

## 웹 기반하의 국어의 로마자 전사 표기 자동 변환 시스템

김 홍 섭\*

### Automatic translation system for hangul's romanization Based on the World Wide Web

Hong-sop Kim\*

#### 요 약

국어의 로마자 표기법 기본원칙을 모르더라도 웹 상에서 한글단어, 문장, 문서를 문자열(String)로 입력받아 국어의 로마자 표기 자동변환이 가능하게 설계 및 구현하였다. 특히, 표음 중심의 전사(Transcription) 표기법의 기계적 변환을 위해 알고리즘(algorithm)화된 음운법칙을 적용하여 소리글자로 자동 변환 후 정부 표준안인 국어 로마자 표기법 테이블에서 대응 글자를 참조하여 기계화 코드 변환이 가능토록 자동변환 알고리즘 설계하였으며, 유니코드(Unicode)와 아스키(ASCII)테이블에서 잘 쓰이지 않는 코드를 반달표(~)와 어긋점(') 등 특수부호로 할당하여 글꼴을 제작하였다. 또한 예외어 사전 관리를 통한 예외 처리 문제에 대한 해결 방안을 제시하였으며, UML 표기와 C++ 언어를 이용하여 사용자 편리성과 구현에 대한 모델을 제안하였다.

#### Abstract

After automatic translation system for hangul's romanization based on the World Wide Web converting korean-word, sentence, document to Transliteration letters by applying algorithm based phonological principles, even though a user do not know the basic principles of the usage of Korean-to-Romanization notations, It refers to corresponding character table that has been currently adopted the authority's standard proposition for Korean-to-Romanization notation rule concurrently, add to make possible to convert a machinized code as well. It provides font for toggling Korean-English mode, insert-edit mode by assigning ASCII codes and Unicode are hardly used to them.

This program could be made in C++ progamming language and Unified Modeling Language to implement various font, font-expanding and condensing, alternative printing.

---

\* 오산대학 정보관리과 조교수

## I. 서론

한글을 로마자로 표기해야 하는 국제 협약에 따라 1959년 2월에 '국어의 로마자 표기법'을 제정한 이후, 여러 차례 개정을 거쳐 문화관광부고시 제2000.8호(2000년 7월 7일)로 새 로마자 표기법 개정시안을 발표하고, 도로 표지판 및 문화재 안내판 등은 오는 2005년 12월 말까지, 변경을 계획하고 있다.[1]

즉, MR(Mccune\_Reischauer(1939))방식과 표음 중심의 전사(Transcription) 표기법(문교부:1984)안, 임의적인 혼합 방식(7), 관습적인 표기방식(8)을 혼용 사용을 인정함으로써 단일화 및 웹(Web) 상에서 자동변환과 웹 서비스를 지원 할 시스템 개발의 필요성이 증대되었다. 최근 ISO(International Standard Organization)에서는 남북한 단일안 및 인터넷 서비스가 가능하도록 기계적 변환을 요구하고 있으나 반달표(~)와 어깨점(') 표기 곤란 등의 사유로 난항을 거듭하고 있어 국가적으로 시급히 해결해야 할 과제로 제기 되었다. [12]

## II. 관련 연구

### 1. 국어의 로마자 표기

국어 로마자 표기법은 문자 중심의 전자법(轉字法)과 발음 중심의 전사법(轉寫法)으로 구분 할 수 있으며, 개정된 로마자 표기법은 국립 국어 연구원에서 제시한 어문규정의 표기의 기본 원칙과 표기 일람에 따라 기술한다.[6] 개정안에 따라, 도로 표지판 및 광고물, 문화재 안내판, 교과서 및 출판물 등에서 혼용 사용을 인정하고 있는 전사 표기법 처리, 즉 음운 법칙에 따른 소리글자로 1차 변환 후 로마자로 2차 변환해야 하나 표준(Qwerty) 키보드는 반달표와 어깨점 폰트를 제공하지 못하는 사유로 전자법에 관한 연구가 없었던 현실을 파악하고 웹 상

