



신출한자 국제 표준화 연구

연구책임자 김 흥 규



국립국어원



문자코드연구센터
CHARACTER CODES RESEARCH CENTER

제 출 문

국립국어원장 귀하

“신출한자 국제 표준화 연구”에 관하여 귀 원과 체결한 연구용역 계약에 의하여 연구보고서를 작성하여 제출합니다.

2008년 12월 19일

연구책임자 : 김홍규(고려대학교 국어국문학과)

고려대학교 산학협력단

연구책임자 김 홍 규

목 차

목 차	i
I. 연구개요	1
1. 연구 목표	1
2. 연구 필요성	2
II. 연구내용	4
1. 비표준한자 수집 및 정리	4
2. 신출한자 검정	4
1) 원전 대조를 통한 오류 여부 검정	5
2) 『SuperCJK Ver14.0』 코드북 대조	6
3) 『CJK Ext.C Ver10』 코드북 대조	7
4) ‘Annex S(IRGN951)’에 의한 이체자 통합 및 선별	7
3. IRG 제안을 위한 <한국측 제안한자 목록> 작성	7
4. 전문가 자문회의 개최	8
5. IRG 회의 참석 및 국제 표준 활동 대응	8
6. 문자코드 정보 공유 및 지원	9
III. 연구결과	10
1. 비표준한자 수집 및 정리	10
2. 신출한자 검정	11
3. IRG 제안을 위한 <한국측 제안한자 목록> 작성	12
1) <한국측 제안한자 목록> 한자정보 작성	12
2) <한국측 제안한자 목록> IDS 정보 작성	13
3) <한국측 제안한자 목록> 트루타입 폰트 제작	14
4) <한국측 제안한자 목록> 전거정보 원문이미지 자료 구축	14
4. 전문가 자문회의 개최	15
5. IRG 회의 참석 및 국제 표준 활동 대응	15
1) <Ext.B 한국측 한자> 트루타입 폰트 제작	15
2) <Ext.C 한국측 한자> 트루타입 폰트 제작	16

3) <Ext.C 한국측 한자> 선거정보 원문이미지 자료 구축	17
4) <Ext.D 한국측 제안한자> 선거정보 원문이미지 자료 구축	17
6. 문자코드 정보 공유 및 지원	18
1) 소식지 발간	18
2) 웹사이트 운영	18

【연구 결과물 첨부】

※첨부1 <한국측 제안한자 목록>(IDS 정보 포함)	21
※첨부2 <한국측 제안한자 목록> 트루타입 폰트	79
※첨부3 <한국측 제안한자 목록> 선거정보 원문이미지 자료(부분)	107
※첨부4 <Ext.B 한국측 한자>(166자) 트루타입 폰트(최종 수정)	161
※첨부5 <Ext.C 한국측 한자>(404자) 트루타입 폰트	165
※첨부6 <Ext.C 한국측 한자>(404자) 목록	169
※첨부7 <Ext.C 한국측 한자> 선거정보 원문이미지 자료(부분)	195
※첨부8 <Ext.D 한국측 제안한자>(32자) 목록	239
※첨부9 <Ext.D 한국측 제안한자> 선거정보 원문이미지 자료	243
※첨부10 <문자코드연구센터> 소식지 21호, 22호	277

I. 연구개요

1. 연구 목표

이 연구는 2003년 국제 표준화 기구에 제안된 ‘Ext.C1 한국측 제안한자’ 이후 새로 산출된 신출한자를 수집·정리하고, 이를 검정하여 국제 표준코드에 등록하기 위한 방안을 모색하는 데에 그 목표를 둔다.

우리나라는 2000년 이후 국책사업으로 고전 전산화 사업을 연차적으로 진행하였고, 각 분야의 다양한 자료들이 웹 데이터베이스로 구축되어 관련 분야 연구자는 물론 관심 있는 사람이라면 누구나 손쉽게 자료를 이용할 수 있게 되었다. 그런데 이 과정에서 옛한글과 신출한자 등 국제 표준코드에 없는 문자들의 전산 처리에 여러 가지 어려운 문제가 발생하여 이를 해결하고자 하는 각계의 노력과 연구가 진행되고 있다.

