

ISO 10646 부호계를 이용한
옛한글 문서 편집기 개발에 관한 연구

이 중화, 김 경석
부산 대학교 전자계산학과

A Study of developing an Old-Hangul Editor
supporting the international standard Hangul code

Lee, Junghwa and Kim, Kyongsok
Department of Computer Science, Pusan National University

요 약 문

국제 표준화 기구 (ISO) 에서는 1992년에 국제 표준 글자 부호계 ISO 10646 을 확정하였는데, 거기에 들어 있는 국제 표준 한글 부호계는 새로운 조합형과 완성형 두 가지로 이루어져 있으며, 새로운 조합형은 옛한글까지 완벽하게 지원할 수 있다. 그러나 아직까지 이 새로운 조합형의 실용성에 대한 실증적인 연구가 없었다.

본 논문에서는, 새로운 조합형 한글 부호계를 지원하는 옛한글 문서 편집기 개발 과정을 살펴 보고, 또한 편집기 개발을 통하여 확인한 다음 두 가지 점에 대해서도 살펴 보겠다. 첫째, 새로운 조합형을 써서 한글 정보 처리를 하는 데 아무 문제가 없으며, 둘째, 새로운 조합형으로 옛한글을 지원할 수 있다는 사실을 확인할 수 있었다.

옛한글 문서 편집기의 성공적인 개발로, 이 때까지 어려움을 겪었던 옛한글 자료의 전산화가 가능하게 되어, 우리의 고전에 대한 연구가 활발해질 것으로 예상된다.

I. 들머리

전자계산기가 점차 보급됨에 따라 우리 일상 생활에 많은 부분까지 전자계산기의 도움을 받고 있다. 그 중 문서 편집기는 전자계산기 사용자가 가장 많이 쓰는 응용 프로그램이다. 지금까지 우리는 많은 문서 편집기를 접해 왔는데, 이들은 사용자의 다양한 요구에 부응하여 그 기능면이나 내용면에서 획기적인 발전을 이루고 있다. 이러한 시점에서 전문적인 분야에 적합한 전자계산기의 이용에 대한 한 보기로 옛한글 문서 편집기 개발에 대해 살펴 보고자 한다.

지금까지 볼 수 있었던 거의 모든 문서 편집기에서는 옛한글에 대한 고려가 되지 않았고, 최근의 몇몇 편집기가 옛한글에 관심을 나타내고 기능을 마련하였지만, 옛한글을 자유로이 표현하기에는 부족한 것이 사실이다. 특히 국내에서 쓰는 거의 모든 한글 부호계가 옛한글을 충분히 고려하지 않았기 때문에, 옛한글 문서를 만든다 하더라도 호환성이 거의 없다. 이러한 문제를 한꺼번에 해결할 수 있는 방안이 국제 표준 한글 부호계에 들어 있는 새로운 조합형 한글 부호계라고 보며, 이 부호계를 써서 옛한글 문서 편집기를 개발하였다.

국제 표준 한글 부호계는 새로운 조합형과 완성형으로 크게 나눌 수 있는데, 이 가운데 새로운 조합형 한글 부호계에는 현재까지 밝혀진 옛한글 글자(자모)를 모두 넣었기 때문에, 옛한글 부호계로 아주 바람직하여, 본 편집기의 내부처리 코드로 썼다. 새로운 조합형에서 첫소리, 가운뎃소리, 끝소리 글자는 각각 2 바이트로 나타내며, 또한 끝소리 채움 글자를 쓰지 않고 부호값을 보고 끝소리 글자(받침)이 있는지 없는지를 알아 낸다. 다시 말해서, 편집기 안에서 두 바이트라는 일정한 길이를 가지는 부호계(완성형 또는 두 바이트 조합형)를 쓰는 문서 편집기와는 아주 다르다.

옛한글을 처리하는 문제에 있어서 본 문서 편집기에서는 옛한글을 효과적으로 입력시킬 수 있는 옛한글 입력 자판과 입력 방법을 제안한다. 또한 방점을 음절의 단위에 포함시켜 운용함으로써 한글 처리단위를 바이트 단위가 아닌 의미적 단위로 처리하고자 하였다.

맺음말에서는 본 작업에서 얻을 수 있는 기대효과 및 활용 방안에 대해 논해 볼 것이다.

