

한글대장경 웹 검색 시스템의 성능 개선

구현우*, 선수림*, 박미화*, 이재수**,
이용규***, 이금석***, 홍영식***, 한보광****

목 차

1. 서 론
2. 한글대장경 전산화 7차 사업
 - 2.1 한글대장경의 입력 · 교정 · 색인작업
 - 2.2 데이터베이스 저장
 - 2.3 웹 검색 인터페이스
 - 2.4 누락문자 관리시스템 확장
3. 결론 및 향후 과제

요 약

본 연구는 한글대장경 전산화 7차 사업으로 한글대장경 30책 분량을 전산화하여 검색 시스템을 구축하는데 목적이 있다. 고려대장경의 우리말 번역본인 한글대장경을 전산화하기 위해 개역된 고문헌을 입력하여 데이터베이스로 구축하고, 인터넷을 통해

* 동국대학교 컴퓨터공학과

** 동국대학교 불교학과, 전자불전 · 문화콘텐츠연구소 전임연구원

*** 동국대학교 컴퓨터공학과 교수

**** 동국대학교 선학과 교수, 전자불전 · 문화콘텐츠연구소 소장

그 내용을 검색할 수 있도록 한다. 한글대장경 고문헌은 확장한자, 누락문자, 특수문자 등을 포함하고 있어서, 본 연구에서는 효과적인 입력과 저장을 위해 유니코드(Unicode)를 사용하며, 유니코드로 표현하지 못하는 문자들은 이미지 폰트를 생성하여 표현한다. 데이터베이스를 구축하기 위해서 DBMS로는 MS-SQL 7.0을 사용하고, 운영체제로는 윈도우 2000 서버를, 웹 서버로는 IIS(Internet Information Server)를 사용하여 검색 시스템을 구축하였다. 또한 다양한 검색 방법을 제공하는 검색 엔진을 개발하여, 유니코드로 저장된 한글대장경 고문헌의 내용을 웹(<http://ebtc.dongguk.ac.kr/>)을 통해 보다 쉽게 전 세계에서 접근할 수 있도록 한다.

1. 서론

본 연구는 한글대장경 전산화 7차 사업으로 한글대장경 30책, 110경을 전산화하여 전 세계에서 활발하게 사용되고 있는 인터넷을 통하여 검색할 수 있도록 하는 것이다.

불법이 인도에서 꽃을 피워 인류의 정신문화를 꽃피워 왔다. 불교의 가르침은 보통 사람들이 구사하는 언어를 통해 전해져왔는데, 초기에는 부처님으로부터 신성한 가르침을 직접 듣는 것이 가능하였고, 입에서 입으로 구전되어 왔다.

부처님의 입멸 후 그러한 가르침의 전통은 인도에서 결집(結集)을 통해 문자화되어 보다 많은 인류를 깨달음의 길로 이끄는 지침이 되었다. 부처님의 가르침은 동아시아의 거의 모든 국가에 전해졌고, 시대와 역사를 초월하여 각국의 찬란한 정신문화를 이끌어 왔다.

불교사는 경전의 역경을 통해 이루어져왔다고 해도 과언이 아니다. 경전의 번역과 수집 및 출간은 국가의 가장 큰 문화사업으로 중요시되어 왔다.

세계문화유산으로 등록된 고려대장경은 몽고의 침입으로 국가가 위기에 처했던 시기에 부처님의 가르침으로 국가의 안녕과 백성의 평안을 기원하기 위해 전 국가적으로 역량을 결집한 우리의 문화유산인 것이다.

조선시대에 이르러서는 훈민정음의 창제로 일반 백성들도 우리나라 말과

글을 널리 사용할 수 있게 되었다. 한문불경을 훈민정음으로 번역해 민간에 널리 유포시키기 위하여 간경도감에서 한글로 된 불경이 제작되기 시작하였다. 이는 지식인만의 불교에서 일체중생을 위한 불교로의 전환을 의미하게 된다. 조선 말기에서부터 가속화된 불경의 한글화는 일제의 강점기에 민족의 정신을 일깨우는 작업으로 진행되어 오늘에 이르게 되었다.

동국대학교의 역경원 설립과 함께 본격화되기 시작한 한글대장경 사업은 현대문명의 발달에 발맞추어 새롭게 전산화의 길을 모색하고 있다. 이는 한글대장경을 디지털화하여 인터넷을 통해 전세계의 인류에게 제공함으로써 시간과 장소를 초월하여 불법의 진리를 홍보하는 것이며, 또한 우리나라의 뛰어난 정신문화를 전세계에 알리는 새로운 전범활동이라고 할 수 있다.

본 연구는 한글대장경 전산화 7차 사업의 성과를 기술하고 그동안 사업의 성과를 토대로 웹 인터페이스 성능을 부가하여 검색 성능을 강화하였다.

2. 한글대장경 전산화 7차 사업

2.1 한글대장경의 입력·교정·색인작업

동국역경원에서 각 분야의 전문적인 능력을 지닌 역경위원들의 엄정한 번역을 담당하고, 그 번역된 결과를 입력하였다. 이에 대해 교정과 윤문을 거쳤고, 3차에 걸쳐서 엄밀한 교정작업을 수행하였다.

한글대장경 전산화 제 7차 사업은 2007년 6월부터 2008년 5월까지 수행하였다. 이번 사업에서 입력교정한 대장경의 목록은 총 30책 분량, 290경으로 다음과 같다. (※ K번호는 고려대장경의 경전고유번호임.)

대각등계집

K.0002 방광반야경(放光般若經)

K.0007 소품반야바라밀경

K.0018 실상반야바라밀경

K.0020 반야바라밀다심경

전자불전 제10집(2008)

- K.0027 아촉불국경
- K.0030 불설포태경
- K.0031 운수사리불토엄정경
- K.0032 법경경
- K.0033 옥가라월문보살행경
- K.0044 불설여환삼매경
- K.0052 불설유일마니보경
- K.0052 불설유일마니보경
- K.0056 대방등대집경
- K.0057 대승대집지장심륜경
- K.0058 대방광심륜경
- K.0059 대승대집경
- K.0063 허공장보살신주경
- K.0067 반주삼매경
- K.0068 불설반주삼매경
- K.0069 발피보살경
- K.0095 불설장엄보리심경
- K.0103 도세품경
- K.0106 불설대반니원경
- K.0107 대반열반경후분
- K.0108 불설방등반니원경
- K.0109 사동자삼매경
- K.0110 대비경
- K.0117 정법화경
- K.0118 청품묘법연화경
- K.0119 불설유마힐경
- K.0121 설무구칭경
- K.0122 불설대방등정왕경
- K.0123 대승정왕경
- K.0124 선사동자경
- K.0138 불필정입정입인경
- K.0140 등집중덕삼매경
- K.0141 집일체복덕삼매경
- K.0142 지심범천소문경
- K.0143 사익범천소문경
- K.0146 지세경
- K.0149 운수사리현보장경
- K.0150 대방광보협경
- K.0161 대승입능가경
- K.0162 불설보살행방편경계신통변화경
- K.0164 대방등무상경
- K.0171 불설보여래삼매경
- K.0176 불설약사여래본원경
- K.0177 약사유리광여래본원공덕경

한글대장경 웹 검색 시스템의 성능 개선(구현우 외)

- K.0178 약사유리광칠불본원공덕경
- K.0181 월등삼매경
- K.0182 불설월등삼매경
- K.0183 불설월등삼매경
- K.0190 후출아미타불계
- K.0193 칭찬정토불섭수경
- K.0194 불설관미륵보살상생도솔천경
- K.0195 불설미륵대성불경
- K.0196 불설미륵래시경
- K.0197 불설미륵하생경
- K.0198 불설미륵하생성불경
- K.0204 손권방편경
- K.0205 낙영락장엄방편품경
- K.0216 불설노여인경
- K.0217 불설노모경
- K.0218 불설노모녀육영경
- K.0219 월광동자경
- K.0220 신일아본경
- K.0221 덕호장자경
- K.0222 문수사리문보리경
- K.0223 가야산정경
- K.0224 불설상두정사경
- K.0225 대승가야산정경
- K.0226 불설장자자제경
- K.0227 불설보살서경
- K.0228 불설서동자경
- K.0231 불설무구현녀경
- K.0232 불설복중녀청경
- K.0233 불설전녀신경
- K.0234 신일경
- K.0241 입법계체성경
- K.0253 아사세왕수결경
- K.0254 채화위왕상불수결호묘화경
- K.0258 여래시교승군왕경
- K.0260 불설간왕경
- K.0264 불설대방등수다라왕경
- K.0265 불설문수사리순행경
- K.0267 불설문수시리행경
- K.0270 불설용시녀경
- K.0275 불설용시보살본기경
- K.0287 불공견삭신변진언경
- K.0354 불설보망경
- K.0356 불설범지녀수의경
- K.0358 사가매경

전자불전 제10집(2008)

- K.0359 보살생지경
- K.0365 불설백불명경
- K.0371 불설칭양제불공덕경
- K.0372 불설수진천자경
- K.0373 마하마야경
- K.0376 관세음보살수기경
- K.0377 불설해룡왕경
- K.0381 부사의광보살소설경
- K.0386 불설초일명삼매경
- K.0389 대위덕다라니경
- K.0390 불설불명경
- K.0391 과거장엄겁천불명경
- K.0392 현재현겁천불명경
- K.0393 미래성수겁천불명경
- K.0394 오천오백불명신주제장멸죄경
- K.0395 불설부사의공덕제불소호념경
- K.0401 불설관불삼매해경
- K.0406 보살종도술천강신모태설광보경
- K.0407 불설홍도광현삼매경
- K.0409 무소유보살경
- K.0411 불설명도오십교계경
- K.0412 문수사리문경
- K.0414 중음경
- K.0415 월상녀경
- K.0415 지인보살경
- K.0417 대승밀엄경
- K.0418 문수사리문보살서경
- K.0420 연화면경
- K.0432 소실지갈라경
- K.0443 아미타고음성왕다라니경
- K.0453 불수반열반약설교계경
- K.0461 불설심명경
- K.0464 불설멸시방명경
- K.0469 불설현수경
- K.0470 불설문수사리반열반경
- K.0480 반니원후관랍경
- K.0483 불림열반기법주경
- K.0484 수지칠불명호소생공덕경
- K.0487 불설견고녀경
- K.0489 유덕녀소문대승경
- K.0491 이구혜보살소문에불법경
- K.0492 차마바제수기경
- K.0494 불설대승유전제유경
- K.0495 불위해룡왕설법인경

