

로마자 표기 변환에 관한 연구

최규정^U 유성석 박기홍
군산대학교 컴퓨터 정보학과
{gyujung, sssyou, kh.park}@cs.kunsau.ac.kr

A Study on Transformation System for Roman Notation

Gyu-jeong Choi^U Seong-Seok You Ki-hong Park
Dept. of Computer Information Science, Gunsan National University

요 약

본 논문은 한글 고유 명사를 새로 개정된 한글 로마자 표기법에 의해 로마자 표기로 변환하는 시스템의 구성에 대해 기술하였다. 1995년에 문화체육부에서 배포한 프로그램 이후로 현재까지 한글을 로마자 표기로 자동 변환하는 프로그램이 나오지 않았고 이 프로그램 또한 새로 개정된 로마자 표기법이 아닌 예전의 로마자 표기법을 따르는 이유로 본 논문은 새로 개정된 표기법을 따르는 한글 고유명사의 로마자 자동 변환 시스템을 DB를 이용하여 구축하는 방안을 제시하였다. 그 결과 DB를 이용한 한글의 로마자 자동변환시스템은 개발자 측면에서는 프로그램 구현과 정확성 면에서 효과적인 장점을 얻을 수 있으며, 사용자 측면에서는 새로 개정된 개정안을 쉽게 익히고, 사용할 수 있게 한다.

1. 서론

그동안 말많은 국어의 알파벳 표기원칙에 대해 국립국어 연구원은 지난 1999년 11월 17일 국어 로마자 표기법 개정 시안을 발표했다. 이는 지금까지 나온 로마자 표기안 가운데 가장 합리적인 것으로 평가되고 있다.

이는 새로 발표된 개정안이 표기형식 위주의 반달표(˘)와 어긋점(ˆ)을 없애고, 실제 발음을 중요시했기 때문이다.

이에 현재 정부가 고시한 국어의 로마자 표기법을 온 국민들이 쉽게 익히고, 외국인들에게는 우리말에 더욱 친숙해질 수 있도록 하기 위해 국어의 로마자 변환 프로그램이 필요한 실정이다.

본 논문에서는 새로 발표된 개정안에 따른 국어의 로마자 변환프로그램을 위한 한글 고유명사의 로마법 표기 자동 생성 알고리즘을 제시하고자 한다.

본 논문의 구성은 2장에서는 기존 로마자 표기법의 문제점을 제시했고, 3장에서는 본 논문이 제시하는 시스템에서 적용하는 규칙과 시스템의 구성, 구현에 대해 기술하였고, 4장

에서는 결론에 대해서 기술하였다.

2. 기존 로마자 표기법의 문제점

기존 표기법은 첫째 정보화 시대에 맞지 않다. 반달표와(˘)와 어긋점(ˆ)을 그대로 사용하여 컴퓨터에서 사용하기 불편하고 표기법과 상관없이 마음대로 표기하여 로마자 표기가 혼란스러워졌다.

둘째 종전의 표기법은 너무 어려운 표기법이다. 유무성의 차이를 느끼지 못하는 한국인에게 유성음과 무성음을 나누어 표기하는 종전의 표기법은 이해하기 어려웠다.

셋째 종전의 표기법에서는 ‘ㄱ, ㄷ, ㅌ, ㅈ’ 과 ‘ㅋ, ㅌ, ㅍ, ㅊ’이 제대로 구별되지 않았다.

이에 새로운 개정안이 발표되었는데 이를 적용하는 시스템이 아직 구현되지 않고 있다.

3. 한글 로마자 표기 변환 시스템

본 시스템에서는 1999년에 고시한 국어 로마자 표기법을 적용하여 한글의 고유명사를 로마자로 자동 생성하는 프로그램의 적용규칙과 구성에 대해 기술하고자 한다.

3.1 적용규칙

먼저, 한글의 고유명사를 로마자로 변환하려면 다음의 2가지 규칙을 적용하여야 한다.

