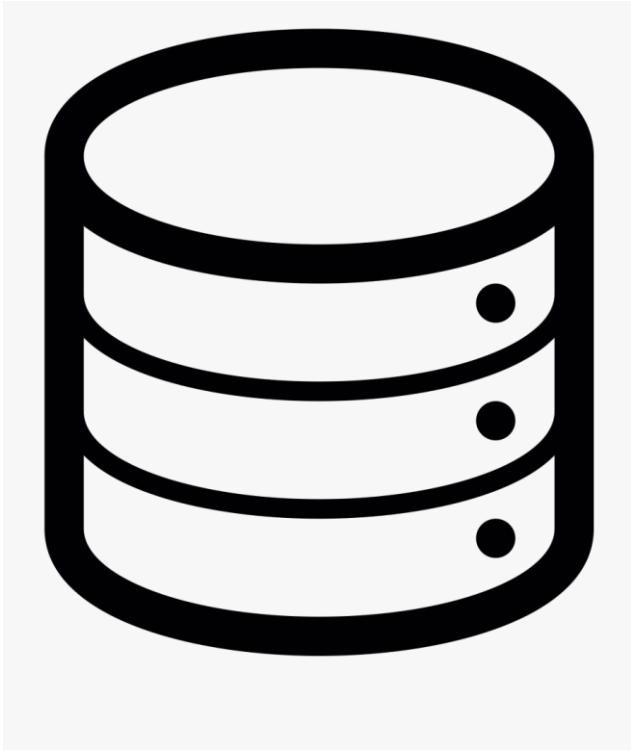


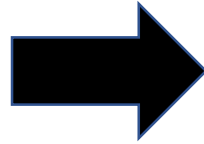
트위터 감정분석 위한 DB구축

조국 전 법무부 장관 관련 트윗

1714670 통계학과 노유진



SNS 데이터



감정에 Score를 매김

크롤링 대상 및 과정

크롤링 대상

조국 전 법무부 장관 관련 트윗

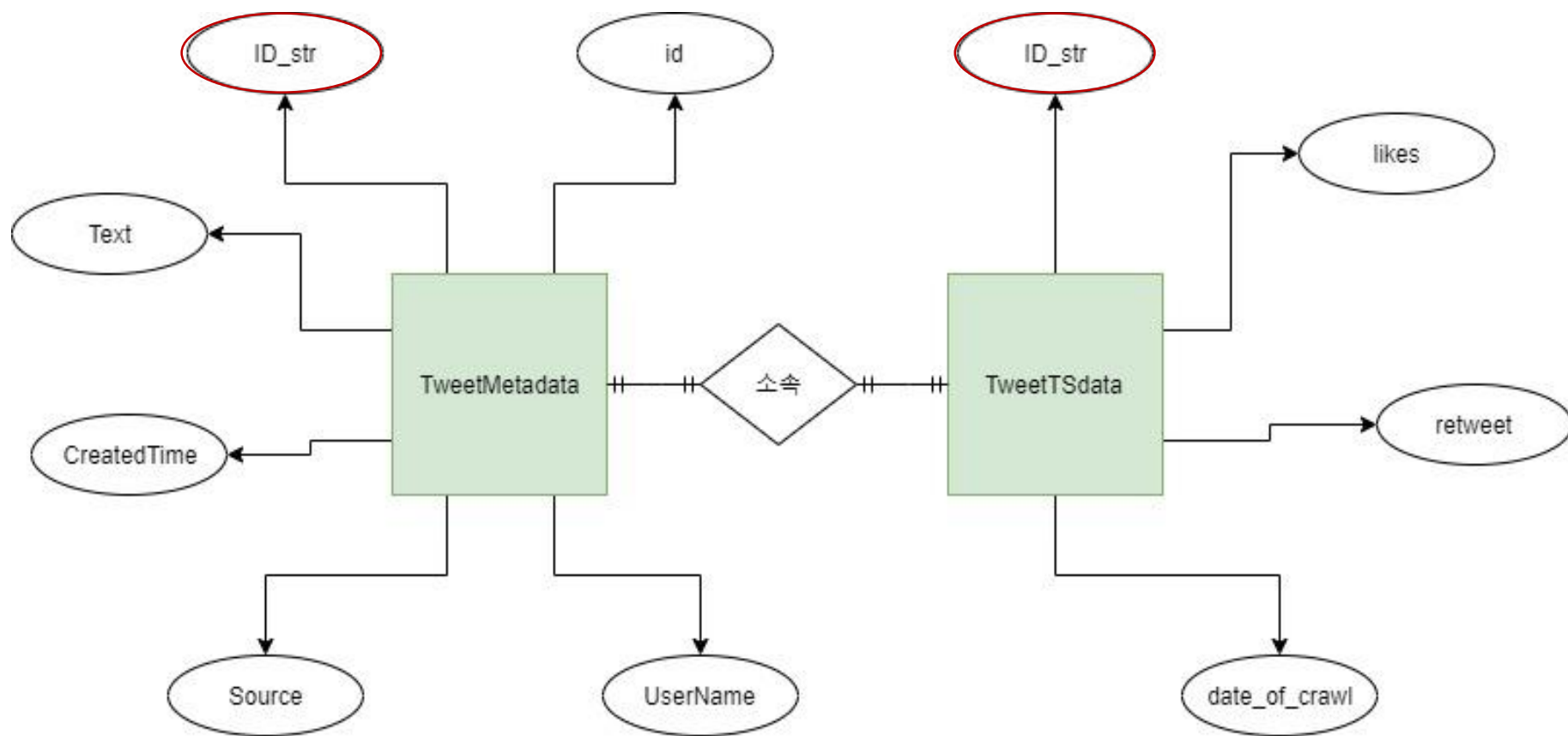
크롤링 과정

Tweepy 패키지를 통해 트위터 crawling

리트윗으로 인한 중복 제거

Pymysql 패키지를 통해 MySQL에 저장하기

DB의 ER Diagram



R Shiny 구현

Tweet DB

시작날짜 지정:
2019-12-08 - 2019-12-09
Apply Changes

Sentimental Analysis

Analysis:

sent	score	magni
조국은 영리하고 교활하다!....	2.00	1.00

현재 DB구조 및 저장 전략

Metadata

ID	ID_str	UserName	Upload_time	Text
1831	1203115520058281984	icho3599	2019-12-07 00:54:27	조국만 남았다..종착지 향하는 '유재수 감찰무마 의혹' 수사 다음뉴스 https://t.co/s6MoDOSfDx 조국을 털기 위해...
1832
1833

Time series data

ID_str	Date_of_crawl	Retweet	like
1203115520058281984	2019-12-08 04:04:19	4514	394
...
...

