

한글대장경 전산화

Digitization of Hangeul Tripitaka

노진홍*, 유응구*, 박성은*, 이용규*, 이금석*, 홍영식*, 한보광**
Jin Hong No, Eung Gu Ryu, Sung Eun Park, Yong Kyu Lee, Keum Suk Lee,
Young Sik Hong, Bo Kwang Han

*동국대학교 컴퓨터공학과

Dept. of Computer Engineering, Dongguk University

**동국대학교 선학과

Dept. of Seon Studies, Dongguk University

요약

본 연구는 한글대장경 전산화 2차 사업으로 한글대장경 30권 분량을 전산화하여 전 세계에서 활발하게 사용되고 있는 인터넷을 통하여 검색할 수 있도록 하는데 목적이 있다.

이는 고려대장경의 우리말 번역본인 한글대장경의 개역작업과 함께 전산화 사업을 수행함으로써 고문헌을 컴퓨터에 입력하여 데이터베이스에 저장하고 인터넷 및 CD-ROM을 통하여 검색할 수 있도록 하는 작업으로 구성된다. 한글대장경 고문헌은 이체자(異體字)와 오자(誤字), 파자(破字), 확장한자 및 특수문자 등을 포함하고 있으므로 효과적인 입력과 저장을 위해 본 연구에서는 유니코드(Unicode)를 사용하였으며, 고문헌의 문서구조 표현과 효율적인 검색을 위해 XML을 적용하였다. 또한 유니코드로 저장된 고문헌은 국내통신망과 인터넷을 통해 전세계에서 접근할 수 있으며, 효율적이고 편리한 검색을 위해서 경명 검색과 용어 검색, 쪽수 검색 등의 다양한 검색 방법을 영어 및 한글 인터페이스로 제공하는 검색엔진을 개발하였다. 이와 함께 인터넷이 지원되지 않는 환경에서도 손쉽게 한글대장경 문헌을 검색하고 열람할 수 있도록 CD-ROM을 이용한 검색시스템을 제공

한다.

본 연구는 한글 워드 프로세서를 통해 입력된 한글 대장경을 대상으로 하였으며, 인터넷 검색 환경은 윈도 NT에서 운영되는 마이크로소프트사의 SQL Server와 IIS(Internet Information server)를 사용하여 구현하고, CD-ROM 검색 시스템은 마이크로소프트사의 MS-ACCESS와 Visual Basic으로 작성된 프로그램을 통해 구현하였다.

I. 서 론

불법이 인도에서 전래되어 인류의 정신문화를 이끌어 온 지도 어언 3000여 년이 지났다. 불교의 가르침은 보통사람들이 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, 그 국민들에게 안심락도의 삶을 제시하였다. 이후 각 나라의 전법승들은 부처님의 가르침을 그 나라의 언어로 전하여 널리 일체중생을 이롭게 하는 역경사업에 진력하였다. 이는 국가의 지원을 받는 경우도 있었고, 전법승만의 불타는 신념에 의한 개인적인 사업인 경우도 있었다. 마침내 전법의 발길이 닿은 국가들에서는 불전을 자국어로 번역하여 편찬유포하게 되었다. 우리나라에 불교를 전해준 중국에서는 한문(漢文)불경이 편찬유포된 것이다. 우리나라에서는 중국의 불경을 전해받아 국가와 국민의 정신적 지주로 삼아왔다. 이는 역사에서도 확인되는 바이다. 세계문화유산으로 등록된 고려대장경은 몽고의 침입으로 국가가 위기에 처했을 시기에 부처님의 가르침으로 국가의 안녕과 백성의 평안을 기원하기 위해 전 국가적으로 역량을 결집한

우리의 문화유산인 것이다. 조선시대에 이르러서는 불경을 민간에 널리 유포시키기 위하여 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 전법활동이라고 할 수 있다.

II. 한글대장경 전산화 2차 사업

본 연구에서는 한글대장경을 전산화하여 전 세계에서 손쉽게 검색할 수 있도록 하는데 그 목적이 있다. 이러한 연구 목적을 달성하기 위해 크게 3가지 기술이 필요하다. 한글대장경을 컴퓨터에 입력하고 이를 편집하여, 데이터베이스에 저장하고, 데이터베이스에 저장된 내용들을 웹에서 검색할 수 있는 인터페이스와 검색 기술 및 웹이 지원되지 않는 환경에서 CD를 통해 검색할 수 있는 인터페이스와 검색 기술이 필요하다. 이들을 위해 본 연구에서 개발한 기술 내용 및 사용방법에 관해 먼저 1절에서는 한국불교전서의 입력 및 교정 방법에 대해 기술하고, 2절에서는 유니코드로 데이터베이스에 저장하는 방법을 살펴보고, 4절과 5절에서는 데이터베이스에 저장되어 있는 내용들을 검색할 수 있는 인터페이스 및 검색방법에 대해 설명한다.

1. 한글대장경의 입력 및 교정

한글대장경의 입력은 (주)동국전산에 외주를 주어 입력하고, 3차에 걸쳐서 엄밀한 교정 작업을 수행하였다.

2002년도 한글대장경 전산화 제2차 사업에서 입력교정한 대장경의 목록은 총 46경으로 다음과 같다. (※ K번호는 고려대장경의 경전고유번호임.)

- K.1 대반야바라밀다경 (1~300권)
- K.116 묘법연화경 7권
- K.119 유마힐소설경 3권
- K.127 금광명최승왕경 10권
- K.129 돈진타라소문여래삼매경 3권
- K.130 대수긴나라왕소문경 4권
- K.154 해심밀경 5권
- K.160 입능가경 10권
- K.175 문수사리보초삼매경 3권
- K.179 아사세왕경 2권
- K.271 팔길상신주경 1권
- K.273 자서삼매경 1권
- K.274 여래독증자서삼매경 1권
- K.276 팔양신주경 1권
- K.282 팔불명호경 1권
- K.286 팔길상경 1권
- K.378 수능엄삼매경 2권
- K.387 현겁경 8권
- K.390 불명경 12권
- K.452 천불인연경 1권

- K.471 불인삼매경 1권
- K.493 팔부불명경 1권
- K.508 십길상경 1권
- K.521 금강삼매경 1권
- K.571 현양성교론 20권
- K.573 현양성교론송 1권
- K.582 광백론본 1권
- K.610 대승유식론 1권
- K.613 유식론 1권
- K.615 대장부론 2권
- K.632 일수로가론 1권
- K.637 백자론 1권
- K.649 증일아함경 (1~51권)
- K.650 잡아함경 (1~50권)
- K.744 치선병비요법 2권
- K.801 정법염처경 (1~70권)
- K.802 불본행집경 (1~60권)
- K.955 아비달마구사론 (21~30권)
- K.984 수행도지경 7권
- K.1006 오문선경요용법 1권
- K.1406 범원주림 (38~100권)
- K.1433 대승과유론 1권
- K.1441 육십송여리론 1권
- K.1445 대승이십송론 1권
- K.1465 금광명경 4권
- K.1503 조당집 (1~10권)

1) 한글대장경의 입력 및 교정

입력과 교정을 마친 30권의 한글대장경은 전자불전연구소에서 페이지·대제목·소제목·각주·색인에 대하여 각각 태깅 작업을 수행하였다.

- (1) 페이지 태그작업 : 페이지를 검색하여 해당 원문을 보여준다.
- (2) 제목 태그작업 : 경전의 대제목과 소제목을 검색할 수 있으며, 경전의 제목을 통하여 해당 원문을 확인할 수 있게 한다.
- (3) 색인 태그작업 : 중요한 불교용어와 고유명사를 검색할 수 있게 한다.

2) 색인작업

색인어 선정 작업은 최종교정이 끝난 다음 최종교정지를 넘겨받아 해당하는 색인어에 형광펜으로 알아보기 쉽게 표시한다. 문장을 자세히 읽어 적합한 색인어를 추출하였다.

색인작업은 다음과 같은 원칙을 두고 수행하였다.

- (1) 최종 교정이 이루어진 다음에 색인어를 선정한다.
- (2) 한번 등록된 색인어는 중복표시하지 않는다. 즉, 한번 등록된 색인어는 계속해서 색인어의 기능을 가진다.
- (3) 불교의 중요한 술어·경론명·인명·법명·지명·사찰명 등의 고유명사를 중심으로 색인어를 선정하였으며, 또한 해당 경론의 교리나 사상에서 주요하게 다루고 있는 특징적인 용어를 중심으로 선정하였다.

2002년도 한글대장경 제2차 전산화사업에서 등록된 색인어는 약 5,000 단어에 달하며, 색인어는 한글대장경 전산화사업이 진행됨에 따라 계속

증가될 것이다. 색인어를 통한 대장경 검색은 불교를 전문으로 연구하는 학자나 학생은 물론 한글대장경을 열람하고자 하는 일반 사용자들에게도 상당한 편리성과 정확성을 제공할 것이다.

2. 데이터베이스 저장

한글대장경의 원문은 제목, 원문, 주석 등으로 구성되어있고, 각각에 해당되는 내용은 태그로 구별하기 때문에 이를 이용하여 데이터베이스를 구축할 수가 있다. 이때 원문에 나타나는 한문은 기존 문자 셋으로 표현하는데 한계가 있어서 유니코드로 변환하여 저장한다. 원문을 데이터베이스에 저장하기 위해서는 먼저 각 부분을 구별해 주는 태그들이 유효해야한다. 따라서 각 태그들이 유효한지 검증하는 작업이 필요하고 이 작업을 마친 다음에는 원문으로부터 제목, 원문 내용, 키워드를 추출하여 유니코드로 변환한 값을 각 테이블에 저장한다. 즉, 키워드는 키워드 사전으로부터 추출하여 키워드 테이블에 저장하고, 원문은 유니코드로 변환하여 “edocdata” 테이블에 권별로 저장한다. 이러한 한글대장경의 데이터베이스 구축 단계를 정리하면 다음과 같으며 자세한 내용은 본문을 통해서 살펴보도록 한다.

- ① 태그가 삽입된 원문의 유효성 검증 작업
- ② 키워드 및 원문 저장
- ③ 인덱스 추출
- ④ 파일 처리

1) 태그가 삽입된 원문의 유효성 검증 작업

텍스트 파일로 변환된 원문에 페이지, 제목, 주석 등을 구별하기 위하여 각각 <PAGE>, <JMOK>, <COMMENT>라는 태그들을 삽입한다. 이러한 태그들은 여는 태그(<...>)와 닫는 태그(</...>)가 쌍으로 이뤄져

야 하며, 만일 그렇지 않으면 잘못된 데이터가 데이터베이스에 들어갈 수 있으므로 반드시 확인을 해야 한다. 이러한 태그들의 검증 작업의 순서는 대략 다음과 같다.

- (1) “*.txt”로 저장된 파일들을 “*.xml”로 확장명을 바꾼다.
- (2) 웹 브라우저에서 해당 XML 문서를 불러들인다.
- (3) 웹 브라우저에 에러 메시지가 나타나지 않으면 유효한 문서이고, 그렇지 않으면 오류가 발생한 부분을 찾아 원문 내용을 수정해야 한다.
- (4) 모든 태그들은 여는 태그와 닫는 태그가 쌍으로 이루어져야만 유효한 문서를 생성할 수 있다.
- (5) 최종적으로 이렇게 생성된 유효한 문서를 데이터베이스 구축에 사용한다.

