

유니코드 한자처리를 위한 입력기와 편집기의 설계 및 구현

# 유니코드 한자처리를 위한 입력기와 편집기의 설계 및 구현

Design and Implementation of An Editor Supporting Unicode Chinese Characters

한인\*, 이용규\*, 이금석\*, 홍영식\*, 한보광\*\*

In Han, 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

한국고문헌과 같이 한자를 많이 사용하는 문서를 KSC-5601 문자코드로 입력하기에는 지원하는 한자의 수가 크게 부족하기 때문에 보다 많은 한자를 지원할 수 있는 문자코드를 사용하여야 한다. 또한 전세계를 연결하는 인터넷을 통하여 입력된 문서를 공개하기 위해서는 한 나라에 국한된 문제코드로 문서를 입력할 경우 많은 문제점을 내포하게 된다. 이런 두 가지 측면을 고려하여 본 연구에서는 유니코드를 사용하였다. 그러나 유니코드 한자를 입력하고 편집할 수 있는 도구가 없기 때문에 본 연구에서는 유니코드 한자입력기와 문서편집기를 개발하였다. 유니코드 한자입력기는 유니코드 내의 모든 한자를 한글 음의 순서로 재배열하였으며, 유니코드 한자를 입력하기 위해서는 한글을 먼저 입력하고 해당 음의 한자로 변환하는 음전환방식을 사용하였다. 본 유니코드 한자입력기와 문서편집기를 사용하여 한국고문헌인 한국불교전서를 입력하였다.

## 1. 서론

이미 오래 전부터 미국과 중국어권 나라들을 중심으로 고문헌에 대한 전산화, 특히 불전 전산화에 대한 연구가 활발히 진행되었으며, 연구 목적뿐만 아니라 상업용 그리고 문화 유산의 가치를 드높인다는 측면에서 다양한 결과물들을 속속 발표하고 있다. 이에 비해 우리 나라는 찬란한 5천년의 역사 속에서 탄생한 소중한 고문헌 자료들의 전산화에 대해 등안시 해왔으며 해인사 고려대장경 연구소와 같은 일부 연구소에서 몇몇 전산화 과제가 진행 중이다[1].

한국 고문헌의 전산화는 다른 중국어권 나라와 마찬가지로 고문헌의 입력 관점에서 약 2~3만 자에 이르는 한자를 사용하여 기록된 고문헌을 처리하기 위한 폰트와 코드 문제를 들 수 있다. 고문헌 한자 코드에 대한 문제는 이체자(異體字)와도 관련이 깊은데, 이체자란 뜻과 음은 동일하지만 서로 다른 형태로써 기록되는 한자를 말한다. 이체자는 오자(誤字)나 탈자(脫字)를 포함하는 파자(破字)와 함께 부가적인 폰트와 코드를 필요로 하므로 고문헌 전산화에 있어 반드시 해결해야 할 문제이다[1].

많은 연구에서 입력 가능한 한자의 수를 늘리기 위해 자체적으로 폰트를 제작하고 임의로 코드를 부여하여 사용하는 예가 있다. 그러나, 이러한 경우 입력된 고문헌을 검색하기 위해 새로운 폰트를 설치하여야 하고, 이는 인터넷 서비스 특히 웹을 이용한 서비스 제공에도 큰 문제가 발생할 수 있다.

따라서 본 연구에서는 인터넷의 웹 서비스로 고문헌을 제공할 수 있도록 하기 위하여 유니코드[2,3,4]를 사용하였으며, 현재 한국인들이 사용하는 음에 의한 한자 입력 방식으로 유니코드를 입력할 수 있도록 유니코드 한자 입력기와 유니코드 문서 편집기를 개발하였다.

본 논문의 2장에서 문자코드와 기존 한자 입력시스템, 그리고 문서 편집기를 살펴보고 3장에서 본 연구에서 개발한 시스템을 소개한 후, 4장

에서 결론을 맺도록 하겠다.

## 2. 기존 음전환방식의 한자 입력시스템과 유니코드 문서편집기

### 1) 코드체계

불교전적을 전산화하기 위해서는 한자의 사용이 필수적이고, 한자를 사용하기 위해서는 한국뿐만 아니라 중국과 일본의 문자 코드를 고려할 필요가 있다. 따라서 본 연구는 한국에서 사용하는 문자코드인 KSC-5601에서 사용할 수 있는 한자의 수가 4888자이기 때문에 중국(대만 포함)과 일본에서 사용하는 문자코드 체계와 전세계에서 통용되고 있는 유니코드에 대해 살펴본다.

#### ① 일본 JIS 코드

JIS C 6226-1978은 일본에서 처음으로 제정한 표준 코드로서, 한자와 같은 표의문자를 컴퓨터에서 처리하기 위하여 제작되었다. 이 코드는 동아시아에서 사용되는 코드 표준, 특히 Big5와 같은 코드에 많은 영향을 주었다. 대략 6500 글자를 표현할 수 있으며, 1990년도 개정판은 JIS X 0208-1990으로 추가적인 5,800자의 한자와 함께 발표되었다. 또한 일본에서는 JIS 코드 이외에도 많은 업체에서 새로운 부호화(encoding) 방법과 함께 새로운 코드를 추가하고 있다.

