

한국어 음운 현상과 단어 유형을 고려한

한글-로마자 변환 시스템 구현

*박수호^U *김광영 **Robert J. Fouser *권혁철
*부산대학교 전자계산학과, **카고시마 대학교

(suhopark, kykim)@solge.cs.pusan.ac.kr, rjfouser@msg.biglobe.ne.jp,
hckwon@hyowon.pusan.ac.kr

An Implementation of Hangul Romanization System Based on Korean Phonotactic Rules and Categories of Words

*Su-ho Park^U *Gwang-yeong Kim **Robert J. Fouser *Hyuk-Chul Kwon
*Dept. of Computer Science, Pusan National University, **Kagoshima University

요 약

본 논문은 새로 개정된 국어의 로마자 표기법(2000.7.7.고시)에 따른 로마자 자동 변환 시스템의 구조와 변환 알고리즘을 제시한다. 특히 새 로마자 표기에 따른 규칙을 반영하여 사용자가 쉽고 편리하게 사용할 수 있도록 인명, 고유명사, 행정구역, 일반, 학술응용의 5개 항목으로 나누어 변환할 수 있게 했다. 또한 로마자표기에 적용한 표준 한글 음운변동 규칙과 로마자 표기의 원칙에 따른 5가지 변환 규칙을 도움말로 보여준다. 그리고 로마자 표기 원칙에 따라 발음상 혼동의 우려가 있는 부호의 규칙성을 조사하여 로마자 자동 변환 시스템에 적용함으로써 새 로마자 표기에서 발생할 수 있는 혼동을 제거하여 로마자 변환 시스템의 정확성을 높였다.

1. 서론

개정 이전의 로마자 표기법에서는 반달표(˘)와 어긋점(˙)의 같은 불편한 기호 때문에 표기법을 쉽게 사용할 수 없어 로마자 표기의 혼란을 가져왔다. 더구나 과거 표기법에 있는 반달표나 어긋점이 붙은 기호는 컴퓨터와 인터넷에서 쉽게 사용할 수 없어서 로마자 표기의 어려움이 가중되었다. 새로 개정된 국어의 로마자 표기법은 이러한 단점을 보완하여 쉽게 사용할 수 있으면서 비교적 단순하고 간결하게 만들었다.

이 논문은 국어의 로마자 표기법에 따른 자동 변환 시스템을 위한 변환 알고리즘을 제시한다. 또 새로 개정된 로마자 표기법의 가장 큰 문제인 발음상 혼동을 가져올 수 있는 표기를 해결하기 위해 도입한 불입표(-)의 사용규칙을 정의한다. 또 인명, 고유명사, 행정구역(교통관련지명 포함), 일반, 학술 응용으로 로마자 표기에 따른 원칙을 세분화하여 로마자 변환 시스템을 쉽게 사용할 수 있게 한다.

2. 로마자 변환 시스템

2.1 기본원칙

로마자 변환 시스템은 문화관광부고시 제2000-8호(2000.7.7.)를 따른다. 제1장 표기의 기본 원칙 제1항 국어의 로마자 표기는 국어의 표준 발음법에 따라 적는 것을 원칙으로 하며, 제2항에 따라 로마자 이외의 부호는 되도록 사용하지 않는다.

2.2 표기 일람

국어의 로마자 표기법 제2장 표기 일람 제1항과 제2항에 따

라 국어의 모음과 자음은 다음과 같이 적는다.

단모음	ㅏ	ㅑ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅣ	ㅞ	ㅟ	ㅝ	ㅞ
	a	eo	o	u	eu	i	ae	e	oe	wi				

이중모음	ㅘ	ㅙ	ㅚ	ㅜ	ㅝ	ㅞ	ㅟ	ㅠ	ㅡ	ㅢ	ㅣ	ㅤ	ㅥ	ㅦ
	ya	yeo	yo	yu	vae	ye	wa	wae	wo	we	ui			

파열음	ㄱ	ㄲ	ㅋ	ㄷ	ㄸ	ㅌ	ㅍ	ㅑ	ㅑ	ㅑ	ㅑ	ㅑ	ㅑ	ㅑ
	g,k	kk	k	d,t	tt	t	b,p	pp	p					

파찰음	ㅈ	ㅉ	ㅊ	마찰음	ㅅ	ㅆ	ㅎ	비음	ㄴ	ㄹ	ㅇ	유음	ㄹ
	j	jj	ch		s	ss	h		n	m	ng		rl

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

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

2.3 표기상의 유의점

국어의 로마자 표기법 제3장 제1항 음운 변화가 일어날 때에는 변화의 결과에 따라 다음과 같이 적는다.

제1항 음운 변화가 일어날 때에는 다음과 같이 적는다.

