

데이터베이스기술 (II)

Technical approach in Database

정리/최 중태

목 차

1. 개요
2. 데이터베이스제작과 생산기술
3. 데이터베이스유통기술
4. 데이터베이스이용기술
5. 데이터베이스검색기술
6. 데이터베이스시스템기술
7. 격언

제 3 절 데이터베이스유통기술

1. 정보통신망 이용 기술

정보통신망 서비스 현황

80 년대의 PSTN, PSDN 등을 통한 개별적인 통신망 서비스는 90년대 들어서서는 ISDN, LAN, WAN 등의 고속 복합통신망으로 확산되고 있으며, 2000년대에는 B-ISDN(광역 종합 정보통신망)을 통하여 모든 통신 서비스가 하나의 망으로 제공될 것으로 예상되고 있다.

국내에서의 정보 통신망의 구축은 ISDN이 일부 지역에서 시범서비스를 하는 단계에 있으며 상용서비스는 1994년에 실현하는 것을 목표로 하고 있다. 통신망 사용료, 관련 종합정보서비스 기술의 낙후, 전송속도의 제한으로 인한 통신서비스 매체의 제한 등 문제는 산재해 있지만 고속의 디지털 단일망인 B-ISDN망이 2000년대에 이루어질 예정이다. 멀티미디어 관련 기술은 통신망에서의 서비스를 배제한 상태에서 PC에 CD-ROM을 장착하여 제공하고 있으나 비디오



기술 등 멀티미디어의 다양한 서비스를 제공하기에는 아직 초창기 수준이다. 특히 통신망에 접속하여 제공되는 서비스는 아직 일부 대학이나 연구소에서 극히 제한된 기능을 구현해 보는 정도이다.

이러한 정보통신망과 더불어 고속 반도체 집적 기술, 하드웨어 구현기술, 소프트웨어의 지능화에 따른 컴퓨터 분야의 발달은 상호 교환되는 정보도 텍스트를 중심으로 하는 개별적인 것로부터 텍스트, 그래픽, 음성 및 오디오, 화상 및 영상 등이 복합적인 형태로 축적, 처리 및 교환되는 멀티미디어 정보서비스 시대를 맞이하게 될 것이다. 따라서 차세대 정보서비스 데이터베이스 서버는 멀티미디어 정보를 처리, 축적할 수 있는 능력을 전제로 구현되어야 하며, 이러한 멀티미디어 정보를 서비스할 수 있는 시스템이 구축되어야 한다.

비디오텍스 시스템

최근 정보화 사회로의 진입이 가속되면서 점차 다양한 정보통신서비스가 요구되고 있다. 그 중

에서도 새로이 시작된 HiTEL 서비스 등에 영상 정보 서비스인 비디오텍스가 많은 관심을 받고 있다. 이는 영상매체가 문자만의 정보서비스보다 더 큰 흥미와 친밀감을 주고 있음을 말해 주는 것으로서 고도 정보통신 서비스에서 비디오텍스의 중요성이 강조되고 있다.

국내의 비디오텍스 관련산업은 비디오텍스의 서비스를 지원하는 정보검색단말기와 PC용 비디오텍스 Emulation 소프트웨어는 사용가능한 수준에 있으나 속도의 개선, 정확한 표준 데이터, 구문의 복부호에 의한 화면구성 방법의 개선 등이 필요하다. 비디오텍스 저작도구 소프트웨어 및 서버는 외국의 시스템을 들여와 한글을 이식하여 제공하는 단계이므로 한글의 정확한 입출력과 데이터베이스상에서의 한글표현 및 한글의 색인화가 아직 미비한 실정이다. 현재 비디오텍스 서버는 국내 자체기술로 개발하고 있는 중이다.

그리고, 사용하기 편리하고 어떤 시스템에서도 비디오텍스 화면을 제작할 수 있는 저작도구를 개발 보급하여 창조력 있고 재능있는 많은 정보 제공자의 참여를 유도해야 한다. 또한, 창조적인 소규모의 정보제공자가 비디오텍스 서비스에 많이 참여할 수 있도록 하기 위해서는 저가의 비디오텍스 시스템의 공급이 필요하다.