#### ② 대만 Big5 코드

Big5는 13,051개의 중국 한자를 지원하며 사용빈도에 따라 두 개의 부분으로 나뉘어진다. 또한 각 부분에 있는 글자들은 획순에 따라 정렬되어 있다. 그러나 Big5 코드는 매우 서둘러서 만들어졌기 때문에 획순을

오인식하거나 동일한 한자가 중복되는 등 현재까지 많은 문제점들이 발생되고 있다.

### ③ 대만 CNS 코드

이 코드는 대만의 중국 국가 코드(CNC : Chinese National Code)로써 발표되었으며 48,027개의 한자를 위한 상형 문자 형태, 획수, 초두 머리 등을 정의하고 있다. 이 표준에 정의된 한자의 특징은 모두 40 x 40 픽셀의 폰트를 정의하고 있다는 것이다. 이 코드에서 사용된 방법은 코드 배정시 7개의 계층 구조로 나누었는데, 하위 계층일수록 자주 사용되는 문자들을 배정하였고, 상위 계층일수록 사용 빈도가 낮은 문자를 배정하였다. 또한 추가적인 문자 배정을 위하여 총 9개의 계층을 추가로 마련하였다. 이중 5개의 계층은 추가적인 표준 한자 지원을 위한 것이고, 4개의 계층은 추가적인 비표준 한자를 위한 영역이다. 현재 이 코드에는 55,000개의 한자를 위한 코드가 배정되어 있다.

### ④ 대만 CCCII와 EACC 코드

이 코드는 중국문자연구단체(CCRG : Chinese Character Research Group)에서 개발된 코드로서 초기 개발 목표는 동아시아의 모든 한자와 중국의 모든 한자를 포함할 수 있는 코드를 만들기 위함이었다. CCCII는 현재 75,000 개의 한자를 수용할 수 있는 코드 체계를 가지고 있으며, 특히 이체자에 대해 집중적으로 관심을 가지고 제작되었다고 할 수 있다. EACC는 약 16,000개의 한자를 포함할 수 있도록 한 CCCII의 축소판 코드 체계라고 할 수 있다.

### ⑤ 유니코드(Unicode)

유니코드는 Apple, Metaphor, RLG, Sun, Xerox 등 여러 컴퓨터 관련 업체가 결성한 유니코드 작업 그룹(UWG : Unicode Working Group)에

서 동아시아권 국가뿐만 아니라 다른 세계 모든 국가의 문자들을 포괄하는 새로운 코드 체계를 제정할 목적으로 만들어진 코드이다. 유니코드의 발표로 인하여 이때까지 지지부진하고 있던 ISO-10646 UCS의 다국어 코드에 유니코드를 사용하기에 이르렀다.

#### ⑥ ISO/IEC-10646 UCS

ISO/IEC-10646 UCS(ISO/IEC-10646:1993 Information technology -- Universal Multiple-Octet Coded Character Set, 이하 ISO-10646이라 칭한다)는 4바이트로 한 문자를 나타냄으로써, 보다 일반적인 문자 처리가 가능한 문자 집합과 그 부호화를 정의한다.

ISO-10646은 문자 집합의 보다 일반적인 적용을 위해서 canonical form이라는 개념을 사용한다. 각 문자는 문자들끼리 서로 구분되기 위해 4 옥텟의 16진수 이름을 갖게 되는데, 이것을 canonical 태그라 한다. 각 문자는 실제로 사용되기 위해 canonical 태그로부터 적당한 방법으로 부호화 될 수 있다.

이것은 ISO-2022와 같은 종전의 문자 표준이 코드영역에 직접 문자를 대응함으로써 문자 체계 자체에 제약이 주어진 것과는 달리, canonical 태그로 문자 집합을 정해 놓고 그 사용환경에 따라 부호화 하는 방법을 바꿈으로써 코드 사용환경이 바뀌더라도 문자집합 자체는 변경되지 않도록 한 것이다.

ISO-10646은 현재의 기술 상황을 고려하여 한 글자를 4-octet으로 표현하도록 하였다. 그러나 현재 환경은 한 글자를 한 바이트로 가정하고 있고, 또한 그 가정을 기반으로 한 표준들도 있는 상황이다. (대표적인 예 ISO-2022) 따라서, ISO-10646에서 규정한 문자집합을 실제로 사용하려면 어느 정도의 변환이 필요하다. 이것은 UTF(UCS Transformation Format) 이라 한다. 현재 제정되었거나 제정되고 있는 UTF들은 UTF-1, UTF-7, UTF-8, UTF-16이 있다.

