

국어 로마자 전사표기 웹 컴포넌트 설계 및 구현

김 홍 섭 *

The design and implementation of Web Component for Korean to Roman transcription

Hongsop Kim*

요 약

개정된 국어 로마자 표기법의 기계적 변환을 위하여 웹 기반하의 컴포넌트를 설계하고 구현하였다. 특히 음운법칙을 적용하여 소리글자로 변환 후 로마자 표기로 자동 변환하는 과정과 전사법 변환에 관한 알고리즘을 제시한다. 컴포넌트는 웹 서버의 스크립트 메커니즘에 의해 작동하도록 하였고, 예외어 사전 등은 웹 서버 내부 혹은 외부에서 지원하는 보조 기능으로 설계하였으며, 전체적인 시스템 구조는 UML 명세서와 의사 코드를 이용하여 작성하였다. 제안된 구조는 웹 컴포넌트 형태의 캡슐화 서비스를 제안함으로써 소프트웨어 개발의 경제성, 라이브러리 재사용 및 확장성 등을 향상시킬 수 있을 것으로 기대된다.

Abstract

In this paper, a web-based automatic transcription component is designed and implemented for mechanical conversion of revised Korean-to-Romanization notation rule. Specially, we proposes system architecture and algorithms that transcript Korean to Roman automatically after transliterate Korean to phonetic symbol applying phonological principles. The components operate under the web server's script mechanism, and the dictionary for exceptional usage is designed as an accessorial function supported either operating at web server internally or externally. The overall system architecture is presented by UML specification and pseudo code. The proposed architecture can be implemented in encapsulated service by object oriented component and that can be easily adapted and modified on the internet environment and this system may have many advantages to improve efficiency, library reuse, extensibility at software development.

▶ Keyword : 로마자(romanization), 한글로마자(Korean-to-Roman), 전사법(transcription)

• 제1저자 : 김홍섭

• 접수일 : 2004.10.20, 심사완료일 : 2004.11.16

이 논문은 2003년도 오산대학 학술비지원 연구에 의하여 연구되었음.

I. 서론

한글에 관한 정보는 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 로마자 표기법 개정시안을 발표하고, 도로 표지판, 광고물 및 문화재 등의 안내판은 2005년 12월말까지 변경을 추진하고 있다.[1]

또한 ISO(International Standard Organization)에서는 남북한 단일안 및 기계적 변환을 요구하고 있어 웹 서비스의 필요성이 제기되었다.[4]

본 연구는 국어 로마자 전사표기 자동 변환을 위해 웹 컴포넌트 설계 및 구현에 관한 참조모델을 제시하였으며 구성은 다음과 같다. 2장은 관련 연구이며 3장은 로마자 전사표기 변환과정과 예를 제시하고 4장에서는 웹(Web) 상에서 국어 로마자 전사표기 웹 컴포넌트 구성요소 설계와 의사 코드, 예외어 처리, 특수부호 폰트 제작 과정을 제안하여 재사용성 및 확장성을 고려하였다. 5장에서는 결론 및 향후 연구 과제를 기술하였다.

II. 관련 연구

2.1 국어의 로마자 표기

국어의 로마자 표기는 MR(Mccune_Reischauer:1939)표기법과 표음 중심의 전사(Transcription) 표기법(문교부:1984), 전자법(남·북한안:1986), 임의적인 혼합과 관습적인 표기방식을 혼용하여 사용함으로써 단일화 필요성이 제기되어왔다.[5] 로마자 표기 방식은 문자 중심의 전자법(轉字法)과 표음 중심의 전사법(轉寫法)으로 구분 할 수 있으며, 개정된 로마자 표기법은 MR 표기법을 근간으로 표준 발음법에 따라 적는 것을 원칙으로 하며, 국립 국어 연구원에서 제시한 어문 규정의 표기의 기본 원칙과 표기 일람에 따라 기술한다.[2][6][9]

