

# 고려대장경 인터넷 검색 및 열람

윤 용 석

고려대장경연구소

## 목 차

1. 서 론
  - 1) 개 요
  - 2) 고려대장경 인터넷 서비스의 목적
2. 본 론
  - 1) 고려대장경 전산화 작업 진행경과
  - 2) KTCS 코드체계
  - 3) 고 려 워 드
  - 4) 교정보조 프로그램
  - 5) 고려대장경 인터넷 서비스의 특징
  - 6) 고려워드 인터넷 버전
  - 7) 향 후 계 획
3. 결 론

## 1. 서 론

### 1) 개 요

세계문화유산인 고려대장경은 그동안 수 차례에 걸친 전산화 작업이 이루어졌다. 원문의 입력과 교정이 수행되었고, 고려대장경에 사용되고

있는 한자를 모두 수록하기 위해 4바이트 코드체계(KTCS)가 채택되었으며 폰트 및 관련 툴 들이 개발되었다. 이러한 고려대장경 전산화의 성과를 바탕으로 99년 12월말에 고려대장경 인터넷 서비스(<http://www.sutra.re.kr>)를 개시한다.

## 2) 고려대장경 인터넷 서비스의 목적

전산화된 고려대장경을 인터넷상에 공개함으로써 불교자료의 대중 보급을 꾀하고 국내외 불교학 연구를 위한 불교문헌정보 인프라를 조성하며, 불교정보의 연구기반이 되는 DB검색 시스템을 구축할 뿐 아니라 다양한 검색 서비스 제공으로 불교연구의 활성화를 꾀하여 고려대장경이라는 전통문화 유산의 현대화, 세계화를 하는데 있다.

## 2. 본 론

### 1) 고려대장경 전산화 작업 진행경과

현재까지 고려대장경 전산화 작업은 원문 입력 이후 이체자 폰트 추가와 함께 수 차례의 교정을 실시하여 완성된 경판 데이터를 얻게 되었다.

연도에 따른 전산화 진행 상황은 다음과 같다.

- ▶ 95년 고려대장경 원문 입력: 고려대장경 전산화 작업은 삼성의 훈민정음으로 1차 전산 입력을 하였고 이체자 폰트를 사용할 수 있는 환경의 고려워드 파일로 변환하여 편집 및 교정을 하였다.
- ▶ 97년 KTCS(4 바이트 코드 체계) 설계
- ▶ 98년 - 고려워드 설계
  - 고려대장경용 한자 FONT 제작: 교정 및 수정 작업에 필수

적으로 소요되는 8만여자의 폰트를 개발하여 이를 고려위드에 탑재하여 원문 교정 및 수정작업에 활용하고 있다.

- 교정 보조 프로그램 제작

▶ 99년 - 인터넷 검색엔진 개발

- 불교용어사전, 한자 자전 개발

12월 고려대장경 인터넷 서비스 실시

## 2) KTCS 코드체계

### ① KTCS 코드체계의 설계목적

KTCS 코드체계 설계의 주목적은 고려대장경에 출현하는 모든 한자에 대한 코드부여, 저장, 검색, 출력을 가능하게 하는데 있다.

고려대장경의 최초 입력시에는 이체자에 대한 정보가 없었고 그래서 문서 편집기로 사용되던 훈민정음으로는 예상을 뛰어넘는 이체자의 출현량을 감당하기 어려웠고 그에 상응하는 폰트를 수용하기 어렵다는 결론에 이르렀다. 그런 까닭에 앞으로 어떤 문자로 출현하더라도 그것을 수용할 수 있는 코드 시스템이 필요하게 되었고 팔만시스템은 앞으로 발췌될 모든 이체자들을 수용할 수 있는 4바이트 코드 시스템(KTCS) 개발하였고 또 이 코드 시스템에 적용되어질 수 있는 폰트들을 제작하였다. KTCS 코드체계는 지금까지 발견된 이체자 이외에도 새롭게 출현될 문자들에 대해서 검색과 정렬의 기준을 부여할 수 있으며 기존 문자에 대한 정보의 편집 또는 새로운 문자를 코드 시스템 내에 포함할 수 있는 방법을 제공하였다. 또 이를 적용한 에디터를 제작함으로써 KTCS 코드체계가 실제 적용될 수 있었다.

## ② KTCS 코드체계의 특징

### 가. 4 바이트 체계

KTCS 코드체계에서 각 문자는 32비트로 표현된다. 현재 한자권 국가에서 채용하고 있는 16비트 문자 체계는 최대 65536자이고, 실제로는 약 2만자 정도를 포함할 수 있을 뿐이다. 이것은 고려대장경에 수록된 한자의 일부만 표현할 수 있다. 따라서 KTCS 코드체계는 2바이트 체계의 부족한 문자 공간을 해결하기 위해 4바이트 구조를 채택하였고 그럼으로써 KTCS 코드체계에서 표현할 수 있는 문자의 총 수는 약 43억 자에 달한다.

### 나. 플레인 텍스트(plain text)

플레인 텍스트는 언어 식별자, 폰트 크기, 색깔, 하이퍼텍스트 링크 등의 추가적인 정보가 포함되지 않은 일련의 단순한 문자코드들로서 공공의 목적으로 사용되며, 표준화되어 있고, 쉽게 읽을 수 있다.

### 다. 유니코드 기반, KS, JIS, GB, CNS, BIG-5 코드 체계 호환

KTCS 코드체계 설계에는 다른 코드체계와의 호환성을 고려하였다. 이것은 KTCS 코드체계로 구성된 문서가 이 코드체계를 사용할 수 없는 환경에서 정보손실을 최소화할 수 있도록 하며, 현존 코드 체계를 지원하는 폰트의 사용성을 높일 수 있게 한다. KTCS 코드는 유니코드를 기반으로 하고 있고 유니코드와 KSC 5601을 포함한 한자권 국가의 기본 문자 체계간의 변환 테이블이 존재한다.

### ③ KTCS 코드의 장점

#### 가. 이체자를 포함한 고려대장경의 모든 한자 수용

기존의 2바이트 코드 체계로는 이체자를 포함한 고려대장경의 한자들을 모두 수용할 수 없었지만 KTCS 코드체계를 채택하였을 때는 고려대장경의 한자뿐 아니라 현재 지구상의 존재하는 모든 형태의 문자들을 모두 표현할 수 있을 정도로 방대하며 이후 확장성에 있어서도 아무런 제약이 없다.

#### 나. 효율적인 정렬 수행

KTCS에서 한자는 모두 상위 2바이트를 그 정자체 코드로 가지고 있다. 유니코드에서의 한자배열 기준은 첫 번째가 부수, 두 번째가 부수를 제외한 획수이다. 따라서 KTCS 코드로 구성된 문자열을 코드 값의 순으로 정렬할 경우, 유니코드 한자 배열순으로 나열된다. 또한 상위 2바이트 다음에 획수 부분이 오게 되므로 유니코드에 정의된 한자 이외에 새로 추가된 한자들에 대해서도 유니코드 정렬방식으로 정렬되는 특징이 있다. 이러한 정렬 특성은 도움말의 목록 혹은 점진적 검색(incremental search)에 한자가 사용될 때 그 구현을 용이하게 만든다. 부수별 정렬 또한 쉽게 구현될 수 있다. 모든 한자의 부수가 그 정자의 부수와 동일하다고 가정할 경우, 유니코드에서 같은 부수의 한자는 연속된 코드 할당 값을 가지는 블록으로 존재한다. 따라서 각 부수에 대한 블록의 시작 코드 위치 및 마지막 코드 위치를 데이터베이스화하면 부수별로 한자를 정렬할 수 있다. 반면 음에 따른 검색은 그리 효율적이지 못하다. 동일한 음의 한자 코드 값들간에는 아무런 연관 관계도 존재하지 않으므로, 음에 따른 한자들 데이터베이스의 구축이 깔끔하지 못하다. 이 점은

KTCS의 설계에 있어서 음에 따른 논리적 구조의 중요성에 낮은 비중을 두었기 때문이다. 그러나 음에 따른 분류는 국가별로 다를 수 있으므로, 국제화의 관점에서 그리 중요한 기준이 될 수 없다. 또한 음에 따른 논리적 구조의 지원이 미약하다는 점이 구현이 불가능하다는 것을 뜻하지는 않는다.