에서 전자법의 기계적 변환에 관한 문제점을 해결하였다. 문화관광부고시(제2000.8호)로 확정 발표한 새 로마자 표기법의 주요 내용은 다음과 같다.[1]

- 1) 특수부호 반달표(~)와 어깨점(')은 제외.
- 2) ㄱ, ㄷ, ㅂ, ㅅ은 위치에 상관없이 g, d, b, j로  
< 개정 전 말머리 k, t, p, ch 말중간 g, d, b, j >  
(예: 김포<Gimpo> 대구<Daegu> 부산 <Busan >)
- 3) ㅋ, ㅌ, ㅍ, ㅊ은 k, t, p, ch로  
< 개정 전 k', t', p', ch >
- 4) 모음 어, 으는 eo, eu로 <= 개정 전 ü, ö.>
- 5) ㅅ은 항상 s로 <= 개정 전 ㅅ과 모음>가 어울릴 때는 sh로 ( 예: <Shilla> 개정 후 => Silla )
- 6) 인명(人名), 회사명, 단체명은 관용(慣用)을 허용  
(예: 김<Kim> 김치<Kimchi> 삼성 <Samsung >)

## III. 시스템 분석

인터넷기반의 분산환경에서 사용자가 현행 정부 표준안인 국어의 로마자 표기법의 기본원칙을 모르더라도 웹 상에서 전자법과 전사법을 선택한 후, 전자법은 새 로마자 표기법 개정안을, 전사법은 음운법칙을 모르는 일반인 까지 사용 가능하도록 알고리즘화된 음운법칙을 적용하여 소리글자로 자동변환 후 MOE(문교부)안에 적용하여 로마자로 변환이 가능하다.[2]

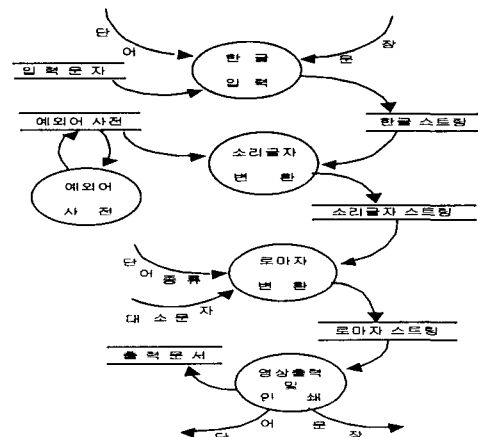


그림 2. 자료흐름도

(그림 1)은 웹 상에서 한글단어, 문장, 입력문서를 문자열(string)로 받아서 한글 스트링을 만들어내고 음운 법칙을 적용하여 소리글자로 변환하고 변환 과정에서 예외어를 참고하여 소리글자 스트링은 단어 종류를 참조하여 대·소문자 선택 여부를 참고하여 로마자 테이블에 저장되어있는 로마자 폰트와 대응되어 변환되고, 변환된 로마자는 사용자 요구에 따라 단어, 문장, 문서 단위로 영상, 프린터, 파일 출력하는 과정을 나타낸다. UML 규약에 따른 객체 흐름과 활동을 도해한 문서는 (그림 2)와 같다.

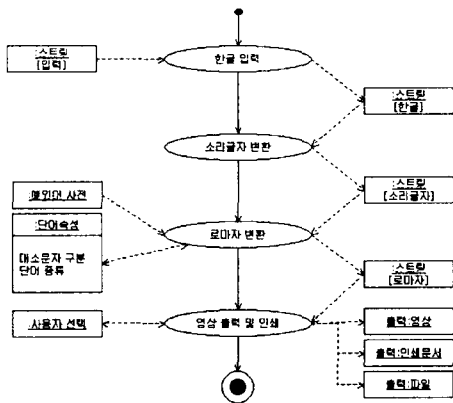


그림 2. 활동도(Activity Diagram)

#### IV. 시스템 설계

##### 1. 웹 상에서 국어 로마자 자동변환 모형

(그림 3)은 웹 상에서 국어의 로마자 표기 자동변환 처리과정과 예외어 사전 관리 Data Base모형이다.

