

한의학고전문헌 DB구축과 신출자 처리

- 한의학지식정보자원웹서비스를 중심으로 -

이정화¹⁾, 김홍준

한국한의학연구원 학술정보부

Abstract

Improvement plan for 'Newly found ideographs(新出漢字)' in the digitalizing business of the old Korean Medicine documents

- with 'knowledge of oriental web service' -

*Lee Jeonghwa, Kim Hong Jun
Korea Institute of Oriental Medicine*

As we enter the 21st century, the Information Era, we are making a national effort to digitalize the information resources of Korean Studies, based on our leading digital technology.

However, there is much difficulty of computerizing Chinese characters used in Korea, China, and Japan, with the of technologies developed by the West.

This paper gives an example of how to register and process the Newly found ideographs(新出漢字) put forth by Digitalization of Knowledge information resource on Korean oriental medicine.

¹⁾ 교신저자 : 이정화 - 한국한의학연구원 학술정보부, 대전시 유성구 전민동 461-24 (305-811) / 017-224-2708

I. 들어가는 말

21세기 정보화 시대로 접어들면서 우리나라에서는 앞선 디지털기술의 바탕위에 한국학 지식 정보자원을 국가적인 차원에서 디지털화하는 사업들을 활발하게 진행하고 있다. 이는 발달된 인터넷 전산망을 통하여 더욱 더 활기를 띠게 되었으며, 축적된 데이터들은 전 세계에 서비스되고 있다. 이러한 통신매체의 발달과 데이터의 축적 기술은 종이매체에서 불가능했던 일들을 가능케 하고, 엄청난 양의 자료들을 데이터베이스화하여 실시간으로 서비스함으로써 많은 사람들이 자료를 공유하게 되었다.

그러나 서구 주도하에 시작된 정보통신의 발달은 한·중·일 삼국에서 사용하고 있는 한자를 전산화하기에는 그리 용이하지 않으며, 완벽한 재현을 위해서는 많은 어려움이 있다. 컴퓨터는 작동의 명령어와 각종 프로그램에 의해 운용되는데 이러한 컴퓨터 언어가 영어로 되어있고 그 프로그램 체계 또한 영어에 맞게 개발되었다. 때문에 서구 중심의 컴퓨터 기술은 한자의 입출력을 배려하지 못하였으며, 이로 인한 디지털화의 한계는 영어 이외의 문자로 이루어진 자료의 디지털이징과 구현에 어려움으로 작용하고 있다. 우리나라의 경우도 마찬가지로 국가지식정보자원관리사업의 경우, 한자 또는 고어가 기록되어 있는 고전 자료를 디지털화하게 되는데, 그 작업 과정에서 한자나 고어를 완벽하게 처리하여 서비스하기란 쉽지 않은 게 현실이다.

이러한 상황을 개선하기 위해 국제적으로는 한자 문화권에서 사용하고 있는 한자 코드를 통합 정리된 국제 표준 문자 코드(Unicode)²⁾로 완성하기 위해 노력하고 있다. 이 코드 체계가 완성되면, 국제적으로 통용되는 코드로 전 세계가 이 코드를 이용하게 될 것이며, 한자에 대한 정보를 처리하게 된다. 이를 우리는 국제표준문자코드(USC 10646-1,1993)라고 부르는데, 이 코드를 부여 받은 한자는 현재 27,484자에 달하여, 한·중·일 각국의 한자는 각국의 사용 관례를 존중하여 그 고유 자체가 각각 달리 등재되어 있다.

본 글에서는 한의학 지식정보자원관리사업에서 산출되는 신출한자³⁾의 등록과 처리절차의 사례를 들어보고, 고전 자료 디지털화에서의 신출한자 처리에 대한 개선방안을 찾아보고자한다.

²⁾ 컴퓨터에서 세계 각국의 언어를 통일된 방법으로 표현할 수 있게 제안된 국제적인 문자 코드 규약. 국제 표준으로 제정된 2바이트계의 만국 공통의 국제 문자부호 체계(UCS : Universal Code System)를 말한다. 애플컴퓨터 · IBM · 마이크로소프트 등이 컨소시엄으로 설립한 유니코드(Unicode)가 1990년에 첫 버전을 발표하였고, ISO/IEC JTC1에서 1995년 9월 국제표준으로 제정하였다. 공식명칭은 ISO/IEC 10646-1(Universal Multiple-Octet Coded Character Set)이다.

³⁾ 신출한자는 현재 확정된 표준 용어는 아니다. 사용하는 사람에 따라 신출자, 벽자, 비표준한자, 이체자 등 여러 가지로 쓰이고 있다. 본고에서는 한문으로 이루어진 고전자료의 전산화 과정에서 새롭게 검출된 한자 중 전산처리가 어렵거나 불가능하여 입출력이 불가능한 한자들의 의미로 '신출한자'를 사용하고자 한다.

II. 신출한자의 개념과 유형

1. 국제표준규격과 한자코드

국제표준코드의 필요성은 정보통신기술의 발달로 국가 간 정보유통이 활성화 되면서 중요한 문제로 대두하였다. 국제 표준코드로서 유니코드(Unicode)가 실용화되면서 세계는 바야흐로 단일코드 시대에 접어들고 있다. 유니코드는 1996년 V2.0이 발표되어 일부 운영체계의 표준코드로 채택된바 있고, 1999년에는 V3.0이, 2001년 5월에는 V3.1이, 2003년 DPSMS 4.0이 발표되었다.⁴⁾ 향후 대부분의 운영 체제나 응용소프트웨어가 유니코드를 표준코드로 채택할 것이 확실시 된다.

