

Construction of Retrieval-Based Medical Database

Yong-Won Shin^{1†}, Bong-Oh Koo² and Byung-Rae Park³

¹Department of Healthcare Management, ²Department of Physical Therapy,

³Department of Radiological Science, Catholic University of Pusan, Busan 609-757, Korea

In the current field of Medical Informatics, the information increases, and changes fast, so we can access the various data types which are ranged from text to image type. A small number of technician digitizes these data to establish database, but it is needed a lot of money and time. Therefore digitization by many end-users confronting data and establishment of searching database is needed to manage increasing information effectively. New data and information are taken fast to provide the quality of care, diagnosis which is the basic work in the medicine. And also It is needed the medical database for purpose of private study and novice education, which is tool to make various data become knowledge. However, current medical database is used and developed only for the purpose of hospital work management. In this study, using text input, file import and object images are digitized to establish database by people who are worked at the medicine field but can not expertise to program. Data are hierarchically constructed and then knowledge is established using a tree type database establishment method. Consequently, we can get data fast and exactly through search, apply it to study as subject-oriented classification, apply it to diagnosis as time-depended reflection of data, and apply it to education and precaution through function of publishing questions and reusability of data.

Key Words: Medical Informatics, Medical database

서 론

최근 의학지식은 매년 폭발적으로 증가하고 있다. 계산에 의하면 1945년부터 1970년 사이에 의학지식은 3배로 증가했으며, 임상지식만을 예로 들어도 1930년부터 1975년 사이에 무려 11배로 증가하였다 (Cho, 1997). 양적인 증가 외에도 화상, 그래프와 같은 이미지 자료에서 소리와 같은 멀티미디어 자료까지 다양한 자료를 접할 수 있게 되었다. 이렇게 자료가 다양화되고 양적으로 증가하면서 의료 정보 및 지식 또한 급속도로 변화하고 있다 (Kim et al., 1995).

이러한 변화에 대처하기 위해서는 양적으로 증가하는 자료 속에서 필요 정보를 신속하게 찾아내어 지식화 할 수 있는 기술이 요구된다. 또한 다양한 유형의 자료를 효율적으로 저장하고 이용하기 위한 방법 또한 매우 중요시되고 있다. 따라서 많은 종류의 정보자원들이 정보의 생산과 입수를 위해 디지털화 되고 있으며 심지어 이러한 디지털화 된 자

료들로만 이루어진 디지털 도서관이 구축되고 있다. 디지털 도서관은 텍스트를 비롯하여 이미지, 소리 등과 같은 멀티미디어 형태의 다양한 자료들을 처리할 수 있고, 시간과 공간에 관계없이 언제, 어느 곳에서나 필요한 자료를 쉽게 검색하고 입수할 수 있는 장점이 있다 (<http://www.lg.or.kr/digilib/lecture>). 그러나 이러한 자료의 디지털화를 위한 어려운 자료 가공 과정은 주로 소수의 전문가에 의해서 이루어지므로 많은 시간과 노력을 필요로 하고 있다. 따라서 프로그램 생산성을 적용한 자동 프로그래밍 기법 등을 이용하여 어려운 가공과정을 해결하고 있으나 미흡한 상황이며 소수의 전문가로 증가하는 많은 양의 자료 처리하는 것은 불가능하다.

그러므로 자료를 직접 접하는 일반 사용자들에 의해서만이 급속도로 증가하는 자료와 다양한 정보를 디지털화 할 수 있다. 즉 복잡한 프로그래밍의 과정 없이도 일반 사용자가 직접 많은 양과 다양한 형태의 자료를 쉽게 디지털화 할 수 있는 방법을 필요로 하고 있다. 또한 디지털화 된 정보의 공유와 유통을 위해서는 HTML (Hyper text markup language)이나 상위의 기술인 SGML (Standard generalized markup language)과 같은 전자 문서 표준안을 이용할 필요성이 있다 (Jeong et al., 1998).

의료 분야는 진료, 연구, 예방, 교육과 같은 다양한 목적으로 자료와 지식을 이용하므로 일반적인 자료와 지식의 구조

*논문 접수: 2004년 9월 7일

수정재접수: 2004년 12월 9일

†교신저자: 신용원, (우) 609-757 부산광역시 금정구 부곡3동 9번지, 부산가톨릭대학교 보건과학대학 병원경영학과
Tel: 051-510-0615, Fax: 051-510-0618
e-mail: kevin@cup.ac.kr

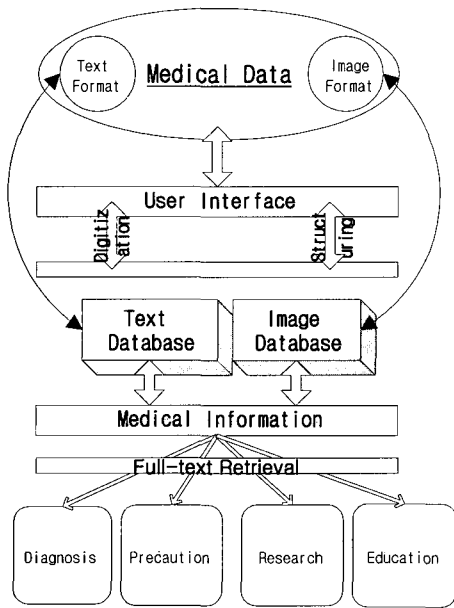


Fig. 1. Medical database model.

와는 다른 모습을 보이고 있다. 시변성을 반영하는 진료 자료의 특징과 연구 목적을 위해 주제별로 자료를 정리하는 등의 여러 목적으로 자료와 지식을 구조화해야만 한다. 하지만 이러한 자료와 지식의 구조화는 진료시의 의사결정과 같은 정보처리 중심보다는 자료를 직접 접하는 현장에서의 데이터를 중심으로 이루어져야만 한다. 그러므로 실제 여러 가지 자료를 접하는 의사와 같은 의료 현장의 종사자들이 직접 자료를 구조화하고 쉽게 디지털화하여 저장할 수 있는 데이터베이스가 필요하게 되었다.

