

자연어 처리를 활용한 한국어 유튜브 댓글 종류 분류



목차

1

프로젝트 소개

2

진행 상황

3

향후 계획

Part 1

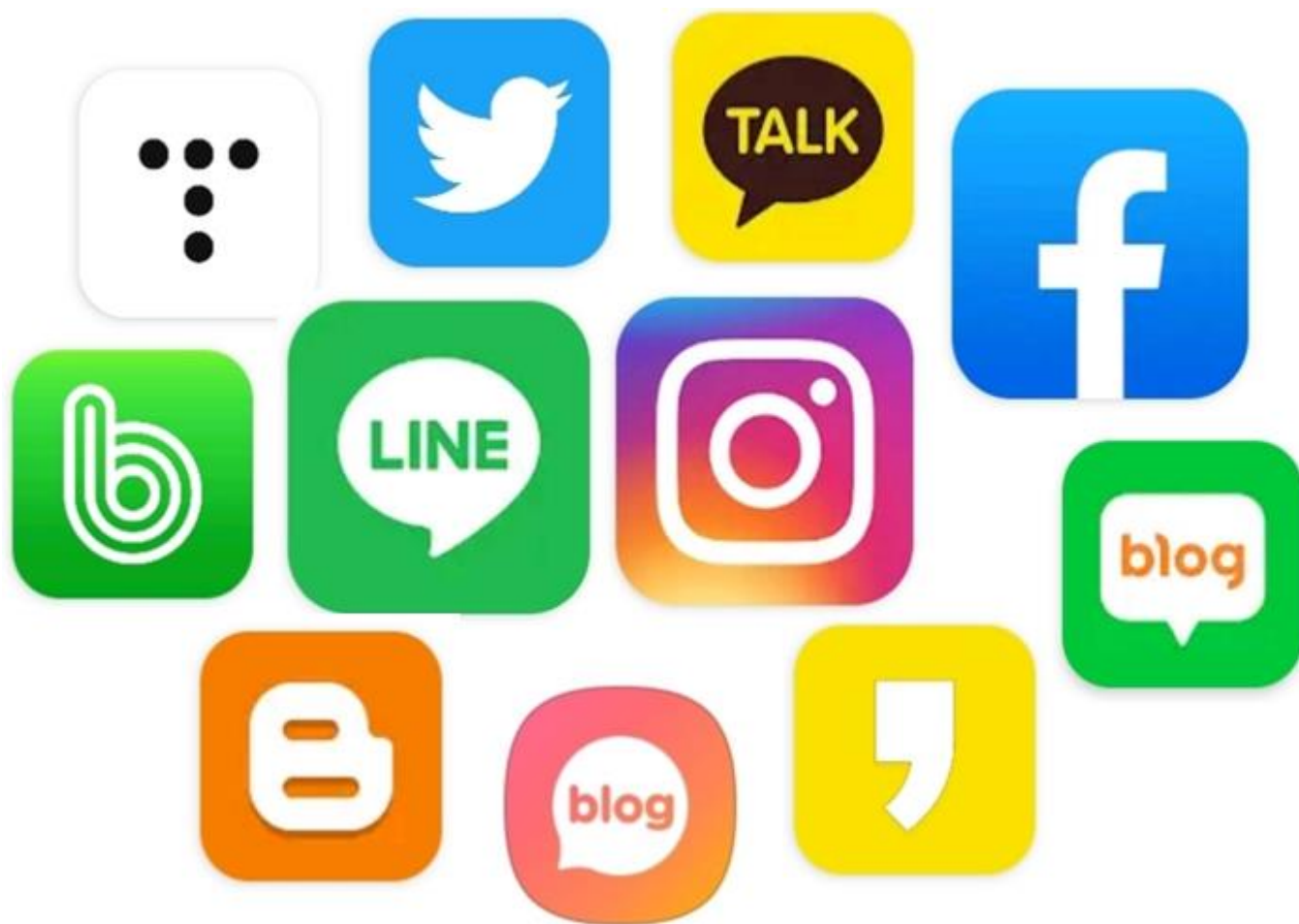
프로젝트 소개

1.1 개요

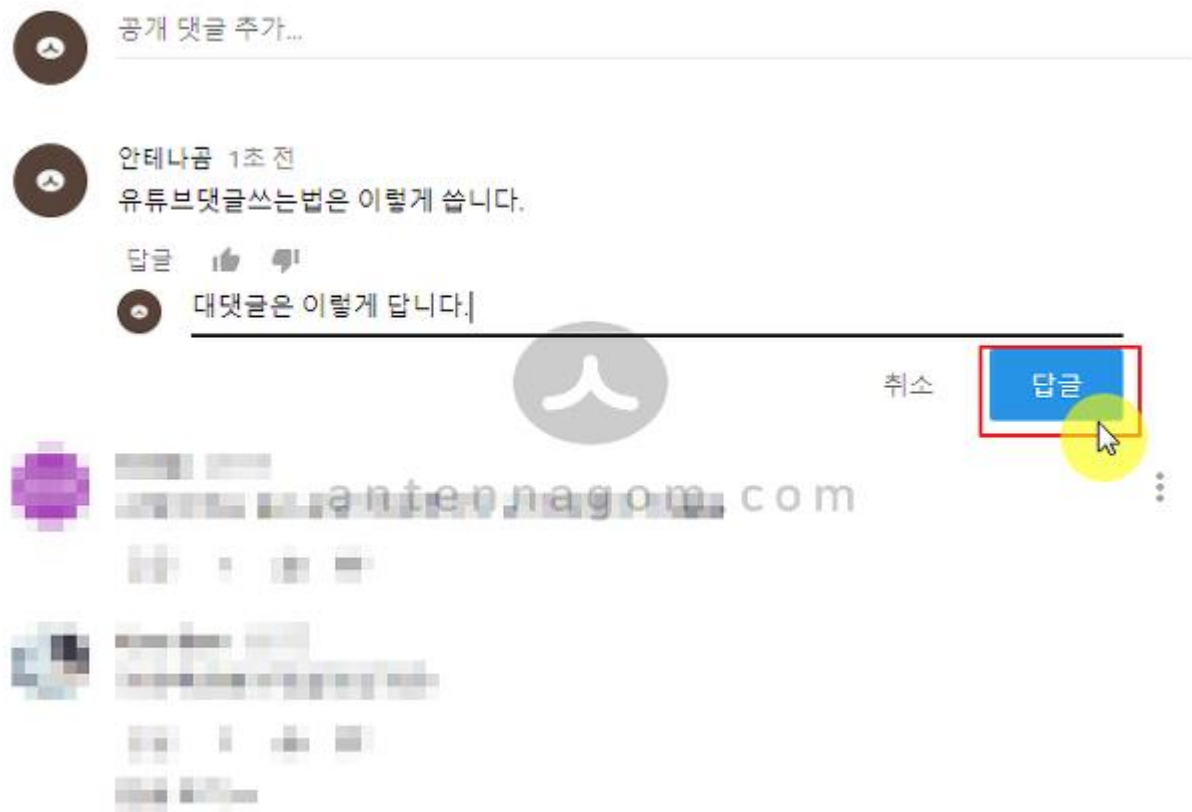
1.2 프로젝트 일정



© 2006 The Authors
Journal compilation © 2006 Blackwell Publishing Ltd



Part 1.1 개요



Part 1.2 프로젝트 일정

001

유튜브 영상 한국어
댓글 데이터 수집

2023.3~4

002

데이터 셋 전처리

2023.4~5

003

목적에 따른
데이터 분류
(주로 오픈소스 활용)