2) 키워드 및 원문 저장

키워드 저장은 키워드가 저장된 텍스트 파일로부터 키워드를 추출하여 키워드 테이블을 구축하는 방법을 이용한다. 키워드 파일에는 한글과 한글에 대한 한자가 저장되어 있어, 실제 “keyword” 테이블에는 한글과 한자에 대한 유니코드 값이 저장된다.

원문 저장은 유니코드 편집기에서 작성된 유니코드 원문을 그대로 테이블에 저장한다. 원문 파일을 라인별로 읽어 저장하면서 페이지 태그를 검사하여 페이지 당 라인(line)수와 “ncontinue” 등의 부가 정보를 생성한다. 원문을 저장할 때 원문에서 한 라인이 테이블의 저장공간을 초과하였을 경우에 “ncontinue”를 100자씩 끊어서 저장한다.

3) 원문 파일에서 인덱스 구축

키워드 테이블을 사용해서 원문을 검색하며, 검색 방법으로는 순차검색 방법을 사용한다.

(1) 유니코드의 특징

유니코드를 검색하기 위해서는 기존의 아스키파일과 구분되는 몇 가지 특징이 있다. 그 특징은 다음과 같다.

- ① 유니코드 파일의 처음이 “FF FE(16진수)”이다.
- ② 유니코드는 2바이트로 구성되기 때문에 문자를 비교할 때 2바이트 단위로 비교해야 한다.

(2) 키워드 인덱스 구축

키워드 인덱스를 구축하기 위해서는 키워드 파일에서 순차적으로 키워드를 읽고 각각의 키워드를 원문과 비교하는 방법을 적용했으며, 자세한 과정은 다음과 같다.

- ① “edocdata”와 “keyword” 테이블을 사용한다.
- ② 각각의 유니코드로 저장된 테이블의 처음 시작 코드를 읽는다.
- ③ 키워드 테이블에서 하나의 키워드를 추출하고, 이 키워드를 “edocdata” 테이블의 전체 내용과 비교한다. “keyword” 테이블에서 선택한 키워드와 같은 코드를 발견하면, 일련번호, 키워드 번호, 키워드, “edocdata” 테이블에서의 페이지의 정보를 저장하고, 원문 파일의 다음을 비교한다.
- ④ “edocdata” 테이블 수만큼 비교한 다음에 다음 키워드를 읽어 앞의 과정을 수행한다. “keyword” 테이블 전체를 비교했을 때, 프로그램은 종료한다.

(3) 제목 테이블 구조

원문을 검색할 때 <JMOK>과 </JMOK> 키워드를 검색하여 해당되

는 부분이 발견되면 제목을 “tag_jmok_table” 테이블에 저장하고, 제목이 위치한 위치 정보를 각각 “book”, “page”, “line”에 저장한다.

제목 태그는 트리 구조의 형태를 갖는다. <JMOK 다음에 나타나는 숫자가 제목의 레벨을 나타내고, <JMOK4> 태그의 속성인 “SERCH=“TRUE””는 경제목을 의미하며 그 값은 “tag_kyung_table”에 저장한다. “tag_page_table”에는 “tag_jmok_table”에 저장된 일련번호를 가지고 페이지 번호만 따로 추출하여 저장한다.

4) 테이블 생성 방법 및 구조 설명

데이터베이스를 구축하기 위해서는 MS SQL Server 7.0을 사용하며, 본 절에서는 테이블 생성 방법과 주요 테이블 구조에 대한 설명을 한다.

(1) 전체적인 테이블 생성 방법

먼저 색인 파일을 읽어서 “keyword”와 “hkeyword” 테이블을 만든다. 그런 다음 원문에서 한 줄을 읽어 제목 태그가 나타나면 각 해당 정보를 “tag_jmok_table”, “tag_page_table”, “tag_kyung_table” 테이블에 저장한다. 그리고 원문을 읽어가면서 해당 키워드가 존재하면 그 키워드는 “keyword_index” 테이블에 저장한다.

(2) 주요 테이블 구조

- keyword 테이블

열 이름	데이터형식	길이	정밀도	축소	Null 허용
keynum	int	4	10	0	✓
keyword	nvarchar	300	0	0	✓
keyword2	nvarchar	300	0	0	✓

[그림 1] keyword 테이블의 레코드 형식

keynum	keyword	keyword2
1	00AC00AC	E68FE68F
2	00AC00AC	B65BB65B
3	00AC00AC98B0	E68FE68FA390
4	00AC00AC98B04	3D4F3D4FA39051
5	00AC00AC7CB7	E68FE68F857FF2
6	00AC00AC31C19	B65BB65B56800E
7	00AC70AC1CC8	3D4F02F9D063
8	00AC8CAC	4C6B4850

[그림 2] keyword 테이블의 내용

- ① 테이블 명 : keyword
- ② 테이블의 역할 : 키워드 사전으로부터 입력받은 키워드를 저장
- ③ 필드의 역할
 - keynum : 각 키워드에 대한 유일키를 저장
 - keyword : 한글 키워드에 대한 유니코드 값을 저장
 - keyword2 : 한자 키워드에 대한 유니코드 값을 저장

■ keyword_index 테이블

열 이름	데이터형식	길이	정밀도	축소	Null 허
uid	int	4	10	0	<input checked="" type="checkbox"/>
keynum	int	4	10	0	<input checked="" type="checkbox"/>
pagenum	int	4	10	0	<input checked="" type="checkbox"/>
nlinenum	int	4	10	0	<input checked="" type="checkbox"/>
nbooknum	int	4	10	0	<input checked="" type="checkbox"/>

[그림 3] keyword_index 테이블의 레코드 형식

uid	keynum	pagenum	nlinenum	nbooknum
1	263	1	2	1
2	2	1	5	1
3	263	1	6	1
4	2	1	7	1
5	263	1	7	1
6	263	1	8	1
7	4	2	9	1
8	2	2	10	1
9	7	2	10	1
10	263	2	10	1

[그림 4] keyword_index 테이블의 내용

- ① 테이블 명 : keyword_index
- ② 테이블의 역할 : 각 권별로 키워드 인덱스 테이블을 유지하고, 키워드가 발견된 원문의 권, 페이지, 라인정보를 저장
- ③ 필드의 역할
 - uid : keyword_index 테이블의 유일키를 저장
 - keynum : “keyword” 테이블의 “keynum”과 일치한 값을 저장
 - pagenum : 키워드가 발견된 곳의 페이지 번호를 저장
 - linenum : 키워드가 발견된 곳의 라인 번호 저장
 - nbooknum : 현재의 권 번호를 저장

■ edocdata 테이블

열 이름	데이터형식	길이	정밀도	축소	Null 허용
nlinenum	int	4	10	0	<input checked="" type="checkbox"/>
sdocdata	nvarchar	1800	0	0	<input checked="" type="checkbox"/>
npagenum	int	4	10	0	<input checked="" type="checkbox"/>
npageline	int	4	10	0	<input checked="" type="checkbox"/>
ncontinue	int	4	10	0	<input checked="" type="checkbox"/>
nbooknum	int	4	10	0	<input checked="" type="checkbox"/>

[그림 5] edocdata 테이블의 레코드 형식

nlinenum	sdocdata	npagenum	npageline	ncontinue	nbooknum
1	3C004A004D004F	1	1	1	1
2	3C004A004D004F	1	2	1	1
3	3C004A004D004F	1	3	1	1
4	74C707B88CAC2	1	4	1	1
5	B4C590B24CB52	1	5	1	1
6	F8AD200045B540	1	6	1	1
7	C8B9C8B22000F	1	7	1	1
8	20005CD5BCAEE	1	8	2	1
9	31C144C72000B4	2	1	1	1
10	200070AC98CC5	2	2	2	1
11	58C7200004C7EC	2	3	2	1
12	F8AD20004CB52	2	4	1	1

[그림 6] edocdata 테이블의 내용

- ① 테이블 명 : edocdata
- ② 테이블의 역할 : 각 권별로 원문을 저장
- ③ 필드의 역할
 - nlinenum : 원문에 대한 유일키를 저장
 - sdocdata : 원문을 유니코드 형태로 저장
 - npagenum : 페이지 번호를 저장
 - npageline : 페이지 라인을 저장
 - ncontinue : 라인이 1800자를 초과할 경우 증가
 - nbooknum : 현재 권 번호를 저장

■ tag_bookname_table 테이블

열 이름	데이터형식	길이	정밀도	축소	Null 허용
booknum	int	4	10	0	<input checked="" type="checkbox"/>
bookname	nvarchar	800	0	0	<input checked="" type="checkbox"/>
hbookname	nvarchar	800	0	0	<input checked="" type="checkbox"/>
korea	nvarchar	800	0	0	<input checked="" type="checkbox"/>
shinsu	nvarchar	800	0	0	<input checked="" type="checkbox"/>

[그림 7] tag_bookname_table 테이블의 레코드 형식

	booknum	bookname	hbookname	korea	shinsu
▶	1	00B329BC11AD88BD5	대방광불화엄경 80권	4B002E003800300	54002E00320037C
	2	00B329BC11AD88BD5	대방광불화엄경 40권	4B002E003100320	54002E00320039C
	3	C4BCEDC5A1C744CE	발역잡마합경	4B002E003600350	54002E00310030C
	4	A1C7F4BCA5C7BDAC	잡보장경	4B002E003100300	54002E00320030C
	5	A1C744BE20C7BDAC	잡비유경	4B002E003100300	54002E00320030C
	6	A1C744BE20C7BDAC	잡비유경	4B002E003100300	54002E00320030C
	7	A1C744BE20C7BDAC	잡비유경	4B002E003100300	54002E00320030C
	8	6CADA1C744BE20C7	구잡비유경	4B002E003100300	54002E00320030C

[그림 8] tag_bookname_table 테이블의 내용

- ① 테이블 명 : tag_bookname_table
- ② 테이블의 역할 : 책이 고려대장경과 신수대장경의 어느 부분에 해당되는지 나타냄
- ③ 필드의 역할
 - book_num : 각 책에 대한 유일키를 저장
 - bookname : 책의 제목에 대한 유니코드 값을 저장
 - hbookname : 책의 제목의 한글 독음 값을 저장
 - korea : 고려대장경의 해당부분을 나타냄
 - shinsu : 신수대장경의 해당부분을 나타냄

■ tag_jmok_area_table

열 이름	데이터형식	길이	정밀도	속소	Null 허용	기본
▶ tag_num	int	4	10	0	<input checked="" type="checkbox"/>	
jmok	nvarchar	800	0	0	<input checked="" type="checkbox"/>	
nlevel	int	4	10	0	<input checked="" type="checkbox"/>	
npagenum	int	4	10	0	<input checked="" type="checkbox"/>	
nlinenum	int	4	10	0	<input checked="" type="checkbox"/>	
nbooknum	int	4	10	0	<input checked="" type="checkbox"/>	
startnum	int	4	10	0	<input checked="" type="checkbox"/>	
endnum	int	4	10	0	<input checked="" type="checkbox"/>	
startpage	int	4	10	0	<input checked="" type="checkbox"/>	
endpage	int	4	10	0	<input checked="" type="checkbox"/>	

