있다. 고성능 관계 데이터베이스 시스템의 실현을 위한 관계 질의의 병렬처리 시도는 데이터베이스 처리에 적합하게 설계된 특수용도 컴퓨터인 데이터베이스 컴퓨터의 개념이 생기면서부터 시작되었다. 1970년대에 들어서면서 가속화된 기술적인 발전과 더불어 하드웨어 가격의 급격한 하락으로 하드웨어 개발에 대한 부담이 줄어들게 되자 데이터베이스 컴퓨터에 대한 연구가 시작되었다. 또한 1970년대 말에 괄목할만한 성장을 보이게된 관계 데이터베이스의 시대적 배경조성은 데이터베이스 컴퓨터에 대한 연구를 구체화시키는데 많은 도움을 주었으며, 근래에 들어 처리 데이터의 대용량화가 연구를 가속화시키는 원인이 되었다. 데이터베이스 컴퓨터의 구조가 저가

의 고성능 마이크로프로세서를 사용한 다중 프로세서 형태로 전환되면서 부터 대량의 마이크로프로세서를 사용한 대규모 병렬 데이터베이스 컴퓨터의 구조에 대한 연구가 활발히 진행되고 있다.

바. 데이터 및 객체 저장관리자

현대사회가 고도의 정보화 산업시대로 변화됨에 따라 많은 양의 데이터를 효과적으로 저장하고 관리할 수 있는 기술이 모든 분야에서 요구된다. 이러한 많은 양의 데이터를 체계적이고 효율적으로 관리하기 위하여 데이터베이스 관리 시스템은 크게 데이터를 저장하는 부분과 질의어를 처리하는 부분으로 나눌 수 있는데 이중 하부구조에 해당하는 데이터를 저장하고 관리하는 부분을 저장 시스템을 저장시스템이라 한다. 저장시스템은 데이터를 디스크에 저장하고 검색하는 부분을 직접관리함으로써 데이터베이스 관리시스템의 성능에 직접적인 영향을 준다. 데이터베이스 개발과 더불어 이를 효과적으로 지원하기 위한 저장 시스템의 연구 역시 활발하게 진행되고 있다. 저장시스템은 디스크의 액세스와 버퍼의 관리, 객체의 저장 및 관리 객체를 액세스하는 방법들을 제공한다. 이를 위해서는 물리적 디스크를 관리하는 디스크 관리자, 디스크 관리자와의 인터페이스를 통하여 버퍼의 내용을 디스크에 읽고 쓰는 역할을 수행하는 버퍼 관리자, 데이터 페이지와 객체를 관리하는 객체 관리자, 가변길이 키를 지원하는 B⁺트리 관리자가 필요하다. 또한 사용자가 이 저장시스템을 보다 쉽게 사용하도록 편리한 인터페이스를 제공하는 스캔 관리자도 필요하다.

데이터베이스 관리 시스템을 국산화 하는데 가장 핵심이 되는 부분이 저장시스템이기 때문에 이런 저장시스템을 시급히 개발 보급해야 한다. 보급된 저장시스템을 바탕으로 국산 DBMS를 개발하게 되면서 저장시스템에 대한 중복 투자를 제거할 수 있고 저장시스템에 대한 기술을 쉽게

습득할 수 있으므로 다른 소프트웨어 산업에도 파급효과를 미칠 수 있다.

사. 한국정보 검색지원 DBMS

최근들어 국내 데이터베이스 관리시스템 시장의 규모는 폭발적으로 증가하고 있다. '92년에 약197억원 정도의 시장을 나타내더니 '93년에 들어서는 더욱 가속도가 붙어 약340억원 정도 시장으로 확대될것으로 예측되고 있다. 국내 DBMS시장이 외국에 비하여 앞으로 더 급신장할것이 예상되는 이유는 현재 대부분의 정보가 데이터베이스화 되어 있지 않은 상황에서 데이터베이스 산업이 태동기와 정보화 사회를 맞고 있다는 점이다. 세계 시장도 DBMS의 호황은 예외가 아니어서 UNIX 전문 잡지인 UNIX/WORD에 따르면 '92년도 UNIX시장에서 최대 매출액을 기록한 소프트웨어 업체들중 1위에서 4위까지는 데이터베이스 시스템 전문회사(ORACLE, INFORMIX, SYBASE, INGRES)가 독차지 하고 있음을 나타내고 있다. 이러한 때에 국내 DBMS시장의 현황을 보면 오라클을 비롯한 외국 유명사의 상품이 거의 전 시장을 점유하고 있다. 이러한 상황은 국제 수지면에서 쏠만 아니라 국내 데이터베이스 산업 기수의 해외 기술의존도를 심화시키는 결과를 초래하고 있다. DBMS기술의 중요성은 모든 컴퓨터 시스템이 사용되는 곳의 약 80%는 컴퓨터를 사용하기 위한 것이라기 보다 데이터베이스를 사용한다는 통계와 DBMS가격이 시스템 가격을 상회한다는 사실에서 알 수 있다.

근래에 들어 국내 연구소 및 기업에서 이러한 중요성을 인식하고 바다를 비롯한 국산 DBMS의 개발 및 상품화에 나서고 있다. 그러나 개발력의 미숙 및 경험부족으로 해외 유명제품과의 어려운 경쟁을 벌이고 있다. 현재 상품화되고 있는 정보검색시스템이나 대부분의 DBMS는 외국에서 개발한 제품에 한글을 지원하는 형태이다.

따라서 우리의 실정에 충족되지 않으며 한글에 다양한 처리를 할 수 없는 실정이다. 앞으로 데이터베이스 산업이 육성되면 이는 당연히 대부분이 한글자료가 될 것이며 이에 대한 검색과 관리가 이루어져야 할 것이다. 지금까지의 시스템은 정보검색 시스템과 DBMS가 별개로 존재하였으나 이제는 정보검색 기능을 갖고 한글을 처리할 수 있는 국산DBMS의 개발이 요구된다. 이러한 DBMS는 한글에 대한 자연스런 입출력과 한글 질의, 한글정보검색, 멀티미디어 서비스, 음성인식 등의 기능을 갖추어야 할 것이다. 한글검색지원을 하는 국산 DBMS를 개발하게 되면 상당한 수입대체 효과를 얻을 수 있고 데이터베이스 산업의 근간을 이루는 DBMS에 대한 인력양성도 도모할 수 있다. 또한 개발된 국산DBMS를 값싸게 공급하게 되면 국내 데이터베이스 산업 육성의 붐을 조성할 수 있으리라 믿는다.

7. 결 언

정보화 사회를 맞이하여 데이터베이스의 중요성이 나날이 증가하고 있으나, 아직 우리나라는 초기단계라고 볼 수 있으며 데이터베이스 관련 기술분야가 매우 취약한 상태이다. 정보의 원활한 유통을 위한 표준화와 데이터베이스의 제작, 이용, 유통분야의 기초기술 개발이 필요하며 조작, 운영 기술뿐만 아니라 정보수집, 가공, 유통 서비스 등 제기술의 발전이 필요하다.

본문에서 데이터베이스 관련 기술을 제작 및 생산기술, 유통기술, 이용기술, 검색기술, 시스템 기술로 나누어 정리하였으며 각각의 기술에 대해 국내외에서 가구되고 개발된 현황을 소개하였다. 이러한 기술의 발전을 통해 유용한 데이터베이스를 사용자에게 제공하며 국내 데이터베이스 전문업체가 데이터베이스를 효과적으로 사용할 수 있게 되어 데이터베이스 이용의 확대가 데이터베이스 산업 활성화에 기여했으면 한다. **DB**