

한국어 문형 패턴 조사기의 설계 및 구현

송유석, 이상곤, 이인홍
전주대학교 컴퓨터공학과

e-mail : bobbobgo@naver.com, {samuel, ihlee}@jj.ac.kr

Design and Implementation of Frame Pattern Analyzer in Korean

Yusuck Song, Samuel Sangkon Lee, and In-Hong Lee
Dept. of Computer Science & Engineering, Jeonju University

요 약

본 논문에서는 한국어에서 출현하는 일반적인 형태의 문장 패턴을 조사하여 제2 외국어로서 한국어를 배우는 외국인들에게 우선적으로 가르쳐야 할 한국어의 문장 패턴을 검색하는 프로그램을 개발하였다. 이를 위해 지난 10년 동안 조사 구축된 21세기 세종 계획의 결과물에 출현하는 한국어에 적합한 문장 패턴을 조사하는 프로그램을 설계하였다.

1. 서론

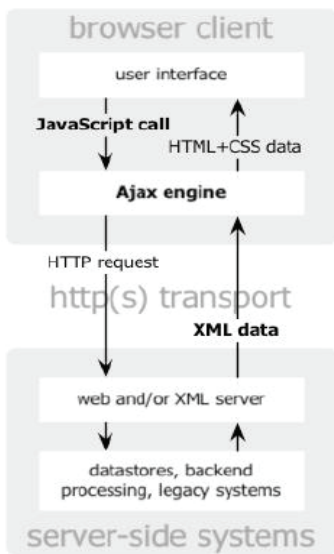
21C 세종 결과물의 말뭉치를 보면 국어정보화 측면에서 귀중한 자료들이 많이 구축되어 있다. 이 말뭉치를 살펴보면 개별 형태소들이 어떠한 양태로 출현하는지 파악할 수 있도록 도와주는 조사 도구가 필요함을 알 수 있다. 말뭉치에서 출현하는 다수의 태그 정보에서 조사자가 원하는 내용의 태그 조합을 입력하면 입력 질의어에 적합한 형태를 가진 어절이 존재하는지를 판별하여 잘못된 태그 열이 검색되면 수정하여 무결성을 보장하고, 한국어에서 다수로 출현하는 문형 패턴을 조사하는 연구에 활용하고자 한다.

한 형태의 페이지가 아니라 Web 2.0 개념을 반영하여 웹이 곧 플랫폼이라는 이론적 배경하에 제작하였다. 주요한 요소 기술로는 Ajax의 개발 기법을 통해 클라이언트와 서버의 역할을 분리하였는데, 서버의 역할은 단순한 데이터의 조회이며, 반면에 클라이언트의 역할은 서버로부터 받은 데이터를 처리하는 단순한 형태에서 분리하여 페이지의 재실행이 아닌 페이지의 재구성으로 No-Splash 페이지를 구성하여 빠른 반응 속도와 동적 페이지를 가진 웹 플랫폼으로 제작하였다.

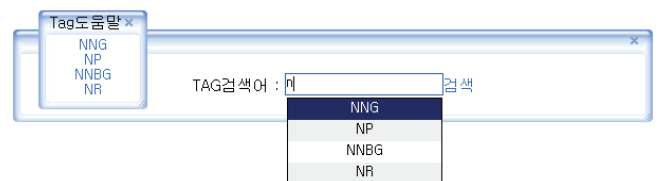
2. 문형 패턴의 검색

2.1 화면 구성

본 논문에서 개발된 프로그램은 한국어에서 나타나는 문형을 조사 검색하여 많이 출현하는 문형을 찾아내 외국인들에게 한국어에 가장 적합한 문형을 찾아내는 프로그램이다. 메인 화면은 위의 (그림 2)와 같이 브라우저 상단에 위치하며, 원하는 형태의 태그 명령어를 입력하여 조회할 수 있다. 문형의 각 레이어를 자유롭게 움직일 수 있어 문형 패턴을 검색하고자 하는 검색자가 작업하기 편한 곳으로 이동시켜 두고 작업할 수 있다.