[그림 9] tag_jmok_area_table 테이블의 레코드 형식

tag_num	jmok	nlevel	npagenum	nlinenum	nbooknum	startnum	endnum	startpage	endpage
1	00B329BC11AD88BD54D6A0	0	0	0	1	1	32426	0	2062
2	74D520001CC82800E38940	1	10001	2	1	2	161	10001	10017
3	00B3FCC8E0C2EDC500B3	1	20001	162	1	162	197	20001	20003
4	00B329BC11AD88BD54D6A0	1	1	198	1	198	32426	1	2062
5	00B329BC11AD88BD54D6A0	2	1	199	1	199	308	1	16
6	31002E00200038C1FCC81A	3	1	202	1	202	308	1	16
7	00B329BC11AD88BD54D6A0	2	17	309	1	309	979	17	46
8	31002E00200038C1FCC81A	3	17	312	1	312	979	17	46
9	00B329BC11AD88BD54D6A0	2	47	980	1	980	1720	47	78
10	31002E00200038C1FCC81A	3	47	980	1	980	1720	47	78

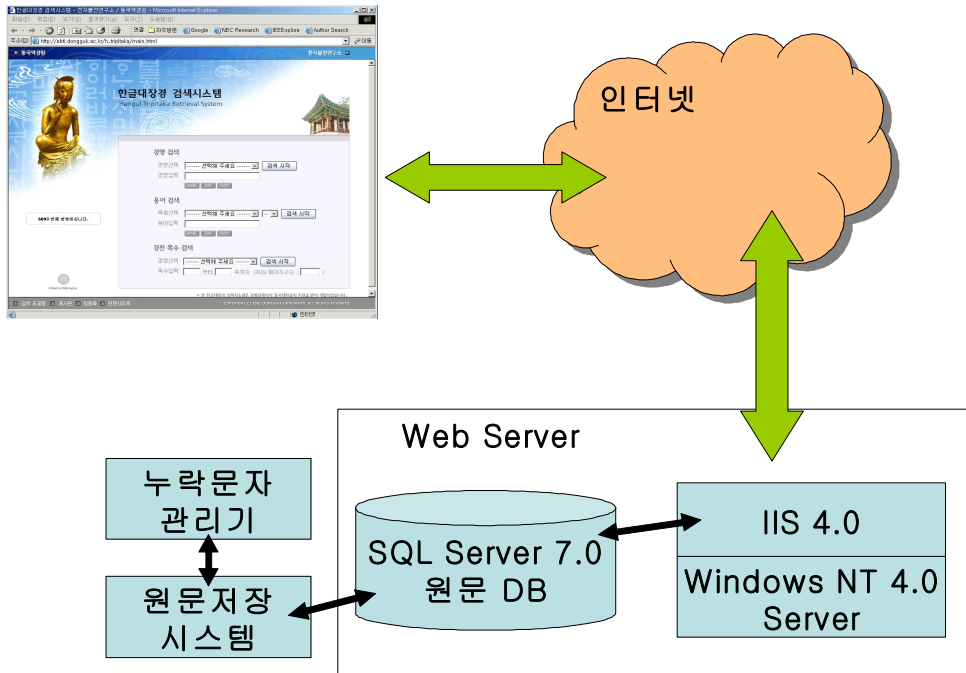
[그림 10] tag_jmok_area_table 테이블의 내용

- ① 테이블 명 : tag_jmok_area_table
- ② 테이블의 역할 : 각 레벨에 해당하는 제목들의 위치 정보 저장
- ③ 필드의 역할
 - tag_num : 각 제목에 대한 유일키 값을 저장
 - jmok : 책의 제목에 대한 유니코드 값을 저장
 - nlevel : 각 제목의 레벨 정보를 저장
 - npagenum : 각 제목이 위치한 페이지 정보를 저장
 - nlinenum : 페이지 안에서 각 제목이 위치한 라인 정보를 저장
 - nbooknum : 현재 권의 번호를 저장
 - startnum : 각 제목에 해당되는 원문 내용이 “edocdata” 테이블에서 시작되는 정보를 저장
 - endnum : 각 제목에 해당되는 원문 내용이 “edocdata” 테이블에서 끝나는 정보를 저장
 - startpage : 각 제목이 원문에서 시작되는 페이지에 대한 정보를 저장
 - endpage : 각 제목이 원문에서 끝나는 페이지에 대한 정보를 저장

3. 웹 검색 시스템

1) 한글대장경 웹 검색시스템 구성

[그림 11]은 본 연구에서 개발한 한글대장경 웹 검색시스템의 구성을 나타낸다.



[그림 11] 한글대장경 웹 검색시스템의 구성도

사용자는 웹 브라우저를 이용하여 웹 서버에 접속하고, 사용자의 요청은 IIS를 통해 ASP로 구현된 웹 검색시스템에 전달된다. 웹 검색시스템은 구축된 데이터베이스와 연동하여 결과를 사용자에게 반환한다.

한글대장경 웹 검색시스템의 개발 및 운영 환경은 다음과 같다.

- 운영체제 : Microsoft Windows NT 4.0
- 데이터베이스 : Microsoft SQL Server 7.0
- 웹 서버 : Microsoft Internet Information Server 4.0
- 개발 언어 : ASP, Javascript, Visual Basic 6.0
- 클라이언트 환경 : Internet Explorer 5.0 이상

2) 웹 검색시스템의 주요 기능

웹 검색시스템은 한글대장경의 효율적이고 편리한 검색을 위해서 다양한 검색 방법을 제공하며 사용자의 편의를 위하여 검색 도움말 및 게시판 등을 제공하고 있다. [표 1]은 한글 대장경의 웹 검색시스템에서 제공하는 검색 방법들을 나타낸 것이다.

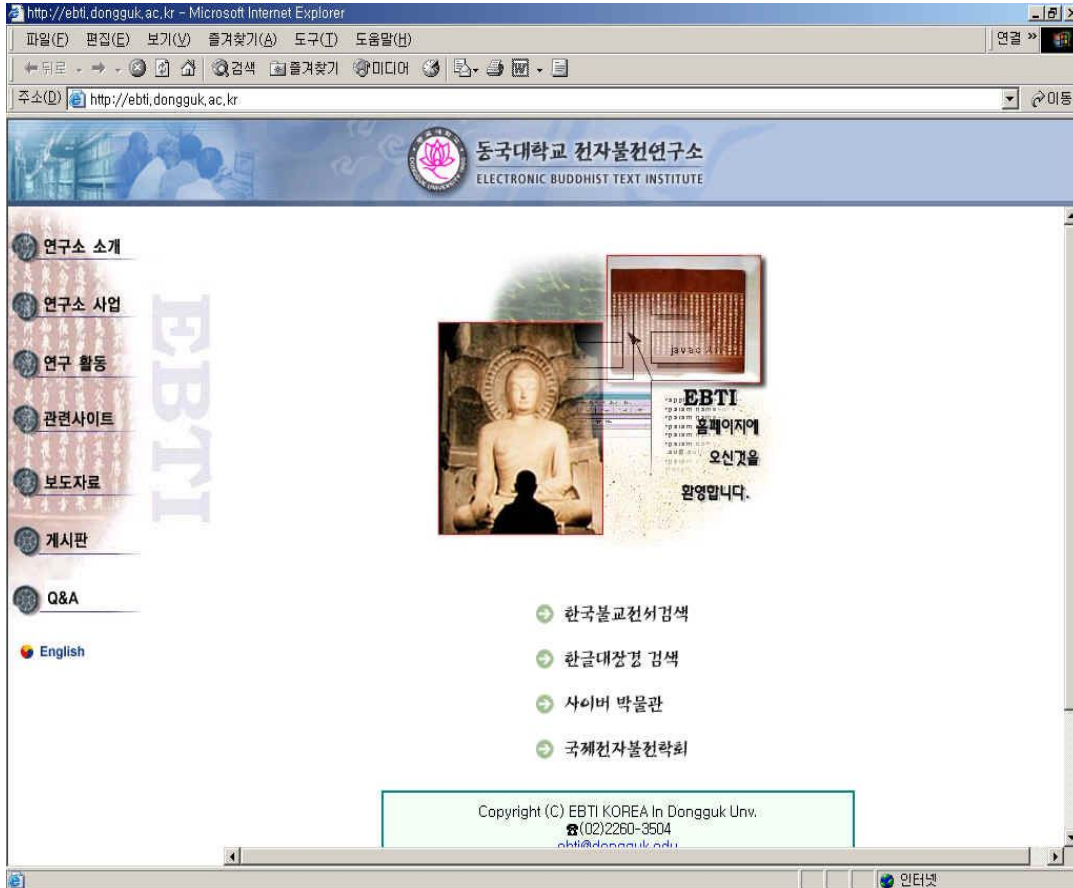
[표 1] 웹 검색시스템에서 제공하는 검색 방법

검색 방법	기능
경명 검색	경명을 입력하거나 리스트로 제공되는 경명의 목록을 이용하여 원하는 경을 검색한다.
용어 검색	하나의 경에 대하여 키워드로 지정된 용어를 가나다 순서로 검색하거나, 검색하고자 하는 용어를 입력하여 한글로 검색하거나 한자 검색, 결과내 검색, 히스토리 검색을 한다.
쪽수 검색	하나의 경에 대하여 검색을 원하는 페이지를 시작 쪽수와 끝 쪽수를 입력하여 검색한다.

또한 학술적인 참고 자료로서 의미가 있도록 한글대장경 원문과 동일하게 검색 결과에 들여쓰기를 적용하였다.