II. 옛한글 문서 편집기에 쓴 한글 부호계: 새로운 조합형

한글이 1443년에 처음 만들어진 뒤, 1900 년대에 들어와 맞춤법이 만들어질 때까지 한글은 많은 바뀌어 왔다. 특히 조선시대나 개화기에 쓰이던 우리 말과 글은 지금 우리가 쓰고 있는 것과는 많이 다른데, 이를 가리켜 옛한글이라고 한다. 물론 현재의 맞춤법에 맞는 글자나 소리마디도 옛 문헌에서 썼지만, 여기서는 일반적으로 생각하듯이 옛날에만 쓴 글자를 가리키기로 한다.

2.1. 국제 표준 한글 부호계(ISO 10646)

1992년 6-7월에 열린 ISO-IEJ/JCT1/SC2/WG2 제 22차 회의에서 국제 표준 글자 부호계를 확정하였는데, 거기에는 국제 표준 한글 부호계도 들어 있다. 이 국제 표준 한글 부호계는 크게 완성형과 새로운 조합형의 2 가지로 이루어져 있다.

완성형은 총 6,656 소리마디가 크게 3부분으로 나누어져 수용되었는데 구성은 다음과 같다.

- i) KSC 5601에 들어 있는 2,350 소리마디
- ii) KSC 5657에 들어 있는 1,930 소리마디
- iii) 그 밖에 어느 KSC 에도 들어 있지 않은 2,376 소리마디

상용 조합형에서는 첫소리, 가운데소리, 끝소리 글자를 부호값으로 구별하지 않고, 두 바이트 안의 자리에 따라서 구분하는 했는 것과는 다르게, 새로운 조합형에서는 그 부호값 자체가 다르다. 새로운 조합형은 요즘한글 글자뿐만 아니라 옛한글 글자(자모)에 대해서도 부호값을 주었다. 또 두 바이트 조합형이 2 바이트라는 길이의 제약 때문에 국제 표준이 될 수 없었는데 비해서, 새로운 조합형은 첫소리, 가운데소리, 끝소리 글자를 각각 2 바이트로 나타내고 한 소리마디는 이들의 조합인 4 또는 6 바이트로 나타내기 때문에 국제 표준으로 들어 갈 수 있었다. 옛한글에 나오는 많은 수의 겹글자를 제대로 수용하기 위해서는 이와 같은 새로운 조합형이 아니고는 거의 불가능하다. 새로운 조합형을 위한 240 개의 글자에 대한 부호값은 다음과 같다.

첫소리 글자 : 0x1100 - 0x1159 (90 개)
확장용 : 0x115a - 0x115e
첫소리 글자 채움 : 0x115f

가운데소리 글자 채움 : 0x1160
가운데소리 글자 : 0x1161 - 0x11a2 (66 개)
확장용 : 0x11a3-11a7

끝소리 글자 : 0x11a8 - 0x11f9 (82 개)

확장용 : 0x11fa - 0x11ff

또한 옛한글에서 중요하게 쓰는 방점 (홀점과 겹점) 의 부호값은 다음과 같다. 다만, 10646 문서에 이 방점은 combining 으로 분류되어 있어서 문제가 있을 뿐만 아니라, 이 방점의 용법에 대한 설명도 전혀 없어서 그대로는 도저히 쓸 수 없다. 따라서 앞으로 우리 나라에서 combining 이 아닌 새로운 방점과 그 용법을 정의하여 10646 에 넣기 위한 준비를 하고 있으며, 다음에 인쇄될 10646 에는 용법이 명확한 방점이 새로 들어갈 것이다.

홀점(.) : 0x303e : hangul single dot tone mark

겹점(:) : 0x303f : hangul double dot tone mark

2.2. 새로운 조합형에 들어 있는 옛 글자

본 문서 편집기에서는 국제 표준 한글 부호계에 들어 있는 새로운 조합형 부호계를 썼다. 새로운 조합형 한글 부호계에는 요즘한글 뿐만 아니라, 다음과 같이 모두 171 개의 옛한글 글자가 들어 있으며 그 부호값은 다음과 같다.

옛한글 첫소리 글자 : 71 개 (0x1113 - 0x1159)

옛한글 가운뎃소리 글자 : 45 개 (0x1176 - 0x11a2)

옛한글 끝소리 글자 : 55 개 (0x11c3 - 0x11f9)

따라서 옛한글 글자에 대해서 부호값을 따로 주는 문제가 전혀 없으며, 또한 소리마디를 새로운 조합 방식으로 나타내기 때문에, 옛한글도 요즘한글과 꼭 같은 방법으로 나타낼 수 있다는 장점이 있다.