특히, 한자문화권의 경우 공통의 문화적 기반으로 인하여 각국 사이의 원활한 정보 소통의 필요성은 더욱더 절실하다. 때문에 이러한 필요성에 부응하여, 한자문화권 국가에서 사용되는 한자코드에 대한 국제표준규격이 필요하게 되었고 이에대한 국제 표준규격은 'ISO/SC2/WG2' 산하에 소속된 IRG (Ideographic Rapporteur Group, 표의문자코드전문가회의)를 통하여 진행되고 있다.

한국 · 중국(대만포함) · 일본은 ISO 산하에 표의문자코드전문가회의(IRG)를 결성하여 회합과 연구를 거듭한 끝에 당초 각국이 수용을 희망한 약 5만4천자의 한자를 20,902개의 한자군으로 통합하는 최종안(ISO 10646-1)을 확정하였다. 여기에 1997년 Extension A가 추가되어 총 27,484자가 국제표준문자코드로 제정되었고, 이는 유니코드 V3.0에 수용되었다. 이후 2000년 12월 서울에서 열린 IRG 회의에서 Extension B 영역의 한자 42,711자가 추가로 표준코드로 확정되었다. 그러나 표준 코드로 제정된 CJK 한자 가운데 현재 운영체제나 각종 응용프로그램 개발에서 실용화 된 것은 Unicode CJK⁵⁾ Unified Ideographs Extension A 27,484자이다.

유니코드 한자는 한 · 중 · 일 통합 한자(Super CJK)를 기반으로 하였기 때문에 각 나라마다 문자의 관습과 의미는 차이가 날 수밖에 없다. 게다가 컴퓨터의 보급이 확산되어 사용이 일반화 되면서 한 · 중 · 일 삼국이 같은 한자를 서로 다른 전산체계에서 한자의 자체와 부호체계를 사용함으로써, 인터넷 전산망을 통한 신속한 정보 교류에 많은 장애가 발생되고 있다. 이 같은 현실을 고려할 때 한 · 중 · 일이 협의를 통해 한자를 표준화하는 작업을 하여야 하나 그리 간단치 않다.

4) 유니코드 V3.0 한중일 통합한자(Super CJK) 27,484자이며, 1991년 이후 추진되어온 다국어 문자 처리를 위한 유니코드 표준 (Unicode Standard) 이 2003년 4월 버전 4.0.0을 업그레이드되었다. 이번 업그레이드는 V3.2에서 업그레이드 된 것으로 중요한 변화가 있었다. 우리나라와 동아시아 관련된 추가 사항으로는 ① CJK 이체 선택자 블록 추가 : 향후에 등재될 CJK 이체자들을 위한 이체 선택자(block of variation selectors)라는 영역이 추가되었다. ② 사전 용도의 추분절음 부호 부호 문자(약물)들이 추가되었다. ③ 동아시아 전각 · 반각 문자의 글자폭에 관한 규정이 유니코드 표준 부속 문서에 삽입되었다.

5) CJK란 China(중국), Japan(일본), Korea(한국)를 가리킨다.

따라서 이러한 문제를 극복하고 유니코드 한자를 우리의 어문생활과 정보 환경에서 보다 효율적으로 이용하기 위해서는 다양한 기초 연구가 이루어져야 할 것이다. 한 나라와 문자 정책은 쉽게 바뀔 수 있는 것이 아니다. 이를 위해서는 먼저 우리 한자의 자체, 독음, 교육용 한자를 표준화함은 물론이거니와 동시에 컴퓨터에서 사용하는 한자도 표준화되어야 한다.

2. 신출한자의 개념

신출한자는 한문자료의 고전 전산화 작업 과정에서 새롭게 조사·추출된 한자로서, 기존의 국제표준규격인 유니코드 확장한자 문자세트에서 아직 등록되어 있지 않은 한자를 말한다. 현재 한자문화권 국가에서 사용되는 한자 코드에 대한 국제 표준규격은 표의문자 코드전문가회의(IRG)라는 국제기구를 통하여 진행되고 있다.

국제 문자 부호계에 포함된 국제 표준규격의 한자를 정리하면 다음과 같다.

- 기본 다국어 평면(BMP, Basic Multilingual Plane)	: 20,902자
- 한 중일 통합한자 확장 A(CJK Unified Ideographs Extension A)	: 6,582자
- 한 중일 통합한자 확장 B(CJK Unified Ideographs Extension B)	: 42,711자
	합계 : 70,195자

따라서 신출한자의 범위는 위의 목록에 들어있지 않은 한자 곧, Extension B 영역 이외의 한자를 말한다. 단, 우리나라의 경우 중국이나 대만과는 달리 컴퓨터의 범용 OS와 주요 워드프로세서에서 Extension A 영역의 한자까지만 입출력을 지원하기 때문에 Extension A 영역외의 한자부터 신출한자로 처리하고 있다. 다시말해, Extension B 영역의 한자부터 신출한자로 처리하고 있는 것이다. 물론 소프트웨어를 사용하여 임시적으로 Extension B 영역의 한자를 입출력할 수 있지만, 현재로서는 그것도 일부의 프로그램에서만 가능하기 때문이다.