2. CD-ROM 기술

데이터베이스 서비스는 크게 온라인 서비스와 오프라인 서비스로 나눌 수 있다. 현재는 대형시스템에 이용자의 PC를 통신회선을 통해 연결하는 온라인 방식이 주류를 이루고 있으나 CD-ROM과 같은 대기역용량의 저장매체의 등장과 함께 데이터베이스의 개인화가 이루어지고 있기 때문에 오프라인 서비스에 대한 수요도 증가하고 있다.

지금까지 국내 데이터베이스의 주요 유형은 문헌초록 형태의 2차정보가 주류를 이루고 있었다.

그러나 저장기술이 발달하고 이용자들이 사실형의 1차정보에 대한 요구가 증가함에 따라 전문(Fulltext) 데이터베이스 개발 구축이 요구되고 있다. 이러한 요구사항에 CD-ROM을 사용하고 있으나 그 이용도나 효율면에서 매우 저조한 수준이다. 따라서 CD-ROM의 정보도 온라인으로 제공할 수 있는 기술과 효율적인 검색 및 저장 기술이 요구된다.

3. 기타 정보 전달 기술

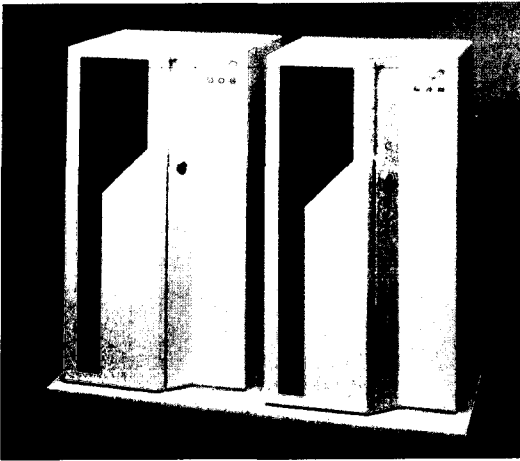
원격 데이터베이스 접근(RDA : remote database access)

컴퓨터가 널리 보급되고 컴퓨터 통신 기술이 발달함에 따라서 지역적으로 산재해 있는 컴퓨터 시스템들을 LAN과 WAN으로 연결하여 사용자에게 편리한 서비스를 제공하는 것이 중요하게 되었다. 또한, OSI 통신망을 이용하여 원격 데이터베이스를 액세스하는 요구가 증가하게 되었다.

원격 데이터베이스 액세스(RDA)는 응용계층에서 운용되는 다양한 응용과 원격지의 다양한 데이터 관리 시스템을 OSI 통신망을 이용하여 연결하는 기능을 담당하는 시스템이다.

원격 데이터베이스(RDA) 표준화는 ECMA(European Computer Manufacturers Association)에서 연구를 시작하였으며, 1985년 12월에 그 연구 결과를 ECMA Technical Report 30에 발표하였다. 같은 해에 ISO는 RDA를 새로운 표준화 항목으로 채택하여 지금까지 표준화 작업을 수행하고 있다.

현재 ISO에서 수행하고 있는 RDA표준화 프로젝트는 일반 RDA표준화와 RDA SQL 명세표준화 분야로 세분화하여 표준화 작업을 수행하고 있다. 일반 RDA표준화는 다양한 DBMS 환경에서 적용되는 공통적인 RDA 서비스 인터페이스를 표준화하고 있으며, 실제 RDA 시스템을 구현하기 위해서는 일반 RDA표준 규격은 관계형 DBMS뿐만 아니라 다양한 유형의 데이터관리시스템에 대



한 원격 액세스에 적용되는 일반적인 규격이므로, 이를 바탕으로 다양한 유형의 데이터관리시스템을 위한 RDA 특수 표준들이 개발될 것으로 예상되나 현재는 관계형 DBMS를 액세스하기 위한 RDA SQL 명세 표준화만 진행중이다.

