한국어 로마자 변환기 개발에 관한 연구

Development of Romanization System of Korean

윤방원*, 정태충

경희 대학교 전자계산공학과 인터넷 & 지능시스템 연구실

요 약

본 논문은 한국어 로마자의 표준안 확립을 위해 로마자 변환 테이블을 정리 및 종합하고 이를 실제 변환기로 개발하는데 필요한 과정에 대해 기술하였다. 테이블 구성을 위해 국어의 표준 발음의 전체 구성을 분석하고, 불필요한 규칙이 있거나 없을 경우의 규칙을 추가, 삭제하여, 상충이 있을 경우 우선 순위에 의해 규칙을 선택하여 규칙을 수정하여 적용하였다. 이렇게 생성된 표준 발음 테이블은 한글을 받아 적용한 후 표기별 자음과 모음의 테이블을 통해 로마자로 변환된다. 일관성 없는 규칙, 표준화의 설득력 및 홍보의 미숙 등과 같은 과제로 활용의 예가 적었던 한국어 로마자 변환기는 공공기관의 지명 및 문화재 로마자의 표준화 등에 사용될 수 있으며, 인터넷을 통한 웹 로마자 사전 제작을 통하여 표준화의 문제에 가장 중요한 객관적인 지표로서 활용될 수 있다.[3]

1. 서론

처음 1948년 문교부에서 <한글을 로마자로 적는 법>을 제정·고시(최초의 정부안)는 1959에 공포된 <한글의 로마자 표기법>, 1984년 <국어의 로마자 표기법>에 의해, 1999년 <국어의 로마자 표기법(시안)>으로 발전되어 왔다.

1999년 <국어의 로마자 표기법(시안)>에서 두드러진 현행 표기의 문제인 특수 부호 사용과 자음 유·무성의 구별을 지적하고 새로운 표기로 나타내는 것을 주장하고 있다[1]. 따라서, 본 논문은 이러한 한국어 로마자 변환이 일정하지 않고 예외를 포함하여 불완전한 규칙에 대해 표준 발음법에 의거한 규칙들을 모델링하고 정리, 종합하여 테이블을 구성하는 방법을 제시하고, 여러 한국어 로마자 변환 테이블로 로마자화하는 시스템을 구현하는 것으로써, 2장에는 표준 발음법에 의한 기존의 방법의 문제점과 새로 구성될 표준 발음법 표현에 대한 문제를 설명하며, 3장에서는 시스템의 설계 방법에 대하여 소개하고, 4장에서는 각 표준 발음법 규칙을 재적용한 테이블 및 로마자 변환 테이블을 마지막 5장에서는 결론과 향후 과제에 대해 설명한다.

2. 기존 로마자 표현에 있어 표준 발음의 처리

로마자 변환에 있어 로마자 변환 이전에 지금까지는 if-else로 자모와 형태소를 비교해 음운변동을 처리하여 표준발음 음가를 생성시키고자 하는 연구가 있어왔다[5]. 수 많은 비교문에 의해 많은 코드가 필요하였지만, 배열을 통한 테이블로부터 변경된 표준 발음 코드를 필요로 한다.

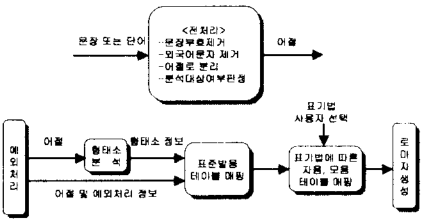
3. 시스템 설계 및 구현

한국어 로마자 변환기는 "표준어 발음법"을 바탕으로 기존의 주장되었던 로마자 표기에 대한 내용으로 개발된다. 그에 필요한 모듈은 다음과 같다.

- '표준어 발음법'에 따른 테이블 구성
- 다양한 로마자 변환 규칙을 수용
- 공개된 형태소 분석기와 예외사전을 통한 한글 문장 로마자 변환
- 한글의 일반 명사 및 단어를 입력하면 로마자로 출력 모듈

이러한 모듈로 구성되는 한국어 로마자 변환기의 시스템의 구조는 (그림1)에서와 같이 전처리부, 예외처리부, 형태소 분석기, 표준발음 테이블 매핑, 표기법에 따른 자음, 모음 매핑, 로마자 생성부로 나누어진

다. 또한, 본 변환기는 일반 사용자를 위해 인터넷을 통하여 사용할 수 있도록 동적인 프로그래밍이 가능한 ASP(Active Server Page)로 구현하며, Windows NT IIS(Internet Information Server)를 사용한다.