본 연구에서는 다양하고 많은 자료를 컴퓨터 사용에 익숙하지 않은 일반 의료인들도 쉽게 디지털화하고 의료의 여러 목적에 이용할 수 있는 의료용 데이터베이스를 구축하고자 한다. 구축되는 데이터베이스는 전문과 같은 일차 자료와 다양한 진료 자료를 의사와 같은 일반 의료인들이 직접 디지털화하여 저장할 수 있다. 뿐만 아니라 의료 분야의 데이터를 중심으로 체계적으로 분류하여 구조화함으로써 의료분야의 다양한 목적에 이용할 수 있을 것이다.

연구 방법

본 논문의 검색중심의 의료용 데이터베이스 구축을 위해 의료분야의 다양한 자료를 텍스트 형식과 이미지 형식으로 분리하여 디지털화하고 이미지와 텍스트를 각각 관계형 데이터베이스인 Paradox에 나누어 저장한다. 생산성이 높은 4세대 언어인 Delphi를 이용하여 쉬운 사용자 인터페이스를 구성함으로써 텍스트 형식의 자료는 검색이 가능한 아스키 형식

로 직접 입력한다.

이미지 형식은 재사용을 위해 표준화된 형태로, 그리고 검색을 위해 내용을 대표할 수 있는 단어와 함께 저장하며 스캐너 등을 이용하여 생성된 이미지 파일을 직접 입력한다. Fig. 1에서 보는 모델과 같이 데이터를 어려운 가공과정 없이 사용자 인터페이스를 통해서 구조화, 디지털화하여 지식화된 의료정보를 전문 검색을 통해 검색한다. 그리고 쉬운 검색과 유용한 정보의 종류 및 형태를 파악하고, 진료, 예방, 교육, 연구목적에 필요한 자료의 구조화 방식을 조사하여 일반 의료인들도 쉽게 구축 가능한 의료용 데이터베이스를 제시한다. 본 연구에서 구축된 데이터베이스의 특성과 기능은 임상 미생물 단행본의 전문을 직접 입력하여 각 실행 예를 통해 상세히 설명한다.

관련 연구

1. 의료용데이터베이스

의료분야에서 정보는 생명을 다루는 의료인에게는 필수 도구이기 때문에 데이터베이스와 컴퓨터의 이용은 일찍부터 연구되어 왔다 (Cho, 1997). 의료용 데이터베이스는 목적에 따라서 여러 가지로 나누어진다. 여러 목적의 의료용 데이터베이스 중에서 의료 관계종사자들이 업무에서 자주 접하는 것은 병원용 데이터베이스로서 병원에서 행하는 진료에서 얻어지는 각종 데이터를 데이터베이스의 관점에서 정리한 것이다. 따라서 현재 병원용 데이터베이스는 병원에서 발생하는 많은 양의 데이터를 처리할 수 있고, 데이터의 신뢰성을 증가시킬 수 있는 범용의 관계형 데이터베이스를 이용하고 있으며 주로 병원의 각 업무 정보를 통일적으로 취급하는 병원 관리의 목적으로만 쓰이고 있다. 그러므로 병원용 데이터베이스는 진료의 목적보다는 신속한 환자관리와 원무 행정을 중심으로 하고 있다. 그리고 의료 자료의 처리가 큰 비중을 차지하고 있기 때문에, 데이터 검색은 환자의 성명 또는 번호와 같은 극히 제한된 Key 항목에 의해서만 행해지고 있다 (Kim et al., 1995).

그러나 진료시에는 해당 환자의 나이와 같은 기본 정보뿐만 아니라 환자의 검사결과나 병력 등과 같은 다양한 항목을 자세히 검색하고 인지할 필요성이 있다. 또한 환자 데이터는 관계형 데이터베이스에서 다룰 수 없는 구조적인 특징들이 있다 (Kim et al., 1995). 첫째, 환자 데이터는 검사 결과와 같은 여러 가지 데이터들의 단위 데이터 요소들로 이루어진다. 예를 들면 성명, 연령과 같은 기본 정보와 검사항목, 결과, 투약 등이 한 사람의 환자정보를 구성하는 요소로 생각된다. 따라서 환자정보를 구성하는 여러 데이터들을 해당 환자를 중심으로 계층적으로 수집, 정리해야 한다. 둘째, 환자 데이터는 시변적이다. 즉 검사결과, 투약정보 등과 같은 여러 데

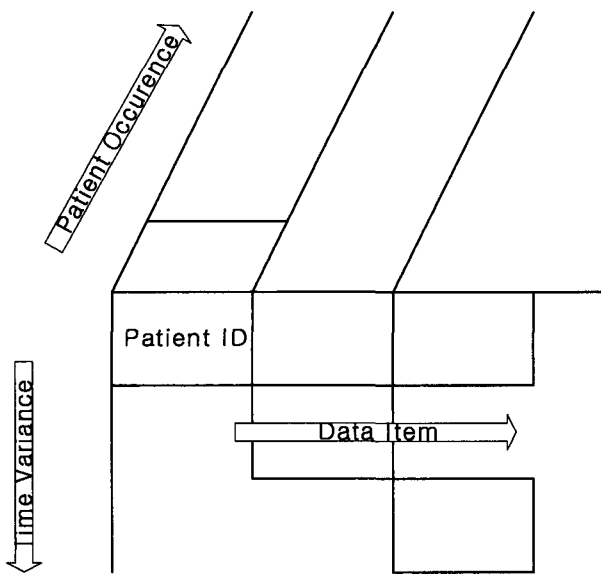


Fig. 2. Feature of patient data structure.

이터들은 진료의 횟수와 시간에 따라 반복적으로 발생하므로 시변성을 반영하게 된다. 예를 들면 어느 날 환자가 진료를 받은 경우 진료가 1회로 끝나는 경우는 드물고, 1주일 짝게는 하루 후에 다시 진료를 받게 된다. 입원 환자에 대해서는 이러한 현상이 매일 발생하게 된다. 물론 데이터가 시간적으로 변화하고 발생하는 것은 다른 분야에서도 흔한 경우이나, 대부분의 경우 갱신된 최후의 기록만이 남는 것이 일반적이다. 하지만 환자를 정확하게 진단하기 위해서는 과거의 증상이나 처방과 같은 병력 정보가 꼭 필요하기 때문에 환자 및 의료분야의 데이터는 시간에 따라 발생한 데이터를 모두 보관해야 한다.

