

한의학 문헌정보의 전자도서관 구축

한중현 · 윤용갑* · 한성국, 주수중**

I. 연구개발의 목적 및 필요성

1. 연구개발의 목적

한의학은 전통의학으로써 유구한 역사를 통해서 방대한 문헌이 축적되어 왔으나, 이들 문헌자료가 체계적으로 정리되고 집합되어 있지 않아 한의학 연구와 교육에 여러 문제를 야기하고 있다. 컴퓨터 기술의 급속한 발전과 정보화 시대의 도래에 따라 한의학 문헌정보를 디지털화 하여 문헌정보의 활용을 극대화하여 한의학 연구를 활성화하고 과학화하는 한편, 인터넷 등 초고속 정보통신망을 통해서 체계화된 한의학 연구 정보를 제공함으로써 한의학의 세계화와 중추국으로서의 국가 경쟁력을 제고할 필요성이 증대되고 있다.

본 연구에서는 한의학 문헌정보 중에서도 핵심이 되는 한약재와 처방정보를 총체적으로 집약한 멀티미디어 기반의 전자도서관을 구축한다. 구축된 전자도서관은 멀티미디어 데이터베이스와 지능형 정보 검색기를 기반으로 고도의

기능과 사용자가 필요로 하는 한의학 정보를 신속, 정확하게 서비스하며, 국내외 초고속 정보 통신망과 연계하여 한의학 정보 센터로의 역할을 하도록 구현한다.

정보가 무한한 가치를 갖는 정보 사회에서는 정보의 가공과 유통체계가 국가 발전과 직결하고 있으므로, 우리만이 갖고 있는 전통 한의학 정보는 경쟁력 있는 고유 정보 상품으로 세계화와 정보화가 시급한 분야이다. 국가 초고속 정보통신망과 인터넷 등으로 멀티미디어 기반의 한의학 정보를 제공하는 한의학 전자도서관을 구축하여, 한의학의 중추국으로서의 면모를 과시하고 한의학의 과학화와 현대화에 기여하는 한편, 정보화 사회의 새로운 서비스 창출에 기여하고자 한다.

2. 연구 개발의 필요성

한의학은 한자문화권의 동양의학으로 한국, 중국, 일본에서만 사용되는 전통의료 기술이다. 특히 전통 한의학은 수천 년을 통해 이론과 기술이 전수되어 오면서 체계화한 생명의학으로써 세계 여타의 나라에서도 그 유래를 찾기 어

※ 본 논문은 1997~1998년도 한국한의학연구원에서 시행한 한방치료기술연구개발사업지원 연구비에 의해 수행된 것임.

* 원광대학교 한의과대학 한의학과

** 원광대학교 공과대학 컴퓨터 공학과

려운 고유의술이다. 특히 현대의 분석적 서양의 학이 당면하고 있는 제반 문제를 해결할 수 있는 분야로 양의학계에서도 그 유효성이 설득력 있게 부상되고 있다. 한의학 전자도서관 구축의 필요성은 다음과 같다.

- (1) 수천 년간 전수되어온 방대한 양의 한의학 자료를 체계화
- (2) 한의학 자료를 첨단 저장 매체에 저장, 활용성 극대화
- (3) 초고속 정보통신망을 통해서 멀티미디어 한의학 정보를 제공함으로써 한의학의 현대화와 세계화
- (4) 한의학 정보 공유를 통한 한의학 연구의 활성화와 과학화
- (5) 한의학을 차세대 의학으로 발전시키기 위한 토대 구축
- (6) 고유 과학 문화 전통을 세계화함으로써 국가 경쟁력 제고
- (7) 사회 각 분야에서 요구되고 있는 전자도서관 관련 요소기술 확보
- (8) 전자매체 활용을 통해 정보화사회 구현

II. 국내·외 기술개발 현황

본 과제와 관련하여 전자도서관 구축에 관한 선진 외국의 추진중인 연구·개발의 실태와 현황을 살펴보고, 더불어 국내의 추진현황 및 앞으로 방향을 살펴본다.

먼저 국외의 기술현황으로 가까이 있는 일본의 경우, 1994년 8월 총리를 책임자로 하고 각료를 구성원으로 하여 '고도정보화통신사회의 추진실'이라는 기구를 설립하여 멀티미디어 전자도서관 연구추진을 실시하였다. 1995년 8월 이후, 과학, 문화, 체육분야에 정보구축 지원해오는 한편, 전자도서관 구축을 국가정보화 정책의 핵심사업으로 강력하게 추진하고 있다. 참고로 1994년 ~ 1997년까지 대학, 일본과학기술원

(NAIST) 및 일본학술정보센터(NACSIS) 전자도서관 구축을 위해 일본 정부는 15억2천7백9십만 엔을 투자해 왔다. 미국 또한 DLI 프로젝트를 시발점으로 하여 대학을 모체로 하여 전공분야 별 전자도서관 개발 사업이 추진되고 있으며, 엄청난 예산을 투입하고 있다. 유럽의 경우, 특히 영국의 Initiative of Access, 독일의 통합 가상 도서관 체계, 프랑스의 Tolbiac 프로젝트들이 자국의 국립 도서관 자료들을 전자도서관화를 위해 추진되고 있다. 각국의 대표적인 전자도서관 프로젝트의 현황은 아래와 같다.

