

국제 컴퓨터 망을 이용한 한글 정보의 전송

○
김 홍운, 김 영현, 이 균하
인하대학교 전자계산학과

Transmission of Hanguel Information Using International Computer Network

요 약

본 논문에서는 국제 컴퓨터 망을 통하여 한글 정보를 전송할 수 있는 한 방안을 제안하였다. 한글 문서를 서구 문자로 바꾸어 서구문만의 전송이 가능한 컴퓨터 망에서도 전송이 가능하도록 하였고 한글과 영문이 혼용된 문서를 서구 문자로 전자하는데 한/영 구분기호를 사용하였으며 전자된 한글이 포함되어 있음을 표현하는 통신문 서식을 만들어 사용하였다. 또한 한글을 서구문자로 전자하고 복원하는 소프트웨어를 작성하였다.

1. 서 론

컴퓨터가 대량으로 보급됨에 따라 사회 각 분야에서 정보화가 급속히 진행되어 정보 자원의 교환을 위한 컴퓨터 망 시스템이 발전하게 되었다. 컴퓨터 망에 연결된 각종의 컴퓨터들 사이에 정보 교환이 이루어지려면 정보를 표현하는 방식이 표준화 되어야 한다. 따라서 컴퓨터를 이용한 정보 자원을 원활히 교환하기 위하여 ISO(International Standard Organization)에서는 세계 각 문자를 고려한 표준 부호를 권장하고 있다[3].

그러나 한글의 경우 현재 한글 표준 부호가 모든 컴퓨터 망에 100% 보급되어있지 않았을 뿐 더러 표준 부호 사용을 권장하기 이전에 구축되어 사용되고 있는 망도 존재하고 있으며 16 비트 이상을 필요로 하는 한글, 한자와 같은 문자들의 전송을 고려하지 않은 망도 존재하고 있다. 이와 같이 서구 문화권에 적합하도록 구축된 컴퓨터 통신망을 통한 정보 교환은 서구문 표준 부호로 작성된 정보는 전송이 가능하지만 한글로 작성된 정보를 그대로 전송하는데에는 문제가 있다.

본 논문에서는 서구문만의 전송이 가능한 컴퓨터 망에서도 한글 정보를 전송할 수 있는 한 방안을 제시하고자 한다.

2. 통신망과 한글 부호에 대한 분석

앞에서 설명한 바와같이 서구문 전용의 컴퓨터 망을 이용하여 한글 정보를 전송 하는데는 다음과 같은 문제가 야기될 수 있다. 현재 세계 각 지역에는 이미 서구문 전용의 통신망이 구축되어 있으므로 이를 이용하여 한글 정보를 전송할 수 있다면 한글 정보화 영역이 손쉽게 확대된다. 이러한 접근 방법은 예를 들어 컴퓨터가 보급되기 이전에 이미 전 세계적으로 망을 구축하여 사용하고 있는 전화망을 이용하여 컴퓨터 통신을 시도하는 경우와 비유할 수 있다. 즉 그림 1과 같은 전화망은 300 - 3400 Hz 범위의 주파수만을 전송하는 통신망에 적합 하므로 컴퓨터 신호를 변.복조하여 전 세계적인 컴퓨터 통신망을 구축할 수 있는 방안과 같다고 볼 수 있다.

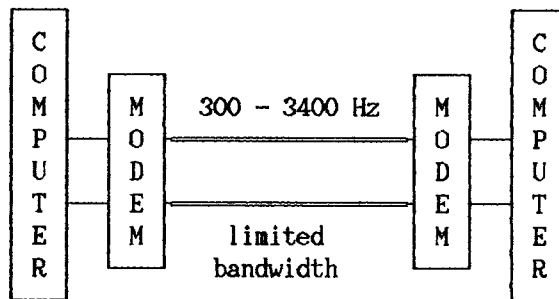


그림 1. 전화 통신망

서구문 전용의 컴퓨터 통신망은 7 - 8 비트로 충분히 표시되는 서구문 전송을 염두에 두고 설계하였기 때문에, 한글 한자와 같이 16비트 이상으로 이루어진 정보를 전송 하는데는 그림 2 와 같이 허용되는 부호의 폭(code width)이 상대적으로 작아진다.