Fig. 2는 환자 데이터의 특징을 구조화한 그림으로서 환자에 대한 여러 가지 데이터 요소와 시변성이 각 환자마다 발생하는 구조적 특징을 볼 수 있다. 하지만 병원에서 쓰이는 관계형 데이터베이스는 복잡한 구조의 실세계를 단순히 애트리뷰트 (Attribute) 값의 릴레이션 집합으로만 모델링하므로 개체 의미나, 개체간의 관련성을 데이터베이스에 직접 유지하지 못하고 (Kim, 1996) 시변성과 계층적인 구조를 반영할 수 없다. 따라서 현재의 데이터베이스는 단지 병원의 관리를 위한 것으로, 데이터를 검색하는 등의 진료용으로는 사용하지 못하고 있으므로 환자 데이터를 전문 검색을 이용하여 자세히 검색하고 환자에 관한 데이터를 계층적으로 저장할 수 있는 데이터베이스가 필요하다.

의사를 비롯한 의료인들은 진료뿐만 아니라 보건 교육과 홍보를 통한 예방, 초보자를 위한 교육, 의료인들의 지식과 기술 발전을 위한 연구 등을 직무로 하고 있다. 먼저 지역 주민의 보건 교육과 전염병 홍보 등의 예방의 목적을 달성하기

위해서는 데이터베이스 내에서 필요한 데이터를 검색하고 검색된 데이터를 쉽게 가공하여 교육용 교재 등과 같은 정보화된 산물을 창출해냄으로써 합리적인 선택 및 결정을 유도해야만 한다.

초보자의 교육에 있어서는 개인적으로 자신의 능력과 적성에 기초하는 주제적 학습 (Self-learning)이 전개되어야 한다 (Kim, 1994). 교육에 필요한 다양한 자료를 검색하고 교육공학의 개념적 수업모형 (Concept model)을 적용하여 수업모형에 사용되는 자료들을 제작해야 한다. 수업활동을 흥미롭게 구성하며 학생들의 학습의욕을 고취시키기 위하여 난이도 수준을 고려하고, 학습상태를 알아보기 위한 적절한 평가를 위한 절차 및 도구가 필요하다. 한편 컴퓨터를 이용한 CAI (Computer aided instruction)는 상호작용적 교수공학의 대표적인 경우로서 피교육자의 수준별 학습 등의 다양한 학습 모델을 제시할 수 있다 (Kim, 1994).

의료인들의 지식과 기술 발전을 위한 연구 목적을 위해서는 데이터 웨어하우스 관점의 주제 지향적 (Subject-oriented)인 데이터의 구조화 (Inmon et al., 1996)가 필요하다. 즉 환자 데이터가 저장된 병원용 데이터베이스에서 학술적으로 중요한 기록만을 추출하여 연구의 주제 중심으로 재구성함으로써 의료인의 연구과정을 효율적으로 수행 할 수 있어야 한다. 또한 다양한 문헌 자료 중에서 개인적인 수집을 통해 얻어진 필요 자료들을 각 연구의 주제를 중심으로 구조화 할 수 있어야 한다. 따라서 단지 제한적인 검색만 가능한 관리 목적의 데이터베이스가 아닌 다양하고 방대한 의료 자료를 의료 분야의 각 목적에 맞게 자세히 검색하고 구조화하여 지식화 할 수 있는 의료용 데이터베이스를 구축해야 한다.

2. 의료데이터검색

현재 의사를 비롯한 의료관계 종사자들은 대학병원의 특수한 경우를 제외하면 학술회의나 학술문헌 혹은 정기간행물에서 새로운 지식과 정보를 얻고 있다. 이러한 문헌 정보가 양적으로 팽창하게 되면서 예전과는 다르게 사람들에게 질의하는 방법만으로는 원하는 내용을 빠르고 정확하게 찾을 수 없게 되었다 (Park, 1996). 따라서 다양한 문헌 정보를 신속하고 정확하게 찾기 위해서는 효율적인 정보 검색이 필요하다. 정보검색이란 수집된 정보 또는 정보자료의 내용을 분석한 뒤 적절히 가공하여 축적해 놓은 정보파일로부터 이용자의 정보요구에 적합한 정보를 탐색하여 찾아내는 일련의 과정을 의미한다.

최근에는 많은 비용과 인력의 부담을 줄일 수 있는 소프트웨어 공학의 프로그램 생산성을 적용한 여러 가지 기술로 대량의 자료를 처리하려하고 있다. 이러한 프로그램 생산성을 이용한 기술로 전문 검색 데이터베이스 구축시의 복잡한 과정과 데이터 입력시의 어려움을 해결하고 있다. 하지만 폭

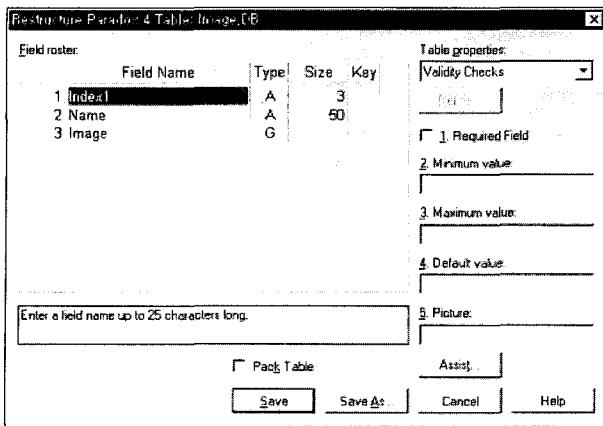


Fig. 3. Image database structure of medical database for full-text retrieval.