- (1) 일본 정보처리진흥청(IPA) 전자도서관 프로젝트
 - ① 전자도서관 시범 사업
 - ② 시범사업 창출을 위한 시범사업
 - ③ 교육용 소프트웨어 개발사업
- (2) 일본 과학 기술원(NAIST)의 Mandala Project
 - ① 단행본 기술 보고서, 잡지 등의 디지털화
 - ② 도서관을 미디어 센터화
- (3) 일본학술정보센터(NACSIS) 전자도서관 시스템(ELS)
 - ① 기존의 서지 데이터베이스 검색 시스템과 디지털화된 학술잡지 및 회의자료를 통합한 정보 시스템
 - ② 일본 정보처리학회 모든 간행물의 디지털화
- (4) 일본 Ariadne Project
 - ① 일본 전자도서관 개발 그룹과 FUJITSU 공동 연구
 - ② 5대 특징 : Digitigation, Networking, Interactive, Multimedia, Scalable
 - ③ 영어와 일어의 기계 번역 가능
- (5) 미국 NRL 전자도서관 프로젝트
 - ① 미국 해군 연구소(NRL)의 연구도서관과 기술정보센터 중심
 - ② 캠퍼스 전역 정보 시스템 Info-Net 구축
 - ③ 광 이미지 시스템 구축
 - ④ TORPEDO 시스템도 구현
 - ⑤ 메뉴방식의 사용자 인터페이스
 - ⑥ 18개 CD-ROM 데이터베이스

(6) 미국 과학 재단 (NSF) 기금 중심 전자도서관

① Carnegie Mellon University

- 4백 8십만 달러
- 인포미디어 상호 작용의 온라인 전자도서관
- 과학과 수학 분야 자료

② UC Berkeley

- 4백만 달러
- 환경 정보
- 자동 색인, 지능형 검색

③ University of Michigan

- 4백만 달러
- 지구 과학과 우주 공학 분야
- 멀티미디어 기반으로 주제 정보 체계화

④ UC, Santa Barbara: The Alexandria Project

- 4백만 달러
- 지도, 이미지, 사진 등 지리 정보 시스템
- 공간 색인 정보(Spatially - indexed information)

⑤ Stanford University

- 통합된 가상 도서관 기술 개발
- 대규모 공유 데이터 베이스

(7) 미국 의회 도서관 NDLP

- ① 미국의 고유역사의 디지털화 - America Memory Project
- ② 국회도서관의 예산과 기부금으로 추진

(8) 영국의 디지털 도서관 프로젝트

- ① 1993년부터 시작하여 20여가지 개발 계획
- ② 특허 정보 관리용 고속 유크 박스 운영
- ③ 고대 서사시 필사본을 디지털 형태로 변환
- ④ 전자 사진 열람 시스템 운영
- ⑤ 온라인 네트워크 목록(Network OPAC)제공

현재, 국내에서도 정부나 민간단체의 관심이 고조되면서 전자도서관 사업이 본격적으로 진행되고 있으나 디지털 콘텐츠 제작부분의 투자는

외국에 비해 뒤떨어지고 있다. 국내의 주요 전자도서관으로는 민간부분에서 국내 최초의 전자도서관인 LG 상남 도서관, 대학에서는 홍익대학교, 공공기관에서는 컨소시엄 형태로 추진되는 국가 전자도서관, 한국과학기술원의 첨단 과학기술 전자도서관 구축 프로젝트들이 있다. 국내의 전자도서관의 프로젝트 및 개발 현황을 살펴본다.

(1) LG 상남 도서관

- ① 1996년에 국내 최초로 개관한 전자도서관 (ELIT)
- ② 1997년에 웹 기반의 인트라넷 시스템으로 바뀜
- ③ 학술잡지, 회의자료에 수록된 원문의 이미지(Tiff)로 디지털화 서비스

(2) 홍익대학교 전자도서관

- ① 최초의 대학의 전자도서관
- ② 학술자료(소장자료, 기사, 학위논문 원문, VOD, 학술CD)의 온라인 검색
- ③ 디지털 콘텐츠는 여타 기관과 같이 빈약한 상태임

(3) 국립중앙 도서관

- ① 1995년부터 디지털 콘텐츠의 대량으로 개발
- ② 국자 전자도서관 시범/연계사업의 지원으로 수행
- ③ 디지털 대상자료 : 인문과학분야의 박사학위논문, 한국 역사 및 고전

(4) 한국과학기술원의 과학기술 전자도서관

- ① 한국과학기술의 주관으로 4년의 개발기간 (1997-2001)에 순 개발 투자액으로 130억 지원
- ② 디지털 콘텐츠 중심의 지식기반 미래형 과학기술 전문 전자도서관
- ③ 정보 재창출의 지식 저장소의 역할 수행
- ④ 새로운 패러다임의 전문 도서관의 모델을 제시
- ⑤ 사용되는 요소기술 : 지능형 통합 멀티미디어 정보베이스 기술, 정보보안 기술, 상호가동성 기술, 제도 및 표준화 연구

(5) 기타 프로젝트들

- ① 학술정보센터의 RISS
- ② 국회도서관
- ③ 산업기술정보원 전자도서관
- ④ 한국과학기술의 과학도서관

국외의 경우는 물론, 국내에서 프로젝트로 수행 또는 개발된 한의학 문헌정보의 전자도서관의 구축 사례는 언급된 적이 없으며, 한국 한의학 연구원의 지원에 의한 본 과제가 처음이라 사료된다. 한의학 문헌은 오랜 역사를 가지고 있으며, 그 분량 또한 워낙 방대하다. 문헌의 자료자체도 대부분 세로쓰기의 탁본이나 복사된 원전들이 대부분이다. 문헌에 쓰이는 일부의 한자들이 현 시스템 환경에서 지원되지 않는 관계로 문자인식 과정을 거치지 않고 스캔된 문서를 .pdf 형태로 변환할 경우, 단어를 추출할 수 없으며 단순히 하나의 이미지 자료로 처리 및 사용할 수밖에 없었다. 이러한 문제는 한의학 문헌들로 자동 인덱스의 추출이나 시소러스 구축에 걸림돌이 되고 있다. 다시 말하면, 한의학 전문인들의 지원이 불가피하다. 문헌정보 중 약물이나 약제 등의 비정형 데이터에 대한 인덱스 구축 및 관리유지가 어려웠다. 그러나 본 연구개발에서는 멀티미디어 데이터베이스의 구축과 검색을 위한 인덱스 구축을 위해 객체-관계형 모델을 이용하였고, 한의학 전문교수들의 지원으로 한의학 문헌들(방제, 약리 관련자료)을 계층적 구조로 계통 체계를 분류시킨 노력으로 인덱스들의 추출 문제를 해결하였다. 또한 지능형 정보검색 엔진을 개발하여 한의학 문헌정보의 특성을 고려한 검색을 지원할 수 있도록 하였다.

