

정음 부호세트의 코드 표준안에 관한 연구

이강원*, 김우선*, 김철*

*용인대학교 컴퓨터정보학과

e-mail : kizmo@interspace.ne.kr

A Study on Standardization of the OPA Character set

Gang-Won Lee*, Woo-Sun Kim*, Chul Kim*

*Dept. of Computer Science & Information Processing

Yong-In University

요약

오늘날 우리가 사용하는 한글은 24자로 구성되어있고, 한국어만을 적기 위한 문자체계이다. 그러나 훈민정음은 28자소로 구성되어 있으며 한국어 이외에도 세계 모든 언어와 소리를 적을 수 있는 소리 기호 체계이다. 현재 한글은 인터넷이나 컴퓨터에서 사용이 어렵고, 자판 구도가 불합리하며, 남북 및 세계 표준이 아직 정해지지 않고 있으며 정렬방식이 불일치하고, 내부 처리 코드가 상이한 문제점을 가지고 있다. 기존의 한글처리로는 부적합한 것들을 종합적으로 해결하기 위하여 새로운 방식을 모색하고, 미래의 방향을 제시할 목적으로 정음부호(OPA : Ortho Phonic Alphabet)의 코드 제안에 따른 기존 KS C 5601과 KS C 5700과의 코드 변환시 문제점과 OPA 표준화를 위해 이 연구를 수행한다.

1. 서론

한글은 자소나 자모, 마디글자까지 표시할 수 있는 양면성과 탁월한 유연성을 발휘할 수 있는 문자로서, 현재 세계공용어로 쓰이는 영어는 말과 글이 일치하지 않아 학습하기에 매우 어려우며 자모 중심이어서 한자나 일본어처럼 마디글자는 구분되지 않아 음절한계가 모호한 특성이 있는 반면에 한글은 세계 공용 문자에 적합하다고 할 수 있다. 정음부호(OPA : Ortho Phonic Alphabet)는 국제 발음기호로 사용될 수 있기 때문에 사라져 가는 언어에 대한 기록수단으로 사용될 수 있다. 로마자나 국제음성기호로의 표기도 가능하지만 학습하기에 어렵고 불편하다. 따라서 국제정음기호는 새로운 기록수단을 제공할 수 있다.

본 논문은 다음과 같이 구성되어 있다. 2절에서는 부호화 문제로 인하여 통일된 내부처리코드의 필요성과 현재 사용되는 부호 체계(KS C 5601 완성/조합형, KS C 5700)에 대해 설명하고, 정음부호(OPA : Ortho Phonic Alphabet - 세계의 모든 말을 정음으로 적기 위한 음성기호 체계)의 개념에 대해 설명한다. 3절에서는 정음부호(OPA)의 자판 구도(안)을 알아본다. 4절에서는 정음부호(OPA) 7비트 N비트 구성방법을 5절에

서는 정음부호와 KS C 5601/부속서4 와 비교하고 6절에서는 정음부호와 KS C 5700의 비교를 통해서 기존 부호 체계의 문제점과 정음부호(OPA)의 타당성을 제시한다. 마지막으로 7절에서는 본 논문의 결론을 맺는다.

2. 본론

가. 부호화 문제

컴퓨터가 영문을 사용하는 서구권에서 개발되어 왔으므로 부호체계도 서구문자를 다루기에 적합하게 개발되어 왔다. 따라서 이제까지 컴퓨터에서 한글을 다루는 부분에 대해서 여러 가지 문제들이 제기되어 왔다. 현재 한글 부호는 2바이트를 사용하고 있으며 로마자의 ASCII는 7비트로 구성되어 있다. 그러나 한글 7비트는 KS C 5700 부속서 4에 규정하고 있으나 일대일 대응이 되지 못하여 한글의 코드 변환이 어렵게 되어있다. 한편 '모듬이 한글꼴' 모듬은 2바이트 완성형과 조합형으로 구성되어 있다. 일본 JIS를 본뜬 구조이다. 완성형이든 조합형이든 국제 규격을 만족하지 못하고 있다.

그 외 한글의 내부처리코드가 회사마다 또는 기종마다

4. 정음부호(OPA)의 7비트 N바이트 구성방법

한글 낱자 부호체계와는 다른 구조이고 자판에 규정된 타건과 1:1로 대응되게 하였다. (그림 2)의 OPA 자판배치(안)을 기준으로 하였다. 여기에서는 ‘뎀’과 ‘이음’이란 소리경계를 표시하기 위함이다. 아래 (표 1) 정음 문자의 7비트 부호틀(안)에서 보면 KS C 5601/부속서 4의 코드 1101101 위치의 ‘나’에 ‘뎀’을 코드 1101110 위치의 ‘내’에 ‘이음’을 정음기호 부호체계(안)에서는 대응시켜 놓았다.

예를 들면 ㄱ + 뎀 + ㅏ 는 ‘가’와 다르다. ‘ㄱ사다’는 ㄱ + 이음 + ㅏ + 이음 + ㅏ 처럼 구성된다.