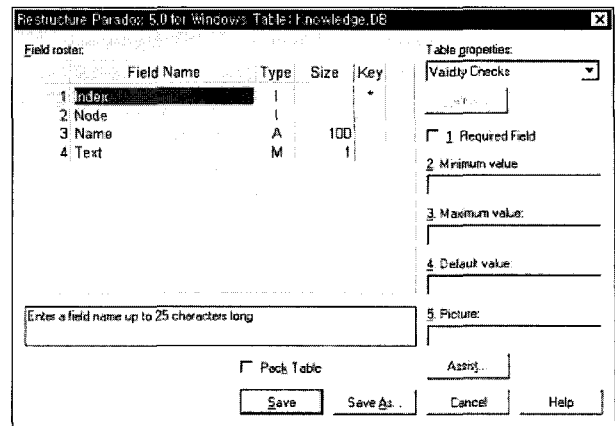


Fig. 4. Text database structure of medical database for full-text retrieval.

발적으로 증가하는 많은 양의 정보를 소수의 전문가만으로 자료를 디지털화하고 입력하는 데는 많은 시간과 노력을 필요로 하고 있다.

따라서 정보의 양적 증가에 효과적으로 대처하기 위해서는 정보를 직접 접하는 일반 사용자들에 의한 자료의 디지털화와 데이터 입력이 필요하다. 프로그래밍에 대해 전혀 모르는 일반 사용자를 위해서는 Non-programing 형태의 End-user computing 환경이 필요하며 텍스트 형식은 전문 검색이 가능한 형식으로 입력되어야 한다. 텍스트 외에 전문에 포함되거나 참고적으로 첨가된 이미지의 검색을 위한 방법이 필요하며 다른 응용 프로그램과의 상호 교환 할 수 있는 형식으로 저장해야한다. 검색은 다양한 검색기법과 연산자 및 명령어를 제공하고 있어 이용 시 검색 전문가의 도움을 받아야 하므로 컴퓨터에 익숙하지 않은 일반 사용자들도 검색 전문가의 도움 없이 다양한 검색 기법과 연산자를 이용할 수 있는 환경이 필요하다.

의료분야의 학술자료와 논문과 같은 문헌 정보와 같은 여러 데이터를 효율적으로 검색하고 이용하기 위해서는 전문적인 프로그래머가 아닌 일반 의료인들이 직접 필요 데이터를 디지털화하여 데이터베이스에 저장하고, 저장된 데이터를 검색 전문가의 도움 없이도 정확하게 검색하고 손쉽게 입수할 수 있는 전문 검색환경을 구축해야 한다.

3. 데이터베이스 구축

의료 분야에서는 의학 연구와 적용에 필요한 양질의 정보를 습득해야 한다는 인식이 증가하면서 의료정보의 저장, 관리, 검색을 매우 중요하게 생각하고 있다 (Kim et al., 1995). 따라서 의료 분야의 진료, 예방, 교육, 연구와 같은 다양한 목적을 만족할 수 있는 의료용 데이터베이스를 구축함으로써 각종 학술잡지와 같은 문헌을 프로그래밍에 대해 전혀 모르

는 일반 의료인들도 쉽게 디지털화하고 전문 검색을 통해 이를 정확하고 효율적으로 검색할 수 있다. 의료용 데이터베이스는 의료 분야의 가장 큰 목적인 진료를 위해 환자 데이터의 시변성을 반영하고 연구 목적의 주제 지향적인 구조를 만족할 수 있도록 데이터를 계층적으로 구조화하여 저장해야 한다.

전문 검색 데이터베이스 구축을 위해서는 먼저 필요자료를 수집하고 디지털화 하여 저장한다. 디지털화는 어려운 프로그래밍의 과정 없이 텍스트 형식은 검색이 가능한 기계 가독형인 아스키 형식으로 직접 입력한다. 텍스트를 다루는데 이용하는 글과 같은 워드프로세서는 모든 기능이 버튼과 메뉴의 GUI 형식으로 이루어져 있어 많은 사용자들이 이를 이용하고 있다. 따라서 일반 사용자들이 텍스트 형식의 자료를 다루는데 가장 많이 이용하는 워드프로세서 형태의 사용자 인터페이스를 구축함으로써 쉽게 전문 자료 및 다양한 텍스트 형식을 디지털화 하고, 입력된 전문은 곧 바로 데이터베이스에 저장한다. 아스키 형식으로 저장된 텍스트는 다양한 형식의 텍스트를 표현 할 수 있어 최근 많이 이용되고 있는 RTF (Rich text format)와 인터넷의 전자 문서인 HTML 과 같은 다른 텍스트형식으로 쉽게 변환 가능한 장점이 있다 (Jeong et al., 1998). 또한 이미 디지털화 된 텍스트형식의 자료는 파일 가져오기 기능을 이용하여 데이터베이스에 쉽게 입력하도록 한다.

이미지 형식은 재사용을 위해 표준화된 비트맵 (Bitmap) 형태로 저장하며 검색을 위해 이미지의 의미를 표현할 수 있는 단어와 함께 저장한다. 이미지 형식의 자료 또한 파일 가져오기 및 OLE 객체 기술을 이용하며 빠른 검색과 저장을 위해 텍스트형식과 이미지 형식을 나누어서 데이터베이스에 저장한다. Fig. 3과 Fig. 4는 텍스트와 이미지를 저장하는 데이터베이스의 구조로서 각 데이터베이스의 필드명, 형식, 크기를 볼 수 있다.

검색용 데이터베이스 구축 시 전문은 제목과 본문으로 나누어 입력하여 제목과 본문의 형식으로 이루어진 인쇄본의 형식을 그대로 이용하며 입력되어진 자료의 인덱스 생성과 업데이트와 같은 복잡한 과정은 자동화한다. 일반 사용자들은 인덱스 생성과 같은 복잡한 작업을 수행하지 못할 뿐 아니라 데이터 추가마다 인덱스의 업데이트 과정을 매번 사용자가 수행하는 것은 비효율적이기 때문이다. 따라서 전문 자료 중 제목과 같이 짧은 텍스트는 데이터베이스의 자동화된 인덱스를 이용하여 자동화하고, 본문과 같은 많은 양의 텍스트는 SQL (Structured query language)과 같은 데이터 처리 언어를 이용하여 전문의 모든 단어를 검색한다. 이러한 데이터베이스의 인덱스는 중요 단어들의 위치를 표시된 리스트를 포함하고 있어 데이터를 쉽게 검색한다 (<http://www.snu.ac.kr/~course/DB/ch2/ch2.3>). 또한 새로운 데이터가 추가되거나 삭제될 경우 인덱스도 이에 맞게 자동적으로 변경된다.

