

## 응용 계층에서의 과학적 데이터 관리기법과 연구사례

- 의학 분야의 데이터 관리를 중심으로 -

Management Techniques of Scientific Data on Application Layer and Case Study

- A Focus the Management of Medical Data



김영준

Kim, young-jun.

Instructor, Hankuk University of Foreign Studies.

### I. 머리말

이전에도 데이터베이스관리 시스템(Database Management Systems: 이하 DBMS)은 의학, 화학, 천문학, 지질학 등과 같은 과학분야에서 사용하는 데이터를 관리하는데 간혹 사용되기도 했지만, 실제적으로 그것은 데이터베이스 관점에서의 사용이라기 보다는 화일관리(file management) 측면의 요소가 더 강했다고 할 수 있다.

과학분야의 데이터를 관리하는 데이터베이스는 몇가지 특성을 가지고 있다. 첫째, 주변장치가 많이 필요하며, 둘째 표현해야 할 객체(object)의 수가 많고, 그들사이의 관계를 포함하는 정보구조 자체가 복잡하다. 예를 들어, 전형적인 검색체 데이터베이스는 40~50개의 객

本稿는 과학분야에서 사용하는 데이터 그중에서도 특히 의학분야의 데이터베이스 설계에 필요한 데이터베이스 관리기법을 소개하고 그 연구사례를 기술한다.

This paper describe database management techniques for scientific applications- especially, medical applications- and case study in U.S.A.

뿐만 아니라 그들의 다양한 배열 형태와 그들간의 관계들을 포함하고 있다. 세째, 데이터의 유래 즉 각각의 데이터들이 서로 다른 것으로부터 추출된 것인지 아닌지에 관한 정보와 또 그들이 언제 어디서 수집되고 생성되는가 등의 정보가 수록되어 있는 데이터셋(dataset)의 수가 많다.

첫번째 특성은 다양한 인터페이스의 개발에 의해서, 두번째 특성은 새롭고 강력한 패스(path)를 제공함으로써 처리할 수 있다. 그리고 세번째 특성은 메타 데이터 베이스를 이용해 데이터셋에 관한 정보를 모델링함으로써 처리할 수 있다.

本稿에서는 이러한 특성을 가지고 있는 과학분야의 데이터 그중에서도 특히 의학분야의 데이터베이스 설계에 필요한 데이터베이스 관리기법과 적용방법을 기술하고 이와 관련해 현재 진행되고 있는 미국대학의 연구사례를 소개한다.

## II. 데이터베이스 관리를 위한 계층적 접근

데이터 관리기법에 대한 접근은 크게 화일계층(file layer), 데이터시스템계층(data system layer), 객체계층(object layer), 응용계층(application layer) 등 4계층으로 구분해 단계적(hierarchical)으로 접근할 수 있다.

화일계층은 각 화일단위로 데이터를 관리하는 계층으로, 이 계층에서의 사용자는 각 화일이

저장되어 있는 장소에 대한 상세한 정보를 알 수 없으며, 다른 기법으로의 전환시 그 적용방법을 명확히 알 수 없는 단점이 있다. 화일계층의 상위가 데이터시스템계층으로, DBMS는 이 계층에서 복잡한 응용들을 기술하는데 필요한 다양한 데이터 모델들을 제공하기 위해 개발됐다. 객체계층은 데이터시스템계층 상위의 단계이다. 관계형 DBMS가 상업적으로 성공을 했지만, 응용과학자들이 그들의 전문분야에 적용하기에는 여전히 부적합한 시스템이라 할 수 있다. 그 주된 이유로는 사용자가 데이터베이스를 이용하기 위해서는 릴레이션(relation)의 정규화 등과 같은 데이터베이스 개념에 대한 상세한 이해가 필요하고, 조인(join) 오퍼레이션 등과 같은 오퍼레이션을 많이 사용해야 하기 때문이다. 따라서 이러한 개



님들은 데이터베이스 비전문가인 응용과학자들이 사용하기에는 적합하지 않다고 할 수 있다.