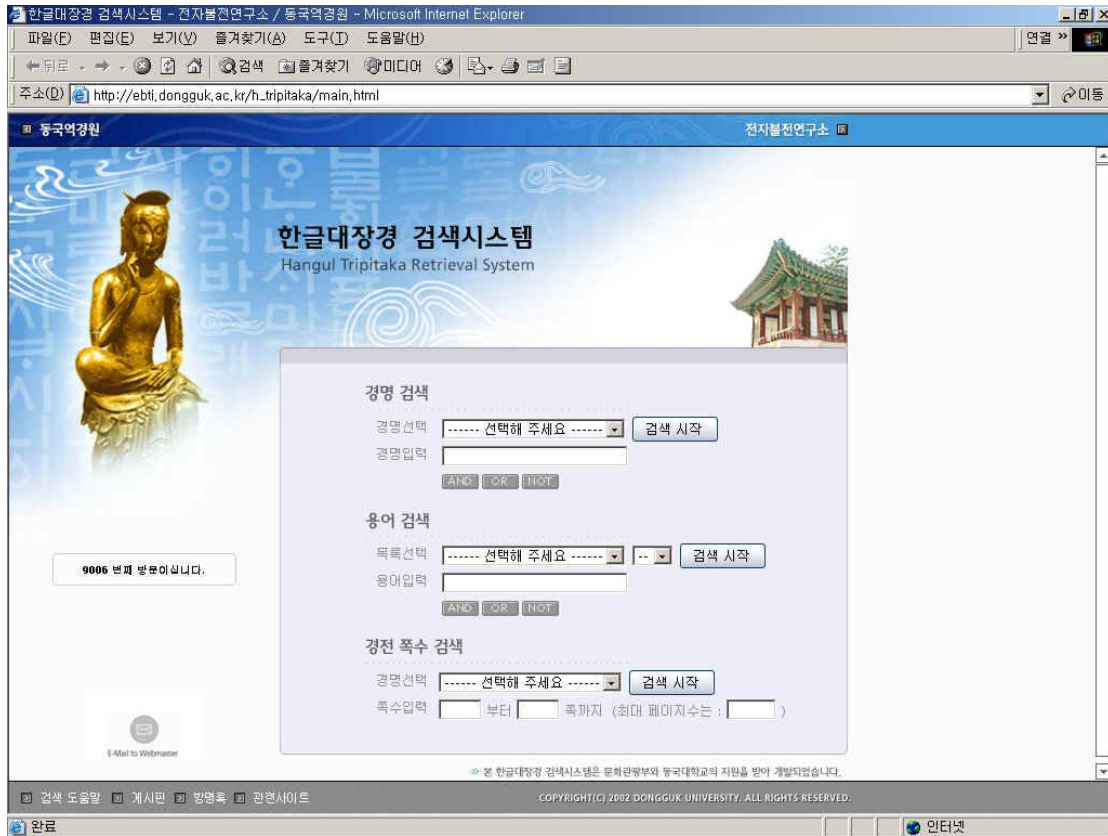
3) 검색 방법

웹을 통하여 한글대장경 검색시스템을 사용하기 위해서 사용자는 웹 브라우저를 이용하여 동국대학교 전자불전연구소 (<http://ebti.dongguk.ac.kr>)에 접속한다. [그림 12]는 동국대학교 전자불전연구소의 초기화면을 나타낸다.



[그림 12] 동국대학교 전자불전연구소 웹 사이트 초기화면

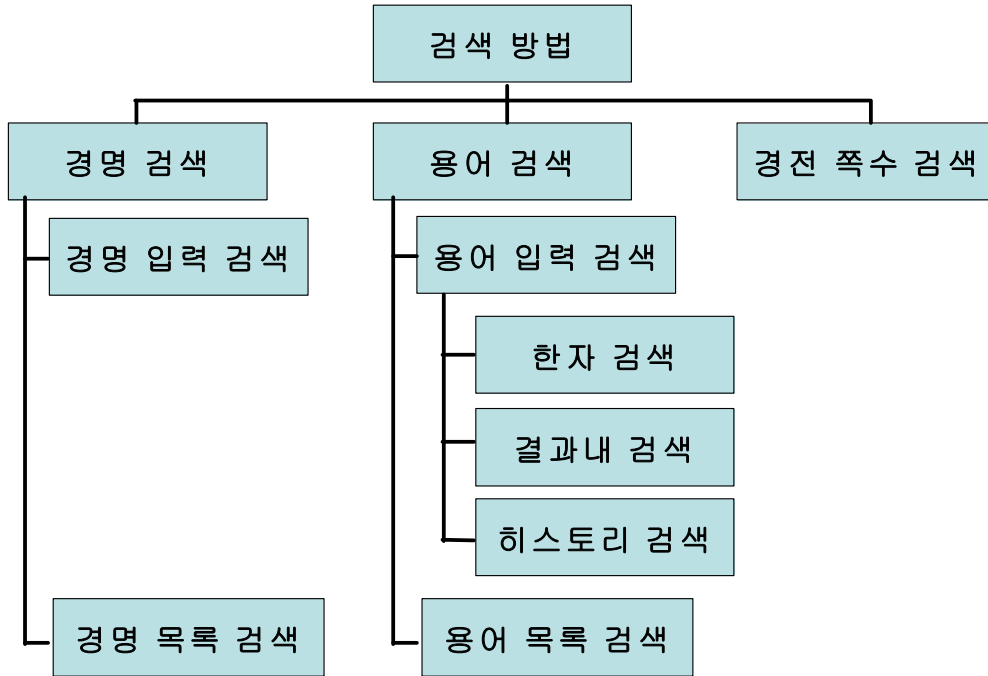
[그림 12]에서 ‘한글대장경 검색’을 선택하면 새로운 창에 한글대장경 웹 검색시스템이 나타난다. [그림 13]은 한글대장경 웹 검색시스템의 초기화면을 나타낸다.



[그림 13] 한글대장경 웹 검색시스템 초기화면

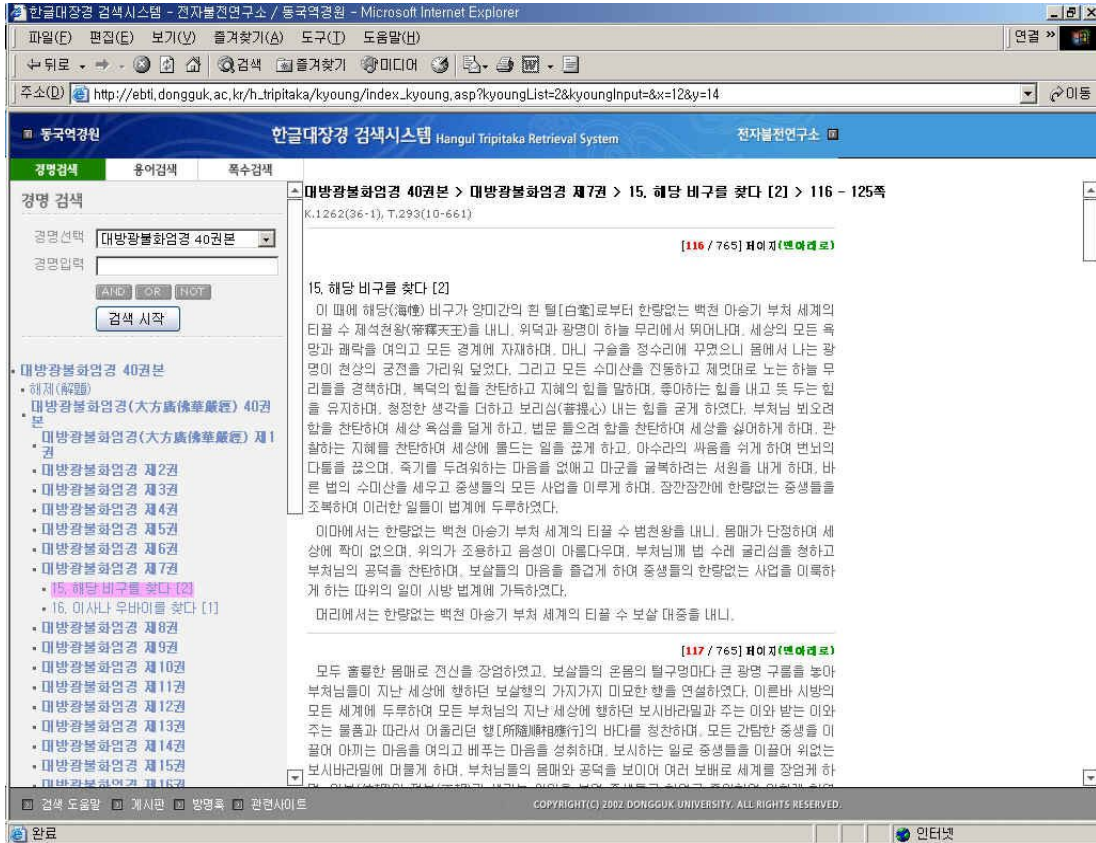
초기화면은 검색 메뉴와 다양한 링크로 구성된다. 상위 프레임은 전자불전연구소, 동국역경원 링크로 구성되고, 하위 프레임은 검색도움말, 게시판, 방명록, 관련사이트 링크로 구성된다. 중간 프레임은 경명 검색, 용어 검색, 쪽수 검색으로 구성되고, 각 검색 메뉴에서 조건을 선택하거나 입력하고, '검색 시작'을 클릭하면 검색 방법에 맞는 해당 검색 페이지로 이동한다.

한글대장경 웹 검색시스템의 기본 검색 방법은 입력된 경명이나 용어의 한글 독음을 이용하여 검색하는 것이다. 또한 불리언 검색 기능을 제공하여 사용자가 정확한 질의를 할 수 있도록 한다. [그림 14]는 한글대장경 웹 검색시스템에서 제공하는 검색방법의 목록을 나타낸다.



[그림 14] 한글대장경 웹 검색시스템에서 제공하는 검색 방법

한글대장경은 수많은 경(經)으로 이루어져 있기 때문에 경단위의 검색은 한글대장경에서는 필수적인 기능이라 할 수 있다. 경의 이름을 입력하여 검색하는 ‘경명 입력 검색’이나 경명으로 구성된 목록에서 선택하여 검색하는 ‘경명 목록 검색’은 기존의 고문헌 검색 시스템에서 제공하고 있다. 하나의 경은 많은 제목으로 구성되어 있기 때문에 제목을 색인 목록으로 만들면 신속한 검색이 어렵다. 따라서 본 검색시스템은 경을 구성하는 제목을 트리형태로 만들었다. [그림 15]는 트리 형태로 구성한 제목을 나타낸다.

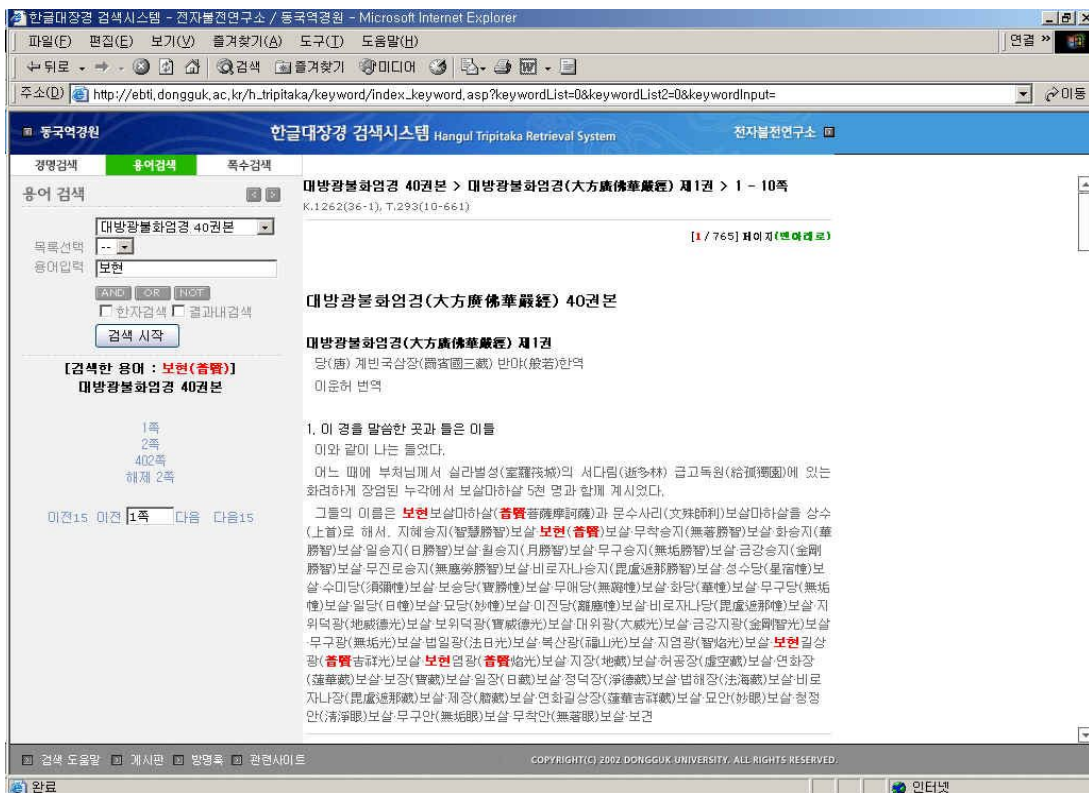


[그림 15] 트리 형태로 구성한 제목

용어 검색은 등록된 용어를 목록에서 선택하거나 직접 한글 독음을 입력하여 해당 용어를 포함하는 페이지를 검색하는 방법이다. 한글대장경 검색 시스템은 현재 50,000 여개의 불교 용어를 이용하여 경전별로 용어의 위치를 색인화 함으로써 빠른 검색이 가능하도록 하고, 사용자의 다양한 검색 요구에 맞도록 결과내 검색, 히스토리 검색 및 한자 검색을 지원한다.