신출한자는 크게 이체자⁶⁾류와 신출자류로 구분할 수 있다. 이체자류는 기존 유니코드에 등록된 한자와 자형만이 다른 이체자들을 가리키고, 신출자류는 기존에 등록되지 않은

6) 異體字는 문자를 사용하는 과정에서 여러 가지 요인으로 발생한 正字와 音義는 같지만 字形이 다른 자를 가리킨다. 이체자는 혹, 음의가 같지만 서로 자체가 다른 개념만으로 이해되어, 단순히 동일 음의를 가진 자체가 서로 다른 두 가지 관계만을 규정 지은 용어로 해석될 수 있다. 그러나 일반적으로 이체자는 “某字의 이체자”라는 뜻으로 해석되어, 두 자체의 관계를 나타내는 것에 그치지 않고 정자에 대응하는 자체임을 나타낸다. 이체자라는 용어는 이와같이 동일자의 다양한 자체관계를 나타낼 뿐만 아니라, 바로 어떤 한자의 정자의 다른 字體라는 뜻을 함유하고 있다. 別字, 別體字, 異文, 字體之異, 僞體, 譌字, 謬體, 繆體, 俗字, 俗體 등으로 불리기도 한다. (『漢字 異體字 辭典 編纂 방법론 연구』 『국제 문자 코드 제안 한자의 표준화에 대한 연구』. 李浚碩. 문화관광부. 1998.)

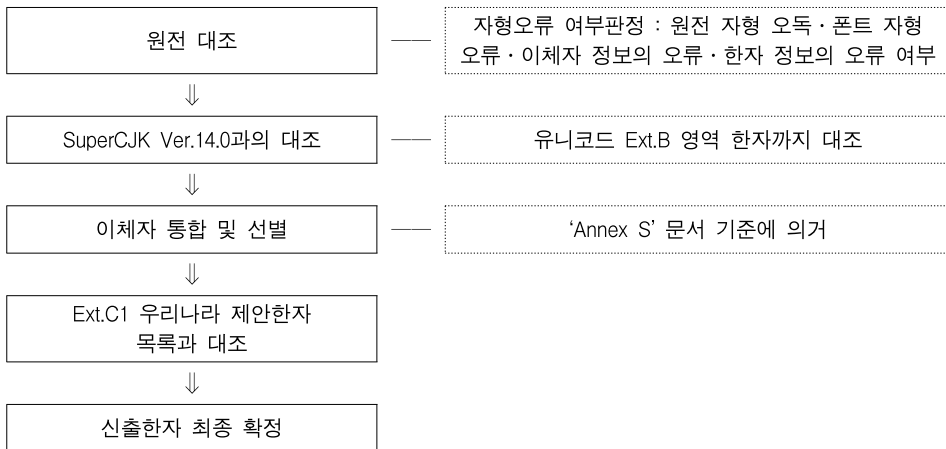
한자로서 새롭게 발견된 한자를 말한다. 이체자를 유형별로 나누면, BMP 이체자, Extension A 이체자, Extension B 이체자로 나눌 수 있으며, 신출한자의 유형은 Extension B, Extension C1 제안 한자(한국), 신출한자로 나눌 수 있다. 이처럼 다양한 신출한자의 유형을 정확히 선별해내는 작업은 그리 간단치 않으며, 전문가의 의견과 판단이 필요하다.

3. 신출한자 검정 과정⁷⁾

신출한자를 정확하게 판정하고 이를 국제기구에 제안하기 위해서는 전문기관에서 여러 단계의 검정과정을 거쳐야 한다. 이는 무분별하게 마구 수집하여 많은 양을 확보하기 보다는 가능한 정확한 데이터를 확보해 후보 목록으로 작성하여 국제 표준기구에 제안하는 것이 바람직하기 때문이다. 국제적인 공감과 인정을 얻지 못하는 신출한자를 정확성 없이 제안한다면, 국제 표준으로 수용되기 어려울 뿐만 아니라 이후 한자 표준화사업에 대한 국가적 신뢰를 잃을 수도 있기 때문이다. 그 만큼 신출한자 처리와 제안은 가능한 신중하고 정확하게 이루어져야 한다.

우리나라 신출한자의 대표적인 검정기관이라 할 수 있는 문자코드연구센터⁸⁾에서 이루어지고 있는 신출한자의 검정 과정을 도식화하면 아래와 같다.

[신출한자 검정 과정]



7) 이는 신출한자 검정의 일반적인 절차이며, 정통부 과제인 지식정보자원관리사업에서 2000-2004년까지 각 사업별로 산출된 신출한자를 통합하여 문자코드연구센터에 검정을 의뢰하여 보고받은 보고서에 기록된 실제 검정과정이기도 하다.(전산원 수탁과제)

8) 문자코드연구센터는 1998년에 비표준문자 연구와 등록 지원을 담당하는 국가 공인 대표기관으로 출범하여, 비표준문자의 국제표준화 작업 및 문자코드 관련 제반 사업을 지속적으로 추진해 왔으며 2003년 비표준문자등록센터에서 문자코드연구센터로 명칭을 변경하여 오늘에 이른다.