반면, 객체계층에서의 상호작용은 데이터시스템계층의 상호작용에 비해 몇가지의 잇점을 가지고 있다. 첫째, 사용자가 더욱 간편하고 더욱 직관적인 개념들을 처리할 수 있으며, 둘째 데이터시스템계층에 비해 데이터의 정의와 그에 관련된 질의의 표현이 더욱 간결하다. 그리고 셋째, 객체계층의 각 응용 프로그램들은 그것의 기본이 되는 데이터베이스 시스템에 대해 서로 독립적이다. 그러나, 불행하게도 현재의 객체계층 모델들은 데이터베이스 설계 도구(tools)의 개발에 초점을 맞추고 있고, 더우기 그 초점은 관계형 데이터베이스를 질의하기 위해 데이터베이스 언어를 사용하는 사용자에게 여전히 필요하다.

그러나, 어떤 특정 응용들의 요구는 데이터베이스 객체들을 사용해 직접적으로 기술할 수 없는데, 그것을 제공하는 것이 응용계층이다. 비록 객체계층의 기능이 확장된다고 하더라도 여전히, 예컨대 화학적 또는 생물학적 구조들을 표현하기 위한 세분화된 그래픽 인터페이스를 제공하는데는 어려움이 있다. 그러나, DNA 염기 배열 (DNA sequence)같은 응용계층의 인터페이스들은 객체계층의 지원과 다양한 각 계층 사이의 인터페이스의 제공에 의해 그 세부적인 데이터베이스 기술의 사용에 있어 독립성을 가져다 줄 수 있다. 그리고 이러한 각 계층적 모델들이 데이터베이스 설계에 있어서 단지 한 계층의 모델만을 사용하는 것보다 특정 계층의 지원아래 그 계층의 상위에서 모델이 개발될 때 더욱 효율성을 갖게 되는 것이다.

다음 절에 의학분야에서 사용하는 데이터들을 관리하는데 필요한 데이터베이스 설계에 대해 연구하고 있는 미국대학의 연구사례들을 기술

한다.

### III. 연구사례

#### 1. 3차원 의학적 데이터베이스 QBISM

첫번째 연구사례로 미국 UCLA 의과대학 신경 이미지 실험실(Neuro Imaging Lab)에서 추진하고 있는 3차원 의학적 이미지 데이터베이스의 프로토타입(prototype)인 QBISM(Query By Interactive Spatial Multimedia)을 소개한다. 이 연구의 목표는 시각적 환경(visual environment)상에서 많은 양의 공간적 데이터 사이의 상호작용을 가능하게 하고, 그것을 효율적으로 관리하기 위한 데이터베이스 기술을 확장하는 것이다.

이 연구에서는 2차원 의학적 이미지로부터 생성된 3차원의 공간적 데이터들을 처리하기 위해 논리적이고 물리적인 데이터베이스 설계에 초점을 맞추고 있다. 그것의 구체적 응용은 기능적 두뇌 매핑(Functional Brain Mapping)으로, 두뇌속의 특정 활동과 말하고 팔운동하는 따위의 기능적 행위사이의 상호관계를 알아내는 것이다. 두뇌속의 활동은 일정한 구조에 의해 균일하게 분포되어 있기보다는, 오히려 두뇌 구조층이나 어떤 특정 부분에 균일하지 않게 집중적으로 분포되어 있는 특성을 가지고 있다. 따라서 두뇌 활동의 정확한 위치를 파악하는 것과 그들의 상관관계 그리고 기능적 두뇌의 구조도를 작성하고, 더 나아가 그것을 이해하는 것은 과학적, 교육적 가치 뿐만아니라 진단과 치료 측면에서도 여러가지 임상적 응용의 효과를 가져다 줄 수 있다.

