

남북한 언어 비교 사전 검색 시스템의 설계 및 구현

김금영, 조시성, 안동언, 정성종, *두길수
전북대학교 컴퓨터공학과, *서남대학교 전기전자멀티미디어공학부
전화 : 063-270-2416

Design and Implementation of South-North Korean Dictionary Retrieval System

Gum-Young Kim, Si-Sung Cho, Dong-Un An, Sung-Jong Chung, Gil-Su Doo
Dept. of Computer Engineering, Chonbuk National University
E-mail : com94@dreamwiz.com

요약

본 논문은 세종계획의 일환으로 개발되었던 '남북한 언어 비교 사전 검색 시스템'을 개선한 것이다.

'남북한 비교 사전 검색 프로그램'은 남북한의 언어 격차를 해소하고자 일반인에게 남북한의 언어정보를 검색할 수 있도록 개발된 시스템이다.

본 논문은 또한 일반인들이 남북한이 이질성을 가진 단어에 관심을 가지고 있는 점에 착안하여 이질성을 가진 단어를 검색할 수 있도록 별도로 '남북한 이질화된 언어 검색 시스템'을 개발하였다.

'남북한 언어 비교 사전 검색 시스템'은 Microsoft Visual C++로 개발되었으며 Database는 Microsoft Access를 사용하였다.

1. 서론

'한민족 언어 정보화'는 국내적으로 모든 국민들에게 국어정보화의 저변 인구가 확대될 수 있도록 유도하여 국민들로 하여금 스스로 올바른 국어 생활을 영위할 수 있는 여건을 마련하여 주고, 국외적으로는 전세계에 국어와 한글의 우수성을 널리 알려서 궁극적으로 문화의 시대인 21세기에 한국문화를 선양할 목적으로 진행되고 있다.

이를 위한 한 과제로서 남북한으로 분단되어 있는 우리 나라의 현 상황에서, 남과 북이 다르게 추진하고 있는 국어정보화를 조화시키고 통일시킬 필요가 있으며, 더 나아가서는 우리말과 우리글을 사용하는 전세계의 해외 동포들에게 국어정보화의 틀을 마련하여 주어야 할 필요가 있다. [1][2][3][4]

그동안 남과 북의 언어 자료를 체계적으로 모아 일반 사용자가 검색하고 비교할 수 있는 실질적인 시스템은 개발되지 않았다. [6][7]

본 논문은 '남북한 언어 비교 사전 검색 시스템'을 개발하여 이러한 필요를 만족시키기 위한 가장 기본적인 작업이다. 국어학자들이 어문 규정 자료를 수집하고 검증하여 데이터베이스를 구축하였으며, 전산학자가 이러한 정보를 효율적으로 사용자에게 전달할 수 있도록 시스템화하였다. [1][2]

2. 남북한 언어 비교 사전 검색 시스템 설계

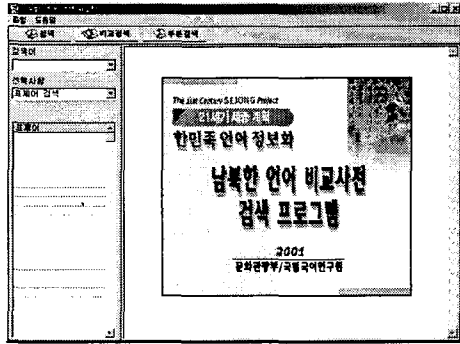
'남북한 언어 비교 사전 검색 시스템'은 Microsoft Visual C++로 개발되었으며 Database는 Microsoft Access를 사용하였다.

'남북한 언어 비교 사전 검색 프로그램'은 남북한 언

어 비교 사전 데이터베이스와 데이터베이스를 검색할 수 있는 검색기로 구성되어 있다. 검색의 유형은 크게 3가지이다. 표제어에 대한 완전일치와 부분일치 검색을 제공함으로써 표기를 정확히 알지 못하는 검색어를 찾을 수 있도록 하였다. 남북한 차이에 의한 검색을 제공하는데 표기, 발음, 문법, 의미별로 남북한의 차이 또는 공통인 것을 찾을 수 있다. 또한, 표제어가 아닌 해설 부분에서 단어를 찾을 수 있는 내용 검색을 제공한다.

2.1 시스템의 구성

<그림 1>에서 볼 수 있듯이 3개의 메뉴탭으로 구성되어 있으며, 각 메뉴의 기능은 다음과 같다.



<그림 3> '남북한 언어 비교 사전 검색 프로그램' 초기 화면

- ① 검색
사용자가 입력한 남한 또는 북한 표제어에 대하여 '표기', '발음', '문법', '의미', '기타' 정보를 제시한다.
- ② 비교검색
사용자가 지정한 조건을 고려하여 검색을 실시한다. 사용자가 원하는 언어 분야의 남과 북의 언어를 제시한다.
- ③ 전체검색
시스템의 데이터 베이스에 등록되어 있는 표제어를 제시한다.

3. 남북한 언어 비교 사전 검색 시스템의 구현

3.1 데이터베이스의 구성

남북한 언어 비교 사전 검색 시스템에서 사용하는 데이터베이스는 [표1]과 같다.