문자코드연구센터는 이러한 연구의 일환으로 1998년부터 표준코드에 아직 수용되지 않은 비표준문자를 등록하고 처리하는 방법을 표준화하는 데에 일조를 해 왔다. 그리고 유니코드와 같은 대규모 국제 표준코드의 원활한 수용과 활용에 필요한 지식·기술·자원을 개발하여 그 결과물을 학술·문화·정보산업·출판 분야의 각 기관과 연구자들에게 지원해 왔다. 이러한 지식과 기술 축적을 바탕으로 2002년부터는 고전 전산화 사업에서 산출된 신출한자를 수집·정리하고, 신출한자의 검정과 이체자의 효율적 처리 문제 등 한자 처리에 대한 제반 정보와 기술을 관련 기관 및 산업체에 제공하고 공유하였으며, 신출한자 검정과 이체자 처리에 관하여 향후에도 지속적으로 협의를 할 수 있는 협력 체계를 갖추었다.

이와 같은 노력의 결과로 2007년에 본 센터의 ‘옛한글정보처리연구위원회’에서 정리한 <옛한글자모 확장목록>을 유니코드 BMP 영역에 최종 등록할 수 있었으며, ‘Ext.C1 한국측 제안한자’ 가운데 404자를 ISO/IEC 10646에 Ext.C 한자로 최종 등록할 수 있었다. 그러나 2003년에 수집·정리하여 제안된 Ext.C1 한자 이후 새로 산출된 신출한자의 국제 표준화는 그다지 원활하게 진행되지 않았다.

따라서 2003년 이후 새로 산출된 신출한자에 대해서 국제 표준코드 체계에 등록하여 향후 컴퓨터에서 한자를 원활하게 처리할 수 있는 기반을

마련하는 데에 이바지할 수 있도록 한다. 그리고 유니코드와 같은 국제 표준코드의 원활한 수용과 활용에 필요한 지식·기술·자원을 개발하여 학술·문화·정보산업·출판 분야의 각 기관과 연구자들에게 지원함과 아울러 국제 문자코드의 표준 활동에 선도적 역할을 담당하고자 한다.

2. 연구 필요성

신출한자는 국제 표준 문자코드 체계, 특히 국제 표준코드로 통용되고 있는 유니코드(혹은 ISO/IEC 10646)에 등록되어 있지 않아 컴퓨터에서 전산 처리를 할 수 없는 한자를 말한다. 컴퓨터에서의 효율적인 한자 처리는 우리나라에서만 아니라 한자문화권 전체의 난제로, 현재 유니코드에는 약 70,000여 자의 한자가 등록되어 있음에도 불구하고, 지속적으로 신출한자가 발견되고 있어 이들 한자의 처리 방법을 강구할 필요가 대두하게 되었다.

이러한 문제를 해결하기 위해 국제 표준화기구에서는 ISO/IEC 10646 제정 초기부터 한자문화권 국가들이 별도로 협의기구를 구성하여 한자 처리와 관련된 제반 논의를 하게 하였고, 그 뒤 이 기구는 JTC1/SC2/WG2 산하에 IRG(Ideographic Rapporteur Group)로 편입되게 된다. IRG는 해마다 두 차례씩 회의를 개최하여 한자 처리와 관련된 정보와 의견을 서로 교환하면서 그 해결 방안을 모색해 나가고 있다.

우리나라도 2000년 이후 국책사업으로 진행된 고전 전산화 사업에서 많은 양의 신출한자가 산출되었다. 그러나 이들 신출한자는 주로 다음과 같은 이유에 의해서 곧바로 국제 표준코드에 반영할 수가 없다.

첫째 주관기관이나 사업연도에 따라 각기 다른 방식으로 신출한자를 처리하였기 때문에 신출한자 판별을 위한 공통의 표준적 지침이 확립되어 있지 않아 주관기관이나 사업자의 임의적 판단에 의해 신출한자를 처리한 경우가 많다.

둘째 원전 자형을 오독하여 표준코드에 있는 한자를 신출한자로 처리한 사례로, 특히 Ext.B 영역의 한자인 경우가 많다.

셋째 입력 과정에서 원전을 잘못 입력한 사례로, 입력하고자 하는 한자

가 특정 한자의 이체자로 표준코드에 등록되어 있음[특히 Ext.B 영역 한자]에도 정자로 바꾸어 입력하는 경우가 많다.

이러한 문제를 해결하기 위해서는 기존에 신출한자로 판정된 한자에 대한 보다 정밀한 재검정이 필요하며, 그 결과를 IRG가 정한 규격에 맞추어 국제 표준에 제안할 목록을 작성하여야 할 필요가 있는 것이다. 이것은 국가적 차원에서 중복 연구를 막을 수 있을 뿐 아니라 물질적 측면에서 국제 문자자원을 확보함으로써 향후 막대한 비용과 인력을 절감할 수 있을 것이다.

