

국어 어문 규정 검색 프로그램 개발

오형진, 이신원*, 두길수**, 정성중, 안동언
전북대학교 전자정보공학부, *정인대학, **서남대학교
전화 : 063-270-2416 / 핸드폰 : 016-9860-2489

Development of Korean Language Regulation Retrieval System

Hyung Jin Oh, Shin-Won Lee*, Kil-Su Du, Sung Jong Chung, Dong Un An.
Computer Engineering of Chonbuk National University, Chongin College*, Seonam
University
E-mail : hyungjin@duan.chonbuk.ac.kr

Abstract

국어 어문 규정 검색 프로그램은 21세기 세종계획 프로젝트의 한민족 언어 정보화 분과에서 개발한 것으로, 한글 맞춤법, 표준어 규정, 외래어 표기법, 국어의 로마자 표기법 등의 어문 규정과 검색어에 대한 어문 규정 설명을 HTML 문서로 보여준다. 본 논문에서는 국어 어문 규정을 검색하기 위한 데이터베이스의 구조와 국어 어문 규정 검색 프로그램의 검색 방법에 대해서 논한다.

I. 서론

21세기 세종계획 프로젝트의 한민족 언어 정보화 분과에서 개발한 국어 어문 규정 검색 프로그램은, 문교부 고시 제88-1호(1988. 1. 19.)에 따라 같은 해 3월 1일부터 시행하게 된 '한글 맞춤법 규정'과 문교부 고시 제88-2호(1988. 1. 19.)에 따라 시행하게 된 표준어 규정'과 문교부 고시 제85-11호(1986. 1. 7.), 문화부 고시 제1992-31호(1992. 11. 27.), 문화체육부 고시 제1995-8호(1995. 3. 16.)에 따라 시행하게 된 외래어 표기법, 그리고 문화관광부 고시 제2000-8호(2000. 7. 7.)에 따라 시행하게 된 국어의 로마자 표기법을 검색할 수 있도록 만들어졌다[1][2][3][4].

국어 어문 규정 검색 프로그램은 한글 맞춤법

검색 시스템, 표준어 규정 검색 시스템, 외래어 표기법 검색 시스템, 국어의 로마자 표기법 검색 시스템 등의 네 개의 시스템으로 구성되어 있지만, 독립된 시스템으로 구성되어 있는 것이 아니라 네 개의 시스템을 통합하여 단일 인터페이스에서 규정과 표제어를 검색할 수 있도록 설계하였다. 통합 인터페이스로 설계한 이유는 일반인은 자신이 찾고자 하는 검색어와 그와 관련된 규정이 어느 규정에 속하는지 잘 알지 못하기 때문이다.

국어 어문 규정 검색 프로그램의 통합된 인터페이스는 다음과 같은 메뉴로 구성되어 있다.

첫째, '한글 맞춤법', '표준어 규정', '외래어 표기법', '국어의 로마자 표기법'의 규정과 이를 설명한 부분을 메뉴 창에서 "규정"으로 표시하였다.

둘째, 이용자가 알고자 하는 표제어를 직접 검색하여 옳고 그름을 알 수 있도록 하여 주는 부분을 메뉴 창에서 "검색"으로 표시하였다. 이용자가 직접 검색어를 입력하면 그 검색어에 적용되는 모든 어문 규정을 찾아볼 수 있도록 하였다.

셋째, 이용자가 검색 영역과 검색 범위를 선택한 후에 검색 결과에서 나열된 검색어들 가운데 한 검색어를 선택하면 그 검색어에 관한 규정을 찾아볼 수 있는 부분을 메뉴 창에서 "부분 검색"으로 표시하였다.

넷째, 이용자가 한글 맞춤법, 표준어 규정, 외래어

표기법에 관한 자신의 실력을 직접 알아 볼 수 있도록 한 부분을 메뉴 창에 “게임”으로 표시하였다.