(그림 1) 전체 시스템의 개요

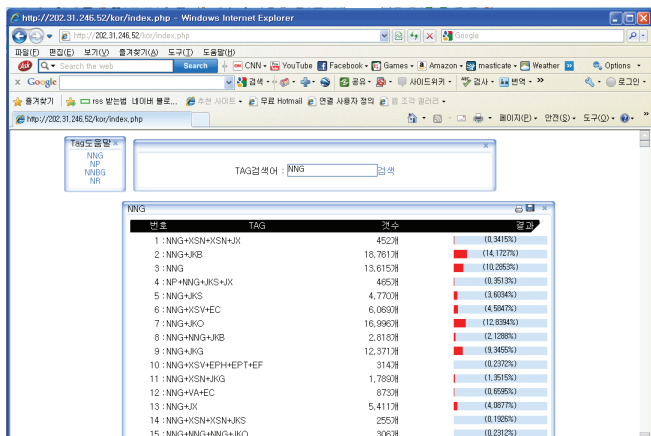


(그림 2) 부분 검색어의 자동 완성 기능

본 논문에서 소개된 프로그램은 (그림 1)에서와 같이 단순

2.2 검색 기능

본 논문에서 주요 기능은 동일하거나 혹은 유사한 태그열을 검색하는 기능이다. 태그열 중 부분적인 문자열을 입력하면 그 입력된 글자를 기준으로 자동완성 기능을 제공하도록 설계하였다. 각 태그의 설명과 추천하는 단어를 선택할 수 있어 사용자를 유도할 수 있다. 또한 화살표 키를 이용하여 예를 들어 Up(↑)과 Down(↓) 그리고 엔터키를 통해 원하는 단어의 선택 입력이 가능하다. 자동 완성 기능을 잘 이용하면 사용자가 특정 태그가 기억이 나지 않을 경우에도 태그 중 한 글자만 입력하여도 입력된 태그를 포함하는 전체 태그열을 검색할 수 있다. (그림 2)에서 영문자 'n'을 입력하면 일반 명사(NNG), 대명사(NP), 의존 명사(NNNG), 수사(NR) 등 네 가지 태그 정보를 모두 보여 주고 있다.



(그림 3) 검색의 결과 화면

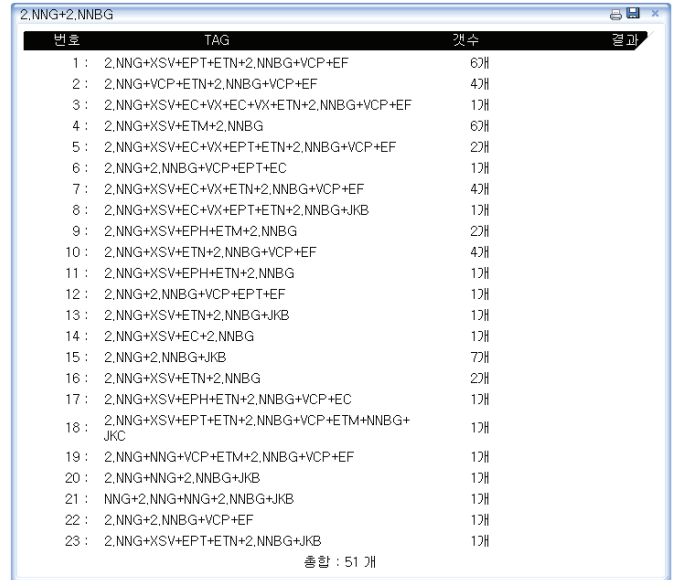
전체 태그의 구성표는 본 논문의 부록에 제시하였다. 태그 검색을 하면 (그림 3)과 같이 출력되는데, 예를 들어,