(1) 원전 대조를 통한 검정

‘원전 대조를 통한 오류 여부 검정’ 작업은 신출한자 후보 목록의 개별 한자를 원전과 직접 대조하면서 오독으로 인한 오류나 폰트의 오류 여부 등을 조사하여 검정하는 것으로 신출한자 검정의 처음 단계이다. 원전대조를 통하여 신출한자를 검출할 때에 나타나는 오류는 대부분 원전자료의 입력과 교정 등과 같은 1차 전산화 과정에서 발생한다. 이 단계에서의 오류는 한자자동입력기를 이용하거나 작업자가 수작업으로 입력을 하면서 발생하는 것이다. 따라서 초기 작업 인력의 질적 수준에 따라 그 오류의 비율이 큰 차이가 나게 된다. 때문에 1차 작업자의 질적 수준을 높이고 이를 전문화 할 필요가 있다.

이를 유형별로 살펴보면 다음과 같다.

① 원전 자형의 오독 검정

원전 자형의 오독은 원전 한자의 자형에 대한 착각이나 이해의 부족으로 기존 표준코드에 등록되어 있는 한자인데도 불구하고 이를 변별하지 못한 경우에 많이 발생한다.

② 폰트 자형의 오류 검정

폰트 자형의 오류는 원전자형의 오독에서 비롯되는 경우가 많으며 신출한자를 폰트로 새롭게 제작하는 과정에서 자획의 일부가 잘못 제작되는 경우에 발생한다.

③ 이체자 처리의 오류 검정

이체자 처리의 오류는 이체자가 아닌데 이체자로 잘못 파악한 경우에 발생한다. 이는 필사자의 개인적 취향이나, 필체로 인해 나타나는 경우와, 판본의 경우 활자나 판의 글자가 떨어지거나거나 뭉그러져 정확하게 인쇄되지 못한 경우이다. 이와 같은 경우 오류를 범하지 않기 위해서는 각 전문분야별 전문가의 안내와 판단이 요구된다.

④ 한자 정보의 오류 검정

한자 정보의 오류는 해당 한자의 부수 정보, 총획수 정보, 잔여획수 등을 잘못 파악하여 나타나는 현상으로 이미 유니코드에 등재되어 있는데 이를 놓치는 경우라 하겠다.

(2) SuperCJK Ver.14.0 과의 대조를 통한 신출한자 여부 판별

이 과정은 앞 단계에서 원전대조 검정의 과정을 통해 걸러진 신출한자 후보들을 SuperCJK Ver. 14.0과 대조를 통하여 신출한자 여부를 판별하는 2단계 작업이다. 이 한자목록은 Extension B 영역의 한자까지를 포함하고 있기 때문에 신출한자 판정에 매우 유용하다. 이 목록의 배열 방법은 표제자를 『강희자전』 부수 순서로 구분하고, 같은 부수 내에서는 획수 순서로 나열하는 전통적인 자서의 체제를 따르고 있다. 그리고 표제자에는 부수, 잔여획수, 유니코드 코드의 정보, 『강희자전』의 위치정보, 『한어대자전』의 위치정보, 각국의 한자 코드 등의 정보를 제공한다.

또한 SuperCJK Ver. 14.0은 웹상으로 서비스되기 때문에 신출한자를 이 한자목록과 대조하는 작업은 해당 한자의 부수와 잔여획수를 확인하여 수록 여부를 통해 신출한자 여부를 판정하는 비교적 단순한 일이다. 단, 이 목록으로 신출한자를 대조하는 과정에서 주의해야 할 점은 검색작업 시 획수 계산에서 $\pm 1 \sim 2$ 획까지 대조해야한다는 것이다. 이는 나라마다 자형이 다르고 부수가 다르고 획수를 계산하는 방법이 다를 수 있기 때문이다.

(3) IRG의 'Annex S' 문서 기준에 의거하여 이체자 통합 및 선별

'Annex S(IRG N951)' 문서는 IRG의 기술문서로서, 유사 자형 한자와의 구분과 이체자 통합 및 선별에 대한 기준이 제시된 문건이다. 이 문건의 취지는 기본적으로 각국에서 이체자를 무분별하게 신출한자로 제안하는 것을 통제하기 위한 것이다. 이 문서는 기존에 등록된 이체자는 수용하고, 새롭게 제안되는 신출한자에만 적용되는 기준으로써 통합해도 무방한 이체자와 구분해야 할 이체자를 예시를 통하여 기준을 정하고 이를 표준화하였다.

(4) Ext.C1 우리나라 제안한자 목록과 대조

여기서 말하는 Ext.C1 제출한자 목록은 2002년 4월 IRG에 제출한 우리나라의 신출한자 목록이다. Ext.C 영역의 한자는 C1과 C2로 나누어 제안하게 되는데, 현재는 각국에서 C1 영역의 한자를 제안된 목록을 최종적으로 검토하는 단계에 있다. 따라서 신출한자를 Ext.C1 한자 목록과 대조하여 이곳에 등재되지 않은 한자를 선별하는 단계이다. Ext.C1 우리나라 제안한자 목록과 대조하여 중복되지 않는 한자를 선별하고, C2 영역의 우리나라 제안한자 목록에 추가로 등록하게 된다.

(5) 신출한자 문자정보 확인

신출한자로 검출된 한자에 대한 문자정보를 확인하는 단계이다. 신출한자에 대한 부수, 잔여획수, 총획수 등의 정보를 중심으로 재검토하는 단계로 국제 표준 제안한자 양식에 맞게 필요한 정보를 확인하여 정리·입력하는 과정이다. 신출한자의 표현 방안은 부호를 사용하는 방법, 자형의 조합식을 사용하는 방법, 한글음을 사용하는 방법 등이 입력 교정 단계에서 사용되고, 웹 화면 출력은 태그(Tag)를 사용하는 방법으로 이루어진다.

