

# 문서 입력을 위한 한/영 자동 토글 시스템

최연우, 서동린, 김용수, 박영호, 이금석(동국대학교)  
김철호(국방정보체계연구소)

Hangul/English Automatic Toggle System for Document Input

Y. W. Choi, D. R. Seo, Y. S. Kim, Y. H. Park, K. S. Lee(Dongguk University)  
C. H. Kim(Institute for Defense Information Systems)

## 요 약

컴퓨터의 보급과 활용이 늘어남에 따라 한글과 영문이 혼합된 문서 작성의 필요성이 계속 증대되고 있다. 컴퓨터 하드웨어 및 소프트웨어에서 사용자의 편리성과 효율성을 최대한 배려하는 "사용자 인터페이스"에 대한 연구가 증대되고 있어서 보다 편리한 한/영 혼용문서의 작성 방법이 요구되고 있다.

키보드를 통한 한/영 혼용문 입력시에 입력된 문서의 내용을 입력 오토마타 및 사전 수록 정보를 토대로 분석하여 자동적으로 한글 또는 영문으로 변환하는 자동 토글 시스템 개발에 대해 설명하고, 현재의 시스템이 해결하지 못한 문제점 및 개선 방향을 기술한다.

## I. 서 론

최근 컴퓨터 하드웨어 및 소프트웨어 개발에 있어 시스템과 사용자간의 상호작용 즉, "사용자 인터페이스"에 대한 관심이 증대되고 있다. 사용자 인터페이스 연구의 궁극적인 목표는 보다 편리하고 효율적인 인간과 컴퓨터간의 상호작용 수단을 제공하는 것이며, 최근에는 단지 사용자에게 친숙한 또는 즐거움을 주는 인터페이스 단계를 넘어 음성 인식, 컴퓨터 시각, 자연어 처리 등 인공지능 분야에서 연구된 기술들을 응용한 지능형(intelligent) 사용자 인터페이스 개발이 연구되고 있다[1][2].

한편, 컴퓨터 보급과 활용이 늘고 정보화 사회의 성숙에 따라 한글과 영문이 혼합된 문서 작성의 필요성이 계속 증대되고 있다. 현재 대부분의 한/영 혼용문서 작성 방식은 한글과 영문을 각각 입력할 때마다 한/영 모드 변환을 지정하는 이른바 토글 키(toggle key)를 사용하고 있다. 즉, 문서 작성중 토글 키를

이용하여 "지금 입력하는 것은 한글" 또는 "지금 입력하는 것은 영문"이라는 것을 매번 지정한 후 문서 내용을 입력하는 방식이 이용되고 있다. 그러나 이와 같은 한/영 혼용문서 작성 방식은 토글 키 지정에 소요되는 시간과 작업의 번거로움으로 인해 문서 작성의 생산성 저하와 오류 증대 등의 제반 문제를 초래한다. 따라서 이와 같은 문제에 대한 개선을 통한 보다 편리하고 효율적인 한/영 혼용문서 작성 방안이 요구된다[3].

본 연구에서는 키보드를 통한 한/영 혼용문서 작성시, 입력된 문서의 내용을 오토마타 및 사전 수록정보를 토대로 분석하여 자동적으로 한글 또는 영문으로 변환하는 즉, 자동토글 기능을 갖는 시스템을 개발하고자 한다.

## II. 한/영 자동 토글 시스템 설계

### 1. 자동 토글의 개념

자동 토글은 현재 사용자에게 일임되어 온 토글 키를 통한 입력 모드 변환 기법을 개선한 것으로, 문서 입력 모드를 변환할 때 토글 키를 사용하지 않고 입력된 문서 내역을 분석하여 컴퓨터 내부에서 자동적으로 변환하도록 하는 기법이다.

자동 토글의 관건은 토글변환 시점을 정확히 알아내는 것이다. 이를 위해서는 지금까지 인간의 지능으로 수행되던 입력 모드 변환을 컴퓨터가 자동적으로 판별할 수 있도록 기존의 입력 오토마타를 수정해야 하며 검사될 단어들을 수록한 여러 사전들을 준비해야 한다.

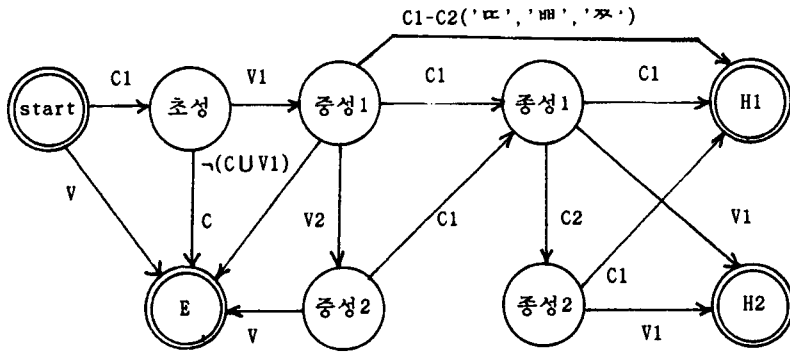
### 2. 자동 토글을 위한 한/영 혼용 문서 특성 분석

자동 토글 기법을 구현하기 위해 한글과 영문이 혼용된 문서 특성을 단어를 중심으로 살펴보면 다음과 같이 크게 네 가지로 구분된다.