II. 연구내용

1. 비표준한자 수집 및 정리

그동안 본 센터에서는 표준코드에 등록되지 않은 비표준한자를 지속적으로 다양한 방법으로 수집하여 정리하였다. 특히 지난 2001부터 2007년까지는 국책사업으로 <국가지식정보자원관리사업>이 진행되었고, 그 일환으로 고전자료 전산화 사업이 추진되었는데, 이 과정에서 다량의 비표준한자가 산출되었다. 본 센터에서는 이 사업을 수행한 각 기관과 연계하여 다음과 같이 비표준한자를 수집하여 정리하였다.

[비표준한자 수집 현황(2004-2007)]

연 도	수 집 대 상	수집문자
2004	·『사마방목』 ·고전국역총서 ·한국문집총간	1,096자
2005	·2001-2004년간 <국가지식정보자원관리사업> 산출 신출한자	10,247자
2006	·2005년도에 <국가지식정보자원관리사업> 산출 신출한자	3,926자
2007	·2006년도에 <국가지식정보자원관리사업> 산출 신출한자	2,197자
합 계		17,466자

2. 신출한자 검정

신출한자 여부의 정확한 판정을 위해 본 센터에서는 여러 단계의 검정 과정을 수행하고 있다. 신출한자를 무분별하게 추출하여 많은 양을 확보하는 것보다는 가능한 정확한 목록을 작성하여 국제 표준기구에 제안하는 것이 바람직하기 때문이다. 국제적으로 공감하고 인정할 만한 신출한자 목록을 제안하지 못하면, 국제 표준으로 수용되기 어려울 뿐만 아니라 이후 한자 표준화 사업에 대한 국가적 신뢰를 잃을 수도 있다. 그만큼 신출한자의 처리는 가능한 한 신중하고 정확하게 진행되어야 한다.

본 연구에서 수행하는 신출한자 처리 과정은 다음의 여섯 단계로 이루어진다.



1) 원전 대조를 통한 오류 여부 검정

이 검정 작업은 본 센터에서 1차로 수집한 신출한자를 원전과 직접 대조하면서 오독으로 인한 오류나 폰트 오류 여부 등을 검정하는 일이다.

○ 원전 자형의 오독

원전 자형의 오독은 원전 한자의 자형에 대한 착각이나 이해의 부족으로 기존의 표준코드에 등록되어 있는 한자임에도 이를 판별하지 못한 경우에 많이 발생한다. 이 오류는 원전 자료를 입력하는 과정에서 잘못된 것으로 보이며, 교정 및 교열 작업이 아직 이루어지지 않았거나 충실하게 수행되지 못하였음을 나타낸다. 신출한자의 정확한 검출을 위해서는 원전의 자형이나 획의 미세한 차이도 자세하게 대조하고 변별해야 한다.

○ 폰트 자형의 오류

폰트 자형의 오류는 앞서 살펴본 원전자형의 오독에서 많이 나오지만, 신출한자를 폰트로 제작하는 과정에서 자획의 일부가 잘못 제작되기도 한다. 이 예는 비교적 사소한 오류로 생각할 수 있으나 대표자 자형 선정에 부정확한 정보를 줄 수 있으므로 주의해야 한다. 신출한자를 국제 표준한자에 제안할 때에는 폰트를 제작해서 목록과 함께 제출해야 하기 때문에 보다 정확한 폰트 자형의 제작이 필요하다.

○ 이체자 처리의 오류

이체자 처리의 오류는 이체자가 아닌데도 이체자로 잘못 파악하거나, 이체자가 표준코드에 등록되어 있음에도 정자로 바꾸어 입력한 경우에 발생한다. 이 경우는 목활자로 간행된 문집류의 한적에서 흔히 나타나는 사례로서, 대체로 1차 입력자가 원전의 목판 활자 자형에 익숙하지 않아 생긴 것일 가능성이 크다. 이밖에도 목판이 오래되어 자획의 일부가 떨어져 나간 경우나 필사자의 개인적 필체에 의한 부분적인 자형의 차이 등을 이체자로 처리한 사례도 나타나는데 이러한 한자들은 모두 이체자로 보기 어렵다. 또 다량의 이체자가 표준코드에 등록되어 있는데, 이를 인식하지 못하여 입력하기 쉬운 정자로 바꾸어 입력하는 경우도 종종 보인다.