한글대장경 웹 검색 시스템의 성능 개선(구현우 외)

- K.0497 사자장엄왕보살청문경
- K.0499 불설대의경
- K.0500 우바이정행법문경
- K.0501 불설금강삼매본성청정불괴불멸경
- K.0502 불설사자월불본생경
- K.0503 불설살라국경
- K.0505 수제가경
- K.0511 불설장자법지처경
- K.0514 불설장자녀암제차사자후요의경
- K.0518 불설당래변경
- K.0520 불설법멸진경
- K.0529 불장경
- K.0537 불설문수사리정를경
- K.0538 불설문수회과경
- K.0544 삼만타발타라보살경
- K.0548 십선업도경
- K.0549 대지도론[1-40]
- K.0562 승사유범천소문경
- K.0591 섭대승론
- K.0592 섭대승론본
- K.0593 섭대승론석론
- K.0594 섭대승론석
- K.0657 불개해범지아발경
- K.0759 불설마등녀경
- K.0761 불설아난문사불길흥경
- K.0768 오모자경
- K.0770 불설아난분벌경
- K.0776 불설장자음열경
- K.0782 불설내녀기역인연경
- K.0784 불설칠녀경
- K.0785 불설병사왕오원경
- K.0786 불설팔사경
- K.0787 불설유리왕경
- K.0791 아사세왕문오역경
- K.0794 월난경
- K.0796 불설정반왕반열반경
- K.0799 생경
- K.0802 불본행집경[1-60]
- K.0806 불설대안반수의경
- K.0807 음지입경
- K.0814 선행법상경
- K.0815 불설장자자오뇌삼처경
- K.0816 불설수마제장자경
- K.0817 불설아난사사경

전자불전 제10집(2008)

- K.0819 불설흑씨범지경
- K.0821 불설미생원경
- K.0822 불설아구류경
- K.0824 불설분별경
- K.0826 건타국왕경
- K.0831 불설대가섭본경
- K.0834 가조아나함경
- K.0836 아난칠몽경
- K.0837 불설부인우고경
- K.0839 불설마하가섭도빈모경
- K.0843 불대승대경
- K.0844 야기경
- K.0845 불설마달국왕경
- K.0846 불설전다월국왕경
- K.0848 불설말라왕경
- K.0849 불위연소비구설정사경
- K.0850 라운인육경
- K.0851 불설사갈비구공덕경
- K.0854 변의장자자경
- K.0855 무구우바이문경
- K.0859 노지장자인연경
- K.0863 불설오왕경
- K.0869 장조범지청문경
- K.0873 비구청시경
- K.0874 신관경
- K.0877 보달왕경
- K.0878 불멸도후관령장송경
- K.0879 불설범마난국왕경
- K.0884 불설손다야치경
- K.0886 비구피녀악명욕자살경
- K.0887 불설선행삼십칠품경
- K.0909 사분율비구계본
- K.0922 담무덕부사분율산보수기갈마
- K.0956 아비달마순정리론 제33권-제80권
- K.0957 아비달마장현종론 제12권-제40권
- K.0960 잡아비담심론
- K.0986 도지경
- K.0991 좌선삼매경
- K.0994 가섭부불반열반경
- K.0996 보살가색욕법경
- K.0997 불입열반밀적금강역사애련경
- K.1004 선법요해
- K.1008 법관경
- K.1011 불설내신관장구경

한글대장경 웹 검색 시스템의 성능 개선(구현우 외)

- K.1012 선요경
- K.1013 아육왕경
- K.1028 소도지경
- K.1030 육보살역당송지경
- K.1031 찬관세음보살송
- K.1071 대당대자은사상장법사전
- K.1075 속고승전 제1권-제30권
- K.1081 광홍명집 제16권-제30권
- K.1082 남해기귀내법전
- K.1088 불설대승장엄보왕경
- K.1090 불설성불모소자반야바라밀다경
- K.1095 대방광총지보광명경
- K.1099 불설대승선견변화문수사리문법경
- K.1103 금요동자경
- K.1140 불위사가라용왕소설대승경
- K.1145 대승보월동자문법경
- K.1146 불설일체여래안상삼매의계경
- K.1153 불설보장신대명만나라의계경
- K.1154 불설성보장신의계경
- K.1155 일체여래대비밀왕미증유최상미묘대만다라경
- K.1156 불설대마리지보살경
- K.1171-2 불설치성광소재경
- K.1174 금광왕동자경
- K.1187 불설성다라보살법찬
- K.1190 불설제석반야바라밀다심경
- K.1192 불설오십송성반야바라밀경
- K.1194 불설제불경
- K.1197 일체여래설불정륜왕일백팔명찬
- K.1199 불설대승무량수장엄경
- K.1210 불설환화망대유가교심분노명왕대명관상의계경
- K.1211 불설묘길상보살소문대승법라경
- K.1212 불설팔대보살경
- K.1220 불설유가대교왕경
- K.1227 보수보살보리행경
- K.1246 불설선락장자경
- K.1250 불설결정의경
- K.1269 천수천안관세음보살대신주본
- K.1278 금강공포집회방광계의관자재보살삼세최승심명왕경
- K.1279 출생무변문다라니경
- K.1283 백천송대집경지장보살청문법신찬
- K.1309 문수문경자모품제심사
- K.1357 불위우전왕설왕법정론경
- K.1381 대승이취육바라밀다경
- K.1382 대화엄장자문불나라연력경

전자불전 제10집(2008)

- K.1395 근본설일체유부비나야갈치나 의사
- K.1416 불설제석암비밀성취의계
- K.1417 불설승군왕소문경
- K.1427 불설성불모반야바라밀다경
- K.1437 광명동자인연경
- K.1448 불설초분설경
- K.1454 불설비밀삼매대교왕경
- K.1456 불설성관자재보살불공왕비밀심다라니경
- K.1457 불설시일체무외다라니경
- K.1469 불설정생왕인연경
- K.1472 불설수용존자경
- K.1476 불설제개장보살소문경
- K.1479 불설대승대방광불관경
- K.1483 금색동자인연경
- K.1486 불설여래부사의비밀대승경
- K.1498 일체경음의 제1권-제10권
- K.1506 대방광불화엄경수현분제통지방계
- K.1507 십구장원통기
- K.1508 석화엄지귀장원통초
- K.1509 화엄경삼보장원통기
- K.1513 화엄경탐현기 제1권-10권

2.1.2. 태그 작업

입력과 교정을 마친 30책의 한글대장경은 본 연구소 입력팀에서 페이지·대제목·소제목·해제·서론·각주·진언이미지에 대하여 각각 태깅 작업을 수행하였다.

- 1) 페이지 태그작업 : 페이지를 검색하여 해당 원문을 보여주었다.
페이지의 태그는 ‘<PAGE PAGENUM='경번호-10001'/>’로 하였다.
- 2) 제목 태그작업 : 경전의 대제목과 소제목을 검색할 수 있으며, 경전의 제목을 통하여 해당 원문을 확인할 수 있게 하였다.
경 제목의 태그는 ‘<JMOK1 SEARCH='TRUE'>경 제목</JMOK1>’로 하였다.
경의 소제목을 경우 태그는 ‘<JMOK2>경 소제목</JMOK2>’과 그 이하의 소제목은 ‘<JMOK3>경 소제목</JMOK3>’로 하였다.
- 3) 각주 태그작업 : 한글대장경의 각주에 나타나 있는 원문을 확인할 수 있도록 하였다.

각주 태그는 원문의 각주는 ‘<COM NUM='1'/>’로 하며, 각 해당 페이지의 하단에

‘<COMMENT> 각주의 내용</COMMENT>’으로 입력하였다.

- 4) 진언 이미지 태그작업 : 한글대장경에 나타난 진언은 이미지로 처리한 후 이미지로 확인할 수 있도록 하였다.

진언 이미지의 태그는

‘<JIN 경 번호-페이지-페이지의 진언 일련번호>

진언내용 </JIN>’로 하였다.

2.2 한글대장경 데이터베이스 관리 시스템의 개선

한글대장경 데이터베이스 시스템은 한글대장경을 데이터베이스로 저장하고 관리하는 시스템이다. 데이터베이스 관리 시스템은 한글대장경을 저장하기 위해 원문을 구별해주는 각 태그들의 유효성을 검증한 후 원문으로부터 제목, 원문 내용, 키워드를 추출하여 유니코드로 변환하고 해당 테이블에 값을 저장하는 작업을 진행한다. 또한 데이터베이스 저장과 관리를 위해 원문을 검색하고 수정·삭제하는 기능을 포함한다.

한글대장경 제7차 사업에서는 새로운 한글대장경을 데이터베이스로 구축하였으며 기존에 구축한 데이터베이스에 대한 재검증 작업을 진행하였다. 또한 기존에 사용하던 데이터베이스 자동 입력 프로그램을 개선하였다. 이를 좀 더 자세히 기술하면 다음과 같다.