1. 입력된 한글 문자열이 음운 변화가 일어날 때 결과에 따라 발음되는 형태로 변환한다.
2. 발음되는 형태로 바뀐 한글 문자열을 국어 로마자 표기법에 따라 로마자 문자열로 변환한다.

예) 신라 ⇒ 실라 ⇒ Silla

규칙1. 음운변화에 따른 발음 형태로의 변환 규칙[1][2]

국어의 로마자 표기는 표준 발음을 적는 것을 원칙으로 한다. 물론 글자가 곧 발음인 경우는 문제가 없다. 글자와 발음이 상이한 경우가 문제인데 그런 경우에 글자를 기준으로 적지 않고 발음을 기준으로 표기한다. 글자와 발음이 상이한 경우는 크게 다섯 가지로 나눌 수 있다.

첫째, 자음 사이에서 동화 작용이 일어나는 경우이다. 이 경우도 몇 가지로 나누어 볼 수 있다.

(가) 'ㄱ, ㅂ'이 비음인 'ㄴ, ㄹ' 앞에서 비음인 'ㅇ, ㄹ'으로 바뀌는 비음화 현상이다.

예) 백마 → [뵤마], 감문 → [감문]

(나) 'ㄹ'은 비음인 'ㄹ, ㅇ' 다음에서 반드시 비음인 'ㄴ'으로 바뀐다.

예) 탐라 → [탐나], 종로 → [종노]

(다) 'ㄹ'은 'ㄱ, ㅂ' 다음에서 'ㄴ'으로 바뀌는데 이 때는 'ㄴ'으로 바뀌는데 이 때는 'ㄱ, ㅂ'도 'ㅇ, ㄹ'으로 바뀐다.

예) 심리 → [심니], 독립 → [동닙]

(라) 'ㄹ' 다음에 'ㄴ'이 올 때에는 'ㄴ'이 'ㄹ'로 바뀐다.

예) 별내 → [별래]

(마) 'ㄴ' 다음에 'ㄹ'이 올 때에 앞의 'ㄴ'이 'ㄹ'로 바뀌기도 한다.

예) 신라 → [실라]

둘째, 'ㄴ, ㄹ'이 덧나는 경우이다. 합성어에서 둘째 요소가 '야, 여, 요, 유, 애, 예' 등으로 시작되는 말이면 'ㄴ, ㄹ'이 덧난다.

예) 학여울 → [항녀울], 물약 → [물략]

셋째, 구개음화가 되는 경우이다. 'ㄷ, ㅌ' 다음에 '이'가 오면 'ㄷ, ㅌ'이 'ㅈ, ㅊ'으로 바뀐다.

예) 굳이 → [구지], 같이 → [가치]

넷째, 'ㄱ, ㄷ, ㅂ, ㅈ'이 'ㅎ'과 합하면 거센소리가 된다.

예) 좋고 → [조코], 놓다 → [노타]

다만, 'ㄱ, ㅂ' 다음에 'ㅎ'이 오는 체언의 경우 어원의식을 존중하여 'ㅎ'을 살려서 적는다.

예) 목호 → [무코] → Muko → Mukho

다섯째, 된소리현상

예) 울산 → [울싼] → Ulsan → Ulsan

팔당 → [팔땅] → Palttang → Paldang

규칙2. 국어 로마자 표기법에 따른 로마자 표기 규칙

1. 표기의 기본원칙[1][2]

(가) 국어의 로마자 표기는 국어를 표준발음에 따라 적는다.

(나) 로마자 이외의 부호는 되도록 사용하지 않는다.

(다) 1음운 1기호의 표기를 원칙으로 한다.

2. 표기일람

(가) 모음은 다음과 같이 적는다.

단모음:

ㅏ(a) ㅑ(eo) ㅓ(o) ㅕ(u) ㅡ(eu) ㅣ(i) ㅞ(ae)

ㅚ(e) ㅜ(oe) ㅟ(wi)

이중 모음:

ㅟ(ya) ㅟ(yeo) ㅟ(yo) ㅟ(yu) ㅟ(yae) ㅟ(ye)

ㅟ(wa) ㅟ(wae) ㅟ(wo) ㅟ(we) ㅟ(ui)

[붙임] 장모음의 표기는 따로 하지 않는다.