새로운 부호계의 사용은 기존의 코드체계와의 호환성 문제를 야기 시킨다. 이는 현재 가장 많이 사용하고 있는 완성형이나 조합형 부호계(상용조합형이나 KSC 5601 부호계)와의 상호 코드변환기를 제공함으로써 해결한다. 그러나 기존의 부호계에서는 옛한글 글자에 대한 코드가 고려되어 있지 않으므로 상호 부호값 변환은 요즘한글로 국한된다.

III. 옛한글 문서 편집기에 쓴 한글 자판

3.1. 옛한글 자판: 세벌식과 두벌식

1) 세벌식: 옛한글에서는 옛한글 흘글자를 넣는 방법과, 많은 수의 옛한글 겹글자를 넣는 방법을 정해야 한다. 일단은, 옛한글 자판으로 가장 바람직한 세벌식을 우선적으로 지원하고 있다. 세벌식 가운데서도 이미 많은 사람들에게 익숙한 390 자판을 기본으로 하여, 옛한글 흘글자를 배열하였다.

옛 흘글자의 글쇠 위치를 정할 때는, 요즘한글과 모양이나 성격이 비슷한 글자는 서로 가까운 곳에 배치하여 옛한글 자판을 쉽게 외울 수 있도록 하였다. 이것은 아직 입력 속도를 무시한 채, 외우기 쉬운 면만 강조했으며, 꼭 이상적인 옛한글 자판이라고는 생각하지 않으며, 앞으로 자료를 바탕으로 좀 더 나은 옛한글 자판을 개발할 수 있으리라고 본다. 현재 쓰이고 있는 자판에는 모든 키에 이미 글자가 배당되어 있기 때문에 어쩔 수 없이 윗글자쇠 (SHIFT 키)와의 조합으로 옛한글 글자를 나타내게 하였는데, 보기를 들면 '옛 ㅇ (꼭지 달린 이응)'은 SHIFT+'ㅇ'로 하였다.

옛 겹글자는 그 겹글자를 이루는 각 흘글자를 하나씩 치면 된다. 보기를 들어, 옛한글의 첫소리 글자 'ㅂㅅ ㅓ'는 자판에서 'ㅂ' + 'ㅅ' + 'ㅓ'와 같이 친다. 가운데소리 글자도 이와 마찬가지로인데 옛 가운데소리 겹글자나 옛 끝소리 겹글자도 마찬가지이다.

[부록 1]은 본 문서 편집기에서 기본적으로 제공되는 자판이다.

2) 두벌식: 두벌식으로 옛한글을 지원하는 것이 상당히 복잡하기 때문에, 아직은 지원하지 않고 있다. 그러나, 워낙 두벌식 자판을 쓰는 사람이 많기 때문에, 앞으로 시간이 나는 대로, 두벌식을 바탕으로 한 옛한글 자판도 지원할 예정이다.

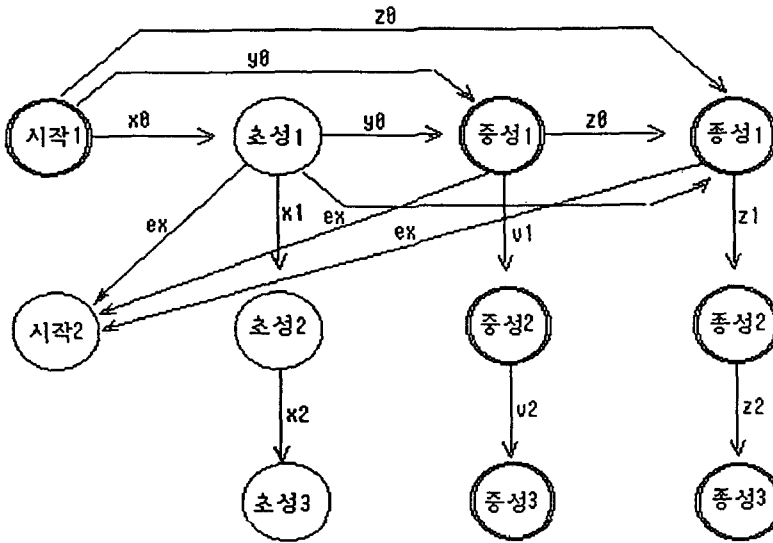
3.2. 사용자 정의 자판

두벌식 자판을 쓰든 세벌식 자판을 쓰든, 옛한글 글자의 배열을 사람에 따라서 다르게 할 수도 있고, 또 앞으로 능률적인 옛한글 자판 개발을 위해서 여러 가지의 자판 배열을 시험해 볼 수 있어야 하므로, 사용자가 원하는 대로 글자 (요즘한글 및 옛한글 모두 포함)를 배열하여 쓸 수 있도록 하는 기능을 제공하게 된다.