이 연구에서는 이미지상에서 요구되는 질의를 지원하기 위해 IBM社의 Almaden 연구센터에

가져다 줄 수 있다.

이 연구에서는 이미지상에  
서 요구되는 질의를 지원하기  
위해 IBM社의 Almaden 연  
구센터에서 개발한 starburst  
DBMS를 이용하여 프로토타  
입을 만들었고, 이 프로토타  
입은 시각화 패키지인 IBM  
社의 Data Explorer/6000의  
도움을 받고 있다. 사용자는  
조직의 구조와 관심의 형태  
및 범위에 따라 질의가 가능  
하며, 이 시스템이 제공하는  
몇가지 방법 중 하나를 선택  
해 사용할 수 있다. 또한, 사  
용자는 그들의 관점의 변화나  
적절한 조작(예, 표면의 오려  
붙임) 등에 의해 결과를 다시 산출할 수 있고,  
다른 환자 또는 다른 연구로부터 데이터를 선택  
하기 위해 자체적으로 질의를 수정할 수 있다.

이미 기술하였듯이 QBISM은 데이터의 조작  
과 기능적 두뇌 매핑 연구의 시각화를 지원하는  
시스템으로, 상호작용적이고 반복적인 구조상  
에서 복합적인 의학적 이미지에 대한 질의를 제  
공해야 한다는 관점에서 출발했다. 이러한 시각  
화에 대한 연구는 질의 뿐만 아니라 통계적 결  
과의 산출도 가능하게 한다.

이제 기술하였듯이 QBISM은 데이터의 조작  
과 기능적 두뇌 매핑 연구의 시각화를 지원하는  
시스템으로, 상호작용적이고 반복적인 구조상  
에서 복합적인 의학적 이미지에 대한 질의를 제  
공해야 한다는 관점에서 출발했다. 이러한 시각  
화에 대한 연구는 질의 뿐만 아니라 통계적 결  
과의 산출도 가능하게 한다.

QBISM 시스템에서 질의의 표현은, 먼저 의  
학 연구자들이 그들에게 표준이 되는 구조도를  
바탕으로 두뇌의 구조를 표현하고 그 구조를 명  
확히 하는 것으로부터 시작한다.

그리고 연구자는 그 구조도의 위치를 적절히  
조절할 수 있고, 각 구조 중 두드러진 활동들은  
PET 데이터에 근거하여 연구자가 식별할 수  
있는 색깔로 표시된다. 이렇게 표시된 각 활동