(나) 자음은 다음과 같이 적는다.

파열음:

ㄱ(g,k) ㅋ(kk) ㆁ(k) ㄷ(d,t) ㅌ(tt) ㅌ(t) ㅃ(b,p)

ㅍ(pp) ㅍ(p)

과찰음:

ㅈ ㅉ ㅊ

j jj ch

마찰음:

ㅅ ㅆ ㅎ

s ss h

비음:

ㄴ ㄹ ㅇ

n m ng

유음:

ㄹ

r, l

[붙임1] 'ㄱ, ㄷ, ㅂ'은 모음 앞에서는 'g, d, b'로 적고, 자음 앞이나 어말에서는 'k, t, p'로 적는다.

(보기)

구미 Gumi 영동 Yeongdong 백암 Baegam
옥천 Okcheon 합덕 Hapdeok 태백 Taebaek
갑곶 Gapgot 한밭 Hanbat 호법 Hobeop

[붙임2] 'ㄹ'은 모음 앞에서는 'r'로, 자음 앞이나 어말에서는 'l'로, 'ㄹㄹ'은 'll'로 적는다.

(보기)

구리 Guri 설악 Seorak 칠곡 Chilgok
임실 Imsil 울릉 Ulleung 대관령 Daegwallyeong

3. 표기상의 유의점

제1항 음운 변화가 일어날 때에는 변화의 결과에 따라 다음과 같이 적는다.

1. 자음 사이에서 동화 작용이 일어나는 경우

(보기)

약목 Yangmok 만머느리 manmyeoneuri
읍내 eumnae 신문로 Sinmunno 설날 seollal
종로 Jongno 왕십리 Wangsimni 신라 Silla

2. 'ㄴ, ㄹ'이 덧나는 경우

(보기)

꽃잎 kkonnip 담요 damnyo
들일 deullil 물약 mullyak

3. 구개음화가 되는 경우

(보기)

해돋이 haedoji 같이 gachi 묻히다 muchida

4. 'ㄱ, ㄷ, ㅂ, ㅈ'이 'ㅎ'과 어울려 나는 경우

(보기)

목호 Muko 놓다 nota

집현전 Jipyeonjeon 맞히다 machida

[붙임1] 단어 내부에서의 경음화는 표기에 반영하지 않는다.

(보기)

압구정 Apgujeong 낙동강 Nakdonggang
죽변 Jukbyeon 낙성대 Nakseongdae
합정 Hapjeong 석굴암 Seokguram
팔당 Paldang 울산 Ulsan
물고기 mulgogi 눈동자 nundongja
셋별 saetbyeol 산새 sansae

[붙임2] 인명 표기에서 한자어 이름의 사이와, 행정 구역 표기에서 앞말과 단위명 사이에서 일어나는 음운 변화는 표기에 반영하지 않는다.

(보기)

한복남 Han Bok-nam 홍경래 Hong Gyeong-rae
사북면 Sabuk-myeon 구산리 Gusan-ri

제2항 발음상 혼동의 우려가 있을 때에는 붙임표(-)를 쓸 수 있다.

(보기)

중앙 Jung-ang 준강 Jun-gang

제3항 고유명사는 첫 글자를 대문자로 적는다.

(보기)

부산 Busan 세종 Sejong 삼국사기 Samguksagi

제4항 인명은 성과 이름의 순서로 띄어 쓴다. 한자어 이름일 경우에는 ㄱ을 원칙으로 하고 'ㄴ'과 'ㄷ'도 허용한다.

(보기)

ㄱ ㄴ ㄷ

윤숙영 Yun Suk-yeong Yun Sukyeong
Yun Suk Yeong

남성렬 Nam Seong-ryeol Nam Seongryeol
Nam Seong Ryeol

한하나 Han Hana 송꽃님 Song Kkonnim

독립문 Dongnimmun 오죽헌 Ojukeon
촉석루 Chokseongnu 종묘 Jongmyo
다보탑 Dabotap 광희문 Gwanghimun

[붙임] 한글 표기에서 띄어 쓰는 말은 띄어 쓰는 단
위마다 붙임표(-)를 쓸 수 있다.