사용자가 자판을 정의하는 프로그램을 써서 각 글자의 배열을 정하여 파일 형태로 자판 배열을 저장한 뒤, 옛한글 문서 편집기에서 원하는 자판을 골라서 쓰면 된다.

3.3. 입력 오토마타의 설계

위의 그림에 나온 세벌식 옛한글 자판으로 글자를 입력하면, 아래의 [그림1]에 나오는 입력 오토마타가 동작하여 소리마디로 만들어 준다.



< 상태설명 >
 초성1: 첫소리 글자 입력상태
 초성2: 첫소리 겹글자 상태
 초성3: 첫소리 세 겹글자 상태
 (예, 'ㅂ+ㅅ+ㄱ')
 중성1: 가운뎃소리 글자 입력상태
 중성2: 가운뎃소리 겹글자 상태
 중성3: 가운뎃소리 세 겹글자 상태
 (예, 'ㄷ+ㄱ+ㅡ')
 종성1: 끝소리글자 글자 입력상태
 종성2: 끝소리글자 겹글자 상태
 종성3: 끝소리글자 세 겹글자 상태
 (예, 'ㅁ+ㅁ+ㅁ')
 시작2: 입력을 유지하면서
 시작1 상태로로 분기

< 입력 설명 >
 x_0 : 첫소리 글자
 x_1 : 첫소리 글자 중 겹글자가 가능한 글자
 x_2 : 첫소리 글자 중 세 겹글자가 가능한 글자
 y_0 : 가운뎃소리 글자
 y_1 : 가운뎃소리 글자 중 겹글자가 가능한 글자
 y_2 : 가운뎃소리 글자 중 세 겹글자가 가능한 글자
 z_0 : 끝소리 글자
 z_1 : 끝소리 글자 중 겹글자가 가능한 글자
 z_2 : 끝소리 글자 중 세 겹글자가 가능한 글자
 ex: 그 외의 글자 글자들

[그림 1] 세벌식 옛한글 입력 오토마타

사실 입력 오토마타라는 말을 세벌식에서는 거의 뜻이 없다. 세벌식에서는 첫

소리 글자와 끝소리 글자가 다르게 들어 오기 때문에, 그냥 들어오는 대로 처리하면 된다. 그에 비해서, 두벌식에서는 닿소리 글자로 들어온 것을 첫소리 글자와 끝소리 글자를 나누려다 보니, 입력 오토마타라는 말이 나오게 되었는데, 이런 점에서 보더라도 두벌식은 한글의 특성에 맞지 않는 것 같다.

IV. 옛한글 문서 편집기에 쓴 출력 방식

현재까지 많이 쓰는 문서 편집기는, 하드웨어의 도움없이 프로그램내부에서 한글 출력을 자체적으로 처리한다. 이 방법은 하드웨어적으로 한글을 처리하는 것 보다 처리속도가 떨어지나, 추가 비용이 들지 않기 때문에, 본 옛한글 문서 편집기는 하드웨어를 쓰지 않기로 하였다.

옛한글 출력을 위한 폰트는 프로토 타입으로 세벌체, 24 x 24 비트맵 폰트를 만들어 사용하였다.

V. 옛한글 문서 편집기를 쓸 수 있는 피씨

본 문서 편집기는 MS-DOS version 3.XX이상의 운영 체제를 가지는 IBM-PC AT 이상의 피씨를 대상으로 "C"언어를 사용하여 구현하였다. 그러나 한글 출력 루틴이나 그래픽환경에 필요한 그래픽 루틴과 같이 속도가 중요한 부분에서는 어셈블리 언어로 구현된 부분도 더러 있다. 그래픽 카드는 이미 VGA 그래픽 카드가 일반화 되어있음을 감안하여 이를 대상으로 하였지만, 옛한글 자모를 표현할 수 있는 최소 단위를 24 x 24 픽셀로 하였기 때문에, 최종적으로는 보다 높은 해상도를 나타낼 수 있는 SVGA 그래픽 카드를 장착한 피씨를 주 대상으로 할 것이다. 문서 편집기의 개발 환경과 운용 환경은 같다고 보는데, 가장 좋은 성능을 위해서는 다음과 같은 환경을 권한다.

- . IBM PC-386 이상
- . 1 MB 이상의 주기억 장치
- . VGA 이상의 그래픽 카드
- . 하드 디스크를 장착한 피씨
- . MS-DOS version 3.XX 이상

VI. 옛한글 문서 편집기의 개발로 기대되는 효과 및 활용방안

본 옛한글 문서편집기의 성공적인 개발은, 이 때까지 제한적으로만 이루어지고 있었던 옛한글 자료의 전산화 및 계산기를 이용한 옛한글 연구가 훨씬 더 효율적으로 또한 광범위하게 이루어질 수 있는 환경을 만들었다. 따라서 다음과 같은 효과가 있을 것으로 기대한다.