이상과 같이 한자를 사용할 수 있는 많은 코드 체계가 있다. 하지만, 많은 코드들이 각 나라에서만 사용할 수 있도록 제정되어 있기 때문에 본 연구에서는 인터넷을 이용해 전세계인들이 공용할 수 있도록 전산화하기 위해 유니코드의 사용을 긍정적으로 평가하고 유니코드를 사용하였다.

## 2) 한자 입력 방식

한자를 사용할 수 있는 많은 문자코드가 있듯이 한자를 입력하기 위해서 각 나라마다 많은 입력 방법을 사용하고 있다. 본 연구에서는 우리나라에서 가장 많이 사용하는 음전환 방식을 채택했다. 음전환방식이란 우선 입력하고자 하는 한자의 음을 한글(완성형)으로 입력하고 해당음의 한자를 찾아 입력하는 방식이다.

일반적으로 음전환 방식으로 한자를 입력하기 위해서는 크게 두 가지 방법이 사용되고 있다. 첫 번째는 운영체제인 Windows에서 제공하는 IME를 이용하는 방법이고, 두 번째는 편집기에서 제공하는 독자적인 한자 입력 시스템을 사용하는 것이다.

첫 번째 방법인 IME를 이용하는 방법은 Windows에서 제공하는 KSC-5601 코드 한자인 4888자만을 입력할 수 있다. 따라서 한국 고문헌을 입력하기 위해 요구되는 한자의 수에 비해 크게 부족하다는 단점을 갖고 있다. 다음 [그림1]은 한글 Windows NT에서 하나의 한자를 IME를 이용해 입력할 때 나타나는 모습이다.

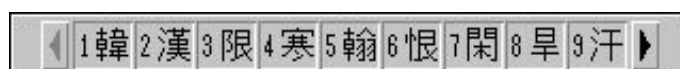


그림 22 MS-Windows에서의 한자 IME

두 번째 방법인 응용프로그램에서 제공하는 한자입력 시스템은

KSC-5601에서 지원하는 한자의 수보다는 상당히 많은 한자를 지원하지만 각 응용 프로그램에서 독자적으로 만든 폰트와 코드를 사용하기 때문에 다른 응용 프로그램과 호환이 되지 않는다는 단점을 갖고 있다. 다음 [그림 2]는 한글과 컴퓨터의 한글 응용 프로그램에서 제공되는 한자 입력시스템을 이용한 한자를 입력할 때 나타나는 모습이다.

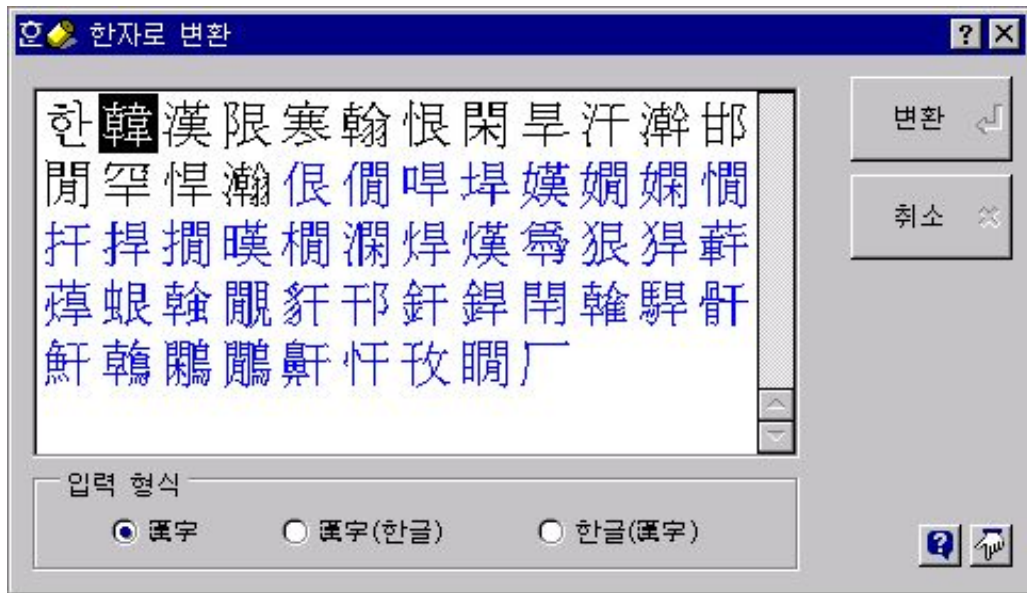


그림 23 한글에서의 한자 입력시스템 모습.

그러나 위의 한자 입력 시스템에서 KSC-5601이외의 한자는 해당 프로그램에서 사용할 수 있으나 인터넷을 위한 문서를 작성하더라도 KSC-5601 코드의 문자를 제외한 다른 한자들은 올바르게 출력되지 않는다.