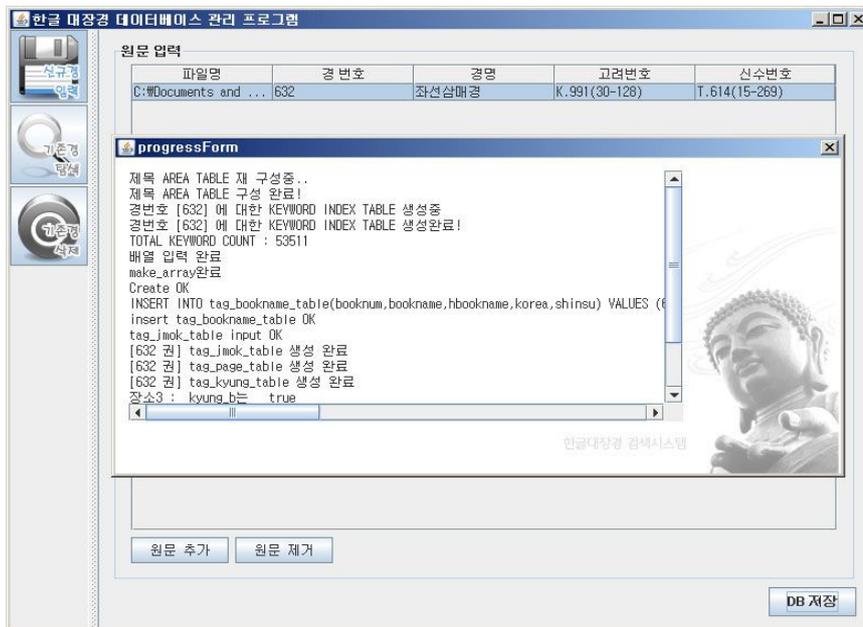
첫째, 새로운 한글대장경 191경을 데이터베이스로 구축하였다. 현재 한글대장경 사업을 통해 구축된 데이터베이스에는 한글대장경 655경이 포함되어 있다.

둘째, 기존에 구축한 한글대장경 원문을 대상으로 전체적인 재검증 작업을 수행하였다. 재검증 작업을 통하여 오류가 있었던 44경의 한글대장경 원문을 수정하였다. 기존 경의 재검증 작업은 다음과 같이 수행되었다.

- ① 누락된 고려·신수번호의 보완 작업

- ② 특수 기호의 깨짐 현상 보정
- ③ 잘못된 경명으로 인한 검색 오류 보완 작업
- ④ 중복된 경과 누락된 페이지 정정 작업
- ⑤ 누락 및 중복 페이지 내용 수정

셋째, 데이터베이스 입력 작업 개선을 위해 데이터베이스 자동 입력 프로그램을 재검증 및 수정하였다. 이를 통해 [그림 1]과 같은 화면으로 구성된 데이터베이스 자동 입력 프로그램을 작성하였다.



[그림 1] 한글대장경 데이터베이스 관리 프로그램

넷째, 한글대장경 자연어 검색 기능을 지원하기 위해 전문 검색용 카탈로그 테이블을 데이터베이스에 추가하였다. 한글 대장경 원문을 저장한 edoctable에는 Unicode로 변환된 대장경 내용이 입력되어 있어 전문 검색(Full Text Search)을 지원하는데 어려움이 있었다. 이에, 별도의 전문 검색용 카탈로그 테이블을 추가함으로써 전문 검색 기능을 지원 하도록 하였으며, 이로 인해 자연어 검색 속도가 개선되는 효과를 얻을 수 있었다.

2.2.1 한글대장경 데이터베이스 구축 단계

한글대장경의 원문은 제목, 원문, 주석 등으로 구성되어 있고, 각각에 해당되는 내용은 태그로 구별하여 데이터베이스를 구축한다. 원문에 나타나는 한자는 기존 문자 집합으로 표현하는데 한계가 있어 유니코드로 변환하여 저장한다. 한글대장경 데이터베이스를 구축하기 위해 먼저, 원문을 구별해주는 각 태그들의 유효성을 검증하는 작업을 진행한다. 유효성 검증 작업 후 원문으로부터 제목, 원문 내용, 키워드를 추출하여 유니코드로 변환한 후 그 값을 해당 테이블에 저장한다.

(1) 태그가 삽입된 원문의 유효성 검증 작업

텍스트 파일로 변환된 원문에는 제목, 페이지, 주석, 진언 이미지 등을 구별하기 위하여 각각 <JMOK>, <PAGE>, <COMMENT>, 라는 태그들을 삽입한다. 이러한 태그들은 여는 태그(<...>)와 닫는 태그(</...>)가 쌍으로 구성되어야한다. 태그가 잘못 작성되면 잘못된 데이터가 데이터베이스에 입력될 수 있으므로 유효성 검증 작업이 필요하다. 원문의 유효성 검증 작업은 다음과 같은 순서로 이루어진다.

- ① “*.txt”로 저장된 원문 파일을 “*.xml”로 확장자명을 변경한다.
- ② 웹 브라우저에서 해당 XML 문서를 불러들인다.
- ③ 웹 브라우저에 에러 메시지가 나타나지 않으면 유효한 문서이고, 에러 메시지가 나타나면 해당되는 내용을 찾아 원문을 수정한다.

(2) 키워드 인덱스 생성 및 저장

키워드 추출 및 저장 단계에서는 원문 내에서 키워드로 지정된 단어를 찾아 그 위치와 단어를 keyword_index 테이블에 저장한다. 미리 지정된 키워드 목록 정보는 keyword 테이블에 유니코드로 저장되어 있다. 유니코드를 검색하기 위해서는 기존의 아스키파일과 달리 2바이트 단위로 문자를 비교해야 한다. 또한 유니코드 파일의 처음이 16진수 “FF FE”임에 유의해야 한

다. 키워드 인덱스 구축에는 원문이 저장된 “edocdata”와 미리 지정된 키워드 정보를 갖는 “keyword” 테이블이 사용된다. 키워드 인덱스를 구축하는 절차는 다음과 같다.

- ① 키워드 테이블에서 하나의 키워드를 추출하고, “edocdata” 테이블의 전체 내용과 비교한다.
- ② “keyword” 테이블에서 선택한 키워드와 같은 코드를 발견하면, 일련 번호(uid), 키워드번호(keynum), 키워드(keyword), “edocdata” 테이블에서의 페이지(pagenum) 정보를 keyword_index 테이블에 저장한다.
- ③ “edocdata” 테이블의 모든 레코드를 비교한 다음에 다음 키워드를 읽어 키워드 비교 과정을 반복 수행한다.
- ④ “keyword” 테이블 전체를 비교했을 때, 프로그램을 종료한다.

(3) 제목 인덱스 구축

원문에 <JMOK>과 </JMOK> 태그가 나타나면 해당 제목의 레벨과 위치를 “tag_jmok_area_table” 테이블에 저장한다. 다음과 같이 제목 태그는 트리 구조의 형태를 갖는다. <JMOK 다음에 나타나는 숫자인 1, 2, 3, 4는 제목의 레벨을 나타내고, <JMOK4> 태그의 속성인 “SEARCH=TRUE”는 경 제목을 의미하며 그 값은 “tag_kyung_table”에 저장한다.

```
<JMOK1> ... </JMOK1>
<JMOK2> ... </JMOK2>
<JMOK3> ... </JMOK3>
<JMOK4 SEARCH=TRUE> ... </JMOK4>
```

(4) 원문 저장

원문 저장은 유니코드 편집기에서 작성된 유니코드 원문을 테이블에 저장한다. 원문 파일을 줄 단위로 읽어 “edocdata” 테이블에 저장한다. “edocdata” 테이블에는 페이지 당 라인(line)수와 단 번호 등의 부가 정보를 저장한다. 부가 정보는 원문에 대한 인덱스 역할을 한다. “ncontinue”는 원문의 라인이 계속된다는 정보이다. 원문을 저장할 때 원문에서 한 라인이 레코

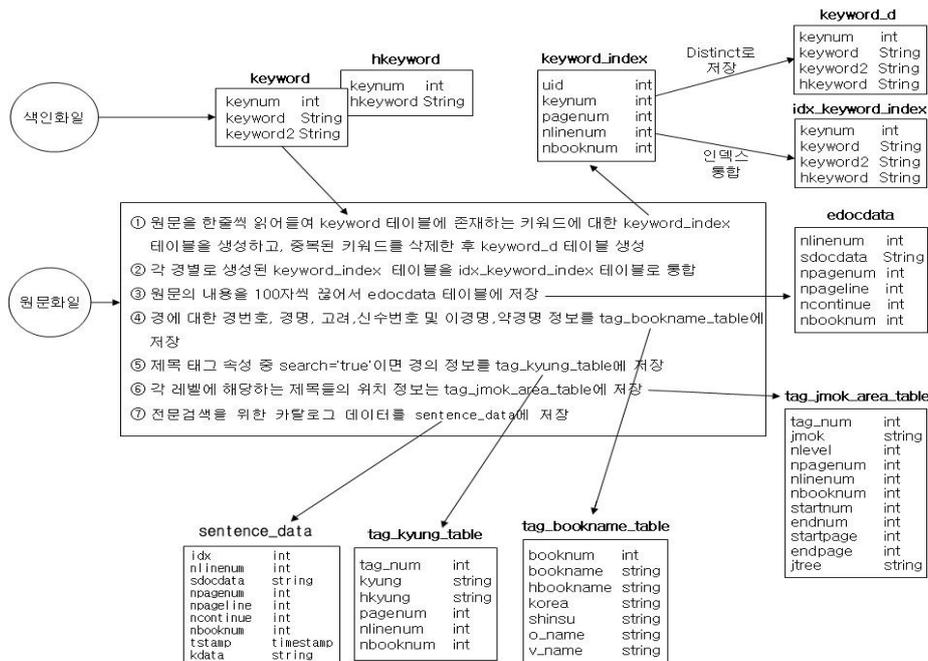
드의 저장크기를 초과할 경우에 100자 단위로 나눠서 저장하게 되는데 이때 라인이 이어짐을 “ncontinue”에 표현한다.

(5) 카탈로그 데이터 생성 및 저장

카탈로그 데이터는 한글대장경 자연어 검색 기능을 지원하기 위한 데이터이다. 한글 대장경 원문을 저장한 edoctable에는 Unicode로 변환된 대장경 내용이 입력되어 있어 전문 검색(Full Text Search)을 지원하는데 어려움이 있었다. 이에, 별도의 전문 검색용 카탈로그 테이블인 sentence_data 테이블을 추가함으로써 전문 검색 기능을 지원하도록 하였으며, 이로 인해 자연어 검색 속도가 개선되는 효과를 얻을 수 있었다.