#### 다. 효율적인 검색 가능

KTCS에서는 기본적으로 정자에 대한 정보를 코드 내에 포함하고 있다. 또한 약간의 메타정보 구축만으로 부수 정보도 쉽게 얻을 수 있다. 따라서 다양한 검색에 있어서 큰 효율을 보일 수 있다. 먼저 찾고자 하는 특정 한자 혹은 한자열에 대한 검색은 문자 코드의 유일성이 유지되는 한 아무 문제 없이 수행된다. 또한 찾고자 하는 한자 혹은 문자열에 대한 정자 한자(문자열)만을 입력하더라도, 이를 정자로 하는 모든 이체자들도 검색결과에 포함되도록 할 수 있다. 이는 상위 2바이트 값만을 비교함으로써 쉽게 수행된다. 특정 부수를 가진 한자에 대한 검색도 마찬가지로 쉽게 구현될 수 있다.

#### 라. 뛰어난 확장성

KTCS 코드체계에서는 새로운 한자의 추가가 매우 쉽게 이루어질 수 있다. 통합 코드관리 시스템에서는 각 정자체, 획수 쌍에 대한 최신 인덱스를 유지하게 된다. 최신 인덱스는 현재까지 정의된 한자들 중 해당 정자체, 획수 쌍에 대응하는 한자들의 개수를 나타낸다. 새로 추가될 한자는 정자체 유니코드 값, 획수 값의 정보를 통합코드 관리 시스템에 입력함으로써 새로운 코드 값을 부여받는다. 이 때 시스템 내부에서는 인덱스 값이 하나 증가하게 된다.

마. 정자체를 쉽게 판별

모든 한자는 자신의 정자체에 대한 유니코드 값을 상위 2바이트 값으로 가지고 있으므로 코드 값을 관찰하면 바로 정자체를 판별할 수 있다.

바. 각종 논리적 구조에 대한 메타 정보 구축 용이

KTCS에서 코드 값 자체에 포함된 논리적 구조는 정자체에 따른 구조가 있다. 새로 추가되는 한자는 높은 인덱스 값을 갖게 되며, 유니코드 정의 한자는 0의 인덱스 값을 가진다. 새로 추가되는 한자의 경우, 빈도가 높을수록 낮은 인덱스를 갖도록 정의한다면, 인덱스가 빈도 정보를 표현하게 된다. 부수에 따른 구조는 각 부수에 대한 한자 블록 시작과 끝 위치 정보를 구축하면, 단 한번의 테이블 검색으로 쉽게 지원된다. 총 획수에 따른 구조 또한 쉽게 지원된다.

④ 통합 코드관리 시스템

유니코드를 바탕으로 한 KTCS 코드를 관리하기 위해서는 유기적으로 연결된 관리 시스템이 필요하다. 통합 코드 상의 이체자의 추가가 입력 작업과 병행되기 때문에 고정된 코드 생성만으로는 입력이 불가능하고 따라서 확장 가능한 코드를 관리할 수 있는 기능이 필요하다.

이 시스템은 고려대장경 코드를 입력하는 입력기와 고려대장경 코드를 관리하는 코드 관리기, 고려대장경 코드의 관련 정보를 관리하는 데이터베이스로 구성된다.

고려대장경의 한자를 입력하는 과정에서 KTCS 코드가 확장 변경되기 때문에 입력기와 코드 관리기는 서로 밀접한 관련을 가진다. 입력기는 현재까지 완성된 고려대장경 코드를 바탕으로 문서를 입력할 수 있어야

하고 코드 관리기는 입력기가 항상 정확한 코드로 문서를 편집할 수 있도록 입력기의 현재 코드를 갱신할 수 있도록 하여야 한다.

#### ⑤ 코드관리기(코드관리 데이터베이스 시스템)

코드관리기는 정의된 모든 한자들에 대한 데이터베이스를 유지하고, 새로운 문자의 정의 및 기존 문자 정보의 수정 및 삭제 기능을 제공하는 프로그램이다. 또한 한자에 대한 각종 논리적 구조(정보)를 담고 있는 메타 정보 파일을 생성하여 에디터에 제공하는 역할도 담당한다. 메타정보 파일에는 각 한자코드를 음 기준으로 정렬한 테이블, 부수 기준으로 정렬한 테이블, 입력방식(정마)기준으로 정렬한 테이블 등이 포함된다. 이 메타 정보 파일은 에디터에서 한자를 입력하거나 검색하는 데 있어서 데이터베이스를 거쳐야 할 때 들게 되는 오버헤드를 줄이는 역할을 한다. 특히 음 테이블, 부수 테이블, 정마 테이블 등은 한자 입력시에 효율적으로 사용되며, 폰트 매핑 테이블 한자 출력에 있어서 필수적이다.

그림 1은 고려대장경 코드 관리기의 실행모습을 보인 것이다. 리스트 윈도우가 전체 한자들에 대한 정보를 일정한 기준으로 정렬하여 보여주고 각 라인은 하나의 문자에 대한 정보를 담고 있다. 각 칼럼은 문자 정보들을 구분한다. 문자에 대한 정보는 데이터베이스 내에서의 식별자(ID), 정마 코드값(JungMa), 정자체 정보(Std), KTCS 코드 값의 16진수 표현, 최근 변형날짜(Date), 총획수(Cnt), 부수(Radix), 부수를 제외한 잔여 획수(RCnt), 정자체 여부(IsStd), 음(Snd), 폰트 파일 식별자(FID), 폰트 파일 내에서의 2바이트 코드값(FIndex), 유니코드 여부(IsUni) 등이다.

코드관리기는 다음과 같은 기능들을 가지고 있다. .