○ 한자 정보의 오류

한자 정보의 오류는 해당 한자의 부수 정보, 총획수 정보, 잔여획수 등을 잘못 파악하여 나타난다. 물론 같은 한자라도 각 나라별로 관습적으로 부수를 다르게 보는 경우가 더러 있지만, 위의 예는 적절하지 않은 사례이다. 부수의 판단이 애매한 경우에는 조합의 방식이 같은 한자의 부수를 참조할 필요가 있다. 부수가 다르게 정해지면 잔여획수가 아울러 달라지기 때문에 유의해야 한다. 일반적으로 통용되는 방법으로 부수를 정해야 해당 한자를 정확하고 쉽게 검색할 수 있다.

2) 『SuperCJK Ver14.0』 코드북 대조

이 과정은 제출된 신출한자 목록을 IRG에서 제정한 『SuperCJK Ver14.0』 코드북과의 대조를 통하여 신출한자 여부를 판별하는 작업이다. 이 코

드북은 CJK 통합한자 Ext.B 영역의 한자까지를 포함하고 있기 때문에 신출한자 판정에 매우 유용하다. 그리고 ISO/IEC 10646과 유니코드에는 한자문화권 각국이 제안한 최초의 한자를 통합하여 대표자형만을 신고 있지만, 『SuperCJK Ver14.0』에는 각국이 제안한 최초 자형을 모두 신고 있어 이형자를 판단하는 데에 중요한 자료를 제공하고 있다.

3) 『CJK Ext.C Ver10』 코드북 대조

‘CJK Ext.C’ 한자는 2006년에 IRG 회의에서 결정되어 ISO/IEC 10646과 유니코드에 등록하기 위한 과정을 거치고 있다. 현재 그 마지막 단계를 논의 중에 있으며, 최종 버전이 『CJK Ext.C Ver10』으로 정리되어 있다. 그러므로 ‘CJK Ext.C’ 한자까지 대조를 하여야만 실질적인 신출한자라 할 수 있을 것이다.

4) ‘Annex S(IRGN951)’에 의한 이체자 통합 및 선별

‘Annex S(IRGN951)’ 문서는 IRG의 기술문서로서, 유사 자형 한자와의 구분과 통합에 대한 기준이 제시된 문건이다. 이 문건의 취지는 기본적으로 각국에서 이체자를 무분별하게 제안하는 것을 통제하기 위한 것이다. 이 문서는 기존에 수용된 이체자는 그대로 두고 이후 제안되는 신출한자에만 적용되는 기준인데, 통합해도 무방한 이체자와 구분해야 할 이체자를 예시를 통하여 표준화한 것이다.

3. IRG 제안을 위한 <한국측 제안한자 목록> 작성

이상의 과정을 거쳐 최종적으로 판정된 신출한자는 <한국측 제안한자 목록>으로 작성하여 IRG에 제안하게 된다. 목록 작성을 위해서는 IRG에서 제시하는 ‘국제 표준 제안한자 포맷(IRG N881)’에 맞게 필요한 정보를 확인하여 입력하여야 하는데, 이 과정에서 수행되는 작업 내용과 통합목록의 포맷은 다음과 같다.

[IRG N881 포맷 양식]

KXI Index XXXX.YYZ (강희자전 위치정보)	KXradical UCS code(table 76+77)+flag XXXXY (강희자전 부수의 유니코드값)	Stroke count of phonetic (잔여획수)	First stroke of the phonetic (오필획)	Source Information and ID SSSNNNNN (소스 정보)	Variant Ucode U1,U2 (이체자 정보)
1-8(byte)	9-13	14-15	16	17-24	25-35

<한국측 제안한자 목록>이 완성되면, 이 목록을 바탕으로 한자를 결구에 의해 분해한 ‘IDS(Ideographic Description Sequence) 정보’를 추가로 입력하고, 트루타입 폰트를 제작할 것이다. 그리고 최근에는 해당 한자의 전거자료를 요구하고 있으므로 최종 목록의 ‘원전 이미지 자료’를 추가할 것이다.

4. 전문가 자문회의 개최

이상에서 작성된 <한국측 제안한자 목록>은 해당 분야의 전문가에게 감수를 의뢰하고, 자문회의를 통해 보다 구체적으로 검토한다. 이 평가를 통하여 검토의 정확성을 평가 받고, 이후 작업을 수정 및 보완 작업을 진행할 계획이다.

5. IRG 회의 참석 및 국제 표준 활동 대응

그 동안 본 센터에서는 IRG 회의에 연구책임자가 한국측 대표로 참석하여 한자와 관련하여 국제 표준화 동향을 파악하고, 비표준문자의 국제 표준화를 지원하였다. 2008년에는 6월 9일부터 13일까지 부산에서 제30차 IRG 회의가 개최될 예정이며, 제31차 회의는 11월 10일부터 14일까지 중국 쿤밍에서 개최될 예정이다. 이와 같은 IRG 회의에 참석하여 국제 표준화를 지원할 예정이다.