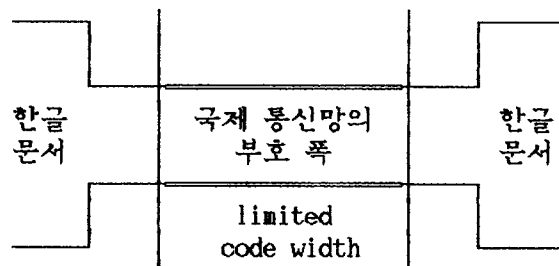


그림 2. 한글 문서 전송시 부호의 제약

국제 통신망 중에는 서구문자, 숫자, 특수문자들 이외의 문자 일부들을 통신망 제어문자로 사용하기 때문에 이러한 문자들을 전송할 경우 통신망의 장애가 발생할 수 있어 전송이 곤란하다. 또한 7 비트로 문자를 전송하는 망에서는 전송시 문자의 8 번째 비트를 잘라내므로 문자의 의미를 상실하게 된다. 이러한 환경에서 허용되는

부호의 폭이 서구문에 적합하도록 설계되어 서구문으로 작성된 정보는 전송되지만 한글로 작성된 정보를 전송할 경우 사용할 수 있는 한글 부호의 폭이 좁아진다. 서구문 화면에서도 이치(binary) 화일을 전송할 때는 우리와 같은 문제에 당면하게 되어 전송이 보장되지 않은 문자는 부호 변환후 변환 표시 문자와 함께 전송하는 등의 방식으로 이 문제를 해결하고 있다. 한글 문서의 전송도 이러한 방식으로 처리할 수 있으나 한글 정보를 이치 화일로 취급해야 하는 문제가 있다. 한글 정보의 전송은 문서 전송시 뿐만 아니라 문자 전송시에도 보장되어야 하므로 이치 화일 전송 방식 만으로는 처리할 수 없다.

서구문 부호의 폭에 포함되는 N 바이트형 한글 부호와 7 비트 완성형 한글 부호는 이러한 환경에 적합한 좋은 예이지만 한글 단말기가 없을 경우 가독성이 떨어진다. 가독성이 높은 부호로는 예전 부터 사용 되었던 "(구)문교부안"의 로마자 표기법이 좋은 방법일 수 있으나 전자된 문서를 복원시 이중 해석이 될 수 있는 모호성이 발생한다[2]. 현재 논의 중에 있는 한글 로마자 표기법안[1,4]은 모호성 문제를 해결 하였지만 ISO 표준안으로 채택 되어있지 않은 상태이다.

그러나 이 표기법에서도 문자 단위의 전자에 머무르고 있어 아직 해결 해야할 점이 있다. 특히 한글과 영문이 혼재된 정보를 전자하는 방안은 포함되어 있지 않다. 또한 전자된 문서를 전송하기 위한 통신문 서식도 없는 실정이다. 본 논문에서는 이러한 점을 고려하고 국제 컴퓨터 통신망에서 한글 정보를 전송할 수 있는 한 방안을 제시하고자 한다.

3. 한글 부호의 로마자 표기

로마자 표기법은 전사법 과 전자법이 존재하며 전사법이 실제 발음에 충실하지만 복원시 본래의 철자를 복원하기 어려운 단점이 있다. 반면에 전자법은 철자법을 그대로 반영하기 때문에 원래의 철자를 기계적으로 복원하기 쉬운 장점이 있지만 실제 발음에 충실하지 못하다는 단점이 있다.

국제 통신망에 사용될 한글 로마자 표기법은 기계화가 가능 하여야 하므로 전자법을 사용해야 한다. 국제 컴퓨터 통신망에 사용할 한글 로마자 표기법(전자법)으로 다음과 같은 세가지 방식을 고려해볼 수 있다.

(1) KBS-89 (표 1 참조)

공업진흥청 한글 로마자 표기법 회의[4]에서 모호성 배제를 위하여 발표한 표기법으로 예1)과 같은 경우는 원문으로 복원시 /가짜/와 /갓자/로 이중 해석이 가능하므로 예2)와 같이 음절 분리용 delimiter(')를 두어 로마자 전자시 /가짜/는 /GA'JJA/로 표기하고 /갓자/는 /GAJJA/로 표기하는 방식.

예1) GAJJA => 가짜, 갖자

예2) 가짜 => GA'JJA

갖자 => GAJJA

한글 자음	로마자	한글 모음	로마자
ㄱ	G	ㅏ	A
ㄴ	N	ㅑ	YA
ㄷ	D	ㅓ	EO
ㄹ	R,L	ㅕ	YEO
ㅁ	M	ㅗ	O
ㅂ	B	ㅛ	YO
ㅅ	S	ㅜ	U
ㅇ	' ,NG	ㅠ	YU
ㅈ	J	ㅡ	EU
ㅊ	C	ㅣ	I
ㅋ	K	ㅐ	AE
ㅌ	T	ㅒ	YAE
ㅍ	P	ㅖ	E
ㅎ	H	ㅙ	YE
ㅇ	GG	ㅛ	OE
ㅌ	DD	ㅜ	WI
ㅍ	BB	ㅠ	YI
ㅈ	SS	ㅑ	WA
ㅊ	JJ	ㅓ	WEO
		ㅕ	WAE
		ㅗ	WE

표 1. 공업진흥청에서 발표한 한글 로마자 전자표

(2) INHA-90

KBS-89 방식의 표기법을 포함하고 있는 방식으로서 KBS-89 방식에서 수용하지 않았던 ①-④의 경우 delimiter(')를 생략하여 경제성을 더욱 향상시킨 방식.