그 내용 중의 하나가 컴퓨터에서 입력 검색이 불편했던 특수 부호인 반달표와 어갯점 표기를 제외시키며(로마자 이외의 부호는 되도록 사용하지 않는다.) ㄱ, ㄴ, ㄷ, ㄹ은 모음 앞에서는 g, d, b, r로, 자음 앞이나 어말 앞에서는 k, t, p, l로 적는다. 모음도 '어, 으'를 'ö, ü'로 적던 것을 'eo, eu'로 적는다. 개정안에 따라 웹 기반하의 자동 변환이 요구되며, 혼용 사용을 인정하고 있는 전사 표기법 처리, 즉 음운 법칙에 따른 소리글자로 1차 변환 후 로마자로 2차 변환이 요구되었으나 표준(Qwerty) 키보드는 반달표(˘)와 어갯점(˙) 폰트를 제공하지 못하는 사유로 전자법에 관한 연구가 없었던 현실을 파악하고 웹 상에서 전자법의 기계적 변환에 관한 문제점을 해결하였다.

III. 로마자 자동 변환 제안

3.1 로마자 변환 개념

국어의 로마자 표기법의 기본원칙을 모르더라도 인터넷 사용자가 웹 상에서 로마자 표기법 개정안(2000)과 전자법(Transcription)을 선택한 후, 로마자 변환을 처리하며, 전자법은 음운법칙을 적용하여 소리글자로 자동 변환 후 문교부안(1984)에 적용하여 로마자로 변환이 가능하다. [7][8]

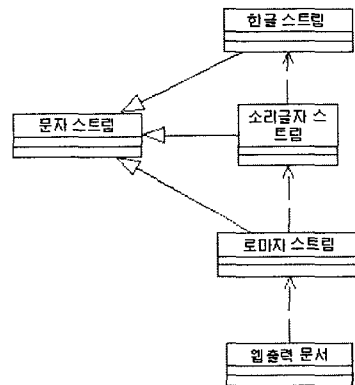
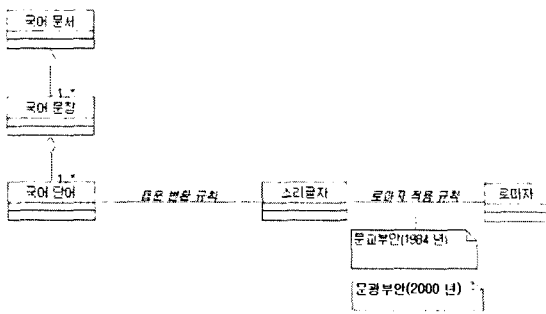


그림 1. 개념도
Fig. 1 Conceptual diagram

(그림 1)은 웹 상에서 한글단어, 문장, 문서파일을 문자열(string)로 받아서 한글 스트림을 만들어내고 음운법칙을 모르는 일반인까지 사용 가능하도록 음운 법칙을 적용하여 소리글자로 변환하고 변환 테이블에서 로마자 대응 글자를 참

조하여 소리글자 스트링으로 기계적 변환이 가능하도록 하였으며 변환 과정에서 예외어 사전에 참고하여 적용하고, 단어의 종류를 참조하여 대·소문자 선택 여부를 결정하고, 변환된 로마자는 사용자 요구에 따라 단어, 문장, 문서 단위로 로마자 변환과 보기, 파일을 출력하는 과정을 나타낸다. 국어 로마자 자동변환 적용 시나리오 UML 규약에 따른 객체 흐름과 활동을 도해한 명세서는 (그림 2)와 같다.



음운법칙
(자음동화, 구개음화 ...)
예)

진달래	Jindallae
독립문	Dongnimmun
해돋이	Haedoji

종성 초성	→	종성 초성
ㅅ + ㄷ	→	ㄴ + ㄷ
ㄱ + ㄴ	→	ㅇ + ㄴ
ㄲ + ㄴ	→	ㄱ + ㄴ
ㄱ + ㄹ	→	ㅇ + ㄹ
.		.
.		.