현재 한글 윈도우즈 운영체제에서 사용되고 있는 유니코드 편집기로는 한글 윈도우즈 NT의 메모장이 있다. 이 프로그램은 일반적인 텍스트 문서를 만들 수 있기 때문에 다른 응용 프로그램과 호환이 된다는 장점이

있다. 본 연구가 진행될 당시에는 어떠한 유니코드 한자 입력기와 문서 편집기가 존재하지 않았다. 그러나 마이크로오피스 2000이 출시되면서 유니코드를 지원하게 되었고, 이를 이용한 문서 작성이 진행 중이다.

### 3. 음전환방식의 유니코드 한자입력기와 유니코드 문서 편집기

#### 1) 음에 의한 유니코드 한자 입력기

한글/한자변환 모듈은 크게 한글/한자 정보를 저장한 Database, Database를 빠르게 접근하기 위해서 구성되는 CTable class, 마지막으로 CTable를 통해서 얻어온 정보를 대화상자로 표현하고 사용자의 입력에 따라서 선택된 한자코드를 변환시켜주는 CKT Class의 세 가지 모듈로 구성된다.

##### ① 한글/한자 Database의 자료구조

한글/한자 변환테이블 Database에 입력될 자료들은 파일 형태 (CKT.db)로 저장되어 있다. 이러한 파일 형태의 자료를 만들기 위해 먼저 유니코드에 해당하는 한자들에 대해 음을 달고, 음별로 분류하여 해당하는 유니코드를 파일에 저장하였다. 각 한자들을 음을 달아 분류할 때 고려해야 할 점은 한자가 뜻 글자이므로 상황에 따라 여러 가지 음으로 읽혀질 수 있다. 따라서 이러한 한자의 특성을 살려서 한 글자의 음이 여러 가지인 경우 모든 음에 해당 한자가 나타날 수 있도록 하였다. 그리고 음들을 분류하여 입력할 때 임의의 무작위 순서로 파일을 구성하지 않고 한글음의 사전식 배열에 따라 정렬하여 저장하였다.

이렇게 한글음을 정렬하여 저장함으로써 한자로 변환하고자 하는 유니코드 한글음을 외부 파일에서 탐색하여 테이블에 적재할 때, 효율적으로,



그리고 빨리 탐색할 수 있는 알고리즘을 수행할 수 있도록 한 것이다. 물론 한자의 음을 달 때는 일차적으로 한자 옥편을 이용해서 달았다. 그러나 그 옥편에 나와있는 한자가 유니코드에 해당하는 한자들을 모두 포함하고 있는 것은 아니다. 최대한 많은 한자들에 대한 음을 달아 분류하려고 하였으나, 현재 많이 사용하고 있는 한자가 아니거나 혹은 우리나라에서는 거의 사용되지 않고 주로 일본이나 중국에서 사용되고 있는 한자들에 대한 음은 옥편에 나와있지 않았다. 그래서 이런 점들은 불교학 전문가들과 상의하여 가능한 많은 한자들에 대한 음을 파악하여 분류하였다.

한글음에 따라 유니코드의 분류된 파일의 내용을 보면 아래 [그림 3]과 같이 나타나게 되는데, 모두 16진수를 표현하는 4바이트 크기의 문자열로 저장되어 있다. 각 코드는 blank로 구분되어 진다. 한글의 코드는 0xAC00(가)에서부터 0xD7A3(힉)까지의 범위이고 한자는 0x4E00에서 0x9FA5까지의 범위에서 표현된다.

/ 가 / AC00
8EFB 52A0 8CC0 9050 53EF 5BB6 5CA2 5E4F 5FA6
93B5 6935 6687 73CE 73C8 9D1A 70A3 7241 726B
7271 728C 75C2 7615 7638 767F 7822 7A3C 7B33
801E 8175 8238 827D 82DB 8304 9553 9160 83CF

그림 3 한글음 '가'에 해당하는 유니코드 화일

파일형태로 표현된 위 Database는 독이성(readability)을 위해서 주석도 표현할 수 있도록 하고 있다. 첫 번째 '/'가 나타나면 주석의 시작이고 다시 한번 '/'가 나타나면 주석의 끝으로 인식하도록 표현한다. 그러므로 주석사이에 '/'를 표현할 수 있는 방법은 없으니 주의가 필요하다.

② CTable Class

CTable class는 한글/한자 변환테이블(CKT.db)를 통해서 자료를 입력 받아서, 그 자료를 한글과 그 한글에 해당하는 한자들의 집합으로 자료 구조를 변형시킨다. CTable이 CKT.db를 모두 Loading하여 저장하면 다음 [그림 4]와 같은 형태로 저장된다.

Table[1]	Table[2]	Table[3]	Table[4]	.....
Table[KORSIZE]				
AC00	AC01	AC02	AC03	..... D7A3
CCCC	CCCC	CCCC	CCCC	..... CCCC
CCCC	CCCC	CCCC	CCCC	..... CCCC
.....				
.....				
.....				
.....				