- 코드의 추가, 삭제(그림 2)
- FONT 및 META 정보관리



- 코드 테이블(CDB) 생성
- 새로운 한자 및 이체자 등록
- KTCS를 사용하는 APPLICATION(WEB DB, 고려워드)에 표준화된 코드 테이블 제공

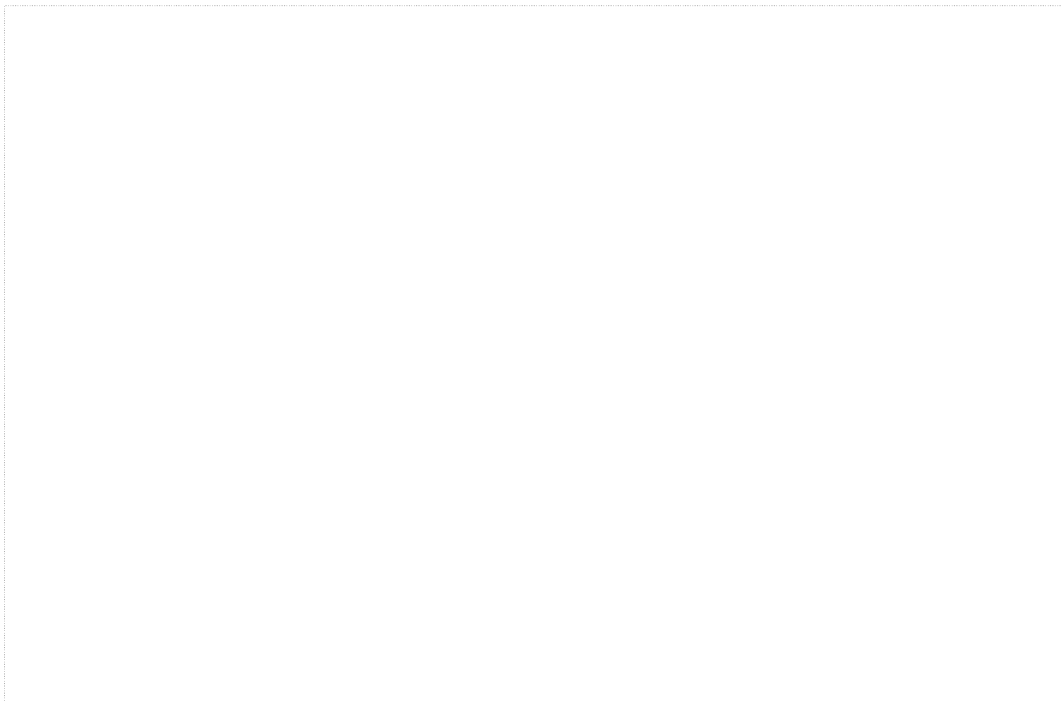


그림 1. 코드관리기



그림 2. 코드 추가

### 3) 고려워드

고려워드는 8만자에 이르는 한자폰트의 인코딩 기술을 세계최초로 개발, 이를 탑재하고 KTCS 코드를 바탕으로 이체자를 비롯한 한자 및 유니코드 상의 문자를 입력, 편집, 출력하기 위한 워드 프로세서로서 고려대장경 원문의 교정 및 수정작업에 필수적인 어플리케이션이다.

고려워드의 특징은 효율적인 시스템 구조 설계로 메모리 관리에 중점을 기울였다. 고려워드는 16만여개의 고려대장경 파일을 동시에 관리해야 할 뿐만 아니라 약 10만여자에 달하는 단일 폰트를 사용해야 한다. 고려대장경 파일은 본문만 약 2기가 바이트 이며 DB 검색을 위한 색인어를 포함한다면 10기가 바이트 이상이 된다. 그리고 10만여자에 달하는 폰트는 그 자체용량만도 30~50 메가 바이트에 달하는 방대한 분량으로서 일반적인 설계구조로는 메모리에 로딩할 수 없다는 난점이 있다. 고려워드는 이의 기술적인 해결에 중점을 두었고 기존의 워드프로세서 프로그램들이 단일적, 독립적, 1회적인 파일 포맷 형태를 지님으로써 생산된 파

일들을 DB 자료로 활용할 수 없었던 단점을 개선하여 고려워드 파일포맷형태 그대로 DB로 전환되도록 하였으므로 방대한 분량의 고문서 데이터베이스에 적합하도록 설계되었다.

### ① 고려워드의 기능

가. 워드프로세서 기본기능 내장(글꼴변경, 서체지정 등)

고려워드는 일반적인 워드프로세서의 기능을 보유하고 있으면서, 고려대장경 전산화를 위한 특화된 어플리케이션이다.

나. 10 만 자 이상의 한자 표현 가능

수십만자의 한자를 무리 없이 사용할 수 있도록 대용량 출력엔진을 설계, 탑재한 프로그램으로 동양권의 모든 고전 전적을 전산화 할 수 있다.

다. kwp 문서의 이미지 출력 기능

고려워드로 작성한 파일을 이미지화함으로써 고려워드를 사용하지 않는 고려대장경 인터넷 서비스 이용자들이 이미지 파일로 서비스를 제공한다.

### 고려워드 문서의 이미지맵화

고려워드를 가지지 않은 사용자가 경판 자료를 고려워드 사용자와 똑같은 형태로 볼 수 있기 위해서 워드 문서의 출력 형식을 Device Context 전환법을 이용하며 이미지화 하는 기능의 구현한다. 이 부분에 대해서 팔만시스템은 이미 1998년에 텍스트와 데이터베이스 연동을 실험하였으며 앞으로의 과제는 이렇게 생산되어진 이미지 파일의 이미지 맵핑 기술(이미지화된 문서내의 각 문자단위로 분할을 하고 분할된 문자에 코드값, 메타 정보 등의 정보 가중치를 맵핑하는 기술)을 적용하는 것이다. 이미지 맵핑이 되었을 때 고려워드를 가지고 있지 않은 사용자는 경판 이미지 안에서 고려워드에서와 같이 검색 등의 기능을 이용할 수가 있다.

#### 라. 가로쓰기 및 세로쓰기 지원

고려워드는 가로쓰기 외에도 세로쓰기를 기본적으로 지원하여 고려대장경뿐 아니라 어떤 고문서 전산화에도 적합하다.

#### 마. 다양한 한자 입력 방식 지원

##### 가) 정마방식 입력

한자의 음, 부수, 획수를 몰라도 한자의 모양으로 입력하는 방법으로 입력방식을 익히고 나면 쉽게 입력할 수 있다. 고려워드에서 고려대장경을 입력하는 주된 방법이다. 고려워드에서 이 방식을 선택하면 정마입력 윈도우가 나타난다. 입력은 A~Z까지 최대 4개의 문자와 0~9까지의 숫자로 입력을 하게 되며, 입력하는 도중에 현재까지의 조합에 해당하는 후보 글자들이 나타난다. 이때 후보 글자를 숫자로 선택하거나, 계속해서

정마를 입력한다.



그림 3. 고려워드의 정마입력 방법

#### 나) 한글음 입력

입력의 형태는 정마 입력법과 같으며 정마값 대신에 한글입력을 하는 차이가 있다. 그러나 한글음을 알고 있는 한자에 대해서만 입력할 수 있다는 단점이 있다. 대부분의 워드프로세서에서 채택하고 있는 입력방식이지만, 음을 알고 있다고 하더라도 해당 글자를 찾는 시간이 걸리는 문제가 있어 고속입력에는 부적합하다.

#### 다) 부수획수 입력

가장 보편적인 입력방법이고, 고려대장경 인터넷 서비스에서도 부수획수 입력법을 사용하여 브라우저 내에서 한자를 입력할 수 있게 하고 있다.

#### 라) 코드표 입력

##### 바. 클립보드 지원

고려워드는 운영체제와 다른 코드체계를 따르고 있으므로 클립보드 기능은 필수적이다. 고려워드에서는 클립보드를 지원함으로써 4바이트를 2바이트로, 2바이트를 4바이트로 전환하여 프로그램간에 데이터 교환이 가능하도록 하였다. KTCS 코드 체계에서는 상위 2바이트를 정자체 코드를 가지고 있으므로 2바이트로 전환하였을때 정자체로 전환이 된다. 그리고 반대로 2바이트를 4바이트로 전환하였을때는 하위 2바이트 0000를 추가하여 4바이트로 전환한다.