데이터베이스에 저장된 데이터의 전문 검색 및 다양한 검색은 전문 검색사의 도움 없이 이용하도록 한다. 전문 검색 시 많은 명령어와 다양한 연산자의 이용은 검색의 정확성을 높일 수 있으나 사용자는 모든 명령어와 연산자를 기억해야만 하는 단점을 가진다. 따라서 윈도우의 메뉴 방식을 이용하여 불리언 (Boolean) 검색의 연산자를 부호화 하지 않고 AND, OR를 직접 보면서 일반 사용자도 쉽게 입력한다. 불리언 검색은 불 대수 (Boolean algebra)를 이용하여 탐색문을 만족시키는 정보를 검색하는 것으로 탐색어간의 관계는 AND, OR 등으로 표현되며 실제로는 *(AND), +(OR) 등의 부호로서 표현되어 효과적인 검색과 비교적 쉬운 질의어로 널리 사용되고 있다 (Jeong, 1993). 하지만 불리언 검색은 탐색어의 상대적인 중요도를 나타내지 못하는 단점이 있다 (Lee et al., 1996).

한번 검색이 실행된 후에는 필터링 기능을 이용하여 2차 세부 검색을 실시하며 SQL의 부분매치 연산자 LIKE를 이용하여 부분 검색을 실행하도록하며 각 탐색어의 출현 위치를 제목과 본문으로 제한하는 제한 탐색이 가능하도록 한다. 이

러한 검색은 모두 단지 탐색어를 입력하고 메뉴에서 연산자를 선택하기만 하면 검색이 실행된다. 검색 시에는 사용자에게 의해 입력된 탐색어와 메뉴의 연산자를 변수로 처리, 번역하여 SQL의 쿼리로 변환하는 방식을 이용함으로써 Non-programming을 구현한다.

SQL은 데이터베이스의 표준 언어로서 데이터베이스에서 데이터 정의, 조작, 제어 기능을 제공하고 있다. SQL문의 가장 일반적인 명령은 SELECT 문으로서 데이터를 검색하고 이 데이터를 사용자에게 되돌리는 역할을 한다. SQL에서는 다양한 내장 함수와 연산자를 제공하고 있어 다양한 검색을 행할 수 있다 (Perkins et al., 1996).

다음 예제는 구축에 이용되어진 불리언 검색을 위한 구문으로서 불리언 연산자를 조건문에 첨가하고 부분매치를 LIKE 문과 부분매치 연산자 %를 이용한 구문이다.

```
SELECT*
FROM 전문
WHERE 제목 LIKE '데이터베이스'
AND 전문내용 LIKE '의료용%'
OR 전문내용 LIKE '전문 검색'
```

이와 같이 SQL을 이용한 전문 검색은 다양한 내장 함수와 연산자를 이용하여 여러 가지 검색 기법을 구현 할 수 있는 장점이 있다. 전문을 검색한 결과는 사용자가 쉽게 인지할 수 있도록 한다. 정확한 검색이 이루어졌다고 해도 사용자가 입력한 탐색어의 위치를 알기 위해 전문을 다시 읽는 것은 비효율적이기 때문이다. 따라서 검색 후 전문의 프레젠테이션 모드에서는 탐색어를 기존 글과 다른 색으로 표시함으로써 사용자가 쉽게 탐색어를 발견한다.

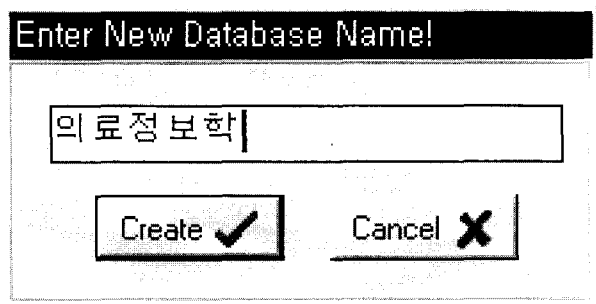


Fig. 5. Window for new database.

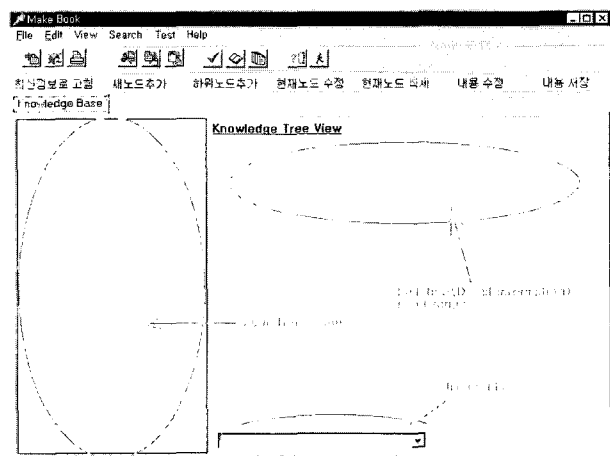


Fig. 6. Intial window.

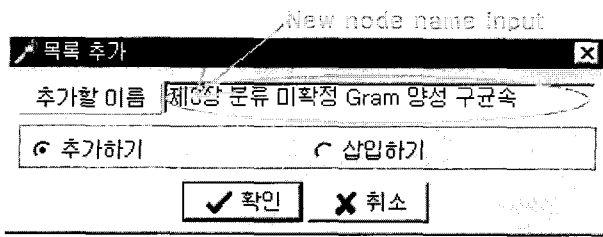


Fig. 7. Input form of appending tree node.



Fig. 8. Tree which is appended node.

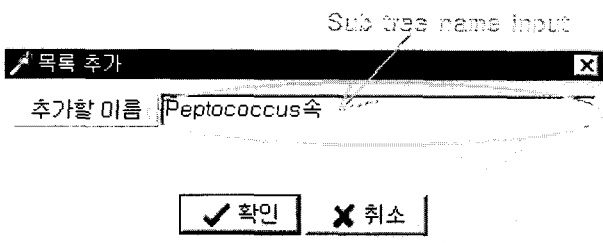


Fig. 9. Append sub node.