결과적으로 본 연구개발을 방대하지는 않지만 다양한 형태의 한의학 자료(황제내경, 고급의방집성, 동의보감, 화한약물학)를 이용하여 객체-관계형 데이터베이스를 구축하였다. 이들 정형 및 비정형의 문헌정보들을 검색하기 위해서 키 인덱싱과 내용기반 인덱싱을 제공할 수 있도록 하였다. 본 전자도서관을 서비스를 제공

하기 위한 시스템 환경은 유닉스 기반에서 최근 기술인 객체-관계 모델의 데이터베이스를 구축하기 위해서 객체-관계 데이터베이스 시스템인 Informix-Universal Server를 이용하였으며, 이 데이터베이스 시스템과 웹과의 연동을 위해 Informix Web DataBlade Module을 사용하여 웹 기반의 서비스를 제공하도록 하였다.

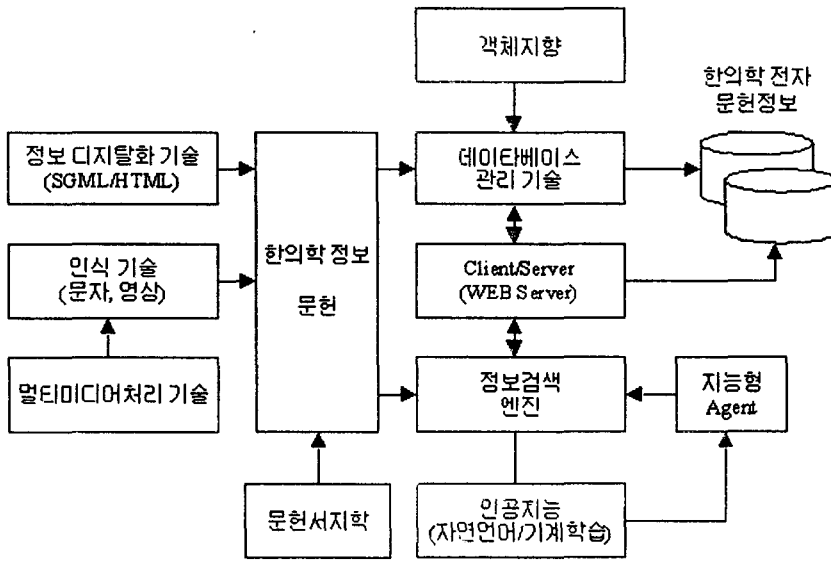
Ⅲ. 연구개발의 내용 및 결과

전자도서관의 구축은 시스템 측면과 콘텐츠(contents) 측면의 두 가지로 구분하여 생각할 수 있다. 시스템 측면은 전자도서관의 하드웨어적 특성을 규정하고 콘텐츠 측면은 소프트웨어적 특성을 명시하게 된다. 따라서, 본 과제의 수행에 관한 역할 분담 및 내용은 다음과 같다.

우선 전자도서관 구축의 내용적 측면을 담당하는 콘텐츠 측면은 한의학의 문헌정보 중에서 처방에 관련한 문헌정보 시스템의 구축과 한약재의 정보 시스템 구축을 그 내용으로 하였다. 즉 한의학의 문헌정보를 전자도서관으로 구축하는 데 있어서 가장 중요한 부분이 바로 처방의 문헌정보와 한약재의 정보 시스템 구축이기 때문이다. 이 두 가지 내용을 전자도서관 구축의 콘텐츠 측면으로 활용하게 한 것이다.

다음으로 한의학 전자도서관 구축의 시스템 측면은 컴퓨터공학과에서 담당하였다. 공학적인 측면을 담당하는데 있어서는 컴퓨터 시스템 상의 운영적 측면을 고려하여 한의학 문헌정보의 멀티미디어 데이터베이스를 구축하는 부분과 한의학 문헌정보를 위한 지능형 검색엔진을 개발하는 부분으로 나누어 과제를 수행하게 하였다. 이것은 다시 말하면 사용자의 편의를 도모하기 위한 상하 개념의 복합적 접근이 가능하도록 하기 위한 것이다. 즉, 멀티미디어 데이터베이스를 구축함으로써 위로부터 아래쪽으로는 통합적인 정보 제공이 가능하게 할 수 있고, 지능형 검색 엔진을 개발함으로써 아래에서 위쪽

으로 검색해 나갈 수 있는 시스템을 구축하는 것이다.



[그림 1] 연구개발 기술 및 방법

1. 콘텐츠 측면

1) 연구개발 내용

한의학 의료 정보연구에 있어서 한약처방에 관한 연구는 매우 중요하다. 특히 한약처방에 관해서는 그 동안 축적한 경험적 문헌자료와 최근의 임상과학의 실험적 연구 결과 등을 함께 조화롭게 다루어야 한다. 그러나 지금까지는 한약처방에 관한 연구자료수집과 산발적으로 흩어져 있어 관련 연구자료수집과 한약처방에 관한 연구 등에 많은 문제점이 제시되고 있다. 연구에 있어서의 주안점은 다음과 같다.

- (1) 한약재와 처방문헌의 긴밀성
- (2) 한의학 전자도서관을 위한 기본 핵심 데이터
- (3) 다양한 처방 자료의 수집과 정리
- (4) 처방 문헌의 과학적 분류

한의학에 관한 서적이 BC 이전의 내용에서부터 현대에 이르기까지 다양하며 내용이 대부분 한문으로 쓰여져 있어 이에 대한 해독력과 이해에 많은 어려움이 있다. 또한 내용 중에 인용된 한약재의 종류가 다양하고 쓰이는 용도가 질병에 따라 다양하여 한의서적으로부터 찾아 응용하기에 많은 어려움이 있는 실정이다. 따라서 처방에 대한 자료(서적)의 수집 및 분석, 처방에 대한 자료(실험결과)의 수집 및 분석을 통하여

- (1) 처방의 전자정보시스템
- (2) 멀티미디어 데이터베이스제작
- (3) 멀티미디어 처리기술
- (4) API 인터페이스
- (5) 사용자 인터페이스의 개발을 연구개발내용으로 한다.

