

한국형 텔리텍스 정보교환용 부호 및 문서 관련 규격작성을 위한 제언

洪范基 · 李東明 · 李泰勳 / 電子裝置研究室

I. 서론

텔리텍스는 공중통신망을 이용하여 신속정확하게 문서를 송수신하는 국제적인 서어비스라 할 수 있다.¹⁵⁾ CCITT에서 이미 1980년에 그 규격을 발표한 바 있고, 당 연구소에서는 한국형 텔리텍스의 규격안 작성을 위해 1982년부터 3개년 project로 연구활동을 수행하고 있다. 그중에서 특히 구미와는 상이한 문자인 한글, 한자의 부호화 및 그 표준화는 시급한 과제가 아닐 수 없다.

본고에서는 먼저 한글 coding scheme, 한자수용문제, code extension에 관하여 기술하고 page format 및 문자 열과 행의 pitch에 관한 방안을 제시하고자 한다.

II. 문자집합과 부호화표현

1. 기본방침

텔리텍스 서어비스는 국내는 물론 국제 통신이 가능해야 하므로 구미문자 및 그 부호화에

관하여 권고하고 있는 CCITT T.61과 호환성을 가져야 한다.¹³⁾ 그리고 우리나라에서의 텔리텍스 서어비스는 T.61에서 정의하고 있는 기본문자 set 이외에도 한글처리가 가능해야하므로 한글 set 및 부호화는 가능한한 국내의 규격을 따르도록 고려한다.

한글문자 set의 수용방법으로서 여러문자 set을 사용할 경우에 확장방법인 ISO-2022를 준함으로써 미래에 어떤 Telematic service 들과의 interworking시에도 확장성을 보장한다.

2. 한글기본 부호체계

한국형 텔리텍스의 기본부호체계로 7단위, 혹은 8단위 부호를 들수 있는데 KS-C-5601 '정보교환용부호'는 양자가 모두 정의되어 있다.¹⁶⁾

CCITT T.61에는 기본 텔리텍스 서어비스를 위해 8단위 부호의 사용을 권고하고 있으며 여러 종류의 문자집합을 사용할 때 8단위 부호를 채택하면 G-set이 2개 마련이 되므로 지시(Designate) 및 호출(Invoke)의 빈도수가 7단위 부호체계 보다 줄어들게 된다.¹⁰⁾ 이와같은 이유로 한국형 텔리텍스의 기본 부호체계는 8단위가 적당할 것으로 생각된다.¹²⁾

기본 부호체계에 한글을 수용하는 방안으로서 자소를 각각 분리하여 code를 부여하는 단수 byte 부호방식과, 한글을 마치 한자와 같이 문자단위로 code를 부여하는 복수 byte 방식이 있다.^{16),17)}

단수 byte 그래픽 캐릭터 집합에서는 캐릭터를 최대 96개까지 지정할 수 있고, 글자구성이 자소합성 즉, 초성, 중성, 종성에 의해 모든 글자의 형성이 가능한 한글은 전체 자소수가 96개를 넘지 않으므로 한글수용이 가능하다. 한국공업규격 KSC 5601의 한글 자동용 부호표를 이용하여 한글 부호를 규정하고 있다.¹⁶⁾

복수 byte 그래픽 캐릭터 집합용에서는 최대 94×94(행×열)까지 캐릭터 수용이 가능하며 따라서 G-Set에 한글부호를 지정할 수 있다.

이와 관련된 규격이 KS-C-5619가 있는데 여기에서는 사용빈도에 의해 한글 1316자, 한자 1692문자를 규정하고 있다.¹⁷⁾

두 방식의 장단점을 비교하면 <표 1>과 같다.

| 단수 Byte부호 | 복수 Byte부호 |
|---|--|
| 1. 대부분의 System이 채택 2. 한자수용의 어려움 3. 한글 처리용이 4. 한글 1자당 평균 2.5 Byte 소요 ¹³⁾ | 1. 국내 적용경험 부족 2. 한자수용가능 - 사용빈도에 의해 ¹⁷⁾ 한글 1316자 한자 1692자 3. 제한된 글자만 가능 - Table에 없는 글자 처리 고려 a. 여러 Leve의 Set 준비 b. DRCS 이용 4. Pattern과의 Mapping Algorithm이 필요 5. Memory 낭비 - 24×24 인쇄시 한글패턴 약 100K 소요 |

<표 1> 단수 Byte, 복수 Byte 부호의 장단점

<표 1>과 같이 복수 byte 사용시에는 memory 문제, 구현경험 부족문제, 외자처리 문제 등 해결해야 할 과제가 많으나, 단수 byte 부호를 사용할 경우 큰 무리가 없는 것으로 판단된다.