용어 검색은 크게 ‘입력 검색’과 ‘목록 검색’으로 나눌 수 있다. ‘입력 검색’은 경을 선택하고 용어입력란에 검색할 용어를 입력하여 검색하는 방법이다. 용어를 입력하여 검색하면 입력한 용어가 포함된 페이지들을 15개씩 잘라서 보여준다. 그 중에 검색하고자 하는 페이지를 선택하면 그 페이지에서 검색한 용어가 있는 부분을 빨간 글씨로 나타내서 결과를 표현한다.

또 다른 용어 검색 방법인 ‘목록 검색’은 경을 선택하고, ‘목록 선택’에서 ‘ㄱ’부터 ‘ㅎ’중에 하나를 선택하여 검색하는 방법이다. 목록 중에서 한 가지를 선택하면 해당하는 용어 전체 목록이 나타나고, 검색된 용어 목록들 중에서 찾고자 하는 용어를 선택하면, 해당 용어가 있는 페이지들의 목록이 15개씩 화면에 나타난다. 선택한 페이지의 내용은 ‘입력 검색’과 같이 검색한 용어가 있는 부분이 빨간 글씨로 표현된다. [그림 16]은 ‘보현’이란 용어를 용어 검색한 화면이다.

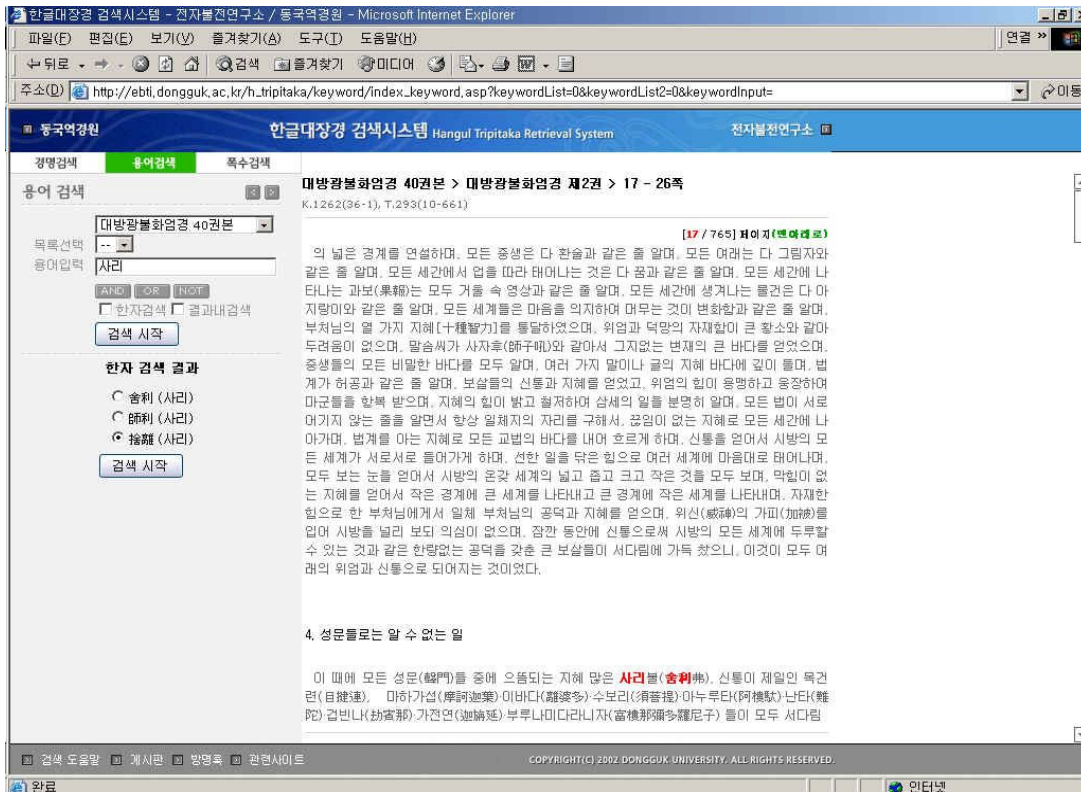


[그림 16] ‘보현’이라는 용어를 용어 검색한 결과 화면

지금까지 살펴본 단순한 검색 방법 외에도 본 한글대장경 검색 시스템에서는 ‘결과내 검색’, ‘한자 검색’, ‘히스토리 검색’등의 다양한 검색 방법을 제공하고 있다. 이미 검색한 결과와 찾고자하는 결과가 관련이 있는 경우, 검색한 결과 안에서의 재검색은 보다 빠르고 정확한 검색을 가능

하게 한다. ‘결과내 검색’은 사용자가 검색을 진행해 나감에 따라 점차적으로 사용자가 원하는 방향으로 검색의 범위를 좁혀갈 수 있어 빠르고 정확한 검색을 할 수 있다. ‘결과내 검색’은 이전의 검색 결과 페이지에 한하여 이후의 검색하려는 용어를 검색하는 방법이므로, 이전의 검색 용어와 이후의 검색 용어의 AND연산으로 검색을 처리하였다.

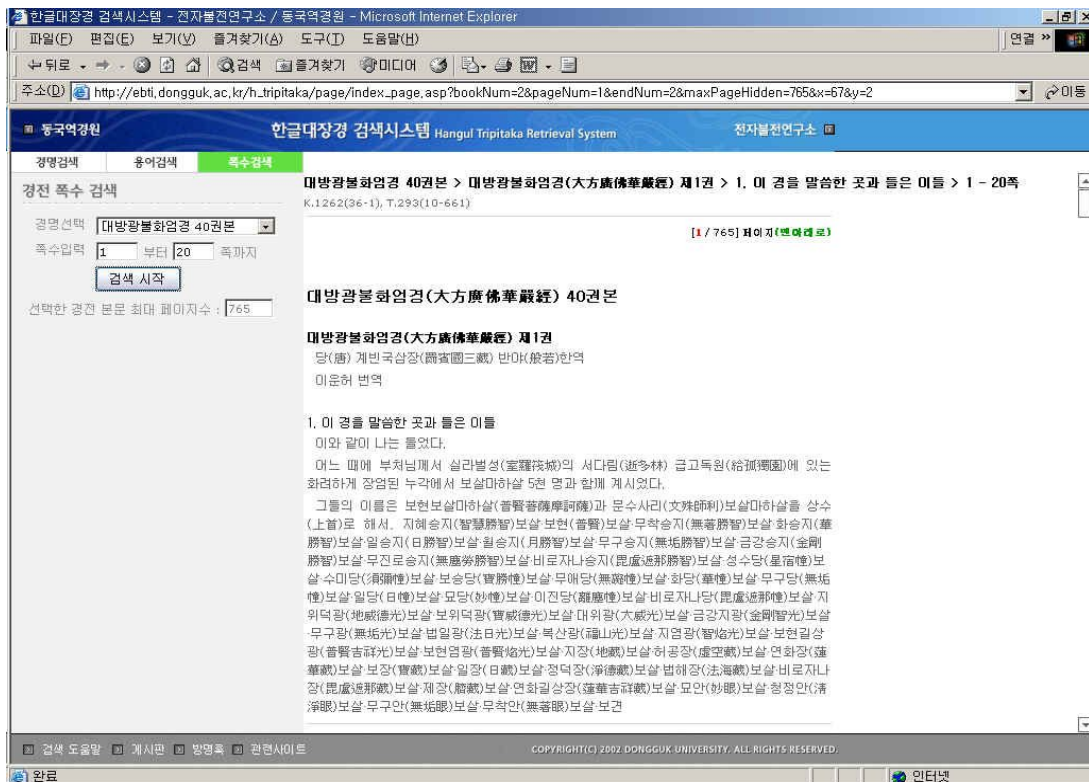
독음을 이용한 용어 검색은 독음은 같지만 한자가 다른 용어에 대해서는 정확한 검색 결과를 보장하지 못한다. 때문에 정확한 검색을 위해 한자를 이용한 검색 방법이 필요하다. ‘한자 검색’ 체크박스를 선택한 후 용어를 입력하면 한자와 한글독음이 함께 저장된 데이터베이스인 키워드 테이블에 접근하여 독음에 해당하는 한자를 검색한다. 검색 결과의 한자가 하나 이상이면 라디오 버튼을 제공하여 검색하고자 했던 정확한 한자를 선택하여 검색할 수 있도록 한다. [그림 17]은 ‘사리’라는 독음의 용어를 한자 검색한 화면이다.



[그림 17] ‘사리’란 독음을 가진 한자를 검색한 화면

히스토리 검색은 검색한 결과를 유지하여 이용하는 방법으로 웹 브라우저의 히스토리 기능과 동일하게 동작한다. 즉, ‘이전’ 또는 ‘다음’ 버튼을 이용하여 이미 검색한 결과를 신속하게 다시 제공한다. 이 검색 방법은 단독으로 사용할 수도 있지만 ‘결과내 검색’과 연동하여 사용할 수도 있다.

한글대장경 검색 시스템은 기본 검색 결과로 10쪽을 제공하고 있지만 사용자가 10쪽 이상의 검색 결과를 원하거나 특정 쪽으로 이동하고자 할 경우 ‘경전 쪽수 검색’을 사용할 수 있다. [그림 18]은 ‘대방광불화엄경 40권본’의 1-20쪽을 경전 쪽수 검색한 결과이다.



[그림 18] 쪽수 검색 결과

검색을 위해 사용자는 경을 먼저 선택하는데 사용자는 선택한 경이 몇 쪽으로 구성되어 있는지 알 수 없기 때문에 잘못된 쪽수를 입력할 수 있다. 따라서 본 시스템에서는 사용자가 경을 선택하면 경의 최대 페이지

수 정보를 제공하여 검색 페이지 입력의 오류를 막는다. 경을 선택한 후 시작 쪽수와 마지막 쪽수를 입력하고 ‘검색시작’을 누르면, 입력한 쪽 번호가 경의 범위에 속하는 지를 확인하고 경명과 쪽수를 이용하여 검색한다.