본 논문에서는 검색어와 그와 관련된 국어 어문 규정을 검색하기 위한 데이터베이스의 구조와 국어 어문 규정 검색 프로그램의 검색 방법에 대해서 논하며 구성은 다음과 같다. 2장에서는 국어 어문 규정 검색 프로그램의 데이터베이스 구축에 대하여 살펴보고, 3장에서는 국어 어문 규정 데이터베이스의 검색에 대하여 논의 하며, 끝으로 4장에서 결론을 맺는다.

II. 국어 어문 규정 검색 프로그램의 데이터베이스 구축

2.1 국어 어문 규정 데이터베이스의 구조

한글 맞춤법, 표준어 규정, 외래어 표기법, 국어의 로마자 표기법의 데이터베이스는 각각 독립된 테이블에 저장되어 있지만 같은 구조를 가지고 있다.

표 5 규정 테이블의 구조

필드명	설명
ID	행 번호, 원시 자료를 입력할 때 자동으로 생성한다.
Title	표제어
EntryWord	규정에 맞는 표기와 틀린 표기에 대한 목록
Related	관련항, 검색어에 관련 국어 어문 규정
Example	예문
Description	설명

네 개의 규정이 같은 구조를 가져야 통합된 인터페이스를 구성하고, 검색 프로그램을 체계적으로 작성할 수 있다. [표 1]은 규정 테이블의 구조를 보여주고 있다.

이들 규정 테이블의 EntryWord에 있는 목록들을 각각 검색할 수 있어야 한다. 따라서, 네 개의 규정 테이블 각각에 대해서 EntryWord에 있는 목록을 검색할 수 있도록 색인 테이블을 정의하여 작성하였다. 색인 테이블의 구조는 [표 2]와 같다.

표 6 색인 테이블의 구조

필드명	설명
ID	행 번호, 원시 자료를 입력할 때 자동으로 생성한다.
SearchWord	검색어
Key	규정 테이블의 ID

SearchWord는 EntryWord에 있는 규정에 있는 표기와 틀린 표기에 대한 검색어이다.

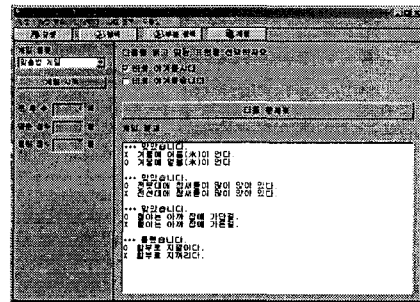
2.2 게임 데이터베이스의 구조

국어 어문 규정 검색 프로그램에는 사용자가 자신이 얼마나 국어 어문 규정을 알고 있는지 스스로 테스트해 볼 수 있는 게임이 있다. 게임 데이터베이스의 구조는 [표 3]과 같다.

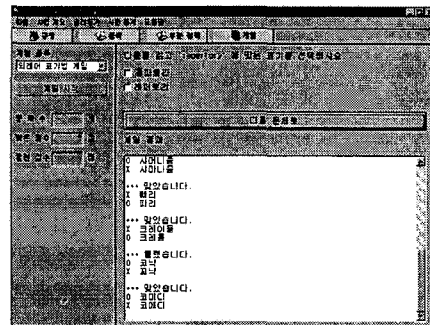
[표 3] 게임 데이터베이스의 구조

필드명	설명
ID	행 번호, 원시 자료를 입력할 때 자동으로 생성한다.
Right1	맞는 표현 1
Right2	맞는 표현 2
Wrong	틀린 표현
Foreign	외래어 원문

다음 <그림 1>과 <그림 2>는 한글 맞춤법 게임과 외래어 표기법 게임을 수행하는 화면이다.