3. 한자수용문제

1945년 해방이후 한글전용의 역사가 꾸준히

진행되어 왔으나 1975년부터 다시 한자 병용이 시작되었다. 학교, 회사, 특히 관공서, 언론기관 등에서는 여전히 한글 한자 혼용으로 문서를 작성하는 것이 그 실정이다.

한국에 텔리텍스가 1986년부터 상용화가 시작되어서 1990-2000년대가 사용절정기라 볼때 그 시기의 사용자 및 관리자의 연령층인 30대 내지 40대는 한글세대이므로 한자사용은 급격히 줄어들 것으로 예측되나 완전한 한글전용은 어려운 것으로 보인다. 따라서 한글set는 필수 그래픽 캐릭터 set으로 정의하고 한자를 optional set으로 채택하는 것이 바람직 할 것으로 판단된다.

3항에서 설명한 바와 같이 단수 byte 체계에서 한자수용은 어려우며, 복수 byte로 지정해야 하는데 이때 선정되지 않은 한자의 정보교환은 불가능해진다.¹⁷⁾ 이의 해결책으로 <그림 1>과 같이 선정된 한자 캐릭터 집합 이외에 하나 또는 둘 이상의 set를 더 마련하여 부호화 하는 방법을 들 수 있다.

| 행 \ 열 | 1 | ----- | 94 | |
|---------------|-----------|-------|----|--------------|
| 1 : 15 | | | | First Level |
| 16 : 29 | 한글 1316 자 | | | |
| 30 : 47 | 한자 1692 자 | | | |
| 48 : 61 | 한글 1316 자 | | | Second Level |
| 62 : 79 | 한자 1692 자 | | | |
| ⋮ | | | | |

<그림 1> 정보교환용 한자부호계

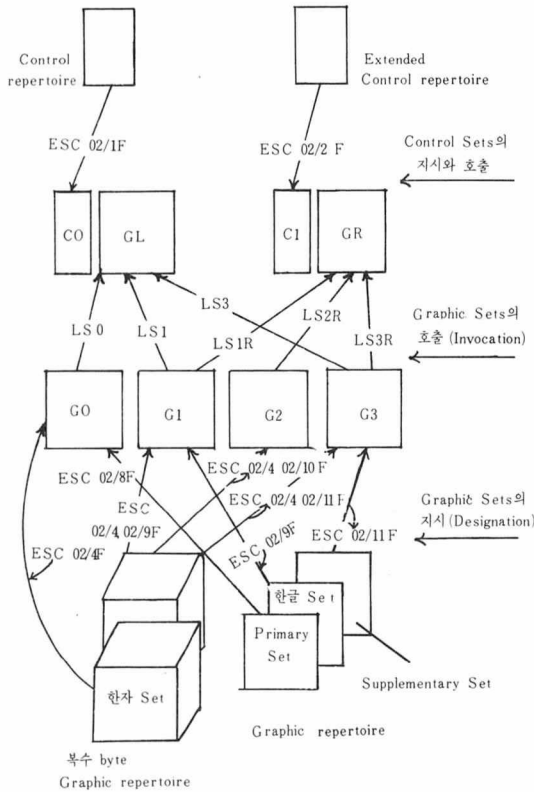
4. 부호확장

앞에서 언급한 바와 같이 여러 종류의 문자 set를 사용할 경우 ISO-2022에 따라서 부호확장이 가능하다.

부호확장은 <그림 1>과 같이 지시(Designate) 및 호출(Invoke)에 의해 이루어지는데 '지시'란 미리 준비된 복수 개의 캐릭터 집합속에서 정해

진 escape sequence 방법에 의해서 특정한 캐릭터 집합을 선택하는 것을 말하며, '호출'은 지시된 캐릭터 집합을 표현가능한 형태로 만드는 것을 말한다. Escape sequence는 부호 확장 순서에 있어서 사용하는 bit열로 2개 이상의 bit 조합으로 구성이 되며, 최초의 bit 조합은 ESC 캐릭터이고 n개 (n=0,1,2,...)의 중간 캐릭터 (I) 그리고 1개의 종단 캐릭터 (F)의 조합으로 구성되어진다.^{[8],[10]}

구미문자를 위한 primary set, supplementary set는 <그림 2>에서 G0, G2에 지시되고 통상 GL, GR에 호출된다.^[11] 한글 set는 G1, G3에 지시될 수 있으나 G1에 지시되고 난 후 GL 호출시 켄달 제어기능 LSI (Locking Shift One)는 실제 code가 다른 제어기능 보다 1 byte적은 2 byte로 구성되므로 한글 set가 GR에 호출되는 것으로 가정하면 G3에 지시하는 것이 바람직하다. G3에 지시할 때의 제어코드는 ESC 02/11F인데 final character는 ISO 2375에서 정한 절차에 따라 등록을 한다.^[11]