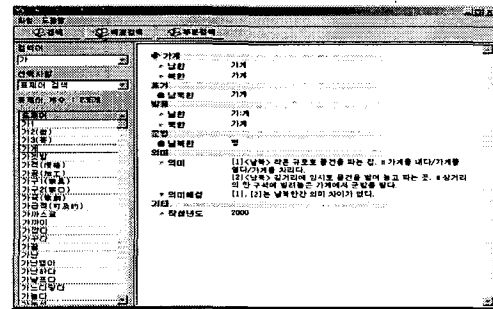
[표 1] '남북한 언어 비교 사전 검색 프로그램' 데이터베이스

| 테이블명 | 테이블 항목 | 항목 설명 |
|-----------|-------------|-----------|
| InputDic | DGram | 문법 설명 |
| | SGram | 남한 문법 |
| | NGram | 북한 문법 |
| | SNGram | 남북한 공통 문법 |
| | DPronoun | 발음 설명 |
| | SPronoun | 남한 발음 |
| | NPronoun | 북한 발음 |
| | SNPronoun | 남북한 공통 발음 |
| | DSyntax | 표기 설명 |
| | SSyntax | 남한 표기 |
| | NSyntax | 북한 표기 |
| | SNSyntax | 남북한 공통 표기 |
| | Head | 대표 주제어 |
| | STitle | 남한 주제어 |
| | NTitle | 북한 주제어 |
| | EditDate | 수정한 연월일 |
| | EditPeople | 수정한 사람 명 |
| | MngDate | 관리 연월일 |
| | MngPeople | 관리자 명 |
| | NumSemantic | 의미의 개수 |
| | Semantics | 의미 저장 |
| | State | 단어의 완성 여부 |
| | Etc | 기타 필드 |
| DSemantic | 의미 설명 | |
| TitleID | 일련 번호 | |

3.2 세부 기능

(1) 검색

검색 메뉴에서는 사용자가 입력한 남한 또는 북한 표제어에 대하여 '표기', '발음', '문법', '의미', '기타' 정보를 보여준다.



<그림 4> 검색 실행 예

최근 검색한 검색어 리스트를 저장하고 있는 '검색어 히스토리' 기능, 남한과 북한의 표제어를 따로 검색할 수 있도록 선택할 수 있다.

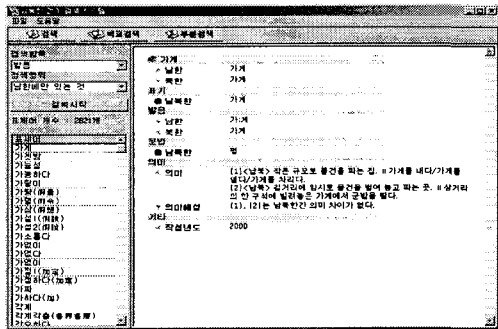
<그림2>는 검색을 실행한 예이다.

(2) 비교검색

비교검색은 선택사항을 지정하여 그 조건을 만족하는 일부 표제어들만 검색할 수 있어, 남북한의 언어 차이를 검색을 통해 쉽게 확인할 수 있다.

상단의 검색항목 선택사항에서 '표기', '발음', '문법', '의미'의 4가지 항목 중에서 하나를 선택하고, 하단의 검색영역 선택사항에서는 '남북한 공통으로 있는 것', '남한과 북한에 있는 것', '남한에만 있는 것', '북한에만 있는 것'의 4가지 항목 중에서 하나를 선택하면, 두 가지 조건을 만족하는 결과를 제시한다.

<그림 3>은 비교검색을 실행한 예이다.

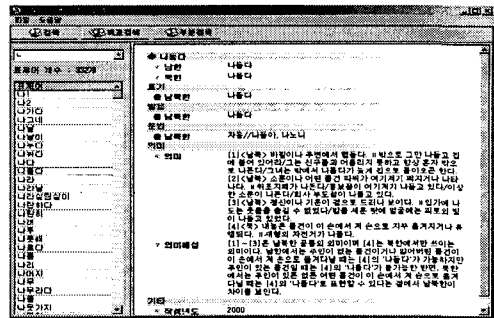


<그림 5> 비교검색 실행 예

(3) 부분검색

부분검색은 '남북한 언어 비교 사전 검색 프로그램'의 데이터베이스에 등록된 표제어를 검색범위에 해당하는 목록을 뽑아 보여준다. 아래의 그림에서 볼 수 있듯이 모든 표제어가 왼쪽 창에 리스트로 나열되고, 사용자가 원하는 표제어를 선택하면, 항목이 활성화되면서 오른쪽 창에 세부 정보가 나타나게 된다.

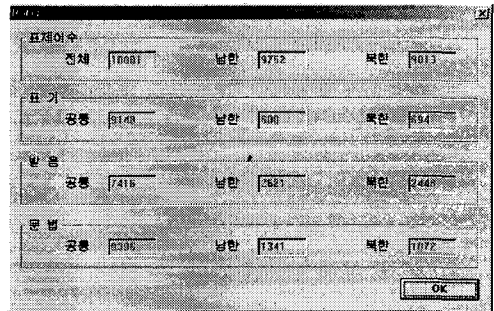
<그림4>는 부분 검색을 실행한 예이다.