한글대장경 검색 시스템은 반복적인 검색에서도 현재 검색한 결과가 한글대장경의 어느 부분에 속하는지 쉽게 알 수 있도록 검색 결과의 경내 위치 정보와 연관된 고려대장경, 신수대장경의 정보를 부가 정보로 제공한다. [그림 19]는 검색 결과가 속한 경내 위치정보와 연관된 고려대장경 및 신수대장경 정보를 나타낸다.



[그림 19] 검색 결과의 경내 위치 정보와 연관된 부가 정보

4. CD-ROM 검색 시스템

1) CD-ROM 검색 시스템의 개요

한글 대장경은 인터넷과 웹 브라우저를 이용한 검색 시스템의 제공뿐만 아니라 CD-ROM을 이용한 검색 시스템을 제공하고 있다. 이는 인터

넷이 지원되지 않는 환경에서도 손쉽게 한글대장경 문헌을 검색 및 열람이 가능하도록 하고자 함이다. 따라서 CD-ROM 검색 시스템은 인터넷이 지원되지 않는 PC에 설치하여 사용하며, 한글 대장경 전산화 홈페이지에서의 검색과 똑같은 기능을 제공한다.

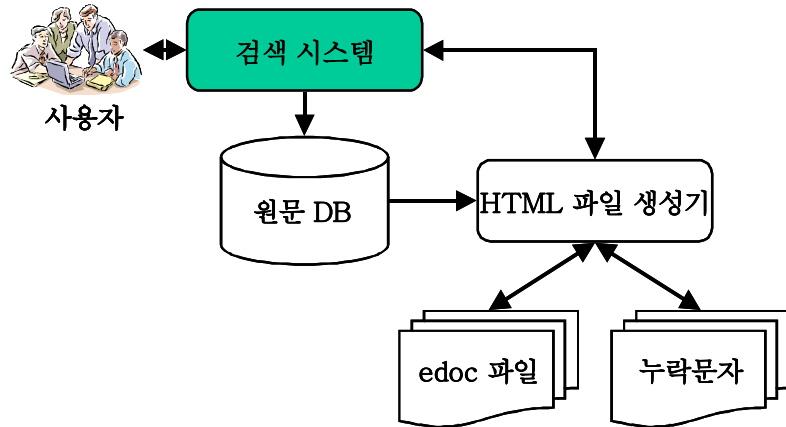
또한 모든 검색 방법 및 사용방법을 웹-기반 프로그램과 동일하게 개발하였으며, 윈도우 환경에서 자주 사용되는 인터페이스인 도움말 구조를 사용하였다. 그러므로 사용자에게 쉽고, 일관성 있는 사용방법을 제공하고 있다.

2) CD-ROM 검색 시스템의 구조

CD-ROM 검색 시스템은 웹-기반 검색 프로그램과 많은 점에서 유사한 구조를 가진다. 다른 점이 있다면 실제 원문 데이터는 파일 시스템 상에 html 문서의 형태로 존재한다는 것과 DB접속 및 기본 인터페이스 Visual Basic으로 구현되었다는 점 등이다. 이외에 검색에 사용되는 검색 인덱스는 웹-기반 프로그램과 거의 동일한 구조를 취하고 있다.

데이터베이스는 SQL 서버 대신 독립적인 파일 시스템으로 구동이 가능한 MS-ACCESS를 사용하였다. 이것은 작은 데이터베이스에 효율적으로 사용이 가능하며, CD-ROM등의 파일 시스템의 검색 인덱스를 저장하기에 효율적인 구조라고 볼 수 있다.

[그림 20]은 CD-ROM 검색 시스템의 구조를 나타낸다.



[그림 20] CD-ROM 검색 시스템의 구조

이와 같은 검색 시스템에서 검색이 이루어지는 과정을 살펴보면, 먼저 사용자는 Visual Basic으로 작성된 프로그램을 통하여 여러 검색조건(경명, 용어, 쪽수 등)을 입력하게 되며 이것은 데이터베이스의 검색 조건으로 사용이 된다. 이러한 검색조건은 MS-ACCESS로 저장된 문서 전체의 인덱스를 참조하여 검색을 수행하게 되며 그 결과로 검색 조건을 만족하는 원문에서 생성한 edoc 파일 이름을 HTML 파일 생성기에 전달하게 되고 HTML 파일 생성기는 검색 시스템에서 사용자가 볼 수 있는 HTML 형식의 문서를 작성한 후 검색 시스템에 반환한다. 반환된 HTML 형식의 문서는 검색 시스템의 Web 컨트롤을 통해 검색 결과를 사용자에게 보여준다.

3) 검색 방법

CD-ROM 검색 시스템의 한글대장경 검색 방법은 웹을 통한 한글대장경 검색 방법을 기초로 한다. 단, 검색할 원문의 내용이 데이터베이스에 저장하지 않고 edoc 파일로 저장하여 이 파일을 처리하여 결과를 보여주는 차이점을 보인다.

CD-ROM 검색 시스템을 실행하면 [그림 21]과 같은 스플래시 화면이 나타나고 3초간 화면에 표시된 후, [그림 22]의 검색 초기 인터페이스로 자동으로 변환된다.

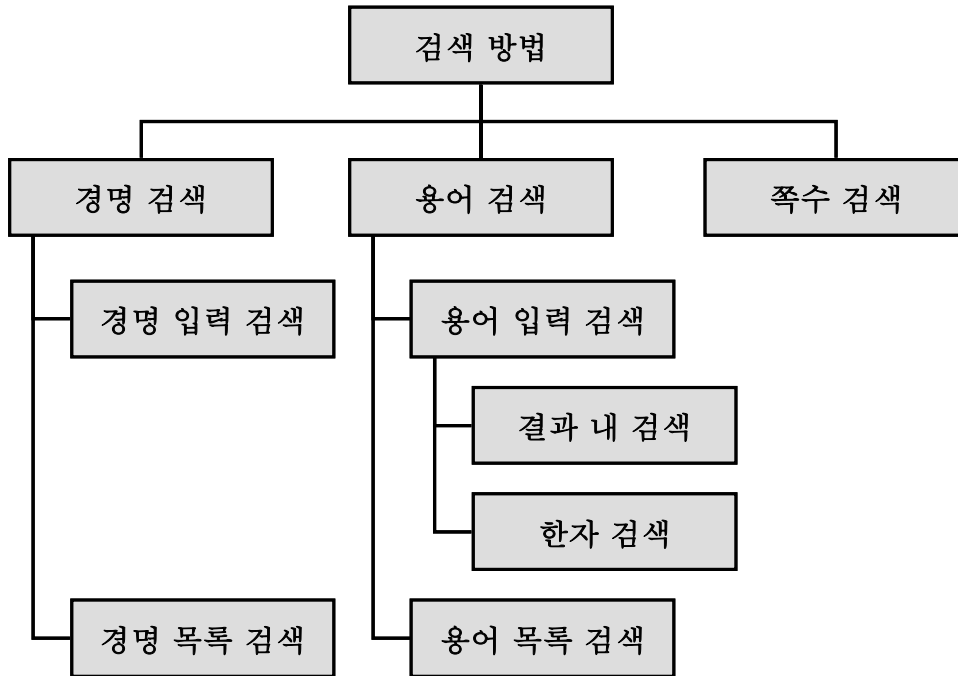


[그림 21] 스플래시 화면



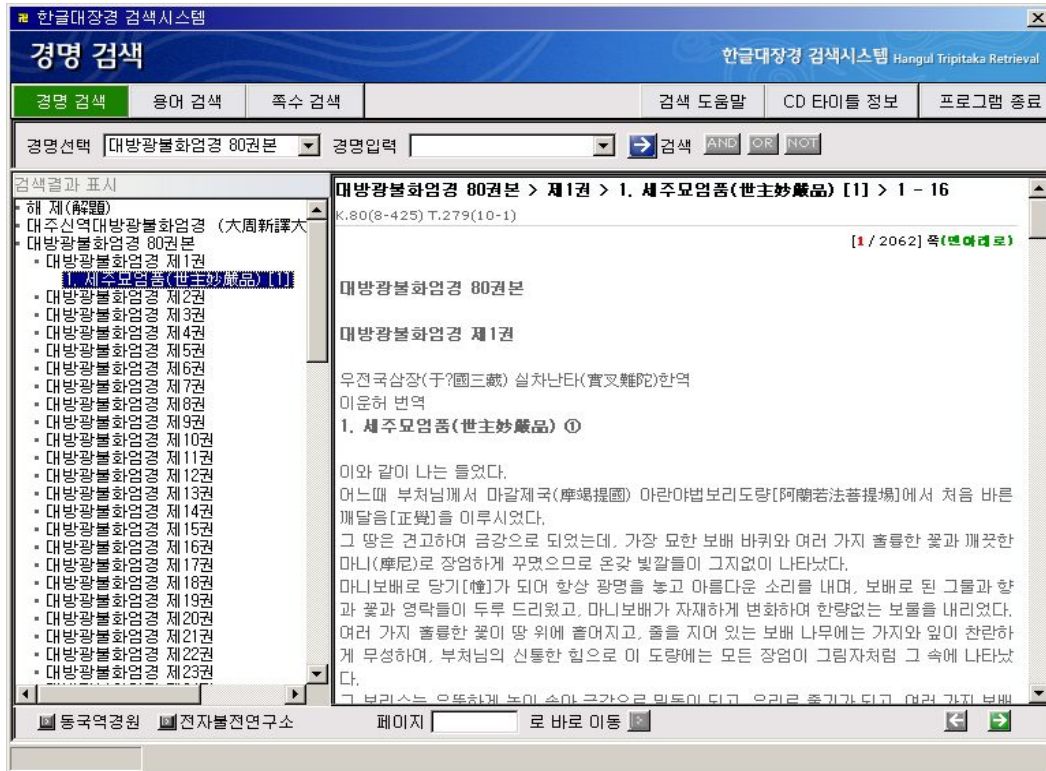
[그림 22] 초기화면

CD-ROM 검색 시스템의 검색 방법은 저장된 경명이나 색인어의 한글 음을 이용한 검색을 기본으로 한다. [그림 23]은 검색 시스템의 검색 방법의 목록을 보여준다.



[그림 23] CD-ROM 검색 시스템에서 제공하는 검색 방법

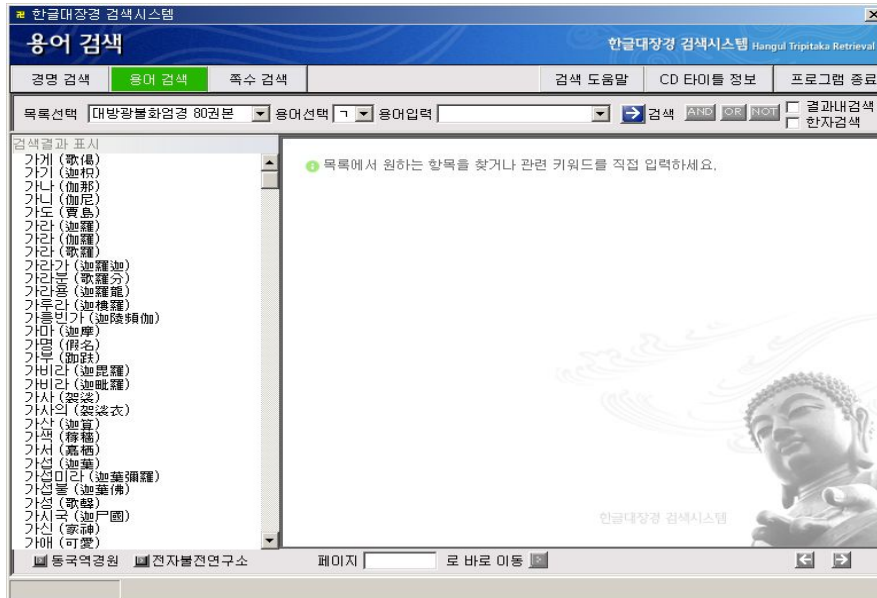
[그림 22]의 초기 화면에서 경명 검색을 버튼을 선택하면 [그림 24]와 같은 경명에 의한 검색을 할 수 있는 인터페이스가 표시된다. 경명 선택 콤보 박스에서 원하는 한글 대장경 문헌을 직접 선택하거나, 경명 입력 텍스트 박스에 경명의 한글 음을 입력하고 검색 버튼을 눌러 입력된 한글음의 포함된 경명 중에서 한글대장경 문헌을 선택할 수 있다.