- 조직사업으로 : 조직사업/NNG + 으로/JKB (한 어절에서 최소 한 번이라도 포함)의 경우,
- 필승불패의 것으로 : 필승불패/NNG + 의/JKG + 것 /NNBG + 으로/JKB)의 어절에서 /NNG가 앞에 출현하고 뒤이어, /NNBG가 뒤따라 출현하는 태그열의 경우를 검색하고자 하는 경우,

“NNG + NNBG”를 검색하면 이와 같은 문형 패턴을 추출할 수 있다. 본 논문에서 제공하는 프로그램의 가장 큰 장점은 일반 명사가 두 글자이고, 동시에 의존 명사가 두 글자인 어절을 검색하고자 하는 경우의 검색이 가능하다는 점이다. 이와 같이 글자의 특정 길이를 지정한 검색이 가능하여 한국어에서 잘 나타나는 문장 패턴을 검색할 수 있다. 이 때 지정 방법은 다음과 같다. 글자수(숫자로).(구분 기호)태그명 과 같이 지정한다. 아래의 예와 같은 검색식을 입력한 경우, “휴식/NNG + 때문/NNBG + 예/JKB”이 검색된다.

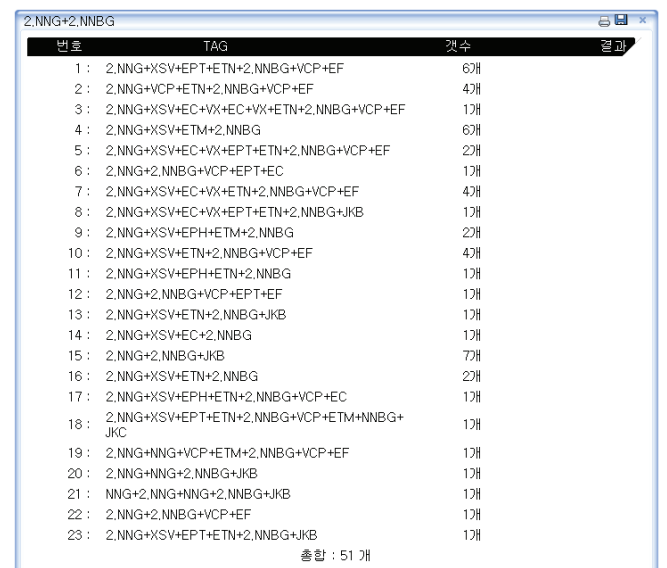
- 예 : 2.NNG + 2.NNBG

모든 검색어는 원하는 태그 형태로 조합하거나, 글자의 개수를 지정하고, 사용자가 원하는 형태로 자유로운 검색이 가능하다.



(그림 4) 제목 표시줄에 표시된 검색 창

위의 (그림 4)에서와 같이 왼쪽 위에 검색 제목이 출력되고, 현재 검색어를 인쇄하거나, 검색 결과를 엑셀 파일로 추출하여 저장할 수 있다. 그 메뉴의 하위에 (그림 5)와 같은 하위 메뉴들을 준비하였는데, 번호순, 태그의 이름순, 다수 출현 개수 순 등의 세 가지 순서로 정렬이 가능하도록 하였다. 이 그림에서와 같이 표시한 부분을 마우스 클릭을 하면 각 표시에 따라 정렬을 하게 된다. 선택 시 오름차순과 내림차순을 번갈아 가면서 소팅한다.



(그림 5) 하위 메뉴

검색의 결과는 DB에 저장되어 있는 모든 데이터를 비교

해 나오는 수치이며, 태그 문자열의 생성 양태를 파악할 수 있다.

