

국립국어원 2007-01-17

발간등록번호

11-1370252-000064-10

연구보고서 【2007년】

21세기 세종계획
문자코드 표준화 연구

연구책임자 : 정우봉(고려대학교)

국립국어원

제 출 문

국립국어원장 귀하

21세기 세종계획 “문자코드 표준화 연구”에 관하여 귀 원과 체결한 연구 용역 계약에 의하여 연구 보고서를 작성하여 제출합니다.

2007년 12월 10일

총괄 연구 책임자 : 정우봉(고려대학교)

세부과제명 : 문자코드 표준화 연구

연구책임자 : 정우봉(고려대학교)

연 구 원 : 박종우(고려대학교)

연구보조원 : 신상현 외 1명

요 약 문

1. 사업 명

본 사업의 명칭은 <문자코드 표준화 연구>이다.

2. 연구의 목적

본 센터 운영의 제3단계에 해당하는 4년간(2004~2007)의 목표는 제1단계(1998~2000)와 제2단계(2001~2003) 성과를 기반으로 새로운 국제 표준 코드인 유니코드의 원활한 운용을 위한 국가적 차원의 지식과 기술을 확보하여 학술·문화·정보산업·출판 분야 등 각 기관과 연구자들에게 지원하는 데 있다.

2007년도는 본 센터 운영의 제3단계 마지막 해임과 동시에 전체 사업의 마지막 해에 해당하므로 그동안의 사업성과를 정리하고, 문자코드 표준화의 향후 과제와 전망을 살펴보기로 한다. 이를 통해 유니코드 환경에서 우리 어문 생활이 원활하게 이루어질 수 있도록 하는 데 기여하고자 한다.

3. 연구의 필요성

국제 표준코드인 유니코드가 본격적으로 실용화되면서 세계는 단일코드의 시대에 접어들게 되었고, 범용 운영체제와 응용 소프트웨어에서 유니코드가 표준코드로 채택되어 사용되면서 인터넷을 통해 수많은 정보가 언어와 국경을 초월하여 실시간으로 소통되고 있다. 유니코드의 실용화는 언어간 정보 소통의 편리함을 가져온 반면, 표준코드에 수용되지 못한 문자의 정보처리 문제를 남겨두었다.

21세기 변화된 정보 환경을 수용하고 이를 적극 대응하기 위해서 유니코드를 우리의 어문 생활과 정보 환경에 효율적으로 활용할 수 있는 방안을 마련하는 일이 무엇보다 시급한 과제이다.

4. 연구내용 및 결과물

1) 비표준문자 수집·정리 및 국제 표준화 지원

2007년도에는 2006년도 <국가지식정보자원관리사업>에서 산출된 신출한자를 각 사업에 참여하고 있는 관계 기관과 협의하여 추가로 수집하였다. 그리고 이를 정리하여 KIRG에 제출함으로써 <Ext.DUN 한국측 제안한자 목록>으로 제29차 IRG 회의에 제안할 수 있도록 지원하였다.

2) 문자코드 표준화 연구

① <Ext.B 한국측 한자> 트루타입 폰트 제작 지원

기존에 국제표준 문자코드인 유니코드에 등록된 42,711자의 Ext.B 한자 가운데 한국측에서 제안한 166자의 한자에 대한 트루타입 폰트 제작 요청이 KIRG로부터 있었다. 이에 본 센터에서는 <Ext.B 한국측 한자> 166자의 트루타입 폰트를 제작하여 지원하였다.

② <Ext.D 한국측 제안한자> 트루타입 폰트 제작 지원

2006년도에 IRG에서 ‘Ext.C 한자’ 4,219자가 확정되면서 ‘Ext.C2 한자’ 15,000여 자가 Ext.D로 명칭이 바뀌었다. 여기에는 한국에서 제안한 한자 135자가 포함되어 있으며, KIRG로부터 이들 한자의 트루타입 폰트 제작 요청이 있었다. 본 센터에서는 <Ext.D 한국측 제안한자> 135자의 트루타입 폰트를 제작하여 지원하였다.

③ <Ext.DUN 한국측 제안한자 목록> 작성

본 센터는 2006년도에 KIRG 지원을 위해 <Ext.C2 한국측 제안한자 목록>을 작성하였으며, IRG에 공식적으로 제출할 계획이었다. 그러나 제27차 IRG 회의에서 전체 20,000여 자의 ‘Ext.C1 한자’ 중 4,219자가 Ext.C로 확정되고, ‘Ext.C2 한자’에서 다룰 예정이었던 15,000여 자가 Ext.D로 명칭이 바뀌었다. 그리고 ‘Ext.D 한자’는 더 이상 제안을 받지 않는 등 IRG의 정책 변화로 2006년도에 작성한 <Ext.C2 한국측 제안한자 목록>은 <Ext.E 한국측 제안한자 목록>으로 제안할 수밖에 없는 상황이 되었다.

그런데 지난 6월 4일부터 중국 시안[西安]에서 개최된 제28차 IRG 회의 결과, 각국에서 국가적으로 시급한 한자를 Ext.D 작업과 별도로 ‘긴급 필요[Urgently Need] 한자 추가 접수’를 의결하였다. 따라서 본 연구의 최초 계획이었던 <Ext.E 한국측 제안한자 목록>을 총 2,288자의 <Ext.DUN 한국측 제안한자 목록>으로 작성하여 문자코드 표준화를 지원하였다.

④ <Ext.DUN 한국측 제안한자> 한자정보 작성

이상과 같은 과정을 거쳐 작성된 <Ext.DUN 한국측 제안한자> 목록은 “국제 표준 제안한자 양식(IRG N881)”에 맞추어 필요한 정보를 확인하여 입력하여야 한다. 이러한 원칙에 따라 최종 확정된 2,288자의 <Ext.DUN 한국측 제안한자> 목록의 한자에 대해서는 ‘강희사전 위치정보’·‘강희사전 부수정보’·‘잔여획수’·‘오필획 정보’·‘해당 한자의 원문정보’·‘이체자 정보’ 등을 모두 작성하였다.

⑤ <Ext.DUN 한국측 제안한자> IDS 정보 작성

‘Ext.DUN 한자’는 각국에서 ‘긴급하게 필요[Urgently Need]한 한자’에 해당되는 목록이기 때문에 제안 초기에 모든 정보가 완벽하게 갖추어져 있어야 한다는 전제가 있었다. IDS 정보를 비롯하여 트루타입 폰트, 그리고 출전선거 원문 이미지자료 등을 일괄적으로 제출할 것을 요구하였다. 이러한 방침에 따라 <Ext.DUN 한국측 제안한자> IDS 정보를 작성하여 KIRG에 제출하였다.

⑥ <Ext.DUN 한국측 제안한자> 트루타입 폰트 제작

<Ext.DUN 한국측 제안한자> 2,288자의 트루타입 폰트를 제작하여 KIRG에 제출하였다.

⑦ <Ext.DUN 한국측 제안한자> 출전선거 원문 이미지자료 구축

2006년에 작성한 <Ext.C2 한국측 제안한자> 1,175자 목록의 출전선거 원문 이미지자료에 2007년도에 새로 수집한 신출한자 1,221자의 출전선거 원문 이미지자료를 추가로 구축하였다. 그리고 <Ext.DUN 한국측 제안한자> 목록에 맞추어 파일 이름을 변경하고 다시 정렬하여 KIRG에 제출하였다.

⑧ 전문가 자문회의

제29차 IRG 회의에 제안된 <Ext.DUN 한국측 제안한자 목록>은 해당 분야의 전문가에게 감수를 의뢰하고, 자문회의를 통해 보다 구체적으로 검토하였다.

3) ‘문자코드 표준화 연구’ 사업 정리 및 향후 과제와 전망

본 센터는 1998년부터 제1단계(1998~2000), 제2단계(2001~2003), 제3단계(2004~2007)의 3단계로 나누어 ‘문자코드 표준화 연구’를 진행해 왔으며, 각 단계별로 그 운영 목표에 부합하도록 원활하게 진행되었다.

2007년도는 제3단계 마지막 해에 해당하므로 그 동안 본 센터에서 진행해 온 사업 내용과 진행과정을 체계적으로 정리하였다. 특히 비표준문자의 국제 표준화 등록 현

황과 과정, 관련 기록과 자료를 유형별로 분류하여 시간순으로 정리함으로써 연구자들에게 편의를 제공하고자 하였다. 그리고 국제 문자코드 표준화의 전망과 이에 대응하기 위해 향후 국가적 차원에서 반드시 진행해야할 과제를 정리하고자 하였다.

4) 옛한글의 효율적 활용을 위한 지원체계 확립

본 센터에서는 지난 2006년도 7월 두 차례에 걸쳐 개최된 옛한글정보처리연구위원회 회의를 통해 국제 표준코드에 제안하기 위한 117자의 <옛한글자모 확장목록>을 최종 결정하였으며, ISO에 공식적으로 제안되었다.