6. 문자코드 정보 공유 및 지원

- 웹사이트 운영 : 현재 운영 중(<http://www.codes.or.kr/>)
- 소식지 발간(2회)

Ⅲ. 연구결과

1. 비표준한자 수집 및 정리

그동안 본 센터에서 수집하여 정리한 비표준한자는 다음과 같다.

[비표준한자 수집 현황(2004-2008)]

연 도	수 집 대 상	수집문자
2004	·『사마방목』 ·고전국역총서 ·한국문집총간	1,096자
2005	·2001-2004년간 <국가지식정보자원관리사업> 산출 신출한자	10,247자
2006	·2005년도에 <국가지식정보자원관리사업> 산출 신출한자	3,926자
2007	·2006년도에 <국가지식정보자원관리사업> 산출 신출한자	2,197자
2008	·‘한국고전번역원’ 정리 이체자	14,028자 (정자 4,838자)
합 계		31,494자

2008년에는 ‘한국고전번역원’으로부터 『한국문집총간』에서 산출된 이체자 14,028자를 수집하여 정리하였다. 이 한자들은 정자 4,838자에 대응하는 이체자 14,028자를 정리한 것으로, 현재 ‘이체자검색시스템’ 홈페이지를 통해 제공하고 있다.¹⁾ 그러나 ‘한국고전번역원’으로부터 제공받은 이체자는 배포상의 문제가 있을 뿐만 아니라 異寫字와 異形字와 같이 복잡한 형태로 구성되어 있어 신출한자 추출을 위해서는 새로운 기준이 필요로 하다는 결론에 이르게 되었다. 이로 인하여 향후 좀 더 구체적인 연구를 위해 수집만 하기로 하고, 이 연구에서 신출한자 검정을 위한 대상 한자에서는 제외하였다.

1) 고전국역원 ‘이체자검색시스템’ 홈페이지(<http://www.itkc.or.kr/DCH/index.jsp>) 참조.

2. 신출한자 검정

이와 같이 수집한 비표준한자는 2006년과 2007년 2차례에 걸쳐 <한국측 제안한자 목록>을 작성하였다. 2006년에는 2004년부터 2006년까지 수집한 비표준한자 15,269자를 대상으로 모두 2차례의 검정 과정을 거쳐 신출한자 5,736자를 추출하였고, 이 가운데 전거정보와 각종 부가정보가 확실한 한자 1,175자를 확정하여 <Ext.C2 한국측 제안한자 목록>을 작성하여 KIRG에 제출하였다.²⁾ 그러나 제27차 IRG 회의에서 Ext.C1 한자의 일부를 Ext.C로 확정하고, 그 나머지를 Ext.D로 결정하게 되면서 Ext.C2 계획 자체가 취소되어 결국 제출하지 못하게 되고 말았다. 이에 따라 2006년도에 본 센터에서 작성한 <Ext.C2 한국측 제안한자>는 어쩔 수 없이 <Ext.E 한국측 제안한자>로 제안하기로 KIRG에서 결정하였다.

그리고 2007년 중국 시안[西安]에서 개최된 제28차 IRG 회의에서 각국에서 국가적으로 시급한 한자를 Ext.D 작업과 별도로 ‘긴급 필요[Urgently Need] 한자 추가 접수’를 의결하였는데, KIRG에서는 <Ext.E 한국측 제안한자>로 준비 중인 한자 목록을 <Ext.DUN 한국측 제안한자>로 제출하기로 결정하였다. 이에 따라 그동안 진행해 오던 <Ext.E 한국측 제안한자 목록>을 <Ext.DUN 한국측 제안한자 목록>으로 변경하여 작성하게 되었다. 이 목록은 2006년에 작성한 <Ext.C2 한국측 제안한자> 1,175자와 2007년에 새로 수집하여 추출한 신출한자 1,221자를 통합한 것으로 최초의 목록은 모두 2,396자였다. 이 목록을 바탕으로 모두 2차례의 검정과 1차례의 자문회의를 거쳐 중복되는 글자를 삭제하였으며, 최종적으로 확정된 한자 2,288자를 <Ext.DUN 한국측 제안한자 목록>으로 작성하여 KIRG에 제출하였다.³⁾

위에서 2004년부터 2006년까지 추출한 신출한자 5,736자 가운데 전거정보와 각종 부가정보가 확실한 한자 1,175자를 <Ext.C2 한국측 제안한자 목록>으로 작성하였다고 언급하였다. 이 1,175자를 제외한 나머지 4,561자는 전거정보가 없거나 각종 부가정보가 미비한 경우가 많았는데, 이를 다시

2) 『21세기 세종계획 문자코드 표준화 연구』 2006년 연구보고서, 국립국어원, 2006, 18-20면 참조.