<그림 3> 한글 맞춤법 게임 화면



<그림 4> 외래어 표기법 게임 화면

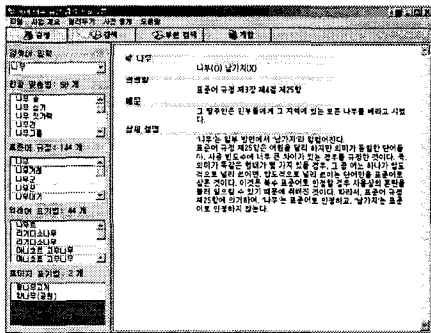
II. 국어 어문 규정 데이터베이스의 검색

국어 어문 규정 검색 프로그램은 Microsoft Visual C++로 작성되었으며[7], 데이터베이스는 Microsoft Access를 사용하였다.

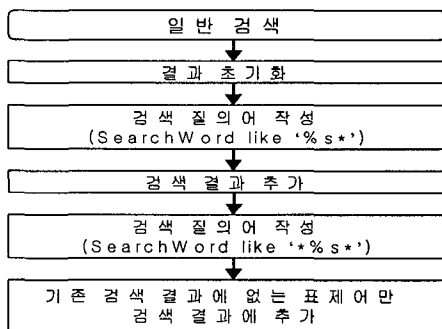
데이터베이스를 검색하는 방법에는 두 가지가 있다. 첫 번째 방법은 검색어를 입력하여 해당 검색어를 포함하는 모든 표제어를 검색하는 일반 검색 방법이다. 두 번째 방법은 검색어의 범위를 지정하여 검색을 수행하는 부분 검색 방법 방법이다.

3.1 일반 검색

일반 검색은 검색 메뉴를 이용하여 수행하며, 검색 결과로 해당 검색어와 일치하거나 검색어가 앞에 나오는 표제어를 먼저 표시하고, 검색어를 포함하는 표제어를 뒤에 표시한다. <그림 3>에 검색어 “나무”를 입력하여 검색한 결과이다.



<그림 3> 일반 검색 결과 화면



<그림 4> 일반 검색 흐름도

<그림 4>는 일반 검색의 흐름도이다. 두 번

에 걸쳐서 검색 질의어를 작성하여 검색한다. 첫 번째 질의어는 해당 검색어와 일치하거나 검색어가 앞에 나오는 표제어를 찾기 위한 것이고 두 번째 질의어는 검색어가 포함된 모든 표제어를 찾기 위한 것이다. 이때 두 번째 질의어에 의해서 검색된 결과가 첫 번째 질의어에 의해서 검색된 결과와 중복될 수 있기 때문에 첫 번째 검색 결과에 없는 표제어만 검색 결과에 추가한다. 두 경우 모두 검색어와 정확하게 일치하는 표제어뿐만 아니라 부분적으로 일치하는 표제어까지 보여주는 부분 일치 검색 방법을 사용하고 있다.

3.2 부분 검색

부분 검색은 검색의 범위를 지정하여 검색을 수행하는 것으로 메뉴에서 검색 범위를 선택하는 사전식 검색과 검색식을 사용하여 질의를 수행하는 두 가지 방법으로 나눌 수 있다.

첫 번째 방법은 메뉴에 있는 ‘ㄱ’에서 ‘ㅎ’ 중에서 하나를 지정하여 사전식 검색을 수행할 수 있다. 예를 들어 ‘ㄹ’을 선택한 경우 검색 범위는 ‘마’ 보다 크고 ‘바’ 보다 작은 모든 단어들을 검색한다.

또 다른 검색 방법은 OR 연산과 AND 연산으로 구성된 검색식을 사용하는 것이다. 검색어들을 공백 문자(' ')로 분리하여 나열한 경우에는 OR 연산을 사용한다. 예를 들어 “소리 나무”라고 검색어를 입력하면 “소리”로 시작하는 모든 표제어와 “나무”로 시작하는 모든 표제어를 검색하라는 의미이다. 프로그램 내부의 검색 질의어 작성에서는 검색어에서 공백 문자를 기준으로 각 단어를 분리한 후 각 검색어들을 OR로 연결하여 검색식을 작성한다. 예를 들어 “소리 나무”라고 검색어를 입력한 경우 검색식은 다음과 같이 만들어진다.

