

The background of the slide is a dark, textured surface covered with numerous 3D question marks. Most of these question marks are black and have a matte finish. One question mark, located on the right side of the slide, is a vibrant red color and has a glossy, reflective surface. The lighting creates highlights and shadows on the 3D shapes, giving them a sense of depth and volume.

# 일간베스트 댓글에서 나타나는 혐오 표현 분석

소프트웨어학과 20213021 신채원

# 목차



## 01

### 데이터 추출, 전처리 및 형태소 분석

- 일간베스트 데이터 중 댓글만을 추출
- 띄어쓰기 및 맞춤법 검사를 통한 전처리 작업
- KLT2000 혹은 konlp를 통한 형태소 분석

## 02

### 빈도수 계산 및 워드 임베딩, 유사어 그래프화

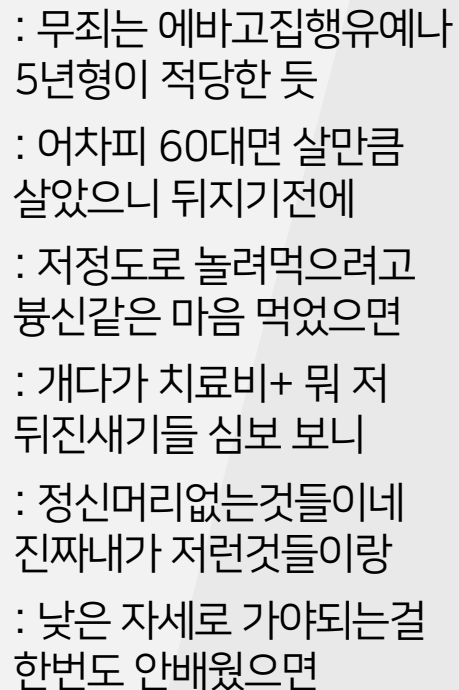
- pandas를 사용한 단어 출현 횟수 계산
- word2vec을 통한 워드 임베딩
- 그래프 생성을 통한 유사어 네트워크 형성

## 03

### 혐오 표현 분석 결과

- 커뮤니티 내의 단순한 욕설을 넘어선 혐오
- 일간베스트의 혐오의 대상이 무엇인지

# 데이터 추출, 전처리 및 형태소 분석

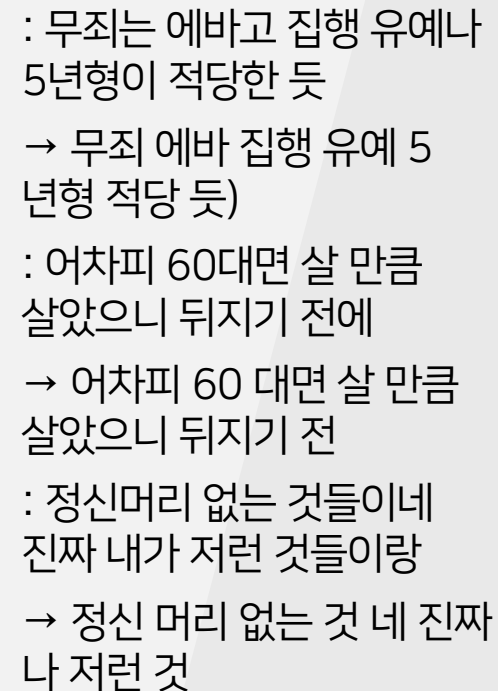


: 무죄는 에바고 집행유예나  
5년형이 적당한 듯  
: 어차피 60대면 살만큼  
살았으니 뒤지기전에  
: 저정도로 놀려먹으려고  
붕신같은 마음 먹었으면  
: 개다가 치료비+ 뭐 저  
뒤진새기들 심보 보니  
: 정신머리없는것들이네  
진짜내가 저런것들이랑  
: 낮은 자세로 가야되는걸  
한번도 안배웠으면

py-hanspell을 사용한 맞춤법 검사



KLT2000을 사용한 형태소 분석



: 무죄는 에바고 집행 유예나  
5년형이 적당한 듯  
→ 무죄 에바 집행 유예 5  
년형 적당 듯)  
: 어차피 60대면 살 만큼  
살았으니 뒤지기 전에  
→ 어차피 60 대면 살 만큼  
살았으니 뒤지기 전  
: 정신머리 없는 것들이네  
진짜 내가 저런 것들이랑  
→ 정신 머리 없는 것 네 진짜  
나 저런 것