#### 4) 교정보조 프로그램

교정보조 프로그램은 방대한 분량의 고려대장경 원문 교정에 소요되는 시간, 인력, 비용을 절감하기 위해 개발한 전문적이고 특수한 목적의 프로그램이다. 기본적으로 영인본과 입력본의 판형을 일차 입력시에 동일하게 유지하였기 때문에 각 라인별 문자열의 개수는 동일하다. 그리고 문장 사이의 별첨 또한 입력본이 기본 판형을 충실히 하였기 때문에 동일하다. 따라서 교정 보조 프로그램에 의한 정확한 데이터 교정을 기대할 수 있다.

### ① 교정보조 프로그램의 필요성

팔만대장경은 총 수록 글자 52,801,771 글자에 달하는 방대한 분량이다. 이런 방대한 분량의 단일문서를 단순대조교정방식을 채택하여 교정을 수행했을 때 교정횟수가 증가할수록 난이도가 증가하여 교정 작업을 까다롭게 하였다. 이에 교정 보조 프로그램을 설계하게 되었고 그로 인해 교정에 따른 많은 비용과 시간을 줄일 수 있게 되었다.

### ② 교정보조 프로그램의 목적

전산화된 대장경 판본의 체계적 교정을 위하여 각 판본 파일에 기록된 한자들을 전체 이미지로부터 분리해내어 대응되는 한자 코드에 자동으로 연결시켜줌으로써 이후 교정작업을 하거나 문자의 형태에 대한 정보를 얻어내는 등 원본이미지를 참조하는 대량 작업에 있어 중요한 정보를 제공해 줄 수 있다. 특히 교정작업에 있어서는 한자코드별로 일괄교정을 가능케 하며 교정을 하는 전문가의 역할을 분리시킬 수 있어 작업의 효율과 질을 증가시킬 수 있다.

### ③ 교정보조 프로그램의 교정 방법

- 교정방식은 segment화 되어 있는 고려대장경 비트맵원본과 고려대장경을 입력한 kwp 파일을 한 화면에서 보면서 교정한다.
- kwp 문서를 열 때 비트맵원본이 함께 열린다.
- 비트맵원본과 kwp 문서의 글자 이동은 동시에 함께 한 글자씩 이동할 수 있고, 각각 따로 글자를 이동할 수도 있게 한다. 그리고 사용자가

지정한 임의의 글자로 이동할 수 있다.

- 교정은 kwp의 일반입력방식(정마, 음가, 부수획수)을 이용한다.
- 비트맵원본과 kwp 문서의 글자 한 자씩 확대하여 볼 수 있고, 비트맵원본 전체 보기와 kwp 문서 전체 보기를 할 수 있다.
- 교정이 끝난후 교정된 문서는 kwp로 저장한다.
- 비트맵원본 파일 위치와 kwp 문서파일 위치를 지정하여 사용할 수 있도록 환경설정을 할 수 있다.



그림 4. 교정보조 프로그램



## 5) 고려대장경 인터넷 서비스

### ① 고려대장경 인터넷 서비스의 특징

- 웹 브라우저만으로 검색하는 기능과 고려워드를 이용한 검색기능을 제공한다.

- Windows, UNIX, OS/2, Macintosh, Linux 지원
- 가장 보편적인 2 종류의 웹 브라우저 지원  
(IE 4.0, Netscape 4.0 이상)

- KTCS 코드 지원함으로써 인터넷 상에서 모든 문자 표현 가능

현재 인터넷 상에서는 KS 코드인 4,888자 외에는 표현할 수 없는 것이 현실이지만 고려대장경 웹 데이터베이스는 KTCS 코드를 지원함으로써 8만여자에 달하는 고려대장경 문자들을 인터넷을 통해 제공, 열람, 검색할 수 있다.

- 동적인 경판 페이지 생성

각 홈페이지는 ASP 등과 같은 기본 폼 형태로 작성되어 검색 결과에 따라 홈페이지의 내용이 작성되어진다.

- 대장경 열람 및 검색 기능

원문은 고려워드 파일과 고려워드 파일의 이미지, 두 가지 방법으로 서비스된다. 고려워드 사용자는 고려워드 파일을 전송받아서 수정 또는 편집이 가능하다.

- 교학 연구자를 위한 전문검색 기능

검색 방법으로는 한글검색, 한자검색(음가, 부수획수 입력), 조건검색이 있다.

- KTCS 코드 지원 검색 엔진

웹 DB 검색을 위해 색인어(일자색인), 단어 검색, 인덱스 검색을 하는데 필요한 검색엔진이 KTCS를 지원하여, 이체자를 포함하여 사용자가 다양하게 검색을 수행할 수 있도록 한다.

- 웹상에서의 한자 입력 기능

정자 뿐만 아니라 이체자까지 입력가능한 구조로 되어 있다. 검색방법에 있어 한자를 보유하지 않은 국가에서 접속했을 경우에도 한자의 입력 및 검색이 가능하도록 설계하였다. 웹상에서의 한자 입력 방식으로는 부수획수 입력과 음가 입력을 제공한다. 음가 입력에서 불전맞춤법사전에 기록된 단어는 단어단위로 한자변환이 가능하다.



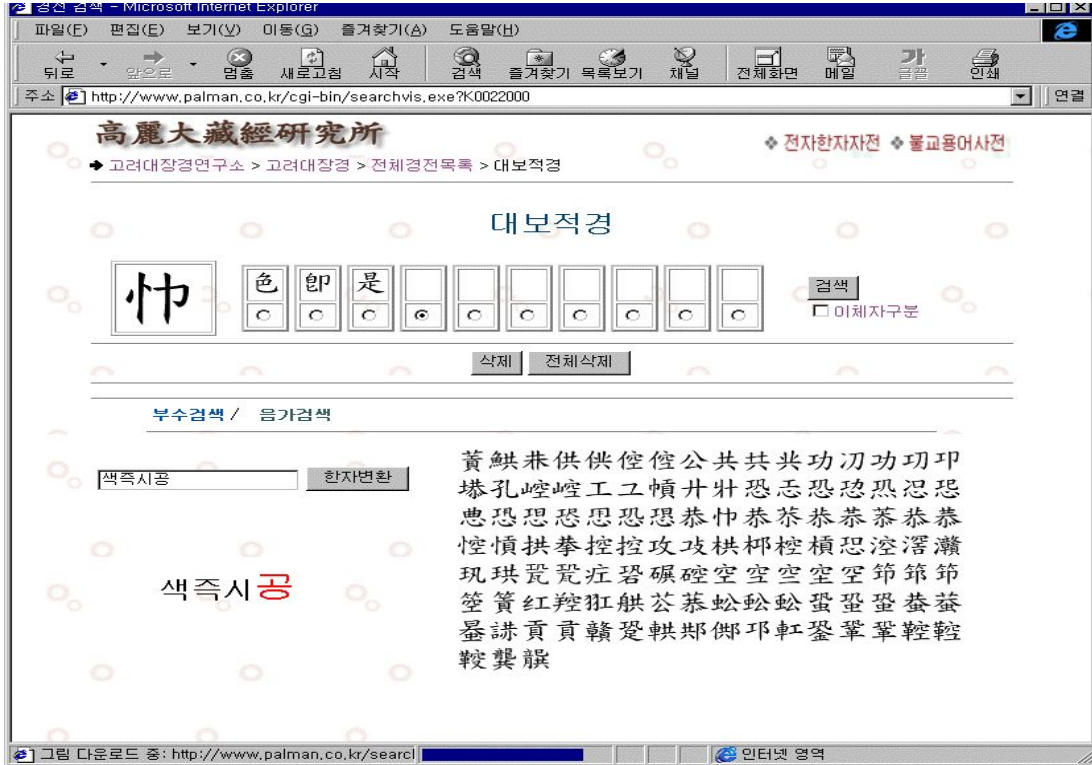


그림 6. 음가 한자 입력

- 정자 및 이체자 구분 검색 가능

한자검색에서 정자를 입력했을 때 해당 이체자까지 검색할 수 있고 또는 옵션을 주어 정자만 검색한다거나 이체자만을 검색할 수 있다.