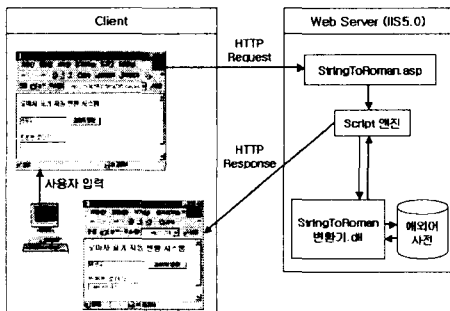


그림 3. 메시지의 전달과 원격 호출 과정

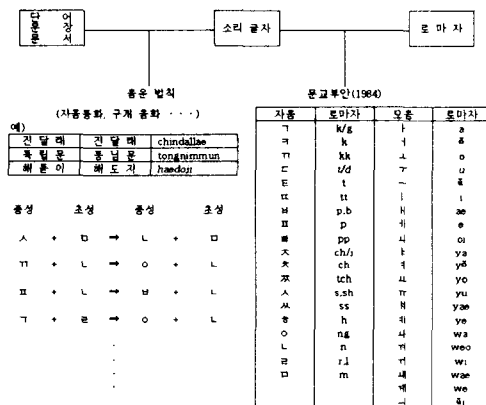
제안된 모델은 입력된 한글을 로마자 표기로 변환하기 위한 것으로 사용자는 웹 브라우저를 통하여 시스템에 접속한 후 변환하고자 하는 한글에 대한 로마자 표기 정보를 얻을 수 있다.

서비스를 제공하는 웹 서버의 경우 IIS5.0을 기반으로 사용자는 변환하고자 하는 한글 스트링을 매개변수로 서버쪽 스크립트인 StringToRoman.asp를 원격 호출한 후 서버쪽 스크립트 엔진이 StringToRoman.dll을 처리하여 매개 변수로 받은 한글 스트링을 로마자 표기로 변환하여 호출한 클라이언트에게 넘겨주게 된다.

StringToRoman.dll은 데이터베이스에 저장된 예외어 사전에서 예외어를 검색하고 예외어가 있을 경우 처리해 주게 된다. 예외어 사전을 위한 DataBase로는 SQL Server 7.0을 사용한다.

##### 2. 전사법 변환과정

국어 로마자 전사법 자동변환 과정은 (그림 4)에서 볼 수 있듯이 단어, 문장, 문서 등을 입력받아 한글문자열(string)화하고, 한 음절을 초, 중, 종성으로 분리하여 자음동화, 구개음화, 유음화 등 알고리즘(algorithm)화된 음운법칙을 적용시켜 소리글자 스트링으로 변환시킨 후 MOE(문교부)안 테이블에 대응시켜 로마자를 생성하



교적 우수하다. 그러므로 한글 음운체계의 정보처리로 실용성과 원어환원성의 효과를 높이기만 하면 전사법이 더 우수하게 될 수 있다.[2]

### 3. 특수부호 폰트 제작

〈표 1〉에서 볼 수 있듯이 ‘ㄱ, 一, ㅋ, ㆁ’는 잘 쓰이지 않는 ASCII 코드를 대치시켜 한글 라이브러리를 설계, 반달표 와 어깨점 폰트를 제작하여 사용하였다.

표 1. 반달표 와 어깨점 폰트 ASCII 대응

반달표 폰트	아스키 폰트	아스키 코드	참 고
0	[@]	64	폰트 에디터로 반달표 폰트를 작성하여 아스키 코드에 대치시켜 사용
U	[ ]	95	
o	[']	96	
h	[Ctrl-Backspace]	127	

### 4. 시스템 구성도

국어의 로마자 전사 표기법은 크게 도움말, 변환, 환경의 메뉴로 구성되며, 도움말에는 로마자 표기법의 개요 및 개정 경위 및 표기 기본원칙과 기능기 및 운영방법이 설명되고, 종료 메뉴로 구성하였다. 변환기능은 단어, 문장, 문서 변환으로 분류하였고, 단어 변환을 선택하면 단어종류 및 대소문자를 선택 소리글자로 변환 후 고유명사, 일반단어, 인명, 행정구역의 단어 변환이 가능하다. 문장변환을 선택하면 주소, 일반문장 등을 소리글자 및 로마자로 변환할 수 있고 예외어 사전에 등록되어 있는 단어가 발견될 경우 적용 여부 선택이 가능하다. 이외도 문서변환이 가능하며, 환경 메뉴에는 예외사전에 추가등록, 삭제가 가능하도록 설계하였다.