2) 처방에 관한 문헌정보 총괄

우선 한의학 전자도서관 구축에 필요한 문헌 영역을 한방생리, 한방병리, 처방분석의 3단계로 나누었다. 한의학에 관한 연구를 진행하기 위해서는 한방생리를 분석하는 것이 선행되어야 하고, 질병을 다루기 위해서는 한방병리를 연구해야 하며 그리고 나서는 처방과 약재를 결부시켜야만 질병에 대한 종합적 접근이 가능하기 때문이다.

한방생리 분야에서는 黃帝內經을 데이터베이스화하는 것이 전자도서관 구축에 필수적이라고 생각하였다. 黃帝內經은 중국의 춘추전국시대에 나온 책으로, 한의학계에서는 최고의 원전으로 간주되고 있으며 현재까지도 한의학 입장에서 그 가치를 높이 인정받고 있다.

한방병리 분야에서는 조선의 許俊이 펴낸 東醫寶鑑을 데이터베이스화하였다. 東醫寶鑑은 許俊의 수십 년간의 의료경험과 아울러 중국, 한국 등에서 발간된 모든 의학 서적들을 총망라한 백과 사전적 저작으로 질병에 대한 자세한 분석이 곁들여져 있어서 중국에서도 그 가치를 인정받고 있다.

처방 분야에 있어서는 黃帝內經 시대부터 근대까지 대표적인 처방 10,000 여 종을 수록하고 있는 古今醫方集成的 처방과 6,000 여 방을 수록하고 있는 東醫寶鑑 처방을 중심으로 데이터베이스화를 시도하였다.

3) 처방에 관한 문헌 및 실험 자료

(1) 한의학 문헌정보 데이터베이스 구축

① 황제내경의 데이터베이스화

黃帝內經은 素問 81편과 靈樞 81편으로 구성되어 있으며, 전 편을 다음과 같은 형식으로 데이터베이스화 하였다.

- ㉠ 黃帝內經을 素問과 靈樞로 크게 나누었다.
- ㉡ 素問 81편을 각 1편 당 1개의 hwp file로 만들었다.

㉢ 靈樞 81편을 각 1편 당 1개의 hwp file로 만들었다.

② 東醫寶鑑의 데이터베이스화

東醫寶鑑은 신형편, 외형편, 내경편, 잡병편, 탕액침구의 5개의 편으로 크게 분류할 수 있으며 하나의 편 안에는 80여 개의 문으로 다시 분류할 수 있고, 각각의 문에서는 다시 700 여 개의 증상 및 병증과 그에 대한 6,000여 개의 처방으로 나누어 데이터베이스화하였다.

- ㉠ 東醫寶鑑을 5개의 편으로 분류하여 hwp file로 만들었다.
- ㉡ 80 여 개분에 따른 소주제를 hwp file로 만들었다.
- ㉢ 700여 증상 및 병증을 file로 만들었다.
- ㉣ 東醫寶鑑에 나오는 6,000여 처방들을 따로 분류하여 다른 처방들과 연계할 수 있는 file로 만들었다.

(2) 문헌의 인덱스 추출 및 분석

한의학 문헌을 전자 정보로 이용하기 위해서는 인덱스를 추출하는 것이 반드시 필요하며 인덱스 작업이 갖는 의미는 한의학 문헌의 방대한 분량을 핵심적인 한의학 용어 위주로 인덱스 함으로써 사용자들이 편리하게 사용할 수 있게 만드는 것이다.

(3) 처방 문헌의 수집 및 분석

현존하는 한의학 처방을 문헌 정보화하는 것이 주된 목표이며 따라서 본 연구를 진행하면서 많은 서적을 참조하고 다양한 실험결과와 기존에 작성된 논문을 검토, 수정, 보완 및 분석하여 고금의방집성 및 동의보감처방 16000여 처방을 데이터베이스화하였다. 그 과정에서 하나의 처방이 여러 문헌에 나오는 경우도 있었고, 또 같은 처방이라도 출전에 따라, 또 의가의 학설에 따라 다양하게 변화하는 것을 관찰할 수 있었다. 그 경우에는 처방을 창조한 의가의 본래의 뜻에 따르는 것을 원칙으로 하였으며, 동일한 처방이 중복되는 경우에는 보다 더 오래된 문헌을 출전으로 간주하였고, 같은 처방

명을 가진 처방이라도 구성 내용이 다른 경우에는 다른 처방으로 간주하고 그 출전이나 의가의 이름을 기록하는 것을 원칙으로 하였다.

① 고급의방집성 처방

- ㉠ 처방명에 따른 인덱스 분류
- ㉡ 처방의 주치에 따른 인덱스 분류
- ㉢ 처방의 효능에 따른 인덱스 분류
- ㉣ 처방구성 약물 인덱스 분류

② 동의보감 처방

- ㉠ 처방명별 인덱스 분류
- ㉡ 처방의 주치별 인덱스 분류
- ㉢ 증상 및 병증에 따른 처방 인덱스
- ㉣ 처방의 구성 약물에 따른 인덱스 분류

4) 처방 정보를 활용한 한의학 전자도서관 전문 모델

본 연구를 진행하면서 수집, 분석된 처방은 총 16,000여 처방이었다. 그러나, 이들 처방을 단순히 데이터베이스화하는 것은 그 활용도의 측면에서 볼 때 크게 의미가 없다고 할 수 있다. 하나의 처방을 찾아내는 것도 물론 중요하지만 한의학 문헌정보의 가치에서는 한의학의 이론에 입각하여 그 생리, 병리를 알고 그 후에 해당하는 처방을 찾을 수 있도록 연계하는 과정이 더욱 중요하기 때문이다. 그래서, 본 과제에서는 처방 정보의 활용을 극대화할 수 있도록 다음과 같은 구조로 전자도서관의 전문 모델을 구성하였다.

- (1) 한의학적인 질병의 원인이 되는 기본 생리 이론을 황제내경을 이용하여 찾는다.
- (2) 찾아낸 기본 이론의 병리적 현상을 동의보감을 이용하여 찾는다.
- (3) 동의보감 내에서 병리 현상의 치료 처방을 찾는다.
- (4) 동의보감 이외의 다른 처방을 연계하여 찾는다.