이렇게 제안된 <옛한글자모 확장목록>은 2007년 4월 23일부터 27일까지 독일 프랑크푸르트에서 개최된 ISO/IEC의 JTC1/SC2/WG2 제50차 회의에서 유니코드 BMP 영역에 추가로 등록되었음을 공식적으로 확인하였다.

5) 문자코드 정보 제공 및 지원체계 강화

- ① 산업계 및 관련기관 지원
- ② 웹사이트 운영
- ③ 소식지 간행(2회) : 19호(7월), 20호(12월)
- ④ 중간보고서 발간(1회)
- ⑤ 종합결과보고서 발간(1회)

5. 연구결과의 활용

1. 수집된 비표준한자는 자체 검정과 정리 작업을 거쳐 IRG에 <Ext.DUN 한국측 제안한자 목록>으로 제안함으로써 국제 표준화에 기여할 것이다.
2. <Ext.B 한국측 한자>와 <Ext.D 한국측 제안한자>에 대한 트루타입 글꼴은 학술계·출판계·정보산업계 등 여러 분야에서 국제 표준한자를 편리하게 활용하는 데 도움을 줄 수 있을 것이다.
3. <Ext.DUN 한국측 제안한자 목록> 작성은 한자의 국제 표준화에 기여할 수 있을 뿐만 아니라 국제 표준화 활동에서 한국의 위상을 제고할 수 있을 것이다. 그리고 향후 국제 표준규격으로 채택이 될 경우, 우리나라의 문자자원을 충분히 활용할 수 있어 경비 절감 효과를 가져 올 수 있을 것이다.
4. 본 센터에서 그동안 진행한 ‘문자코드 표준화 연구’ 사업의 정리 및 향후 과제와 전망은 국제 문자코드 표준화에 대응하기 위해 국가적 차원에서 반드시 진행해

야할 과제를 제시함으로써 국가 정책에 활용할 수 있을 것이다.

5. ISO에 공식적으로 제안되어 유니코드 BMP 영역에 추가로 등록된 <옛한글자모 확장목록> 117자는 향후 옛한글을 완전하게 전산적으로 처리할 수 있을 것이며, 유니코드 환경에서 옛한글을 검색할 수 있는 기초를 마련한 것이라 할 것이다.
6. 센터 홈페이지 운영 및 소식지 발간은 학계·출판계·정보산업계 관계자에게 문자 코드 관련 정보를 접근하고 활용하는 데 도움을 줄 수 있을 것이다.

목 차

요약문	i
목 차	vii
제1장 연구의 목적 및 필요성	1
제1절 연구의 목적	1
제2절 연구의 필요성	2
제2장 연구의 최종 목표 및 당해년도 목표	3
제1절 3단계(2004~2007) 최종 목표	3
제2절 당해년도(2007년) 목표	3
1. 비표준문자 수집·정리 및 국제 표준화 지원	3
2. 문자코드 표준화 연구	3
3. ‘문자코드 표준화 연구’사업 정리 및 향후 과제와 전망	5
4. 옛한글의 효율적 활용을 위한 지원체계 확립	8
5. 문자코드 정보 제공 및 지원체계 강화	8
제3장 현재까지의 추진실적(1998~2006)	9
제4장 연구 결과	18
제1절 비표준문자 수집·정리 및 국제 표준화 지원	18
제2절 문자코드 표준화 연구	18
1. <Ext.B 한국측 한자> 트루타입 폰트 제작 지원	18
2. <Ext.D 한국측 제안한자> 트루타입 폰트 제작 지원	19
3. <Ext.DUN 한국측 제안한자> 목록 작성	19
4. <Ext.DUN 한국측 제안한자> 한자정보 작성	20
5. <Ext.DUN 한국측 제안한자> IDS 정보 작성	20
6. <Ext.DUN 한국측 제안한자> 트루타입 폰트 제작	20
7. <Ext.DUN 한국측 제안한자> 출전전거 원문 이미지자료 구축	21
8. 전문가 자문 및 자문회의	21
제3절 ‘문자코드 표준화 연구’사업 정리 및 향후 과제와 전망	22
제4절 옛한글의 효율적 활용을 위한 지원체계 확립	23
제5절 문자코드 정보 제공 및 지원체계 강화	24
1. 소식지 발간	24
2. 웹사이트 운영	24
3. 산업계 및 관련기관 지원	25

[부록]

※부록1 Ext.B 한국측 한자 트루타입 폰트	29
※부록2 Ext.D 한국측 제안한자 트루타입 폰트	33
※부록3 Ext.DUN 한국측 제안한자 목록	37
※부록4 Ext.DUN 한국측 제안한자 트루타입 폰트	215
※부록5 Ext.DUN 한국측 제안한자 출전전거 원문 이미지자료(부분)	227
※부록6 전문가 자문 및 자문회의 회의록	281
※부록7 21세기 세종계획 문자코드 표준화 연구 사업정리 및 향후 과제와 전망	285
※부록8 ISO/IEC의 JTC1/SC2/WG2 회의 문서 N3242	367
※부록9 문자코드연구센터 소식지 제19호·제20호	379

제1장 연구의 목적 및 필요성

제1절 연구의 목적

본 센터의 목적은 국가 제정 표준코드에 아직 수용되지 않은 비표준문자를 등록하고 처리하는 방법을 표준화하는 한편, 유니코드와 같은 대규모 국제적 표준코드의 원활한 수용과 활용에 필요한 지식, 기술, 자원을 개발하여 학술·문화·정보산업·출판 분야의 각 기관과 연구자들에게 지원하는 것이다.

바야흐로 국제적 표준코드가 실용화되었지만 세계 각국의 특수성으로 인하여 표준코드만으로 연구, 교육, 출판 및 정보처리에 필요한 모든 문자들을 소화할 수는 없다. 특히 한국·중국·일본과 같은 한자문화권 국가에서는 한자 자체의 특수성 때문에 더욱 그러하다. 특히 한국의 경우는 한자와 더불어 훈민정음 창제 이래 한글 자소와 표기법이 다양하게 변천해 오는 이중적 문자 사용의 역사적 흐름을 겪었기에, 표준규격 이외의 문자를 정보처리에 수용하는 문제가 다른 나라보다 더 절실하다. 또한 한국은 중국·일본과 달리 비표준한자의 컴퓨터 사용 환경 내 수용이라는 공통적인 과제에 덧붙여, 자체 표기 체계인 한글과 이두, 구결문자의 입출력 문제를 해결해야 한다.

한글, 한자, 이두, 구결문자 등 국어의 활용 자형 범위는 실로 광범위하며, 표준코드 이외의 문자가 앞으로도 발견될 가능성은 매우 크다. 따라서 새로 찾아내는 문자들을 효과적으로 수집하고 기술 규격 등을 표준화하여 보급함으로써, 개인과 기관마다 별도의 방법으로 정보화를 추진하는 데서 오는 혼란을 방지할 필요가 절실하다. 또한 비표준문자 등록 및 표준화를 위한 창구를 단일화, 집중화함으로써 비표준문자 수집에 소요되는 비용과 시간을 절약할 수 있고, 개별적으로 처리된 자료를 교환·공유할 수 있다.

이와 함께 우리에게 주어진 또 하나의 과제는 새로운 국제 표준코드의 원활한 수용과 운용을 위한 국가적 차원의 지식, 기술, 자원을 확보하는 일이다. 국제 표준코드는 국제적 범용성이 일차적으로 고려되기 때문에 경우에 따라 우리의 어문 생활에서 불편한 측면이 적지 않다. 따라서 이를 우리의 어문 생활과 정보 환경에서 원활하게 운용하기 위해서는 우리의 실정에 맞는 지식 및 기술 기반이 구축되어야 한다.

이러한 과제들을 모든 개인과 기관들이 개별적, 분산적으로 처리할 경우 개별적, 임시적 처리에 따른 인력, 시간, 재정의 낭비, 추가된 문자 세트가 서로 다른 데 기인하는 자료 호환성의 결여, 비전문적 분석과 서체 제작으로 인한 정보 처리상의 혼란 등이 야기될 수 있다.

문자코드연구센터는 이러한 난점과 비능률을 해소하고, 국어 정보처리에 관련된 각급 기관과 정보산업체 및 개인들의 생산성을 높이는 <문자코드에 관한 정보 수집, 연

구 및 지원의 국가적 네트워크 중심>으로서 활동하게 된다. 그리하여 국어정보 학술 활동의 효율을 높이며, ISO 등의 국제적 부호계 표준 활동에 긴요한 지적 자원 기반 역할을 담당할 수 있다. 나아가 이러한 성과를 토대로 유니코드 환경에서 우리 문자 생활이 원활하게 이루어질 수 있도록 하는 데 기여하고자 한다.

