(NATURAL SCIENCE)

Vol. 62 No. 3 JUCHE105 (2016).

주체105(2016)년 제62권 제3호

조선어고속입력체계의 한가지 실현방법

리충혁, 방도일

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《새로운 과학기술분이를 개척하기 위한 사업도 전망성있게 밀고나가야 합니다. 나라의 과학기술을 세계적수준에 올려세우자면 발전된 과학기술을 받아들이는것과 함께 새로운 과학기술분이를 개척하고 그 성과를 인민경제에 적극 받아들여야 합니다.》(《김정일선집》 중보관 제11권 138~139폐지)

세계적으로 매개 민족어의 형태론적 및 통계론적특성분석에 기초하여 본문편집체계에서의 고속성을 보장하기 위한 연구 및 개발사업이 진행되고있다.

조선어입력체계와 관련한 연구에서는 조선어자모와 글자들의 확률통계적특성과 관련한 연구, 조선어문자판의 최량설계와 관련한 연구들이 진행되여 도입[2, 3]되고있지만 입력작업의 고속화와 관련한 연구는 진행된것이 없다.

론문에서는 초성렬패턴의 정합방법에 의한 조선어고속입력체계를 실현하기 위한 한가지 방법에 대하여 론의한다.

1. 조선어문자입력체계와 관련한 선행연구분석

조선어문자들은 자음글자(초성)와 모음글자(중성), 받침(종성)의 각이한 결합에 의하여 이루어진다.

초성종류는 19개로서 여기에는 《ㄱ, ㄴ, ㄷ, ㅁ, ㅂ, ㅅ, ㅇ, ㅈ, ㅊ, ヲ, ㅌ, ㅍ, ㆆ, ㄲ, ㄸ, ㅃ, ㅆ, ㅉ»가 속하며 중성종류는 21개로서 여기에는 《ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ, ㅣ, ㅐ, ㅒ, ㅔ, ㅖ, ㅚ, ㅟ, ㅚ, ㅚ, ㅚ, ㅚ, ㅚ, ㅚ, ㄲ 속한다. 그리고 종성종류는 27개로서 홑받침과 둘받침으로 구분되는데 흩받침은 16개로서 《ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅇ, ㅈ, ㅊ, ȝ, ㅌ, ㅍ, ㅎ, ㄲ, ㅆ»가, 둘받침은 11개로서 《ㄳ, ㅆ, ㄶ, ㄲ, ㄲ, ㄲ, 朮, 朮, 朮, 朮, 朮, ㅆ, ㅂ, 朮 朮, 朮, 朮, 朮, ㅆ, ▷다.[1]

한편 매몰형체계에서 조선어입력을 위한 문자판설계에서는 《평양》손전화기에서의 9개 누름건반방식과 일반쿔퓨터 및 《아리랑》손전화기에서의 26개 접촉식건반방식이 연구되여 리 용되고있다.

콤퓨터와 손전화기를 비롯한 매몰형체계들에서의 조선어고속입력건반형식으로서는 이미 실천에서 리용되고있는 26건반방식이 합리적이라는것을 알수 있다. 그리고 초성렬입력 방식으로 건반입력방식을 절체하였을 때 건반구역을 초성건반구역만을 가지고 확대하여 건반령역을 넓게 하는 방법을 쓸수 있으며 또는 조선어건반입력함수를 개조하여 초성건반입력에서 둘받침으로 인식하는 현상을 제거할수도 있다.

우의 선행연구로부터 얻어지는 결론은 조선어자모음들을 다른 나라 문자에서처럼 더 기

초적인 필획으로 분리하는것은 의의가 없으며 1개 조선어문자입력과 관련하여서는 건반배 치와 건반입력속도의 견지에서 이미 최량화되여있다는것이다.

따라서 조선어고속입력과 관련한 연구과제는 초성렬입력에 의한 단어입력방법을 실혀 하는데 귀착된다.

2. 초성렬패턴입력에 의한 조선어고속입력체계의 설계

초성렬입력에 의한 단어추출문제는 고전적인 패턴인식문제로 귀착시킬수 있다.[1] 여기서 주어진 초성패턴렬 X 에 대하여 최대사후확률 P(W|X) 를 가지는 해당한 W 렬을 찾 아야 하는데 이것은 베이스공식에 의하여 다음과 같이 표현할수 있다.

$$\hat{W} = \underset{W}{\operatorname{arg max}} P(W | X) = \underset{W}{\operatorname{arg max}} \frac{P(W)P(X | W)}{P(X)}$$

1) 체계대면부 설계

우선 기존의 입력방식이 있는데 이것은 《초성+중성(+종성)》입력에 의한 편집방식으 로서 이때에는 해당한 문자들이 초성렬편집창에 현시되다가 공백건이 입력되면 본문편집 창에 해당하 단어가 넘어간다. 하지만 단어후보렬목록창에는 단어후보들이 뜨지 않는다.