-자음 사이에서 동화 작용이 일어날 때

[보기]백마 Baengma 신문로 Sinmunno 신라 Silla

-'ㄴ, ㄹ'이 덧붙일 때

[보기]학여울 Hangnyeoul 알약 allyak

-구개음화가 될 때

[보기]해돋이 haedoji 같이 gachi 맞히다 machida

-구개음화가 일어날 때
 [보기]해돋이 haedoji 같이 gachi 맞히다 machida
 -ㄱ, ㄷ, ㅂ, ㅈ이 'ㅎ'과 합하여 거센소리로 될 때
 [보기]종고 joko 놓다 nota 잡혀 japyeo
 단, 체언에서 'ㄱ, ㄷ, ㅂ' 뒤에 'ㅎ'이 따를 때에는 'ㅎ'을 밝혀 적는다.

[보기]목호 Mukho 집현전 Jiphyeonjeon
 [붙임]된소리되기는 표기에 반영하지 않는다.
 [보기]압구정 Apgujeong 팔당 Paidang 울산 Ulsan
 제2항 발음상 혼동의 우려가 있을 때에는 음절 사이에 붙임표(-)를 쓸 수 있다.

[보기]중앙 Jung-ang 해운대 Hae-undae
 제3항 고유 명사는 첫 글자를 대문자로 적는다.

[보기]부산 Busan 세종 Sejong
 제4항 인명은 성과 이름의 순서로 띄어 쓴다. 이름은 붙여 쓰는 것을 원칙으로 하되 음절 사이에 붙임표(-)를 쓰는 것을 허용하며, 이름에서 일어나는 음운 변화는 표기에 반영하지 않는다. 단, 종성의 대표음 발음은 인정한다. 또한 성의 표기는 따로 정한다.

[보기] [원칙]한복남 Han Boknam 홍빛나 Hong Bitna
 [허용]한복남 Han Bok-nam 홍빛나 Hong Bit-na
 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 붙임표(-)를 넣는다. 붙임표(-) 앞뒤에서 일어나는 음운 변화는 표기에 반영하지 않으며, '시, 군, 읍'의 행정 구역 단위는 생략할 수 있다.

[보기] 경기도 Gyeonggi-do 의정부시 Uijeongbu-si
 제6항 자연 지물명, 문화재명, 인공 축조물명은 붙임표(-) 없이 붙여 쓴다.

[보기] 남산 Namsan 독도 Dokdo
 제7항 인명, 회사명, 단체명 등은 그 동안 써 온 표기를 쓸 수 있다.

제8항 학술 연구 논문 등 특수 분야에서 한글 복원을 전제로 표기할 때에는 한글 표기를 대상으로 적는다. 이때 글자 대응은 제2장을 따르되 'ㄱ, ㄷ, ㅂ, ㄷ'은 'g'd'b'l'로만 적는다. 음가 없는 'ㅇ'은 붙임표(-)로 표기하되 어두에서는 생략하는 것을 원칙으로 한다. 기타 분절의 필요가 있을 때에도 붙임표(-)를 쓴다.

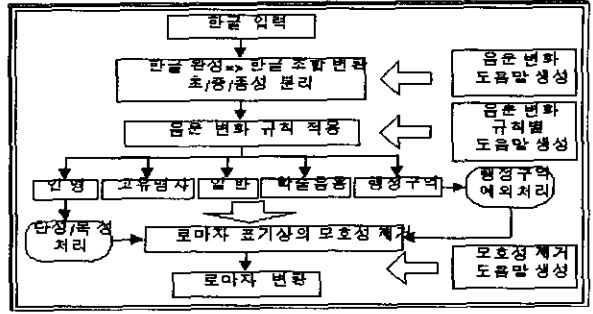
[보기] 없었습니다 eobs-eoss-seubnida 집 jib 쥘 jip

2.4 로마자 변환 시스템 구현
 로마자 변환 시스템을 인명, 고유명사, 행정구역(교통관련지명 포함), 일반, 학술응용 5개 항목으로 분류하여 사용자가 국어의 로마자 표기법에 익숙하지 않아도 쉽게 사용할 수 있게 하였고, 변환 과정에서 일어나는 음운 변화와 기타 예외적인 사항들을 도움말에서 자세히 알 수 있도록 하여 일반 사용자에게 로마자 표기법을 쉽게 이해할 수 있도록 하였다. 웹(urimal.cs.pusan.ac.kr)에서 일반 사용자에게 공개(2000.9~2001.2)하여 얻어진 결과는 다음과 같다.