‘N바이트 조합형’은 초성-중성-종성의 세 갈래로 음운을 나누는 글자는 아마 한글 이외에는 없을 것이기 때문이다. (죽은 글자로는 원나라의 파스파 문자가 3별식 음운 체계를 가졌었다.) ISO-2022를 따르면서, 한 바이트당 94 코드점을 쓸 수 있다. 이 가운데 2비트(32 코드점)는 초성/중성/종성/특수목적용 가리키는데 배정하면, 나머지 62 코드점을 쓸 수 있기 때문에, 중요한 옛 음운이나 한자/특수기호까지 포함시키는 것은 쉬운 일이다. (실제로 첫가끝 자모에 들어가 있는 상당수의 초성-중성 음운들은 중국 한자음을 음역할 때나 쓰는 ‘특수한’ 자모로써, 일반적인 자모로 인정하기가 힘들다.)

5. 정음부호와 KS C 5601/부속서4

(표 1) 정음 문자의 7비트 부호틀(안)에서 보면 KS C 5601/부속서 4에서는 정음부호의 ‘ㄴㄴ’, ‘ㄹㄹ’의 코드값이 정의되어있지 않다. 하지만 정음기호의 부호체계에서는 기존의 ‘ㄱ, ㄷ, ㅂ, ㅅ, ㅈ’와 동일하게 코드값을 부여하고 있다.

비트	한글 낱자 부호체계 (KSC 5601/부속서 4)				정음기호의 부호체계 (안)				로마자 부호체계 (ISO 546)			
	1	1	1	1	1	1	1	1	1	1	1	1
비	0	0	1	1	0	0	1	1	0	0	1	1
비	0	1	0	1	0	1	0	1	0	1	0	1
	4	5	6	7	4	5	6	7	4	5	6	7
0	채움	라			ㅏ				P		p	
1	ㄱ	ㄴ			ㄱ	ㄴ	ㅌ		A	Q	a	q
2	ㄱ	ㅂ	ㅅ		ㄱ	ㅂ	ㅅ		B	R	b	r
3	ㅈ	ㅊ	ㅌ		ㅈ	ㅊ	ㅌ		C	S	c	s
4	ㄴ	ㅂ	ㅅ		ㅈ	ㅊ	ㅌ		D	T	d	t
5	나	사	ㅌ		ㅌ	ㅌ	ㅌ		E	U	e	u
6	내	싸	ㅌ		ㅌ	ㅌ	ㅌ		F	V	f	v
7	ㄷ	ㅌ	ㅌ		ㅌ	ㅌ	ㅌ		G	W	g	w
8	ㄷ	ㅌ			ㅌ	ㅌ	ㅌ		H	X	h	x
9	ㄹ	ㅌ			ㅌ	ㅌ	ㅌ		I	Y	i	y
10	ㄹ	ㅌ	ㅌ		ㅌ	ㅌ	ㅌ		J	Z	j	z
11	ㅌ	ㅌ	ㅌ		ㅌ	ㅌ	ㅌ		K		k	
12	ㅌ	ㅌ	ㅌ		ㅌ	ㅌ	ㅌ		L		l	
13	ㅌ	ㅌ	ㅌ		ㅌ	ㅌ	ㅌ		M		m	
14	ㅌ	ㅌ	ㅌ		ㅌ	ㅌ	ㅌ		N		n	
15	ㅌ	ㅌ	ㅌ		ㅌ	ㅌ	ㅌ		O		o	

(표 1) 정음 문자의 7비트 부호틀(안)

(그림2)의 자판배치 제안에서처럼 정음부호에 있는 ‘ㄴㄴ’, ‘ㄹㄹ’을 기존의 자판을 이용하여 입력하려면 두 번 타건해야 하기 때문에 2바이트가 입력되어야 한다. 이것은 사용자 하여금 혼동을 줄 수 있으므로 동일한 방식으로 하나의 자리를 차지하게 자판을 구성할 필요가 있다.

6. 정음부호와 KS C 5700

KS C 5700에서는 UCS(Universal Multiple - Octet Coded Character Set)를 규정하여 제정하였는데, 그 부호계에는 자모형과 240자 자소형, 그리고 11172자 완성음절형이 포함되어 있다. 이것은 결정적으로 표준화의 근본 취지에 어긋난다. 까닭은 부호계의 단일화 보다 오히려 부호계의 다중화로 유도하는 꼴이기 때문이다. 특히 완성형은 자소 정보를 제공하지 못하며 이것은 한글문자의 음소문자 특성을 없애는 것이며, 옛 한글 표현을 배제하고 있다.