- 단어가 모두 한글로 구성된 경우
- 단어가 모두 영어로 구성된 경우
- 단어가 영어+한글로 구성된 경우
- 단어가 한글+영어로 구성된 경우

첫째, 단어가 모두 한글로 구성된 경우는 자동 토글 기법의 기본이 되는 입력 모드가 된다. 둘째, 단어가 모두 영어로 구성된 경우를 위해 기존의 입력 오토마타를 수정이 필요하고, 영어가 한글로 조합이 가능한 경우에 대비하여 이를 검사할 영어 단어 사전을 구성해야 한다. 셋째, 영어+한글로 구성된 단어일 경우에는 둘째 경우의 영어 단어 사전 외에 뒷부분의 한글을 검사할 한글 단어 사전을 준비해야 한다. 영어+한글로 구성되는 단어는 영어 단어에 한글 조사나 어미가 붙는 경우가 대부분이다. 넷째는 한글+영어로 구성되는 단어인데 거의





- Arc Description

- C1: 단자음, 쌍자음
- C2: 복자음
- V1: 단모음
- V2: 복모음

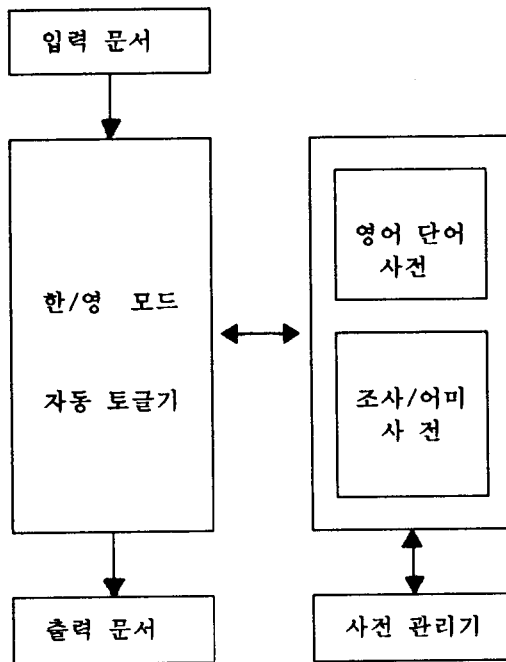
- State Description

E: 한글 조합이 실패하여 현재 입력된 문자열을 영문으로 변환하고 영문모드로 바뀌는 상태.

[그림 2] 수정된 입력 오토마타

4. 시스템 구성

한/영 모드 자동 토글을 위한 시스템 구성도는 그림 3과 같다.