Ⅲ. 한의학 지식정보자원 관리사업에서의 신출한자

국가지식정보자원관리사업의 하나로 수행된 한의학지식정보자원 관리 사업에서는 2003

년1차년도 사업에서 총 163종의 고문헌을 디지털화하였으며, 2차 사업에서 157종, 3차 사업에서는 44종의 한의학 고전문헌을 디지털화하였다. 3차년도 사업동안 총 364종의 고문헌을 디지털화 하였으며 여기서 신출한자가 산출된 문헌은 207종에 달한다. 문헌당 적게는 1종으로부터 많게는 한 문헌에서 448개의 신출한자가 산출되기도 하였다.

한의학지식정보자원 관리 사업을 개괄하고, 신출한자 처리 절차를 살펴보고, 신출한자의 출현 현황을 정리하면 다음과 같다.

1. 한의학지식정보자원 관리 사업

명칭	사업내용	추진기간	구축내용		사업비 (천원)
한의학 지식정보 자원 디지털화 사업(1차)	한의학 지식정보 자원 서비스 기반마련 -초기 데이터 구축 -초기 사이트 런칭	2003/05 ~ 2003/11	DB	고문헌 텍스트 8만7천면 이미지 5만면 해제1천8백매 치료기술 텍스트 5천면 이미지 1만점	754,900
한의학 지식정보 자원 디지털화 사업(2차)	한의학 지식정보 자원 서비스 본격화 -데이터 추가 구축 -검색 및 부가 기능 개선	2004/04 ~ 2004/11	DB	고문헌 텍스트 8백만자 이미지 1만4천면 해제 1천8백매 치료기술 텍스트 1,300만자/ 용어사전 6백만자 용어시서스 디스크립터 3천개	1,334,424
한의학 지식정보 자원 디지털화사업(3차)	한의학 지식정보 자원 서비스 확충 -정보 유형 다양화 -검색부가 및 관리기능 개선	2005/03 ~ 2005/11	DB	고문헌 텍스트 150만자 이미지 7천5백면 해제 1천매 전문의료정보 텍스트 4천매 이미지 2천점 2D 50점 건강의료정보 텍스트 2천매 이미지 5백점 용어시서스 디스크립터 3천개(추가)	594,500

2. 신출한자 처리 절차

한의학 지식정보자원 디지털화사업에서의 신출한자 처리 과정을 자세히 살펴보면 아래의 도표와 같다.

단계	단계	처리내용	비고
----	----	------	----

입력	최초입력	입력 단계에서 입력 불가능한 자들을 추출 -한자의 정확한 판독이 요구됨 -이체자는 정자로 입력	-신출예정자 관리테이블 기록(엑셀시트에 목록화) -교정단계에서 정확성을 높인다.
	입력불능한자출현	마크업 부가 혹은 신출자 관리 번호 부여 신출예정자 관리테이블 기록	
교정	마크업 발견	원본한자의 정확한 교정 이체자는 정자로 교정	
수정입력	마크업 발견	이체자는 정자로 수정입력 정자는 정확한 자형대로 입력 입력불가능한자 집계, 리스트업	신출자관리테이블 갱신
신출자 1차 정의	리스트 검토	이체자 / 정자 판정 신출자 여부 임시 판정 신출자의 임시 부가정보 생성 신출자관리테이블 갱신 임시코드 배정	-신출자관리시스템 적용 -신출자 관리위원회 운영
신출자 판정	신출자 판정	신출자 목록에서 중복자 삭제 역통, 규장각, 비표준문자센터 등 기존 신출자들과의 대조를 통해 확인. 이미 존재하는 문자는 기존 코드값 부여	-엑셀시트에 표시 -주관기관 신출한자 판정
코드 부여	임시코드 부여	최종 신출자를 대상으로 임시코드값 부여. 이때 부수/ 획수 순으로 코드값 부여. 기존 문자는 해당 이미지 폰트 활용	-엑셀시트로 작업
신출자 제작	리스트 적용	코드배정, 폰트제작 유니코드 테이블 및 부가정보 테이블 업그레이드	-상용 입력기 (HWP)는 업그레이드가 불가능하므로 XML Editor상에서 최종 산출물인 XML 문서만 업데이트 -신출자관리테이블 갱신
폰트 제작	이미지폰트 제작	임시코드값이 부여된 신출자를 대상으로 이미지 폰트 제작	-엑셀시트에 이미지폰트 명표시 -16×16, 32×32 두 가지 사이즈로 제작
	서비스 반영	이미지 폰트 제작이 완료된 신출자는 XML 데이터에 이미지 링크를 점검한 후 서비스에 반영	
	트루타입 폰트 제작	최종 신출자를 취합하여 트루타입 폰트 제작, 주관기관에 제출	-서비스에는 반영하지 않음

3. 신출한자의 출현 현황

[2003년도 한의학 지식정보자원 디지털화사업에서의 신출한자 현황]