<그림 6> 부분검색 실행 예

(4) 통계

통계는 '도움말'에서 선택한다. <그림 5>은 통계 메뉴의 화면을 보여주고 있다.



<그림 7> 통계 실행 예

통계 메뉴를 선택하면, '남북한 언어 비교 사전 검색 프로그램'의 데이터베이스에 등록된 남북한의 표제어 수, 표기 수, 발음 수, 문법 수 등에 대한 통계 정보를 볼 수 있다.

4. 남북한 이질화된 언어 검색 시스템 설계와 구현

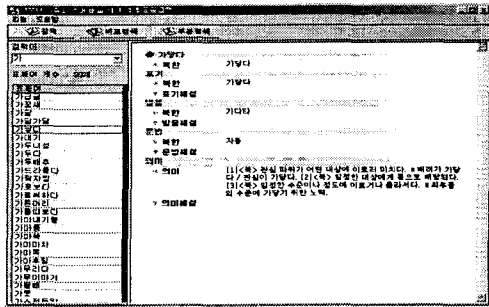
'남북한 이질화된 언어 검색 프로그램'은 Microsoft Visual C++로 개발되었으며 Database는 Microsoft Access를 사용하였다.

'남북한 이질화된 언어 검색 프로그램'은 남북한 이질화된 언어 데이터베이스와 데이터베이스를 검색할 수 있는 검색기로 구성되어 있다. 검색의 유형은 크게 3가지이다. 표제어에 대한 완전일치와 부분일치 검색을 제공함으로써 표기를 정확히 알지 못하는 검색어를 찾을 수 있도록 하였다. 남북한 차이에 의한 검색을 제

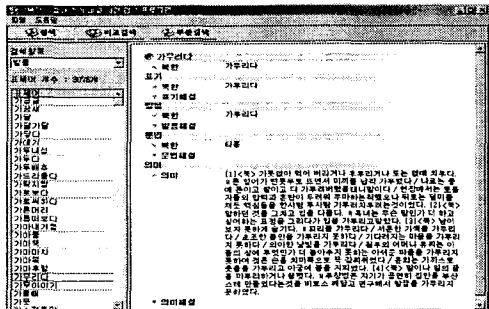
공하는데 표기, 발음, 문법, 의미별로 남북한의 차이 또는 공통인 것을 찾을 수 있다. 또한, 표제어가 아닌 해설 부분에서 단어를 찾을 수 있는 내용 검색을 제공한다.

'남북한 이질화된 언어 검색 시스템'의 구성 및 기능은 '남북한 언어 비교 사전 검색 시스템'과 같으므로, 각 기능별 실행 예만을 제시한다.

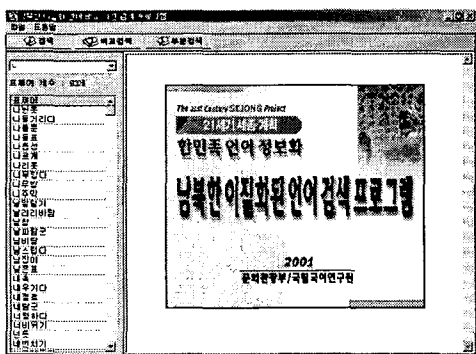
4.1 각 기능별 실행 예



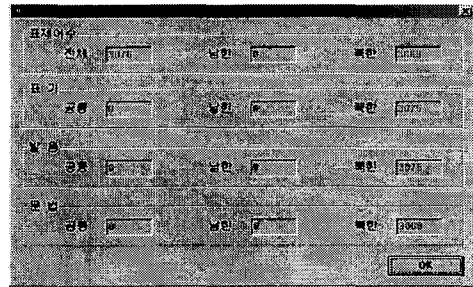
<그림 8> 검색 실행 예



<그림 9> 비교검색 실행 예



<그림 10> 부분검색 실행 예



<그림 11> 통계 실행 예

5. 결론

본 논문에서는 남과 북의 언어 정보를 하나의 시스템으로 표현하고, 일반인의 남북의 언어 정보를 쉽게 이해할 수 있도록 '남북한 언어 비교 사전 검색 시스템'과 '남북한 이질화된 언어 검색 시스템'을 개발하였다.

남북한간 언어 특성에 대한 정보의 체계화에 대한 기초를 마련하고 이를 전산화함으로써 한민족 언어 정보화의 중요한 기초를 마련하였다.

참고문헌

- [1] 문화관광부, 국립국어연구원, 21세기 세종계획 한민족 언어 정보화 연구보고서, 2000
- [2] 문화관광부, 국립국어연구원, 21세기 세종계획 한민족 언어 정보화 연구보고서, 2001
- [3] <http://www.sejong.or.kr/>
- [4] <http://www.korean.go.kr/>
- [5] 강승식, 한국어 정보처리의 문제점 및 방법론 고찰, 제 10회 한글 및 한국어 정보처리 학술대회, pp329-334, 1998
- [6] 국립국어연구원, 제1회 국어정보화 아카데미 자료집, 2000