① 받침이 1개인 음절 + 초성 쌍자음(앞음절의 받침으로 이동이 불가능한 경우)

예) 참깨 => CAMGGAE

굴뚝 => GULDDUG

식빵 => SIGBBANG

흰쌀 => HYINSSAL

살짝 => SALJJAG

② 특수 문자 또는 숫자 + 초성(ㅇ)

예) (1)에서 => (1)ESE0

38의 => 38YI

③ 숫자 + 초성 쌍자음

예) 153쪽 => 153JJOG

④ 발침이 2개인 음절 + 초성 쌍자음

예) 읍락 => OLGSSAG

(3) COMPACT-1 (표 2 참조)

음가와 관계 없이 한글 자모와 로마자를 1:1로 대응시켜 정보 전송 효율을 향상시킨 표기법. 국제 컴퓨터 통신망에서 영문 소문자 전송도 가능하므로 이 표기법을 사용하면 가독성은 거의 불가능 하지만 한글 정보 전송 효율은 높아진다.

초성	로마자	중성	로마자	종성	로마자
ㄱ	A	ㅏ	T	ㄱ	A
ㄴ	B	ㅑ	U	ㄴ	B
ㄷ	C	ㅓ	V	ㄷ	C
ㄹ	D	ㅕ	W	ㄹ	D
ㄴ	E	ㅗ	X	ㅁ	E
ㅁ	F	ㅛ	Y	ㅂ	F
ㅂ	G	ㅜ	Z	ㅅ	G
ㅅ	H	ㅠ	a	ㅇ	H
ㅇ	I	ㅡ	b	ㅈ	I
ㅈ	J	ㅣ	c	ㅊ	J
ㅊ	K	ㅊ	d	ㅋ	K
ㅋ	L	ㅌ	e	ㆁ	L
ㆁ	M	ㅍ	f	ㄷ	M
ㄷ	N	ㅑ	g	ㅌ	N
ㄹ	O	ㅓ	h	ㅍ	O
ㅁ	P	ㅕ	i	ㅑ	P
ㅂ	Q	ㅗ	j	ㅛ	Q
ㅅ	R	ㅜ	k	ㅜ	R
ㅇ	S	ㅠ	l	ㅠ	S
		ㅡ	m	ㅡ	
		ㅣ	n	ㅣ	

표 2. 전송 효율을 향상시킨 로마자 전자표

4. 통신문 서식 및 변환 소프트웨어

앞에서 제안한 방식으로 한글이 포함된 문서를 컴퓨터 망을 통하여 전송하고자 하려면 한글을 로마자로 전자하여야 한다. 이러한 작업은 수작업으로 할 수 있지만 한글 사용이 가능한 컴퓨터에서는 변환 소프트웨어를 작성하여 기계적으로 해결할 수 있다. 그러나, 로마자로 전자된 문서를 다시 원문으로 복원할 경우 한글 만으로 작성된 문서는 복원이 비교적 쉽게 되지만 한글과 영문이 혼재하는 문서는 다음과 같은 문제가 따른다.

첫째, 로마자로 전자된 문서의 내용이 한글로만 작성된 것인지 한글과 영문이 혼용된 것인지를 식별하기 어렵고 둘째, 앞에서 논의한 바와 같이 한글을 로마자로 전자시키는 방식은 여러가지가 존재하고 있는데 문서 복원시 어떤 방식을 적용해야 하는지 식별하기가 어려울 뿐더러 셋째, 전자된 문서에서 한글이 로마자로 전자된 내용과 영문이 함께 존재하므로 이 것을 구별하여 파악하기가 어렵다.

본 논문에서는 이러한 문제점을 해결하기 위하여 첫째, 문서에 한글이 로마자로 전자되어 있다는 의미를 나타내는 설명문과 로마자 전자 방식에 관한 정보를 그림 3과 같이 문서의 머릿부분(header)에 기록하도록 하는 방법을 제안하고자 한다. 그러나, 변환 소프트웨어 사용시 이 머릿부분은 원문으로 복원시키는데 포함되지 않도록 하여야 한다. 둘째, 문서에 한글과 영문이 혼재한 경우 이를 로마자로 전자시 한글에서 영문으로, 영문에서 한글로 바뀌는 위치에 기호를 표시 하도록하고 그 기호를 머릿부분에 포함되도록 하여 원문으로 복원시 쉽게 식별이 가능하도록 하는 기호를 사용할 것을 제안한다. 현재 논의되고 있는 한글의 로마자 표기법을 표준화하는 작업에서 이에 대한 논의가 있어야 할 것으로 본다.