번호	TAG	갯수	결과
1:	2.NNG+XSV+EPT+ETN+2.NNBG+VCP+EF	67개	
2:	2.NNG+VCP+ETN+2.NNBG+VCP+EF	47개	
3:	2.NNG+XSV+EC+VX+EC+VX+ETN+2.NNBG+VCP+EF	17개	
4:	2.NNG+XSV+ETM+2.NNBG	67개	
5:	2.NNG+XSV+EC+VX+EPT+ETN+2.NNBG+VCP+EF	27개	
6:	2.NNG+2.NNBG+VCP+EPT+EC	17개	
7:	2.NNG+XSV+EC+VX+ETN+2.NNBG+VCP+EF	47개	
8:	2.NNG+XSV+EC+VX+EPT+ETN+2.NNBG+JKB	17개	
9:	2.NNG+XSV+EPH+ETM+2.NNBG	27개	
10:	2.NNG+XSV+ETN+2.NNBG+VCP+EF	47개	
11:	2.NNG+XSV+EPH+ETN+2.NNBG	17개	
12:	2.NNG+2.NNBG+VCP+EPT+EF	17개	
13:	2.NNG+XSV+ETN+2.NNBG+JKB	17개	
14:	2.NNG+XSV+EC+2.NNBG	17개	
15:	2.NNG+2.NNBG+JKB	77개	
16:	2.NNG+XSV+ETN+2.NNBG	27개	
17:	2.NNG+XSV+EPH+ETN+2.NNBG+VCP+EC	17개	
18:	2.NNG+XSV+EPT+ETN+2.NNBG+VCP+ETM+NNBG+JKB	17개	
19:	2.NNG+NNG+VCP+ETM+2.NNBG+VCP+EF	17개	
20:	2.NNG+NNG+2.NNBG+JKB	17개	
21:	NNG+2.NNG+NNG+2.NNBG+JKB	17개	
22:	2.NNG+2.NNBG+VCP+EF	17개	
23:	2.NNG+XSV+EPT+ETN+2.NNBG+JKB	17개	

총합 : 51 개

(그림 6) 태그 문자열의 양태

위의 (그림 6)에서 4번의 레코드(파란선)를 선택하면, 형태소 태그에 정확히 일치하는 데이터를 열람하여 그와 동일한 태그 집합을 확인할 수 있도록 프로그램을 설계하였다.

번호	TAG	갯수	결과
1:	확약/NNG+하/XSV+L/ETM+대/NNBG	1개	
2:	매견/NNG+되/XSV+L/ETM+대/NNBG	3개	
3:	생산/NNG+하/XSV+는/ETM+족/NNBG	2개	

총합 : 6 개

(그림 7) 열람 후 모습

번호	TAG	갯수	결과
1:	확약/NNG+하/XSV+L/ETM+대/NNBG	1개	
2:	매견/NNG+되/XSV+L/ETM+대/NNBG	3개	
3:	생산/NNG+하/XSV+는/ETM+족/NNBG	2개	

총합 : 6 개

(그림 8) 수정 가능 기능

컴퓨터로 형태소 분석을 수행하면 잘못된 태그가 부착되거나 부착된 태그를 수정해야 하는 경우가 발생하는데 이 때 (그림 8)과 같이 본 프로그램을 이용하면 쉽게 수정할 수 있다. 개수에 따라서 모든 형태소의 수정이 가능하다.

번호	TAG	갯수	결과
1:	특권/NNG+출/XSN+를/XSN+만/JX	17개	
1:	집/NNG+은/JX	10개	
1:	활기/NNG+를/XSA+L/ETM	17개	
2:	환희/NNG+를/XSA+은/ETM	37개	
3:	희락/NNG+하/XSA+L/ETM	37개	
4:	허무/NNG+하/XSA+L/ETM	17개	
5:	황기/NNG+를/XSA+은/ETM	17개	
6:	행복/NNG+하/XSA+은/ETM	17개	

(그림 9) 창 겹치기 기능

사용자가 형태소를 분석을 하다보면 불가피 하게 많은 창을 띄워놓고 작업을 해야 하는 경우가 있는데, 브라우저를 여러 개 실행시키지 않고 레이어를 여러 개 중첩하여 띄움으로써 (그림 9)와 같이 사용자에게 편의성을 제공한다.