2.2.2 한글대장경 데이터베이스 구축

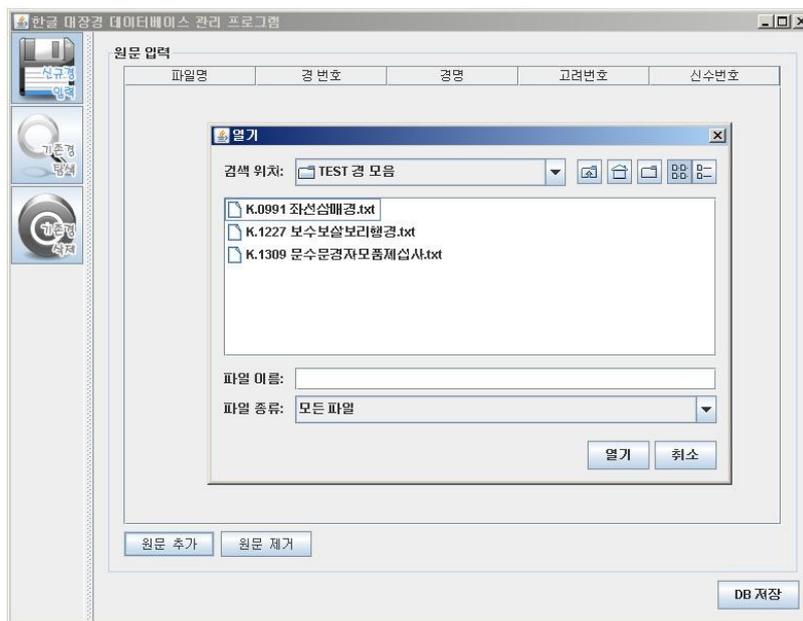
데이터베이스 구축에 사용된 DBMS는 마이크로소프트 SQL Server 7.0을 사용하였다. [그림 2]는 색인 파일과 원문 파일을 이용하여 각각의 테이블을 생성하는 방법과 테이블 구성정보를 보이고 있다.



[그림 2] 테이블 생성 방법

테이블 생성 과정을 요약하면 다음과 같다.

- 색인 파일을 읽어서 “keyword”와 “hkeyword” 테이블을 만든 다음 원문에서 읽어 들인 한 줄에 keyword 테이블의 키워드가 존재하면 그 키워드는 “keyword_index” 테이블에 저장한다.
- 원문은 100자씩 끊어서 “edocdata” 테이블에 저장한다.
- 고려·신수번호와 이경명·약경명 정보는 “tag_bookname_table”에 저장한다.
- 제목 태그 중 “SERCH=“TRUE”” 인 제목은 “tag_kyung_table” 테이블에 저장한다. 각 제목 태그 정보는 각 레벨 정보와 위치 정보를 “tag_jmok_area_table”에 저장한다.



[그림 3] 한글대장경 데이터베이스 관리 프로그램

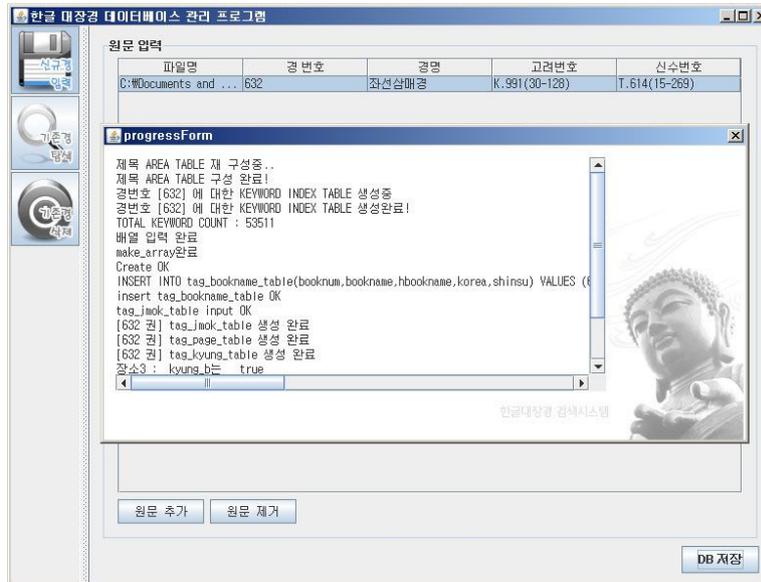
2.2.3 한글대장경 데이터베이스 관리 시스템

한글대장경 데이터베이스 관리 시스템은 한글대장경 원문의 내용을 입력·수정·삭제하는 기능을 지원하며, 원문을 부분적으로 검색하여 수정할 수 있

한글대장경 웹 검색 시스템의 성능 개선(구현우 외)

도록 한다. 원문을 입력하기 위해서는 [그림 3]의 중간에 있는 ‘원문 추가’ 버튼을 눌러 ‘원문 입력 목록’에 추가한 뒤, ‘DB 저장’ 버튼을 누른다.

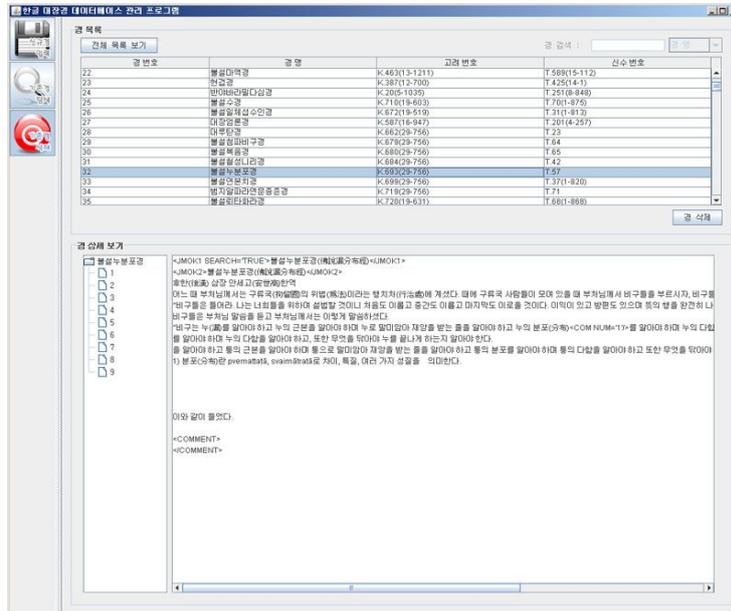
원문이 저장되는 과정은 다음 화면을 통해 확인할 수 있다.



[그림 4] 원문 저장 과정 1.

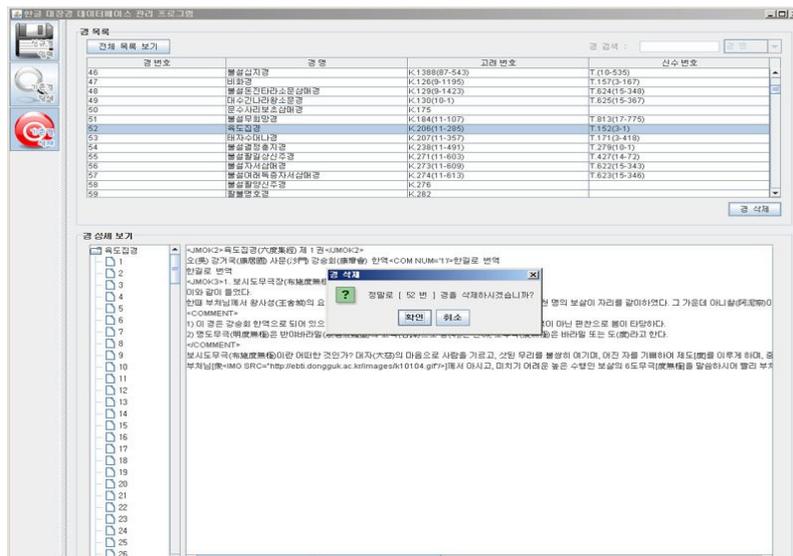
기존 프로그램에서 ‘제목 인덱스 생성’ 버튼을 눌러 각 레벨에 해당하는 제목들의 위치 정보를 갖는 제목 인덱스가 생성했던 과정은 ‘DB저장’과정속에 포함되었다. 즉, 원문 입력이 끝난 후 자동적으로 ‘고려,신수번호’ 및 ‘이경명,약경명’이 제목 인덱스 테이블에 입력된다. 데이터베이스 저장 및 관리 시스템에는 또한 저장된 내용을 검색하거나 수정할 수 있는 기능이 제공된다. 다음 화면에서 검색 및 수정하고자 하는 경의 이름을 입력한 뒤에 오른쪽 상단에 있는 ‘검색’ 버튼을 누르면 해당 제목을 포함한 모든 경의 원문을 확인할 수 있다.

전자불전 제10집(2008)



[그림 5] 원문 저장 과정 2.

원문 삭제를 위해서는 기존 경 이름을 검색한 뒤, 오른쪽 중간에 있는 ‘경 삭제’ 버튼을 눌러 기존의 원문 전체를 삭제한다.



[그림 6] 원문 저장 과정 3.

2.3 웹 검색 인터페이스

한글대장경 웹 검색 인터페이스는 사용자가 웹을 통하여 한글대장경을 보다 편리하게 검색할 수 있도록 다양한 검색 방법을 제공하고 있다. 검색 방법은 크게 경명검색, 용어검색, 쪽수검색으로 구성되어있으며 각각의 검색 방법으로 검색한 결과가 한글대장경의 어느 부분에 속하는지를 쉽게 알 수 있도록 위치 정보를 제공하고 있다. 또한 사용자가 한글대장경을 검색한 후 그 페이지에 해당하는 고려대장경과 신수대장경의 권수와 페이지 정보를 제공하고 있다.

또한 사용자 접근성을 높이기 위한 검색 사용자 인터페이스 및 기능을 지속적으로 개선하였다. 개선된 검색 사용자 인터페이스 및 기능은 첫째로 용어검색에서 통합되어 있던 전체 용어 검색과 상세 용어 검색 기능을 분리하여 전체 용어 검색 대신에 전체 본문에 대한 문장 검색 기능을 추가하여 기존에 수동으로 작업한 색인화된 DB를 검색하지 않고 DBMS의 전문 검색 서비스를 이용하여 검색 속도와 정확도를 높였다. 둘째로 상세 검색을 분리함으로써 다양한 검색 조건을 검색하는 과정에서 발생하는 오류를 줄이고, 속도 및 정확도를 향상시켰다. 셋째로 본문 내용의 보안성을 높이기 위해 마우스 오른쪽 버튼 및 복사를 위한 기타 키 입력을 방지하는 기능을 모든 본문 페이지에 적용하였다. 마지막으로 검색을 위해 입력했던 용어를 검색 결과에 보다 정확하게 표시하고, 사용자의 가독성을 높이는 표현방식으로 기능을 개선하였다.

