2014년 10월 20일 워드클라우드 결과

1. 프로세스

- 엑셀 파일로 저장된 트위터 크롤링 결과물을 데이터프레임에 불러온다.

- 한글로 된 날짜를 datetime 형식에 맞게 변환해서 저장한다.

- [이름, ID, 날짜, 트윗] 한글로 된 컬럼명을 영어 [nickname, id, date, tweet] 으로 바꾼다.

- tweets(::list) 라는 변수에 트위터 글을 모조리 저장한다.

- 이를 pytagcloud 에 활용하기 위해 string 형식으로 바꿔준다. twt\_str(::str)

- konlpy의 tag 중 Twitter()를 활용하여 명사만 추출해서 저장한다.

- 그리고 이를 collections 모듈의 Counter 함수를 통해 각 단어의 빈도를 저장.

- pytagcloud 라이브러리로 워드 클라우드 만들어서 png 파일로 저장.

2. 결과물 1



- ‘세월호’ 를 포함한 상위 100개 단어를 그린 결과

- ‘세월호’의 빈도가 너무 높아 다른 단어들을 제대로 볼 수 없다.

3. 결과물 2



- ‘세월호’를 제외한 상위 100개 명사로 그린 워드클라우드.

- 비슷한 시기에 있었던 판교 환풍구 추락 사고와 관련된 키워드들이 보인다; 이재명, 성남, 판교, 환, 풍구,