

# Perl script를 이용한 의존구문분석

# • perl script 사용하기

- ../SKKU-BK21/stanford 폴더 내부에 대상 파일을 위치시킨다.

2021_scale_linguistics > SKKU-BK21 > stanford			
이름	수정된 날짜	유형	크기
2runPaser4English	2021-06-17 오후 2:20	텍스트 문서	1KB
AustenL7-TNT	2021-06-09 오전 9:52	텍스트 문서	39KB
len9	2021-07-26 오후 1:03	텍스트 문서	6KB
SSausten-EN-20	2020-08-18 오후 6:24	텍스트 문서	7KB
StanfordCoreNLP.properties	2017-01-20 오전 9:54	PROPERTIES 파일	4KB
stanford-corenlp-3.9.1	2018-02-28 오후 7:24	ALZip JAR File	7,903KB
stanford-corenlp-3.9.1-models	2018-02-28 오후 7:24	ALZip JAR File	354,075KB
StanfordCoreNLP-english-all.properties	2017-05-26 오전 1:58	PROPERTIES 파일	5KB
StanfordCoreNLP-english-fast.properties	2017-05-26 오전 1:58	PROPERTIES 파일	4KB
stanford-parser	2018-03-01 오후 9:04	ALZip JAR File	4,312KB
stanford-parser-3.9.1-javadoc	2018-03-01 오후 9:04	ALZip JAR File	5,221KB
stanford-parser-3.9.1-sources	2018-03-01 오후 9:04	ALZip JAR File	3,440KB

# • perl script 사용하기

- cmd 창을 활성화시킨다.

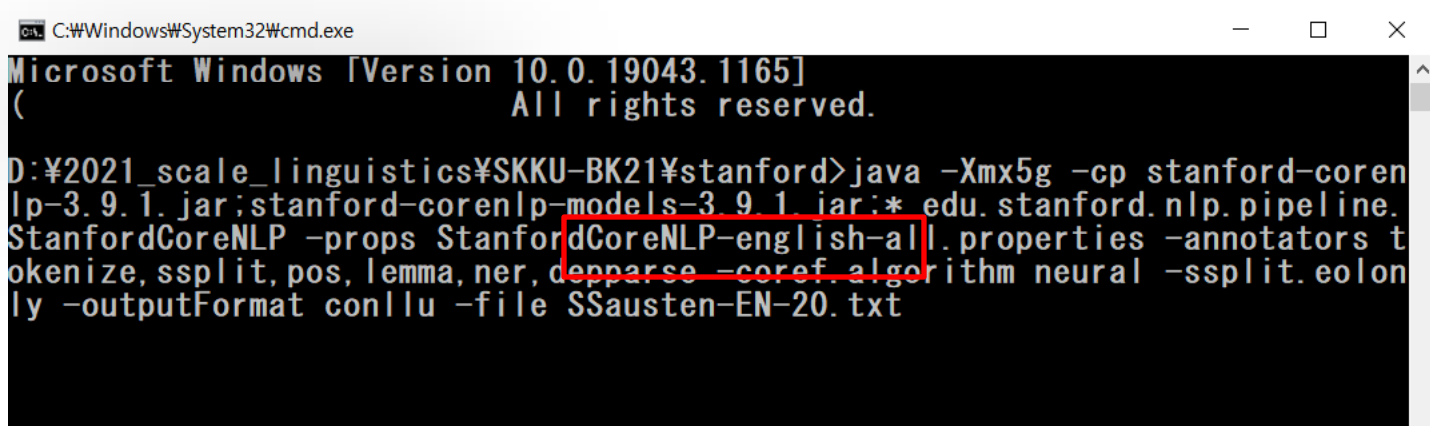
Link	cmd				
	cmd				
	"cmd" 검색				
	2runPaser4English	2021-06-17 오후 2:20	텍스트 문서	1KB	
	AustenL7-TNT	2021-06-09 오전 9:52	텍스트 문서	39KB	
	len9	2021-07-26 오후 1:03	텍스트 문서	6KB	
	SSausten-EN-20	2020-08-18 오후 6:24	텍스트 문서	7KB	

# • perl script 사용하기

- cmd 창에서 아래의 명령어를 입력한다.

```
>>> java -Xmx5g -cp stanford-corenlp-3.9.1.jar;stanford-corenlp-models-3.9.1.jar;* edu.stanford.nlp.pipeline.StanfordCoreNLP -props StanfordCoreNLP-english-all.properties -annotators tokenize,ssplit,pos,lemma,ner,depparse -coref.algorithm neural -ssplit.eolonly -outputFormat conllu -file 대상_파일_이름
```

※대상 파일은 확장자까지 입력해야 합니다.



```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.19043.1165]
(
All rights reserved.

D:\2021_scale_linguistics\SKKU-BK21\stanford>java -Xmx5g -cp stanford-corenlp-3.9.1.jar;stanford-corenlp-models-3.9.1.jar;* edu.stanford.nlp.pipeline.StanfordCoreNLP -props StanfordCoreNLP-english-all.properties -annotators tokenize,ssplit,pos,lemma,ner,depparse -coref.algorithm neural -ssplit.eolonly -outputFormat conllu -file SSausten-EN-20.txt
```

# • perl script 사용하기

- 완료되면 같은 위치에 [파일이름].conllu 파일이 생성됨.

2021\_scale\_linguistics > SKKU-BK21 > stanford



이름	수정한 날짜	유형	크기
2runPaser4English	2021-06-17 오후 2:20	텍스트 문서	1KB
AustenL7-TNT	2021-06-09 오전 9:52	텍스트 문서	39KB
len9	2021-07-26 오후 1:03	텍스트 문서	6KB
Ssausten-EN-20	2020-08-18 오후 6:24	텍스트 문서	7KB
Ssausten-EN-20.txt.conllu	2021-06-17 오후 2:19	CONLLU 파일	48KB

# • perl script 사용하기

- conllu format
- .conllu 파일은 메모장, notepad++ 또는 Editplus에서 열 수 있습니다.

ID	FORM	LEMMA	PLEMMA	POS	PPOS	HEAD	DEPREL	PHEAD	PDEPREL
1	This	this	—	DT	—	4	nsubj	—	—
2	was	be	—	VBD	—	4	cop	—	—
3	the	the	—	DT	—	4	det	—	—
4	page	page	—	NN	—	0	root	—	—