웹 검색 인터페이스의 주요 기능에 대한 자세한 내용은 본문을 통하여 살펴보도록 한다.

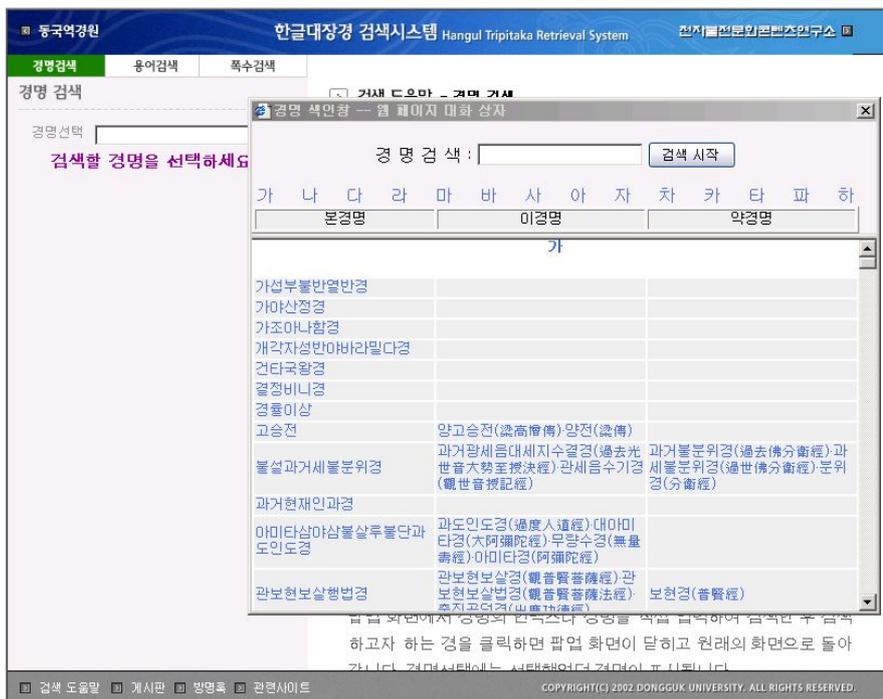
2.3.1 웹 검색 인터페이스 주요 기능

한글대장경 웹 검색인터페이스는 사용자가 편리하게 검색 할 수 있도록 여러 가지 검색 방법을 제공하고 있다. 주요 검색 방법은 경명검색,

용어검색, 그리고 쪽수검색 등이 있다. 보다 자세한 내용은 다음과 같다.

(1) 경명검색

경명검색은 [그림 7]과 같이 경명 목록에서 특정 경명을 선택하여 그 경에 대한 내용을 검색하는 방법이다. 한글대장경의 많은 경전들은 본경명(本經名) 뿐만 아니라 이경명(異經名)과 약경명(略經名)도 가지고 있다. 본경명은 대장경을 잘 아는 전문가 위주의 경명이고, 이경명이나 약경명은 일반 사용자에게 친숙한 경명이다. 본 검색시스템은 본경명은 물론 이경명이나 약경명으로도 쉽게 검색 할 수 있는 서비스를 제공하고 있다. 이와 같은 서비스제공으로 사용자는 보다 편리하게 이경명이나 약경명으로도 검색할 수 있다.



[그림 7] 경명 선택을 위한 경명 색인창 화면

(2) 용어검색

본 검색 시스템에서는 용어검색을 위하여 각 경전별로 자주 검색하는

한글대장경 웹 검색 시스템의 성능 개선(구현우 외)

용어, 약 5만여 단어를 선정하여 빠른 검색이 가능하도록 색인화 하였다. 또한 사용자의 다양한 검색 요구에 맞도록 단순히 검색어만을 입력받아 전체 본문 내용과 비교하여 검색하는 전체문장검색과 다양한 옵션을 추가하여 검색하는 경별상세검색 기능으로 나뉘어져 있다.



[그림 8] 전체 경을 대상으로 ‘광대한’을 검색한 화면

[그림 8]의 전체문장검색은 검색어를 입력하고 검색을 하면 전체 경에서 해당되는 문장이 일치하는 본문 내용을 모두 검색한다. 기존의 문장 검색 기능은 미리 색인화된 용어로만 검색이 가능하였다. 그러나 사용자의 검색 요구를 수용하기에는 색인화 작업에 한계가 있으므로, 전체 본문에 대한 문장 검색 기능이 계속적으로 요구되어 왔다. 이를 해결하기 위해 655개의 경을 모두 통합하는 DB를 생성한 후 DBMS 자체의 색인 기법을 이용하여 본문을 직접 검색할 수 있도록 구성한 후 전체문장검색 기능을 구현하였다. 입력한 검색어를 본문과 직접 비교하여 검색하기 때문에 정확한 검색이 가능하며, DBMS의 전문 검색 서비스를 이용하여 속도를 향상시켰다.

경별상세검색은 경 선택 후 용어를 입력하면 상세검색을 하게 되는데 이때 ‘단어검색/불리언검색’, ‘한자검색’, ‘결과내검색’, ‘자연어검색’, 그리고 ‘문장검색’ 등의 검색 옵션을 지원하고 있다. [그림 9]와 같이 상세검색은

크게 ‘입력검색’과 ‘목록검색’으로 나눌 수 있다. 먼저, ‘입력검색’은 경전을 선택한 후 ‘용어입력’ 상자에 검색할 용어를 직접 입력하여 검색한다.



[그림 9] 경별상세검색에서 ‘가루라’ 용어로 검색한 화면

용어를 직접 입력하는 ‘방법과 한글자음을 선택하는 방법 이외에도 ‘단어/불리언검색’, ‘결과내검색’, ‘한자검색’, ‘자연어검색’, 및 ‘문장검색’ 등을 제공하고 있다. 각 검색 방법에 대한 자세한 설명은 [표 1]에 나타내었다.

[표 1] 다양한 용어 검색 방법 및 기능

검색 방법	기능
단어/불리언검색	AND(&), OR(!), NOT(!) 연산자를 조합하여 용어를 입력하여 검색을 할 수 있어 다양한 조건의 검색이 가능하다.
결과내검색	이미 검색한 결과와 찾고자하는 결과가 관련이 있는 경우, 검색한 결과 내에서의 재검색은 빠르고 정확한 검색을 가능하게 한다.
한자검색	독음을 이용한 용어 검색은 음은 같지만 한자가 다른 용어가 있기 때문에, 정확한 검색을 위해 한자를 이용한 한자 검색 방법이 필요하다. ‘한자 검색’ 옵션 버튼을 선택한 후 용어를 입력하면 한자와 한글독음이 함께 저장된 키워드

한글대장경 웹 검색 시스템의 성능 개선(구현우 외)

	테이블(keyword_table)에 접근하여 독음에 해당하는 한자를 검색한다.
자연어검색	문장 검색과 달리 입력된 문장을 형태소 분석을 통해 검색 가능한 용어를 추출하고 경 단위로 입력한 단어를 유니코드로 변경한 후 유니코드의 패턴매칭을 통해 추출된 용어들을 검색한다.
문장검색	위의 용어 검색 방법에서는 용어 선택이나 용어 입력에 의한 검색만이 가능 하였지만 문장 검색에서는 입력한 문장 전체가 포함된 페이지를 검색하는 것이 가능하다.

(3) 쪽수검색



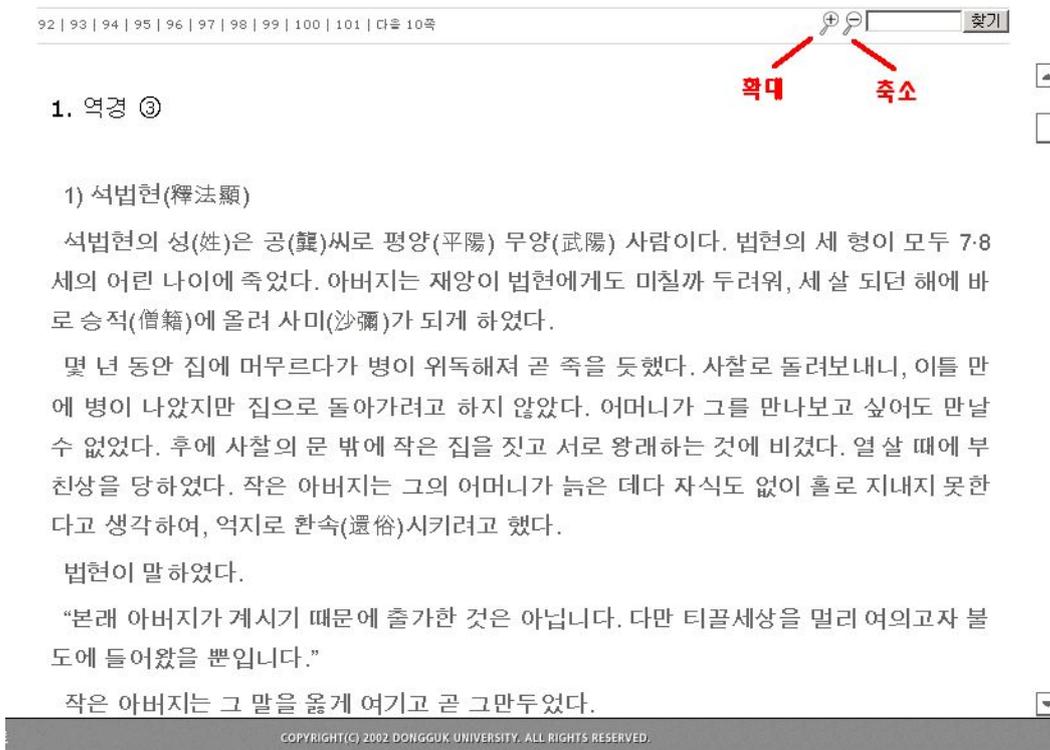
[그림 10] 쪽수 검색의 결과 화면

[그림 10]의 쪽수검색은 선택한 경의 쪽을 입력하면, 해당 쪽으로 바로 이동하는 검색 방법이다. 선택한 경에 따라 본문, 해제, 그리고 서문 정보가 나타나는데, 사용자는 경에 따라 이들 옵션 단추를 선택할 수 있다.