다음 초성렬입력에 의한 단어입력방식이 있는데 이것은 《초성 1+초성 2+(초성 3+ 초성 n)》입력에 의한 편집방식이다.

이때 단어후보모임에서 필요한 단어의 선택은 요구되는 단어에 해당한 번호입력 또는 마우스찰칵, 손접촉방식에 의해 진행한다.

2) 초성렬입력에 의한 단어고속입력체계설계

우선 건반에 의한 초성렬입력형식은 다음과 같다.

〈초성 1+초성 2+[초성 3+초성 n]〉

여기서 보는것처럼 입력때 초성렬의 길이는 최대 4까지로 제한되며 이때 초성 4는 단 어에서 제일 마지막문자의 초성을 의미한다.

한편 초성렬패턴사전의 구조설계는 다음과 같이 한다.

D(초성 1, 초성 2)={WPi |i=1, n}

WPi = (Wi, Pi)

여기서 Wi는 i번째 단어, Pi는 Wi의 우선순위값이다.

초성렬패턴사전을 리용한 후보렬탐색을 다음과 같이 한다.

- ① 초성렬입력패턴 <초성 1+초성 2+[초성 3+초성 n]>이 입력되면 초성렬패턴사전 D(초 성 1. 초성 2)에서 이 입력패턴과 최종일치되는 단어들을 선출하여 1차단어후보렴을 만든다.
- ② 1차단어후보렬가운데서 본문안에 이미 입력된 단어모임에 속하는 단어들의 우선권 을 2준위로 설정한다.
- ③ 1차단어후보렬에 대하여 우선순위값에 의한 순서화를 진행하여 2차단어후보렬을 만 든다.
 - ④ 2차단어후보렬을 단어후보렬목록창에 출력한다.

체계의 성능에서 관건적열쇠는 2차단어후보렬에서 사용자가 요구하는 단어가 적어도 5 ~6번째 후보안에 들어가도록 하는것이다.

같은 초성렬패턴에 속하는 단어후보렬에서 사용자가 요구하는 단어들의 우선순위를 높여주는데 필요한 정보들로서는 다음과 같은것들을 들수 있다.

- ① 편집체계에서 고정적으로 자주 쓰이는 단어부류정보 즉 정적으로 제일 많이 쓰이는 단어들에 대한 정보이다.
- ② 편집중의 본문주제와 관련하여 자주 리용되는 단어부 류정보이다.
 - ③ 단어의 빈도크기순서정보[2]이다.
 - ④ 단어n그람사전정보이다.
 - ⑤ 품사n그람사전정보이다.

우와 같은 론의에 기초한 조선어고속입력체계대면부는 그림과 같다.

본문편집창	단어후보렬 목록창	
초성렬편집창		
건반		

그림. 조선어고속입력체계 대면부

3. 성 능 평 가

제안한 조선어고속입력체계와 일반입력체계의 성능을 비교하기 위하여 여러가지 문서에 대한 입력실험을 진행하였다.

실험결과는 표와 같다.

표. 문서입력실험결과

— — — — — — —			
문서종류	일반건반입력방법에서 입력시간/s	고속입력체계에서 입력시간/s	성능/배
소설(71KB)	315	78	4.0
정보과학(71KB)	402	148	2.7
물리 학(71KB)	457	217	2.1
화학(71KB)	534	295	1.8

표로부터 초성렬패턴의 정합방법에 의한 입력체계에서는 일반입력체계에서보다 입력 속도에서 약 3배의 성능을 가진다는것을 알수 있다.

참 고 문 헌

- [1] 박성호 등; 음성정보처리, **김일성**종합대학출판사, 175, 주체95(2006).
- [2] 정순기: 조선어빈도수사전, 과학백과사전종합출판사, 34~47, 1993.
- [3] 권종성; 조선어정보처리, 과학백과사전종합출판사, 56~77, 1994.

주체104(2015)년 11월 5일 원고접수

A Method of Establishment of Korean Input System with High Speed

Ri Chung Hyok, Pang To Il

On the basis of analysis for procedural study concerned with Korean input system and modelling the system, we designed system interface, and studied a method to implement the Korean input system with high speed by matching initial sound pattern.

Key words: Korean input system, initial sound, vowel sound, final vowel sound