RDA 표준은 응용 프로그램과 원격지 개방시스템에서 운용되는 데이터관리시스템 사이에 연동(internetworking)을 지원하기 위한 통신기능을 제공하는 것이 주목적이다. 일반적인 RDA 사용환경에서는 워크스테이션 환경하에서 데이터베이스 서버와 응용들이 LAN에 연결되어 있으며 게이트웨이를 통하여 WAN에 연결되어 있다. 여기서, 사용자는 자신의 워크스테이션에서 RDA 서비스를 이용하여 LAN 혹은 WAN에 연결된 원격지의 데이터베이스 서버에서 관리되는 데이터베이스를 액세스할 수 있다.

분산 트랜잭션 처리(DTP : distributed transaction processing)

컴퓨터의 대중화 및 통신의 비중이 높아짐에 따라 여러 회사가 각기 고유 기준에 따라 자체 기술로 제품을 개발, 판매하게 되어 상호간에 호환성이 존재하지 않는다. 이 문제에 대처하기 위해서는 다른 기종간의 연결을 가능하게 하는 부수장치를 부착하거나 한 회사 제품만을 구입하는

방법이 주로 사용되어 왔다. 그러나, 이런 단점을 해결하기 위해서 대부분의 컴퓨터 시스템에 공유되는 부분을 국제 표준화하여 통신과 정보의 공유가 가능하도록 하며, 각종 프로그램들간의 호환성을 높여 주도록 하는 개방형 시스템에 관심이 높아졌다.

개방형 시스템이 가장 활발한 분야중 하나는 데이터관리응용분야에서 서로 다른 이기종간의 연결을 지원하는 분산트랜잭션처리(Distributed Transaction Processing:DTP)분야이다.

OSI-DTP는 분산트랜잭션의 처리를 위해 필요한 사이트간의 데이터 교환과 각 사이트에서 행해질 동작들을 표준적으로 규정한 OSI 응용 계층 프로토콜이다. OSI-DTP 프로토콜 표준문서에서는 OSI-DTP 프로토콜을 지원하는 TPSP(Transaction Processing Service Provider)의 구성요소들 사이의 인터페이스, 각 구성요소들의 동작, TPSU와 의 인터페이스, 그리고 서로 다른 사이트에 존재하는 두개 이상의 TPSP간의 메시지 전송과 메시지 포맷에 대한 사항들을 규정하고 있다. 이러한 OSI-DTP 프로토콜을 기본 프로토콜로 이용하여 분산트랜잭션 처리시스템을 구축할 경우, 구축되는 사이트에 관계없이 시스템 확장이 용이하다는 장점이 있다.

이와 같이 DTP 시스템은 하나의 트랜잭션이 서로 다른 컴퓨터의 서로 다른 소프트웨어에서 수행되어야 할 경우 이를 지원하는 시스템이다. DTP 시스템은 여러 곳에서 수행되는 트랜잭션을 인식할 수 있다. 또한, 트랜잭션의 수행을 성공적으로 완료 또는 없었던 것으로 철회하는 결정은 그 트랜잭션이 전체 시스템에 수행한 모든 작업의 상태를 고려하여 내리고, 그 결정은 전체 시스템에 균일한 효과를 갖는다.

제4절 Database이용 기술

1. 사용자 편의 도구

맨 머신 인터페이스

인간공학이란 인간이 사용하는 도구, 기계 또는 환경을 설계하는데 있어 인간의 기능과 특성을 고려함으로써 보다 편리하고 안락하며, 안전한 그리고 효율적인 맨 머신 인터페이스 시스템을 만들고자 연구하는 분야로 주로 심리학, 생리학, 해부학, 공학 등의 학제적 연구를 수행하고 있는데, 최근에 와서는 사용자와 컴퓨터시스템 사이의 인터페이스에서 인간요소와 관련된 다양한 형태의 문제들을 해결하기 위한 많은 연구가 진행되고 있다. 맨 머신 인터페이스의 총체적 효율을 높이고 사용자가 보다 쉽게 컴퓨터시스템을 사용하기 해서는 컴퓨터 시스템으로부터 사용자에게 부과되는 육체적 또는 정신적 부담을 줄이는 것이 바람직하다.