그림 4 한글음으로 분류한 유니코드 테이블

한글 코드가 SetCode()를 통해서 주어지면 코드를 index로 설정하여 이진 탐색 기법을 통하여 해당 한글이 존재하는 Table로 이동하여 table index를 얻어낸다. table index에 연결된 한자 코드들을 모두 얻어내어, 얻어낸 한자의 개수와 한자의 내용을 저장한다.

### ③ CKT Class

입력은 편집기에서 한자로 변환하고 싶은 글자를 먼저 한글음으로 입력한 후 변환하고자 하는 글자뒤에 커서를 위치시키고 F9 키를 누르면 한자 변환 대화상자가 위의 그림 5처럼 나온다. 한자 대화 상자에서 원하는 한자를 커서로 움직여 선택한 후 마우스로 클릭하면 한자 입력이 된다. 사용자 인터페이스는 다른 상용화되어 있는 한자 입력기와 비슷하나 한 윈도우에 많은 글자가 나타나도록 하였다. 이는 한글 한 자의 음에 해당하는 한자의 수가 많은 글자들에 대해 한자 변환 대화상자는 한 화면에 한자가 다 나타날 수 없으며, 그럴 경우 많은 화면 이동이 필요하다. 이러한 인터페이스의 문제점을 해결하여 효율적으로 한자를 입력하기 위해 한 화면에 100자씩 나타나도록 하였다. 다음으로 CTable에 저장된 한자정보를 대화상자를 통해서 사용자에게 보여주는 class이다. Visual C++에서 사용하는 MFC library에서 CDialogbox를 상속받아서 구현된다.

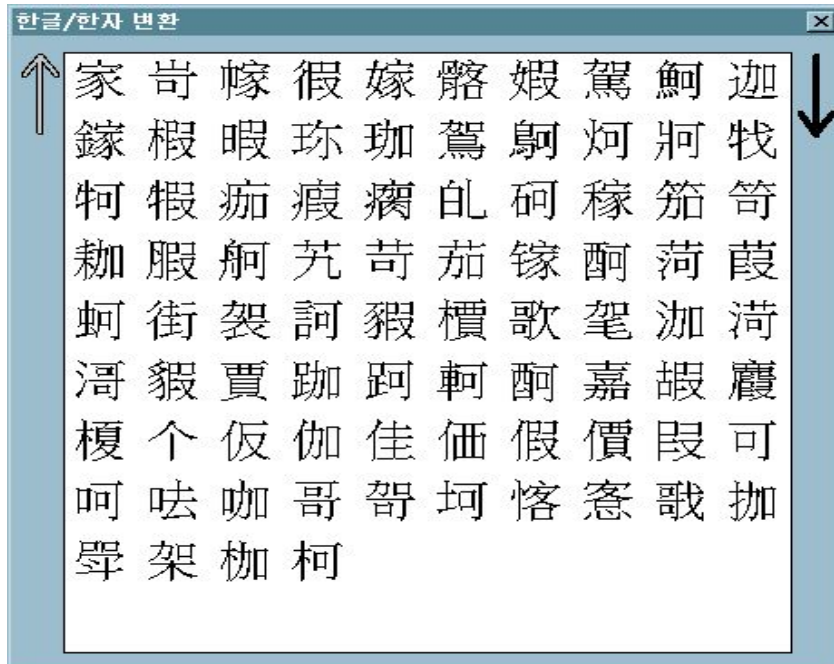


그림 5 한글음 '가'에 해당하는 한자 변환 대화상자

위의 [그림 5]는 한글 '가' 자에 대한 CKT 대화상자가 출력된 그림이다. 대화상자는 크게 두 부분으로 나뉘어져 있다. 한자가 현재 출력창에서 나타낼 수 있는 한자 개수보다 더 클 경우에 나머지 한자를 보여줄 수 있도록 이동시켜주는 화살표와, 한자를 표현하는 창으로 나타난다.

이 클래스는 XmlEdit에서 한글을 입력하고 현재 완성된 IME 조합상태에서 한글/한자 변환키를 누르면 현재 한글코드가 CKT Class로 전송되어 해당 한글코드에 대한 한자창을 출력하게된다.

다음은 사용자 인터페이스를 살펴보겠다. 위 한자 입력창은 키보드는 지원하지 않고, 마우스를 통한 입력만을 지원한다. 한 화면에 한자를 모두 출력하는 것을 기본으로 한다. 하지만, DB의 한자가 한 화면에 모두 보여주지 못할 경우에는 좌우에 있는 화살표를 통해서 다른 화면의 한자 화면으로 이동할 수 있도록 한다. 키보드에 의해서 현재 선택하고자 하는 한자의 위치를 변경한 다음에 Enter를 통해서 한자를 선택하는 방법은 Visual C++에서 제공하는 대화상자의 특성상 구현하기 복잡하다.