제2절 연구의 필요성

컴퓨터는 서구권에서 영문 문자 활용을 위주로 개발되어 왔다. 따라서 한글과 한자 등 우리 문자문화를 수용하기 위해서는 부호체계를 비롯한 각종 환경과 기능상의 여러 가지 문제점을 선결해야 한다. 이러한 상황에서 한글에 대한 국어학적인 연구가 충분히 수행되지도 않은 시점에서 국제 표준 조류에 성급하게 한글을 수용함으로써, 현재의 표준부호계는 결과적으로 국어 생활과 연구에 불충분한 부호계로 머물고 말았다. 이를 시급히 보완하기 위해서는 우리 역사와 문화 분야에 나타나는 다양한 문자들의 쓰임새와 字音, 字義 등 관련 정보에 대한 국가 차원에서의 체계적인 파악을 통하여 충분한 자료를 확보함으로써 국어 정보 생활을 지원하고, 국제 표준화 활동을 위한 상시적 기반을 육성할 필요가 있다.

표준부호계에 수용된 문자만으로 일상적인 언어생활을 영위하는 데에 큰 지장은 없다고 볼 수도 있다. 그러나 좀더 주의를 기울여 본다면 학생들의 국어교육 면에서만 하더라도 비표준문자들은 사용되며, 나아가 한글 및 한자 등 우리의 전래 문자체계를 바탕으로 이루어지는 전문적인 연구와 자료 전산화 등에는 보다 확장된 부호계가 절실하게 요구된다.

그 동안 수차례의 연구나 표준 제정 작업 등을 통해서 우리의 문헌에 등장했던 문자들에 대한 조사가 상당히 이루어졌고, 그 결과는 표준화의 초안 단계까지 이르게 되었지만, 아직도 조사할 내용이 많고 등록되지 않은 문자도 방대하게 남아 있다. 이 문제는 임시방편적이고 단편적인 접근으로 결코 해결될 것이 아니다. 보다 적극적으로 이 작업을 상설기관에서 주관하여, 자료를 모으고, 학자들이나 일반인들의 제안을 받고, 수집된 정보에 대해서는 즉각적으로 공표하여 2중, 3중의 중복 연구를 피하게 하는 일이 필요하다.

이와 함께 새로운 국제 표준코드의 원활한 수용과 운용을 위한 국가적 차원의 지식, 기술, 자원을 확보하는 일이 긴요한 과제도 대두되었다. 이 문제 역시 중복 작업 및 투자를 피하기 위해서는 문자코드 연구에 관한 경험이 풍부한 상설기관이 주관하여, 우리의 어문 생활과 정보 환경에서 국제 표준코드가 원활하게 운용될 수 있는 지식 및 기술 기반을 확보, 지원해야 한다.

제2장 연구의 최종 목표 및 당해년도 목표

제1절 3단계(2004~2007) 최종 목표

본 센터 운영의 제3단계에 해당하는 4년간(2004~2007)의 목표는 1단계(1998~2000), 2단계(2001~2003) 성과를 기반으로 새로운 국제 표준 코드인 유니코드의 원활한 운용을 위한 국가적 차원의 지식과 기술을 확보하여 학술·문화·정보산업·출판 분야 등 각 기관과 연구자들에게 지원하는 데 있다. 이를 통해 유니코드 환경에서 우리 어문 생활이 원활하게 이루어질 수 있도록 하는 데 기여하고자 한다.

본 센터 운영의 제3단계(2004~2007) 제4차년도에 해당하는 2007년도에는 기존의 연구 성과를 바탕으로 다음과 같은 연구를 진행하고자 한다.

- 비표준문자 수집·정리 및 국제 표준화 지원
- 문자코드 표준화 연구
- ‘문자코드 표준화 연구’ 사업 정리 및 향후 과제와 전망
- 옛한글의 국제 표준화를 위한 지원체계 확립
- 문자코드 정보 지원 체계 강화

제2절 당해년도(2007년) 목표

1. 비표준문자 수집·정리 및 국제 표준화 지원

본 센터에서 제3단계 사업 기간인 2004년부터 2006년까지 총 15,269자의 비표준문자를 수집하여 국제표준화를 지원하였다.

2007년도에는 2006년도 <국가지식정보자원관리사업>에서 산출된 신출한자를 사업에 참여하고 있는 관계 기관과 협의하여 추가로 수집할 예정이며(1,000여 자 내외 예정), 이를 정리하여 KIRG에 제출함으로써 향후 국제 표준코드인 유니코드 Ext.E 한자에 제안할 수 있도록 지원한다.

2. 문자코드 표준화 연구

① <Ext.E 한국측 제안한자 목록> 작성

본 센터에서는 2006년도에 <Ext.C2 한국측 제안한자 목록>을 작성하여 KIRG에 제출하였으나, IRG의 ‘Ext.C1 한자’ 4,219자가 Ext.C로 바뀌고, ‘Ext.C2 한자’ 15,000여 자가 Ext.D로 그 명칭이 바뀌었다. 그리고 ‘Ext.D 한자’는 더 이상 제안을 받지 않는 등

IRG의 정책 변화로 2006년도에 본 센터에서 작성한 <Ext.C2 한국측 제안한자 목록>은 <Ext.E 한국측 제안한자 목록>으로 제안할 수밖에 없는 상황이 되었다. 2007년도에는 2006년도에 작성한 <Ext.C2 한국측 제안한자 목록>에 자수를 추가하고 보완하여 <Ext.E 한국측 제안한자 목록>을 작성하여 비표준문자의 국제 표준화를 지원한다.

지난 2006년도에 총 15,269자의 비표준한자 중 5,980자의 신출한자를 확인하였고, 이 가운데서 출전정보와 전거자료 이미지를 확인할 수 있는 한자 1,175자를 추출하여 <Ext.C2 한국측 제안한자 목록>으로 작성하였다. 2007년도에는 2006년도에 목록을 작성하고 남은 한자 4,561자와 2007년도에 새로 수집하는 비표준 한자 가운데서 출전정보와 전거자료 이미지를 확인할 수 있는 한자를 추가로 확인하여 <Ext.E 한국측 제안한자 목록>을 보완하고자 한다. 현재 4,561자의 한자 중에 출전정보와 전거자료 이미지가 확인 가능 여부를 그룹별로 분류하면 다음 표와 같다.

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 당초 신출한자 목록을 넘겨받은 해당 기관과의 협의 하에 출전정보와 전거자료 이미지를 확보할 수 있을 것으로 판단된다. 그리고 2007년도에 신규로 수집하는 비표준한자와 통합하여 신출한자 여부를 검정하여 최종 후보목록을 추출한다. 이렇게 추출된 신출한자는 먼저 IRG의 ‘Annex S’ 문서 기준에 의거하여 이체자를 통합·선별하고, 본 센터에서 2004년과 2005년에 걸쳐 개발한 <SuperCJK 14.0 검색시스템>을 활용하여 유니코드 BMP 영역 및 Ext.A·Ext.B·Ext.C 영역에 이미 등록되어 있는지를 검정하여 <Ext.E 한국측 제안한자 목록>으로 통합·정리한다.

이상과 같이 자형에 대한 검정 과정을 마친 신출한자는 “국제 표준 제안한자 양식 (IRG N881)”에 맞추어 필요한 정보를 확인하여 입력한다. 여기에는 다음 표에서 보는 것처럼 ‘강희사전 위치정보’·‘강희사전 부수정보’·‘잔여획수’·‘오필획 정보’·‘해당 한자의 원문정보’·‘이체자 정보’ 등을 필수로 입력한다. 이렇게 작성된 신출한자 목록을 KIRG에 제출하여 향후 국제 표준코드인 유니코드 Ext.E 한자에 제안할 수 있도록 지원하고자 한다.

[IRG N881 문서 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 SSNNNNNN (소스 정보)	Variant Ucode U1(U2) (이체자 정보)
1-8(byte)	9-13	14-15	16	17-24	25-35

② 전문가 자문회의

이상에서 작성된 <Ext.E 한국측 제안한자 목록>은 해당 분야의 전문가에게 감수를 의뢰하고, 자문회의를 통해 보다 구체적으로 검토한다.

3. ‘문자코드 표준화 연구’ 사업 정리 및 향후 과제와 전망

본 센터는 1998년부터 제1단계(1998~2000), 제2단계(2001~2003), 제3단계(2004~2007)의 3단계로 나누어 ‘문자코드 표준화 연구’를 진행해 왔으며, 각 단계별로 그 운영 목표에 부합하도록 원활하게 진행되었다. 각 단계별로 진행된 성과를 살펴보면 다음 표와 같다.