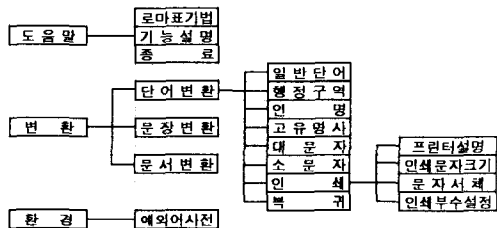


그림 5. 시스템 구성도

## V. 시스템 구현

### 1. 전사표기법 메인 화면 구성

국어의 로마자 전사표기법자동변환의 초기화면은 〈그림 6〉와 같으며 변환 메뉴에서 변환이 가능하다.

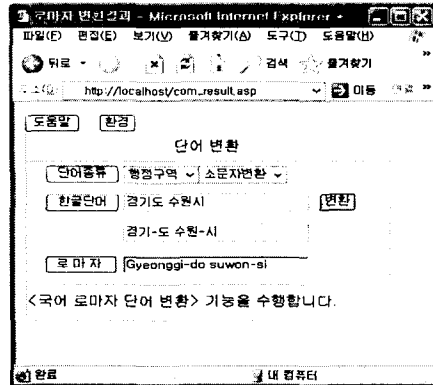


그림 6. 로마자 전사표기법 메인 화면

#### 1.1. 단어변환

초기화면에서 「단어변환」을 선택한 후 사용자가 요청한 「독립문」이라는 단어는 음운법적 적용을 거쳐 「동님문」이라는 소리글자로 변환된 후 다시 MOE테이블에 적용하여 「Tongnimmun」이라는 로마자로 변환된다. 단어 변환의 종류에는 일반단어, 인명, 고유명사, 행정구역이 있으며, 대·소문자 변환 선택과 변환된 단어를 인쇄할 경우 한글 및 로마자에 대한 글자체와 크기를 설정한 후 프린터로 인쇄 할 수 있다. 예외어 일 경우에는 예외어 사전에서 자동으로 검색하여 예외어 적용 여부를 선택 할 수 있다.

#### 1.2. 문장변환

「문장변환」을 선택하여 주소와 같은 문장 입력 후 변환 시 「서울과 같이 예외적으로 변환되는 단어가 발견 될 경우 규칙변환과 예외변환 중 사용자가 선택 사용할 수 있다. 예외어는 예외어 사전에서 관리(추가등록, 삭제)할 수 있으며 「자동변환」을 선택할 경우 발견되는 모든 예외어에 대해서 예외변환을 적용하고 「변환금지」는 예외어 변환을 적용하지 않는다.

#### 1.3. 문서변환

「문서변환」을 선택하면 문서변환 및 인쇄기능을 수행한다. 먼저 「입력문서」란에 한글 문서명을 입력하고 「출력문서」란에 로마자로 변환되어 출력될 파일명을 입력

한 후 '변환' 기능을 수행하면 한글문서 전체가 로마자문서로 변환되며 문서의 내용에 예외어가 포함되어 있을 경우 예외어 처리 화면이 표시된다. 변환된 문서는 예에서 보인 것 같이 혼합. 이단인쇄로 출력이 가능하며 이외도 영상 출력을 위한 기능이 있다.

## 2. 변환 예

### 2.1. 단어변환 인쇄

독립문                      종로구  
Tongnimmun              Chônglogu

### 2.2. 문장변환 인쇄

아름다운 한글              사무자동화학회  
Arūmdaun han-gŭl        Samujadonghwahak'oe

### 2.3. 문서변환 혼합 인쇄

말은 사람과 사람의 뜻을 통하는 것이라.  
Marŭn saramwa saramŭ ttŭsŭl t'onghanŭn kōshira.