## 2) 유니코드 문서 편집기

현재 다른 분야에서 사용되고 있는 대부분의 문서 편집기는 KSC-5601의 코드를 사용하기 때문에 4,888자의 한자만을 입력할 수 있다. 보다 많은 수의 한자를 입력하기 위해서 본 연구에서는 유니코드를 사용하였고, 이런 이유로 유니코드를 화면에 출력하고, 문자의 입력, 수정, 삭제가 가능한 한국불교전서 전용 편집기의 개발이 요구되었다.

### ① 기본 구성

본 연구에서 개발한 유니코드 문서 편집기의 모습은 [그림 6]과 같다.

## 유니코드 한자처리를 위한 입력기와 편집기의 설계 및 구현

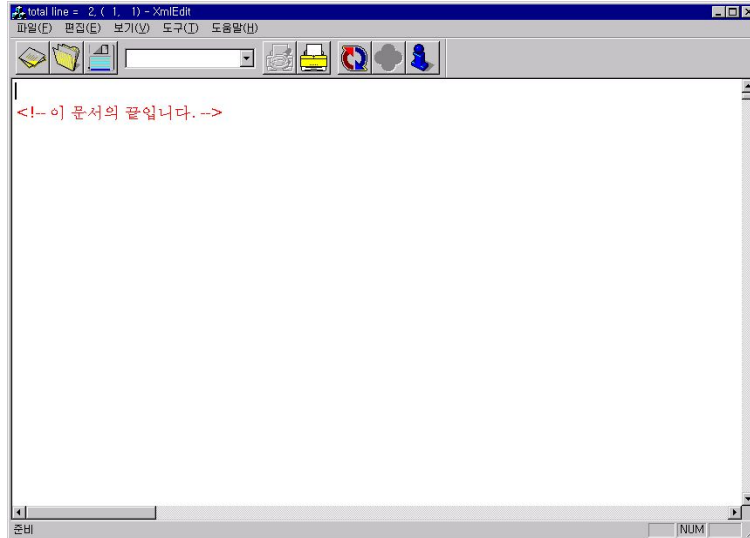


그림 6 유니코드 문서 편집기의 모습

본 유니코드 문서 편집기는 일반 윈도우즈 프로그램과 동일한 인터페이스를 갖는다. 즉, 제목표시줄의 최소화, 최대화(화면복귀), 종료버튼을 갖고 있으며, 메뉴, 도구단추, 이동 막대줄, 그리고 상태표시줄을 갖는다. 본 편집기에서 “<이 문서의 끝입니다>”라는 부분을 초과해서 문서를 입력할 수 없다.

### ② 문자 입력 방법

#### 가. 한/영 입력

한글과 영문의 입력은 일반적인 윈도우즈용 문서 편집기의 입력 방식과 동일하다. 즉, 키보드의 “한/영”-키를 이용하여 상태를 변경할 수 있다. 다음 [그림 7]은 본 편집기를 이용하여 한글과 영문자를 입력한 모습이다.

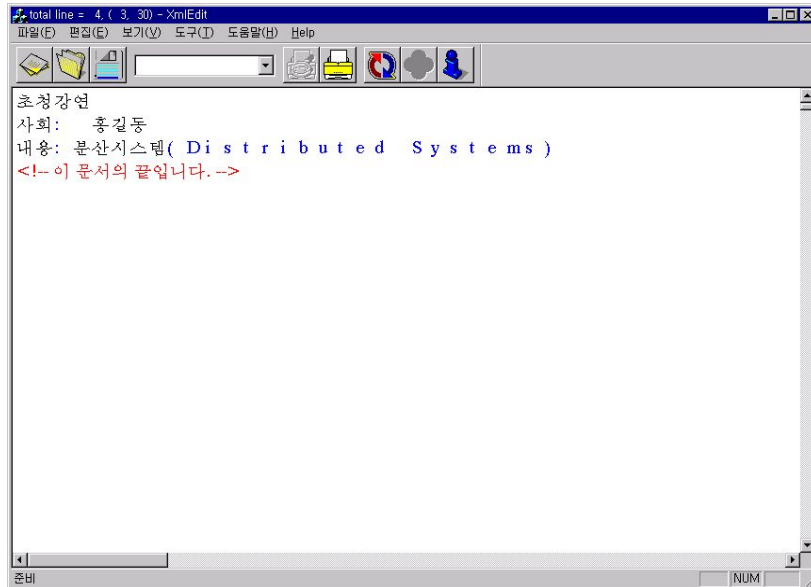


그림 7 한/영 문자의 입력 화면

#### 나. 한자의 입력

본 편집기에서 한자는 음전환방식으로 입력한다. 다음 [그림 8]은 음을 입력하고 그 음에 해당하는 한자를 화면에 출력한 모습을 나타낸다.