2 시스템 측면

1) 검색 시스템을 위한 스키마 설계

(1) 한의학 문헌정보의 특징

한의학은 한자문화권의 동양의학으로 한국, 중국, 일본에서만 사용되는 전통 의료기술이다. 특히 전통 한의학은 수 천년을 통해 이론과 기술이 전수되어 오면서 체계화된 생명과학으로서, 자료의 양이 방대하여 이의 휴대 및 보관과 사용상의 불편함을 감수해야만 했다. 이를 해결하고자 한의학 문헌 정보를 디지털화 할 경우 다음과 같은 문제가 내포되어 있다.

- 한의학 문헌들은 대부분 세로쓰기 형태를 이루므로 문자인식이 어렵다.
- 윈도우즈에서 고문헌에 쓰이는 한자어나 자주 쓰이지 않는 한자어는 지원되지 않는다.
- 스캐너에 의해 입력된 문서는 문자인식 과정을 거치지 않고 PDF형태의 문서로 변환할 경우 각각의 단어를 추출하지 못하고 전체를 하나의 이미지로 인식한다.

상기 제기된 문제점들은 한의학 문헌에 대한 자동 인덱스 추출 및 서소러스 구축에 걸림돌로 작용하고 있다. 또한 한의학 문헌은 약물이거나 약재 그리고 인체의 구조 등에 대한 비정형 데이터를 포함하고 있다. 이들 비정형 데이터에 대한 인덱스의 구축 및 유지는 효율적이지 못하다. 다음은 이를 해결하기 위한 멀티미디어 데이터베이스의 구축과 검색에 관한 객체-관계형 모델링에 대하여 살펴본다.

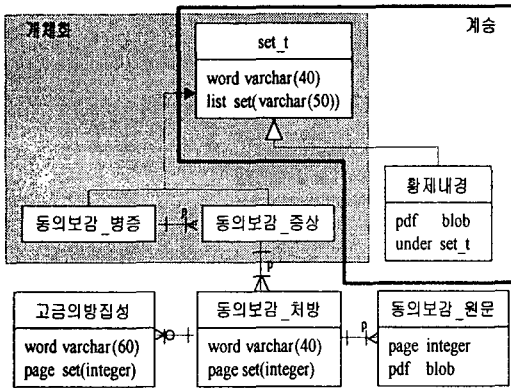
(2) 문헌 자료를 위한 스키마 설계

상기된 한의학 문헌의 특성과 또한 대부분의 전문가들이 원래 문헌의 형태 그대로 문서를 보기 원하므로 문헌 자료에 대해서는 스캔된 이미지를 인터넷에서 범용으로 사용되는 PDF 형태의 자료로 변환하여 사용하였다. 한글 문서의 경우 파일 시스템에 보관하고 검색 엔진과 한글 문서 파서 모듈을 연동하여 사용하면 검색이 가능하며, 이는 병행하여 추진되고 있는

세부과제에서 연구 중에 있다.

동의보감의 내용은 병증, 증상, 처방의 계층적 구조로 표현될 수 있다. [그림 2]는 다른 문헌과의 관계성을 고려한 객체-관계형 데이터 모델을 보인다.

그림에서 각 테이블간의 관계를 나타내기 위해서는 IE 표기법을 사용하였고 사용자 정의 데이터 타입의 상속이나 개체화를 나타내기 위해서는 OMT 표기법을 사용하였다.



[그림 2] 동의보감을 중심으로 본 객체-관계형 데이터 모델

사용자 정의 전역 데이터 타입인 "set_t"는 복합 객체의 하나인 합성 타입을 생성한다. set_t의 레코드인 list는 set타입으로서 동일한 타입을 나열할 수 있어 검색하고자 하는 키워드에 대한 하위 키워드들을 하나의 어트리뷰트에 저장하고 사용자 정의 함수를 이용하여 각 단어를 추출하여 검색할 수 있다. 사용자에 의해 정의된 전역 데이터 타입인 set_t는 동의보감 병증과 동의보감 증상에서 개체화되고 각각의 테이블을 만들며 황제내경 테이블에 상속되어 pdf, word 및 list의 레코드들을 갖는 황제내경 테이블을 생성한다. 객체-관계형 데이터베이스의 특징인 계승은 모델링 시 신중히 고려되어야 한다. 다중의 상위 타입으로부터 계승이 이루어질 때에는 모호성이 발생할 수 있으며, 여러 계층에 거쳐 계승이 이루어질 경우에는 하위 계층에 속해있는 상위 타입의 개체에 대해 추가적인 검사가 이루어져 성능의 저하가

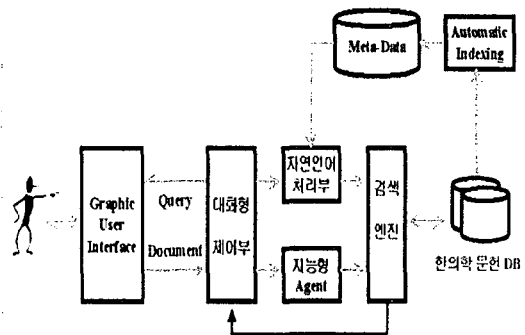
일어나기 때문이다.

2) 시스템 구축 및 실행 결과

(1) 시스템 구축

본 연구는 객체-관계형 데이터베이스 엔진을 사용하고 WEB과의 연동을 위해 Informix Web DataBlade Module을 사용하였다. 각 HTML문서는 하나의 단위 기능을 하는 AppPage 형태로 구성되어 있으며 데이터베이스에 대한 SQL 질의는 Script 형태로 HTML 문서에 저장되어 데이터베이스 서버에 전송된다. 또한 HTML 문서는 그 자체로서 데이터 타입이 되어 데이터베이스에 저장되므로 CGI나 Script가 외부로 도난당할 위험이 적어 보안상 유리한 구조를 이루고 있다.

지능형 검색엔진을 이용한 한의학 문헌 검색 시스템은 다양한 검색기법을 이용해서 결과에 대한 신뢰도와 정확도를 높이는데 역점을 두었다. 전체 시스템 구성도를 살펴보면 [그림 3]과 같다.