- 1) 지금까지 전산화에 어려움을 겪고 있는 옛한글 자료를 전산 처리함으로써, 효율적인 옛한글 연구가 가능하게 되었다.
- 2) 국제 표준 한글 부호계를 써서 문서를 만듦으로써, 옛한글 자료의 교환이 가능하게 되었다.
- 3) 보다 나은 옛한글 문서 편집기와 옛한글 자판을 개발할 수 있는 바탕을 마련하였다.
- 4) 국제 표준 한글 부호계의 실용성 검토 및 장.단점 검토를 위한 자료를 얻을 수 있으며, 이를 바탕으로 국제 표준 한글 부호계를 국내에 받아 들일 때의 방향설정을 위한 기본 자료도 얻을 수 있다.
- 5) 국제 표준 한글 부호계를 지원하는 프로그램을 개발할 때에 필요한 기본적인 한글 처리 루틴을 모듈화하였기 때문에, 만일 쏘쓰 프로그램이 공개될 경우에는, 이에 대한 중복 투자를 막을 수 있다.
- 6) 앞으로 옛한글을 전산처리하고자 하는 학자, 교사, 학생들을 대상으로 공급하는 것이 바람직하다고 본다.

VII. 맺음말

본 논문에서는, 국제 표준 한글 부호계에 들어 있는 새로운 조합형 한글 부호계를 지원하는 옛한글 문서 편집기 개발 과정을 살펴 보았다. 편집기 개발을 통하여 우리는 다음과 같은 두 가지 중요한 사실을 확인할 수 있었다. 첫째, 새로운 조합형을 써서 한글 정보 처리를 하는 데 아무 문제가 없으며, 둘째, 새로운 조합형으로 옛한글을 지원할 수 있다는 사실을 확인할 수 있었다. 본 연구는 개발의 상업적 측면보다는 옛한글에 대한 연구적 측면이 강조되어야 하겠다.

옛한글 문서 편집기의 성공적인 개발로, 이때까지 어려움을 겪었던 옛한글 자료의 전산화가 가능하게 되어, 옛한글에 대한 연구가 활발해질 것으로 예상된다. 또한 이를 촉진시키기 위해서는, 개발된 옛한글 문서 편집기를 옛한글을 연구하는 교수, 교사, 학생에게 보급하는 것이 바람직하다고 본다. 앞으로 이른 시일 안에 옛한글 폰트를 개발해야 하는 과제가 남아 있다.

앞으로의 연구는, 보다 나은 기능을 가진 옛한글 문서 편집기를 개발하는 방향으로 나아가. 옛한글 연구에 많은 기여를 할 수 있게 되기를 바란다. 또한 옛한글에 대한 연구가 활성화 되기를 기대해 본다.

* 본 연구는 국고 지원을 받아서 한글 학회의 주관 아래 이루어지고 있으며, 개발이 끝난 뒤의 옛한글 문서 편집기의 일반 공개 여부는 문화체육부에서 결정할 예정이다.

[참고문헌]

- [1] “한글 사전 편찬 전산화를 위한 터잡이: 옛 한글 가나다 순 및 옛 한글 자판”, 김 경석. 교육 한글 제 2 호 (1989), 5-50 쪽. 한글 학회.
- [2] 표준형 한글 vi 연구 최종 보고서. 연구 책임자 김 경석. 부산대학교 부설 정보 과학 종합 연구소. 1992. 12.
- [3] 홍 운표 외 8인, “한글 옛글자의 컴퓨터 처리 방안에 대한 연구”, 한국어 전산학 창간호, 21 쪽, 1991.
- [4] ISO/IEC 10646-1:1993(E). International Standard. 1st edition. Information technology - Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and Basic Multilingual Plane, May 1, 1993. ISO (the International Organization for Standardization).
- [5] Kyongsok Kim, "A common Approach to Designing the Hangul Code and Keyboard", Computer Standards & Interfaces, Vol. 14, No. 4, pp. 297-325 1992.

[부록 1] 세벌식 옛한글 입력 자판

옛한글 자판 (390 자판 기준)

~	°	!	@	#	\$	%	^	&	*	()	=	+	
q	w	e	r	t	y	u	i	o	p	{	}			
A	S	D	F	G	H	J	K	L	:	;	'	"		
Z	X	C	V	B	N	M	<	>	^	?				