단계	연도	연구 성과 및 실적
1단계	1998년	·자문기구 구성 ·연구 설비 구축 ·등록·관리 체계 수립 ·비표준문자 처리 방안 연구 ·센터 웹사이트 설치 ·기술지침서 발간 ·연차 보고서 발간
	1999년	·한자 정보 데이터베이스(48,027 자) 구축 ·옛글자 5,291자 확보 ·옛한글 정보화 산학협동 심포지엄 개최 ·뉴스레터 발간(2회) ·기술지침서 발간 ·종합 보고서 발간

	2000년	<ul style="list-style-type: none"> ·옛한글 기초 문헌 데이터베이스 구축 ·48,000여 한자의 자형 제작 ·한자 정보 데이터베이스 제공 ·뉴스레터 발간(2회) ·종합 보고서 발간
2단계	2001년	<ul style="list-style-type: none"> ·<옛한글 정보 데이터베이스> 용례 정보 추출 : 2,237개 자형 ·수집·정리된 6,143자(옛한글 6자 포함) 처리 ·한자정보DB훈독 정보 추가 : 총 47,435자 ·뉴스레터 발간(2회) ·종합 보고서 발간
	2002년	<ul style="list-style-type: none"> ·옛한글 문헌 추가 수집 : 총 755,320어절 ·기 구축 자료 교정 : 총 38건 ·워크숍 개최 ·산학협동 심포지엄 개최 ·옛한글정보처리연구위원회 구성 및 회의 개최 ·비표준문자 수집 및 등록 : 총 2,264자 ·뉴스레터 발간(2회) ·사업보고서 발간
	2003년	<ul style="list-style-type: none"> ·옛한글 문헌 수집 및 분석 : 총 473898 어절 ·옛한글 자모 세트 선정 및 배열 방안 워크숍 개최 ·옛한글 표준화 방식 및 규범 시안 제안 ·유니코드 한자(V3.0) 통계 분석을 위한 기초연구 : 총 11,547,746자에 달하는 한문 문헌자료 수집 ·고전 전산화를 위한 한자 이체자 처리 방안 산학협동 심포지엄 개최 ·SuperCJK 통계 추출기 개발 ·통계 결과 및 통계 분석 제안 ·수집된 비표준한자 국제 표준화(Ext. C1) 지원 ·뉴스레터 발간(2회) ·결과 보고서 발간
3단계	2004년	<ul style="list-style-type: none"> ·비표준문자의 수집·등록 및 분석 : 총 1,096자 수집, 신출한자 432자 추출 ·제22차, 제23차 IRG회의 참석 ·SuperCJK Ext.C1 제출한자의 정보 교정·수정 ·SuperCJK Ver.14.0 한자 자형 및 정보 데이터베이스 구축 ·<SuperCJK 14.0 검색시스템> 개발 ·<신출한자의 처리와 국제 표준화> 워크숍 개최 ·국가표준코드의 옛한글 코드 체계 및 표현 방식 분석 ·응용소프트웨어의 옛한글 코드 체계 및 표현 방식 분석 ·옛한글 자료의 코드 표현 기본 원칙 및 방안 수립 ·옛한글 자료의 코드 표현 기본 원칙 및 지침 인준

	<ul style="list-style-type: none"> ·소식지 : 13호, 14호 발간·배포 ·종합보고서 발간·배포 ·관련 기관 지원 : 3회
2005년	<ul style="list-style-type: none"> ·비표준문자 수집 : 총 10,247자 수집, 신출한자 4,787자 추출 ·제24차, 제25차 IRG회의 참석 ·기호, 특수문자 및 한자의 말뭉치 입력과 변환의 표준화 연구 ·SuperCJK Ver.14.0 한자 字音 데이터베이스 구축 ·SuperCJK 14.0 한자 字義 데이터베이스 구축 ·SuperCJK 14.0 四角號碼 데이터베이스 구축 ·<SuperCJK 14.0 검색시스템> 확대·개편 및 웹버전 개발 ·<고전 전산화 사업의 異體字 인코딩 및 DB 구축 방안> 심포지엄 개최 ·주요 워드프로세서의 옛한글 코드 체계 분석 ·주요 워드프로세서의 옛한글 입출력 방식의 비교 분석 ·센터 홈페이지 전면 개편(http://www.code.re.kr) ·소식지 15호, 16호 발간·배포 ·종합보고서 발간·배포 ·관련 기관 지원 : 3회 ·문자코드 처리 관련 교육 실시 : 1회
2006년	<ul style="list-style-type: none"> ·비표준문자 수집 : 3,926자 수집 ·제27차 IRG 회의 참석 : Ext.C 한자 최종 확인 ·Ext.C1 한자 검토 ·Ext.C1 등록 한국측 한자 IDS 정보 입력 지원 ·Ext.C1 등록 한국측 한자 트루타입 폰트 제작 및 출전전거 이미지자료 구축 지원 ·<Ext.C2 한국측 제안한자 목록> 작성 ·옛한글정보처리연구위원회 회의 개최 : <옛한글자모 확장목록> 결정 및 국제 표준기구 제안 ·웹사이트 운영 ·소식지 발간 : 제17호(7월), 제18호(12월) ·산업계 및 관련기관 지원 : 3회 ·종합보고서 발간·배포

2007년도는 제3단계 마지막 해에 해당하므로 그 동안 본 센터에서 진행해 온 사업 내용과 진행과정을 체계적으로 정리하고자 한다. 특히 비표준문자의 국제 표준화 등록 현황과 과정, 관련 기록과 자료를 유형별로 분류하여 시간순으로 정리함으로써 연구자들에게 편의를 제공하고자 한다. 그리고 국제 문자코드 표준화의 전망과 이에 대응하기 위해 향후 국가적 차원에서 반드시 진행해야할 과제를 정리하고자 한다.

4. 옛한글의 효율적 활용을 위한 지원체계 확립

본 센터에서는 지난 2006년도 7월 두 차례에 걸쳐 개최된 옛한글정보처리연구위원회 회의를 통해 국제 표준코드에 제안하기 위한 <옛한글자모 확장목록>을 최종 결정하였으며, ISO에 공식적으로 제안되었다. 제안된 안은 지난 9월에 논의가 되었으나, 유니코드 컨소시엄의 반대로 그 결정이 2007년 4월로 연기되었다.

2007년도에는 2006년도에 제안된 <옛한글자모 확장목록>의 국제 표준코드 채택 일정을 확인하고, 필요한 자료의 요구가 있을 경우에는 이를 적극적으로 지원할 수 있도록 한다.

5. 문자코드 정보 제공 및 지원체계 강화

- ① 산업계 및 관련기관 지원
- ② 웹사이트 운영
- ③ 소식지(2회)
- ④ 중간보고서(1회)
- ⑤ 종합결과보고서(1회)

제3장 현재까지의 추진실적(1998~2006)

제1절 1차년도(1998)

1998년도 사업은 1단계 3개년 사업의 제1차년으로서 전체 사업의 완수를 위한 기반 조성 및 체제 마련이 우선적인 목표였다. 이를 위해 본 센터에서는 1998년도 사업을 다시 3단계로 나누어 체계적인 작업 관리를 기획하고자 하였다.

- 1단계 사업 : 운영체제의 준비 기간으로서 자문기구를 구성하고, 연구 설비를 구축하였으며, 등록·관리 체제 수립 등이 이루어졌다.
- 2단계 사업 : 방법론 연구 및 기초자료 수집 기간으로서 기초자료 수집, 국제정보 수집, 비표준문자 처리 방안 연구를 수행하였다.
- 3단계 사업 : 지속적 조사 및 대외지원·홍보 기간으로서 센터 웹사이트를 설치하였고, 기술지침서 및 연차 보고서를 발간하였다.

전체적으로 각 단계별 사업이 ‘운영체제의 구축과 지원 기반 조성’이라는 1차년도 운영 목표에 부합하도록 원활하게 진행되었다. 또한 5월과 11월에는 본 센터의 활동과 비표준문자 관련 정보를 수록한 뉴스레터가 발간되었다.

제2절 2차년도(1999)

2차년도(1999)의 운영목표는 정보 자원의 확충과 등록·지원 업무의 활성화에 있었다. 등록 및 지원, 조사 및 연구, 보급 및 홍보의 각 세부 항목 별로 그 실적을 정리해 보면 다음과 같다.

- 등록 및 지원 : 웹사이트의 개발·운영을 통한 비표준문자 등록 체제 구축
- 조사 및 연구
 - <한자정보 데이터베이스> 구축(48,027자)
 - <옛한글 정보 데이터베이스> 구축을 위한 한글 옛글자 5,291자 확보
 - 옛한글 정보화 산학협동 심포지엄(6월) 개최
- 보급 및 홍보
 - 뉴스레터 발간·배포 2회(5, 11월)
 - 기술지침서 발간·배포 1회
 - 종합 보고서 발간(11월)

제3절 3차년도(2000)

3차년도(2000)의 운영목표는 운영 체계의 고도화와 체계적·능동적인 데이터베이스 지원에 있었다. 웹사이트의 기능 확장에 따라 기 구축된 데이터베이스는 능동적 지원 방식을 갖추었으며, 보다 체계적인 등록·지원과 보급·홍보를 통해 비표준문자등록센터의 1단계 사업성과를 정리하였다.