# 단어 출현 빈도수 계산

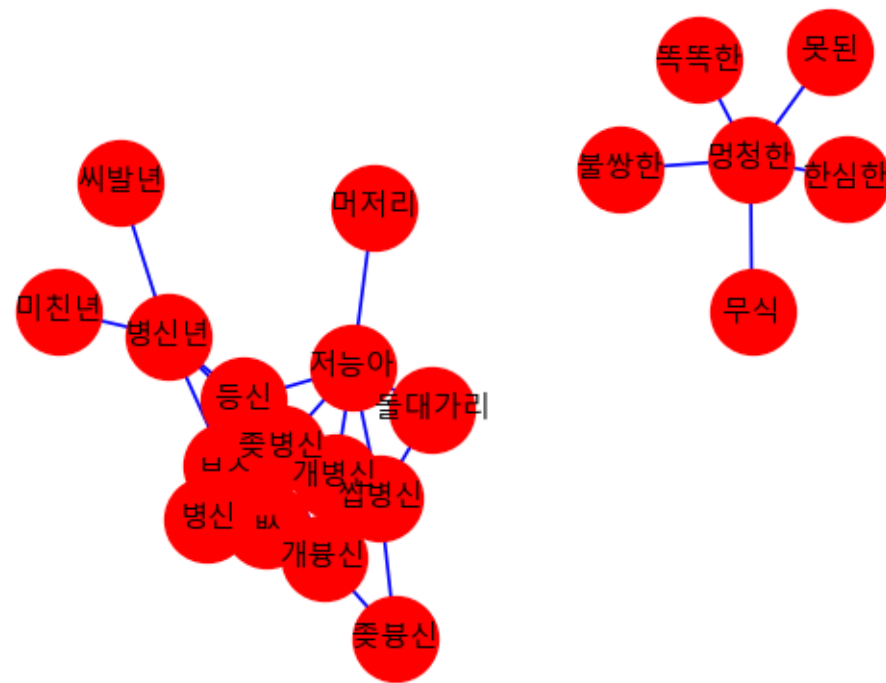
새끼	1위 (전체 1위) 새끼 has appeared 108446 times.	병신	2위 (전체 11위) 병신 has appeared 43295 times.	년	3위 (전체 24위) 년 has appeared 29765 times.
4위 (전체 25위) ㅁㅈㅎ has appeared 29679 times.	ㅁㅈㅎ	5위 (전체 26위) 시발 has appeared 29157 times.	시발	7위 (전체 48위) 홍어 has appeared 19900 times.	홍어

# 워드 임베딩 및 결과 시각화

1. word2vec을 사용한 단어 임베딩
2. 출현 빈도수가 높은 혐오 표현에 대한 유사어 탐색
3. networkx과 matplotlib을 이용한 단어 유사도 시각화

## 혐오 표현 2위 (전체 11위) 병신 유사어 탐색 및 시각화 예시

('좇병신', 0.806771457195282), ('등신', 0.779389500617981),  
('개병신', 0.6732909679412842), ('ㅂㅅ', 0.6521783471107483),  
('저능아', 0.6465569734573364), ('병신년', 0.5834298133850098),  
('씹병신', 0.5703725218772888), ('멍청한', 0.5504761934280396)

