

# 북한 방언 검색 프로그램 개발

두길수<sup>○</sup>, 황호전<sup>○</sup>, 김법관<sup>○</sup>, 안동연<sup>○</sup>, 정성중<sup>○</sup>, 이신원<sup>○</sup>  
\*서남대학교 전기전자공학부, \*\*진북대학교 전자정보공학부, \*\*\*전북과학대학 컴퓨터정보계열  
dgs@tiger.seonam.ac.kr, {hjhwang, kyun}@duan.chonbuk.ac.kr,  
{duan, sjchung}@moak.chonbuk.ac.kr, swlee9237@hanmail.net

## A Development of Retrieval Program for North Korea Dialect

\*Gil-Su Doo<sup>○</sup>, \*\*Ho-Jeon Hwang, \*\*Beob-Kyun Kim, \*\*Dong-Un An, \*\*Sung-Jong Chung, \*\*\*Shin-Won Lee  
\*Faculty of Electric and Electronic Engineering, Seonam University,  
\*\*Department of Computer Engineering, Chonbuk National University,  
\*\*\*Department of Computer Information, Jeonbuk Science College

### 요약

이 연구는 남한의 방언과 북한의 방언 및 세계의 여러 나라에서 쓰이고 있는 한국어 방언을 컴퓨터로 검색할 수 있는 도구를 개발하여, 21세기 지식 기반 사회에서 우리 민족 구성원들이 상호 의사소통을 신속 정확하게 할 수 있도록 하고 또 통일 이후에는 남북한 주민들이 의사소통을 원만하게 할 수 있도록 한다는 데에 목적을 두고 있다. 이를 통해서 민족 구성원간의 유대를 다지고 나아가 민족의 정체성을 확립하며 민족문화의 창달에 기여할 수 있을 것이다. 본 연구는 한국 방언 검색 프로그램 개발 과제의 일부분으로서 2001년에 남한 방언 검색 시스템을 개발하고, 2002년에 북한 방언 검색 시스템을 개발하였다. 본 논문에서는 북한 방언 검색 시스템의 개발 개요와 방언 검색 방식에 대해서 다룬다.

### 1. 서론

북한 방언 검색 프로그램은 남한의 방언과 북한의 방언 및 해외 동포가 살고 있는 지역에서 사용되고 있는 한국어 방언을 컴퓨터로 검색할 수 있도록 하여, 21세기 지식기반 사회에서 국민들의 의사소통을 신속·정확하게 하며 또한 통일 이후에도 우리 민족 구성원들간의 의사소통을 원만하게 함으로써 민족의 정보 공유를 유도하여 민족문화 창달에 기여함을 그 목적으로 한다[1][2][3].

한국 방언 검색 프로그램은 21세기 세중계획 프로젝트의 한 민족 언어 정보화 분과에서 개발하였다. 2000년에는 프로그램 개발을 위한 기초 조사와 연구를 시행하였으며 2001년에는 남한의 방언을 검색할 수 있는 시스템을 개발하였다. 2002년에는 북한의 방언을 검색할 수 있는 시스템이 추가되었으며, 2003년에는 중국을 중심으로 해외의 우리 동포들이 사용하는 방언을 검색할 수 있는 시스템을 개발함으로써 한국 방언 검색 프로그램 개발을 완료할 것이다.

방언은 의사 소통의 장애를 가져오며, 의사 소통의 장애는 우리 나라 문화 발전에 장애를 가져오게 된다. 따라서 각 방언 사용자들의 원활한 의사 소통을 위해서는 각 방언들간의 차이와 대조 관계를 밝혀두어야 할 필요가 있다. 의사 소통의 원활성은 곧 정보 전달의 정확성과 신속성을 의미하게 되며, 더 나아가 문화 전달의 정확성과 신속성을 가져오며, 그것은 곧 문화의 발달로 이어진다[1][2].

방언 관련 연구로서는 SIL International의 음성 부호에 의한 방언 지도 도시 방식이 있으며, 여기에서는 각 지역별 방언의 출현 빈도를 원의 크기로 보여주고 있다. 또한 방언의 발음을 위해 국제 음성 부호(IPA)를 사용한다[18].

한국 방언 검색 프로그램의 데이터베이스를 구축하기 위해서 지금까지 조사 연구된 50여권의 방언 자료집들 중에서 전자 파일로 수집한 방언 자료를 일정한 형태로 가공하여 정리하였다. 2002년도에 개발된 북한 방언 검색 프로그램에 수록

된 표준어 어휘 항목은 12,266개이며, 방언 어휘 항목은 38,027개이다. 또한, 방언 지도에 표시된 군 단위의 세부 지역은 86개이다.