(보기)

금동 미륵보살 반가상

Geumdong-mireukbosal-bangasang

도산 서원 Dosan-seowon

성덕 대왕 신종 Seongdeok-daewang-sinjong

제5항 '도, 시, 군, 구, 읍, 면, 리, 동'의 행정 구역
단위와 '가'는 각각 'do, si, gun, gu, eup, myeon,
ri, dong, ga'로 적고, 그 앞에는 붙임표를 넣는다.

(보기)

충청북도 Chungcheongbuk-do

제주도 Jeju-do

의정부시 Uijeongbu-si 파주군 Paju-gun

도봉구 Dobong-gu 신창읍 Sinchang-eup

주내면 Junae-myeon 인왕리 Inwang-ri

당산동 Dangsang-dong

봉천 1동 Bongcheon 1(il)-dong

종로 2가 Jongno 2(i)-ga

퇴계로 3가 Toegyero 3(sam)-ga

제7항 인명, 회사명, 단체명 등은 그동안 써 온 표기
를 허용한다.

3.2 한글 로마자 표기 시스템의 구현

3.2.1 알고리즘

[붙임] 특별시, 광역시, 시, 군, 읍 등의 행정 구역
단위명은 생략할 수 있다.

(보기)

인천광역시 Incheon 순창읍 Sunchang

제6항 자연 지물명, 문화재명, 인공 축조물명은 붙임
표(-) 없이 붙여 쓴다.

(보기)

남산 Namsan 속리산 Songnisan

금강 Geumgang 독도 Dokdo

해운대 Haeundae 경복궁 Gyeongbokgung

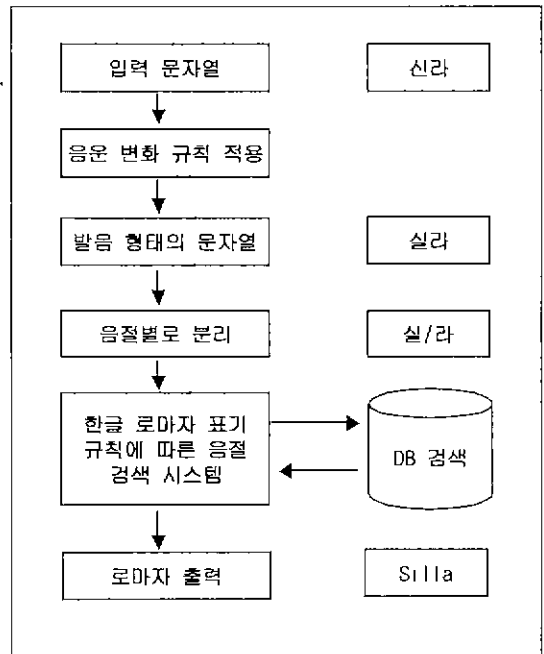
무량수전 Muryangsujeon 돈의문 Donuimun

연화교 Yeonhwagyo 극락전 Geungnakjeon

안압지 Anapji 남한산성 Namhansanseong

화랑대 Hwarangdae 불국사 Bulguksa

희방사 Hibangsa 현충사 Hyeonchungsa



[그림1] 로마자 변환 알고리즘

3.2.2 구성내용

1) 음운 변화 규칙을 적용한 발음 형태의 문자열

생성

입력된 문자열이 곧 발음인 경우 문제가 되지 않지만, 글자와 발음이 상이한 경우 음운 변화 규칙을 적용하여야 한다. 먼저 입력된 문자열이 음운 변화 규칙에 적용되는 경우이면 문자열을 규칙에 따라 발음 형태로 변환하고 그렇지 않으면 입력된 문자열 그대로 사용한다. 변환된 문자열은 DB검색에 맞게 음절별로 분리한다.