자음	로마자	모음	로마자
ㄱ	g, k	ㅏ	a
ㄲ	kk	ㅓ	eo
ㅋ	k	ㅗ	o
ㄷ	d, t	ㅜ	u
ㅌ	tt	ㅡ	eu
ㄷ	t	ㅣ	i
ㅂ	b, p	ㅐ	ae
ㅃ	pp	ㅔ	e
ㅍ	p	ㅚ	oe
ㅈ	j	ㅟ	wi
ㅉ	jj	ㅙ	ya
ㅊ	ch	ㅞ	yeo
ㅅ	s	ㅟ	yo
ㅆ	ss	ㅠ	yu
ㅎ	h	ㅙ	yae
ㄴ	n	ㅟ	ye
ㄹ	m	ㅘ	wa
ㅇ	ng	ㅚ	wae
ㄹ	r, l	ㅜ	wo
		ㅟ	we
		ㅡ	ui

그림 2. 국어 로마자 변환 과정
Fig. 2 Korean to romanization process

3.2 로마자 전사표기 변환과정

(그림 3)는 한글단어, 문장, 문서파일 등을 입력받아 한글 문자열화하고, 한 음절을 초, 중, 종성으로 분리하여 자음동화, 구개음화, 유음화 등 음운법칙을 적용하여 소리글자 스트링으로 변환시킨 후 문장부안 테이블에 대응시켜 로마자를 생성하게 되며 단어 및 문장 그리고 문서 변환, 문서 파일 저

장 등을 선택하게 된다. 반달표(˘)와 어긋점(˙)은 아스키(ASCII)테이블에서 잘 쓰이지 않는 코드를 특수 부호형태의 글꼴로 할당하여 대치성을 고려하였다.

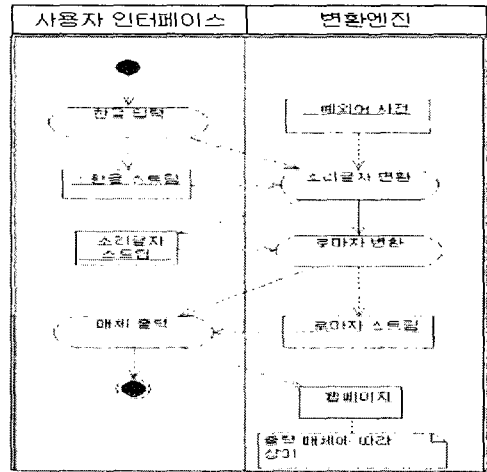


그림 3. 활동도
Fig. 3 Activity diagram

3.2.1 단어변환

단어의 유형에는 일반단어, 인명, 고유명사, 행정구역과, 대·소문자 선택 후 소리글자 및 로마자로 변환할 수 있고 예외어 사전에 등록되어 있는 단어가 발견될 경우 적용 여부 선택과 추가등록, 삭제가 가능하도록 설계하였다.

3.2.2 문장변환

「문장변환」에서 주소, 일반문장 등을 선택 한 후 문장 변환 시 '서울'과 같이 예외적으로 변환되는 단어가 발견될 경우 규칙변환과 예외변환 중 사용자가 선택 사용 할 수 있다. 예외어는 예약어 사전에서 관리(추가, 등록, 삭제)할 수 있으며 「자동변환」을 선택할 경우 발견되는 모든 예외어에 대해서 예외변환을 적용하고 '변환금지'는 예외어 변환을 적용하지 않는다.

3.2.3 문서변환

「문서변환」은 '입력문서'란에 한글 문서 명을 입력하고 '출력문서'란에 로마자로 변환되어 출력될 파일명을 입력한 후 '변환' 기능을 수행하면 로마자문서로 변환되며 문서의 내용에 예외어가 포함되어 있을 경우 예외어 처리 화면이 표시된다. 예외어는 예외사전에서 자동 검색하여 예외어 적용 여부를 선택 할 수 있으며, 변환된 문서는 예에서 보인 것 같이 출력이 가능하다.