- 일반인을 대상으로 한 불교용어사전, 전자한자사전 제공

경관 데이터 화면에서 찾고자 하는 불교용어나 한자의 의미를 쉽게 찾아볼 수 있다.

② 조건 검색: 연산자를 이용한 다양한 검색 기능 제공

## 가. 경전 DB

### 가) 경번호에 의한 검색 기능

- ▶ K-×××로 구분되는 고려대장경 분류번호와 이에 대응하는 경전 명칭을 기본으로 검색

### 나) 경전 명칭에 의한 검색(경명, 이명, 약명 등) 기능

- ▶ 기본 경정 명칭 외에 이명(異名: 다른 이름)으로 사용되는 경전에 대한 데이터 구성 및 검색
- ▶ 기본 경전 명칭외에 약명(略名: 줄임 이름)으로 사용되는 경전에 대한 데이터 구성 및 검색

### 다) 시대순 검색(서력, 간지, 왕\*년<천보3년> 등)

- ▶ 각 경전의 시대별 성립연대에 따른 데이터 구성 및 검색
- ▶ 서력기원에 의한 데이터 구성 및 검색
- ▶ 60간지에 의한 데이터 구성 및 검색
- ▶ 왕력(王歷)에 의한 데이터 구성 및 검색

### 라) 저자, 편·역자에 의한 검색 기능

- ▶ 각 경전의 저자(著者)별 데이터 구성 및 검색
- ▶ 각 경전의 편자(編者)별 데이터 구성 및 검색
- ▶ 각 경전의 역자(譯者)별 데이터 구성 및 검색

### 마) 공간분류에 의한 검색(인도, 중국, 한국 등) 기능

- ▶ 각 경전별 결집 장소 및 국가에 따른 데이터 구성 및 검색

### 바) 왕조별 분류에 의한 검색(당, 송, 기타) 기능

- ▶ 각 경전의 결집 왕조(王朝)에 따른 데이터 구성 및 검색

### 사) 함차(천자문순)에 의한 검색(문헌분류) 기능

- ▶ 각 경전의 함차(函次)에 의한 - 천자문순에 의한 구성 및 검색

### 아) 문헌구분에 의한 검색(신수분류기호) 기능

- ▶ 각 경전에 대한 신수대장경 분류기준에 의한 데이터 구성 및 검색

자) 형식분류에 의한 검색(대승, 소승, 경, 료, 론, 잡) 기능

▶ 각 경전에 대한 형식분류(形式分類)별 데이터 구성 및 검색

차) 종파별 분류에 의한 검색(아함, 법화, 화엄, 선종, 천태, 정토) 기능

▶ 각 경전의 종파(宗派), 종지(宗指)에 의한 데이터 구성 및 검색

카) 주제에 의한 분류

▶ 각 경전의 주제, 내용에 의한 데이터 구성 및 검색

예) 권선징악 등

타) 용도에 의한 분류

▶ 각 경전의 용도에 의한 데이터 구성 및 검색

예) 마음을 안정시키는 경전 등

파) 범주별 분류에 의한 검색 기능

① 누가(說者 who)

② 어디서(설법장소 where)

③ 누구에게(聽者 whom)

④ 무엇에 대하여(설법주제 what)



그림 7. 조건검색

#### 나. 고려대장경 해제 DB

- 고려대장경 해제 기본 데이터 처리
- 경번호에 의한 분류(K 분류, T 분류는 문헌분류 참조)
- 경전 명칭 분류(경명, 이명, 약명 등)
- 경전 명칭 분류(팔리어, 산스크리트어, 티벳어 경명 등)
- 문자별 분류(한글, 한자, 영문, 팔리, 산스크리트어, 티벳, 주음부호, 한어병음, 가타카나, 히라카나)
- 시대순 분류(서력, 간지, 왕\*년<천보 3년> 등)
- 저자, 편·역자 분류
- 공간 분류(인도, 중국, 한국, 일본 등)

- 왕조별 분류(당, 송, 기타)
- 함차(천자문순) 분류(문헌분류)
- 문헌 구분 분류  
(신수 분류기호, 속장경 분류기호, 한글대장경 분류기호)
- 형식분류(대승, 소승, 경, 료, 론, 잡)
- 종파별 분류(아함, 법화, 화엄, 선종, 천태, 정도 등)
- 고려대장경 해제 원문 파일 포맷 전환 처리  
(2바이트 → 4바이트)
- CGI 처리
- SQL 처리(Query 설계)