[그림 3] 한의학 문헌정보 검색 시스템 구성도

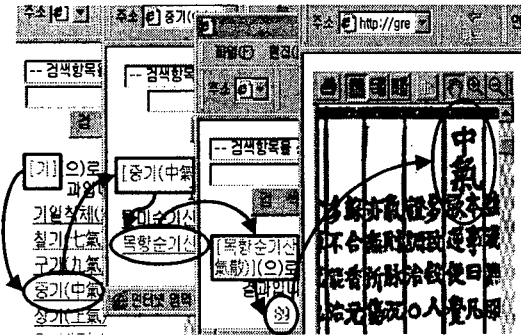
마이크로소프트에서 제공하는 전략적 웹서버인 인터넷 인포메이션 서버를 모체로 하고 Index Server를 검색엔진으로 사용하였으며, Server side script 언어인 Active Server Page를 이용해 프로그래밍 하였다.

어떤 정보의 내용을 기반으로 한 검색에 있

어서 중요한 요소는 사용자가 의도한 검색 결과의 정확도와 반응 속도라고 볼 수 있다. 이러한 결과를 얻기 위해서는 정보검색에 있어 필수적이라 할 수 있는 효과적인 인덱싱이 전체 시스템의 성능을 좌우한다. 인덱싱을 위한 기초 자료를 수집하고, 사용자가 필요한 정보를 찾기 위해 하나의 검색어(자연어) 만으로 해당 주제를 전부 검색할 수는 없다. 그러므로 검색어에 관련된 대소관제나 동의어, 동형이의어, 관련어 등을 적절히 조절하여 모두 검색해야 하는 불편함을 없애기 위해서 정보시스템과 문헌생산자, 색인작성자, 이용자간에 통일적으로 사용할 수 있도록 통제하여둔 용어통제어표인 시소러스(Thesaurus)의 구축이 필요하다.

(2) 실행 결과

본 절에서는 구축된 시스템의 실행 결과를 보인다.



[그림 4] 다단계 인덱싱에 의한 동의보감 문헌 검색 결과

[그림 4]는 첫 번째 단계(병증)의 “기(氣)” 키워드에 대한 검색의 예를 보인다. 이 검색에 대한 결과로 하위 단계(증상)에서 키워드들을 검색해내며, 이런 다단계 인덱싱 검색 과정을 거쳐 사용자의 질의에 답하게 된다.

다음은 지능형 검색 엔진을 이용한 문헌 검색의 결과이다. 고문서나 표현이 어려운 확장한자가 많이 포함된 한의학 문서의 경우에도 원문을 그대로 보여주기 원하는 사용자들의 요구를 반영해 이미지를 사용하였다. 또한, 인터넷

에서의 사용을 고려해 뛰어난 압축률과 함께 이 기종간의 표준 문서교환 포맷 등 여러 가지 장점이 있는 Acrobat PDF를 원문의 문서 포맷으로 채택하였다. 최종사용자의 경우 단지 PDF 뷰어 플러그인을 설치하면 모든 상호작용이 가능하다.

[그림 5]는 각각 한글화된 문서를 검색하여 그 결과를 원문과 일치하도록 변환하는 방법이다. [그림 5]의 방법으로 구조화된 파일 이름을 얻을 수가 있으며, PDF 뷰어를 통해 검색 결과를 볼 수 있다.

```

sOriginPath = RS("vPath")
Full_Path의 길이
nFullPath = Len( sOriginPath )
*문자열을 역순으로 배열 후 첫번째 나오는 / 문자의 위치
nFirstLocation = InStr( 1, StrReverse( sOriginPath ), "/", vbTextCompare )
*Full_Path에서 경로부분과 파일부분을 nFirstLocation을 이용해 분리
sPath = mid( sOriginPath, 1, nFullPath - nFirstLocation + 1 )
sFile = Right( sOriginPath, nFirstLocation - 1 )
*Replace sFile, ".htm", ".pdf"
sFile = left( sFile, Len(sFile)-3 ) & ".pdf"
경로와 파일 사이에 새로운 위치정보를 삽입.
sModifiedPath = sPath & "_pdf/" & sFile
sFileTitle = sPath & sFile
    
```

[그림 5] 결과를 원문으로 보여주기 위한 코드

위에서 언급한 다양한 검색기법으로는 기존의 많은 검색엔진에서도 찾아볼 수 있는 질의의 형식과 질의에 대한 연산기능을 들 수 있다. 본 시스템에서는 두 가지 연산자와 3가지 형태의 질의를 이용한다.

- ① 부울 및 근접 연산자
- ② 대표문자 질의
- ③ 의미 중심 질의
- ④ 백터공간 질의. 단어와 구의 목록과 일치하는 결과를 요구하는 질의.

질의를 공식화 하는 몇 가지 규칙은 연속적인 단어가 구로 취급되는 것과 검색대상에서 제외되는 Noise word로는 알파벳의 관사와 접속사등이며, 마침표(.), 콜론(:), 세미콜론(;), 쉼표(,)와 같은 구두점이 무시된다. 이 외에 앞서 명시한 부울(BOOL) 및 근접(NEAR) 연산자를 통해 추가적인 검색 정보를 지정할 수 있다.