대상자료명	출현자수	대상자료명	출현자수	대상자료명	출현자수
醫方類聚	448	醫寶	9	醫本	3

家庭救急方	2	婦人大全良方 9	3	醫藥鑑	1
簡易辟瘟方	1	婦人大全良方 7	2	宜臬	55
檢要	5	婦人大全良方 9	1	餌方	4
經驗痘方	1	四時纂要抄	2	人夫須知	1
經驗方	7	四醫經驗方	5	仁濟志 1	3
谷青冗語	1	生産秘方	8	仁濟志 2	2
				仁濟志 3	9
廣濟秘笈	11	壽民妙詮	1	仁濟志 4	17
救急簡易方諺解	2	壽世寶訣	3	仁齋直指方	9
救急單方	8	壽養叢書類輯	3	濟生篇	3
救急方	2	袖珍經驗神方	3	舟村新方	3
及幼方 1	2	食物本草	5	增補山林經濟	2
及幼方 2	1	身機踐驗 상	14	增修無冤錄諺解	2
丹谷經驗方抄	3	身機踐驗 하	11	纂圖方論脈訣集成	1
單方秘要·經驗新編	3	神應經	1	察病用藥訣	1
東西醫學要義	3	新註無冤錄	4	瘡疹集	2
麻疹奇方	1	新集御醫撮要方	1	草堂遺訣	1
幕經	2	新纂辟瘟方	3	春鑑錄	7
脈訣理玄秘要	1	新編集成牛馬醫方	12	鍼灸極秘傳	1
名方類證醫書大全	5	兩無神編 1	2	鍼灸集成	5
明善錄	1	兩無神編 2	1	鍼灸擇日編集	2
辟瘟新方	2	良方金丹	3	胎產要錄	1
葆養志	5	諺解臘藥症治方	4	太醫院先生案	3
本草附方便覽 1	12	諺解痘瘡集要	1	八陣方	1
本草附方便覽 2	4	用藥賦	3	漢方醫學小兒專科	4
本草附方便覽 3	14	醫科先生案	14	解惑辯疑	7
本草備要	3	醫科八世譜	2	鄉藥救急方	2
本草類函要領	2	醫方綱要	1	鄉藥集成方 2	6
本草精華	9	醫方集略 1	2	鄉藥集成方 4	4
本草彙英 1	4	醫方集略 6	1	鄉藥集成方 5	31
本草彙英 2	6	醫方撮要	1	鄉藥採取月令	1
本草彙英 3	1	醫方合編 1	5	活人心法	1
本草彙英 6	1	醫方合編 2	4	回春方	3
本草彙英 7	3	醫方合編 3	1		

[2004년도 한의학 지식정보자원 디지털화사업에서의 신출한자 현황]

파일명	대상자료명	출현자수
KIOM_A164_Z_001	家庭療法	3
KIOM_A165_Z_001	經脈氣穴經絡圖	8
KIOM_A166_Z_001	雞壇嚶鳴	1

KIOM_A168_Z_001	老人帖 三班	3
KIOM_A169_Z_001	農書	2
KIOM_A170_Z_001	丹溪纂要 上	1
KIOM_A172_Z_001	東方醫藥 第3卷 第1號	5
KIOM_A172_Z_002	東方醫藥 第3卷 第2號	2
KIOM_A172_Z_003	東方醫藥 第3卷 第3號	3
KIOM_A172_Z_005	東方醫藥 第4卷 第1號	2
KIOM_A173_Z_005	東西醫報 昭和參年十月	1
KIOM_A173_Z_007	東西醫報 昭和四年二月	1
KIOM_A174_Z_005	東西醫學報第六號	3
KIOM_A175_Z_004	東西醫學研究會月報 第四號	2
KIOM_A175_Z_005	東西醫學研究會月報 第五號	3
KIOM_A176_Z_001	東洋醫藥 創刊號	2
KIOM_A176_Z_002	東洋醫藥 第1卷 第2號	1
KIOM_A176_Z_003	東洋醫藥 第1卷 第3號	6
KIOM_A176_Z_005	東洋醫藥 第2卷 第1號	3
KIOM_A177_Z_001	東洋醫學 第一卷第一號	1
KIOM_A178_Z_002	東醫報鑑 第二號	2
KIOM_A179_Z_001	東草單方	7
KIOM_A182_Z_001	撥醮	24
KIOM_A183_Z_001	百草藥學	1
KIOM_A185_Z_001	本草摘要 卷一	3
KIOM_A186_Z_001	傅氏眼科審視瑤函	4
KIOM_A187_Z_001	秘傳萬病通治法	9
KIOM_A189_Z_001	桑韓醫問答	3
KIOM_A190_Z_002	桑韓鏘鏘錄 醫譚	7
KIOM_A191_Z_002	桑韓唱酬集 二	1
KIOM_A192_Z_001	桑韓唱和填箎集 卷三	1
KIOM_A192_Z_003	桑韓唱和填箎集 卷八	6
KIOM_A193_Z_002	桑韓唱和填箎集 卷四	2
KIOM_A193_Z_004	桑韓唱和填箎集 卷九	4
KIOM_A194_Z_001	和韓唱和錄	1
KIOM_A195_Z_001	石山人別錄	1
KIOM_A196_Z_001	仙槎筆談	8
KIOM_A197_Z_001	仙槎筆談抄錄	3
KIOM_A198_Z_001	攝生總要	17
KIOM_A199_Z_001	攝養要訣	3
KIOM_A201_Z_001	壽生新鑑	2
KIOM_A203_Z_001	藥性詩兼註	11
KIOM_A205_Z_001	兩東關語 乾	5