번호	TAG	갯수	결과
1:	집/NNG+만/JX+을/JKO	17개	
2:	허위/NNG+만/JX+을/JKO	17개	
3:	행복/NNG+만/JX+을/JKO	17개	
4:	한길/NNG+만/JX+을/JKO	17개	
5:	자신/NNG+만/JX+을/JKO	17개	
6:	영광/NNG+만/JX+을/JKO	17개	
7:	유리/NNG+만/JX+을/JKO	57개	
8:	수결/NNG+만/JX+을/JKO	137개	
9:	비관/NNG+만/JX+을/JKO	17개	
10:	홍콩/NNG+만/JX+을/JKO	17개	
11:	목욕/NNG+만/JX+을/JKO	17개	
12:	배경/NNG+만/JX+을/JKO	17개	
13:	발전/NNG+만/JX+을/JKO	17개	
14:	만족/NNG+만/JX+을/JKO	17개	
15:	위선/NNG+만/JX+을/JKO	27개	
16:	왕성/NNG+만/JX+을/JKO	17개	
17:	담/NNG+만/JX+을/JKO	57개	
18:	길/NNG+만/JX+을/JKO	17개	
19:	결과/NNG+만/JX+을/JKO	17개	

총합 : 40 개

(그림 10) 인쇄 기능과 엑셀로 문서의 저장 기능

관련 자료의 출력이나 다른 형태로의 문서 저장을 돕기 위해서 현재 열람중인 레이어의 전체 내용을 엑셀 문서로 저장하거나 인쇄 할 수 있는 기능을 두어 작업자들이 뒷사람들에게 보고하는 문서를 자동 생성하는 기능을 제공한다.

3. 결론

본 논문에서 개발한 프로그램은 DB로 저장되어 통합 관리되므로 데이터들끼리의 연산에 의한 유연성이 있으며, 데이터의 무결성을 보장할 수 있는 이점이 있다. 웹 기반 소프트웨어이기 때문에 여러 사람이 동시에 검색 작업이 가능하며, 데이터의 총괄 집계를 효율적으로 하도록 설계하였다.

참고 문헌

[1] "Ajax(Asynchronous JavaScript and XML)"
<http://www.oracle.com/technology/global/kr/pub/articles/schalk-ajax.html>

[2] "Extensible Markup Language"
<http://www.w3.org/XML>

[3] "Analysis of the Performance Evaluation Researches and Implementation of the Website based on Ajax - Intak Hwang"

[4] 안동연, 김재훈, 남영준, 박혁로, 이상근 공역, "최신 정보검색론", pp. 1-514, 교보문고, 2010.

부록

다음의 <표 1>에 본 시스템에서 사용된 40개의 태그 셀을 제시하였다.

<표 1> KFPA에서 제공하는 태그들

번호	태그 명	품사	번호	태그 명	품사
1	EC	연결 어미	21	NP	대명사
2	EF	종결 어미	22	NR	수사
3	EP	선어말 어미	23	SE	출입표
4	ETM	관형형 전성어미	24	SF	마침표, 물음표, 느낌표
5	ETN	명사형 전성어미	25	SH	한자
6	IO	감탄사	26	SL	외국어
7	JO	접속 조사	27	SN	숫자
8	JKB	부사격 조사	28	SO	붙임표
9	JKC	보격 조사	29	SP	쉼표, 가운데점, 콜론, 빗금
10	JKG	관형격 조사	30	SS	따옴표, 괄호표, 줄표
11	JKO	목적격 조사	31	SW	기타 기호
12	JKQ	인용격 조사	32	VA	형용사
13	JKS	주격 조사	33	VCN	부정 지정사
14	JKV	호격 조사	34	VCP	긍정 지정사
15	JX	보조사	35	W	동사
16	MAG	일반 부사	36	XPN	체인 접두사
17	MM	관형사	37	XR	어기
18	NNBG	의존 명사	38	XSA	형용사 파생 접미사
19	NNG	일반 명사	39	XSN	명사 파생 접미사
20	NNP	고유 명사	40	XSV	동사 파생 접미사