구분	인명	고유 명사	행정 구역	일반	학술 응용	외래어	합계
건수	1271	876	713	628	374	29	3891
%	33	22	18	16	10	1	100

[표-1] 웹에서 항목별 사용 빈도

2.5 로마자 변환 시스템 알고리즘



[그림-1] 로마자 자동 변환기 알고리즘

2.5 세부내용
 2.5.1 한글 완성/조합 변환기
 국어의 로마자 표기는 표준 발음법을 따르기 때문에 글자와 발음이 서로 다를 수 있다. 이러한 문제로 한글을 로마자로 변환할 때 사용자에게는 표기의 혼동을 가져 올 수 있으며, 개발자에게는 시스템 구현의 어려움을 준다. 이런 문제를 해결하기 위해서 본 로마자 변환 시스템에서는 한글 완성/조합 변환기를 사용한다. 완성형 한글을 조합형 한글로 자동 변환하여 이를 초·중·종성으로 분리한 후, 음운 변화를 쉽게 적용할 수 있도록 하여 사용자에게는 정확한 로마자 변환을 보여주며, 개발자에게는 쉽게 시스템 구현을 할 수 있게 해준다.

2.5.2 음운 변화 규칙 적용
 한글 완성/조합 변환기를 사용하여 초·중·종성으로 분리한 한글을 앞 글자 종성과 이어진 글자 초·중성에서 일어나는 음운 변동을 적용하고 사용자가 이 음운 변동을 알기 쉽도록 도움말을 생성한다.

2.5.3 인명, 고유명사, 행정구역, 일반, 학술응용 분류
 국어의 로마자 표기법을 좀더 정확하게 나타내기 위해 인명, 고유명사, 행정구역, 일반, 학술응용의 5개항으로 분류하였다.

1)인명:로마자 표기법 제3장 표기상의 유의점 제4항을 따른다. 본 논문에서는 로마자 표기법에서 인명을 [원칙]과 [허용]으로 나누어 [원칙]에서는 성과 이름을 띄어 표기하고, [허용]에서는 성의 관습적 표기와 이름에 붙임표(-) 사용을 보여준다. 또한 270여 개의 한국인의 성씨를 전자사전으로 구축하여 '남궁, 선우, 제갈' 등과 같은 복성과 '김, 이, 박' 등과 같은 단성을 구분하여 표기한다. 그리고, 인명에서 일어나는 규칙을 도움말로 보여 일반 사용자에게 인명에 대한 로마자 표기를 쉽게 익히도록 한다.

[보기]원칙: 남궁광식 Nangung Gwangsik 홍빛나 Hong Bitna
 허용: 남궁광식 Nangung Gwang-sik 홍빛나 Hong Bit-na

2)고유명사:인명과 행정구역(교통관련지명 포함)을 제외하고 나머지 고유명사를 로마자로 변환한다. 발음상 혼동의 우려가 있을 때에는 음절 사이에 붙임표(-)를 쓸 수 있다.

[보기]원칙: 해운대 Haeundae 허용: 해운대 Hae-undae
 3)행정구역:로마자 표기법 제3장 제5항을 따른다. 또한 인명, 회사명, 단체명 등은 그동안 써 온 표기를 쓸 수 있다. 제3장 제7항의 규정을 위해 학교 이름 등이 역 이름으로 쓰거나, 실용상의 편의를 위해 널리 통용되는 영어 번역어 등을 전자 사전으로 구축하여 로마자 표기를 보여준다.

[보기]부산시 Busan-si 부산대앞 Busan National Univ.
광안리 해수욕장 Gwangalli beach

4)일반:간단한 문장을 변환한다.

5)학술응용:로마자 표기법 제3장 제8항을 따르며,한글 복원을 전제로 한다. 어두를 제외한 초성 'ㅇ'은 불임표(-)로 표기하며, 기타 분절의 필요나 발음의 혼동의 여지가 있을 때에 반드시 불임표(-)를 표기한다.

[보기]없었습니다 eobs-eoss-seubnida 값 gabs 굳이 gud-i

2.5.4 로마자 표기상의 모호성 제거

로마자 표기에서 발음에서 혼동을 가져올 수 있는 부호가 있는데 이는 로마자 표기법을 사용하는 사용자에게는 로마자 표기 사용의 어려움을 주고, 시스템을 구현하는 개발자에게는 시스템의 정확성을 떨어뜨리는 원인이 된다. 본 논문은 이러한 문제를 해결하기 위해 다음의 표에서 이들을 정리하여 시스템 구현의 정확성을 높이고, 로마자 표기에서 발음의 혼동을 제거하였다.