[그림 24] 경명 검색 예

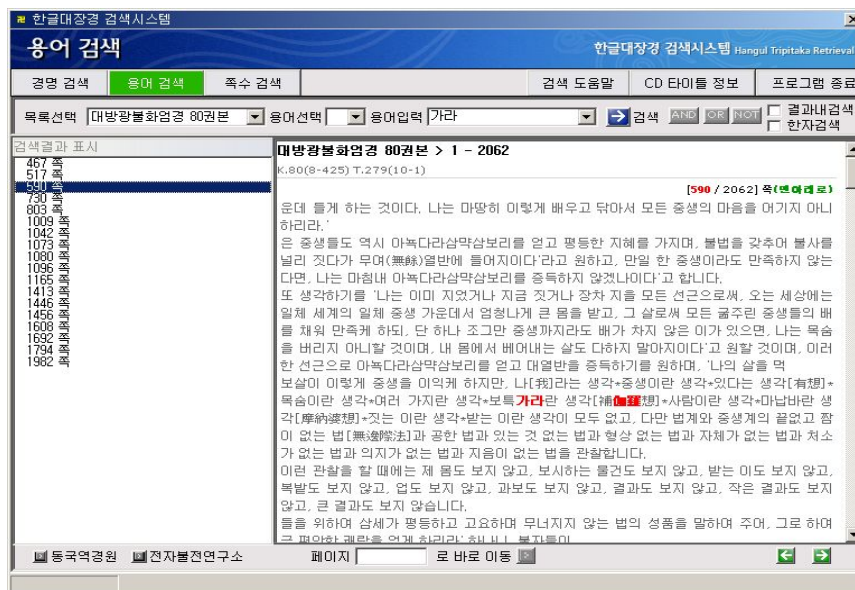
[그림 24]의 좌측에는 사용자가 찾고자 하는 경전의 원문을 쉽게 찾을 수 있도록 Tree형태의 제목 계층 구조를 보여준다. 우측에는 선택한 제목의 원문 내용을 10페이지씩 표시해 주고 우측 하단의 화살표 버튼을 이용하여 이전 혹은 이후 10페이지를 열람할 수 있도록 하는 기능을 제공한다. 또는 화면 하단의 바로 이동의 테스트 박스에 페이지수를 입력하면 해당 경의 페이지로 직접 이동하여 해당 페이지를 검색 할 수 있는 기능도 제공한다.

용어 검색은 초기 화면에서 용어 검색 버튼을 선택하면 [그림 25]와 같은 용어에 의한 검색 인터페이스가 표시된다. 열람하고자 하는 경을 콤보박스에서 먼저 선택한 후, 검색하고자 하는 용어의 시작 자음을 용어 선택 콤보박스에서 선택하거나 용어 입력 텍스트 박스에 색인어의 한글 음을 입력하고 검색 버튼을 누른다.



[그림 25] 용어 선택에 의한 용어 검색 화면

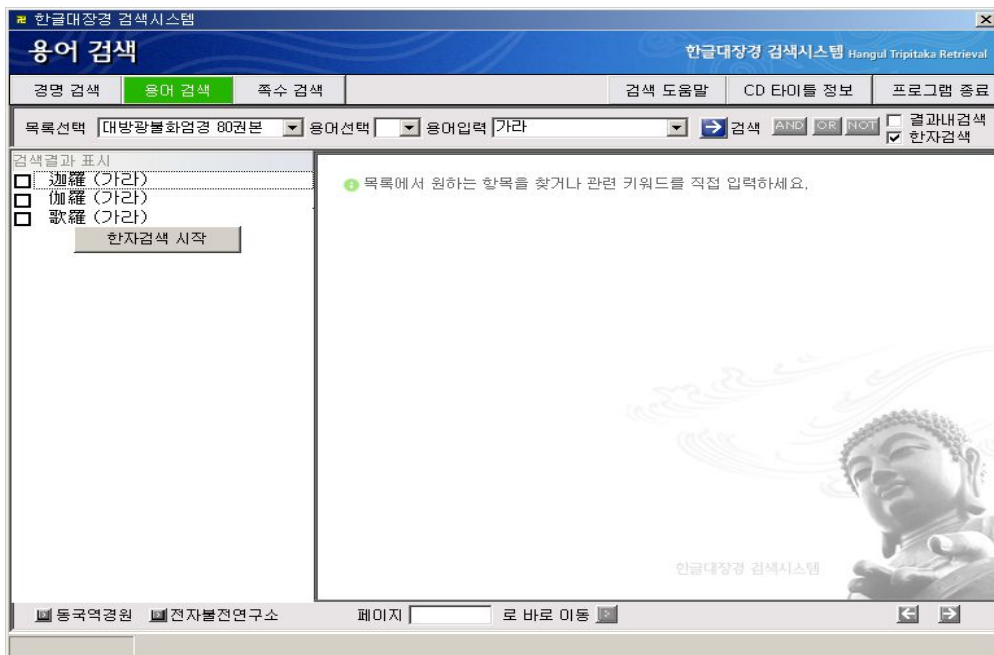
[그림 25]는 용어 선택 콤보박스에서 ‘ㄱ’을 선택한 화면으로 데이터베이스에 등록된 ‘ㄱ’으로 시작하는 색인어들의 목록을 보이고 검색하고자 하는 색인어를 선택하면 색인어가 포함되어 있는 페이지를 [그림 26]과 같이 보여 주고 해당 페이지를 선택하면 화면 우측에 해당 페이지가 나타난다.



[그림 26] 용어 입력에 의한 용어 검색 화면

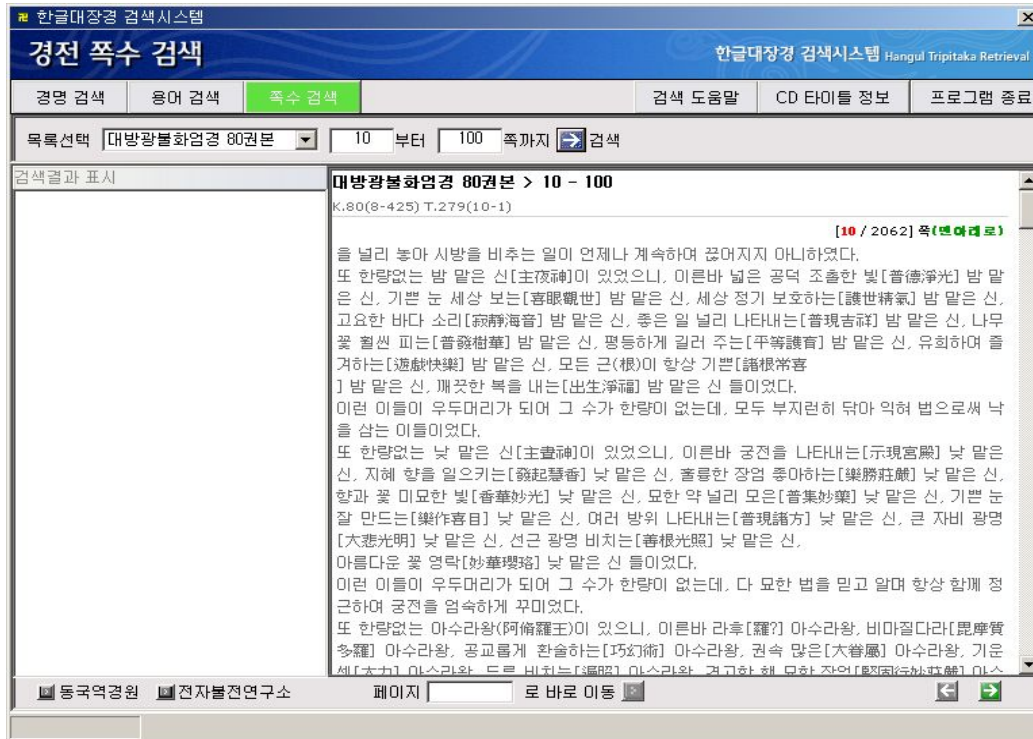
[그림 26]은 용어 입력 텍스트 박스에 검색하고자 하는 색인어인 ‘가라’를 입력하고 검색 버튼을 누르면 색인어를 포함하고 있는 페이지가 화면 좌측에 나타나고 해당 페이지를 선택하면 화면 우측에 내용이 나타난 화면이다. 검색을 하고자 한 색인어는 빨간색으로 표시 된다.

CD-ROM 검색 시스템은 ‘결과 내 검색’과 ‘한자 검색’등의 다양한 검색 방법을 제공하고 있다. ‘결과 내 검색’은 이미 검색한 결과 내에서 검색을 원하는 경우에 사용하고 ‘한자 검색’은 같은 한글 음을 가졌지만 한자가 다른 용어들을 검색하기 위해서 사용되어 진다. [그림 27]은 ‘가라’라는 한글음을 가진 색인어를 사용하여 ‘한자 검색’ 실행한 화면이다.



[그림 27] 한자 용어 검색 예

쪽수 검색은 초기 화면에서 쪽수 검색 버튼을 누른 후 검색을 원하는 문헌을 먼저 선택하고 열람하고자하는 시작 페이지와 끝 페이지의 범위를 입력한 후 검색 버튼을 누르면 원문의 내용이 화면 우측에 해당 범위 안에서 10페이지씩 나타난다. [그림 28]은 ‘대방광불화엄경 80권본’의 10 페이지부터 100페이지까지를 검색한 결과이다.