본 논문에서는 북한 방언 검색 프로그램의 구성과 기능을 설명한다.

### 2. 북한 방언 검색 프로그램의 데이터베이스 구축

#### 2.1 북한 방언 자료의 수집과 정리

북한 방언 검색 시스템에 등재된 표준어 목록은 모두 12,266개이다. 이 프로그램에 수록된 방언 자료는 주로 아래 방언 자료집에서 가려 뽑은 다음, 연구진들이 정리하고 수정한 것이다.

- ㉠ 《한글》誌에 수록된 북한 지역의 방언
- ㉡ 《말과글》誌에 수록된 북한 지역의 방언
- ㉢ 남한에서 간행된 북한 각 지역의 도지, 군지, 면지, 읍지, 도지(島誌)에 수록된 북한 지역의 방언
- ㉣ 단행본에 수록된 북한 지역의 방언
- ㉤ 사전류

각 지역 방언 책임자들이 수집된 방언 자료들을 검토하여, 오류를 수정하고, 내용을 보완하였다. 또한, 이 자료들을 프로그램에서 검색할 수 있고, 데이터베이스로 쉽게 변환할 수 있도록 기본 입력 구조를 만들어서 정리하였다. 기본 입력 구조는 표제어, 품사, 방언 등으로 구성되어 있으며, 필요한 경우에는 뜻풀이와 예문이 포함되어 있다.

#### 2.2 북한 방언 데이터베이스

한국 방언 검색 프로그램의 데이터베이스는 3개의 테이블로

구성되어 있다. 이 테이블들은 방언의 자료를 저장하는 주 테이블인 StandardDB와 각 표준어에 대한 방언형들의 자료를 저장하는 LocalDB와 방언의 지역 코드를 저장하는 MapCode 이다.

StandardDB 테이블의 구조는 [표 1]과 같다. 남한 방언의 경우는 표준어에 대한 설명이 필요없으나, 북한 방언의 경우 표준어 자체가 남한에 없는 경우가 많아 북한 표준어에 대한 설명항(Explain)이 필요하다. <그림 1> 은 StandardDB 의 일부 예를 보여주고 있다.

[표 1] StandardDB 테이블의 구조

필드명	설명
ID	표준어 코드 번호, 원시 자료를 입력할 때 자동으로 생성한다.
Writer	작성자
Date	작성일
SWord	검색어, 표준어에서 공백 문자를 제거하여 검색에서 공백으로 인한 검색 실패를 없앤다.
Standard	표준어
Locals	표준어에 대한 방언형들의 목록
Pos	품사
Explain	표준어의 설명
Mean	표준어의 의미
Title	한자어나 고어가 포함된 표준어, Standard 에는 한자어나 고어가 포함되어 있지 않다.

ID	Standard	Local	Pos	Mean	Explain
1	가	가	조사	가(가)	
2	가	가	조사	가(가)	
3	가	가	조사	가(가)	
4	가	가	조사	가(가)	
5	가	가	조사	가(가)	
6	가	가	조사	가(가)	
7	가	가	조사	가(가)	
8	가	가	조사	가(가)	
9	가	가	조사	가(가)	
10	가	가	조사	가(가)	
11	가	가	조사	가(가)	

<그림 1> StandardDB Sample

LocalDB 는 각 방언형에 대한 세부적인 정보를 포함하고 있다. [표 2]는 LocalDB 테이블의 구조를 보여주고 있다. 또한 <그림 2> 는 LocalDB 의 일부 내용을 보여준다.

[표 2] LocalDB 테이블의 구조

필드명	설명
ID	방언형 코드 번호, 원시 자료를 입력할 때 자동으로 생성한다.
Writer	작성자
Date	작성일
SWord	검색어, 방언형에서 공백 문자를 제거하여 검색에서 공백으로 인한 검색 실패를 없앤다.
Standard	방언형에 대한 표준어의 StandardDB 테이블의 ID를 지정한다.
Origin	원문과 방언형에 대한 참고 자료를 저장한다.
Map	방언형의 지역 분포
Example	방언형의 사용 예
Mean	방언형의 의미

ID	Standard	Local	Pos	Mean	Explain
1	가	가	조사	가(가)	
2	가	가	조사	가(가)	
3	가	가	조사	가(가)	
4	가	가	조사	가(가)	
5	가	가	조사	가(가)	
6	가	가	조사	가(가)	
7	가	가	조사	가(가)	
8	가	가	조사	가(가)	
9	가	가	조사	가(가)	
10	가	가	조사	가(가)	
11	가	가	조사	가(가)	

<그림 2> LocalDB Sample

MapCode 테이블은 방언형들의 자료를 저장하는 LocalDB 테이블의 Map 필드에 나타나는 지역들에 대한 Code를 저장하고 있다. MapCode 테이블의 구조는 [표 3]과 같다.