노하우 & 애로사항

애로사항

이모티콘 저장의 어려움

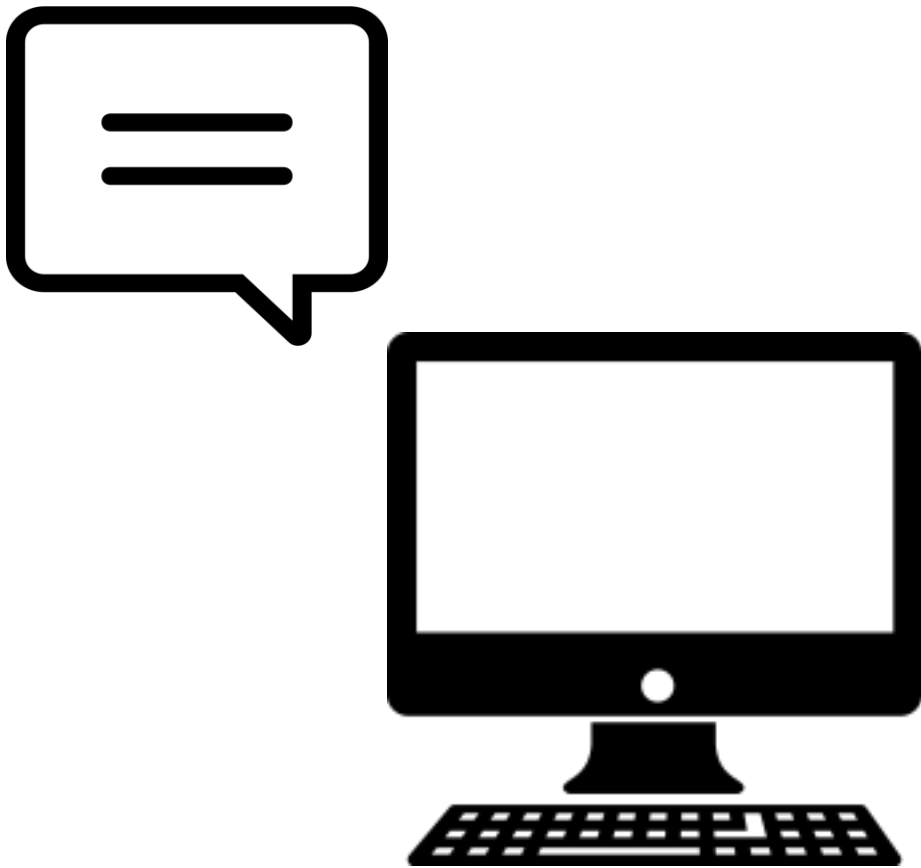
수 많은 리트윗으로 인해 중복 제거 시 적은 양의 데이터만 저장됨

R Shiny 구현의 오류로 인해 더 좋은 결과를 보여주지 못함

제대로 된 감정 분석이 이뤄지지 않음 (군산대 감정분석 사전, 비꼬는 어조를 제대로 분석 못함)

노하우

ALTER TABLE TweetMetadata **MODIFY** Text BLOB; 을 통해 이모티콘 저장



악플 탐지기

트위터나 다양한 SNS에 적혀 있는 악플을 크롤링해 저장하는 프로그램 생성

현재의 프로젝트에서 User에 관한 정보를 더 추가하고 더 세밀한 감정 분석을 통해 이루어질 수 있을 것 같음

왕초짜의+데이터+정복기+#1+다짜고짜+크롤링

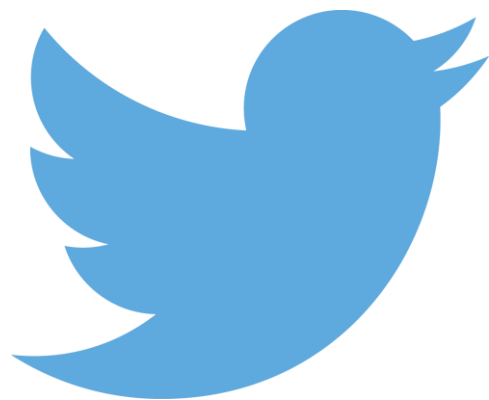
https://github.com/YBIGTA/Deep_learning/blob/master/Scratch%20ML/%EC%99%95%EC%B4%88%EC%A7%9C%EC%9D%98%2B%EB%8D%B0%EC%9D%B4%ED%84%B0%2B%EC%A0%95%EB%B3%B5%EA%B8%B0%2B%231%2B%EB%8B%A4%EC%A7%9C%EA%B3%A0%EC%A7%9C%2B%ED%81%AC%EB%A1%A4%EB%A7%81.md

python tweepy 사용한 트위터 크롤링(crawling)

<https://kslee7746.tistory.com/entry/python-tweepy-%EC%82%AC%EC%9A%A9%ED%95%9C-%ED%8A%B8%EC%9C%84%ED%84%B0-%ED%81%AC%EB%A1%A4%EB%A7%81crawling>

JW's Blog

<https://jeongwookie.github.io/2019/06/10/190610-twitter-data-crawling/>



감사합니다