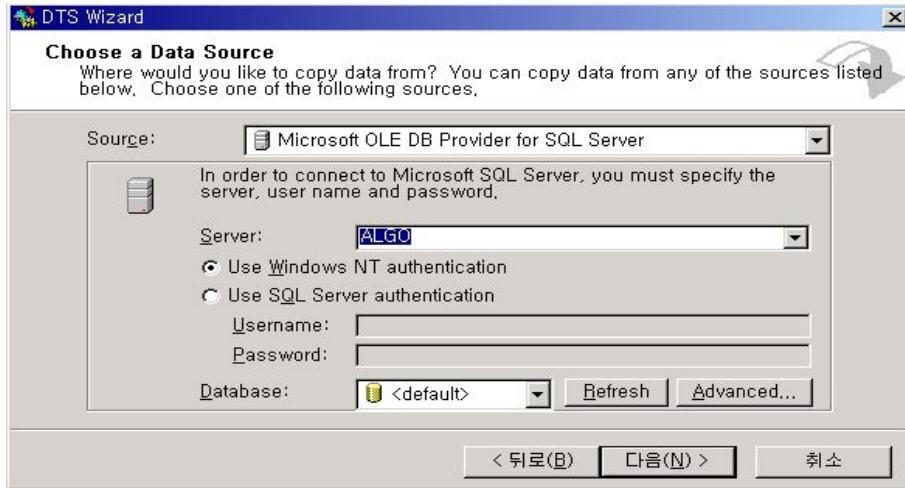


[그림 28] 쪽수 검색 예

4) CD-ROM 검색 시스템을 위한 데이터베이스와 edoc 파일

CD-ROM 검색 시스템을 위한 데이터베이스는 웹을 통한 한글대장경 검색 시스템에서 사용하는 데이터베이스에 있는 테이블 중 edoc_data를 제외한 모든 테이블을 사용한다. 웹 검색 시스템에서 사용하는 데이터베이스와 현재의 데이터베이스의 차이점은 데이터베이스 시스템이 SQL-SERVER 6.5에서 MS-ACCESS로 변경된 점이다.

데이터베이스 변환을 위하여 SQL-SERVER내에 포함되어 있는 'Import & Export Utility'를 사용하였다. 이 프로그램은 SQL-SERVER의 데이터베이스를 다른 형태로 변환해 주거나 다른 형태의 데이터베이스를 SQL-SERVER 형식으로 변환할 수 있도록 해주는 도구이다.



[그림 29] 데이터베이스의 변환

변경된 데이터베이스만을 사용하면 원문의 내용을 저장하고 있는 edoc_data 테이블이 제외되어 있기 때문에 원문의 내용을 검색을 할 수가 없다. 그래서 edoc_data 테이블의 페이지별 내용을 파일에 저장하여 edoc 파일을 생성하여 검색에 사용한다.

하나의 CD에 저장 될 수 있는 데이터의 량이 한정되어 있기 때문에 한글 대장경 원문을 포함하고 있는 전체 데이터베이스를 하나의 CD에 담는 것은 불가능한 작업이다. 그래서 위와 같이 한글 대장경 원문을 페이지마다 하나의 파일을 생성하여 여러 장의 CD에 저장하여 검색 시스템에서 사용된다.

파일 생성 방법은 웹 검색 시스템에서 사용하는 데이터베이스로부터 edoc_data의 모든 필드를 읽은 다음 페이지의 내용만 담고 있는 파일로 출력하여 저장한다.

III. 결론

본교는 불교학을 중심으로 한 한국학과 컴퓨터 정보통신 두 분야를 특성화의 큰 축으로 하고 있으며, 불교자료의 전산화야 말로 본교의 특성

화 방향인 “불교학과 정보통신 기술”의 연계에 가장 적합한 프로그램이라 할 수 있다. 따라서 본 연구에서는 한국불교전적 중 한국불교전서의 일부를 전산화하여 본교의 특성화 사업에 부응하고자 하였다.

현재 우리나라에는 귀중한 불교 문헌들을 포함하여 많은 한문 고문헌들이 있으나 이들에 대한 전산화 작업은 아주 미미한 실정이다. 특히 한국불교 및 한문 고문헌에 대한 연구를 하거나, 필요에 의해 한문 고문헌들을 열람하고 싶을 때 귀중한 자료들이 여러 도서관에 분산되어 있어 손쉽게 이용할 수 없다. 따라서 본 연구를 수행하면 총 한글대장경 60권 분량을 전산화 하여 이를 연구하는 연구자들이나 열람을 원하는 사람들에게 도움이 될 뿐만 아니라 우리의 귀중한 문화유산을 전세계에 널리 알릴 수 있다.

한글대장경의 전산화를 위하여 가장 필요한 것은 워드프로세서 입력형태로 되어있는 한글대장경 원문을 데이터베이스에 저장하는 기술, 저장된 데이터베이스에서 원하는 부분을 검색하는 기술 및, 이를 인터넷 및 CD-ROM상에서 사용할 수 있도록 하는 인터페이스 처리 기술이다. 본 연구에서는 한글 워드 프로세서 형태의 파일을 유니코드 텍스트로 변환하여 데이터베이스에 저장하는 기술을 개발 및 구현하였다. 또한 검색 구조를 위하여 문서의 논리적 구조를 표현할 수 있는 XML을 도입하여 재구성 하였으며, 이러한 XML형태의 문서에서 실제 검색에 필요한 조건들을 추출하여 검색 데이터베이스를 구축하였다. 또한 이렇게 구축된 데이터베이스를 인터넷 상에서 열람 및 검색이 가능하도록 웹-기반 프로그램을 작성하였으며 이를 통하여 인터넷 환경에서 직접 한글대장경을 열람할 수 있도록 하였다. 이러한 인터페이스는 인터넷이 지원되지 않는 환경을 고려하여 CD 형태의 배포본으로도 작성되었으며 이는 직접 PC에 설치하여 인터넷 지원 없이 한글 대장경의 내용을 열람할 수 있게 하였다.

본 연구에서 개발된 한글대장경 30권본에 대해 인터넷을 통해 검색하

고자 한다면 다음의 URL을 이용하면 된다. URL은 <http://ebti.dongguk.ac.kr/> 이다. 향후 연구 과제는 한글대장경의 더욱 많은 부분을 빠른 시일 내에 전산화하는 일이 필요하다. 그리고 데이터베이스에 저장된 내용을 검색할 때 등록된 용어뿐만 아니라 사용자가 원하는 일반 단어들도 검색 가능하도록 해야 하겠다.

참 고 문 헌

- [1] The unicode consortium, The Unicode Standard, Version 2.0, Addison Wesley, 1996.
- [2] 황기태 역, 어드밴스 윈도우 NT, 도서출판 대림, pp757-784, 1995.
- [3] 김응철, “고려장경 및 한자정보전산화에 관련한 문제제기”,
<http://members.iWorld.net/hederein/menu22/Kim.html>
- [4] 심재룡, “정보화 사회와 불교 전산화”,
<http://members.iWorld.net/hederein/menu22/Dogam32.html>
- [5] 강석진, “팔만사천대장경 전산화를 위한 제언, 한자위주 문헌의 워드 프로세서 데이터베이스, 탁상출판 시스템 개발을 위해”,
<http://members.iWorld.net/hederein/menu22/Kang.html>
- [6] 혜묵스님, “세계의 불교자료 전산화 계획과 고려대장경 전산화를 위한 몇가지 문제들”,
<http://members.iWorld.net/hederein/menu22/Hye.html>
- [7] 종림스님, “팔만대장경 전산화 추진경과와 이후 계획”,
<http://members.iWorld.net/hederein/menu22/>
- [8] 노용균, “불전 전산화와 SGML”,
<http://members.iWorld.net/hederein/menu22/Dogam42.html>.
- [9] 이규갑, “고려대장경 전산화에 있어서 이체자의 처리 문제”,
<http://members.iWorld.net/hederein/menu22/Yi.html>
- [10] 정주원, “한글 코드에 대하여”, 1995,

- <http://simac.kaist.ac.kr/~jwjung/seminar/hangul-i18n/ko-code.html>
- [11] Urs App, "A Look at the Korean Tripitaka Input Project",
<http://www.ijjnet.or.jp/iriz/irizhtml/ebit/samsung.htm>
- [12] Urs App, "The Importance of Markup",
<http://www.ijjnet.or.jp/iriz/irizhtml/maketext/foguang.html>
- [13] Urs App, " A Look at the Korean Tripitaka Input Project",
<http://www.ijjnet.or.jp/iriz/irizhtml/ebit/samsung.htm>.
- [14] ISO 8879:1986, Standard Generalized Markup Language, 2nd Edition.
- [15] ISO/IEC 10646-1:1993, "Information Technology - Universal Multiple-Octet Coded Character Set(UCS) - Part I : Architecture and Basic Multilingual Plane".
- [16] Public Unicode Font, <ftp://www.ifcss.org/ftp-pub/software/fonts/unicode>.
- [17] How to View Chinese/Japanese/Korean HTML with Netscape Communication on US version of Windows 95 or NT,
<http://people.netscape.com/ftang/communicatorfont.html>.
- [18] Unicode Support in Win32, Microsoft Developer's Network, 1997.
- [19] Unicode enabling, Microsoft Developer's Network, 1997.
- [20] True Type and Unicode,
<http://truetype.demon.co.uk:80/unicode.htm>
- [21] CJK Codes-Unicode/ISO-10646 Unicied "Ideographs",
<http://www.mit.edu:8001/afs/athena.mit...r/a/k/akbar/www/Unicode-ideographs.html>
- [22] "Installing Bitstream Cyberbit Version 1.1",
<http://www.bitstream.com/cyberbit.html>.
- [23] Panorama, <http://www.softquad.com>, Softquad 사.
- [24] 인터넷으로 만나는 불교,
<http://members.iWorld.net/hederein/menu23/Pogyu121.html>

- [25] 김숙자, SGML의 묻는 것, 성안당, 1997.
- [26] 장희창, 현득창, 이수연, SGML 가이드, 사이버출판사, 1997.
- [27] 현득창, 임광택, 이수연, “SGML 기본 파서를 이용한 SGML문서 편집기의 구현”, 한국정보과학회, 정보과학회 논문지, Vol 25, No. 1, 1998.
- [28] 대장경학술용어연구회, 대정신수 대장경소인, 제1권, 대장경학술용어연구회, 1975.
- [29] 한국불교신문, 시방시계 1월 27일자, 현대불교신문사, 1999.
- [30] 김정숙, 유용구, 이용규, 이금석, 홍영식, “유니코드를 기반으로 한 한자 입력 시스템 개발”, 한국정보과학회, ‘98봄 학술발표논문집, Vol 25, No. 1, 1998.
- [31] 김태규, 유병인, 한인, 이용구, 이금성, 홍영식, “유니코드 한자 지원 문법 지시적 SGML편집기의 설계 및 구현”, 한국정보과학회 ‘98봄 학술발표논문집, Vol. 25, No. 1, 1998.
- [32] 주신탉, 설승진, 이용규, 이금석, 홍영식, “유니코드와 SGML을 이용한 한국 고문헌 데이터베이스 구축”, 한국정보과학회, ‘98 봄 학술발표논문집, Vol. 25, No. 1, 1998.
- [33] 조은정, 신훈철, 이용규, 이금석, 홍영식, “웹에서의 한국 고문헌 검색 시스템”, 한국정보처리학회, ‘98 봄 학술발표논문 CD, 1998.
- [34] 홍영식 외 13인, “웹에서의 한국 고문헌 관리 및 검색 기술 개발”, 정보통신부, ‘97 초고속 정보통신 응용기술개발사업 최종 연구 보고서, 7월 1998.
- [35] 이용규, 홍영식, 이금석, 김정숙, 한인, 설승진, 신훈철, “한국 고문헌 데이터베이스”, 동국대학교, 동국논총 제 三十七, 12월 1998.