<그림 2> 8Bit 부호계에서의 부호확장

Ⅲ. 용지의 크기와 줄 간격 및 문자 간격

텔리텍스 터미널은 한글이나 영문 등의 부호화 문자로 구성된 문서를 상대편 터미널과 송, 수신하는 시스템으로서, 이때 송신측에서 보낸 문서와 수신측에서 받은 문서는 그 내용과 형태가 같아야 한다는 것이 기본 특성 중의 하나이다.^[2]

이렇게 되기위하여 텔리텍스 터미널에서는 서로 약속된 일정한 페이지 형식 (Page format)에 따라 문서를 작성하거나 통신이 이루어져야 할 것이다. 여기에 대하여 CCITT T. 60과 T. 62에서는 이러한 페이지 형식을 결정하는 요인인, 용지의 크기와 줄 간격 및 문자 간격 등에 대해 권고하고 있다.^{[2],[4]}

한글을 수용하는 한국형 텔리텍스에서 이것에 관한 규격이 필요하며, 이러한 규격은 현재 국내에서 많이 보급되어 사용되고 있는 워드 프로세서, 텔렉스 및 타자기 등 OA사무기기의 특성을 충분히 고려하여 제정되어야 할 것이다.

1. 페이지 형식 (Page Format)

텔리텍스 터미널에서의 페이지 형식이란 용지의 크기와 방향을 말하는 것으로 여기에 따라 서류 한 페이지의 가로 한계 (가로 문자 수)와 세로 한계 (줄 수)가 정해진다. 따라서 페이지 형식을 정할 때는, 미래의 국제통신시에 외국 텔리텍스 시스템과의 호환성을 고려해서 국제적으로 이미 규격화 되어있는 페이지 형식과, 현재 국내의 관공서나 기업체에서 문서로 사용하고 있는 페이지 형식 양자를 모두 고려해야 할 것이다.^[12]

<표 2>는 CCITT T. 62에서 권고하고있는 국제적인 페이지 형식을 나타내고 있다. 여기서 A4용지는 210×280mm로서 ISO A4 (210×297mm)와 북미표준 A4 (216×280mm) 용지를 동시에 수용하는 용지 크기이다.

한국형 텔리텍스 터미널에 맞는 페이지 형식을 정하기 위해 당 연구소에서는 84년 8월, 여러 관공서와 기업체를 대상으로 공업 진흥청에서 규격화한 A Series와 B Series 용지에 대해서 이에 대한 조사를 한 바, 그 결과는 <표 3>

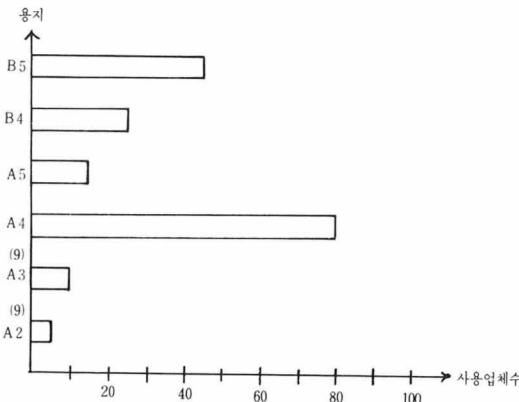
과 같다.^{9),14)} 이에 따라 한국형 텔리텍스의 페이지 형식은 A4와 B5 용지의 가로와 세로방향을 기본으로 하고, 큰 서류나 도표를 고려하여 B4 용지의 가로와 세로방향을 option으로 합이 적절하리라 생각한다.

그러나 실제 국내에서는 국제 규격용지가 아닌 16절지(190×286mm)를 가장 많이 사용하고 있는 실정인데, 16절지를 국내에서 기본으로 채택할 경우는, 이 용지에 대해 새로운 control code를 할당해 주어야 하며, 국내에서만 사용한다는 점을 감안할 때, 16절지는 이와 크기가 가장 유사한 B5 용지에 포함시키는 것이 좋으리라 생각한다.

여기서 용지의 크기와 방향에 대해서 규정하고 있지만, 이것들은 전송로 상의 논리 형태를 말하는 것이지, 인쇄시 사용하는 용지의 크기를 규정하는 것은 아니다.¹²⁾ (인쇄시 사용하는 용지는 사용자의 선택에 의하기도 함)

| Bit 87654321 | Format |
|--------------|--|
| 00000000 | Basic A4, Horizontal and Vertical |
| 00000001 | ISO A4, Hori. and Verti. |
| 00000010 | North American, Hori. and Verti. |
| 10000100 | ISO A4 Extended(ISO 3535), Verti. |
| 01000100 | ISO A4 Extended(ISO 3535), Hori. |
| 10001000 | North American Legal, Verti. |
| 01001000 | North American Legal, Hori. |
| 00000011 | ISO A4, Hori. and Verti. (for use by Kanji Terminal) |
| 00010000 | ISO B5, " " |
| 00100000 | ISO B4, " " |