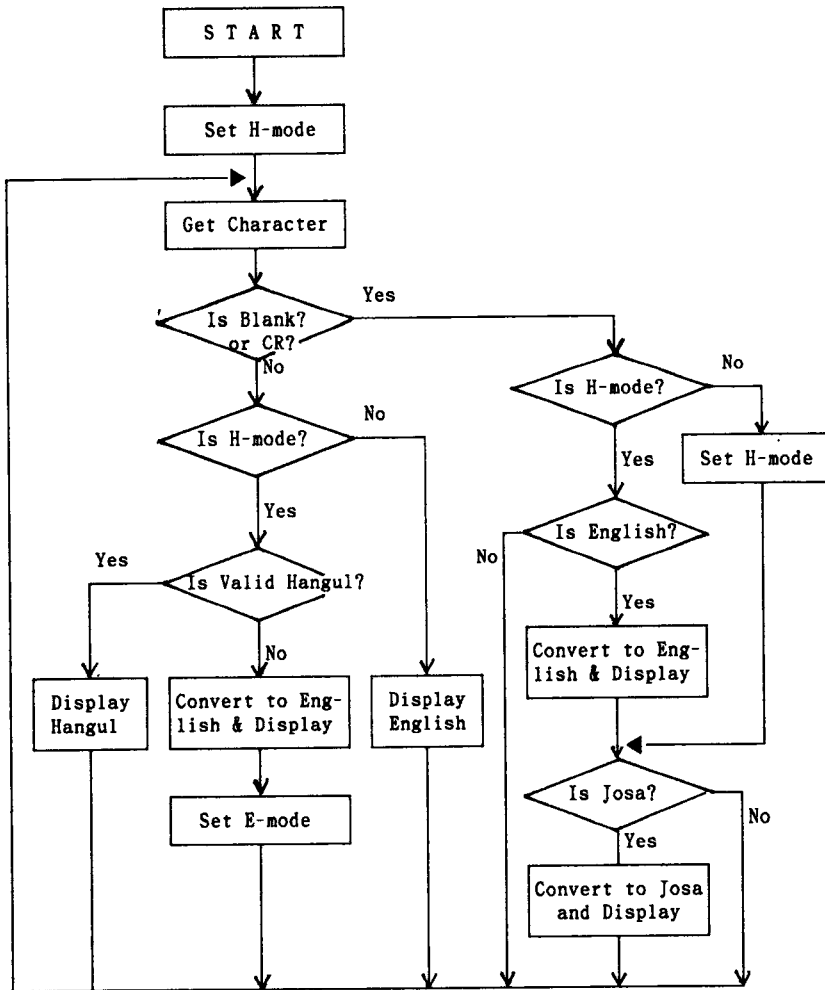


[그림 3] 한/영 모드 자동 토글 시스템 구성도

한/영 모드 자동 토글기는 입력되는 문서를 단어 단위로 분리한 후, 입력 오토마타를 참조하여 한글로 조합이 불가능한 단어는 영어로 출력하고, 한글로 조합이 가능한 영어 단어나 영어와 한글이 혼합된 단어에 대해서는 영어 단어 사전 및 조사/어미 사전을 토대로 자동적으로 한글 또는 영어로 출력한다. 영어 단어 사전에는 모든 영어 단어가 수록되는 것이 아니라 한글로 조합이 가능한 단어만 추출하여 수록된다. 또한, 조사/어미 사전은 영어+한글로 구성되는 단어를 위해 구성되는 사전으로서 한글 부분에 해당하는 조사나 어미 문자열을 수록하게 된다. 사전 관리기는 단어의 입력, 수정, 검색 등의 기능을 제공한다.

5. 시스템 처리과정

한/영 모드 자동 토글기의 전체 처리 과정은 그림 4와 같다.



[그림 4] 한/영 모드 자동 토글기 처리 과정도

그림 4에서 보는 바와 같이 자동 토글기의 기본이 되는 모드는 한글 모드이다. 키보드를 통해 입력된 문자는 버퍼에 저장되고 입력 오토마타에 의해 한글로 조합되며 한글 조합이 실패한 문자열은 버퍼에 저장되어 있던 영어 문자열로 대체된다. 문자열은 스페이스 바(space bar)나 리턴 키(return key)에 의해 토큰으로 분리되고, 한글로 조합된 토큰은 사전에 수록된 단어와 비교하여 동일한 단어가 검색되면 해당 문자열로 변환된다.

### III. 시스템 구현

설계된 시스템은 IBM PC/AT 기종에서 C 언어로 구현되었으며, 논란의 여지는 있지만 현재 가장 많이 사용되고 있는 한글 2벌식 자판과 영문 QWERTY식 자판을 모델로 하였다.

영어 단어 사전은 중학 필수 1200단어를 범위로 하여 구성한 결과, 한글로 조합이 가능한 영어 단어는 51개였으며, 이들 중에서 2개는 한글로도 의미를 가지고 있었다. 또한, 이를 통해 대부분의 영어 단어중 비교적 적은 수의 알파벳으로 구성되는 단어들이 한글 조합에 성공하는 예가 많았으며, 영어로도 의미를 가지고 한글로도 의미를 가지는 단어는 매우 적음을 알 수 있었다.

### IV. 시스템 평가 및 활용방안

#### 1. 시스템 평가

개발된 시스템은 시험 결과 대부분이 초기 설계 의도대로 동작하였으나, 다음과 같은 몇가지 문제점 또한 발견되었다. 앞으로 보완·개선되어야 할 주요 문제들과 그 해결책을 살펴보면 다음과 같다.

##### (1) 한글, 영어 모두 의미를 가지는 단어 처리

한글로 '해', '재가'에 해당하는 영어 문자열은 각각 'go', 'work'이며 이러한 단어들은 한글 및 영어 모두 의미를 가지고 있다. 이와 같이 양쪽 모두 의미를 가지는 단어일 경우에는 입력한 문자가 한글인지 또는 영어인지를 판별하기가 대단히 어려우며, 이는 자동 토글 기법의 완벽한 구현을 가로막는 큰 걸림돌이 되고 있다. 구현된 시스템에서는 이러한 문자열을 모두 영어로 변환하고 있으나, 보다 바람직한 해결을 위해서는 인공지능의 휴리스틱 문제해결 방법을 도입해야 한다.