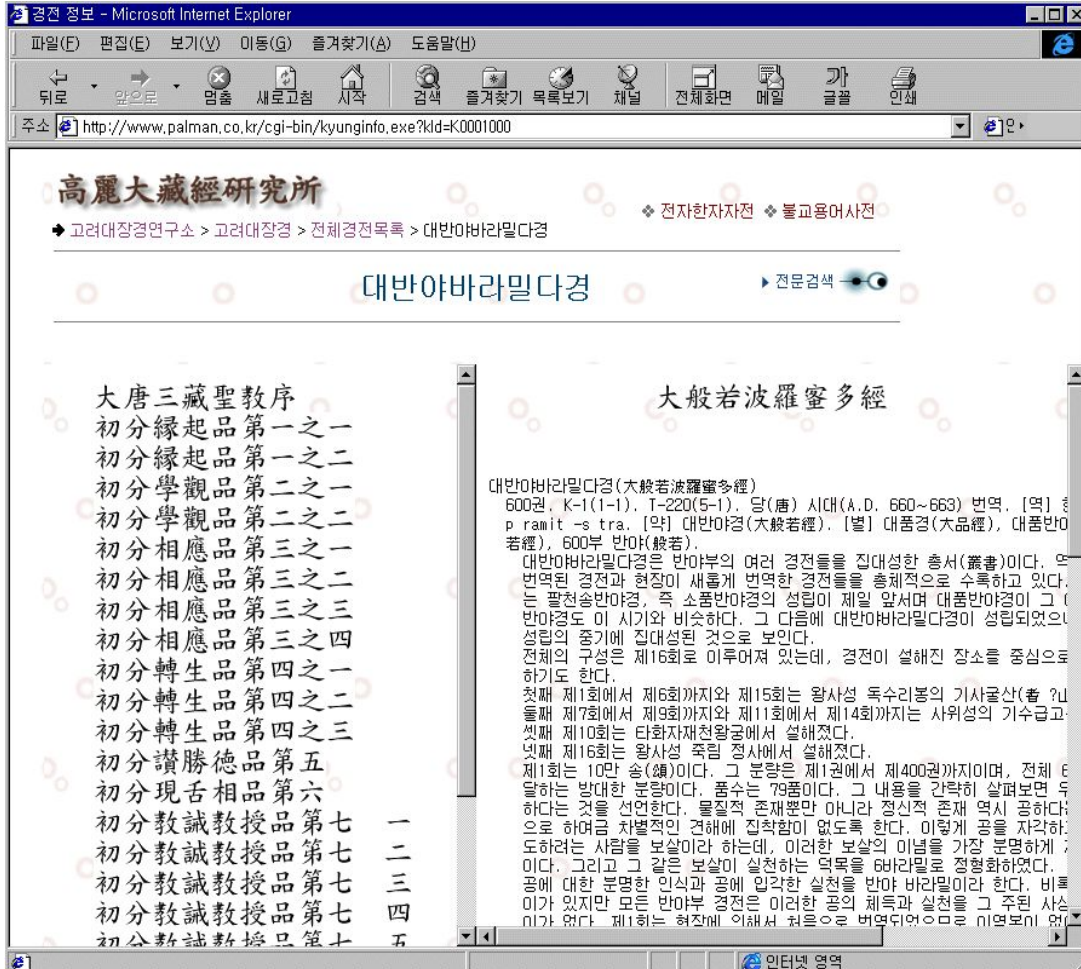


그림 8. 해제

③ 불전맞춤법 : 불교 용어의 특수성에 따른 제약 해결

한글로 경전명 등을 검색을 할 때 정확한 코드 값 하나만 찾게 된다. 그런데 한글로 검색을 했을 때 4바이트 코드에서는 한 글자에 한 음가만 배정되어 있으므로 몇 개의 음을 가진 글자(예를 들면 樂)나 범어를 찾았을 때 검색 결과가 나오지 않을 수도 있다. 그런 문제를 해결하기 위해서 불전맞춤법 기능이 추가되어야 하는데 이 기능은 클라이언트 상에

서 행해지는 것이 아니라 검색 엔진내에서 수행되는 것이다. 예를 들면 검색란에 “요산요수”라고 입력을 했을 때 검색엔진은 락산락수(樂山樂水)로 해석하여 해당 내용을 출력하게 된다.

④ 일자색인: 찾고자 하는 단어의 일부만 알고 있는 상태에서 검색이 가능

색인어의 일관성과 단일성을 위하여 한 글자에 대한 색인 검색을 할 때 앞 뒤 두 글자씩을 함께 출력하도록 하고 해당 색인이 있는 경판으로 이동할 수 있다.

⑤ 한자자전

고려대장경은 8만자 이상의 한자가 사용되었다. 한국에서 발간된 가장

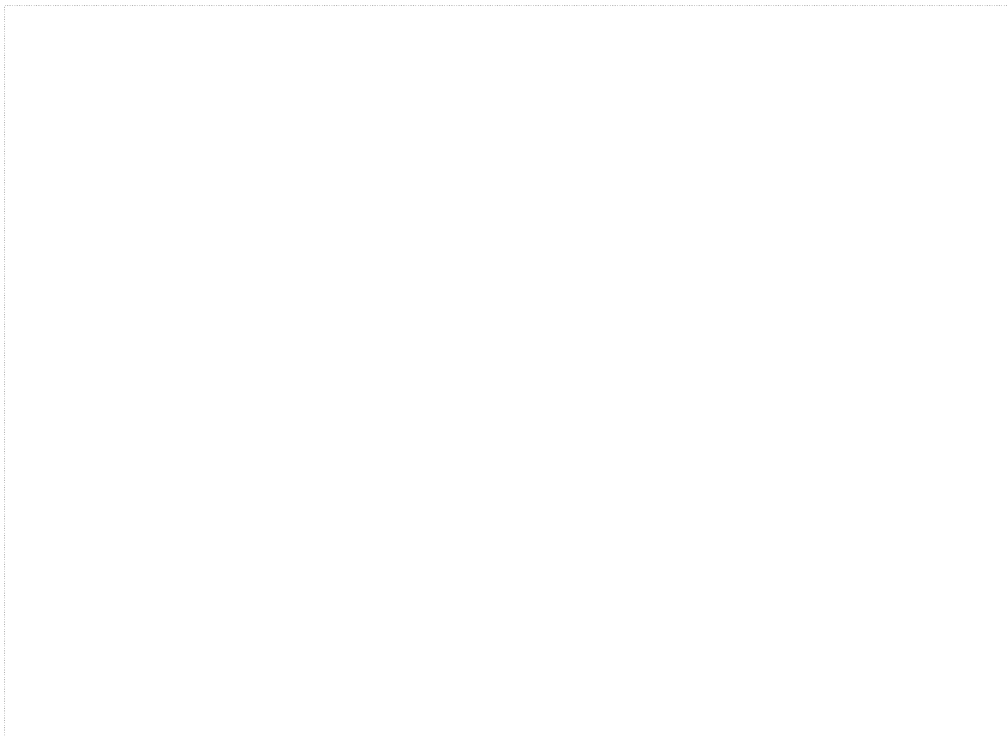


그림 9. 일자색인

큰 한자사전이 대략 3만 5천자를 수록하고 있다는 점을 감안하면 인터넷을 통한 한자사전의 제공은 필수불가결한 것이다. 한자 검색 화면에서 한자 데이터 입력 기능이 있고 검색을 하면 결과 화면이 카드형태로 한자의 음과 뜻을 설명한다.