“SearchWord like ('소리*') OR SearchWord like ('나무*')”

‘-’ 문자를 사용하여 특정한 범위를 지정하여 검색하는 경우에는 AND 연산을 사용한다. 예를 들어 “나무”에서 시작하여 “나무숲”까지를 검색하기를 원한다면 “나무-나무숲”과 같이 검색어를

입력한다. 이와 같은 검색어를 입력하면 검색식은 다음과 같이 생성된다.

“SearchWord>='나무' AND

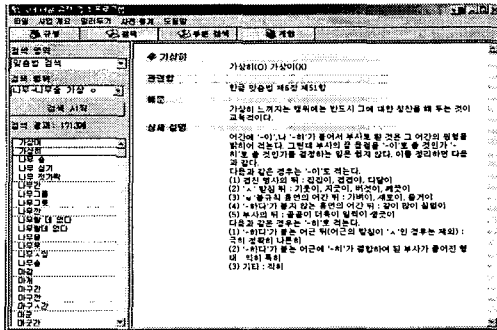
SearchWord<='나무숲' OR SearchWord like ('나무숲*')”

위 검색식에서 AND 연산뿐만 아니라 OR 연산을 사용한 이유는 '나무숲'을 포함하는 모든 표제어까지 함께 검색하기 위한 것이다.

앞에서 서술한 여러 가지 검색 방법을 복합적으로 사용할 수 있다. <그림 8>은 AND 연산, OR 연산, 사전식 검색 방법을 복합적으로 사용한 예이다.

IV. 결론

본 논문에서는 국어 어문 규정 검색 프로그램을 위한 데이터베이스 설계와 검색 프로그램 내부의 질의어 생성에 대해서 알아보았다. 데이터베이스 설계에서는 한글 맞춤법, 표준어 규정, 외래어 표기법, 국어의 로마자 표기법을 위한 규정



<그림 7> 복합 검색의 예

테이블과 색인 테이블을 같은 구조로 설계하였다. 네 개의 규정이 같은 구조를 가져야 통합된 인터페이스를 구성하고 검색 프로그램을 체계적으로 작성할 수 있다. 또한 일반 사용자는 자신이 찾고자 하는 검색어와 그와 관련된 규정이 어느 규정에 속하는지 잘 알지 못하기 때문에 통합된 검색 인터페이스를 제공하는 것이 필요하다.

국어 어문 규정 데이터베이스를 검색하는 방법에는 일반 검색과 부분 검색이 있다. 일반 검색

에서는 입력한 검색어와 일치하거나 검색어가 앞에 나오는 표제어를 먼저 표시하고, 검색어를 포함하는 표제어를 뒤에 표시하는 부분 일치 검색을 수행하였다. 부분 검색을 위한 질의어 생성 시스템에서는 AND 연산과 OR 연산 그리고 사전식 검색 기능을 제공한다. 다양한 검색 방법을 제공함으로써 사용자가 직관적으로 검색어를 입력하여 원하는 표제어들을 쉽게 찾을 수 있도록 하였다.

참고 문헌

- [1] 한글 맞춤법 규정(1988. 1. 19.), 문교부 고시 제88-1호
- [2] 표준어 규정(1988. 1. 19.), 문교부 고시 제 88-2호
- [3] 외래어 표기법(1986. 1. 7.), 문교부 고시 제 85-11호, (1992. 11. 27), 문화부 고시 제1992-31호, (1995. 3. 16.) 문화체육부 고시 제1995-8호
- [4] 국어의 로마자 표기법(2000. 7. 7.), 문화관광부 고시 제2000-8호
- [5] 21세기 세종계획 한민족 언어 정보화 연구보고서(2001), 문화관광부/국립국어연구원
- [6] 세종계획 홈페이지, <http://www.sejong.or.kr/>