3.2.4 개정안 로마자 및 전사표기 변환 예

단어변환	
경기도 수원시	서초구
Gyeonggi-do Suwon-si	Seocho-gu
Kyönggi-do Suwon-shi	SöChö-gu
문장변환	
아름다운 한글	컴퓨터정보학회
Areumdaun Han-geul	Computer Jeongbohakoe
Arümdaun han-gül	ComputerChöngbohak'oe
웹 기반하의 한글 로마자 자동변환	
Web gibanhaui Han-geul Romaja Jadongbyeonhwan	
Web kibanhaui han-gül romaja chadongbyönhwan	
문서변환 혼합	
말은 사람과 사람의 뜻을 통하는 것이다.	
Mareun Saramgwa Saramui Tteuseul Tonghaneun Geosira	
Marün saramgwa saramüi ttüsül t'onghanün köshira.	

표 1. 개정안 로마자 예외어 사전
Table. 1 Dictionary for exceptional usage

한글	로마자	한글	로마자
가랑잎	Garangnip	감다	Gamtta
갈등	Galtteung	같이별	Gamnibyeol
갈증	Galjjeung	값어치	Gabeochi
감고	Gamkko	값있는	Gabinneun

IV. 시스템 설계

4.1 로마자 전사표기 웹 컴포넌트 구성

제안된 모델은 입력된 한글을 로마자 표기로 변환하기 위한 것으로 사용자는 웹 브라우저를 통하여 시스템에 접속한 후 변환하고자 하는 한글에 대한 로마자 표기 범위를 확인하고 필요한 부분을 선택하여 데이터를 수집할 수 있다.

(그림 4)는 웹 상에서 로마자 전사표기 처리과정으로 사용자는 변환하고자 하는 한글 스트링을 매개변수로 서버쪽 스크립트인 StringToRoman.asp를 원격 호출한 후 서버쪽 스크립트 엔진이 StringToRoman.dll을 처리하여 매개 변수로 받

은 한글 스트링을 로마자 표기로 변환하여 호출한 클라이언트에게 넘겨주게 된다. 즉, 웹서버의 경우 반응시간 내 웹브라우저에게 독립적인 컴포넌트를 선택하여 트랜잭션을 수행하고 캡슐화 서비스 제공으로 데이터의 일관성을 보장할 수 있다.

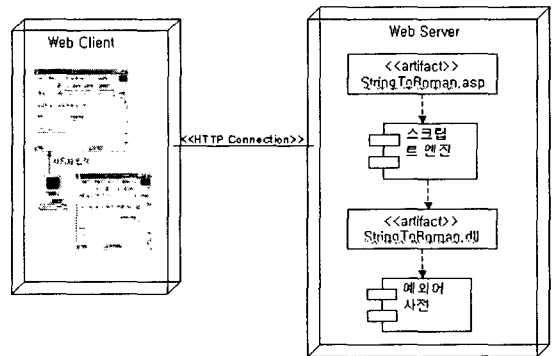


그림 4. 웹 컴포넌트 구성도
Fig. 4 Web component diagram

주요 컴포넌트들은 웹서버를 컨테이너로 한다는 전제하에 설계되었으며, 스크립트 엔진은 웹서버 환경에서 공급되는 일반적인 실행환경으로 가정하였다.

4.2 로마자 전사표기 시나리오

각 컴포넌트들의 상호작용은 다음 그림과 같이 HTTP 프로토콜을 근간으로 호출시 상태의 변화가 일어난다고 가정하였고, 위임(Delegation)과 전달(Communication)의 관계는 (그림 5)와 같다.

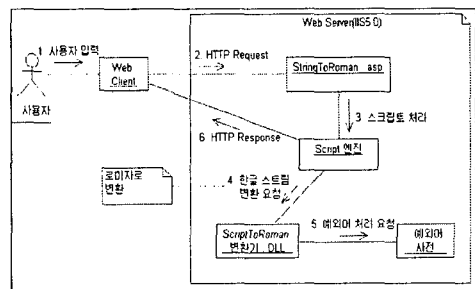


그림 5. 협동 다이어그램
Fig. 5 Collaboration diagram

웹서버 내부 컴포넌트들의 상호 작용은, 일반적인 웹 서버의 스크립트 처리 매커니즘에 따라 정의하였다.