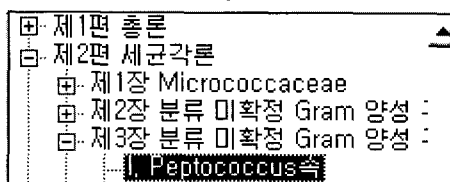


Fig. 10. Tree which is appended sub node.

결 론

본 연구에서는 워드프로세서 형태의 인터페이스를 구성하여 어려운 프로그래밍이 없이도 쉽게 전문과 같은 텍스트를 입력하고, 전문에 포함된 이미지는 의미를 표현할 수 있는 단어와 함께 표준화된 비트맵 형식으로 저장하였다. 따라서 컴퓨터 사용에 익숙하지 않은 일반 사용자에게 의해 전문 검색 데이터베이스를 구축할 수 있도록 하였다. 자동화된 인덱스와 SQL을 이용하여 전문 검색을 구축하고 메뉴 방식을 이용하여 전문가의 도움 없이도 전문을 검색하도록 하였다. 그리고 진료료 위해 환자 데이터를 계층화된 구조로 저장하고, 예방

과 교육을 위해 데이터를 재사용하며, 각종 문헌자료를 주제 지향적으로 구조화하여 연구목적에 이용하도록 하였다.

본 데이터베이스는 전문을 저장하고 검색하기 위한 것이므로 데이터베이스의 생성 시에는 기본적으로 문헌 자료 등의 제목과 본문을 분리하고, 자료의 계층적 구조화를 위한 트리의 노드번호와 레벨을 저장할 수 있도록 데이터베이스가 생성된다. 또한 데이터베이스의 생성시 이미지 데이터를 따로 저장하기 위해서 쌍으로 데이터베이스를 생성하게 된다. Fig. 5에서 보는 것과 같이 새로운 데이터베이스는 개발된 데이터베이스 생성 모듈을 이용하여 자료들의 주제에 맞는 이름을 입력하면 자동으로 생성된다.

Fig. 6은 데이터베이스 생성후의 초기 화면으로서 모든 기능이 메뉴와 버튼으로 조작되는 GUI 환경을 이용하여 사용자가 편리하게 이용하도록 하였으며 창의 왼쪽 영역에는 트리 형태의 전체적인 자료의 구조를, 오른쪽 영역에서는 워드프로세서형식으로 상세 정보를 입력한다.

다음은 임상 미생물학에 관한 단행본을 입력 한 예이다. 단행본은 제목과 전문으로 크게 두 부분으로 이루어져 있으며 여러 가지 도표와 이미지를 포함하고 있다.

먼저 데이터베이스 생성 후 전문을 입력하기 위해 트리에 새로운 노드를 추가한다. Fig. 7과 같이 트리 노드의 이름을 입력하면 새로운 노드가 추가되는데 트리 노드 이름은 각 Subject의 제목을 입력한다. 노드의 추가는 동등레벨에서 하위 위치를 갖는 추가 모드와 상위 위치를 가지는 삽입 모드가 있다. Fig. 8은 Fig. 7에서 입력한 이름의 새로운 노드가 추가된 그림을 볼 수 있다.

단행본은 큰 주제의 제목에 계층적으로 여러 개의 소제목이 있다. 따라서 상위 노드에 대해 여러 개의 하위 노드가 있을 수 있다. 하위 레벨의 노드이름 역시 해당 전문의 제목을 Fig. 9와 같이 입력하면 해당 노드의 하위 트리가 생성된다. Fig. 10은 트리에 추가된 하위 레벨의 노드를 볼 수 있다.

트리는 여러 레벨까지 확장 가능하여 단순 트리 구조에서 여러 가지 복잡한 트리 구조까지 다양하게 표현 할 수 있다. 3단계의 레벨로 구성되어진 트리를 Fig. 11에서 볼 수 있다.

노드 삭제와 노드 이름 변경 등의 여러 가지 기능을 간단한 버튼으로 이용하여 일반 사용자들도 쉽게 계층적 구조의 트리를 구성 할 수 있게 하였다. 노드 위치와 레벨은 미리 작성된 함수에 의해서 추가시나 삭제 시 또는 변경 시에 자동적으로 그 정보가 데이터베이스에 저장되어 전문 데이터와 연결된다. 전체적인 트리를 구성하고 난 후 또는 하나의 노드를 생성한 후 곧 바로 전문 데이터를 입력한다. 데이터의 입력은 Fig. 12에서 보는 것과 같이 워드프로세서 인터페이스를 이용하여 전문을 데이터베이스에 직접 입력하거나 이미 존재하는 텍스트 파일을 가져오기 하여 저장하는 두 가지 방법을 이용할 수 있다.

임상 미생물에 관한 단행본에는 많은 텍스트뿐만 아니라

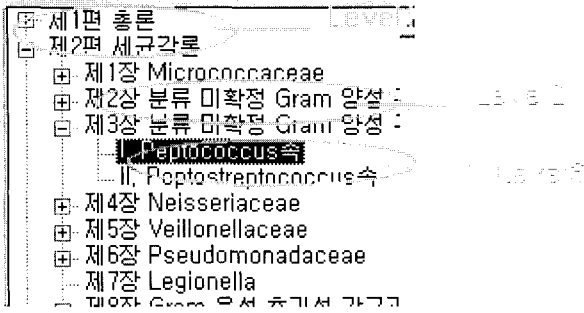


Fig. 11. Multi level hierarchical tree.



Fig. 15. Screen after inputting full-text of microbiology book.

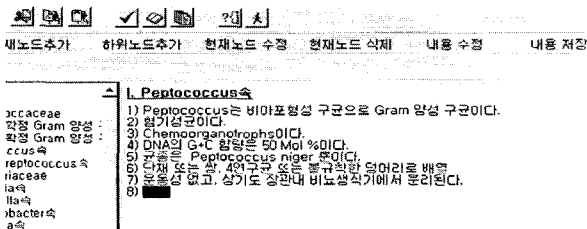


Fig. 12. Input full-text with word processor interface.