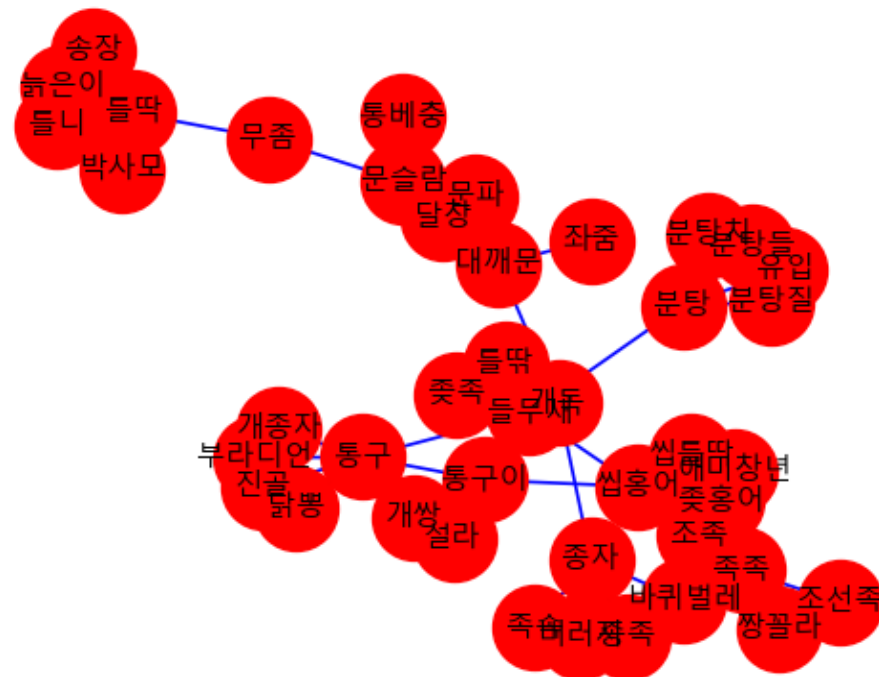


# 워드 임베딩 및 결과 시각화

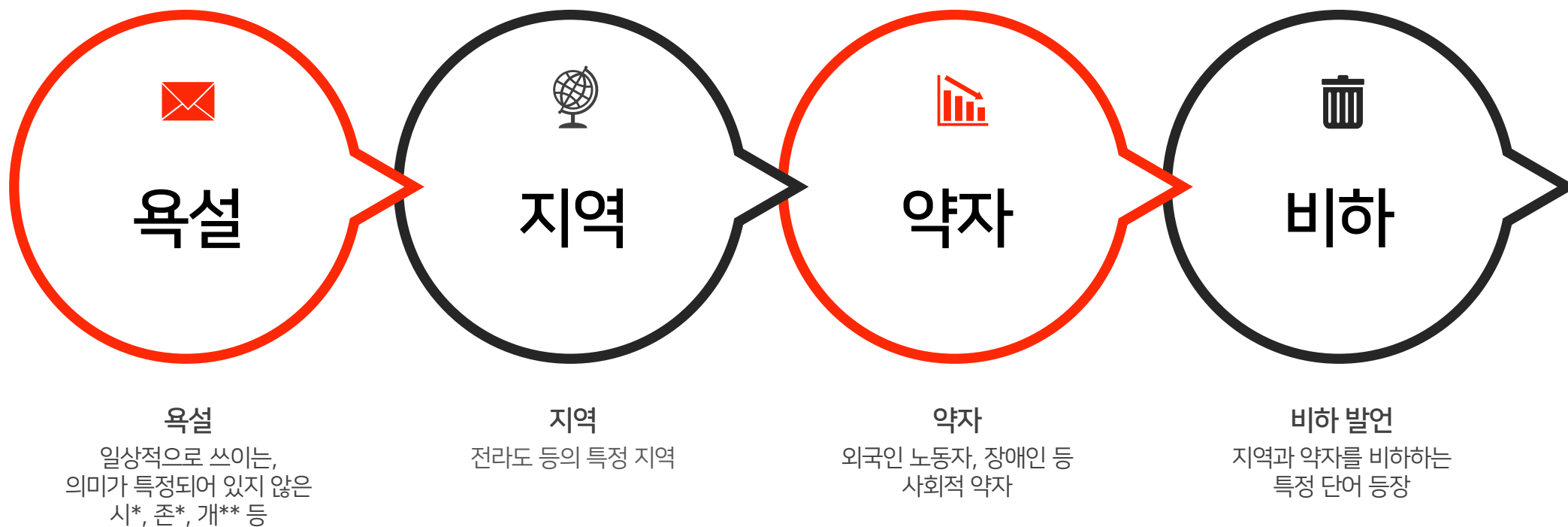
1. word2vec을 사용한 단어 임베딩
2. 출현 빈도수가 높은 혐오 표현에 대한 유사어 탐색
3. networkx과 matplotlib을 이용한 단어 유사도 시각화

## 혐오 표현 7위 (전체 48위) **홍어** 유사어 탐색 및 시각화 예시

('분탕', 0.617926299571991), ('통구이', 0.6037446856498718),  
('들무새', 0.6024805307388306), ('싹홍어', 0.5826643705368042),  
('대깨문', 0.5765116810798645), ('틀딱', 0.5649189352989197),  
('종자', 0.5580267310142517), ('문슬람', 0.5476859211921692)