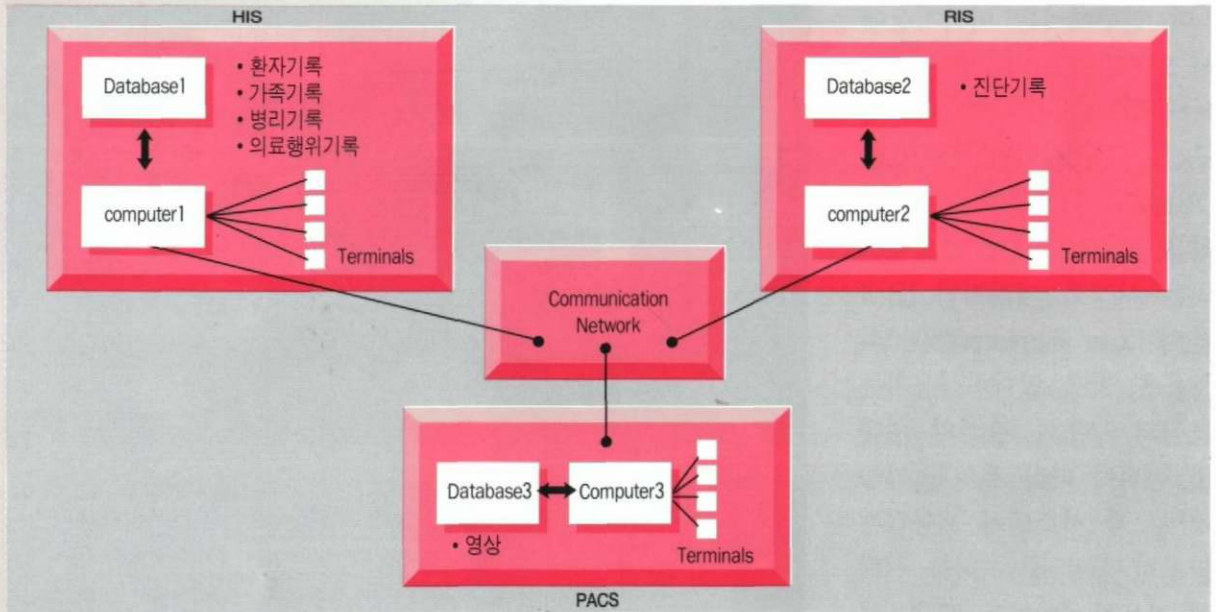


들은 그 형태의 크기를 시각적인 비교를 통해  
나타내 주는 것이고, PET 데이터들은 비정상  
적인 구조에 접근하기 위해 기존의 데이터들과  
통계적으로 비교된다.

이 연구의 데이터베이스는 3차원 의학적 연구  
에 대한 내용과 두뇌의 조직 구조도로 구성되어  
있다. 이 연구는 공간상의 각 위치에서 측정되  
는 어떤 양(예를 들면, 생리학적 활동의 척도  
로서 포도당소모량 등)을 표현하는 3차원 이미  
지 배열이다. 또한 이 연구는 두뇌에 관한 구조  
적 정보를 얻을 수 있는(예, CT, MRI) 그리고  
기능적/생리학적 정보를 얻을 수 있는(예,  
PET, SPECT) 이미지 형태의 분류를 통해  
이루어진다.

연구에 필요한 저장공간은 전형적으로 현재의  
이미지 공간의 분석과 이미지 색깔의 농도를 표  
현하는데 1~30 MByte가 소비되고, 잠재적으  
로는 추가/확장되는 분석과 이미지 농도의 표현

〈그림 1〉 HIS와 RIS 그리고 PACS의 통합환경



어진 환자의 데이터와 거기에 해당하는 구조도 간의 완전한 대응은 이루어지지 않고 있다. 또한, 앞으로 더욱 적절한 질의를 하기 위해 확장성있는 데이터베이스 엔진을 설계해야하고, 사용자 인터페이스로서 보다 나은 질의 시각화 도구의 연결 등이 필요하다. QBISM 시스템 연구의 향후 목표에는 다음이 포함된다.

- 조직 구조도를 위한 다중색인(multi-indexing)기법의 개발과 대규모의 환자집단에 대한 질의를 촉진시키기 위한 환자 연구.
- 공간을 정확히 표현할 수 있는 기법의 개발과 대규모 질의처리를 최적화하기 위한 질의처리전략의 수립.
- 많은 양의 데이터와 연구적인 질의를 지원하기 위한 가설검정 기법들의 결합.
- 이미지 내용에 대한 질의를 제공할 수 있는 통합환경.
- 공간에서 상호작용할 수 있는 메카니즘의 개발(예, 3차원 마우스).

## 2. 의학정보시스템의 통합

두번째 연구사례로 병원 정보 시스템(Hospital Information Systems: HIS)과 방사선 정보 시스템(Radiology Information Systems: RIS) 그리고 영상 정보 시스템인 PACS (Picture Archiving Communication Systems)의 통합을 추진하고 있는 UCLA 전산연구팀, 방사선과학연구팀의 공동연구를 소개한다. 이 연구는 영상의 형태와 환자의 증세사이의 상관관계를 비교하는 것으로, 환자의 성별, 병력, 식이요법, 유전적 구성 등 환자의 프로필에 따라 다양하게 구성되며, 이것은 시간에 따라 변하고 환자의 성숙과정, 질병의 경과, 치료 방법에 따라 다르게 전개된다. 따라서 환자의 진료번호같은 인공적인 키(key)에 의한 것이 아니라 내용에 의해 영상에 접근할 수 있는 능력이 요구된다. 이를 위해서는 의료진단을 위한 영상시각화기법으로 디지털 영상의 컴퓨터지원

진단알고리즘의 개발과 질병상태와 관련된 영상의 형태와 특징을 나타내 주는 데이터의 생성이 필요하다.