2) 한글 로마자 표기 규칙에 따른 자소 검색 시스템

음절별로 분리한 한글 문자들을 로마자 형태의 문자열로 변환할 때 본 시스템은 DB검색을 이용하였다.

그러나 음절별로 분리한 한글 문자를 바로 DB검색에 이용할 수 없다. 그 이유는 다음과 같은 규칙이 적용되어야 하기 때문이다.

- 규칙1. 된소리되기는 표기의 안정을 위하여 발음이 희생되는 한이 있더라도 표기에 반영하지 않는다
- 규칙2. 인명 표기에서 한자어 이름의 사이와, 행정 구역 표기에서 앞말과 단위명 사이에서 일어나는 음운 변화는 표기에 반영하지 않는다.
- 규칙3. 'ㄱ, ㄷ, ㅂ'은 모음 앞에서는 'g, d, b'로 적고, 자음 앞이나 어말에서는 'k, t, p'로 적는다.
- 규칙4. 'ㄹ'은 모음 앞에서는 'r'로, 자음 앞이나 어말에서는 'l'로, 'ㄹㄹ'은 'll'로 적는다.

이러한 이유로 DB 구축 시에 같은 한글 음절이더라도 두 가지 이상의 로마자 표기를 전부 입력하여야 하고, 마찬가지로, DB검색시도 이러한 규칙에 맞는 로마자 표기에 접근하여 추출하여야 한다.

- (보기) 가 → ga 라 → ra
- 각 → gak 라 → la
- 간 → gan

3.2.3 구현환경

운영체제 : 윈도우2000(Server용)

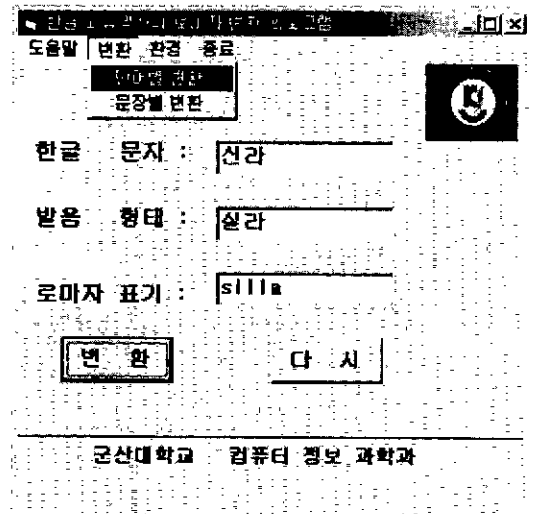
프로그램 언어 : 비주얼 베이직 6.0

데이터 베이스 : Microsoft Data Bound Grid Control 6.0

3.2.4 결과화면

	한글문자	로마자 변환
	가	ga
	각	gak
	간	gan
	갈	gat
	갈	gal
	갈	galk
	갈	galm
	감	gam

[그림2] DB 구성도



[그림3] 로마자 변환 화면

4. 결론

본 논문은 개정안 발표 이후 이를 적용하는 새로운 시스템을 제안하고자 하였다. 기존의 초성, 중성, 종성

조합형 시스템을 탈피하여 새로이 DB를 접목하여 변환함으로써 기존 시스템보다 변환 시간이 빠르고 정확성도 뛰어난 시스템을 구현함으로써 사용자들이 개정안에 좀더 쉽게 접근하고 사용자들에게 통일된 규칙을 적용하고자 한다.

참고 문헌

- [1] 국어의 로마자 표기법 - 국어 국립 연구원, 문화 체육부 1998
- [2] 한글 맞춤법 - 국어 어문 규정집, 문화체육부, 1996
- [3] 기계화를 위한 한글의 로마자 표기법 소프트웨어 개발 - 한국표준과학연구소, 1990
- [4] 국어로마자표기 자동변환시스템 설계 및 구현 -한국OA 학회, 김홍섭, 1993
- [5] 한글로마자 자동표기 전사법 기계적 변환에 관한 연구 - 한글 및 한국어 정보처리학술발표논문. 김홍섭, 1992