1. <옛한글 정보 데이터베이스>의 구축·활용

- 15세기부터 1920년대 근대 국어자료에 이르기까지 옛한글의 다양한 표기형태와 자형을 조사
- 대규모 <옛한글 기초 문헌 데이터베이스> 구축
 - 문헌수 : 1,231종
 - 총음절수 : 21,831,897음절
 - 총옛글자수 : 3,324,952자
- 조사 자료의 전자화된 텍스트에 기반한 통계 분석
- 빈도 통계 : 전체 빈도 / 시대별 빈도 / 문헌성격유형별 빈도 / 장르별 빈도 / 각 시대 문헌성격 유형별 빈도 / 각 시대 유형별 빈도
- 대응 분석을 통한 상관관계 검증 : 시대별, 문헌 성격, 유형별, 장르별, 각 시대 문헌 성격 유형별, 각 시대 장르별 옛한글 출현 빈도의 상관관계 검증

2. <한자 정보 데이터베이스>의 확충 및 응용

- <한자정보 데이터베이스>의 독음 정보 확충 : 데이터베이스 포함 한자 총 51,028자의 약 98%에 해당하는 47,145자의 한글 독음 정보를 국내외 자전을 참고하여 구축하고, 그 전거를 명시함
- SuperCJK 포함 문자정보와의 대응 : 『강희자전』·『한어대사전』 등의 전통 자전의 정보와 베트남·홍콩의 국가 표준 코드까지 포괄, 데이터베이스 포함 정보의 확장
- 48,000여 한자 자형의 BMP·GIF 자형(40×40 규격) 제작, 제공
- <한자정보 데이터베이스>제공과 부수·독음·총획수에 의한 검색 지원

3. 2000년형 워드프로세서 제품군의 코드체계와 문자처리에 관한 연구

- 각 제품군의 확장한자와 옛한글, 구결 등의 구현 범위 조사
- 각 제품군의 옛한글 입출력 방식 비교 검토
- 각 제품군 간의 호환 정도 실험
- 각 제품군의 규격 문서를 HTML 문서로 변환하여 <MS Explorer>와 <Netscape>의

최신 버전에서 불러들일 수 있는지 실험. 이를 통해 각 제품군의 유니코드 호환 여부 판정

4. 보급 및 홍보

- 웹사이트 기능의 확충·확장 : 게시판·자료실의 확장, 관련 사이트 보완
- 뉴스레터 발간 : 2회(5월, 11월호)
- 종합 보고서 발간 : 12월
- 비표준문자의 등록 및 활용 : 팩스나 편지 등을 통한 서면 등록, 연구센터의 방문을 통한 직접 등록, 인터넷 웹사이트를 통한 온라인 등록
- 보급·홍보를 위한 강연 활동과 기업체에 각종 정보 제공

제4절 4차년도(2001)

1. <옛한글 정보 데이터베이스>의 세부 정보 확충 및 인터넷 서비스

- <옛한글 정보 데이터베이스>의 대상 문헌 확대
- 전문적 통계적 방법에 기반한 통계분석
- 각 자형의 실제 용례 정보를 추출 : 2,237개 자형
- 이미 구축된 다양한 정보를 Website와 연동하여 제공

2. 수집된 비표준문자 처리

- 1단계 3년간 수집·정리된 4,634자(옛한글 6자 포함)를 국제 표준 코드 체계 내에서 처리
- 2001년도 수집된 문자 1,509자를 국제 표준 코드 체계 내에서 처리
- 처리 결과 : 수집·정리된 한자 6,137자 가운데 이미 구축한 <한자 정보 데이터베이스>에 포함되지 않은 글자 1,099자를 선별
- 처리결과를 IRG 회의에서 제기할 Extension C 영역에 우리 측 자료로 활용할 수 있도록 한국한자특별전문위원회(KIRG)에 보고
- 본 센터 수집 비표준한자 가운데 CNS에 포함되지 않은 유니코드 문자의 음·훈 조사

3. <한자정보 데이터베이스>의 보완 및 확충

- 세부 정보 확충의 일환으로 새롭게 훈독 정보를 추가 : 자형이 확보된 48,027자 가운데 총 47,435자의 음·훈 확보
- 다양한 코드 정보와 함께 인터넷 전자사전의 기능을 구비

- user interface와 검색 기능 개선

4. 비표준문자 처리 방안의 보급·홍보

- 웹사이트 기능의 확충·확장 : 게시판·자료실 확장, 관련사이트 항목 보완
- 뉴스레터 발간 : 2회(6월, 11월)
- 종합 보고서 발간 : 11월
- 비표준문자의 등록 및 활용 : 팩스나 편지 등을 통한 서면 등록, 연구센터의 방문을 통한 직접 등록, 인터넷 웹사이트를 통한 온라인 등록
- 산업체 지원 2회

제5절 5차년도(2002)

1. 옛한글의 국제 표준화 방안 연구

- 기초 문헌 수집 및 분석
 - 옛한글 문헌 추가 수집 : 총 755,320어절
 - 기 구축 자료 교정 : 총 38건
- 기초 연구
 - 국내에 통용되는 운영체제 및 문서편집기의 옛한글 코드 분석
 - 옛한글의 자모 선정 문제
 - 한글 자모의 배열 순서 문제
- 워크숍 개최
- 산학협동 심포지엄 개최
- 옛한글정보처리연구위원회 구성 및 회의 개최

2. 수집된 비표준문자의 국제표준화 지원

- 비표준문자 수집 및 등록 : 총 2,264자
- 수집된 비표준문자의 국제표준화 지원
- 전체 작업 범위 : 총 3,363자
 - 1단계 수집 한자 : 1,099자
 - 2002년 수집·등록 한자 : 2,264자
- Super CJK Ver.14.0과 대비 : 총 794자 추출
 - 1단계 수집 한자 : 1,099자에서 235자 추출
 - 2002년 수집·등록 한자 : 1,519자에서 559자 추출
- Ext.C1 제출 목록 작성 후 한국한자특별전문위원회(KIRG)에 제출

- 한국한자특별전문위원회(KIRG) 지원 : Ext.C1의 제안 한자의 재교정 작업에서 한국과 북한이 나누어 맡은 부분(7,000자)의 1차 교정 작업 지원.

3. 비표준문자의 정보 지원 및 센터 홍보

- 비표준문자 정보 지원
 - 관련 기관 지원 : 민족문화추진회(SuperCJK 관련 자료 제공)
 - 한국한자특별전문위원회(KIRG)에 비표준한자 제출.
- 뉴스레터·간행물 발간 : 9호(5월), 10호(11월) 제작·발간.
- IRG 회의 참가
- 종합 보고서 발간 : 11월
- 四角號碼 사용법 강좌 개최

제6절 6차년도(2003)

1. 비표준문자의 수집 및 국제표준화 지원

- 국학진흥원 유교정보화사업(1~3차분) 신출한자 211자 수집
- 수집된 비표준한자 국제 표준화(Ext.C1) 지원
- 한국한자특별전문위원회(KIRG) 지원

2. 옛한글 정보처리 표준화 방안 연구

- 옛한글 기초 문헌 수집 : 신문 및 고소설 자료 473,898어절
- 옛한글 자모 세트 선정 및 배열 방안 워크숍 개최
 - 옛한글 표준화 방식, 범위 자모세트, 배열 방안 등에 대한 표준화 시안 제출
- 옛한글 표준화 방식 및 규범 시안 제안
 - 명칭 : ‘옛한글’로 함. 영문표기는 ‘Old Hangeul’로 함.
 - 옛한글 자모의 범위 및 목록
 - 옛한글 배열 순서 기준 마련
 - 옛한글 표현 방안 마련

3. 유니코드 한자(V3.0) 통계 분석을 위한 기초연구

- 통계 설계 및 자료 수집
 - 사서류·문집류·의궤류·일기류 등에서 11,547,746자 수집
- UnicodeCJK 통계 추출기 개발
- 통계 결과 추출

- 출현글자 : 유니코드(V3.0) CJK 한자 27,484자 가운데 11,409자의 용례 확인 (BMP 10,510자, Ext.A 648자, Compatibility 251자)
- 빈도 순위표 작성
 - 순위별 빈도를 백분율로 환산할 경우 순위 5000등까지의 한자가 전체의 99.55%를 차지