3) 『21세기 세종계획 문자코드 표준화 연구』 2006년 연구보고서, 국립국어원, 2006, 19-21면 참조.

원본 자형의 유무와 전후 문맥의 유무에 따라 다음의 표와 같이 분류할 수 있었다.

[검정 대상 한자의 분류]

분 류	원본 자형 유무	전후 문맥 유무	자수
제1그룹	○	○	1,892
제2그룹	○	×	1,810
제3그룹	×	○	33
제4그룹	×	×	826
합 계			4,561

이 가운데 제1그룹과 제3그룹에 해당하는 신출한자 1,925자의 경우에는 해당 한자를 제공받은 기관과의 협의 하에 출전정보와 전거정보 자료 이미지를 확보할 수 있을 것으로 판단하였다. 이러한 판단에 따라 2008년에는 이 1,925자를 대상으로 신출한자 검정을 진행하였고, 그 결과 1,164자의 한자에 대해 전거정보와 각종 부가정보가 확실한 신출한자를 추출할 수 있었다. 이렇게 추출한 1,164자는 다시 재검정 작업을 진행하여 223자의 중복자를 제거하였으며, 전문가 자문회의를 거쳐 최종적으로 931자를 <한국측 제안한자 목록>으로 확정하였다.

3. IRG 제안을 위한 <한국측 제안한자 목록> 작성

1) <한국측 제안한자 목록> 한자정보 작성

이상과 같은 과정을 거쳐 확정된 <한국측 제안한자 목록>은 “국제 표준 제안한자 양식(IRG N881)”에 맞추어 필요한 정보를 확인하여 입력하여야 한다. 이 규정에 따라 최종 확정된 <한국측 제안한자 목록>의 한자 965자는 ‘강희사전 위치정보’·‘강희사전 부수정보’·‘잔여획수’·‘오픈획 정보’·‘해당 한자의 원문정보’·‘이체자 정보’ 등의 각종 부가정보를 입력하였다. 그리고 이렇게 입력된 한자 정보의 오류 여부를 전문가에게 의뢰하

여 자문을 구하였으며, 모든 검토 작업이 끝난 목록에 대해서는 임시코드를 부여하였다.

※ 첨부1 <한국측 제안한자 목록>(IDS 정보 포함) 참조.

2) <한국측 제안한자 목록> IDS 정보 작성

2006년부터 IRG에서는 한자의 구성하는 편방과 부수에 의해 구조적으로 분해한 ‘IDS(Ideographic Description Sequence) 정보’를 입력하여 제출할 것을 요구하고 있다. 이것은 ISO/IEC 10646과 유니코드에 수용되어 표준화되었으며, ‘漢字結構情報’ 또는 ‘漢字構造情報’라고도 한다.⁴⁾ 다음의 표와 같이 IDS는 모두 12가지 형태의 기호로 되어 있으며, 이 기호를 사용하여 일정한 식에 의해 분해와 조합을 할 수 있도록 하고 있다.

[유니코드에서 제시한 12가지의 한자 결구]

번호	IDS	설명	글자 예
1	□	좌우 결구	校 = □木交
2	田	상하 결구	忠 = 田中心
3	田	좌중우 결구	靜 = 田;青爭
4	田	상중하 결구	蓋 = 田+去皿
5	回	완전 포위 결구	回 = □□□
6	回	상삼면 포위 결구	問 = 回門口
7	回	하삼면 포위 결구	凶 = 回土口
8	回	좌삼면 포위 결구	医 = 回口矢
9	回	좌상면 포위 결구	病 = 回广丙
A	回	우상면 포위 결구	匈 = 回凶勺
B	回	좌하면 포위 결구	迄 = 回乞辵
C	回	겹침 결구	爽 = 回大玨

4) The Unicode Consortium, “The Unicode Standard Version 4.0”, Addison-Wesley Professional, 2003, pp.307-309과 ‘ISO/IEC 10646:2003 부속서 F’에서 IDS를 정의하고 그 문법을 기술하고 있다.