IV. 결 론

본 연구를 통해 방대한 처방문헌 자료에 대한 Data Base화 기초의 기반을 확립할 수 있었으며 많은 한의학 정보의 시스템 구축의 전기를 마련하였다고 본다. 즉, 본 연구의 수행결과에서 한의학 문헌정보가 입력될 멀티미디어 데이터베이스의 구조는 객체지향형 모델링 방법을 적용함으로써 중복된 정보의 저장을 회피하고, 효율적인 다단계 인덱싱 방법을 사용하여 빠른 검색할 수 있는 사용자 위주의 시스템 환경을 만들었다. 더욱이 본 데이터베이스 관리시스템은 최근에 출시한 Informix-Universal Server DBMS로서 멀티미디어 정보를 효율적으로 저장(객체-관계 데이터베이스 구조) 및 관리할 수 있고, 본 시스템에서 제공하는 Web Blade 및 Image Blade를 적용하여 복잡한 데이터 구조를 갖는 정보의 검색 뿐 아니라 인터넷과 접목이 가능하도록 구현되었으므로 기존의 웹 브라우저와 관련 정보의 특성에 따라 Plug & play를 이용하여 구축된 한의학 문헌정보를 검색할 뿐 아니라 구축된 색인어는 지능형 검색 시스템에서 제공함으로써 효율적인 검색 지원을 제공할 수 있다. 또한 본 연구 중 지능형 검색엔진 구축은 사용자 인터페이스와 직결된 시스템의 Front-end이다. 구현된 검색엔진은 검색의 정확성과 신속성, 검색결과의 다양한 표현능력 등 검색에 필요한 기본 기능에 충실하고 있으며, boolean 검색기능의 제공과 한국어, 한문 등 다국어 처리 기능 등을 갖는 고성능 검색엔진을 구현함으로써 대량 정보문헌 검색의 기반을 조성하였다. 본 연구에서 구현된 검색엔진은 객체지향기반으로 구현하여 Component함으로써 문헌정보 시스템 구축을 효과적으로 실현할 수 있도록 하였다. 또한 새로운 Pattern Matching Algorithm을 유도하여 고속 Full Text 검색기를 구현하였고 Regular Expression 처리를 가능하게 하여 검색의 가용성을 제고하였다. 본 연구의 수행 결과는 지능형 검색 엔진

구현에 대한 요소 기술로 활용할 수 있다.

본 연구 결과에 대한 종합적인 가치를 논의해보면, 수천 년간 전수되어온 방대한 양의 한의학 자료를 부분적이거나 체계화하는 계기가 되었고, 한의학 자료를 디지털화하여 첨단 저장 매체에 저장함으로써 재활용의 극대화는 물론, 초고속 정보통신망과 연계하여 멀티미디어 한의학 정보를 공유화 함으로써 한의학 연구의 활성화를 촉진시킬 수 있을 것이다. 또한 한국적인 과학 문화 전통을 세계화함으로써 국가 경쟁력을 제고시키고, 나아가 한의학을 차세대 의학으로 발전시키기 위한 토대를 구축하였다고 본다. 기술적인 면에서도 전자도서관 관련 요소기술을 확보함으로써 다양한 분야에 응용할 수 있으리라 생각된다. 앞으로 이러한 개발 기술을 이용하여 한의학 용어의 전자사전 구축, 한의학 전자저널 발간, 한의학 정보교류를 위한 가상공간(cyberspace)구축, 한의학 교육의 멀티미디어화와 같은 응용에 접목을 시켜 나가야 할 것이다.

참 고 문 헌

1. Ehrlich, K., and Cash, D. 1994. Turning information into knowledge: Information finding as a collaborative activity. Proceedings of the Digital Libraries '94 Conference, (College Station, TX, June 19-21), pp. 119-125.
2. Lesk, M. 1991. The CORE electronic chemistry library. Proceedings of the Fourteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, (Chicago, IL).
3. L, ken, S. 1993. Text Representations In Digital Hypermedia Library Systems. M.S. Thesis. Department of Computer Science,

Texas A&M University, College Station,
TX (Dec).

4. Miksa, F., and Doty, P. 1994. Intellectual realities and the digital library. Proceedings of the Digital Libraries '94 Conference, (College Station, TX, Jun 19-21), pp. 1-5.
5. Schnase, J. L., Leggett, J. J., Metcalfe, E. S., Morin, N. R., Cunnius, E. L., Turner, J. S., Furuta, R. K., Ellis, L., Pilant, M., Ewing, R. E., Hassan, S. W., and Frisse, M. 1994. The CoLib project - Enabling digital botany for the 21st century. Proceedings of the Digital Libraries '94 Conference, (College Station, TX, Jun 19-21), pp.108-118.
6. Kahn, Robert and Wilensky, Robert. "A framework for distributed digital object services". May, 1995.
7. Steve B. Cousins, Andreas Paepcke, Terry Winograd, Eric A. Bier, Ken Pier. "The Digital Library Integrated Task Environment" (DLITE) : Digital Libraries '97
8. Thomas A. Phelps and Robert Wilensky, Multivalent Digital Documents in UC Berkeley's Digital Library Project, SIGLINK, Vol. 4, No. 2, September 1995.
9. Thomas A. Phelps and Robert Wilensky, Multivalent Documents : Inducing Structure and Behavior in Online Digital Documents , Best Paper Award - Digital Documents Track, Proceedings of the 29th Hawaii International Conference on System Science (Maui, Hawaii, January 3-6, 1996).
10. Christian Plaunt, A Workbench for Information Retrieval Experimentation, Association of Lisp Users Meeting and Workshop, Workshop Proceedings, 1995.
10.1 Malvern, PA: Association of Lisp Users, 1995. Presentation prepared and given at the workshop in Cambridge, MA, August 1995.
11. Nancy A. Van House, Mark H. Butler, Virginia Ogle, Lisa Schiff, User-Centered Iterative Design for Digital Libraries: The Cypress Experience, D-lib Magazine, Feb 1996, ISSN 1082-9873.
12. Michelle Baldonado, Chen-Chuan K. Chang, Luis Gravano, and Andreas Paepcke. The Stanford Digital Library Metadata Architecture. International Journal of Digital Libraries, 1(2), February, 1997.
13. D. Cutting, J. Pedersen, and P.-K. Halvorsen. "An Object-Oriented Architecture for Text Retrieval," Proceedings of RIAO'91.
14. M. A. Hearst, TileBars : "Visualization of Term Distribution Information in Full Text Information Access," Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, Denver, CO, ACM, May 1995.
15. G. Kopec, "Document image decoding in the Berkeley digital library project," in Document Recognition III, L. Vincent and J. Hull, editors, Proc. SPIE vol. 2660, pp. 2--13, 1996.
16. G. Kopec and P. Chou, "Document image decoding using Markov source models," IEEE. Trans. Pattern Analysis and Machine Intelligence, vol. 16, no. 6, June, 1994.
17. C. Lynch and H. Garcia-Molina, Interoperability, Scaling, and the Digital Libraries A Report on the May 18-19, 1995 IITA Digital Libraries Workshop Reston, VA, August 22, 1995.
18. J. D. Mackinlay, R. Rao and S. K. Card.