(그림1) 전체 시스템의 구조

4. 표준 발음 테이블 및 로마자 변환 테이블 작성

테이블은 중복과 상충을 처리할 수 있는 규칙을 선정하도록 해야 한다. 그렇게 구성된 테이블은 초성/종성 기반으로 테이블을 작성하며 문법 및 형태소 정보가 필요한 규칙의 표기를 작성하게 된다. 또한, 아래와 같이 중요한 중복 및 상충된 부분에 대한 처리를 하게 된다.

· 23항과 25항의 중복 : ㄼ은 23항에 포함되므로 어간받침을 고려할 필요가 없다.

· 18항과 30.2항의 중복 : 18항에서 사이시옷이 아니더라도 ㅅ은 무조건 'ㄴ, ㅁ'과 만나면[ㄴ]으로 발음되는데 30항에서 사이시옷일 경우 처리할 필요가 없다.

· 12.항.4, 13항, 12항.4와 14항의 상충 : 12.4항의 'ㅎ'은 13항에, 'ㄶ,ㅀ'은 14항에 포함되어 있어서 12항 4를 적용하면 이들은 탈락되고 13항과 14항에 따르면 연음이 되어야 한다.

· 13항과 15항에 ㅇ은 예외가 되어야 함 : 예를 들어 "공을"과 "공위"에서 알 수 있듯 받침 ㅇ은 예외가 되어야 한다.

· 30항의 3의 오류 : 사이시옷 뒤에 '이'음이 결 [ㄴㄴ]으로 발음된다고 하였지만 '이', '야', '여', '요', '유'가 모두 해당된다고 되어 있다[7].

<표 1>은 최종 음운변동 규칙 테이블이 되며 실제 시스템 구현에 형태소 분석기로부터의 정보를 입력받아 음운 변동값을 얻을 수 있다[5].

본 시스템은 최근의 발표된 국립 국어 연구원(1999)의 로마자 표기법을 기준으로 작성한다. 표기의 기본원칙은 다음과 같다[1].

· 제1항 국어의 로마자 표기는 국어의 표준 발음에 따라 적는다.