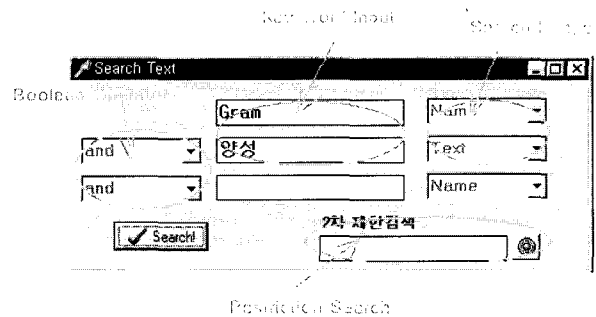


Fig. 16. Search the full-text using boolean operator and restriction search.

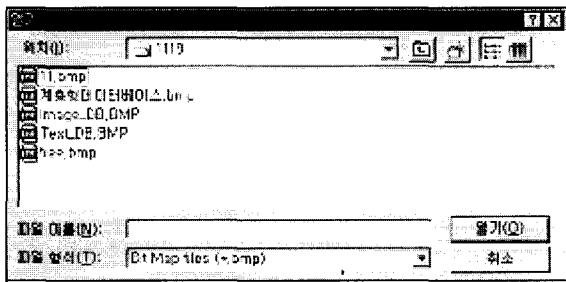


Fig. 13. Dialog of importing image.

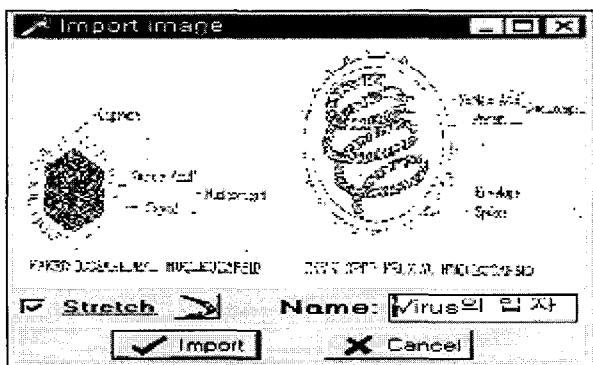


Fig. 14. Save the image with import file.

다양한 이미지와 도표가 존재한다. 이러한 데이터는 같이 스캐너 등을 이용하여 미리 제작된 파일 가져오기를 하여 데

이터베이스에 저장한다. Fig. 13은 이미지 가져오기 대화창으로 입력하고자 하는 비트맵 파일을 선택 한다.

파일 가져오기를 통하여 입력된 의미를 대표하는 단어와 함께 Fig. 14와 같이 저장한다. 입력된 이미지 데이터는 전문 데이터처럼 트리 노드의 위치와 레벨 정보를 함께 저장하므로 각 노드로 이동시 노드에서 볼 수 있는 이미지의 목록을 ListBox에서 볼 수 있으며 이미지의 이름을 선택함으로써 저장된 이미지를 볼 수 있다.

다음 Fig. 15는 임상 미생물 단행본의 전문을 입력한 후의 화면이다. 계층적 구조의 트리를 화면 왼쪽에서 볼 수 있으며 각 트리의 노드를 선택하면 해당 노드의 전문과 이미지의 목록을 화면의 오른쪽 영역에서 볼 수 있다.

데이터베이스에 저장된 자료는 키워드 입력 방식으로 제목과 본문을 구분하여 검색할 수 있는 제한 검색과 AND와 OR를 이용한 불리언 검색을 윈도우의 메뉴를 이용하여 쉽게 검색 할 수 있다. 입력된 탐색어는 부분 매치 연산자를 적용하여 정확한 검색이 이루어지도록 하였다. 또한 검색결과에 대한 검색을 다시 시행하는 세부 검색 기능을 수행 할 수 있다. Fig. 16은 Gram 양성 균에 대한 자료를 찾기 위해 제목에

Name
제 3장 분류 미확정 Gram 양성 구균속
제 14장 Gram 양성 아포형성 간균과 구균군
제 15장 정규 무아포 Gram 양성 간균
제 16장 불규칙한 Gram 양성 무아포 간균
제 21장 호기성, 미호기성, 운동성, 라선형, Vibrioid Gram 음성균

Fig. 17. Hit list of retrieval.

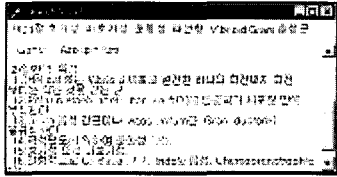


Fig. 18. Search view.

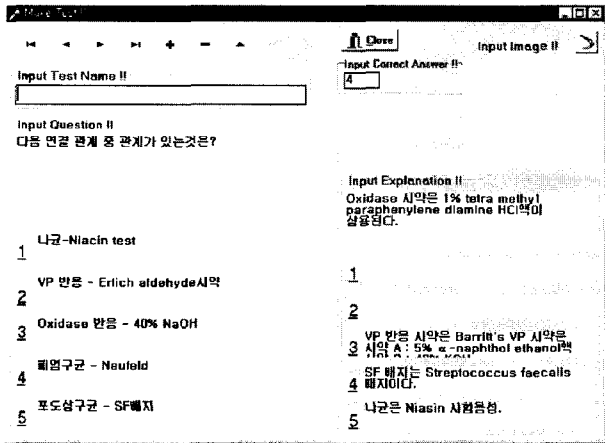


Fig. 19. Module of making test.

"Gram"이라는 단어와 본문 중에 "양성"이라는 단어를 검색하는 그림이다. "Gram"은 데이터베이스에서 만들어진 인덱스를 통하여, "양성"은 전문 검색을 통하여 검색하게 된다.

검색되어진 결과는 Fig. 17에서처럼 간단한 제목과 함께 리스트형태로 나타난다. 사용자는 제목과 함께 리스트를 보면서 적절한 결과를 찾아내며 2차 제한 검색을 통해서 세부 검색을 실시 할 수 있다.