이 연구의 목표는 환자기록, 가족기록, 병리기록, 의료행위기록 등의 정보를 기록하고 있는 HIS와 진단기록인 RIS 그리고 영상정보 시스템인 PACS를 통합하는 것으로 <그림 1>의 형태이다.

즉, 내용에 의해 영상에 접근하고, 단계적 처리과정에 따라 영상을 선택하기 위한 연구로서, 골격의 구조, 표현할 수 있는 질의술어의 개발, 형태의 추출과 특성분석을 위한 방법, 질의 결과에 대한 시각화방법 등의 연구이다.

몇몇 질의의 예를 보면 다음과 같다.

- 비정상적인 두뇌의 형태인 뇌수공동에 해당되는 두뇌의 모습을 찾아라.
- 환자 A의 5번째 손가락의 중앙지골 영상을 찾고, 이 뼈 주축의 길이를 구하라.
- 영상 P1의 비정상적인 두뇌의 모습과 유사한 모습의 객체를 영상 P2에서 검색하라.

현재, 이 팀의 연구중에서 네트워크 환경하의 RIS 시스템의 통합은 완료되었으며, HIS와 PACS 시스템간의 통합은 진행중에 있다

### 3. 로런스 버클리 실험실 연구현황

세번째 연구사례는 미국 캘리포니아 주립대학의 로런스 버클리 실험실(Lawrence Berkeley Lab)에서 추진하고 있는 연구로, 이 연구팀이 지금까지 수행해온 과학분야의 연구경험을 바탕으로 관계형 데이터베이스 시스템 상위에서 객체계층을 이용하여 의학적 분야의 응용들을 수행하는 것이 이 연구의 목표이다.

과거 이 연구팀은 과학분야에서의 기능적인

요구들을 확인하고, 그 특성을 기술하는데 힘을 기울였다.

그리고, 그 특성의 기술을 기초로 과학분야에서의 기능적 요구들을 지원하는 개념적 모델과 물리적 데이터 관리기법들을 개발했다. 즉, 이 연구팀은 현재의 데이터, 통계적으로 요약된 데이터, 순서적 데이터 그리고 다차원(multi-dimension) 데이터를 위한 개념적 모델과 이를 효율적으로 사용하기 위한 다양한 물리적 방법들을 개발했다.

현재, 이와 관련된 주연구인 휴먼 게놈 프로젝트(human genome project)는 다음의 두가지를 포함하고 있다. 첫번째는 염색체 정보 시스템(Chromosome Information Systems)으로, 이 시스템의 객체구조를 정의하는 것은 매우 복잡하고 약 35개의 객체를 포함한다. 따라서 이 연구팀은 관계형 DBMS에서 객체구조를 정의하기 위해 데이터베이스 도구를 개발했고, 이를 이용했다.

실험실 정보 관리 시스템(Laboratory Information Management Systems)은 휴먼 게놈 프로젝트와 관련된 또다른 연구로, 이 시스템의 목적은 다양한 순차적 프로토콜(protocol)들을 지원하는 데이터베이스의 신속한 개발을 위해 소프트웨어 도구들을 제공하는 것이다. 이 도구들은 로런스 버클리 실험실에서 개발된 객체-프로토콜 모델이라 불리는 데이터 모델의 기반이 되며, 객체-프로토콜 모델은 실험실 정보 관리 시스템의 데이터베이스에 의해 구조화된 객체와 프로토콜을 모델링하기 위한 구조를 제공하고 있다.