# 혐오 표현 분석 결과



# 혐오 표현 분석 결과

## 다양한 혐오 대상

단순한 욕설에서 그치지 않고  
여성, 장애인, 외국인 노동자, 민주화 운동 등  
사이트 내 다양한 혐오의 대상이 존재

## 맥락 없는 혐오

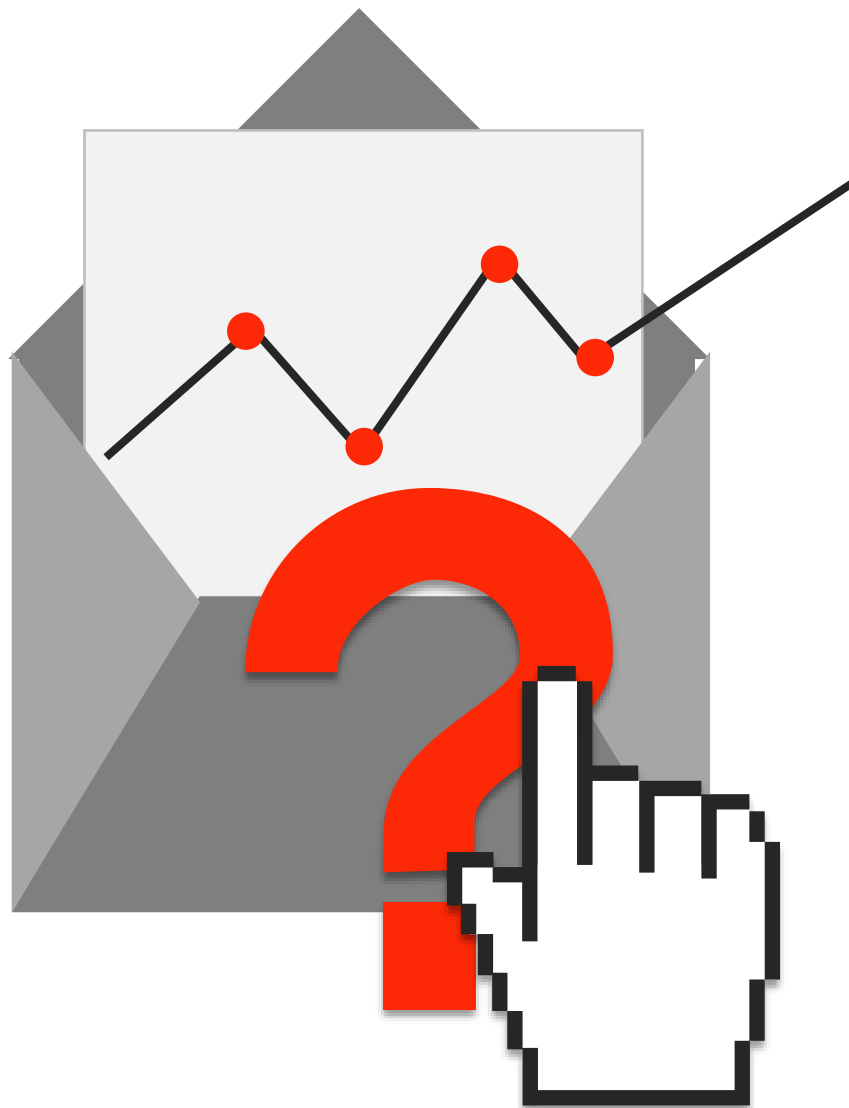
댓글에 동반되는 욕설과 혐오 발언  
상대 + 혐오 표현이 향하는 혐오의 대상  
동시적 비하 및 혐오 발생

## 높은 비율의 혐오 표현

총 1196346문장에 대해 형태소 분석 수행,  
5649988개 형태소 추출  
출현 빈도 상위 100위 → 혐오 표현 16개  
1위 108446회 출현, 16위 11666회 출현

## 혐오를 내포하는 새 용어 창조

틀딱(노인), 뽕지(민주화),  
홍어(전라도 비하) 등 혐오의 뜻을  
내포하고 있는 새로운 용어 창조





감사합니다.

