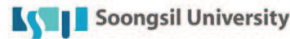


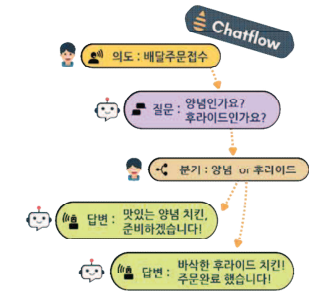
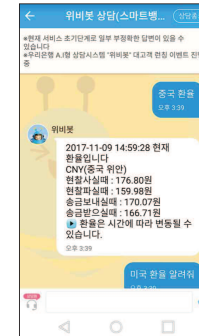
자연어 처리와 워드클라우드



베어드교양대학
강의선
백마관 203호
02-828-7264
iami86@ssu.ac.kr

챗봇(chatbot)

- 채팅(chatting)과 로봇(robot)이 결합되어 만들어진 용어
- 인간의 자연언어를 이해하고 대화가 가능한 채팅 로봇
 - 초기 챗봇 : 입력한 텍스트를 이해하고 답변하는 텍스트 기반
 - 최근 챗봇 : STT(Sound to Text) 및 TTS(Text to Sound) 기술과 접목되어 음성 기반으로 동작



Creative and Smart! LG CNS :: 4차 산업혁명에서 지능형 챗봇의 역할에서 발체
Soongsil University

자연어 처리(NLP)

- Natural Language Processing
- 기계가 자연어를 이해하고 해석하여 처리할 수 있도록 하는 일
- 활용분야
 - 스팸 메일 분류
 - 뉴스 기사의 내용을 기반으로 연애/정치/사회/문화 중 어떤 카테고리에 속하는지 자동으로 분류
 - SNS의 글을 분석하여 글쓴이의 감정을 분석하는 것, 영화 및 제품의 리뷰를 분석하는 것
 - 구글 번역기나 파파고와 같은 번역기
 - 애플의 시리나 삼성 갤럭시의 빅스비, 챗봇 등

한글 자연어 처리 : koNLpy 라이브러리

- 자연어 처리를 위해서는 문장, 형태소, 품사 태깅같은 텍스트 분석이 필요
 - 단어 : 의미를 갖는 문장의 가장 작은 단일 요소
 - 형태소(morpheme)
 - 언어에서 의미가 있는 가장 작은 단위
 - 형태소, 어근, 접두사/접미사, 품사등 언어학적 속성들
 - 품사 태깅
 - 형태소의 뜻과 문맥을 고려하여 품사를 붙이는 것
- 한글 텍스트 분석은 영어에 비해 조사와 복합 명사 등의 분리가 어려움

한글 자연어 처리 : koNLpy 라이브러리

■ koNLpy

- 한글 자연어 처리를 위해 국내에서 개발한 파이썬 기반 오픈 소스
- Hannanum, Kkma, Komoran, Mecab, Okt(Twitter) 와 같은 품사 태깅 패키지 지원
- 예) 아버지가방에 들어가신다

Hannanum	Kkma	Komoran	Mecab	Okt(Twitter)
아버지가방에들 어가/N	아버지/NNG	아버지가방에들 어가신다/NNP	아버지/NNG	아버지/NOUN
이/J	가방/NNG		가/JKS	가방/NOUN
시 L 다/E	예/JKM		방/NNG	예/JOSA
	들어가/W		예/JKB	들어가신/VERB
	시/EPH		들어가/W	
	S다/EFN		신다/EP+EC	

품사 태깅 비교의 예

5

koNLpy 설치하기

```
pip install koNLpy
```

- java 엔진에서 동작하므로 java 설치 필요
- 하지만 colab 에서는 java 엔진 설치 필요하지 않음

6

koNLpy에서 Twitter(Okt) 품사 태깅 패키지 사용하기

■ Okt 라이브러리 로딩하기

```
from konlpy.tag import Okt
```

■ Okt 라이브러리를 사용하기 위한 객체 생성

- 변수명 = Okt()

■ 형태소 분석하기

- 변수명.morphs(문자열)

■ 품사태깅 확인하기

- 변수명.pos(문자열)

■ 명사 데이터 확인하기

- 변수명.nouns(문자열)

7

koNLpy를 워드 클라우드에 적용하기

■ 세종대왕 워드클라우드 만들기

1. 텍스트 화일과 이미지 화일 불러오기

- 세종대왕.txt , maskHeart.jpg

2. 필요한 라이브러리들 로딩하기

- ① 워드 클라우드 생성에 필요한 라이브러리들(wordcloud, STOPWORDS, matplotlib)
- ② 마스크 이미지 처리를 위해 필요한 라이브러리들(Image, numpy)

3. 관련 화일들 open 하고 저장하기

- ① 워드 클라우드 생성을 위한 텍스트 화일을 열어 읽어오기
- ② 마스크 이미지 화일 읽어와서 array로 변경하기

4. 자연어 처리하기

- ① 자연어 처리를 위한 라이브러리 설치 및 konlpy 라이브러리 로딩
- ② Okt를 이용하여 명사 추출
- ③ 명사들을 문자열로 변환하기

5. 워드 클라우드 생성하기

8

수고하셨습니다.