KIOM_A205_Z_002	兩東關語 坤	10
KIOM_A206_Z_001	兩東筆語 卷一·二	5
KIOM_A206_Z_002	兩東筆語 卷三·四	7
KIOM_A206_Z_003	兩東筆語 卷五·六	6
KIOM_A207_Z_001	養病心鑑	7
KIOM_A210_Z_001	倭韓醫談	5
KIOM_A211_Z_001	要略	3
KIOM_A212_Z_001	衛生大鑑	5
KIOM_A214_Z_001	醫心方	164
KIOM_A215_Z_001	醫藥月報 第一卷 第貳號	1
KIOM_A215_Z_002	醫藥月報 第一卷 第三號	2
KIOM_A215_Z_003	醫藥月報 第一卷 第四號	1
KIOM_A215_Z_005	醫藥月報 第二卷 第一號	3
KIOM_A215_Z_007	醫藥月報 第二卷 第三號	3
KIOM_A215_Z_008	醫藥月報 第二卷 第四號	3
KIOM_A215_Z_009	醫藥月報 第二卷 第七八號合編	5
KIOM_A215_Z_010	醫藥月報 第二卷 第九號	3
KIOM_A215_Z_011	醫藥月報 第二卷 第十號	2
KIOM_A215_Z_012	醫藥月報 第二卷 第十一十二號合編	1
KIOM_A215_Z_014	醫藥月報 第二卷 第十四十五號合編	1
KIOM_A218_Z_001	一金方 上	4
KIOM_A218_Z_002	一金方 下	5
KIOM_A221_Z_003	朝鮮醫學界 第參號	2
KIOM_A221_Z_005	朝鮮醫學界 第五號	3
KIOM_A221_Z_006	朝鮮醫學界 第六號	1
KIOM_A221_Z_007	朝鮮醫學界 第七號	1
KIOM_A221_Z_008	朝鮮醫學界 第八號	3
KIOM_A221_Z_009	朝鮮醫學界 第九號	1
KIOM_A221_Z_010	朝鮮醫學界 第拾號	2
KIOM_A221_Z_011	朝鮮醫學界 第拾壹號	1
KIOM_A224_Z_001	朝鮮筆談	4
KIOM_A225_Z_001	朝鮮筆談 乾	2
KIOM_A227_Z_001	忠南醫藥 第五號	1
KIOM_A228_Z_001	韓客贈答 上	1
KIOM_A231_Z_002	漢方醫藥 第十九號	2
KIOM_A231_Z_003	漢方醫藥 第二十號	3
KIOM_A231_Z_004	漢方醫藥 第二十一號	3
KIOM_A231_Z_006	漢方醫藥 第二十三號	1
KIOM_A231_Z_009	漢方醫藥 第三十八號	1
KIOM_A231_Z_010	漢方醫藥 第三十九號	1

KIOM_A232_Z_001	漢方醫藥界 第貳號	1
KIOM_A233_Z_002	韓槎垣篋集 坤	1
KIOM_A234_Z_001	漢藥種商試驗模擬問答集	6
KIOM_A235_Z_001	紅疹神方 單	2
KIOM_A238_Z_001	醫林撮要	7

[2005년도 한의학 지식정보자원 디지털화사업에서의 신출한자 현황]

파일명	대상자료명	출현자수
KIOM_A231_Z_012	한방의약 제27호(漢方醫藥 第二十七號)	3
KIOM_A240_Z_001	남도고취(藍島鼓吹)	1
KIOM_A242_Z_001	도해본초(圖解本草)	1
KIOM_A247_Z_001	의감초집(醫鑑抄集)	1
KIOM_A248_Z_001	의문소초(醫門少鈔)	1
KIOM_A249_Z_001	임증지남의안(臨證指南醫案)	2
KIOM_A250_Z_001	전원필고(田園必考)	1
KIOM_A252_Z_001	침구맥결혈명(針灸脈訣穴名)	1
KIOM_A254_Z_001	유문사친(儒門事親)	1
KIOM_A255_Z_001	화제국방(和劑局方)	1
KIOM_A256_Z_001	황제음부경(黃帝陰符經)	1
KIOM_A259_Z_001	의학십서(醫學十書)	1
KIOM_A260_Z_001	본초몽전(本草蒙筌)	1
KIOM_A274_Z_001	다경(茶經)	1
KIOM_A275_Z_001	제민오술(齊民要術)	1
KIOM_A276_Z_001	사시찬요(四時纂要)	1
KIOM_A277_Z_001	농상집요(農桑輯要)	2

4. 한의학 고전자료에서만 나타나는 신출한자 관련 특이사항

한의학 고문헌에서 자주 쓰이는 특별한 약자나 속자 또는 간자의 경우 그 이해도에 따라 판단의 오류가 있을 수 있는데, 한의학 문헌에서 나타나는 특수한 용례를 이해하지 못하여 다른 글자로 판단한다거나, 또는 신출한자로 규정하는 오류를 범해서는 안 될 것이다. 이는 고전자료를 디지털화하여 우리의 문화유산을 보존하고 이를 연구하고 활용한다는 기본 취지에 반하게 될 뿐만 아니라, 문화유산을 해치는 결과를 초래하게 된다.