(4) 본문 내용 확대/축소 기능

검색된 본문 화면 크기를 4단계까지 확대할 수 있으며 다시 축소 할 수 있다. [그림 11]과 같이 확대/축소 아이콘을 상단 프레임에 배치하여 사용자가 쉽게 접근할 수 있도록 하였다. 확대/축소 아이콘을 클릭하게 되면

텍스트 뿐만 아니라 이미지도 동일한 비율로 확대 혹은 축소된다.

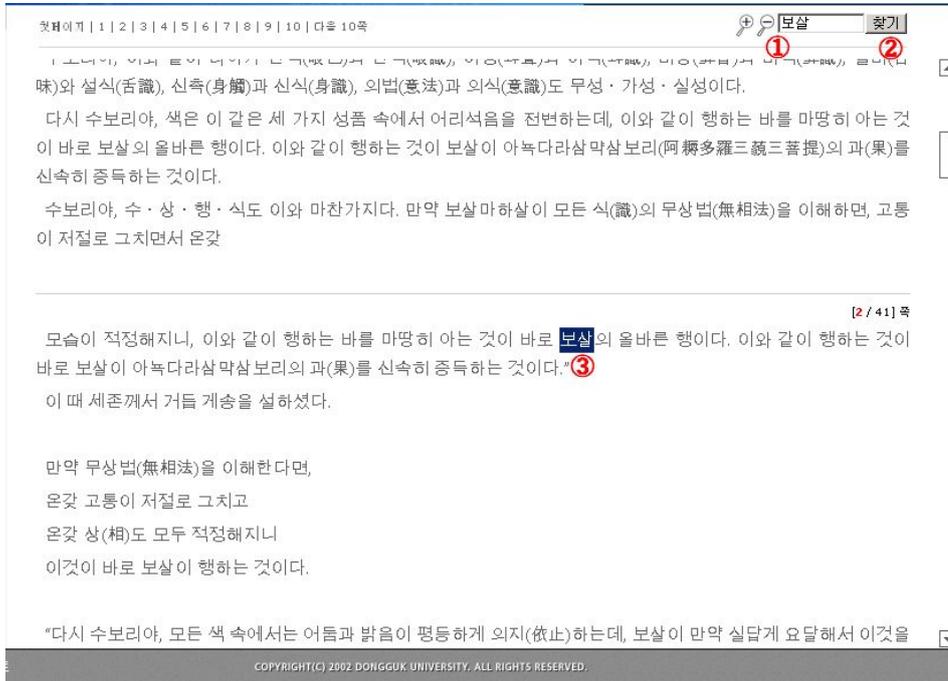


[그림 11] 본문 화면을 1단계 확대

(5) 본문 내 문자열 찾기 기능

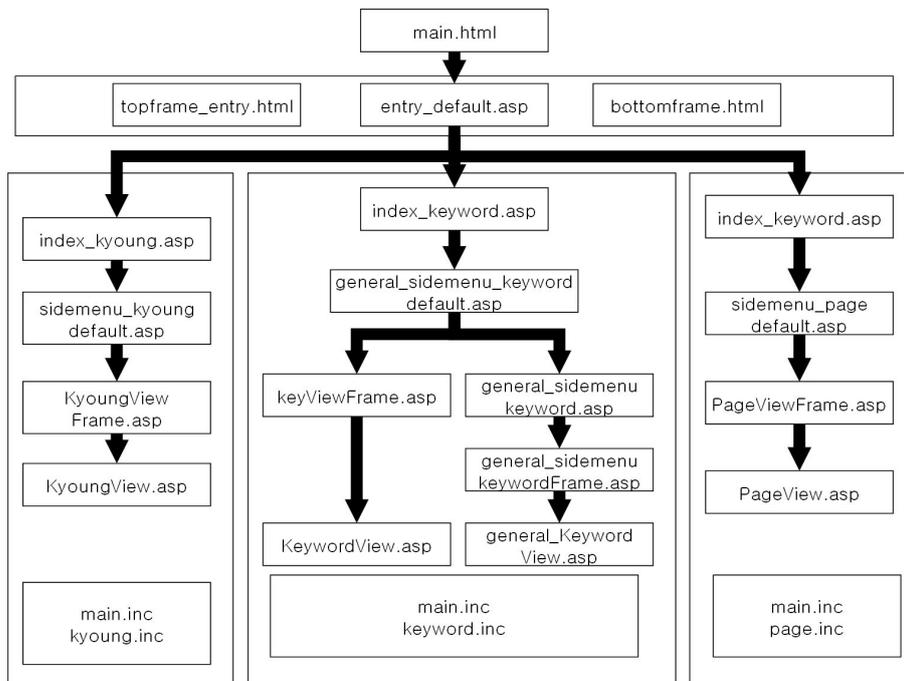
검색된 본문 페이지 내에서 원하는 정보를 바로 찾아 볼 수 있도록 문자열을 입력하면 일치하는 문자열 위치로 이동할 수 있다. [그림 12]와 같이 텍스트 박스에 문자열을 입력 한 후, 찾기 버튼을 클릭 한다. 일치되는 문자열이 있으면 화면이 해당 위치로 이동하고, 일치되는 문자열이 반전된다. 찾기 버튼을 클릭할 때마다 일치하는 문자열로 이동한다.

한글대장경 웹 검색 시스템의 성능 개선(구현우 외)



[그림 12] 본문화면에서 본문 내 검색

2.3.2 웹 검색시스템 구현



[그림 13] 웹 검색시스템의 제어 흐름도

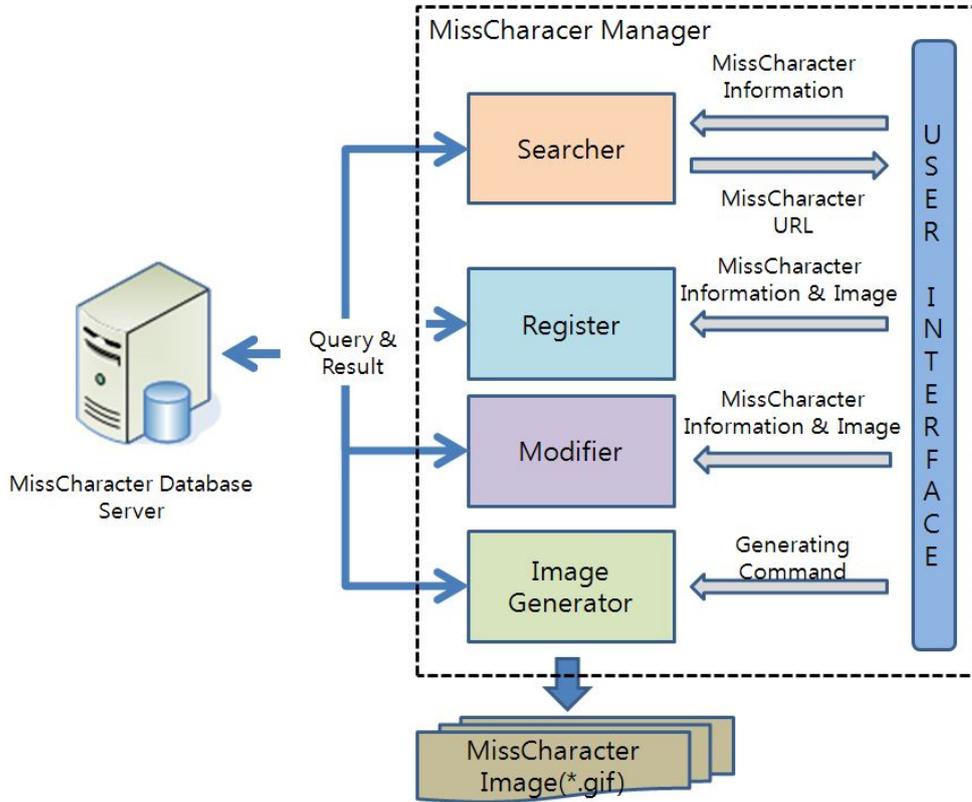
[그림 13]은 한글대장경 웹 검색시스템의 전체적인 제어 흐름을 나타낸 것이다. 한글대장경 웹 검색시스템을 접속하면 main.html이 호출된다. 이것은 세 개의 틀인 topframe-entry.html, entry_default.asp, bottom-frame.html으로 구성되어 있다. 먼저, 상위 틀인 topframe_entry.html은 전자불전연구소와 동국역경원을 링크하도록 구성되어 있다. 다음, 중간 틀인 entry_default.asp는 한글대장경에 대한 간단한 설명과 '검색시작' 단추로 구성되어 있다. 마지막으로, 하위 틀인 bottomframe.html은 검색도움말, 게시판, 방명록, 관련사이트 링크로 구성되어 있다.

사용자가 본 검색시스템의 시작화면에서 검색시작 단추를 누르면, 기본적으로 경명 검색 페이지로 이동한다. 용어 검색과 쪽수 검색으로 이동하기를 원할 때는 해당 탭을 누르면 된다. 사용자가 원하는 검색 결과를 얻는 과정은 다음과 같다. 먼저, 사용자가 해당 검색 페이지에서 검색을 요청하면 본 검색시스템이 사용자의 검색 요청을 질의문으로 변경한다. 그 다음, 원문 저장 데이터베이스에 질의하면 그 결과를 사용자에게 보여준다.