유니코드 한자처리를 위한 입력기와 편집기의 설계 및 구현

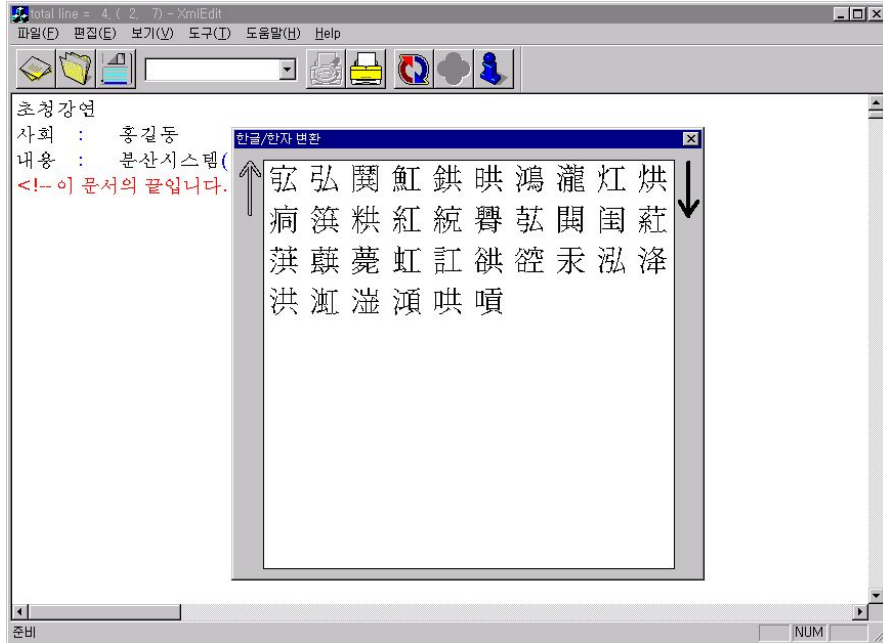


그림 8 한자의 입력 화면

다. 특수 기호의 입력

유니코드 특수 문자는 현재 충분한 자료를 갖고 있지는 않으나 원문자 (①②③...)와 “□”이 단축키로 등록되어 있다. 한국 불교 전서를 입력하기 위한 특수기호를 유니코드 표에서 찾아 등록시키면 확장시킬 수 있다.



Alt + 1 : □  
 Ctrl + 숫자 : 해당 숫자의 원문자

③ 편집기의 기능

본 연구에서 개발한 편집기는 윈도우즈 NT용으로 윈도우즈 기본 프로그램인 메모장 정도의 기능을 갖고 있다. 윈도우즈 NT의 메모장은 유니

코드 문서를 작성할 수 있으나, 화면에 출력되어 사용자에게 보여질 수 있는 문자는 KSC-5601 코드에 있는 것으로 한국 불교 전서를 입력하기에는 부적합하다.

#### 가. 문자의 삭제, 수정, 삽입

일반적인 편집기에서 사용되는 기능으로  키와  키를 이용한 문자의 삭제, 입력할 문자의 위치에 커서를 놓고 새로운 문자를 입력하는 삽입 기능이 있다. 문자의 삭제와 새로운 문자를 입력하는 수정 기능도 있다.

#### 나. 문자 찾기 기능

편집되고 있는 문서에서 사용자는 원하는 문자열을 찾을 수 있다. 다음 [그림 9]은 문자열 찾기를 위한 대화상자이다. 이 대화 상자는 “편집” 메뉴의 “찾기/바꾸기” 항목을 선택하면 된다.

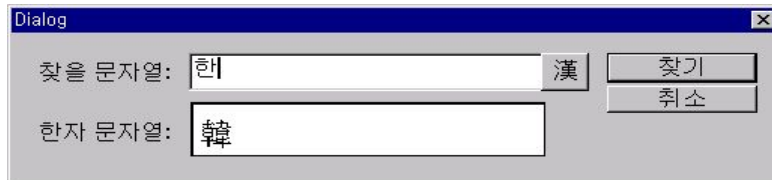


그림 9 찾기 대화상자

찾을 문자열에 사용자는 한글 또는 영자의 문자열을 입력한다. 만약 한자로 변환하려면 입력란의 오른쪽에 있는 “漢”버튼을 누르면 한자를 입력할 수 있는 대화상자가 화면에 나타나고 원하는 한자를 선택하며, 한자 문자열란에 표시된다.

모든 입력을 마치고 문자열을 착지 위해서는 “찾기” 버튼을 누르면 해당 문자열을 편집되는 문서에서 찾아 역상으로 표시해 준다.



## 다. 클립보드 이용 기능

많은 편집기들에서 일반적으로 클립보드를 이용하여 문자열의 이동과 복사를 처리한다. 클립보드를 이용한 복사는 “문자열 지정” → “클립보드로 복사” → “원하는 위치로 이동” → “클립보드에서 복사”의 과정을 거치게 된다.

- 문자열 지정

문자열의 지정은 해당 문자열을 블록으로 지정하는 것을 말한다. 블록으로 지정하는 방법은 일반 윈도우에서 사용하는 방법인 “Shift + 화살표키”를 사용한다. 현재 개발은 수평으로만 문자열을 지정할 수 있다. 또한 마우스를 이용하여 해당 문자열을 블록으로 지정할 수 있다.