<표 1> 규칙을 재적용한 테이블

종성(행) \ 다음초성(열)		ㄱ	ㄲ	ㄴ	ㄷ	ㄸ	ㄹ	ㅁ	ㅂ	ㅃ	ㅅ	ㅆ		ㅎ	어말	모음조사	모음어미	모음접미사	ㅏㅗ기질태	ㅣ사미	히접미사	이야여요유
		0	1	2	3	4	5	6	7	8	9	10		18	19	20	21	22	23	24	25	26
-	0															ㅇ	ㅇ	ㅇ	ㅇ	ㅇ	ㅎ	
ㄱ	1	ㄲ		ㅇ	ㄸ		ㄴ ㅇ	ㅇ	ㅃ		ㅆ			ㅋ		ㄱ	ㄱ	ㄱ	ㄱ	ㅇ ㄱ	ㅎ ㄱ	
ㄲ	2	ㄲ ㄱ	ㄱ	ㅇ	ㄸ ㄱ	ㄱ	ㄴ ㄱ	ㅇ	ㅃ ㄱ	ㄱ	ㅆ ㄱ	ㄱ		ㅋ	ㄱ	ㄲ	ㄲ	ㄲ	ㄱ	ㅇ ㄱ	ㅎ ㄱ	
ㄳ	3	ㄲ ㄱ	ㄱ	ㅇ	ㄸ ㄱ	ㄱ	ㄴ ㄱ	ㅇ	ㅃ ㄱ	ㄱ	ㅆ ㄱ	ㄱ		ㅋ	ㄱ	ㅆ ㄱ	ㅆ ㄱ	ㅆ ㄱ	ㄱ	ㄲ ㄱ	ㅎ ㄱ	
ㄴ	4						ㄹ									ㄴ ㅇ	ㄴ ㅇ	ㄴ ㅇ	ㄴ	ㅇ ㄴ	ㅎ ㄴ	
ㄵ	5	ㄴ	ㄴ	ㄴ	ㄴ	ㄴ	[illegible]	ㄴ	ㄴ	ㄴ	ㄴ	ㄴ		ㅊ ㄴ	ㄴ	ㅈ ㄴ	ㅈ ㄴ	ㅈ ㄴ	ㄴ	ㅇ ㄴ	ㅎ ㄴ	
ㄶ	6	ㅋ ㄴ		ㄴ	ㅌ ㄴ						ㅆ ㄴ					ㅇ ㄴ	ㄴ	ㄴ	ㄴ	ㅇ ㄴ	ㅎ ㄴ	
ㄷ	7	ㄲ		ㄴ	ㄸ			ㄴ	ㅃ		ㅆ			ㅌ		ㄷ	ㄷ	ㄷ	ㄷ	ㅈ	ㅊ	
ㄹ	8			ㄹ												ㄹ	ㄹ	ㄹ	ㄹ	ㅇ ㄹ	ㅎ ㄹ	
ㄺ	9	ㄲ ㄱ	ㄱ	ㅇ	ㄸ ㄱ	ㄱ	ㄴ ㄱ	ㅇ	ㅃ ㄱ	ㄱ	ㅆ ㄱ	ㄱ		ㅋ ㄹ	ㄱ	ㄱ ㄹ	ㄱ ㄹ	ㄱ ㄹ	ㄱ	ㅇ ㄱ	ㅎ ㄱ	
ㄻ	10	ㅁ	ㅁ	ㅁ	ㅁ	ㅁ	ㄴ ㅁ	ㅁ	ㅁ	ㅁ	ㅁ	ㅁ	중	ㅁ	ㅁ	ㅁ ㄹ	ㅁ ㄹ	ㅁ ㄹ	ㅁ	ㅇ ㅁ	ㅎ ㅁ	
ㄼ	11	ㄲ ㄹ	ㄹ	ㅁ	ㄸ ㄹ	ㄹ	ㄹ	ㅁ	ㅃ ㄹ	ㄹ	ㅆ ㄹ	ㄹ	략	ㅍ ㄹ	ㄹ	ㅂ ㄹ	ㅂ ㄹ	ㅂ ㄹ	ㄹ	ㅇ ㄹ	ㅎ ㄹ	
ㄽ	12	ㄹ	ㄹ	[illegible]	ㄹ	ㄹ	ㄹ	ㄹ	ㄹ	ㄹ	ㄹ	ㄹ		ㄹ	ㄹ	ㅆ ㄹ	ㅆ ㄹ	ㅆ ㄹ	ㄹ	ㅇ ㄹ	ㅎ ㄹ	
ㄾ	13	ㄹ	ㄹ	ㄹ ㄹ	ㄹ	ㄹ	ㄹ	ㄹ	ㄹ	ㄹ	ㄹ	ㄹ		ㄹ	ㄹ	ㅌ ㄹ	ㅌ ㄹ	ㅌ ㄹ	ㄹ	ㅊ ㄹ	ㅎ ㄹ	