인간과 컴퓨터 사이의 상호작용이나 이에 관련된 인적요소가 앞으로의 컴퓨터 기술발전에서 보다 중요한 위치를 차지하게 될 것이다. 컴퓨터 사용계층의 일반화 현상으로 컴퓨터의 주된 사용자가 극소수의 전문가들에서 일반대중으로 확산되면서 사용편의성이나 사용자의 수용도와 같은 인간요소가 시스템 설계자들의 중요한 목표가 되고 있다. 또한, 사용계층이 넓어짐에 따라 사용자의 인지능력, 개인취향 및 물리적 작업공간이나 작업 환경에서의 적합성과 같은 인간요소의 다양성을 수용하기 위해서는 맨머신 인터페이스에 대한 연구 및 그 결과의 응용이 점점 더 중요해질 것으로 예측된다.

텍스트 위주의 정보만을 다루다가 다양한 매체를 다루기 위해서는 인간의 감각과 정보처리능력에 대한 새로운 조명과 이해가 필요하게 된다. 새로운 매체의 효율적인 사용을 위해서는 인간의 시각 뿐만 아니라 청각, 촉각 등의 인간감각과 인지과정에 대한 연구가 맨 머신 인터페이스를 위해 중요하게 될 것이다.

지능형 단말기

국내 단말기 보급계획을 보면 2000년까지 300만대를 보급할 예정이다. 그러나, 정보 통신망 서비스에 사용되는 전용단말기는 텍스트 위주의 데이터를 검색하는 기능만을 제공하고 있다. 이와는 달리 정보통신망 서비스의 발달로 고해상도의 화면을 검색하고 동화상의 압축 복원기능을 갖는 단말기가 필요하다. 또한, 필요할 때는 통신서비스와는 독립적으로 단말기에서 작업한 후 통신서비스를 연결할 수 있도록 기능을 확장한 단말기가 필요하다. 이런 단말기를 지능형 단말기라고 한다.

지능형 단말기는 단말기로써의 역할 뿐만 아니라 정보를 저장하고 정보를 가공할 수 있다. 일반적인 정보제공자는 지능형 단말기를 통해서 정보통신망 서비스를 제공하고 정보가공자는 이런 원천 정보를 이용해서 가공한 후 지능형 단말기를 통해서 정보통신망 서비스로 쉽게 제공할 수 있다. 이런 기능외에도 무선기기의 기능을 포함하여 장소에 구애받지 않고 사용할 수도 있고 간단한 입력은 키보드대신 펜으로 할 수도 있다. 또한 터치 스크린 기능을 포함하여 간단한 검색을 할 수 있다.

이런 지능형 단말기 보급은 고부가가치의 데이터베이스 서비스를 활성화하고 데이터베이스 사용자의 저변확대를 도모할 수 있다.

자동 번역

한 언어를 다른 언어로 자동으로 변환시켜 주는 시스템을 자동번역시스템이라 한다. 이러한 시스템은 전화를 통한 자동통역 시스템으로 많이 쓰이거나 개발중에 있으며, 국제간의 교류가 빈번해짐에 따라 그 중요도가 더욱 증가하고 있다.

따라서 데이터베이스에 대한 자료도 국제화될 것이다. 이에 따른 언어로 표기된 자료를 자국의 언어로 표기하여 보거나 저장하는 자동번역시스템이 필요하며, 이에 대한 연구가 진행되어야 한다.

국내에서는 자동통화 전화연구가 한국통신과 ETRI를 중심으로 수행되고 있고, 그 일환으로 학계와 함께 호텔예약을 위한 텍스트베이스 질의 응답시스템 및 이의 고도화 연구, 텍스트 DB의 연구, 대화체계변역 연구 등 기초적인 연구가 진행되고 있다.