- 클립보드로의 복사

블록으로 지정한 문자열을 클립보드로 복사하는 방법은 “편집” 메뉴의 “복사” 항목으로 선택하면 된다. 또는 단축키로 “Ctrl + C”를 눌러도 같은 기능을 수행할 수 있다.

- 클립보드에서 복사 :

클립보드로 복사된 문자열을 해당위치로 복사하는 방법은 “붙여넣기” 기능을 이용한다. 붙여넣기 기능은 “편집” 메뉴의 “붙여넣기” 항목을 선택하거나 단축키 “Ctrl+V”를 누르면 된다.

클립보드를 이용한 문자열 이동은 “문자열 지정” → “클립보드로 이동” → “원하는 위치로 이동” → “클립보드에서 복사”의 과정을 거치게

된다.

복사를 하는 방법과 다른 것은 원래 위치에 있던 문자열은 지워지고 새로운 위치에 문자열을 붙여넣는 것이 다르다. 즉, 원래 위치에서 클립보드로 이동시키는 방법만이 다르다.

- 클립보드로 문자열 이동

블록으로 지정된 문자열을 클립보드로 이동시키기 위해서는 “편집” 메뉴의 “잘라내기” 항목을 선택하거나 단축키 “Ctrl+X”를 누르면 된다.

#### 라. 문서 출력 기능

편집기를 이용해 작성된 문서는 “파일-인쇄” 명령을 통해 다음의 대화상자를 통해 프린터를 이용해 출력할 수 있다.



그림 10 인쇄 대화상자

출력되는 문서는 편집기를 통해 삽입된 태그의 입력된 상태와 태그가 생략된 형태를 선택할 수 있다. “파일-인쇄” 명령은 편집기에서 태그의 입력이 있는 형태를 출력할 수 있고, “파일-태그없는 문서출력” 명령을

이용하면 편집기에서 태그가 입력되지 않은 형태의 문자들만 출력할 수 있다.

#### ④ 파일관리 기능

본 편집기에서 가능한 파일 관리는 파일의 저장과 불러오기 기능이다. 편집기를 통해 입력된 문서는 “파일-저장” 명령을 통해 사용자가 입력한 파일이름으로 저장할 수 있으며, “파일-열기” 명령으로 파일로 되어 있는 문서의 내용을 편집기로 불러 올 수 있다.

## 4. 결론 및 연구과제

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 한국불교전서를 전산화하여 이를 연구하는 연구자들이나 열람을 원하는 사람들에게 도움이 될 뿐만 아니라 우리의 귀중한 문화유산을 전세계에 널리 알릴 수 있다.

한국불교전적을 전산화하기 위해 본 연구를 수행하여 한자를 컴퓨터에 입력할 수 있는 입력 방법 및 입력된 한자들을 편집할 수 있는 편집 시스템의 개발했다. 이러한 기술을 개발하기 위해서는 선행해서 해결되어야 할 문제가 한자 코드와 폰트이다. 본 연구에서는 전 세계에서 사용할 수 있는 유니코드가 지원하는 한자 코드가 대략 26,000자 정도이므로 유니코드와 유니코드 폰트를 사용하였다. 따라서 본 연구에서는 유니코드를 기반으로 한 음전환 방식의 한자 입력 시스템과 입력된 한자를 일반 문서 편집 기능들로 편집할 수 있으면서 구조적 문서 태그를 달 수 있는

편집기를 개발하였다.

본 연구를 통해 개발된 한자 입력 시스템과 문서편집기를 이용해 작성된 문서는 구조적 문서 태그가 포함되어 있어 쉽게 데이터베이스에 저장할 수 있고, 웹을 통해 전세계에 공개할 수 있다.

본 연구의 향후 과제로는 현재 유니코드 문서 편집기의 기능이 일반 편집기의 기능과 동일하지만 점차 기능을 보완하여 워드 프로세서로서의 기능을 수행토록 하는 것이다.

[참고문헌]

- [1] 혜묵스님, "세계의 불교자료 전산화 계획과 고려대장경 전산화를 위한 몇가지 문제들," [<http://members.iWorld.net/hederein/menu22/Hye2.html>], 1997.
- [2] 정주원, "ISO/IEC-10646 Universal Multiple-Octet Coded Character Set(UCS)에 대해서," [<http://simac.kaist.ac.kr/~jwjung/seminar/hangul-i18n/iso10646.html>], 1997.
- [3] The Unicode Consortium, "The Unicode Standard," Version 2.0, 1996.
- [4] ISO/IEC 10646-1:1993, "Information Technology - Universal Multiple-Octet Coded Character Set(UCS) - Part I: Architecture and Basic Multilingual Plane," 1993.
- [5] ISO 8879:1986, "Standard Generalized Markup Language," 2nd Edition, 1986.
- [6] 황기태 역, "어드밴스 윈도우 NT," 도서출판 대림, 1995.
- [7] 동국대학교 출판부, "한국불교전서 신라시대편," 1979.