### 2.4. 문서변환 이단 인쇄

글은 말을 담은 그릇이니, 이치리짐이 없고 자 : Kŕŕŭn marŭl tamŭnŭn kŕŕŕŕŭshŭn. iŭrŕŭnŭrŭ dŕk;  
리를 반듯하게 깎아 글게 선 뒤에야 그 말을 : o chŕnŭl oandŕŕ'asŭ chaba. kutka sŕn twŕy;  
잡지키나니라. 글은 또한 : a kŕ marŭl chŕl chŭk'ŭsanŭra. kŕrŭn ttohan;

## 3. 예외어 처리

「환경」 메뉴에서 '예외어 사전'을 선택하면 <그림 7>와 같은 화면이 표시되며 [이전] 또는 [다음] 버튼을 선택하여 예외어를 검색할 수 있다. 예외어를 추가로 등록할 경우 (추가)버튼을 <그림8>와 같은 화면이 표시되고 예외적으로 표기되는 로마자를 입력하면 새로운 예외어의 추가 등록이 완료된다.

이외도 예외어 삭제와 수정이 가능하다.

예외어 사전은 규칙 변환으로 처리되지 않는 단어를 모아 국립국어 연구소의 자문을 거쳐 등록해 놓은 자료이다.

예) 한글 가 랑 외

- 규칙 변환 Karang-ip(틀림)
- 예외 변환 Karangnip(맞음)

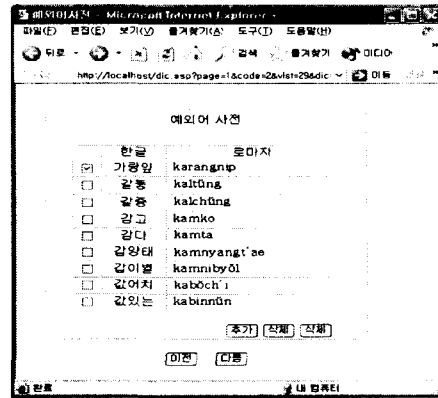


그림 7. 예외어 사전 화면

전자표기법은 국립국어 연구소의 웹사이트를 참고한다.[1]

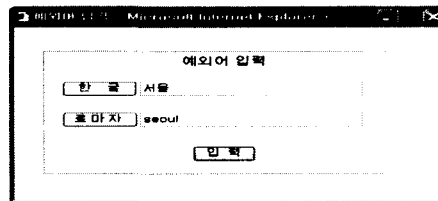


그림 8. 예외어 입력 화면

## 4. 모바일 사용자를 위한 서비스

이동 통신망 및 모바일 단말기 제조 기술의 발달로 무선 랜 및 CDMA2000-1X 서비스 등을 이용한 인터넷 접속이 크게 증가하고 있다.[10] 차세대 인터넷 접속 매체인 PDA나 휴대폰, 스마트폰 등 모바일 단말기가 보편화 되면서 모바일 사용자를 위한 무선 웹 서비스의 구성은 다음과 같다

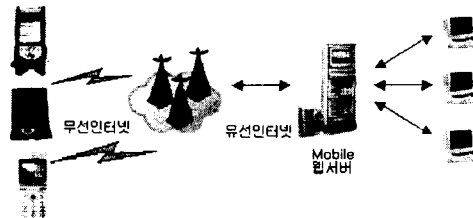


그림 9. 무선인터넷과 무선 데이터 서비스

## VI. 결론 및 향후 연구

인터넷 사용자가 비교적 손쉽게 활용이 가능한 국어의 로마자 표기 자동 변환 시스템을 제안하였다. 국어의 로마자 전사표기법은 음운법칙에 따른 소리글자로 1차 변환 후 로마자로 2차 변환이 요구되며, 표준 키보드는 반달표와 어깨점 폰트를 제공하지 않아 전사표기법 기계적 변환은 어려울 것이라고 여겨져 왔던 문제점들을 해결하므로써 임의 혼합표기 방지 및 단일화 가능성을 제시하고 예외어의 증가 및 소멸에 대비하여 웹 상에서 예외어 사전 관리 기법을 제안하였다.