웹 서버를 컨테이너로 정의하면, 내부 컴포넌트들의 상호 작용은 웹 서버와 함께 항상 표준이 제공될 수 있으므로, 한글 변환 관련 컴포넌트들은 항상 표준화하여 개발될 수 있다. StringToRoman.dll은 데이터베이스에 저장된 예외어 사전에서 예외어를 검색하고 예외어가 있을 경우 처리해 주게 된다. 예외어 사전 관리는 SQL Server를 이용하여 재사용 및 확장성을 고려하여 하였다.

4.3 국어 로마자 변환 컴포넌트 설계

관련 컴포넌트들에 필요한 사용 사례는 한글 소리글자 변환, 한글 로마자 변환, 초성 분리, 종성음 검사, 글자 변환 기능 등으로 파악되었다.

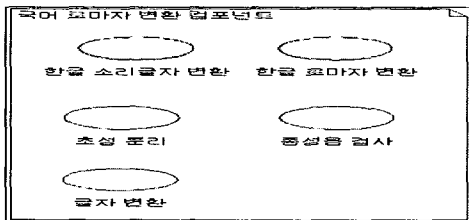


그림 6. 사용 사례 다이어그램
Fig. 6 Use case diagram

4.4 의사 코드(Pseudo Code)

시스템 구조 중 주요한 한글 소리글자 변환, 한글 로마자 변환에 대한 의사 코드는 다음과 같다.

```
char *HangulRespell(char *src, char *des, word kind) {
    // 한글을 소리글자로 변환(예: 독립문 -> 동남문)
    // src:원본, des: 변환될 버퍼
    // kind: 단어의 종류(일반단어, 행정구역, 고유명사)
    while(src != null) {
        if (kind==예외어) 예외어 사전 적용 처리
        else if (영문글자인 경우) 영문글자 처리
        else if (행정구역인 경우) 행정구역 처리
            // 행정구역 접미사(시, 군, 구 등...) 비교 및 적용
        else if (이름인 경우) 이름 처리
            // 성(김, 이, 박, 독고, 남궁, 선우...) 비교 및 적용
        else if (영숫자인 경우) 영숫자 처리
        else { // 한글 일반 단어인 경우
            구개음화 처리
            자음동화(축약) 처리
            유음화 처리
            ...
        }
    }
    return des // 변환된 버퍼 반환
    //des : 한글을 소리글자로 변환된 스트림
}
```

```
char *han2roman(char *src, char *des, word kind, word end) {
    //변환 방법에 따라서 변환 테이블 적용(전사법, 개장안)
    HangulRespell(src,buffer,kind) //한글을 소리글자로 변환
    // 소리글자를 로마자로 변환(예: 동남문 -> Dongnimmun)
    while(문장(buffer)이 끝날 때까지 한 글자씩 읽어) {
        //중간에 '-' 넣기(예: 안국 Anguk(X), An-guk(O))
        //초성과 종성이 모두 'ㄹ'인 경우 r -> l (예:콜라)
        if (영문글자인 경우) 영문글자 변환
    else if (행정구역인 경우) 행정구역 변환
        else if (이름인 경우) 이름 변환
    else if (영숫자인 경우) 영숫자 변환
    else { // 한글 일반 단어인 경우
        무성음 처리
        유성음 처리
        복중성 처리
        ...
    }
}
return des
// des: 로마자로 변환된 스트림
}
```

4.5 전사표기 특수 부호 폰트 제작

〈표 2〉에서 볼 수 있듯이 ‘ㄱ, 一, ㅋ, ㆁ’는 잘 쓰이지 않는 ASCII 코드를 대치시켜 반달표와 어깨점 폰트를 제작하여 대체 지원하였다.[3]

표 2. 특수 부호 폰트
Table. 2 Special symbol font

반달점 폰트	아스키 폰트	아스키 코드	참 고
0	{@}	64	폰트 에디터로 반달점 폰트만 작성하여 아스키 코드에 대치시켜 사용
U	{ }	95	
o	{' }	96	
ü	{Ctrl-Backspace}	127	

4.6 사용자 웹 인터페이스

로마자 전사표기 웹 인터페이스는 (그림 7)과 같으며, 전사표기 예외어 사전은 (그림 8)과 같다.

