

# Assignment1

## 데이터과학응용3-DATA410

1. Gensim 라이브러리를 활용하여 네이버 영화리뷰 데이터셋의 단어를 단어임베딩 (word2vec)으로 표현하세요
  2. 표현된 단어벡터를 시각화하기 위한 차원축소 알고리즘(ex:PCA, t-SNE)을 활용하여 시각화하세요
- 불용어 제거, 명사만 추출 등을 통해 다양한 단어 임베딩 표현이 가능할 것입니다

nsmc 데이터셋: [https://github.com/e9t/nsmc/blob/master/ratings\\_train.txt](https://github.com/e9t/nsmc/blob/master/ratings_train.txt)

gensim library: <https://radimrehurek.com/gensim/>

## 제출물 목록

1. 학번\_이름.zip
  - a. (.py) 소스코드 파일
  - b. 시각화 png 파일