4. 고전 전산화를 위한 한자 이체자 처리 방안 심포지엄 개최

- 학계를 중심으로 정책 당국(문화관광부, 기술표준원 담당자, 국내문자코드위원회 위원장, 국립국어연구원), 정보산업계가 참가하는 심포지엄을 개최하여 주요 현안들에 대한 해결 방향을 모색하고, 공감대를 형성

5. 비표준문자의 정보 지원 및 센터 홍보

- 뉴스레터·간행물 발간 : 11호(5월), 12호(11월) 제작·발간.
- 제21차 IRG 회의 참가 : 2003년 11월 17일~21일(中國 桂林)
- 관련 기관 지원 : 한국정신문화연구원에서 ‘고전 전산화에 있어서 이체자 처리 문제 및 방법’ 발표
- 종합 보고서 발간 : 11월

제7절 7차년도(2004)

1. 비표준문자의 수집 및 국제표준화 지원

- (주)서울시스템 등록 『사마방목』 신출한자 745자
- (주)솔트웍스 등록 ‘고전국역총서 및 한국문집총간’ 신출한자 351자
- 제22차 IRG 회의 참석 : 2004년 5월 24일~28일(중국 청도)
- 제23차 IRG 회의 참석 : 2004년 11월 28일~12월 3일(대한민국 제주도)

2. 유니코드 한자의 효율적 활용을 위한 조사·연구

- SuperCJK Ver.14.0 한자정보 데이터베이스 구축
- SuperCJK Ver.14.0 한자자형 데이터베이스 구축
- SuperCJK Ver.14.0 검색기 개발
- <신출한자의 처리와 국제 표준화> 워크숍 개최

3. 옛한글의 효율적 활용을 위한 방안 연구

- 국가표준코드의 옛한글 코드 체계 및 표현 방식 분석
- 응용소프트웨어의 옛한글 코드 체계 및 표현 방식 분석
- 옛한글 자료의 코드 표현 기본 원칙 및 방안 수립
- <옛한글정보처리연구위원회> 회의 개최
- 옛한글 자료의 코드 표현 기본 원칙 및 지침 인준

4. 문자코드 정보 지원 체계 강화

- 뉴스레터 발간·배포 : 13호(6월), 14호(12월)
- 한국전산원 주최 신출한자 처리회의 참석 발표
- 고려대장경연구소, (주)오픈에스이 신출한자 처리 관련 기술자문 및 정보제공
- 한국한자특별전문위원회(KIRG) 회의 참석
- 종합 보고서 발간 : 12월

제8절 8차년도(2005)

1. 비표준문자의 수집·등록 및 국제 표준화 지원

- <국가지식정보자원관리사업>의 고전적 전산화 과정에서 발견된 10,247자를 수집하여 신출한자 4,787자 추출
- 제24차 IRG회의 참석 : 2005년 5월 24일~27일(일본 京都)
- 제25차 IRG회의 참석 : 2005년 11월 28일~12월 2일(미국 버클리)

2. 기호, 특수문자 및 한자의 말뭉치 입력과 변환의 표준화 연구

- 유니코드 유사 형태의 기호, 특수문자 및 한자 동형자를 조사
- 동형-유사 부호 사용실태 통계 : 1998~2004년간 21세기 세종계획 사업에서 구축된 기초자료 말뭉치 약 1,100MB, 특수자료 말뭉치 약 328MB
- “유사 형태의 기호, 특수문자 및 한자 동형자 처리 권장안” 작성

3. 유니코드 한자의 효율적 활용을 위한 체계적 조사·연구

- SuperCJK Ver.14.0 한자 字音 데이터베이스 구축 : 유니코드 한자 Ext.B 영역까지 국내외 사전 수록 한자음 조사
- SuperCJK 14.0 한자 字義 데이터베이스 구축 : 유니코드 한자 Ext.A까지의 한자 27,484자 중 26,269자에 대한 자의 정보 확인·정리
- SuperCJK 14.0 四角號碼 데이터베이스 구축 : 四角號碼法은 글자 형태의 특성을 보고 특정한 기호를 취하는 형태소 인식법으로, <SuperCJK 14.0 검색시스템>에서

사각호마 검색방법을 지원하기 위해 유니코드 Ext.A까지의 한자 27,484자에 대한 5자리의 사각호마 정보를 구축

- <SuperCJK 14.0 검색시스템> 확대·개편 및 웹버전 개발 : 2004년도 개발한 <SuperCJK Ver.14.0 검색시스템>의 웹 버전으로 字音·字義·四角號碼 정보가 추가 되었으며, 센터 홈페이지(<http://www.code.re.kr>)를 통해 제공
- <고전 전산화 사업의 異體字 인코딩 및 DB 구축 방안> 심포지엄 개최

4. 옛한글의 효율적 활용을 위한 방안 연구

- 주요 워드프로세서의 옛한글 코드 체계 분석
- 주요 워드프로세서의 옛한글 입출력 방식의 비교 분석

5. 문자코드 정보 지원 체계 강화

- 웹 사이트 운영 : 센터 홈페이지 개편
- 뉴스레터 발간·배포 : 15호(6월), 16호(12월)
- 문화재청을 방문하여 <SuperCJK 14.0 검색시스템> 및 관련 정보 제공
- 국사편찬위원회 주최 <고전 전산화 실무자 교육> 지원
- 국사편찬위원회 주최 <고전 전산화 실무자 교육>에 참여한 20여 업체 실무담당자에게 <SuperCJK 14.0 검색시스템> 제공
- 한국한의학연구원에 <SuperCJK 14.0 검색시스템> 및 관련 정보 제공
- 종합 결과보고서 발간 : 1회(12월)

제9절 9차년도(2006)

1. 비표준문자의 수집·등록 및 국제 표준화 지원

- <국가지식정보자원관리사업>의 고전적 전산화 과정에서 발견된 3,926자를 수집하여 신출한자 761자 추출
- 제26차 IRG회의 참석 : 2006년 6월 5일~9일(베트남 후에(Hue))
- 제27차 IRG회의 참석 : 2006년 11월 27일~12월 1일(중국 하이난[海南] 산야[山亞])

2. 문자코드 표준화 연구

- Ext.C1 한자 검토 : 제26차 IRG 회의에서 최종 결정된 Ext.C1 한자 5,000자
- Ext.C1 등록 한국측 한자 IDS 정보 입력 지원 : Ext.C1 한자 5,000자 중 한국측 제안 한자 423자
- Ext.C1 등록 한국측 한자 트루타입 폰트 제작 및 출전전거 이미지자료 구축 지원

- : Ext.C1 한자 5,000자 중 한국측 제안 한자 423자
- <Ext.C2 한국측 제안한자 목록> 작성 대상 자료 분석 : 총 15,269자의 비표준한자 가운데 5,980자의 신출한자 확인
- <Ext.C2 한국측 제안한자 목록>의 최종 검정 : 5,980자의 신출한자 가운데 최종적으로 1,175자 확정
- <Ext.C2 한국측 제안한자 목록> 한자 정보 입력 : 최종적으로 확정된 1,175자

3. 옛한글의 효율적 활용을 위한 지원체계 확립

- <옛한글자모 확장목록> 117자 확정 : 2차례의 ‘옛한글정보처리연구위원회’ 개최
- 향후 유니코드 BMP 영역에 등록할 수 있도록 ISO에 제안

5. 문자코드 정보 지원 체계 강화

- 웹 사이트 운영 : 센터 홈페이지 개편
- 뉴스레터 발간·배포 : 17호(7월), 18호(12월)
- (주)현대정보기술 자료 제공 및 자문 : 2006년 4월, 유니코드 자료와 <유니코드 한자 검색기> 제공
- 경상대학교 <문천각> 자료 제공 및 자문 : 2006년 9월, 신출한자 처리 방법 자문 및 <유니코드 한자 검색기> 제공
- 폰트제작자협회 자료 제공 및 자문 : 2006년 8월, <옛한글 확장자모> 세트 및 유니코드 한자 통계자료 제공, 한자 자형 자문
- 종합 결과보고서 발간 : 1회(12월)

제4장 연구 결과

제1절 비표준문자 수집·정리 및 국제 표준화 지원

그동안 본 센터에서 제3단계 사업 기간 동안인 2004년부터 2007년까지 수집한 비표준문자는 다음과 같다.

