

# 형태소 분석기

## Term Project

Due : 2017년 6월 26일 PM 11:59

과제는 한국어 형태소 분석을 하는 것입니다. Python을 사용해서 어떤 모델이든 사용하여 결과가 잘 나오는 것을 제출하면 됩니다.

제출물 : 소스코드, 보고서

- 소스코드 파일명 : 학번\_이름\_코드
- 보고서 파일명 : 학번\_이름\_보고서

ex) 2013105030\_홍길동\_코드.py, 2013105030\_홍길동\_보고서

보고서에는 다음 내용이 반드시 첨부되어야 함.

1. 사용한 모델에 대한 설명
2. 소스코드 스크린샷 및 설명
3. 실행 결과, 분석

## 주어지는 파일 : Train.txt

```
train.txt
1 아름다운 아름답/Va+L/ETM
2 그 그/MM
3 시작 시작/NNG
4 "이행, "/SS+이/NNP+행/NNG+,/SP
5 잘 잘/MAG
6 가시오. " 가/VV+시/EP+오/EF+./SF+" /SS
7 석훈이 석훈/NNP+이/JKS
8 빙그르르 빙그르르/MAG
9 웃으며 웃/VV+으며/EC
10 악수를 악수/NNG+를/JKO
11 청했다. 청하/VV+았/EP+다/EF+./SF
12
13 손바닥의 손바닥/NNG+의/JKG
14 뜨거운 뜨겁/Va+L/ETM
15 기운, 기운/NNG+./SP
16 힘, 힘/NNG+./SP
17 움직임. 움직임/NNG+./SF
18
19 문득 문득/MAG
20 바라본 바라보/VV+L/ETM
21 하늘은 하늘/NNG+은/JX
22 시원하게 시원/XR+하/XSA+게/EC
23 펼쳐진 펼쳐/VV+어/EC+지/VX+L/ETM
24 룬빛깔이었다. 룬/NNG+빛깔/NNG+이/VCP+였/EP+다/EF+./SF
25
26 그 그/MM
27 순간 순간/NNG
28 진우의 진우/NNP+의/JKG
29 눈가에 눈가/NNG+에/JKB
30 얼핏 얼핏/MAG
31 물기가 물기/NNG+가/JKS
32 어렸다. 어리/Va+였/EP+다/EF+./SF
33
```

- 인풋 파일 구조

어절1 형태소1/형태소태그+형태소2/형태소태그...+형태소n/형태소태그

어절2 형태소1/형태소태그+형태소2/형태소태그...+형태소n/형태소태그

...

어절n 형태소1/형태소태그+형태소2/형태소태그...+형태소n/형태소태그

중간의 줄바꿈되어 문장 경계를 나타냅니다.

## Part-of-Speech tags for Korean

체언	명사	NN	일반명사	NNG
			고유명사	NNP
			의존명사	NNB
	대명사	NP	대명사	NP
	수사	NR	수사	NR
용언	동사	VV	동사	VV
	형용사	VA	형용사	VA
	보조용언	VX	보조용언	VX
	지정사	VC	긍정지정사 부정지정사	VCP VCN
수식언	관형사	MM	관형사	MM
	부사	MA	일반부사	MAG
			접속부사	MAJ
독립언	감탄사	IC	감탄사	IC
관계언	격조사	JK	주격조사	JKS
			보격조사	JKC
			관형격조사	JKG
			목적격조사	JKO
			부사격조사	JKB
			호격조사	JKV
			인용격조사	JKQ
	보조사	JX	보조사	JX
의존형태	어미	E	선어말어미	EP
			종결어미	EF
			연결어미	EC
			명사형전성어미	ETN
			관형형전성어미	ETM
	접두사	XP	체언접두사	XPN
	접미사	XS	명사파생접미사	XSN
			동사파생접미사	XSV
			형용사파생접미사	XSA
	어근	XR	어근	XR
기호	마침표, 물음표, 느낌표			SF
	윺표, 가운뎃점, 콤마, 빗금			SP
	따옴표, 괄호표, 줄표			SS
	줄임표			SE
	불임표(물결, 숨김, 빠짐)			SO
	외국어			SL
	한자			SH
	기타기호(논리 수학기호, 화폐기호 등)			SW
	명사추정범주			NF
	용언추정범주			NV
	숫자			XN
	분석불능범주			NA