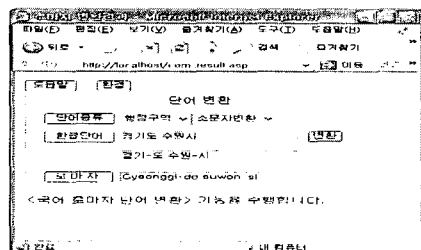


그림 7. 전사표기 웹 인터페이스
fig. 7. Transcription web interface

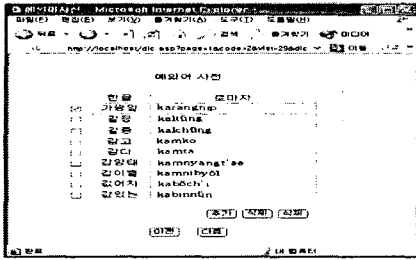


그림 8. 전사표기 예외어 사전
fig. 8. Transcription dictionary for exceptional usage

V. 결론 및 향후 연구

국어 로마자 표기법의 기계적 변환을 위하여 음운법칙을 적용하여 소리글자로 변환 후 개정안 로마자 표기로 자동 변환하는 웹 기반하의 컴포넌트를 설계 및 구현하였다. 또한 표음 중심의 전사표기법 변환에 관한 반달표와 어긋점 폰트를 지원하여 전사표기법 기계적 변환 문제점들을 해결하고 예외어 사전 관리 기법을 제시하였다.

전체적인 시스템 구조는 객체 지향 방법론의 UML 명세서와 의사 코드(Pseudo Code)를 이용하여 제시하였으며, 컴포넌트는 웹 서버의 스크립트 메커니즘에 의해 작동하도록 하였고, 예외어 사전 등은 웹 서버 내부 혹은 외부에서 지원하는 보조 기능으로 설계하였다.

본 논문에서 제안된 구조는 인터넷 환경에서 쉽게 사용하고(plugin-in) 변경할 수 있는 웹 컴포넌트 형태의 캡슐화 서비스를 제안함으로써 소프트웨어 개발의 경제성, 라이브러리 재사용 및 확장성 등을 향상시킬 수 있을 것으로 기대된다.

향후 연구과제로는 XML기반의 플랫폼과 유니코드(Unicode) 적용과 유무선 콘텐츠 분야에서의 웹 서비스 활용 방안 확대와 동시에 관련 컴포넌트들 사이의 상호 작용 표준을 정의하는 기준을 개발할 필요가 있다.

참고문헌

- [1] <http://www.korean.go.kr/search/roma/roman.html>
- [2] <http://www.korean.go.kr/rule/rule10.htm>
- [3] 김홍섭, 웹 기반하의 국어의 로마자 표기 자동변환 시스템 설계, 한국OA학회 논문집, 제6권 제4호, 2001.
- [4] 송기중외, 기계화를 위한 한글의 로마자 표기법, 정신문화원, 1986.
- [5] 도형수, 한글의 로마자 표기에 관한 연구, 효성여대, 1992.
- [6] 김세중, 국어의 로마자 표기법 개정 시안, 국립국어연구원, 1997.
- [7] 정태충, 강혜숙, 음절위주의 한글 로마자 표기형식, 경희대학교, 1998.
- [8] 최규정의, 로마자 표기 변환에 관한 연구, 한국정보과학회 추계학술 발표논문집, Vol. 27, No 2, pp194, 2000.
- [9] 박수호외, 한국어 음운 현상과 단어 유형을 고려한 한글-로마자 변환 시스템 구현, 한국정보과학회 춘계학술 발표논문집, Vol. 28, No 1, pp376, 2001.
- [10] 장연세, 참조 컴포넌트 아키텍처 모델과 UML 명세화에 대한 연구, 한국OA학회 논문집, 제6권 제3호, 2001.

저자소개

김 홍 섭

1984년 동국대학교 학사(전자계산)

1994년 동국대학교 대학원 석사
(전산교육)

2001년 동국대학교 대학원

컴퓨터공학 박사과정 수료

현재 오산대학 인터넷정보관리과

교수

(관심분야) 분산운영체제, 성능평가,

웹 서비스, 유비쿼터스 등