연 도	수집대상	수집문자
2004	· 『사마방목』 · 고전국역총서 · 한국문집총간	1,096자
2005	· <국가지식정보자원관리사업>에서 전산화된 고전적	10,247자
2006	· <국가지식정보자원관리사업>에서 전산화된 고전적	3,926자
2007	· <국가지식정보자원관리사업>에서 전산화된 고전적	2,197자
합 계		17,466자

2007년도에는 총 2,197자의 비표준한자를 수집하였는데, 이것은 2006년도 <국가지식정보자원관리사업>에서 산출된 비표준한자 1,897자와 2007년도 <국가지식정보자원관리사업>에서 산출된 비표준한자 300자를 사업에 참여하고 있는 관계 기관과 협의하여 추가로 수집하였다. 따라서 제3단계 사업 기간 동안인 2004년부터 2007년까지 총 17,466자의 비표준한자를 수집하게 되었으며, 수집된 비표준한자는 모두 2차례의 검정 과정을 거쳐 국제표준문자 규격인 유니코드에 등록되지 않은 7,201자의 신출한자를 확인하였다. 이렇게 수집한 신출한자를 바탕으로 <Ext.DUN 한국측 제안한자> 목록을 작성하였다.

제2절 문자코드 표준화 연구

1. <Ext.B 한국측 한자> 트루타입 폰트 제작 지원

기존에 국제표준 문자코드인 유니코드에 등록된 42,711자의 ‘Ext.B 한자’ 가운데 한국측에서 제안하여 등록된 166자의 한자에 대한 트루타입 폰트 제작 요청이 KIRG로부터 있었다. 이에 본 센터에서는 <Ext.B 한국측 한자> 166자의 트루타입 폰트를 제작하여 지원하였다.

※ <부록1 Ext.B 한국측 한자 트루타입 폰트> 참조.

2. <Ext.D 한국측 제안한자> 트루타입 폰트 제작 지원

지난 2006년 제26차 IRG 회의에서는 그동안 논의가 되었던 ‘Ext.C 한자’ 4,219자가 확정되었다. 그리고 나머지 15,000여 자를 ‘Ext.C2 한자’로 하기로 하였다가 ‘Ext.D 한자’로 명칭을 바꾸어 논의를 진행하기로 결정되었다. 여기에는 한국에서 ‘Ext.C1 한자’로 제안한 573자 가운데 ‘Ext.C 한자’로 결정되어 수용된 404자를 제외한 한자 135자가 포함되어 있으며, KIRG로부터 이들 한자의 트루타입 폰트 제작 요청이 있었다. 이에 따라 본 센터에서는 <Ext.D 한국측 제안한자> 135자의 트루타입 폰트를 제작하여 지원하였다.

※ <부록2 Ext.D 한국측 제안한자 트루타입 폰트> 참조.

3. <Ext.DUN 한국측 제안한자> 목록 작성

본 센터에서는 2006년까지 수집한 비표준한자 총 15,269자를 대상으로 IRG 문서 N1216(CJK_C1 Editorial Group Report)에 의거하여 출전정보와 전거자료 이미지가 확실한 한자 1,175자를 <Ext.C2 한국측 제안한자> 목록으로 작성하였다. 그러나 ‘Ext.C 한자’ 결정 이후 Ext.C1에서 Ext.C2로 변경하기로 한 한자가 Ext.D로 바뀌면서 추가 등록을 받지 않기로 결정되었다. 이에 따라 2006년도에 본 센터에서 작성한 <Ext.C2 한국측 제안한자>는 어쩔 수 없이 <Ext.E 한국측 제안한자>로 제안할 수밖에 없는 상황이 되었다.

그런데 지난 6월 4일부터 중국 시안[西安]에서 개최된 제28차 IRG 회의 결과, 각국에서 국가적으로 시급한 한자를 Ext.D 작업과 별도로 ‘긴급 필요[Urgently Need] 한자 추가 접수’를 의결하였다. 이러한 IRG의 회의 결과에 따라 KIRG에서는 <Ext.E 한국측 제안한자>로 준비 중인 한자 목록을 <Ext.DUN 한국측 제안한자>로 제출하기로 결정하였으며, 본 센터에서는 KIRG의 요청에 의해 그동안 진행해 오던 <Ext.E 한국측 제안한자> 목록을 <Ext.DUN 한국측 제안한자> 목록으로 변경하여 작성하게 되었다.

본 센터에서는 <Ext.DUN 한국측 제안한자> 목록 작성을 위해서 먼저 2006년도 작성한 <Ext.C2 한국측 제안한자> 목록의 1,175자와 2007년도에 새로 수집한 신출한자 1,221자를 통합하였다. 그리고 전거정보와 출전이 확실한 한자 2,396자를 선정하여 <Ext.DUN 한국측 제안한자> 후보목록을 작성하였으며, 이 후보목록을 바탕으로 모두 2차례의 검정과 1차례의 자문회의를 거쳐 최종적으로 2,288자의 <Ext.DUN 한국측 제안한자> 목록을 작성하여 KIRG에 제출하였다.

※ <부록3 Ext.DUN 한국측 제안한자 목록> 참조.

4. <Ext.DUN 한국측 제안한자> 한자정보 작성

이상과 같은 과정을 거쳐 작성된 <Ext.DUN 한국측 제안한자> 목록은 “국제 표준 제안한자 양식(IRG N881)”에 맞추어 필요한 정보를 확인하여 입력하여야 한다. 여기에는 다음 표에서 보는 것처럼 ‘강희자전 위치정보’·‘강희자전 부수정보’·‘잔여획수’·‘오피획 정보’·‘해당 한자의 원문정보’·‘이체자 정보’ 등을 필수로 작성하여야 한다.

[IRG N881 문서 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 (오피획)	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

이러한 원칙에 따라 최종 확정된 2,288자의 <Ext.DUN 한국측 제안한자> 목록의 한자에 대해서는 ‘강희자전 위치정보’·‘강희자전 부수정보’·‘잔여획수’·‘오피획 정보’·‘해당 한자의 원문정보’·‘이체자 정보’ 등을 모두 입력하였다. 그리고 이렇게 입력된 한자 정보의 오류 여부를 전문가에게 의뢰하여 자문을 구하였으며, 모든 검토 작업이 끝난 목록에 대해서는 임시코드를 부여하였다.

※ <부록3 Ext.DUN 한국측 제안한자 목록> 참조.

5. <Ext.DUN 한국측 제안한자> IDS 정보 작성

‘Ext.DUN 한자’는 각국에서 ‘긴급하게 필요[Urgently Need]한 한자’에 해당되는 목록이기 때문에 제안 초기에 모든 정보가 완벽하게 갖추어져 있어야 한다는 전제가 있었다. IDS 정보를 비롯하여 트루타입 폰트, 그리고 출전전거 원문 이미지자료 등을 일괄적으로 제출할 것을 요구하였다. 이러한 방침에 따라 <Ext.DUN 한국측 제안한자> IDS 정보를 작성하여 KIRG에 제출하였다.

※ <부록3 Ext.DUN 한국측 제안한자 목록> 참조.

6. <Ext.DUN 한국측 제안한자> 트루타입 폰트 제작

‘Ext.DUN 한자’는 각국에서 ‘긴급하게 필요[Urgently Need]한 한자’에 해당되는 목록이기 때문에 제안 초기에 모든 정보가 완벽하게 갖추어져 있어야 한다는 전제가 있었

다. IDS 정보를 비롯하여 트루타입 폰트, 그리고 출전선거 원문 이미지자료 등을 일괄적으로 제출할 것을 요구하였다. 이러한 방침에 따라 <Ext.DUN 한국측 제안한자>의 트루타입 폰트를 제작하여 KIRG에 제출하였다.

※ <부록4 Ext.DUN 한국측 제안한자 트루타입 폰트> 참조.

7. <Ext.DUN 한국측 제안한자> 출전선거 원문 이미지자료 구축

‘Ext.DUN 한자’는 각국에서 ‘긴급하게 필요[Urgently Need]한 한자’에 해당되는 목록이기 때문에 제안 초기에 모든 정보가 완벽하게 갖추어져 있어야 한다는 전제가 있었다. IDS 정보를 비롯하여 트루타입 폰트, 그리고 출전선거 원문 이미지자료 등을 일괄적으로 제출할 것을 요구하였다. 이러한 방침에 따라 해당 한자의 출전과 선거정보의 원문 이미지자료를 구축하여 KIRG에 제출하였다.

이를 위해 본 센터에서는 2006년에 작성한 <Ext.C2 한국측 제안한자> 1,175자 목록의 출전선거 원문 이미지자료에 2007년도에 새로 수집한 신출한자 1,221자의 출전선거 원문 이미지자료를 추가로 구축하였다. 그리고 <Ext.DUN 한국측 제안한자> 목록에 맞추어 파일 이름을 변경하고 다시 정렬하여 KIRG에 제출하였다.

※ <부록5 Ext.DUN 한국측 제안한자 출전선거 원문 이미지자료(부분)> 참조.

8. 전문가 자문 및 자문회의

본 센터에서는 <Ext.DUN 한국측 제안한자> 2,288자를 최종적으로 확정하는 과정에서 자체적으로 모두 3차례의 검정과정을 거쳤다. 이 가운데 2차 검정이 끝난 후 검정 결과를 전문가에게 의뢰하여 자문을 구하였다. 그리고 전문가에게 의뢰한 자문 내용은 자문회의에서 토의를 거쳐 확정하였다.