[표 3] MapCode 테이블의 구조

필드명	설명
ID	지역 코드 일련 번호
Local	지역명
Code	코드 번호

MapCode 테이블의 코드 번호는 십진수 4자리로 구성되어 있으며 각 자리는 특별한 용도가 있다. 첫 번째 자리는 남한, 북한, 중국, 일본 등을 구분한다. 남한은 0, 북한은 1, 중국은 2이다. 방언을 사용하고 있는 기타 해외 동포가 살고 있는 지역의 코드를 추가할 수 있다. 두 번째 자리는 특별시·광역시·도를 구분한다. 특별시와 광역시의 코드는 0이며 구체적인 광역시의 코드는 [표 4]와 같이 세 번째와 네 번째 자리를 이용한다. 세 번째와 네 번째 자리는 각 도에 소속된 시·군을 표시한다. 도 전체를 나타낼 때는 00을 사용하며, 전국을 나타내는 지역 코드는 0000이다.

[표 4]는 지역 코드의 일부분을 보여주고 있다.

[표 4] 지역 코드 (일부분)

도	코드	시·군	코드
북한전역	1000	곡산	1301
경기	1100	수안	1302
강원	1200	서흥	1303
황해	1300	맹신	1407
평남	1400	평양	1415
합북	1700	무산	1707

### 3. 북한 방언 검색 프로그램

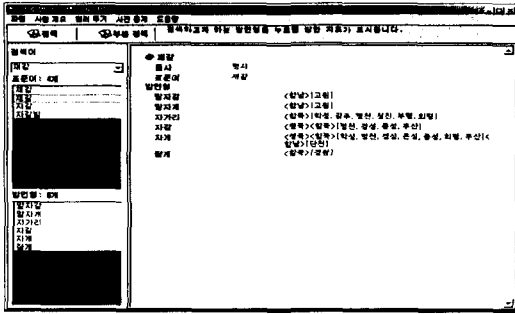
#### 3.1 북한 방언 검색 프로그램의 실행

북한 방언 검색 프로그램은 Microsoft Visual C++로 작성되었으며, 데이터베이스는 Microsoft Access를 사용하였다.

한국 방언 검색 프로그램은 검색창에 표준어 혹은 방언형의 검색어를 입력하면, 입력된 문자열이 포함된 표준어를 보여준다. 입력한 검색어가 방언형인 경우에는 그 방언형에 해당하는 표준어를 보여준다. 검색된 표준어를 선택하면 표준어에 대한 정보를 보여주며 방언형들을 나열하고, 해당 방언형의 사용 지역과 예문을 HTML 문서로 보여준다. <그림 3> 은 북한 방언 검색 프로그램을 실행한 예이다.

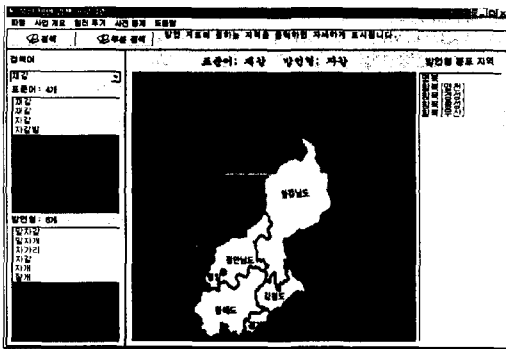
부분검색 기능을 이용하면 검색범위를 지정하여 전체 표제어의 일부를 검색할 수 있다. 선택사항으로 주어지는 ‘\_’에서 ‘\*’까지 중의 하나를 선택할 수도 있고 검색 범위를 입력할

수도 있다.