이러한 예는 한의학 주제뿐만 아니라, 다른 주제나 전공의 경우에도 마찬가지로 발생할 수 있는 문제점으로 각 주제마다 이를 충분히 조사하고 연구하여 정리할 필요가 있다. 이러한 작업들은 한 전문분야에서 통용되는 주관적이고 관습적인 경험의 지식들을 객관화하고 과학화 할 수 있는 좋은 계기라 여겨지며, 이를 취합하여 표준화하고 정보화하여 공유함으로써 인접 학문이 서로 발전할 수 있을 것이며 새로운 부가가치를 창출할 수 있을 것

이다.

한의학 문헌 자료에서만 나타나는 한자의 특수한 용례를 몇 가지 정리하면 아래와 같다.

- 干(방패 간), 召(부를 소) : 薑(생강 강), 棗(대추 조) 대신 사용
/ 예) 강삼조이(薑三棗二 → 干三召二)
- 姜(성 강) : 薑(생강 강) 대신 사용
/ 예) 생강(生薑 → 生姜)
- 攸(돌아갈 귀) : 歸(돌아갈 귀)대신 사용
/ 예) 당귀(當歸 → 当攸)
- 卜(점 복) : 蔔(무우 복) 대신 사용
/ 예) 나복자(蘿蔔子 → 蘿卜子)
- 古(옛 고) : 膏(기름 고) 대신 사용
/ 예) 석고(石膏 → 石古)
- 角(뿔 각) : 殼(껍질 각) 대신 사용
/ 예) 지각(枳殼 → 枳角)
- 別(나눌 별) : 鱗(자라 별) 대신 사용
/ 예) 별갑(鱗甲 → 別甲)
- 癰(악창 웅) : 癰(종기 웅) 대신 사용
/ 예) 응저(癰疽 → 癰疽)

위와 같은 용례는 단순히 웹에서의 구현 문제뿐만 아니라, 검색의 문제, 오해와 오독의 문제 등 많은 문제점들을 지니고 있다. 이에 대한 체계적인 정리와 연구가 필요하다.

IV. 결론

21세기에 접어들면서 고전 전산화 사업이 정부의 국책사업으로 또는 각 대학별 기관별로 자비를 투자하여 연차적으로 진행되고 있다. 그에 따라 많은 결과물이 산출되고 있으며 이러한 데이터의 축적은 한국학 분야 연구자는 물론 어느 누구라도 웹사이트를 통하여 손쉽게 자료를 이용할 수 있게 되었다. 이러한 사업은 우리의 귀중한 전통 문화 자산을 정보화하여 공유함으로써 고전 교육 및 한국학 연구를 지원하는 측면에서뿐만 아니라, 일

반 사용자에게도 매우 의미 있는 일이라 할 것이다.

신출한자 수집과 처리 그 자체가 지식정보자원관리사업이나 또 다른 고전의 DB구축 사업의 목표는 아니다. 그러나 이러한 지식정보자원을 웹상에서 완벽하게 구현하여 서비스 하기 위한 중요한 과제이기도 한다.

그럼으로 표준한자코드를 제정하는 국가기관과 상호협력체제 하에서 효율적인 신출한자의 수집과 관리가 통합적으로 이루어져야 할 것이며, 어느 기관 어느 누구의 이익도 아닌 우리 모두 즉, 국익을 위한 움직임으로 이어져야 할 것이다. 이는 한 사업자나 어떤 협회나 기관에서 대응할 수 있는 문제가 아니다. 국가적인 차원에서의 대응과 적극적인 지원책 마련이 시급하다.

한·중·일 삼국 중 중국과 일본은 이를 위한 많은 작업들을 하고 있으나, 우리나라는 이에 대한 연구와 적극적인 대응이 아직은 많이 부족한 상태이다. 우리 문화유산의 보존 및 연구와 정보화를 위하여, 또 우리의 원활한 정보이용을 위하여 우리가 필요한 한자나 문자가 유니코드에 글자가 없음으로 인하여 입출력이나 재현이 문제가 발생해서는 안될 것이다. 이를 주관하고 검정할 대표기관을 하루빨리 선정하여 국익을 우선하는 국가적인 차원에서의 대응과 적극적인 조치가 필요하다.

[參 考 文 獻]

1. 서경호. 『국제 문자 코드 제안 한자의 표준화에 대한 연구, 상· 하권(연구보고서)』. 서울: 문화관광부, 1998.12.
2. 이재훈. 『국제 문자 코드계의 한자 표준화에 대한 연구(연구보고서)』. 서울: 문화관광부, 2001.12.
3. 정우봉. 『문자코드 표준화 연구 : 21세기 세종계획』. 서울: 문화관광부 · 국립국어연구원, 2003.
4. 민족문화연구원 문자코드연구센터. 문자코드연구센터소식 제13호, 고려대학교 민족문화연구원. 서울: 동센터. 2004. 6
5. 고려대학교 민족문화연구원 문자코드연구센터 『신출한자 처리 관련 업무 협의 - 협의자료』. 서울: 동센터. 2004. 6.
6. 고려대학교 민족문화연구원 문자코드연구센터 『신출한자의 처리와 국제 표준화 - 2004년 문자코드연구센터 워크숍』. 서울: 동센터. 2004. 7.
7. 고려대학교 민족문화연구원. (주)솔트웍스 컨소시엄. 『‘지식정보자원관리사업 신출한자 통합 관리를 위한 목록 및 지침 작성’ 과제 결과 보고서』. 서울: 동연구원. 2004. 12.