이 연구에서는 이러한 국제 기준에 맞추어 최종 결과물인 <한국측 제안한자 목록>에 수록된 한자의 IDS 정보를 입력하였다.

※ 첨부1 <한국측 제안한자 목록>(IDS 정보 포함) 참조.

3) <한국측 제안한자 목록> 트루타입 폰트 제작

2006년부터 IRG에서는 IDS 정보와 함께 트루타입 폰트도 함께 제출할 것을 요구하고 있다. 이에 따라 이 연구에서는 최종 결과물인 <한국측 제안한자 목록>의 한자를 대상으로 트루타입 폰트를 제작하였다. 트루타입 폰트 제작은 본 센터에서 독자적으로 처리할 수 없어 외부의 폰트 제작 전문가에게 의뢰하였다.

※ 첨부2 <한국측 제안한자 목록> 트루타입 폰트 참조.

4) <한국측 제안한자 목록> 전거정보 원문이미지 자료 구축

2008년부터는 IRG에 한자를 제안하기 위해서는 IDS 정보를 비롯하여 트루타입 폰트, 그리고 출전전거 원문이미지 자료 등을 일괄적으로 제출할 것을 요구하고 있다. 그 중에서 출전을 확인할 수 있는 원문 이미지자료가 없을 경우에는 국제 표준 문자코드에 제안 자체가 불가능하도록 규칙을 강화하고 있다. 이러한 방침에 따라 이 연구의 최종 결과물인 <한국측 제안한자 목록>에 수록된 해당 한자의 출전과 전거정보의 원문이미지 자료를 구축하였다.

※ 첨부3 <한국측 제안한자 목록> 전거정보 원문이미지 자료(부분) 참조.

4. 전문가 자문회의 개최

이상에서 작성된 <한국측 제안한자 목록>은 해당 분야의 전문가에게 감수를 의뢰하고, 자문회의를 통해 보다 구체적으로 검토하였다. 이 과정을 통하여 검토의 정확성과 작성된 한자정보가 정확하게 기입되었는지를 평가받았다. 그리고 10자의 한자는 탈각과 오독에 의해 자형이 변형된 것으로 판명되어 새로 제안하기에는 부적절하다는 평가를 받아 <한국측 제안한자 목록>에서 제외시켰다.

5. IRG 회의 참석 및 국제 표준 활동 대응

그 동안 본 센터에서는 IRG 회의에 연구책임자가 한국측 대표로 참석하여 국제 표준화 동향 파악 및 비표준문자의 국제 표준화를 위한 각종 자료와 정보를 제공하였다. 2008년에는 제30차 IRG 회의와 제31차 IRG 회의가 개최되었는데, 제30차 IRG 회의는 6월 9일부터 13일까지 부산에서 개최되었고, 제31차 회의는 11월 10일부터 14일까지 중국 쿤밍에서 개최되었다.

이러한 IRG 회의를 위해 부산에서 개최된 제30차 IRG 회의에 참석하는 한편, 현재 논의 중인 Ext.D 한자에 대한 검토 작업을 실시하였다. 그리고 IRG에서 요구하는 각종 자료와 정보를 정리하여 KIRG에 제공하였다.

1) <Ext.B 한국측 한자> 트루타입 폰트 제작

본 센터는 2007년에 KIRG의 요청에 의해 기존의 국제표준 문자코드에 등록된 42,711자에 달하는 ‘Ext.B 한자’ 가운데 한국에서 제안하여 등록된 166자의 한자에 대해 자형을 검토하고, 이를 바탕으로 한국에서 제안한 한자 자형을 트루타입 폰트로 제작하여 지원하였다. 이것은 국제표준 문자코드에 등록된 각국의 자형을 ‘멀티 컬럼(Multi Colum)’ 형식으로 표현하기 위한 것이다.

2008년에는 지난해에 제작한 <Ext.B 한국측 한자> 트루타입 폰트를 바탕으로 IRG에서 Ext.B 한자를 ‘멀티 컬럼(Multi Colum)’으로 표현하였는데,

이를 검토하는 과정에서 4자의 한자 자형이 최초 제안한 한국의 한자 자형과 서로 다름을 발견하였다. 이에 이 연구에서는 ISO 10646 등과의 비교 검토 작업을 다시 실시하여 4자의 한자 자형을 수정한 <Ext.B 한국측 한자> 트루타입 폰트를 다시 제작하여 KIRG에 제공하였다.

참고로 IRG에서 제공한 ‘Multi Colum Chart(image)[Ext_B-v21]’의 자형과 ‘ISO 10646’의 자형이 서로 다른 ‘Ext.B 한국측 한자’는 다음 표에서 보는 것과 같은 4자이다.

