

# 남북한 언어 비교 사전 검색 시스템의 설계 및 구현

최유경, 황호전, 안동언, 정성중  
전북대학교 컴퓨터공학과  
e-mail : {ykchoi, hjhwang}@calhpl.chonbuk.ac.kr  
{duan, sjchung}@moak.chonbuk.ac.kr

## Design and Implementation of South-North Korean Dictionary Retrieval System

Yoo-Kyung Choi, Ho-Jeon Hwang, Dong-Un An, Sung-Jong Chung  
Dept. of Computer Engineering, Chonbuk National University

### 요 약

본 논문은 남과 북의 언어 정보를 체계적으로 정리하고, 일반인들도 남과 북의 언어 정보의 차이를 쉽게 검색할 수 있도록 '남북한 언어 비교 사전 검색 시스템'을 설계하고 구현한다. '남북한 언어 비교 사전 검색 시스템'은 검색어에 대해서 '표기', '발음', '문법', '의미'의 4가지 항목에서 남북한 언어의 차이점을 알 수 있도록 한다.

'남북한 언어 비교 검색 시스템'은 표제어에 의한 단순한 검색 이외에도 '남한에만 있는 것', '북한에만 있는 것', '남북한 공통인 것', '남북한 차이가 있는 것' 등을 비교하여 검색할 수 있도록 하였다. 또한, 표제어에 대해서 '완전 일치'와 '부분 일치' 검색을 제공하였고 의미 항목에 대해서는 내용 검색이 가능하도록 하였다.

'남북한 언어 비교 사전 검색 시스템'은, 장차 통일이 되었을 때에 남과 북의 언어 생활의 차이로 인한 혼란을 미연에 방지하고 통일된 언어 생활을 할 수 있도록 하는데 기여할 수 있다.

### 1. 서 론

'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]

그 동안 남과 북의 언어 자료를 체계적으로 모아 일반 사용자가 검색하고 비교할 수 있는 실질적인 시스템은 개발되지 않았다.[4][5]

'남북한 언어 비교 사전 검색 시스템'의 개발은 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. 남북한 언어 비교 사전 검색 시스템 설계

남한과 북한의 언어 정보를 하나의 검색 시스템에 포함시켜 사용자의 편의를 도모했으며, 전문 국어학자들이 차후에도 계속 쉽게 데이터베이스 내용을 추가, 삭제, 수정할 수 있도록 시스템을 설계하였다.

본 논문은 문화관광부의 "한민족 언어 정보화" 연구과제의 지원을 받았습니다.

2.1 시스템의 구성

<그림 1> 에서 볼 수 있듯이 7개의 메뉴탭으로 구성되어 있으며, 각 메뉴의 기능은 다음과 같다.

① 검색

사용자가 입력한 표제어에 대하여 ‘표기’, ‘발음’, ‘문법’, ‘의미’, ‘기타’ 정보를 제시한다.

② 비교검색

사용자가 지정한 조건을 고려하여 검색을 실시한다. 사용자가 원하는 언어 분야의 남과 북의 차이를 쉽게 알아볼 수 있다.

③ 전체검색

시스템의 데이터베이스에 등록된 모든 표제어를 제시한다.

④ 통계

시스템의 데이터베이스에 등록되어 있는 표제어 수, 표기 수, 발음 수, 문법 수 등에 대한 통계 정보를 제시한다.

⑤ 편집

전문 연구진들이 시스템의 내부구조를 잘 알지 못 하더라도 데이터베이스의 내용을 쉽게 수정, 추가, 삭제 할 수 있도록 인터페이스를 제공한다.

⑥ 파일

사용자가 지정한 범위의 데이터베이스의 내용을 원하는 파일 형태로 저장한다.

남북한 언어 비교 사전 검색 시스템은 Microsoft Visual Basic 6.0을 사용하여 개발하였고, 데이터베이스는 Microsoft Access 97을 사용하였다.

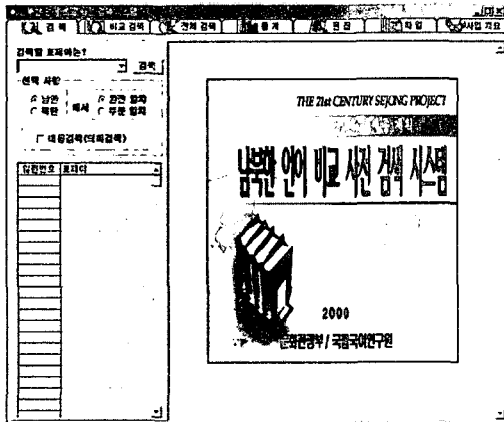
본 시스템은 Windows98 이상 버전의 OS 환경에서 사용이 가능하다.

3.1 데이터베이스의 구성

남북한 언어 비교 사전 검색 시스템에서 사용하는 데이터베이스는 [표 1]과 같다.