또한 이 로마자 표기법이 국제 표준화 기구에서 확정될 경우를 대비하여 변환 소프트웨어의 구조는 이러한 표준안도 쉽게 수용이 가능하도록 기호와 전자표를 수정하기 쉬운 형식으로 작성하여야 한다.

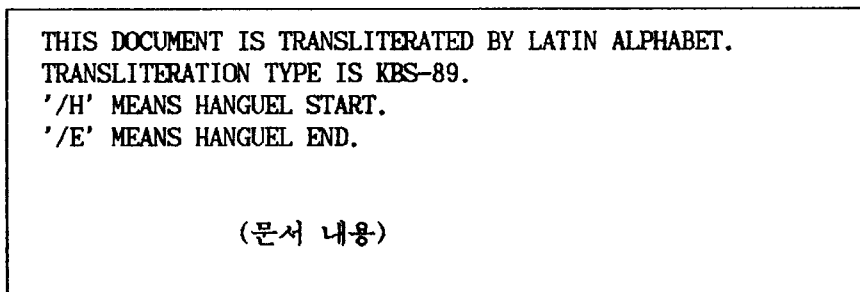


그림 3. 문서의 머릿부분

5. 실험 및 고찰

본 논문에서는 변환 소프트웨어를 PASCAL 언어로 구현하였고, IBM PC 호환기종에서 작성된 한글 문서를 로마자로 전자하여 Data General사의 MV/10000 시스템으로 전송한 다음 세계 학술 연구 정보 망인 BITNET(Because It's Time NETwork)[5]을 통하여 캐나다의 SIMON FRASER 대학과 실험 하였다. 실험을 통하여 발생한 문제점은 다음과 같다.

첫째, 외국에서 한글 정보를 컴퓨터에 입력 시킬때 자판에 한글 자모가 표현되어있지 않아 매우 불편하다는 점을 호소하였다. 이 문제는 미국 현지 교포들이 애용하고 있는 한글 자모 스티커로 해소 하였다.

둘째, 국제간 통신망을 이용한 문서 전송은 한 줄에 표현할 수 있는 문자 수를 80자 이내로 제한하는 망도 존재 하므로 한 줄에 80자 이내로 한글 정보를 작성 하였다 하더라도 로마자로 전자한 후에는 80자가 초과될 수 있기 때문에 문제가 발생하였다. 이 문제는 전자된 문서에서 한 줄의 문자를 80자 이내로 표현하려면 문맥의 의미가 손상 되지 않도록 줄을 분리해야 하는데 80자 이내에서 줄을 분리하고 분리된 줄의 마지막에 다음줄로 연결됨을 의미하는 표시로 연결 기호를 붙여 해결 하였다. 따라서 본 논문에서는 연결 기호로 '/C'를 사용하여 원문에서의 한 줄의 종료와 구별 되도록 하였다. 이 연결 기호를 또한 문서의 머릿부분에 포함되도록 하였다.

6. 결 론

서구문 전용의 국제 컴퓨터 통신망에서 한글 정보를 전송할 수 있는 방법들을 고찰하였다. 한글과 영문이 혼재된 문서를 로마자로 표기하여 전송하는 경우 한글과 영문이 전환되는 지점을 구분할 수 있는 기호가 필요함을 지적 하였고, 이러한 기호와 전송된 문서에 한글 정보가 로마자로 표기 되었다는 의미의 설명문등을 포함하는 통신문 서식에 대한 표준화의 필요성을 느꼈다. 국제간 컴퓨터 통신 망을 이용한 한글 정보 전송 실험에서 얻은 결론은 본 논문에서 제안한 방법을 이용하게될 경우 한글 정보화의 영역이 확대될 것으로 기대 된다.

참 고 문 헌

1. 박동순외, 기계화를 위한 한글의 로마자 표기법 소프트웨어 개발, 한국표준연구소, 1990.
2. 송기중외, 기계화를 위한 한글의 로마자 표기법, 한국정신문화연구원, 1986.
3. 이종학, 박규상, "정보처리용 문자코드 및 한글코드 표준화에 관한 고찰", 전기통신연구 제4권 제3호, 1990, pp 80-71.
4. 한글로마자 표기법 회의, 공업진흥청, 1989.8.
5. B. Plattner, H. Lubich, "Electronic Mail Systems and Protocols Overview and Case Study", IFIP WG 6.5 Conference on Message Handling Systems and Distributed Applications, Oct., 1988.