##### (2) 한글 조사를 포함하는 영어 단어 처리

예를 들어 'mechanisms'와 같은 단어는 한글 조사인 '는'에 해당하는 문자열('sms')을 포함하고 있으며, 화면 상에는 'mechani는'과 같이 출력된다. 이 문제를 해결하기 위해서는 한글 조사 및 어미를 포함하고 있는 영어 단어들을 사전에 등록하여 한글로 바뀌지 않도록 해야 한다.

### (3) 영어 사전에 등록된 단어에 한글 조사나 어미가 붙는 경우 처리

'girl은'이라는 문자열을 입력하고자 할때, 화면에는 '하기은'이 출력되다가 'girl'이란 단어가 영어 사전에 등록되어 있으므로 '하기'가 'girl'로 대체가 되는데 이때 'girl' 뒤의 조사인 '은'까지 영어로 대체되어 결국 'girldms'으로 출력 된다. 이 문제는 영어로 변환된 문자열에 대해서 조사 또는 어미가 포함되는지를 검사하면 해결된다.

### (4) 한글 입력시 오류가 발생하는 경우 처리

예를 들어 '이와'라는 단어를 입력하고자 할때 오류가 발생하여 'ㅏ'를 'ㅑ'로 잘못 입력한 경우 'ㅏ'와 'ㅑ'는 복모음을 형성할 수 없기 때문에 한글 조합에 실패하여 영문으로 변환된다. 즉, 문서 입력자의 의도와는 달리 'dl dhj'로 출력된다. 이를 해결하는 방법으로는 백스페이스 키(back space key)로 틀린 문자를 지우고 (즉 'j'를 지우고) 'dl dh'에 대한 한글 조합인 '이오'로 변환하여 화면에 출력한 다음 한글 모드에서 입력을 기다리도록 하면 된다.

### (5) 사전 확장 및 사전 관리기 개선

현재 중학교 필수 단어 범위에 한정된 사전을 좀 더 확장하고, 사전에 등록된 단어를 효율적으로 관리할 수 있도록 사전 관리기가 개선되어야 한다. 특히, 단어 검색 속도는 시스템의 성능에 큰 영향을 미치는 매우 중요한 요소이므로 이에 대한 효율적인 알고리즘이 연구 적용되어야 한다.

## 2. 시스템 활용 방안

현재의 컴퓨터 기술 수준으로 자동 토글 기법을 완벽히 구현한다는 것은 대단히 어려운 일이다. 따라서 아직은 본 연구에서 제안된 자동 토글 기법이 기존의 토글 방식을 완전히 대체할 수는 없다. 그러나 기존 방식과 상호 보완적으로 즉, 일단 자동 토글 모드에서 문서를 입력한 후 기존 토글 방식을 이용하여 잘못 처리된 문자열을 수정하는 방법을 활용한다면 효율적인 문서 작성이 가능할 것이다. 특히, 약어를 제외한 전문 용어는 한글로 조합되는 경우가 별로 없기 때문에, 이러한 용어를 사용하여 작성된 문서 입력시 효과적이라 생각된다.

## V. 결 론

본 연구에서는 한/영 혼용문서 작성시 토글 키 지정을 통한 문서작성 작업의 생산성 저하와 오류 증대 문제를 개선하기 위한 하나의 방안으로 오토마타 및 사전 수록정보를 토대로 입력되는 문서의 내용을 분석하여 자동적으로 한글 또는 영문으로 변환하는 자동 토글 시스템을 설계하고 그 프로토타입 시스템을 개발하였다.

본 연구를 통해 개발된 자동 토글 시스템 이용은 한/영 혼용문서 작성의 생산성 증대와 오류 감소 효과를 기대할 수 있으나 보다 완벽한 자동 토글 기능을 갖기 위해서는 앞으로 다음과 같은 부분에 대한 추가적인 연구와 보완이 뒤따라야 한다.

첫째, 현재 구축된 영어 단어 및 조사/어미 사전에 보다 많은 단어와 조사/어미 등을 수록함과 동시에 단어 체크에 소요되는 시간을 줄일 수 있는 사전 구성 방법 및 고속의 단어 검색 알고리즘이 고안 적용되어야 한다.

둘째, 궁극적으로 완전한 자동 토글 개념을 실현하기 위해서는 영어로도 의미가 있고 한글로도 의미가 성립되는 단어에 대한 보다 효과적인 처리 방법이 모색되어야 한다.

## 참 고 문 헌

- [1] Ben Shneiderman, Designing the User Interface, Addison-Wesley, 1992
- [2] James Allen, Natural Language Understanding, The Benjamin/cummings Publishing Company, Inc., 1987
- [3] "인텔리전트 유저 인터페이스 구현을 위한 자동 토글 기법 연구", 마이크로 소프트웨어, pp220~223, Oct. 1991
- [4] 이준희, 정내권, 컴퓨터 속의 한글, 정보시대사, 1991