ㄿ	14	ㄲ ㅂ	ㅂ	ㅁ	ㄸ ㅂ	ㅂ	ㄴ ㅂ	ㅁ	ㅃ ㅂ	ㅂ	ㅆ ㅂ	ㅂ		ㅍ	ㅂ	ㅍ ㄹ	ㅍ ㄹ	ㅍ ㄹ	ㅍ	ㅇ	ㅎ ㅂ	
ㅀ	15	ㅋ ㄹ		ㄹ ㄹ	ㄹ						ㅆ ㄹ					ㅇ ㄹ	ㄹ	ㄹ	ㄹ	ㅇ ㄹ	ㅎ ㄹ	
ㅁ	16						ㄴ									ㅁ	ㅁ	ㅁ	ㅁ	ㅇ ㅁ	ㅎ ㅁ	
ㅂ	17	ㄲ		ㅁ	ㄸ		ㄴ ㅁ	ㅁ	ㅃ		ㅆ			ㅍ		ㅂ	ㅂ	ㅂ	ㅂ	ㅇ ㅂ	ㅎ ㅂ	
ㅄ	18	ㄲ ㅂ	ㅂ	ㅁ	ㄸ ㅂ	ㅂ	ㄴ ㅁ	ㅁ	ㅃ ㅂ	ㅂ	ㅆ ㅂ	ㅂ		ㅍ	ㅂ	ㅆ ㅂ	ㅆ ㅂ	ㅆ ㅂ	ㅂ	ㅇ ㅂ	ㅎ ㅂ	
ㅅ	19	ㄲ ㄷ	ㄷ	ㄴ	ㄸ ㄷ	ㄷ	ㄷ	ㄴ	ㅃ ㄷ	ㄷ	ㅆ ㄷ	ㄷ		ㅌ	ㄷ	ㅅ	ㅅ	ㅅ	ㄷ	ㅇ ㄷ	ㅎ ㄷ	
ㅆ	20	ㄲ ㄷ	ㄷ	ㄴ	ㄸ ㄷ	ㄷ	ㄷ	ㄴ	ㅃ ㄷ	ㄷ	ㅆ ㄷ	ㄷ		ㅌ	ㄷ	ㅆ	ㅆ	ㅆ	ㄷ	ㅇ ㄷ	ㅎ ㄷ	
ㅇ	21						ㄴ															
ㅈ	22	ㄲ ㄷ	ㄷ	ㄴ	ㄸ ㄷ	ㄷ	ㄷ	ㄴ	ㅃ ㄷ	ㄷ	ㅆ ㄷ	ㄷ		ㅊ	ㄷ	ㅈ	ㅈ	ㅈ	ㄷ	ㅇ ㄷ	ㅎ ㄷ	
ㅊ	23	ㄲ ㄷ	ㄷ	ㄴ	ㄸ ㄷ	ㄷ	ㄷ	ㄴ	ㅃ ㄷ	ㄷ	ㅆ ㄷ	ㄷ		ㅌ	ㄷ	ㅊ	ㅊ	ㅊ	ㄷ	ㅇ ㄷ	ㅎ ㄷ	
ㅋ	24	ㄲ ㄱ	ㄱ	ㅇ	ㄸ ㄱ	ㄱ	ㄴ ㅇ	ㅇ	ㅃ ㄱ	ㄱ	ㅆ ㄱ	ㄱ		ㄱ	ㄱ	ㅋ	ㅋ	ㅋ	ㄱ	ㅇ ㄱ	ㅎ ㄱ	
ㅌ	25	ㄲ ㄷ	ㄷ	ㄴ	ㄸ ㄷ	ㄷ	ㄷ	ㄴ	ㅃ ㄷ	ㄷ	ㅆ ㄷ	ㄷ		ㅌ	ㄷ	ㅌ	ㅌ	ㅌ	ㄷ	ㅊ	ㅎ ㄷ	
ㅍ	26	ㄲ ㅂ	ㅂ	ㅁ	ㄸ ㅂ	ㅂ	ㄴ ㅂ	ㅁ	ㅃ ㅂ	ㅂ	ㅆ ㅂ	ㅂ		ㅍ	ㅂ	ㅍ	ㅍ	ㅍ	ㅂ	ㅇ ㅂ	ㅎ ㅂ	
ㅎ	27	ㅋ		ㄴ	ㅌ			ㄴ			ㅆ					ㅇ	ㅇ	ㅇ	ㄷ	ㅇ ㄷ		
어간받침ㄴㄵ	28	ㄲ ㄴ			ㄸ ㄴ						ㅆ ㄴ											
어간받침ㅁㄻ	29	ㄲ ㅁ			ㄸ ㅁ						ㅆ ㅁ											
어간받침ㄾ	30	ㄲ ㄹ			ㄸ ㄹ						ㅆ ㄹ											
사이시옷	31	ㄲ			ㄸ				ㅃ		ㅆ											ㄴ ㄴ