2023.5~1학기 종강

004

모델 개선 및
활용 방법 구상

2023.여름방학~2학기

Part 2

진행 상황

21 데이터 셋

22 데이터 전처리



Part 2.1 데이터 셋

	아이디	
0	ㅏㅏ	방장이 얼마나 열심히 일해왔으면 아직까지 빈자리를 채우는 침튜브.. 항상 감사하다
1	박상현	방장이 없어도 돌아가는 침튜브, 이게 바로 제로 아닐까?
2	쓸데없이만재있어	침착맨마저 제로가 된 세상.
3	덕키	실때 든든한 고봉밥 만들어 놓아서 이렇게 침손실없이 영상보내 침순이는 행복합니다
4	플레르드뽀	침착맨 권다니까 침착맨원본박물관,침착맨플러스 정독하고 있음 ... 오히려 좋아 ..밀린 거 많았는데 따라가라고 시
5	부장아재해돌쿤	아니 얼마나 남겨 났길래 유튜부쟁이들은 개방장 방송 쉬는걸 느꼈수가 없네 침수자들 넘모 고마워요 요로분♡
6	xnbdjsak	뇌질과 호들갑에 누구보다 심술을 잘 내면서 관련 컨텐츠는 꼭 짝어 내보내는 역설적인 유튜브
7	오현우	나의 저녁을 책임져준 침착맨 제로 감사하다
8	Nateee	이 날 침하하에서 침착맨 사랑해로 방송시작하자해서방송 시작 할 때 다 같이 침착맨 사랑해 채팅 쳤는데컨텐츠하.
9	우주고양이	이렇게 쉬더라도 끝없이 편집본이 나오는걸보니 이제 격달제로 라이브를 할지도
10	Play- maker	침착맨없어도 굴러가는 침튜브를 봤으니 이제 안심하고 원편데이를 맞이할수있겠다
11	지니	방송을 쉬면서도 계속 올라오는 침튜브 감사합니다
12	콩	웹시제로 라임맛 뒤로 빼는 거 보고 이 사람은 진짜 우리가 어느 포인트에서 열받는지 알고 있다는 생각이 들었음
13	검은냥이 코코& 잡덕집사	ㅋㅋㅋ 맛나게 드시지만 점점 뒤에서 배불러하는 맥콜병진 너무 쟈나요
14	쥬스	방장 없이 돌아가는 침튜브.. 방장 얼마나 일을 하고 간고야
15	김도지	지속가능한 침튜브를 위해 지금같이 열흘하고 품앗이 다니고 휴식 취하는건 어떨까
16	seunghun	제로가 되어버린 방장의 제로 음료 리뷰이게 진짜 제로 아닐까?
17	ㅏㅇ	환타제로 진짜 맛있지. 먹고 놀라서 한박스 쟁여놔었음
18	널디언이사랑해	"제로"침튜브에 올라온 "춘추제로시대".....이건 귀하다...
19	주마등	침착맨은 이제 하나의 기업이다. 그가 없어도 톱니바퀴는 돌아간다...
20	이윤경	폭력적인 귀여움에 가날픈 손가락까지 완벽하다