[표 1] 남북한 언어 비교 사전 검색 시스템의 데이터베이스

테이블 명	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	수정된 연월일
	EditPeople	수정된 사람 명
	MngDate	관리 연월일
	MngPeople	관리자 명
	NumSemantic	의미의 갯수
	Semantics	의미 저장
	State	단어의 완성 여부
	Etc	기타 필드
	DSemantic	의미 설명
	TitleID	일련 번호



<그림 1> 남북한 언어 비교 사전 검색 시스템의 초기화면

3.2 세부 기능

시스템의 기능은 크게 ‘검색어 검색 기능’과 ‘데이터베이스 편집 기능’, ‘파일 기능’으로 나누어진다.

3.2.1 검색어 검색 기능

‘검색’, ‘비교 검색’, ‘전체 검색’의 메뉴 항목에 해당한다.

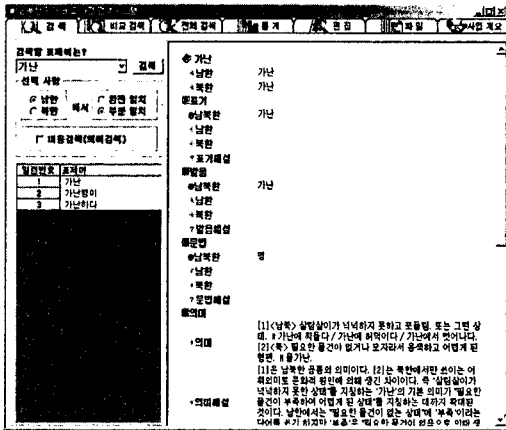
‘검색’ 메뉴에서는 사용자가 입력한 검색어에 대하여 ‘표기’, ‘발음’, ‘문법’, ‘의미’ 등에 관한 세부 정보를 보여준다.

최근 검색한 검색어 리스트를 저장하고 있는 “검색어 히스토리” 기능, 의미해설 필드에서 검색어를

3. 남북한 언어 비교 사전 검색 시스템의 구현

검색하는 “내용 검색” 기능, 검색 옵션으로 사용되는 “완전·부분 일치”, “남·북한 선택” 기능 등을 제공한다.

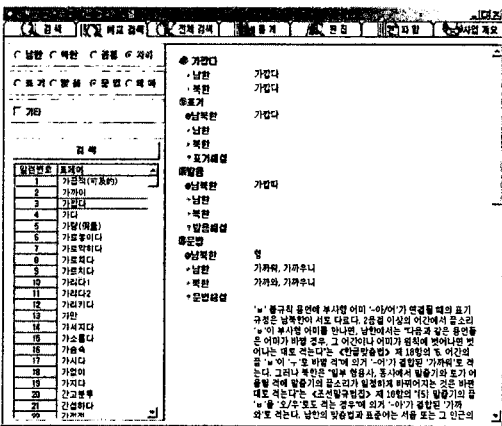
<그림 2>은 검색을 실행한 예이다.



<그림 2> 검색 실행 예

‘비교 검색’ 메뉴에서는 사용자가 지정한 조건으로 검색을 실시한다. 상단의 ‘남한’, ‘북한’, ‘공통’, ‘차이’ 조건항에서 하나를 선택하고, 하단의 ‘표기’, ‘발음’, ‘문법’, ‘의미’의 조건항에서 하나를 선택하면 두 가지 조건을 만족시키는 결과를 제시한다.

<그림 3>은 비교 검색의 예이다.



<그림 3> 비교 검색 실행 예

‘전체 검색’ 메뉴에서는 시스템의 데이터베이스에 등록되어 있는 모든 표제어 목록을 보여주거나, 사용자가 입력한 단어를 포함하는 표제어의 목록을 보

여준다.

목록에서 원하는 표제어를 선택하면 오른쪽 창에서 그에 해당하는 세부정보를 볼 수 있다.

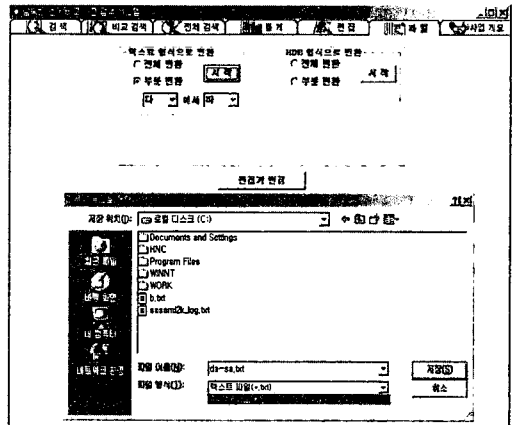
### 3.2.2 데이터베이스 편집 기능

‘편집’ 메뉴에 해당하며 시스템의 데이터베이스에 존재하는 표제어를 수정, 삭제, 추가할 수 있도록 한다. 시스템의 내부 구조를 잘 알지 못하더라도 전문 연구진들이 손쉽게 데이터 작업을 수행할 수 있도록 편리하고 알기 쉬운 인터페이스를 제공한다.

### 3.2.3 파일 기능

‘파일’ 메뉴에서는 시스템의 데이터베이스의 내용을 사용자가 지정하는 파일 형식으로 변환하여 저장한다.