2.4 누락문자 관리 시스템 확장

누락문자를 사용자에게 서비스하기 위해서는 이미지화 과정을 통해 사용자가 문서를 검색하였을 경우 이질감 및 원문과의 통일성을 부여하여야 한다. 따라서 누락문자 관리기는 문자의 등록 및 이미지 태그 입력 기능을 포함하고 있다. 이번 연구에서는 기존 누락문자 관리 프로그램의 기능 확장을 통하여 보다 효율성을 높이고자 한다.

[그림 14]는 누락문자 관리기의 전체 구조도이다. 누락문자 관리기는 크게 누락문자 검색, 등록, 수정, 이미지 생성 부분으로 나뉜다. 이번 연구에서는 누락문자 관리기의 4가지 구성요소중 이미지 생성 요소의 수정을 통해 이미지 생성의 오류율을 줄이고 사용자의 편의성을 증대시키는 방향으로 연구가 진행되었다.



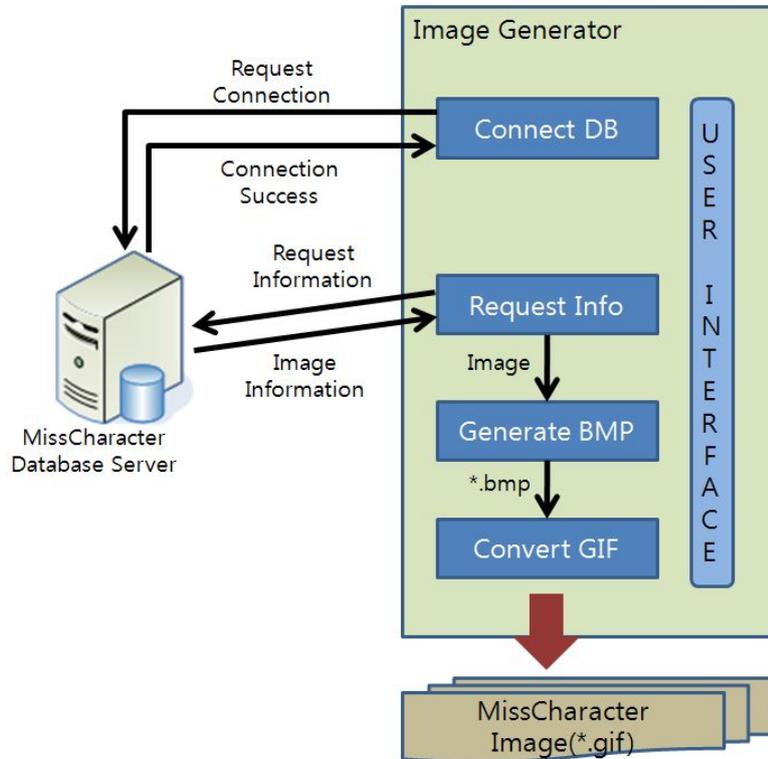
[그림 14] 누락문자 관리기 전체 구조도

2.4.1 누락문자 이미지 생성 프로그램 개선

누락문자 관리기를 통해 등록된 누락문자들은 비트맵 형식의 이미지를 바이너리 데이터 포맷으로 변환되어 저장된다. 이 형식을 사용하면 속도는 빠르지만 웹상에서 바로 게시가 불가능하고 저장해야 할 파일의 용량이 상당히 커지므로 저장 공간의 낭비와 웹 브라우저에서 출력 시 읽어 들이는데 많은 시간을 필요로 하게 된다. 따라서 비트맵 형식의 이미지 파일을 GIF 형식의 이미지 파일로 변환하여 저장 공간에 저장을 한다.

[그림 15]는 누락문자 데이터베이스에 저장된 누락문자 이미지 정보를 GIF 형식의 파일로 변환하는 과정을 간략히 도식화 한 것이다.

이전의 연구에서는 비트맵 형식의 이미지 파일을 GIF 형식으로 변환하기 위해 “alchemy”라는 쉘어웨어 프로그램을 사용하여 포맷을 변환하였다. 그러



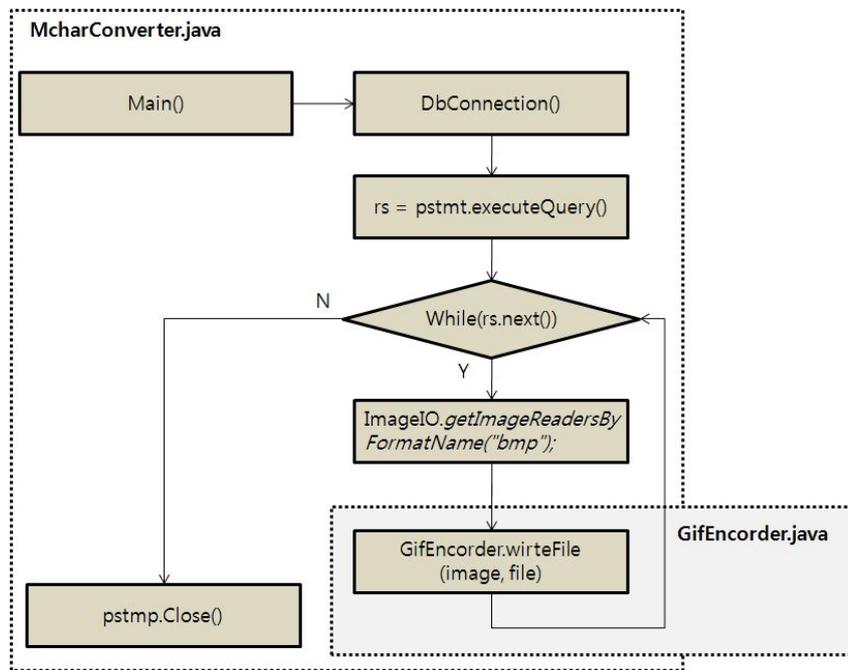
[그림 15] 이미지 변환 과정

나 이와 같이 별도의 외부 유틸리티를 사용하는 경우에 몇 가지 문제점이 발생하였다. 첫째, 사용된 쉘어웨어 프로그램의 경우 개발자들이 자주 사용하는 유틸리티가 아니기에 별도의 사용방법을 숙지해야 한다. 둘째, 이미지를 변환하는 과정을 살펴보면 누락문자 데이터베이스로부터 비트맵 형식의 파일을 모두 로드하고 다시 확장자를 .GIF로 변환한 후에 해당 유틸리티를 사용하여 실제 GIF 형식으로 변환하게 되므로 2단계에 걸쳐 작업이 진행된다. 따라서 변환하는 과정에서 시간이 오래 걸린다. 셋째, 이미지 변환 과정에서 약간의 오류가 발생하였다. 원본인 비트맵 형식의 이미지와 이를 변환하여 생성된 GIF 형식의 이미지를 대조하였을 경우 GIF 형식의 이미지가 이전의 비트맵 형식의 이미지와 다른 경우가 종종 발견되었다.

이러한 문제점을 개선하기 위해서 본 연구에서는 JAVA를 이용하여 비트맵 형식의 이미지 파일을 GIF 형식으로 변환하는 프로그램을 개발하였다.

2.4.2 개선된 이미지 생성 프로그램의 구성

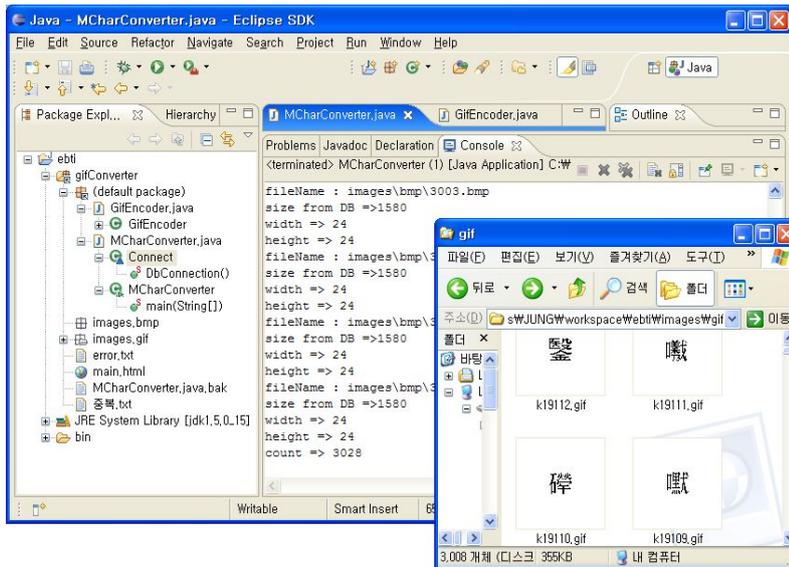
JAVA에서는 이미지를 컨트롤하기 위한 다양한 클래스와 메소드를 제공한다. 따라서 본 연구에서는 JAVA에서 기본적으로 제공하는 이미지 관련 클래스와 메소드를 사용하여 비트맵 형식의 이미지 파일을 GIF 형식으로 변환하는 프로그램을 개발하였다. [그림 16]은 개발한 이미지 생성 프로그램의 구성을 나타낸 것이다.



[그림 16] 이미지 생성 프로그램 구성도

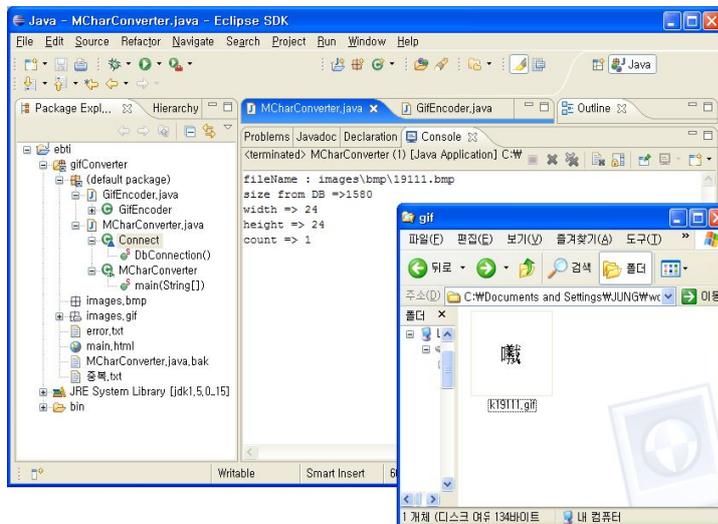
개선된 이미지 변환 프로그램은 기존의 쉘어웨어 프로그램에 비해 비트맵 형식을 GIF 형식으로 변환한 후에 이미지 오류의 발생이 거의 없다. 그리고 이전 프로그램의 경우 우선 데이터베이스에 등록된 누락문자 전체를 비트맵 형식의 이미지 파일로 로드한 후 GIF 형식으로 이미지 변환을 하는 방식이었는데 반해 이번에 개선한 프로그램에서는 비트맵 형식의 이미지 파일을 하나씩 로드하면서 바로 GIF 형식으로 변환하는 방식으로 기존의 방식에 비해 속도에서도 더 좋은 성능을 보인다. 또한, 간단한 SQL문 수정을 통해

특정한 문자만을 변환하는 것이 가능하며, JAVA로 작성되어 간단하게 컴파일과 실행만 하면 자동으로 이미지가 변환되므로 개발자의 입장에서 좀 더 쉽게 사용할 수 있다.