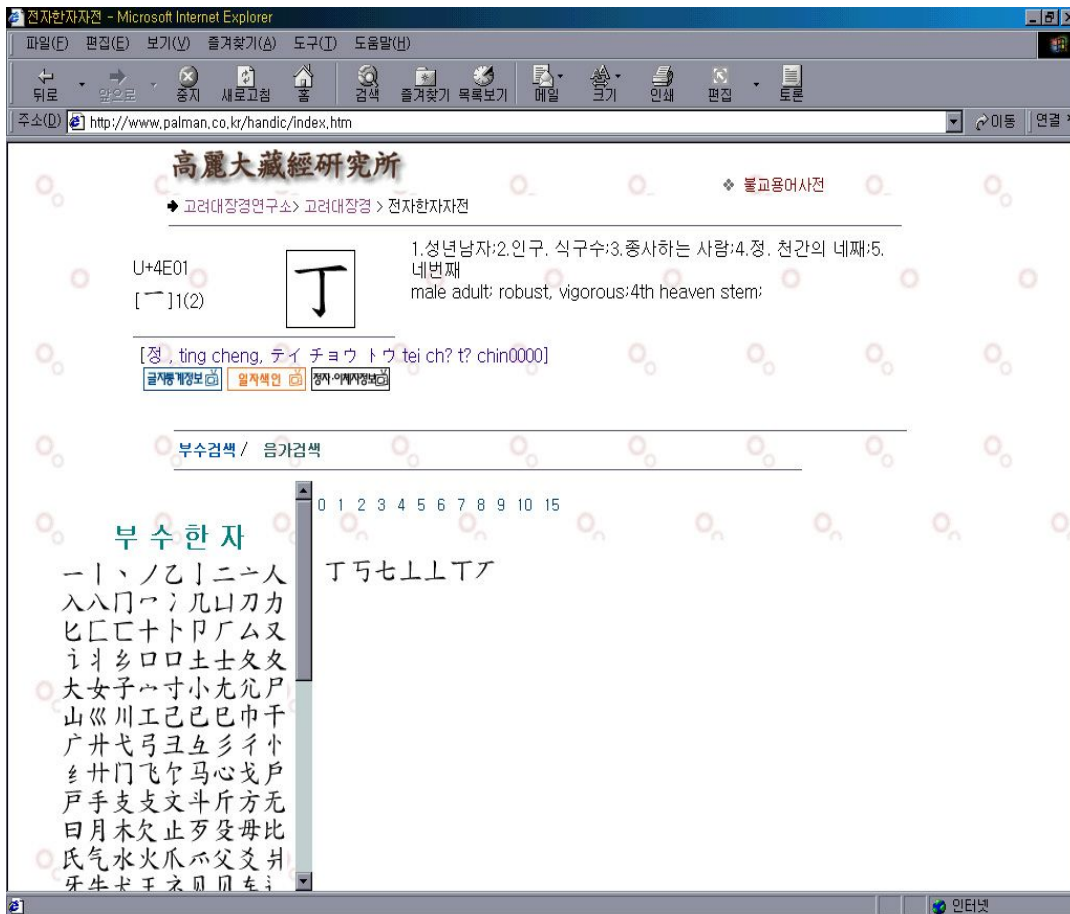


그림 10. 한자사전

### ⑥ 불교용어사전

표제어로는 한자와 한글이 들어가서 검색을 할 수 있다. 그리고 검색 결과에서는 용어에 대한 설명 외에, 산스크리트어, 팔리어, 티벳어까지 불교용어를 포함한다.

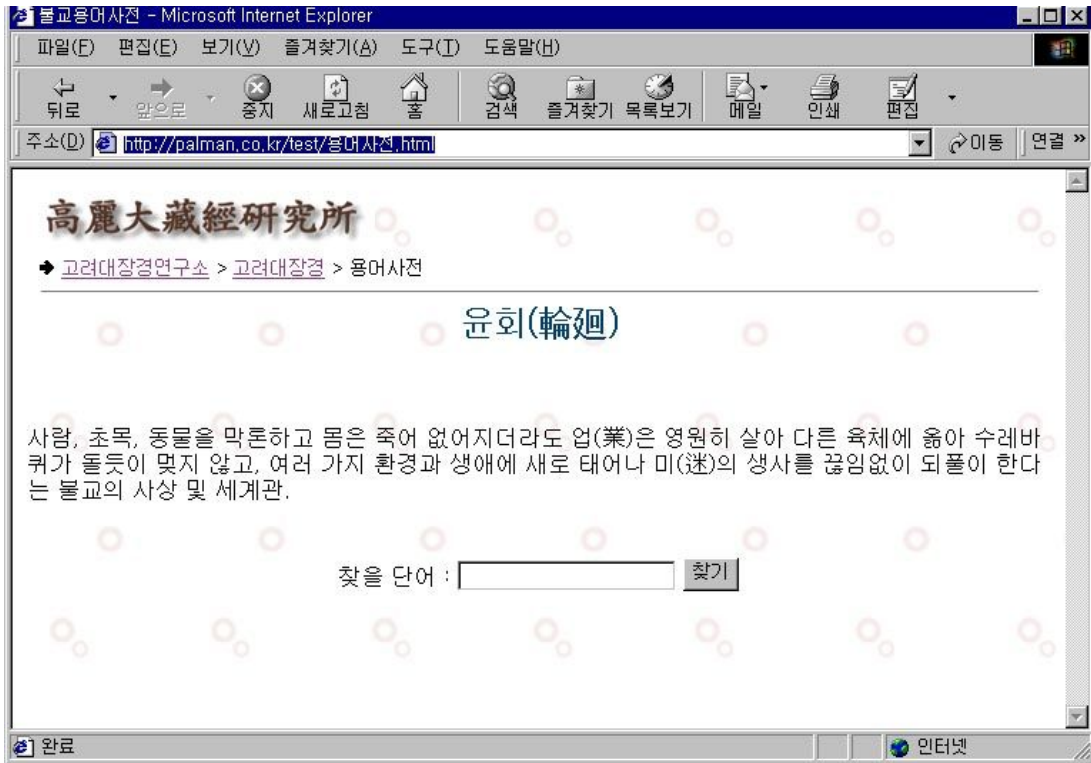


그림 11. 불교용어사전

## 6) 고려워드 인터넷 버전

고려워드 인터넷 버전은 고려워드의 차기 버전으로 고려대장경 인터넷 사이트로 직접 접속하여 검색과 열람을 수행할 수 있는 프로그램이다.

고려워드 인터넷 버전으로 고려대장경 검색 및 열람을 할 때의 장점은 다음과 같다.

- 검색어 입력이 용이

고려워드 자체의 기능을 이용하여 여러 방법(정마, 부수/획수, 음가 입력)으로 검색어를 입력할 수 있다.

- 클립보드를 통한 코드 변환 기능 지원
- 데이터의 신속한 출력

이미지가 아닌 텍스트로 다운로드 받으므로 경판 데이터를 빠른 속도를 볼 수 있다.

- 트루타입 폰트를 사용한 고품질 출력

웹 브라우저를 통해서 접속할 때 경판 데이터의 이미지가 출력되므로 확대를 했을 때에도 같은 품질을 기대할 수 없고 축소 등의 편집을 수행했을 때 품질이 유지될 수도 없다. 반면에 고려워드로 접속을 하였을 경우에는 어떤 경우에도 고품질의 출력을 얻을 수 있다.