자연어 처리

자연어 처리에 있어 가장 공통적으로 사용되는 표현방식으로는 파스트리(parse tree), 어간역할 프레임(thematic role frame), 세계 모델(world model)이 있다. 파스트리와 어간역할 프레임에서 나타난 서술들은 개념(idea)이 단어와 문장들로 표현되는 방식에 의하여 크게 영향을 받는다. 반대로 세계 모델에서의 서술들은 특정한 단어와 문장의 선택에 의하여 영향을 받지 않는다.

파싱 프로시저에서는 단어순서 제약조건들에 관한 지식을 이용하여 파스 트리를 생성하고, 어간 역할 프레임 프로시저에서는 단어, 순서, 전치사, 특수 동사, 문맥에 관한 지식 등을 이용하여 어간 역할 프레임을 생성한다.

따라서 데이터베이스 시스템에 대한 질의나 명령이 자연언어로 기술되어 있을 경우 위에서 기술한 파싱 프로시저나 어간 역할 프레임 프로시저를 이용하여 서술된 문장을 이해하고 그에 대한 적절한 행동을 취해야 한다. 이러한 각 프로시저에 쓰이는 단어순서 제약조건이나 각 품사나 문맥에 관한 지식을 데이터베이스에 저장하여 관리할 수 있다. 이렇게 함으로써 사용자에게 더욱 편리한 인터페이스를 지원하며 일관성 있는 서비스를 지원할 수 있다.

음성 인식 및 합성

음성인식이란 음성속에 내재되어 있는 언어정보를 자동으로 추출하는 과정이다. 음성인식은 대 상음성의 종류(단어나 절을 떼어 발성하는 구분 발성 음성, 연속 발성한 연속음성), 어휘수(수십

단어, 수십~수백 단어, 수천 단어, 무제한 어휘), 화자수(특정 화자, 학습 가능한 다수 화자, 불특정 다수 화자), 환경(방음식, 일반 사무실, 소음이 심한 현장 등) 등의 기술적 요인의 조합에 따라 구현 가능성이 크게 좌우된다.

현재의 음성인식은 기본적으로 음성으로부터 음성 패턴(단어, 음소 등)의 특징을 추출하여 기준 패턴을 찾아내는 인식과정으로 나눌 수 있다. 이러한 알고리즘을 일반적으로 패턴 매칭(pattern matching)알고리즘이라고 부른다.

컴퓨터 환경에 익숙하지 않은 이용자에게 음성으로 명령이나 질의를 받아 들인다면 사용자는 더욱 편리하게 사용할 수 있을 것이며, 데이터베이스의 내용을 음성으로 합성하여 들려 준다면 더욱 편리한 사용자 환경을 제공할 것이다. 따라서 한글문장을 음성으로 변환시키는 시스템이 필요하고 음성 그 자체도 저장하였다가 음성으로 출력할 수 있는 기법도 필요하다. 그러나 지금까지 개발되고 있는 음성합성시스템은 발음의 표현에 있어 분절적이며 강약이나 완급이 조절되지 않는 상태이다. 따라서 사용자가 듣기에 지장이 없는 형태의 음성합성기술이 개발되어야겠다.

이러한 음성 인식과 음성합성기술이 적용되는 음성대화시스템에서 음성/텍스트 데이터베이스의 구축은 매우 중요한 일이다.

2. 표준 명령어 개발

데이터베이스에 관련된 용어 뿐만 아니라 데이터베이스를 사용하는 명령어도 어떤 표준화된 형식을 가져야 한다. 검색질의에 익숙하지 못한 사용자가 검색질의를 하면 질의의 잘못으로 정보가 손실될 수 있고 원하는 검색내용을 정확하게 표현하지 못해 잘못된 결과를 얻을 수 있다. 개발업자마다 독자적인 방법으로 명령어를 개발한다면 사용자가 쉽게 데이터베이스를 접근하여 올바른 내용을 얻을 수 없다. DB