〈표 2〉 CCITT T.60에 권고되어있는 Page Format



〈표 3〉 국내에서 사용하는 문서의 크기 분포

2. 줄 간격과 문자 간격 (Line Pitch and Character Pitch)

줄 간격은 인쇄되는 서류 상에서 임의의 한 줄과 그 다음 줄과의 간격으로 이 파라미터는 한 페이지 내에 인쇄가능한 최대 줄 수를 결정하며, 문자 간격은 임의의 한 문자와 그 다음 문자와의 간격으로 이 파라미터는 한 줄 내에 인쇄 가능한 최대 문자 수를 결정한다. 따라서 이 두 파라미터에 의해 한 페이지의 정보량(한 페이지 내에 수록 할 수 있는 문자 수)이 정해지므로, 줄 간격과 문자 간격은 한 페이지의 정보량과, 인쇄되는 글자의 크기와 모양등을 고려하여 결정해야 할 것이다.

영문의 경우는 CCITT T. 60에서 이에 대해 권고하고 있는데, 줄 간격은 4.23mm (25.4mm의 1/6)의 0.5, 1, 1.5 및 2배를 기본 요건으로 하고, 3.175mm (25.4mm의 1/8)와 5mm의 0.5, 1, 1.5 및 2배를 option으로 하고 있다.¹²⁾

국내의 경우, 대부분의 워드프로세서나 텔렉스, 타자기에서는 4.23mm의 1.5배(6.67mm)와 2배(8.47mm)를 사용하고 있고, 5mm의 2배(10mm)도 관공서에서 많이 사용하고 있다.

이 파라미터에 대해서는 T. 60에서 모두 권고되어 있으므로 한글의 줄간격은 CCITT T. 60에서 권고된 사항을 그대로 적용함이 좋으리라 생각된다.

문자간격에 대해서는 영문의 경우 10pitch를 표준으로, 12pitch와 15pitch를 option으로 하고 있다.

한글은 영문에 비해 활자가 크고 모양이 다양하므로 인쇄할 때는 dot matrix printer를 사용하는데, 현재 국내 대부분의 프린터의 한글 font 구성은 24×24dots로 구성되어 있다.

이 font의 구성은 180dots/inch 프린터를 기준으로 하면 7.5pitch(7.5문자/inch)에 해당되며, 160dots/inch 프린터를 기준으로 하면 6.67pitch(6.67문자/inch)가 된다. 국내에서는 현재 160dots/inch 프린터보다 더 많이 보급되어 사용되고 있으나, 앞으로 주변기기의 발전과 고급화 될 추세를 고려할 때, 문자 pitch 규정시는 180dots/inch 프린터에 기준을 둬야 바람직할 것으로 생각한다.

따라서, 한글의 문자간격은 7.5pitch를 기준으로 하고, 6.67pitch를 option으로 하여 규정할

이 타당할 것으로 사료된다.

IV. 맺음말

텔리텍스는 텔리마틱 서어비스의 일환으로서 OA시대를 선도하는 이정표가 될 것이다.

여러 한글 한자부호를 사용하고 있고, 용지의 규격 및 한글의 행과 열의 pitch에 관한 정의가 불분명한 국내실정을 비추어 볼때 정보화 사회로 나아가는 현시점에서 이에 관한 표준화작업은 시급한 당면과제임이 틀림없고, 본 논문이 조금이나마 도움이 될 수 있기를 바란다.

끝으로 '신규 서어비스 기술개발' 과제 책임자인 강 철희 박사께 이 지면을 통해 감사를 드리는 바이다.

参 考 文 献

1. CCITT Rec. T. 51
2. CCITT Rec. T. 60

3. CCITT Rec. T. 61
4. CCITT Rec. T. 62
5. CCITT Rec. F. 200
6. 한국공업규격 정보교환용부호 KS C 5601 - 1982
7. 한국공업규격 정보교환용 한자부호계 KS C 5617 - 1982
8. 한국공업규격 정보교환용 부호확장법 KS C 5620 - 1977
9. 한국공업규격 장표의 설계 기준 KS A3800 - 1978
10. Code Extension ISO-2022 - 1982
11. Registration Procedure ISO-2375 - 1980
12. 강철희, 'KETRI형 텔리텍스 터미널' Workshop 논문집, '텔리마틱 서어비스의 최근 기술 동향에 대하여', pp 15~pp31, 한국전기통신연구소, 1984년
13. 한국표준연구소, '정보처리 표준화에 관한 연구', KSRI-IR-35
14. 한국전기통신연구소, '새로운 서비스 개발에 관한연구', 연구보고서, 1984. 12.