- 유튜브 영상 하나를 기준으로
약 1100~1200개의 댓글 수집가능

- 작성자 아이디 정보를 제외하고
순수 댓글 내용만 수집

- 같은 주제 영상의 댓글을 수집하여
한번에 분류할 예정

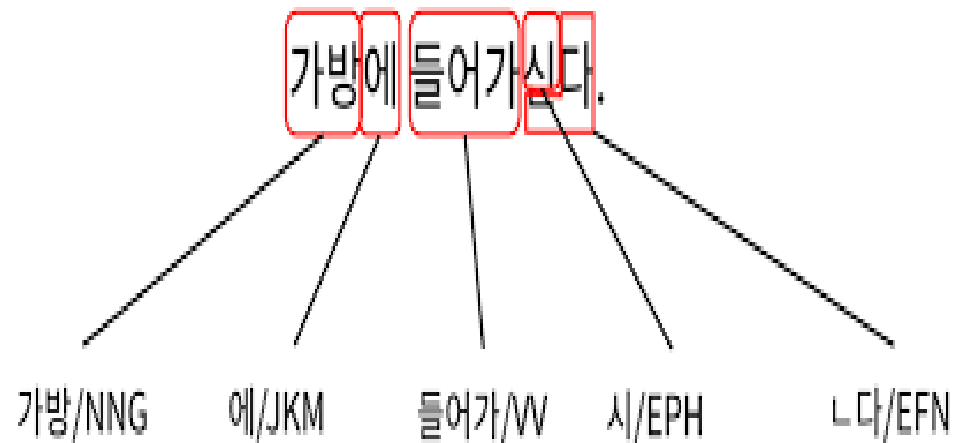
수집 데이터

Part 2.2 데이터 전처리



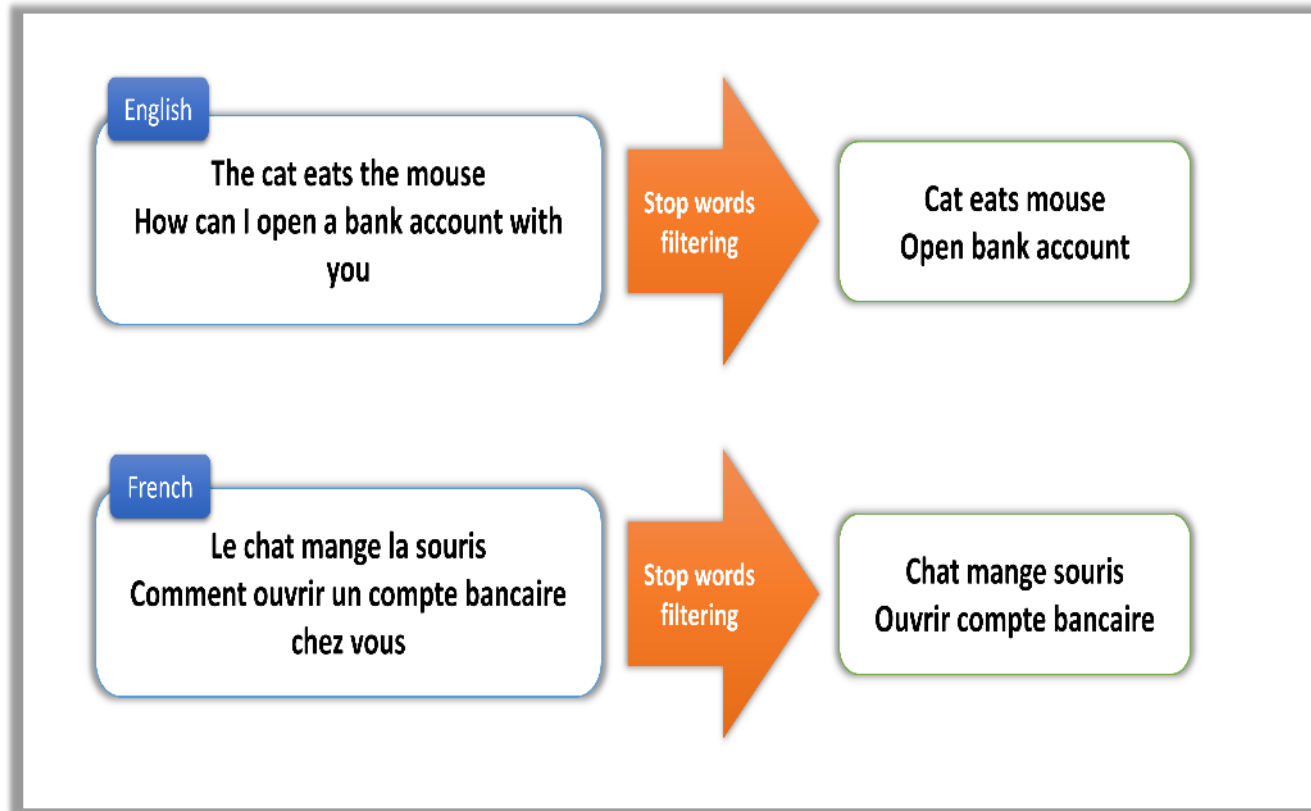
KoNLPy

한국어 자연어 처리 패키지



형태소 분석
(문장 토큰화)

Part 2.2 데이터 전처리



현재 **konlpy** 형태소 분석을 통해 조사, 접속사를 제거하는 방식으로 적용했다.

인터넷 댓글의 경우 구어체이고, 특유의 은어 등의 문제가 있으므로 추후 성능 개선을 위해 추가적으로 불용어 사전을 만들어 보려 함.

불용어 (stop_word) 처리

Part 2.2 데이터 전처리



Part 2.2 데이터 전처리



Part 3

향후 계획

3.1 모델 개선을 위한 기술 적용

3.2 실용적인 적용을 위한 아이디어 구상



Part 3.1 모델 개선을 위한 기술 적용



Part 3.2 실용적인 적용을 위한 아이디어 구상

현재 구상중인 아이디어

1. 구글 크롬 플러그인: 악성, 혐오 댓글 자동 필터링 시스템
2. 유튜브 채널 운영자를 위한 댓글 관리 보조 도구
3. 유튜브에서 벗어나서, SNS(인스타그램) 계정 설명 텍스트를 읽고 광고, 불법거래 등 악성 계정을 분류.



PPT 템플릿 출처:새별의 파워포인트
<http://bit.ly/saebyeol>