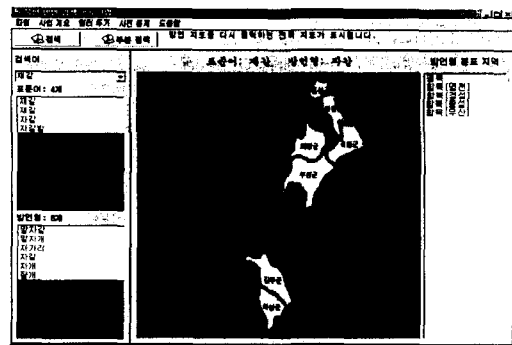
<그림 3> 북한 방언 검색 프로그램 실행 화면

### 3.2 북한 방언 검색 프로그램의 지도 도시



<그림 4> 방언 지도 화면

<그림 3>에서 방언명 중 하나를 선택하면 해당 방언이 분포하는 지역이 지도에 표시된다. 방언 지도는 먼저 전국 지도에서 방언형의 도별 분포를 표시하게 되어 있다. <그림 4>는 “제주”에 대한 방언의 도별 분포를 표시한 예이다.



<그림 5> 세부 방언 지도

방언의 시·군별 분포를 보고자 하는 경우 해당하는 도를 클릭하면 세부적인 내용을 볼 수 있다. <그림 5>은 “제주”에

대한 함경북도 내에서의 분포를 표시한 것이다.

## 4. 결론

본 논문에서는 북한 방언 검색 프로그램의 기능과 데이터베이스 구축과 실행 결과에 대해서 설명하였다. 북한 방언의 지역 분포를 그림으로 확인할 수 있도록 방언 지도를 도시할 수 있도록 프로그램을 작성하였으며, 방언 지도의 행정 구역상의 표시를 위해 지역 코드를 사용하였다. 또한 지역 코드는 현재의 행정 구역 뿐만 아니라 옛 지명이나 읍·면 지역명을 사용할 수 있도록 확장 가능하게 설계했으며, 북한 및 중국과 일본 등에 거주하는 해외 동포가 사용하는 방언의 사용 지역도 표시할 수 있도록 설계하였다. 방언이 분포하는 지역의 확인을 위해 각 지역 코드에 해당하는 마스크 이미지를 사용하였으며, 마스크 이미지를 적용한 컬러링 알고리즘을 구현하였다. 위와 같은 프로그램을 통하여 한국 방언의 분포를 쉽게 확인할 수 있으며, 방언 관련 연구에 기여할 것이다.

## 참고문헌

- [1] 이태영 외(2002), 21세기 세종계획 한민족 언어 정보화 연구보고서, 문화관광부/국립국어연구원
- [2] 이태영 외(2001), 21세기 세종계획 한민족 언어 정보화 연구보고서, 문화관광부/국립국어연구원
- [3] 세종계획 홈페이지, <http://www.sejong.or.kr/>
- [4] 광충구(1995), 《강원도 북부지역의 언어 분화》, 《동대논총》 25.
- [5] 김영환(1982), 《조선어방언학》, 평양: 김일성종합대학출판사, pp.202-272
- [6] 정용호(1988), 《함경남도 방언연구》, 평양: 교육도서출판사, pp.337-380.
- [7] 한영순(1967), 《조선어방언학》, 평양: 김일성종합대학출판사, pp.305-335.
- [8] 한영순(1974), 《조선어방언학》, 평양: 김일성종합대학출판사, pp.416-447.
- [9] 황대화(1986), 《동해안방언연구-함북, 함남, 강원도의 일부 방언을 중심으로》, 평양: 김일성종합대학출판사, pp.161-185
- [10] 김영배(1997), 《平安方言研究(資料篇)》, 태학사.
- [11] 김이협(1981), 《平北方言辭典》, 한국정신문화연구원.
- [12] 김태관(1986), 《咸北方言辭典》, 경기대출판부.
- [13] 국어대사전(금성사)에 등재되어 있는 방언 관련 자료
- [14] 김병제(1980), 방언사전, 과학백과사전출판사
- [15] 이동광, 안동연, 정성중, 김호영, 두길수(2002), “한국 방언 검색 프로그램 개발”, 대한전자공학회 하계 학술발표 논문집
- [16] 김금영, 조시성, 안동연, 정성중, 두길수(2002), “남북한 언어 비교 사전 검색 프로그램 개발”, 대한전자공학회 하계 학술발표 논문집
- [17] 오형진, 정성중, 안동연, 이신원, 두길수(2002), “국어 어문 규정 검색 프로그램 개발”, 대한전자공학회 하계 학술발표 논문집
- [18] <http://www.sil.org/>