<그림 4>는 파일 메뉴의 실행 예이다.



<그림 4> 파일 실행 예

전체변환은 데이터베이스의 내용 전체를 변환하는 것이고, 부분변환은 사용자가 지정한 범위의 데이터베이스의 내용만을 파일로 저장하는 것이다.

저장될 파일 형식으로는 텍스트 형식과 HDB 형식이 있다.

HDB 형식이란, 국립국어연구원에서 지정한 파일 형태로 데이터베이스의 각 필드별 구분자를 <표 2>의 형태로 지정하고 있다.

### 3.3 기타 기능

“복사하기”, “편집기로”, “덧붙이기”의 추가적인 기능을 사용할 수 있다.

<표 2> HDB 파일 형식의 데이터베이스 구분자

HDB 구분자	내용	MDB 항목
#1	표제어 고유번호	TitleID
#2	전체 표제어	Head
#3	엔트리 번호	Entry
#4	남한 표제어	STitle
#5	북한 표제어	NTitle
#6	남한 표기	Ssyntax
#7	북한 표기	Nsyntax
#8	남북 표기	SNsyntax
#9	표기 해설	DSyntax
#10	남한 발음	SPronoun
#11	북한 발음	Npronoun
#12	남북 발음	SNpronoun
#13	발음 해설	DPronoun
#14	남한 문법	SGram
#15	북한 문법	NGram
#16	남북 문법	SNgram
#17	문법 해설	Dgram
#18	의미 개수	NumSemantic
#19	의미	Ssemantic
#20	의미 해설	Dsemantic
#21	교열자	EditPeople
#22	교열일	EditDate
#23	작성자	MugPeople
#24	작성일	MugDate
#25	기타	Etc
#26	완성/미완성	State
#27	한글 의미	SemanHan

“복사하기”는 표제어 목록에서 원하는 표제어들을 선택한 뒤, 마우스 오른쪽 버튼을 눌러 “복사하기” 기능을 실행하면 클립보드에 선택한 표제어들이 복사된다. 따라서 사용자가 사용하기에 편리한 응용프로그램(메모장, 한글 등)에서 “붙여넣기”를 실행하여 작업을 수행할 수 있다.

“편집기로” 기능도 마찬가지로 원하는 표제어들을 마우스로 선택하여 마우스 오른쪽 버튼의 “편집기로” 메뉴를 선택하면, 기본적으로 지정된 응용프로그램이 열리고, 선택한 표제어가 워드패드에 붙여 넣어진다. 기본적으로는 워드패드로 지정되어 있다.

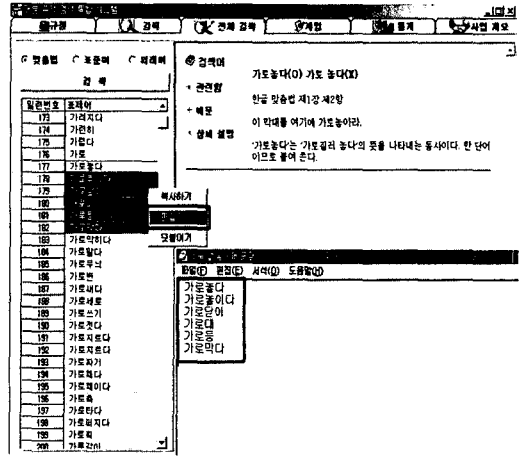
<그림 5>는 “편집기로” 기능의 실행 예이다.

지정되어 있는 편집기를 변경하고자 할 때에는 ‘파일’ 메뉴의 “편집기 변경”을 실행하여 원하는 응용프로그램을 지정하여 주면 된다.

“덧붙이기” 기능은 검색 실행 후 결과로 나온 표제어들 중 원하는 몇몇 표제어의 상세 정보를 하나의 화면으로 연결시켜 한꺼번에 출력하고자 할 때 유용한 기능이다.

이 밖에도 윈도우 웹 브라우저에서 기본적으로 제공되는 “뒤로”, “앞으로”, “인쇄”, “모두 선택”등의

기능도 제공되어진다.



<그림 5> “편집기로” 실행 예

#### 4. 결 론

본 논문에서는 남과 북의 언어 정보를 하나의 시스템으로 표현하고, 일반인도 남북의 언어 정보를 쉽게 검색·비교할 수 있도록 편리한 인터페이스를 제공하는 ‘남북한 언어 비교 사전 검색 시스템’을 개발하였다.

언어 정보들은 시간이 흐름에 따라 변화해 나가기 때문에 시스템에서도 계속적인 데이터의 관리가 필요하다. 전문국어학자들도 쉽게 데이터베이스를 편집할 수 있는 기능을 제공하여 이러한 필요성에 부합하도록 하였다.

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] <http://www.sejong.or.kr/>
- [3] <http://www.korean.go.kr/>
- [4] 강승식, 한국어 정보처리의 문제점 및 방법론 고찰, 제10회 한글 및 한국어 정보처리 학술대회, pp.329-334, 1998
- [5] 국립국어연구원, 제1회 국어정보화 아카데미 자료집, 2000