※ <부록6 전문가 자문 및 자문회의 회의록> 참조.

제3절 ‘문자코드 표준화 연구’ 사업 정리 및 향후 과제와 전망

문자코드연구센터는 국가 표준코드에 아직 수용되지 않은 비표준문자를 수집·등록하고 이를 국가적 차원에서 정리·활용할 수 있는 기반을 마련하기 위해 1998년 문화관광부의 지원을 받아 ‘비표준문자등록센터’란 명칭으로 설립되었다. 그동안 비표준문자를 수집하고 처리하는 방법을 표준화하는 한편 국어 정보화와 활용을 위한 연구, 인쇄·출판·정보 산업 분야 및 국가 표준 활동의 지원 등 공익적 지원 기구로서 역할을 성실히 수행해 왔다.

본 센터는 1998년부터 제1단계(1998~2000), 제2단계(2001~2003), 제3단계(2004~2007)로 문자코드 표준화와 관련된 연구를 진행해왔다. 지금까지 고전 문헌 자료의 조사를 통한 비표준문자의 상시적인 수집·등록, 국제 표준코드의 최신 정보 수집, 국가 표준코드의 보완 및 확장 방안 연구, 각종 응용 소프트웨어 개발 등 당면한 과제를 수행함과 동시에, 국어정보화 관련 기관 및 연구자들과의 지속적인 협력과 지원을 통해 ‘문자코드 연구 및 지원의 국가적 네트워크 중심’으로서의 역할을 충실히 담당하고자 하였다.

이러한 노력에 의해 2006년도에 Ext.C 한자에 한국측이 제안한 한자 404자를 최종적으로 등록할 수 있었으며, 2007년도에는 <옛한글자모 확장목록>을 유니코드 BMP 영역에 최종 등록할 수 있었다. 이에 따라 한자로 된 고전적 자료와 옛한글 자료를 보다 자유롭고 편리하게 정보처리를 할 수 있는 여건을 마련하게 되었다. 이와 같은 결과는 그동안 본 센터가 추진해온 연구과제가 그 결실을 맺은 대표적인 사례가 될 것이다.

이를 바탕으로 국제 문자코드 표준화의 전망과 이에 대응하기 위해 향후 국가적 차원에서 반드시 진행해야할 과제를 제시하고자 하였다.

※ <부록7 21세기 세종계획 문자코드 표준화 연구 사업정리 및 향후 과제와 전망> 참조.

제4절 옛한글의 효율적 활용을 위한 지원체계 확립

본 센터에서는 지난 2006년도 7월 두 차례에 걸쳐 개최된 옛한글정보처리연구위원회 회의를 통해 국제 표준코드에 제안하기 위한 <옛한글자모 확장목록> 117자를 최종 결정하였으며, ISO/IEC에 공식적으로 제안되었다. ISO/IEC에 제안된 <옛한글자모 확장목록>은 2006년 9월에 JTC1/SC2/WG2 제49차 회의에서 논의가 되었으나, 유니코드 컨소시엄의 반대로 확정하지 못하였다. 그러다가 2007년 4월 23일부터 27일까지 독일 프랑크푸르트에서 개최된 JTC1/SC2/WG2 제50차 회의에서 <옛한글자모 확장목록>이 최종 결정되어 유니코드 BMP 영역에 등록되었다.

기존에 옛한글 정보처리를 위해서는 기존에 유니코드 BMP 영역에 등록된 옛한글자모 240자를 조합하는 방식을 사용하고 있었다. 그러나 이것만으로는 옛한글 자료를 컴퓨터에서 완전하게 처리한다는 것은 불가능하였다. 이번에 <옛한글 자모 확장목록> 117자가 BMP 영역에 최종 등록됨으로써 옛한글로 이루어진 자료의 입출력 및 검색의 토대가 마련되었다고 할 수 있으며, 옛한글의 국제 표준화는 사실상 마무리되었다고 할 수 있다.

새로 등록된 <옛한글 자모 확장목록> 117자의 BMP 상의 영역은 다음과 같다.

<초성: 34자>

Hangul Jamo: U+115A ~U+115E

Hangul Jamo Extended - A: U+A960 ~U+A97C

<중성: 28자>

Hangul Jamo: U+11A3 ~U+11A7

Hangul Jamo Extended - B: U+D7B0 ~U+D7C6

<종성: 55자>

Hangul Jamo: U+11FA ~U+11FF

Hangul Jamo Extended - B: U+D7CB ~U+D7FB

※ <부록8 ISO/IEC의 JTC1/SC2/WG2 회의 문서 N3242> 참조.

제5절 문자코드 정보 제공 및 지원체계 강화

1. 소식지 발간

2007년 7월 23일에 제19호를 발간하여, 관련기관 및 산·학 각계에 배포하였다. 소식지에 수록된 내용은 다음과 같다.

[제19호 목차]

- 권두언: 한자 국제 표준화와 인명용 한자
- 특집: 옛한글 표준화 과정과 의의
- 특별기고: 국사편찬위원회의 정보화 사업 소개
- 해외 연구동향: 대만의 ‘國字標準字體’ 제정에 대하여
- 기획연구: 漢字結構情報를 이용한 한자 검색 모델
- 전통 수확서와 역학서에 사용된 부호
- 연구노트: 俗字의 개념과 유형(3)
- 2007년도 사업 계획
- 제28차 IRG 회의 소개
- 센터 동정

[제20호 목차]

- 권두언: 문자코드연구센터의 10년을 마무리하면서
- 표준소개 : ‘KS X 1026-1’의 옛한글 처리
- 특집: 문자코드연구센터의 성과와 과제
- 특별기고1: 한국국학진흥원 정보화사업 소개
- 특별기고2: 한자문화권 상용한자 자형의 통일문제에 관한 제안
- 해외 연구동향1: 現代漢字學의 개념과 연구 내용
- 해외 연구동향2: 京都大 ‘唐代研究 Knowledge Base’
- 제29차 IRG 회의 소개
- 센터 동정

※ <부록9 문자코드연구센터 소식지 제19호·제20호> 참조.

2. 웹사이트 운영

2005년도 전면적인 개편 작업을 거친 홈페이지를 지속적으로 관리 운영 중에 있다. 특히 센터 웹사이트에서 제공하고 있는 <유니코드 한자검색기>는 2005년 12월 서비스를 시작한 이후 현재까지 누적 접속 건수가 50,000건에 이르고 있다.

3. 산업계 및 관련기관 지원

(1) 민족문화추진회와 이체자 관련 자료의 공유와 협력

지난 2월말 민족문화추진회(회장 조 순) 편찬실(실장 백한기) 실무진과 전산화 과정에서 이체자 처리에 대한 의견을 교환하였다. 향후 민족문화추진회와 본 센터는 이체자 관련 자료의 공유와 처리에 대해 협력해 나가기로 의견을 모았다.

(2) (주)맥한도에 자료 제공 및 자문

지난 8월에 폰트 전문 제작업체인 (주)맥한도(대표 금하연)에 유니코드 한자와 관련된 각종 코드 정보를 제공하였으며, 우리 어문 생활에 맞는 표준 자형의 설정 문제와 유니코드 한자 폰트 제작에 대해 자문하였다.

부 록

- 【부록1】 Ext.B 한국측 한자 트루타입 폰트
- 【부록2】 Ext.D 한국측 제안한자 트루타입 폰트
- 【부록3】 Ext.DUN 한국측 제안한자 목록
- 【부록4】 Ext.DUN 한국측 제안한자 트루타입 폰트
- 【부록5】 Ext.DUN 한국측 제안한자 출전전거 원문 이미지자료(부분)
- 【부록6】 전문가 자문 및 자문회의 회의록
- 【부록7】 21세기 세종계획 문자코드 표준화 연구 사업정리 및 향후 과제
와 전망
- 【부록8】 ISO/IEC의 JTC1/SC2/WG2 회의 문서 N3242
- 【부록9】 문자코드연구센터 소식지 제19호·제20호

【부록1】 Ext.B 한국측 한자 트루타입 폰트

【부록2】 Ext.D 한국측 제안한자 트루타입 폰트

【부록3】 Ext.DUN 한국측 제안한자 목록

【부록4】 Ext.DUN 한국측 제안한자 트루타입 폰트

【부록5】 Ext.DUN 한국측 제안한자 출전전거 원문 이미지자료(부분)

【부록6】 전문가 자문 및 자문회의 회의록

**【부록7】 21세기 세종계획 문자코드 표준화 연구 사업정리 및
향후 과제와 전망**

【부록8】 ISO/IEC의 JTC1/SC2/WG2 회의 문서 N3242

【부록9】 문자코드연구센터 소식지 제19호·제20호