[Ext.B 한국측 한자의 자형 비교(4자)]

Ucode	Kcode	Multi Colum Chart(image)[Ext_B-v21]	ISO 10646(image)
25525	K4-0089	𪛗	𪛗
20E0C	K4-0023	𪛘	𪛘
2101F	K4-0026	𪛙	𪛙
27C0F	K4-00D9	𪛚	𪛚

※ 첨부4 <Ext.B 한국측 한자>(166자) 트루타입 폰트(최종 수정) 참조.

2) <Ext.C 한국측 한자> 트루타입 폰트 제작

2007년에 Ext.C 한자 4,219자가 최종적으로 확정되었는데, 한국에서 제안한 한자 가운데에는 404자가 최종 등록되었다. 이에 KIRG로부터 <Ext.C 한국측 한자> 트루타입 폰트 제작을 요청하였고, 이 연구에서는 이를 제작 지원하였다.

※ 첨부5 <Ext.C 한국측 한자>(404자) 트루타입 폰트 참조.

3) <Ext.C 한국측 한자> 전거정보 원문이미지 자료 구축

KIRG는 <Ext.C 한국측 한자> 트루타입 폰트 제작을 요청과 더불어 <Ext.C 한국측 한자>에 대한 전거정보 자료를 함께 구축할 것을 요청하였다. 이것은 최근 IRG에서 한자의 효율적인 처리를 위해 각국별로 전거정보를 제출하도록 그 조건을 강화하였기 때문이다. 이러한 요청에 따라 이 연구에서는 <Ext.C 한국측 한자> 404자에 대한 전거정보 원문이미지 자료를 구축하였다. 그러나 이 가운데 41자는 전거정보를 확인할 수 없었는데, 2003년 최초로 제안할 때 IRG에서 전거정보를 요구하지 않았기 때문에 미처 대응하지 못하여 초래된 것이었다.

※ 첨부6 <Ext.C 한국측 한자>(404자) 목록 참조.

※ 첨부7 <Ext.C 한국측 한자> 전거정보 원문이미지 자료(부분) 참조.

4) <Ext.D 한국측 제안한자> 전거정보 원문이미지 자료 구축

2003년에 KIRG에서는 <Ext.C1 한국측 제안한자>로 총 573자의 한자를 정리하여 IRG에 제출하였다. 이 가운데 Ext.C로 확정된 한자가 404자이고, 나머지 가운데 36자가 기존의 한자와 통합되거나 삭제되고, Ext.D에서 논의하기로 결정된 한자가 133자였다. 그런데 지난 11월 10일부터 14일까지 중국 쿤밍에서 개최된 제31차 IRG 회의에서 Ext.D에서 논의 중인 한자 가운데 전거정보가 없을 경우에는 무조건 삭제한다는 논의가 채택되었다. 이러한 결정에 따라 Ext.D에서 논의하기로 한 133자를 검토하여 전거정보 원문이미지 자료를 구축하였는데, 최종적으로 32자를 확인할 수 있었다.

※ 첨부8 <Ext.D 한국측 제안한자>(32자) 목록 참조.

※ 첨부9 <Ext.D 한국측 제안한자> 전거정보 원문이미지 자료 참조.

6. 문자코드 정보 공유 및 지원

1) 소식지 발간

2008년에는 7월 23일에 제21호를 발간하였고, 12월 15일에 제22호를 발간하여 관련기관 및 산·학 각계에 배포하였다. 소식지에 수록된 내용은 다음과 같다.

[제21호 목차]

- 권두언: IRG 회고와 전망
- 특별기고: 漢字構形學과 그 연구 동향
- 해외 연구동향: 漢字記述言語 ‘CDL’
- 제30차 IRG 회의 소개
- IRG PnP 소개
- 2008년도 사업 계획
- 센터 동정

[제22호 목차]

- 권두언: 유니코드 한자의 수용과 자형 문제
- 특별기고: Unicode 漢字 情報 및 異體字 辭典 편찬의 필요성
- 해외 연구동향: “漢語와 漢字 관계” 국제학술회의
- 연구 노트: 유니코드 한자의 정렬 문제
- 책소개

※ 첨부10 <문자코드연구센터 소식지 제21호·제22호> 참조.

2) 웹사이트 운영

2005년도 전면적인 개편 작업을 거친 홈페이지(<http://www.codes.or.kr/>)를 지속적으로 관리 운영 중에 있다.