- "An Organic User Interface For Searching Citation Links," Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, Denver, CO, May 1995.
19. P. Pirolli and S. K. Card, "Information Foraging in Information Access Environments," Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, Denver, CO, ACM, May 1995.
20. R. Rao, S. K. Card, W. Johnson, L. Klotz, and R. Trigg, "Protofoil: Storing and Finding the Information Worker's Paper Documents in an Electronic File Cabinet," Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, April 1994.
21. R. Rao, J. O. Pedersen, M. A. Hearst, et al., "Rich Interaction in the Digital Library," Communications of the ACM, 38 (4), 29-39, April 1995.
22. 오극잠 : 고급의방집성, 상해대중서국, 1980
23. 許俊 : 東醫寶鑑, 남산당, 1975
24. 이천 : 의학입문, 남산당, 1991
25. 중국의약대계: 대성출판사, 1995
26. 의부전록 : 대성문화사, 1992
27. 신민교 : 임상본초학, 영림사, 1997
28. 장은암 외 : 장마합주 黃帝內經, 고문사, 1993
29. 고목경차량 : 화한약물학, 일본, 1994
30. 임종국 : 침구치료학, 정보사, 1995
31. 이시진 : 본초강목, 일중사, 1982
32. 정보섭외 : 도해 향약 대사전, 영림사, 1990
33. 중의방제대사전 : 인민위생출판사, 1988
34. 강극명 : 방제대사전, 의성당, 1989
35. 신농본초경, 문광도서유한공사, 1969
36. 중약대사전 : 신문출판공사, 1985
37. 윤용갑 : 동의방제와 처방해설, 의성당, 1996
38. Lawrence A. Rowe, John S. Boreczky, and Charles A. Eads "Indexs for User Access to LargeVideo Database", Symp. on Imageing Sci. Tech., San Jose, CA, Feb. 1994.
39. S. Gubbs, C. Breitender, and Dennis Tsicrizis, " Audio/Video Databases: An Obhected Approach", The 9th Data Enginnering Conf. 1993.
40. HongJiang Zhang, Shuang Yeo Tan, Stephen W.Smoliar, and Gong Yihong "Automatic Parsing and indexing of news video", Multimedia Systems, Springer International, Vol. 2 No. 6, 1995
41. Ron Weiss, Andrzej Duda, and David K. Gifford "Composition and Search with a Video Algebra", IEEE Multimedia, 1995 Spring
42. Yoshinobu Tonomura, Akihito Akutsu, Yukinobu Taniguchi, and Gen Suzuki "Structured Video Computing", IEEE Multimedia, 1994 Fall
43. Andrzej Duda, Ron Weiss, and David K. Gifford "Content-Based Access to Algebraic Video", MIT Laboratory for Computer Science
44. Julio Barros, James French, and Worthy Martin "Indexing multispectral images for content-based retrieval", Univ. of Virginia Technical Report, 1994
45. Sreenivas Gollapudi and Aidong Zhang "NetMedia : A Client-Server Distributed Multimedia Environment", Univ. of NewYork at Buffalo
46. H.M.Vin, A.Goyal, and P.Goyal "Algorithms for Designing Multimedia Servers", In First IEEE Intl. Conf. Multimedia Computing and Systems(ICMCS'94), Boston, 1994
47. Aidong Zhang and Sreenivas Gollapudi

- "Multimedia Transaction Management in DataBases Systems", Technical Report CS-TR-95-44, State Univ. of New York at Buffalo, October 1995
48. S. C Joo, C. M. Krishna, and A. Ganz, "The Optimal Reconfiguration of Distributed Systems based on Ring Networks", Proc. of JTC-CSCC '95, Kumamoto, Japan, 1995,7
 49. 주수종, 정출곤, "전역 메타정보를 이용한 DBGS의 설계 및 구현", 한국정보과학회 데이터베이스 학회 논문, 9권 2호, 93. 8.
 50. 주수종, 이재현, 김명희, 이승용, "점진적 장면 변환의 컷 검출법 분석", 한국정보과학회 전문대학 논문집. Vol. 4, No 2, 96. 12.
 51. 노현철, 이근배, 이종혁, "KQML을 이용한 멀티에이전트 환경하에서의 웹 자원 관리에 이전트 설계", In HCI '97학술대회발표 논문집, 1997.
 52. 강현규, 이창렬, 박세영, 백과사전 검색시스템의 설계 및 구현 한국정보과학회 93 가을 학술발표
 53. 김호성, 고희정, 용어 빈도수를 이용한 영문 문헌정보의 점진적인 개념적 집단화, 한국정보과학회 논문지, Vol. 19, No. 1, January 1992. pp. 12~23.
 54. DIGITAL LIBRARIES INITIATIVE,
<http://www.dli.grainer.uiuc.edu>,
<http://www.dli2.nsf.gov>
 55. 용환승 역, "객체-관계 DBMS", 이한출판사, 1996. pp 30-35.
 56. 문성훈, 정성호, 양재동, "내용 기반 스포츠 뉴스 비디오 시스템 설계 및 구현", 한국정보과학회 봄 학술발표논문집 Vol.26, NO. 1, 1999. pp 125-128.
 57. 심춘보, 송광택, 장재우, "사이버 박물관을 위한 내용 기반 멀티미디어 정보검색 시스템의 설계 및 구현", 정보과학회논문지(C) 제 5권 제 1호, pp 88-98, 1999.
 58. Informix DataBlade Module User's Guide, http://www.informix.com/answers/english/pdf_docs/datablade/39_47.pdf, 4362.pdf.
 59. 이원중, 안길수, 주수종, "웹환경에서 확장된 객체지향형 미들웨어를 이용한 멀티미디어 정보 검색 시스템 구현", 한국정보처리학회 논문지, 제 5권 7호, 1998. 7, pp 1847~1853.
 60. 강현규, 이창렬, 박세영, 백과사전 검색시스템의 설계 및 구현 한국정보과학회 93 가을 학술발표
 61. 김호성, 고희정, 용어 빈도수를 이용한 영문 문헌정보의 점진적인 개념적 집단화, 한국정보과학회 논문지, Vol. 19, No. 1, January 1992. pp. 12~23.