· 제2항 로마자 이외의 부호는 사용하지 않는다.

· 제3항 1음운 1기호의 표기를 원칙으로 한다.

여기서, 국어 국립 국어 연구원에서 1984년에 발표한 표기의 기본원칙과 다른 점은 '제2항 로마자 이외의 부호는 사용하지 않는다'이다. 이것은 기존에 [⌣]

나 [']와 같은 부호를 되도록 사용하지 않는다는 원칙에서 로마자 이외의 부호를 사용하지 않아 과거의 컴퓨터 자판에 부호가 없어 실용적이지 못하다는 것을 해결하게 되었다. <표 2>, <표 3>은 새롭게 발표된 로마자 표기법을 테이블 형태로 표현한 것이다.

<표 2> 모음 로마자 표기 테이블

모음	단모음	중모음
	ㅏ ㅓ ㅗ ㅜ ㅡ ㅣ ㅐ ㅔ ㅚ ㅟ	ㅑ ㅕ ㅛ ㅠ ㅒ ㅖ ㅘ ㅙ ㅝ ㅞ ㅢ
로마자	a eo o u eu i ae e oe wi	ya yeo yo yu yae ye wa wae wo we ui

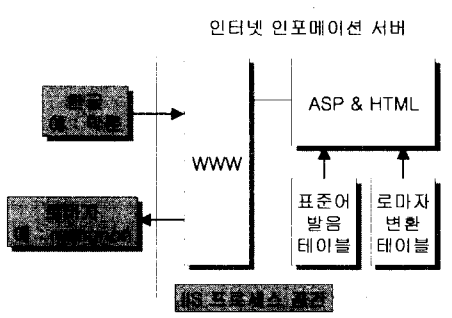
<표 3> 자음 로마자 표기 테이블

자음	기본자음	쌍자음
	ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅇ ㅈ ㅊ ㅋ ㅌ ㅍ ㅎ	ㄲ ㄸ ㅃ ㅆ ㅉ
로마자	g n d r m b s ng j ch k t p h	kk tt pp ss jj
붙임1)	'ㄱ, ㄷ, ㅂ'은 모음 앞에서 'g, d, b', 자음 앞이나 어말에서 'k, t, p'로 적성	
붙임2)	'르'은 모음 앞에서는 'r', 자음 앞이나 어말에서는 'l'로, '르르'은 'll'로 적는다.	

<표 3>에 표현되지 않은 복자음에 대한 것은 기본자음을 단순조합하여 사용한다. 앞서 작성된 내용이 형태소 분석을 통하여 구성된 표준 발음 테이블로 변경된 한글은 로마자 표기 테이블로 완전히 로마자로 변형된다.

4. 한국어 로마자 시스템의 디자인

이 시스템은 결국 (그림 2)와 같은 웹을 통하여 사용자에게 변환될 한글을 입력받아 ASP(Active Server Page)를 통하여 작성된 규칙을 적용하게 된다.



(그림 2) 한국어 로마자 변환기 구성도

여기서, 표준어 발음으로 구성된 문장 혹은 단어의 음절에 대해 초성, 중성, 종성으로 나누기 위해서는 MS Windows계열의 시스템에서는 확장완성형으로 입력된 음절을 유니코드로 변형한 뒤 조합형을 통해 초성, 중성, 종성으로 나누게 되는 과정을 먼저 실행한다[4]. MS Windows계열의 시스템에서는 유니코드 변형과 조합형 변형에 대한 Win-API(Window Application Program Interface) 함수를 가지고 있어 이를 이용하여 초성, 중성, 종성으로 나누게 된다.

본 시스템은 마이크로 소프트 Windows NT 4.0하의 IIS(Internet Information Server)으로 구성되며, 앞서 구성된 표준어 발음 테이블과 로마자 변환 테이블을 통하여 로마자로 Web 문서 형태로 나타내진다.

5. 결론

본 논문에서는 한국어 로마자 변환기 개발 프로젝트 중 빠른 성능 향상을 위해 표준어 발음 테이블 작성과 로마자 표기 테이블을 작성하여 인터넷을 통하여 정확한 한국어 로마자 변환기 개발을 위해 작성되었다. 더욱 효율적인 처리를 위해 테이블 표기의 간소화와 로마자 표기 서비스를 검색엔진에 포함하여 정보 검색 시스템에서 로마자로 표기된 정보를 검색하도록 하여 검색의 재현율을 높이는 연구가 필요하다. 또한, 특정 플래폼에 한정된 현 시스템을 다양한 플래폼으로 이식을 위한 구현이 필요하다.

참고문헌

[1] 국립 국어 연구원, 국어의 로마자 표기법(개정 시안), 1999

[2] 문교부, "표준어 규정", 문교부 고시 제88-2호, 제2부 표준어 발음법, 1988

[3] 정태충, "한국어의 로마자 표기법", 한국어 표기법 연구회, 1998

[4] 한국어 정보처리 연구소, C로 구현한 한글 코드 시스템 프로그래밍 가이드, 1999

[5] 임재걸외 2, 한국어 음가 생성 규칙 테이블 구현, 한국정보처리학회 추계학술발표논문집, 1999

[6] 임재걸외 2, WWW에서의 한국어 표준 음가 생성 시스템 구현, 한국정보과학회 추계학술발표논문집, 1999

[7] 이응백, 이익섭, 이병근, "국어 어문 규정집", 대한교과서 주식회사, 1988/1996