검색결과 리스트에서 필요 자료를 선택하면 해당 자료의 전문을 볼 수 있으며 전문에 출현한 탐색어는 일반적인 텍스트와 다른 색으로 구별함으로써 사용자가 쉽게 탐색어를 발견할 수 있도록 하였다. Fig. 18에서는 앞에서 입력했던 탐색어인 "Gram"과 "양성"이라는 단어를 전문에서 쉽게 볼 수 있다.

Fig. 19와 같이 구축된 데이터베이스의 전문을 이용하여 위한 간단한 Test를 만들 수 있는 모듈은 보건 교육을 통한 예

검사명	화학 검사
일반정보	T-Protein : 5.5
Electrocardiographic Record	Albumin : 6.0
화학 검사	Creatinine : 0.8
응급 화학 검사	Sodium : 150
2002/10/05	Potassium : 4.0
검사명	Chloride : 111
2002/10/05	Amylase : 55
검사기	
일반정보	
일반 혈액 검사	
2002/10/15	
2002/10/0	
2002/11/15	

Fig. 20. Patient data which are hierarchically constructed.

방과 초보자 교육을 위한 목적으로 이용한다.

진료 시 발생하는 다양한 자료를 환자를 중심으로 트리 구조를 이용하여 Fig. 20과 같이 계층화하였다. 환자에 대한 기본 정보에서 일반검사, 심전도 그래프 등을 각 환자를 중심으로 저장하였다. 각 검사 결과와 진료 결과는 노드에 날짜 정보를 입력함으로써 시변성을 반영하도록 하였다.

고찰

전문 검색 데이터베이스 구축시의 어려운 전문 입력 방식은 전문가들에 의해 개발된 프로그램 등으로 해결하고 있으나 소수의 전문가만으로 증가하는 전문 데이터를 입력하는 데는 많은 시간이 소요된다. 따라서 본 연구는 정보의 양적 증가에 대한 대처는 정보를 직접 접하는 일반 사용자들에 의한 데이터의 입력에 의해서만 해결 가능한 점에 착안하였다.

전문 (全文) 검색 데이터베이스 구축을 위해 텍스트를 다루는데 가장 많이 이용하는 글과 같은 워드프로세서 인터페이스를 이용하여 일반 사용자들도 쉽게 데이터베이스에 직접 입력하도록 하였다. 직접 입력된 텍스트는 전문 검색이 가능할 뿐 아니라 텍스트에 다양한 형식을 표현할 수 있는 RTF 형식과 HTML 등으로 쉽게 변환이 가능하도록 하였다. 전문 검색 시 사용하던 복잡한 명령어와 연산자를 GUI형식의 메뉴와 키워드 입력에 따라 자동으로 SQL 구문이 생성되는 non-program 형식을 구현함으로써 컴퓨터 사용이 익숙하지 않은 일반 의료인들도 전문 검색을 실행할 수 있게 하였다. SQL 과 키워드를 이용한 전문 검색 방식은 전문의 모든 단어를 검색 할 수 있으며 적합률이 뛰어난 장점이 있다.

저장되어 있는 이미지의 검색은 현재 패턴 인식 기법 등을 이용하고 있으나 아직은 미흡한 상황이다. 한편 이미지의 내용을 대표할 수 있는 단어는 이미지의 구체적인 의미를 표현 할 수 있다. 따라서 본 연구에서는 이미지에 관련된 단어의 입력, 검색을 통해 이미지 검색에 도움이 되도록 하였다. 외부적으로 나타나는 사용자 인터페이스는 트리 형식으로 지식구조를 명료하게 나타내고, 내부적으로 데이터의 저장

에 있어서는 다량의 데이터를 관리 할 수 있는 관계형 데이터베이스를 이용하였다. 따라서 진료시나 연구 시에 기존의 의료용 데이터베이스에서 표현 할 수 없었던 계층적 지식 구조를 관계형 데이터베이스에 저장하여 표현 할 수 있도록 하였다.

본 논문에서 사용된 SQL과 키워드에 기반한 전문 검색은 동음이의어를 구별하지 못하며, 넓은 영역에서는 키워드 입력 방식의 적합율이 저하되는 단점이 있으며 불리안 검색 방식 외에 상대적 가중치 표시를 위한 가중치 검색 등이 필요하다. OCR (optical character recognition). 기술 발전에 따른 인식율 향상을 텍스트 입력에 이용함으로써 입력의 효율성을 증대시킬 수 있을 것이다.

앞으로 연구에는 생물학적 이미지뿐만 아니라 소리와 동영상과 같은 다양한 멀티미디어 형식의 데이터도 다룰 수 있도록 하며 전문 검색 시 적합성을 증가시킬 수 있는 방법으로 동의어 형태의 단어 확장이 자동으로 발생하는 자연언어 탐색기법의 적용에 대한 연구가 필요하고 정보의 생산뿐만 아니라 정보의 유통과 교환을 위해 전자 문서의 표준안인 SGML로의 변환에 대한 연구가 필요하다.

REFERENCES

Cho H. Management of medical informatics. JeongMoonGak.

1997. 139.
- Inmon WH, Hackthorn RD. Implementation of datawarehousing. Hightech information. 1996. 2-4.
- Jeong YM, An HS. Implementation of electronic library. GooMi Press. 1998. 196-200.
- Jeong YM. The theory of information retrieval. GooMi Press. 1993. pp. 11-12, pp. 22-26, pp. 237-243, pp. 106-118.
- Kim JR. Educational engineering. MoonEumSa. 1994. 194-195, pp. 86-87.
- Kim SI, Kim DW, Park GS. Medical informatics. YeoMoonGak. 1995. pp. 9-10, pp. 115-117.
- Kim YI. Design and implementation of object-oriented database full-text retrieval and search. A master's thesis. 1996.
- Lee MY, Chun ST, Heo DY. Bada III multimedia database in full-text retrieval. Kor inf sci society. 1996. 14: 43-50.
- Park SH. Information retrieval system. Gold Press. 1996. 4. pp. 29-38.
- Perkins Jeff, Morgan Bryan. Easy learned SQL. Samgakhyong Press. 1996. 56-65.
- <http://www.lg.or.kr/digilib/lecture>.
- <http://www.snu.ac.kr/~course/DB/ch2/ch2.3>.