- 고려워드 기능을 이용한 문서 편집 가능

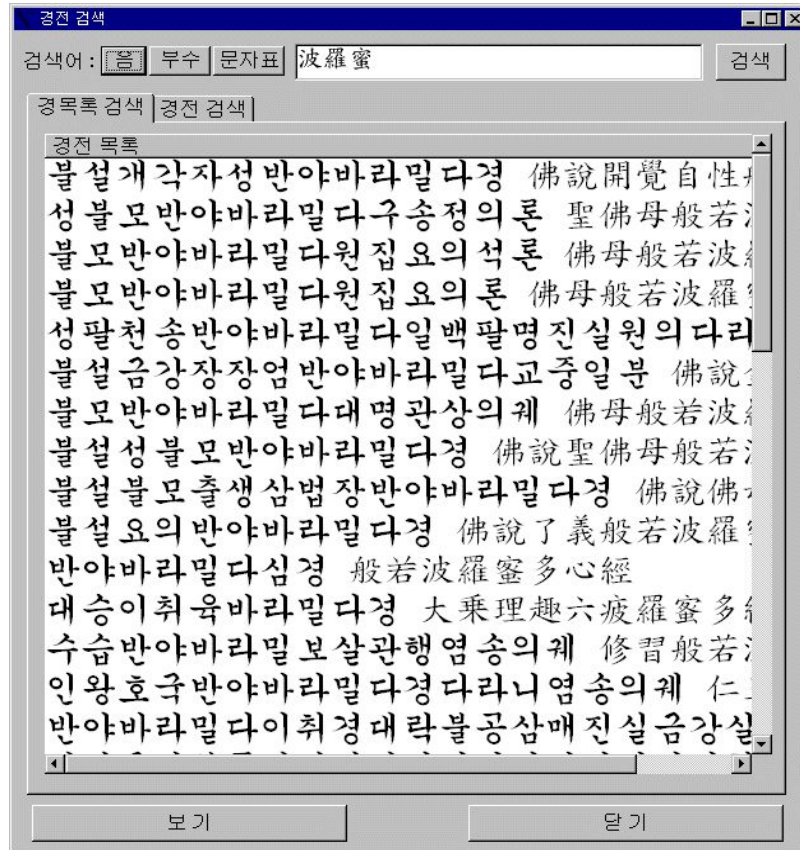


그림 12. 고려워드를 이용한 고려대장경  
사이트 접속 및 경전명 검색



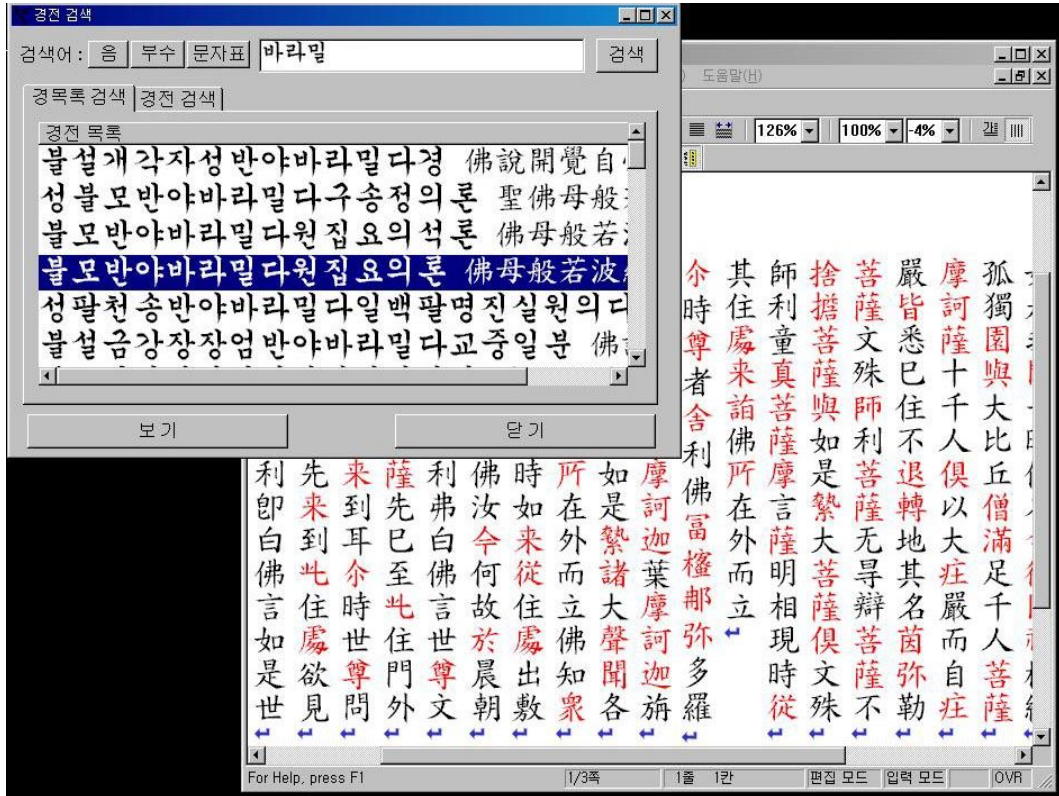


그림 13. 고려워드를 이용한 Full Text 검색

## 7) 향후계획

### ① 2000년 10월 고려대장경 CD-ROM 출시

현재 진행하고 있는 고려대장경 전산화는 웹 서비스를 기반으로 설계되지만, CD-ROM 형태로도 보급함으로써 누구나 고려대장경 경판을 연구 소장할 수 있도록 한다. 이 CD-ROM은 고려대장경을 연구하는 불교학자들에게 반드시 필요할 뿐만 아니라 일반인들에게도 소장 가치가 충분하다.

고려대장경 CD-ROM은 CD-ROM만으로 고려대장경의 모든 원문 및

해제를 검색할 수 있을 뿐 아니라 한자사전, 용어사전, 불전 맞춤법과 같은 다양한 기능들을 PC에서 사용할 수 있도록 한다. 또 함께 제공되는 고려워드와 10만자에 달하는 폰트를 이용하여 각종 문서를 편집할 수 있는 기능도 제공된다. CD-ROM의 성격에 맞도록 고려대장경에 관련한 각종 자료를 텍스트와 멀티미디어 요소로 포함시킬 예정이다.

## ② 이체자 사전 발간

고려대장경 이체자 정리작업은 고려대장경 전산화를 위한 입력과정에서 그 필요성이 제기되었다. 대장경 전산화 작업은 그 원판을 그대로 입력하는 것이 가장 이상적이라고 할 수 있다. 선부른 교감은 이후 연구자나 독자들에게 혼란을 야기할 수 있기 때문이다. 약 10만 여자를 수록한 세계 최대의 사전이라고 할 수 있는 북경판 한어 대사전에도 고려대장경에서 추출되는 이체자들이 대부분 수록되어 있지 않다. 이러 면에서 고려대장경 이체자 사전 발간을 위한 이체자 연구는 세계적인 학문적 성과를 달성할 수 있을 뿐만 아니라 이를 통한 고려대장경 연구의 난점을 해결할 수 있고, 고대 불교자료에 대한 연구를 용이하게 해 준다. 그리고 수년간 지속되어 온 고려대장경 이체자 연구는 상당한 성과를 거두어 이체자 사전을 발간할 수 있는 수준에 이르렀다.

## ③ 학술세미나

고려대장경 전산화를 진행하는 과정에서 이체자에 대한 연구, 검색의 편의를 위한 분류체계의 정립 등 여러 가지 성과가 있었다. 이러한 학술적 성과를 공유하고 더욱 발전시키도록 한다.



④ 지속적인 시스템 업그레이드

전산 환경의 변화에 신속히 대응하고 고려대장경의 연구결과를 즉시 반영할 수 있도록 시스템을 지속적으로 업그레이드한다.

⑤ 위 ①~④ 항을 기반으로 한 통합대장경 검색 시스템 개발

한글대장경을 포함한 통합대장경 검색시스템을 구축함으로써 대장경의 일반화, 대중화를 기대할 수 있다.

### 3. 결 론

지금까지 고려대장경의 전산화를 하는데 있어 인적 물적 자원의 부족 등 많은 어려움을 겪었지만 그만큼 끈질기고 오랜 노력과 연구 끝에 전산화를 할 수 있는 여러 개발툴이 만들어지고 그 외 제반 여건들이 갖추어졌다. 이체자를 포함한 한자의 코드체계, 입력 및 교정 솔루션이 개발되었고 KTCS를 지원하는 웹 검색 솔루션을 갖추게 되었다. 이런 결과물들이 축적되어서 고려대장경의 인터넷 서비스를 시작하게 되었고 이러한 고려대장경의 전산화는 결국 고려대장경을 현대화 세계화 할 수 있는 기반이 되어 줄 것이며 또한 국내외 불교학의 연구에 지대한 공헌을 할 수 있을 것으로 기대되고 있다. 또 향후에 한글대장경을 포함한 통합대장경을 전산화하려 할 때 지금까지의 전산화 작업에서 얻었던 노하우와 성과들이 통합대장경을 구축하는데 있어 기반이 될 것이다. 그리고 무엇보다 고려대장경의 전산화는 일반인들에게 불교를 더욱 가깝고 친근하게 느끼게 하여 불교를 생활불교로서 정착시키는 계기가 될 수 있을 것이다.