ISO/IEC 10646-1 : 1993 ED

TABLE 29 - Row 11 : HANGUL JAMO

	110	111	112	113	114	115	116	117
0	ㄱ	ㄴ	ㄷ	ㄹ	ㅌ	ㅌ	ㅌ	ㅌ
1	ㄱ	ㄴ	ㄷ	ㄹ	ㅌ	ㅌ	ㅌ	ㅌ
2	ㄴ	ㅂ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ
3	ㄴ	ㅂ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ
4	ㄴ	ㅂ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ
5	ㄴ	ㅂ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ
6	ㄴ	ㅂ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ
7	ㄴ	ㅂ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ
8	ㅂ	ㄴ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ
9	ㅂ	ㄴ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ
A	ㅂ	ㄴ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ
B	ㅂ	ㄴ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ
C	ㅂ	ㄴ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ
D	ㅂ	ㄴ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ
E	ㅂ	ㄴ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ
F	ㅂ	ㄴ	ㅅ	ㅈ	ㅊ	ㅌ	ㅌ	ㅌ

(표 2) ISO 10646-1 한글자모 TABLE 29

정음부호(OPA)는 KS C 5700의 한글 자모 테이블의 완성형 부호를 모두 조합할 수 있다.(표 4) 기존의 한글 자모로 되어 있는 자소형 부호계를 유일한 부호계로 두고 조합 방식을 사용하는 방법도 있겠지만 효율적인 활용을 위해서는 한글 완성형 부호계와 호환용 한글 자모 부호계를 없애고 공간을 반납하고 정음부호(OPA)의 조합형식을 따라야 한다고 본다.

ISO/IEC 10646-1: 1993 (E)

TABLE 30 - Row 11 : HANGUL JAMO

	118	119	11A	11B	11C	11D	11E	11F
0	세	새	수	시	세	르	르	스
1	새	새	이	리	리	리	리	리
2	소	소	이	리	리	리	리	리
3	수	수	리	리	리	리	리	리
4	샤	샤	리	리	리	리	리	리
5	배	배	리	리	리	리	리	리
6	피	피	리	리	리	리	리	리
7	포	포	리	리	리	리	리	리
8	피	피	리	리	리	리	리	리
9	가	가	리	리	리	리	리	리
A	개	개	리	리	리	리	리	리
B	고	고	리	리	리	리	리	리
C	계	계	리	리	리	리	리	리
D	구	구	리	리	리	리	리	리
E	파	파	리	리	리	리	리	리
F	피	피	리	리	리	리	리	리

G=00
F=00

(표 3) ISO 10646-1 한글자모 TABLE 30

ㄱ	1100	ㅋ	1163	ㅠ	1101
ㄴ	1102	ㆁ	1165	ㅌ	1103
ㄷ	1103	ㆁ	1167	ㅍ	1108
ㄹ	1105	ㄴ	1169	ㅍ	110A
ㅁ	1106	ㄴ	116D	ㅍ	110D
ㅂ	1107	ㅌ	116E	ㄴ	1114
ㅅ	1109	ㅠ	1172	ㅎ	1158
ㅇ	110B	ㅡ	1173	르	1119
ㅈ	110C	ㅣ	1175	△	1141
ㅊ	110E	ㅑ	1166	○	1159
ㅋ	110F	ㅒ	1162	○	114C
ㅌ	1110			•	119E
ㅍ	1111				
ㅎ	1112				
ㅏ	1161				

(표 4) 정음부호의 ISO 10646-1 한글자모 TABLE 대치 코드표

부호들(안)(표 1)을 KS C 5601/부속서4과 비교해 보면 기존 문자 입력시 2바이트가 입력되어야 하는 몇몇 문제점을 정음부호(OPA)를 통해서 해결하고 있다. 또한 (그림 1) KS C 5700의 BMP 구조에서 보면 한글 자모 뒷부분에 앞으로 표준을 위한 예비 영역이 충분히 있기 때문에 표준으로 정음부호(OPA)를 사용함에 있어서 어려움이 없다고 본다.

정음부호(OPA)는 훈민정음의 원리에 따라 음절자를 구성하므로 모든 부호계를 포함한다. 또한 문자집합이 작으므로 국내의 규격을 만족하고, 모든 한글 부호계의 전체집합이므로 부호간 변환시 호환성이 완벽하게 보장된다. 또한 정음부호는 낱자소 조합으로 음절을 구성하기 때문에 현대 한글은 물론 옛 한글을 자연스럽게 지원한다.

훈민정음 원리에 따르는 정음부호를 기존의 한글처리로는 부적합한 것들을 종합적으로 해결하고, 국어정보처리와 한글 문자 정보처리의 모든 응용 분야에서 만족할 수 있도록 하기 위해서는 표준화가 필요하다.

참고문헌

- [1] 박성득, "국제정음기호표준 제정의 타당성 연구", 한국전산원, 1999.12
- [2] 한국표준협회, 정보교환용 부호 (한글 및 한자) KS C 5601 -1987 완성형, 1987
- [3] 한국표준협회, 정보교환용 부호 (한글 및 한자) KS C 5601 -1992 조합형, 1992
- [4] 한국표준협회, 국제 문자 부호계 (UCS : Universal Multiple-Octet Coded Character Set) KS C 5700 - 1995, 1995

7. 결론

정음부호(OPA)의 자판배치(안)(그림 2)과 OPA 7비트