프로토콜들은 종종 또다른 프로토콜들의 순차적 단계를 포함하기 때문에, 객체-프로토콜 모델은 하위 프로토콜 구성요소의 선택과 순서에 의해 각 프로토콜들을 지원한다. 이러한 객체-

프로토콜 모델을 기반으로 하는 데이터베이스 도구들은 전세계적으로 분자생물학 실험실과 게놈 연구센터에서 DBMS를 널리 사용하고 있기 때문에 상업용 관계형 DBMS의 목표가 되고 있다.

이 도구들은 데이터 관리의 독립성과 명료성, 간결성 등을 제공하며, 객체계층 모델은 공간적, 순간적 그리고 배열구조와 같은 복잡한 구조들을 표현하려는 객체들을 추가적으로 제공할 수 있는 미래의 더욱 강력한 데이터 모델에 의해 더 나은 적합성을 얻을 수 있을 것이다. 그리고 객체 모델은 각 하위시스템들(subsystems)과의 인터페이스를 통해 DNA 염기배열구조와 관련된 일련의 체계를 위해 확장할 수 있을 것이다.


그러나, DBMS외부에서 세분화된 구조들을 지원하는 특정 모듈(module)들을 개발하는데는 일정한 기간이 요구된다.

## IV. 맺는말

현재, 대부분의 과학분야에서 사용하는 데이터베이스들은 공간에서의 다차원 데이터구조

또는 DNA 염기배열같은 배열구조들을 표현하기 위해 몇몇 지원들을 필요로 하고 있다.

그러나, 이 다차원적 데이터구조와 객체들은 상업용 관계형 DBMS에 의해 지원되지 않고 있다. 따라서 확장된 관계형 DBMS나 객체지향 데이터베이스 시스템(Object-oriented Database Systems)같은 데이터베이스 관리기법이 필요한데, 이 시스템들을 사용해 데이터의 구조와 객체들을 구체적으로 정의할 수 있다. 그리고 과학분야에의 적용에 있어 가장 중요한 것은 공간적인 데이터를 저장하고 관리하는 시스템을 구축하는 것이므로, 단기적으로는 기존의 상업용 관계형 데이터베이스를 사용하기 위해 그것을 가능하게 하는 기법과 도구들을 개발해야 하며, 장기적으로는 미래의 새로운 기법을 위한 적용방법을 구축하는 것이 필요하다고 할 수 있다.

本稿에서는 이와 관련되어 과학분야에서 사용하는 데이터 그중에서도 특히 의학분야의 데이터베이스 설계에 필요한 데이터베이스 관리 기법과 적용방법을 기술하고 이와 관련해 현재 미국대학에서 진행하고 있는 연구사례를 소개했다. 

### 〈 참고문헌 〉

- [1] M. Arya, W. Cody, C. Faloutsos, J. Richardson and A. Toga, "QBISM: A Prototype 3-D Medical Image Database System", IEEE Computer Society, Vol. 16, No. 1, pp. 38-42, March, 1993.
- [2] A. F. Cardenas, R. K. Taira, W. W. Chu and C. M. Brean, "Integration and Interoperability of a Multimedia Medical Distributed Database System", IEEE Computer Society, Vol. 16, No. 1, pp. 43-47, March, 1993.
- [3] H. Fuchs, M. Levoy and S. M. Pizer, "Interactive Visualization of 3D Medical Data", IEEE Computer, 22, 8, pp. 46-51, Aug., 1989.
- [4] A. D. Narasimhalu and S. Christodoulakis "Multimedia Information Systems: The Unfolding of a Reality", IEEE Computer, 24, 10, pp. 6-8, Oct., 1991.
- [5] A. Shoshani, "A Layered Approach to Scientific Data Management Projects at Lawrence Berkeley Laboratory", IEEE Computer Society, Vol. 16, No. 1, pp. 4-8, March, 1993.