[그림 17] 개선된 프로그램으로 전체 이미지 변환 실행 결과

[그림 17]은 개선된 이미지 변환 프로그램을 사용하여 누락문자 데이터베이스에 등록된 문자를 모두 변환한 실행 결과 화면이고, [그림 18]은 간단한 SQL문 수정을 통해 특정한 ID의 누락문자를 이미지 변환한 결과 화면이다.



[그림 18] 특정 ID의 이미지 변환 실행 결과

3. 결론 및 향후 과제

본교는 불교학을 중심으로 한 한국학과 컴퓨터 정보통신 두 분야를 특성화의 큰 축으로 하고 있으며, 불교자료의 전산화야 말로 본교의 특성화 방향인 “불교학과 정보통신 기술”의 연계에 가장 적합한 프로그램이라 할 수 있다. 따라서 수년간 연구를 통해 한국불교전적 중 한글대장경을 전산화하여 본교의 특성화 사업에 부응하고자 하였다.

한글대장경의 전산화를 위하여 가장 필요한 것은 워드프로세서 입력형태로 되어있는 한글대장경 원문을 데이터베이스에 저장하는 기술, 저장된 데이터베이스에서 원하는 부분을 검색하는 기술 및 이를 인터넷에서 사용할 수 있도록 하는 인터페이스 처리 기술이다.

본 연구에서는 한글 워드 프로세서로 작업한 형태의 파일을 일반 유니코드 텍스트로 변환하여 이것을 유니코드 형태 그대로 데이터베이스에 저장하는 기술을 개발 및 구현하였다. 또한 검색 구조를 위하여 문서의 논리적 구조를 표현할 수 있는 XML을 도입하여 재구성 하였으며, 이러한 XML 형태의 문서에서 실제 검색에 필요한 조건들을 추출하여 데이터베이스를 구축하였다.

또한 이렇게 구축된 데이터베이스를 인터넷상에서 열람 및 검색이 가능하도록 웹 기반 프로그램을 작성하였으며, 이를 통하여 인터넷 환경에서 직접 한글대장경을 열람할 수 있도록 하였다. 그리고 여러 가지 검색 기능을 추가하여 사용자가 손쉽게 한글대장경을 열람하고 검색할 수 있도록 하였다. 그리고 유니코드로 표현되지 않는 한자를 인터넷에서 사용할 수 있도록 누락문자를 이미지하고, 원문에 해당 이미지의 URL를 입력할 수 있는 누락문자 관리기를 개발하였다. 누락문자 이미지의 손상 또는 잘못된 이미지 생성에 따른 원문의 이질감을 해소하기 위한 누락문자 수정기를 추가 개발하였다.

본 연구에서 개발된 한글대장경 총 30책 110경이며, 지금까지 총 한글대장경 464경에 대해 인터넷을 통해 검색하고자 한다면 URL “<http://ebtc.dongguk.ac.kr>”을 이용하면 된다. 향후 연구 과제는 사용자 편의

성을 고려 및 사용자의 접근에 대한 통계를 통한 다양하고 편리한 검색 기법의 개발이 필요하다. 그리고 현재 서비스되고 있는 한글대장경의 본문 글자체를 확대 및 축소할 수 있는 기능 추가와 검색 시스템의 화면 인터페이스 변경 등이 필요하다.

참고문헌

- [1] Ven. Huimin Bhikkhu, Christian Wittern, and Aming Tu, "CBETA Taisho Electronic Tripitaka," *Electronic Buddhist Text*, Vol. 3, pp.125-129, 2001.
- [2] Ven. Huimin Bhikkhu, Christian Wittern, Aming Tu, Lijuan Guo, and Ray Chou, "A Study on Creation and Application of Electronic Chinese Buddhist Texts: With the Yogācārabhūmi as a Case Study," *Electronic Buddhist Text*, Vol. 3, pp.49-55, 2001.
- [3] Jens Braarvig, "Thesaurus Literaturae Buddhicae (TLB): Its Scope, and a Description of Its Routines," *Electronic Buddhist Text*, Vol. 3, pp.23-32, 2001.
- [4] Dhananjay Chavan, "The Buddha's Words and Electronic Media," *Electronic Buddhist Text*, Vol. 3, pp.101-123, 2001.
- [5] Robert Chilton, "The Asian Classics Input Project (ACIP): Past, Present and Future," *Electronic Buddhist Text*, Vol. 3, pp.69-88, 2001.
- [6] Fred Coulson, "TBRC and Its Model for Linking Text Images with a Bio-Bibliographical Finding Database," *Electronic Buddhist Text*, Vol. 3, pp.131-145, 2001.
- [7] David Germano and Nathaniel Garson, "The Rise of 'Thematic Research Collections' in the Study, Teaching and Transmission of Buddhist Scriptures," *Electronic Buddhist Text*, Vol. 3, pp.147-190, 2001.
- [8] Young Sik Hong, Keum Suk Lee, Yong Kyu Lee, and Tae Sik Han, "Searching Missing Characters from the Hanguk Bulgyo Chonso Database," *Electronic Buddhist Text*, Vol. 3, pp.253-260, 2001.
- [9] C.C. Hsieh, Christian Wittern, and John Lehman, "A Project for Dealing with the Missing Character Problem," *Electronic Buddhist Text*, Vol. 3,

pp.261-269, 2001.

- [10] In Sub Hur, "Report on the Digital Tripitaka Koreana 2001," Electronic Buddhist Text, Vol. 3, pp.89-100, 2001.
- [11] Jae Sung Kim, "A Model of the Unified Tripitaka: Various Versions of the Saddharmapundarika-sutra Processed by XML," Electronic Buddhist Text, Vol. 3, pp.271-278, 2001.
- [12] Ishii Kosei, "Classifying the Genealogies of Variant Editions in the Chinese Buddhist Corpus: N-gram Based System for Variant Document Comparison and Analysis (NGSV)," Electronic Buddhist Text, Vol. 3, pp. 33-47, 2001.
- [13] Michel Mohr, "Linking Chan/Seon/Zen Figures and Their Texts: Problems and Developments in the Construction of a Relational Database," Electronic Buddhist Text, Vol. 3, pp.219-238, 2001.
- [14] Shigeki Moro, "Complex Spatial Digitization Tasks for the SAT Project," Electronic Buddhist Text, Vol. 3, pp.57-68, 2001.
- [15] Charles Muller and Michael Beddow, "Moving into XML Functionality: The Combined Digital Dictionaries of Buddhism and East Asian Literary Terms," Electronic Buddhist Text, Vol. 3, pp.191-218, 2001.
- [16] Christian Wittern, "Charting of Unknown Territory: Application of Topic Maps to Chan-Buddhist Chronicles," Electronic Buddhist Text, Vol. 3, pp.239-251, 2001.
- [17] Unicode enabling, Microsoft Developer's Network, 1997.
- [18] Public Unicode Font, <ftp://www.ifcss.org/ftp-pub/software/fonts/unicode>.
- [19] True Type and Unicode, <http://truetype.demon.co.uk:80/unicode.htm>.
- [20] Urs App, "A Look at the Korean Tripitaka Input Project", <http://www.ijnnet.or.jp/iriz/irizhtml/ebit/samsung.htm>.
- [21] 김무봉, "조선시대 간경도감의 역경사업," 전자불전, 제4집, pp.7-53, 2002.
- [22] 김성철, "『중론』 Śloka의 제작방식과 번역," 전자불전, 제5집, pp.16-36, 2003.
- [23] 김은중, "한글대장경 간행의 의의와 과제," 전자불전, 제4집, pp.79-104, 2002.
- [24] 김재성, "고려대장경 전산화 현황-고려·신수 전산본 일자대조 보고를 중심으로," 전자불전, 제4집, pp.124-154, 2002.
- [25] 노진홍, 유응구, 박성은, 이용규, 이금석, 홍영식, "한글대장경 전산화," 전자불전, 제4집, pp.155-192, 2002.

전자불전 제10집(2008)

- [26] 노진홍, 구현우, 유응구, 박성은, 박영희, 이용규, 이금석, 홍영식, 한보광, “한글대장경 전산화 3차 사업의 현황,” 전자불전, 제5집, pp.108-158, 2003.
- [27] 묘주스님, “한역경전 번역의 개선방향,” 전자불전, 제5집, pp.80-107, 2003.
- [28] 이금석, 이용규, 홍영식, 한태식, “한글대장경 검색시스템,” 전자불전, 제4집, pp.105-123, 2002.
- [29] 전재성, “세계의 현존하는 대장경의 문제점과 일상용어로의 번역,” 전자불전, 제5집, pp.37-62, 2003.
- [30] 한보광, “일제시대 삼장역회의 성립과 역할,” 전자불전, 제4집, pp.54-78, 2002.
- [31] 허인섭, “전산화본 고려대장경 2000 완성의 학술적 의미와 미래전망,” 전자불전, 제2집, pp.95-120, 2000.
- [32] 허일범, “티베트 대장경 번역의 문제점,” 전자불전, 제5집, pp.63-79, 2003.

주제어: 한글대장경, 한글대장경 검색 시스템, 한글 대장경 전산화, 유니코드, XML