1)연속된 단모음

구분	어(eo)	으(eu)	에(e)	혼동하는 글자	
아(a)	aeo	aeu	ae	ae(에)	규칙:e앞에 a,o가 올 수 없다
오(o)	oeo	oeu	oe	oe(외)	
어(eo)	eeo	eeu	ee	ee(외)	

[표-2] e와 a,o 관계

구분	오(o)	외(oe)	혼동하는 글자	
에(e)	eo	oe	oe(에)	규칙:o앞에 e가 올 수 없다
애(ae)	aeo	aeoe	eo(어)	
외(oe)	oeo	oeoe	eo(어)	

[표-3] o와 e의 관계

구분	이(i)	혼동하는 글자	
우(u)	ui	ui(의)	규칙:i앞에 u가 올 수 없다.
으(eu)	eui	e-ui(의)	

[표-4] i와 u의 관계

2)중성'ㅇ'과 단모음

구분	단모음	혼동하는 글자	
중성'ㅇ'	ng+a,e,i,o... (wi제외)	n-ga n-ge ...	규칙:ng뒤에 단모음(위(wi)제외)이 올 수 없다.

[표-5] 중성'ㅇ'과 단모음과의 관계

3)한글 복원을 전제로 한 학술 응용에서 중성에 연속된 초성의 정리

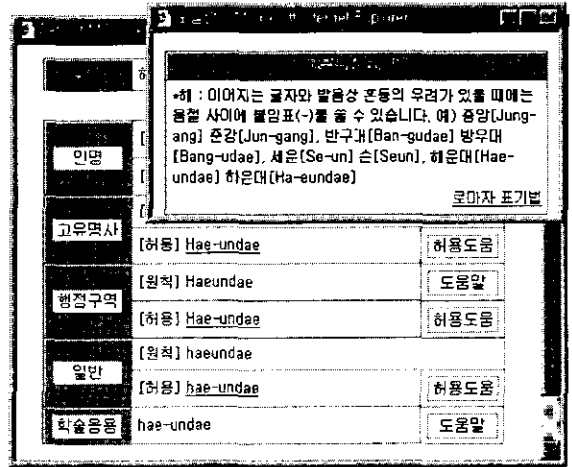
초성	ㄱ(s)	ㄴ(ss)	ㅋ(k)	ㆁ(kk)	ㅍ(p)	ㅌ(t)	ㅈ(j)
ㄱ(s)	ss	sss
ㄴ(ss)	sss
ㄱㄴ(gs)	gss
ㅂㄴ(bs)	bss
ㄹㄴ(ls)	lss
ㅋ(k)	.	.	kk	kkk	.	.	.
ㆁ(kk)	.	.	kkk
ㅍ(p)	pp	.	.
ㄹㅍ(lp)	lpp	.	.
ㅌ(t)	tt	.
ㄹㅌ(lt)	ltt	.
ㅈ(j)	jj
ㄹㅈ(lj)	ljj

[표-6] 중성과 초성과의 관계

2.6 도움말 생성

국어의 로마자 표기는 국어의 표준 발음법에 따라 적는 것을 원칙으로 하기 때문에 글자와 발음이 다를 때 로마자 표기법에 익숙하지 않은 사람들에게 사용의 불편을 끼칠 수 있다. 마찬가지로 인명,고유명사, 행정구역,일반,학술응용의 5개항에서 일어나는 로마자 표기에 차이가 있어 로마자 표기법에 혼란을 가져올 수 있다. 이러한 점을 보완하기 위해 로마자 변환기에서 발생하는 음운 변동 및 인명,고유명사,행정구역,일반,학술응용의 분류 단계와 기타 예외 사항에 대해 도움말을 생성하여 사용자에게 로마자 표기법을 알기 쉽게 한다.

2.6 로마자 변환 시스템 화면



[그림-2] 로마자 자동 변환기(urimal.cs.pusan.ac.kr)

3. 결론

로마자 변환에서 인명, 고유명사, 행정구역(교통관련지명 포함), 일반, 학술 응용으로 세분화하여 사용자에게 로마자 표기를 더욱 쉽게 사용할 수 있게 하고, 로마자 표기의 발음에서 혼동의 여지를 해결하기 위해 모호성 제거를 위한 규칙을 제시하여 로마자 변환 시스템의 정확성을 높였다.

4. 참고 문헌

[1]. 로마자 표기 용례 사전 - 문화관광부, 국립국어연구원, 2000
 [2]. 한글 맞춤법 - 국어 어문 규정집, 문화관광부, 1996
 [3]. 한글로마자 자동표기 전사법 기계적 변환에 관한 연구 - 한글 및 한국어 정보처리 학술발표 논문, 김홍섭, 1992
 [4]. 로마자 표기 변환에 관한 연구 - 한글 및 한국어 정보처리 학술 발표논문, 최규정, 2000
 [5]. http://www.koreandb.net - 한국인의 성씨, 디지털 한국학