국어의 로마자 전자표기법 자동변환 시스템의 설계로 <표 2>와 같은 효과를 기대 할 수 있다.[2]

표 2. 국어의 로마자 자동변환 시스템

표기법 특징	원음 표음력	원어 환원성	실 용 성
효과	상	중	상

이밖에도 행정기관의 공문서나 지하철공사, 철도청, 도로공사의 도로표지판, 신용카드 및 여권 신청 시 인명표기, 교과서, 무역송장서류, 지명, URL과 같은 로마자 표기시 시간 절약 및 예산 절감 효과가 기대된다.

향후 연구과제로는 음성 및 필체인식, XML기반하의 무선 콘텐츠 분야에서의 유·무선 데이터 서비스 활용 방안 확대와 국제적으로 관례화 되어있는 기업체명과 상품명, 인명, 지명 등의 예외어 사전 관리로 처리기능을 확대 적용과 출생 신고 시 인명표기를 의무화한다면 인명 표기에 따른 정보검색 논란을 사전에 방지를 할 수 있는 점을 제안해 본다. 특히 이정표 표기는 외국인이 그 의미를 쉽게 파악할 수 있도록 로마자와 영문단어를 혼합사용 방안 검토가 요망되며, 로마자 표기법을 개정할 때는 정부 관계부처의 실무자들과 외국어, 음성학 전문가들의 충분한 사전 토의를 거친 후에 시행해도 늦지 않다는 점을 인식하고, 지도제작, 백과사전, 항공사, 항만, 도로 표지판 제작자들의 의견을 충분히 수렴하여 참고하여야겠다. 이외도 실용성이 고려된 단일 안이 확정되어 홍보 및 체계적 교육이 병행되어야 하겠다.

## 참고문헌

- [1] <http://www.korean.go.kr/search/roma/roman.html>
- [2] 김홍섭외, 한글 로마자 자동 표기 시스템 설계 및 구현, 제 5회 한글 및 한국어 정보처리 학술대회, 1993.
- [3] Grady Booch, James Rumbaugh, Ivar Jacobson, The Unified Modeling Language User Guide, Addison-Wesely Publishing Co, 1999.
- [4] 정태충, 강혜숙, 음절위주의 한글 로마자 표기형식, 경희대학교, 1998.
- [5] 송재용외, 로마자표기 한글 인명을 위한 검색 모듈 설계와 인명 질의 확장기 구현, 정보과학회 봄 학술발표논문집, Vol. 25, No 1, pp.196, 1998.
- [6] 김세중, 국어의 로마자 표기법 개정 시안, 국립국어 연구원, 1997.
- [7] 도형수, 한글의 로마자 표기에 관한 연구, 효성여대, 1992.
- [8] 송기중외, 기계화를 위한 한글의 로마자 표기법, 정신문화원, 1986.
- [9] 이기문의, 국어 어문 규정집, 교육부, 1992.
- [10] 전영호, 황인준, 모바일 사용자를 위한 서비스 중심의 개인화 시스템, 한국정보과학회 가을 학술 발표논문집, Vol. 29, No 3, pp367, 2002.
- [11] 최규정의, 한글 고유 명사의 로마자 표기 변환에 관한 연구, 한국정보과학회 추계학술 발표논문집, Vol. 27, No 2, pp194, 2000.
- [12] 박수호외, 한국어 음운 현상과 단어 유형을 고려한 한글-로마자 변환 시스템 구현, 한국정보과학회 추계학술 발표논문집, Vol. 28, No 1, pp376, 2001.

### 저 자 소개



김 홍 섭

1984년 2월 동국대학교 학사  
(전자계산)

1994년 2월 동국대학교 석사  
(전산교육)

2001년 7월 동국대학교 박사과  
정 수료(컴퓨터공학)

2002년 12월 오산대학 정보관  
리과 조교수(현)