

캡스톤 디자인

최종 발표

154220 이호진

173050 김사연

154056 박다원

Ccontents

주제

- 주제
- 주제 선정 배경
- 목표

프로젝트 설계

- 데이터
- 딥러닝 모델
- 웹 서비스

데모 영상

Part 1 주제

딥러닝 기반 욕설 탐지 웹 서비스

Part 2 주제 선정 배경



[아나운서 김나정 악플러 고소... “수십 개 욕댓글로 너무 힘들다”](#)

국제신문 | 2019.11.05. | [🔗](#)

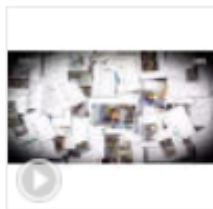
아나운서 김나정이 악플러들에 대해 강경 대응을 하겠다고 나섰다. 4일 자신의 인스타그램 계정을 통해 “오늘 아침까지 32명 검찰 **고소** ππ 심장 떨리고 자꾸 수십 개씩 욕댓글보니까 숨도 못 쉴 거 같고 힘들다”며...



[“악성 댓글 참다가 죽을 것 같아 고소” 악플과 전쟁 분투기](#)

한국일보 **PICK** | 1면2단 | 2020.04.04. | 네이버뉴스 | [🔗](#)

과거 달았던 **댓글**로 **고소**를 당하고, 기소되는 경우가 발생할 수 있기 때문이다. “과제를 내주고 강의를 끝내면서 ‘이제 꽃길만 걸으시길 바랍니다’라고 말하면 일부는 눈물을 흘리기도 해요. 악플러들도 수사를...”



[“악플 때문에 사람이 죽어?” 지금 쓴 댓글 읽어보세요](#)

SBS | 2019.12.16. | 네이버뉴스 | [🔗](#)

또한 심은진은 “모르는 사람도 눈살을 찌푸릴만한 성희롱성 **댓글**을 한 명에게 3년 동안 받은 게 1000개다”라고... 심은진은 “이분은 기간이 길었고 집요했고 심지어 **고소**를 당했는데도 멈추지 않았다. 재미로 느낀다면 그...”

Part 3 목표

- 딥 러닝 활용 욕설 분류
- 욕설 분류 모델을 적용시킨 웹 페이지 제작

Part 1 데이터

- 데이터 수집 방안 : 크롤링
- 수집대상 : 네이버뉴스 댓글

커뮤니티 댓글(디씨 인사이드, 인벤)

keoy**** >

2020.06.09. 11:32

어째건 저째건 효과가 있다는것은 신뢰가 있기 때문이지.

답글 2

👍 286 💬 16

1stc**** >

2020.06.09. 13:24

댓글을 보아하니 사촌들이 많나? 왜들 배아퍼하나? 농민들 도와주자는 의도로 한거 안보임? 오버하지들 말고 그냥 있는대로만 보셔. 아님 배아픈 잉간들이 직접 백종원이처럼 해보등가. 하지도 못할것들이 입만 살아서 나불나불 거리긴.

답글 2

👍 63 💬 10

Part 1 데이터

- 데이터 형태(id, data, label)

13004	재고상품 구매를	0
13005	형이 정리해줄게.	0
13006	기자가 또라이 아	0
13007	55만원 se 2세대.	0
13008	북한 사회에서도	1
13009	찌질이들이 또 자	1
13010	미통탐은 저놈을	1

- 데이터 수 : 약 2만개
- 라벨 분포: 각각 50%

Part 1 데이터

- 데이터 라벨링을 위한 keyword

	keyword
0	욕설1
1	욕설2
2	욕설3
3	욕설4
4	욕설5
5	욕설6

- Keyword 개수 : 약 1700개
- 수집한 데이터를 keyword를 이용하여 라벨링

Part 1 데이터

- 데이터 전처리 과정

한글과 공백 제외한 문자 제거 → 형태소 단위로 토큰화
→ 불용어 제거 → 단어를 정수 벡터화

Part 2 딥러닝 모델

$x_{\text{신}}$, $x_{\text{발}}$
등의 욕설

모델 학습

예측 : 욕설



문장 분석을 통해
욕설 탐지 학습

1.입력

2. Embedding

Word Embedding

[단어 개수, 임베딩 벡터 크기]

Dropout

3.Convolution

Convolution

[배치 크기, 컨볼루션 값, 필터 크기, 커널 사이즈]

4.Pooling

Max-Polling

[배치 크기, 컨볼루션 값, 필터 크기]

Fully-Connected

[배치 크기, 컨볼루션 값, 필터 크기]

Dropout

5.출력 레이어

Dense Layer

[배치 크기, 예측 값]

Part 2 딥러닝 모델

	Test1	Test2	Test3	Test4	Test5
conv 계층	3	5	7	9	9
polling 계층	3	5	7	9	9
drop out ratio	0.1	0.1	0.2	0.2	0.2
split ratio	0.4	0.2	0.1	0.1	0.1
epoc	10	10	10	10	20
train acc	0.926	0.9257	0.9356	0.935	0.9322
train loss	0.562	0.562	0.5703	0.6403	1.072
test acc	0.967	0.9368	0.9476	0.946	0.9406
test loss	0.471	0.4709	0.4864	0.5208	0.8084

Part 2 딥러닝 모델

웹 연결



모델 저장



웹에서 받은 input데이터를 학습



욕설 탐지한 결과를 return

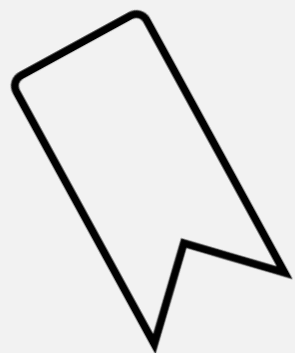
Part 2 딥러닝 모델

라벨링

안녕하세요 즐겁하세요



정상
0



기분 x나 x같네



욕설
1

Part 2 딥러닝 모델



ㅈ랄하네
지랄하네



ㅅ!@#발
ㅅ1발



좇같다
좇같다



좇같다
좇같다

공격적이지만 자연스러운 욕설
변형된 욕설



욕설 탐지

Part 2 딥러닝 모델



한계점

랄



욕설
1

라랄

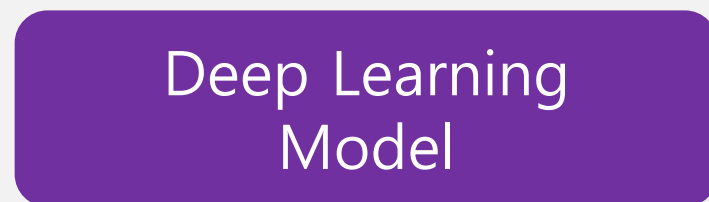
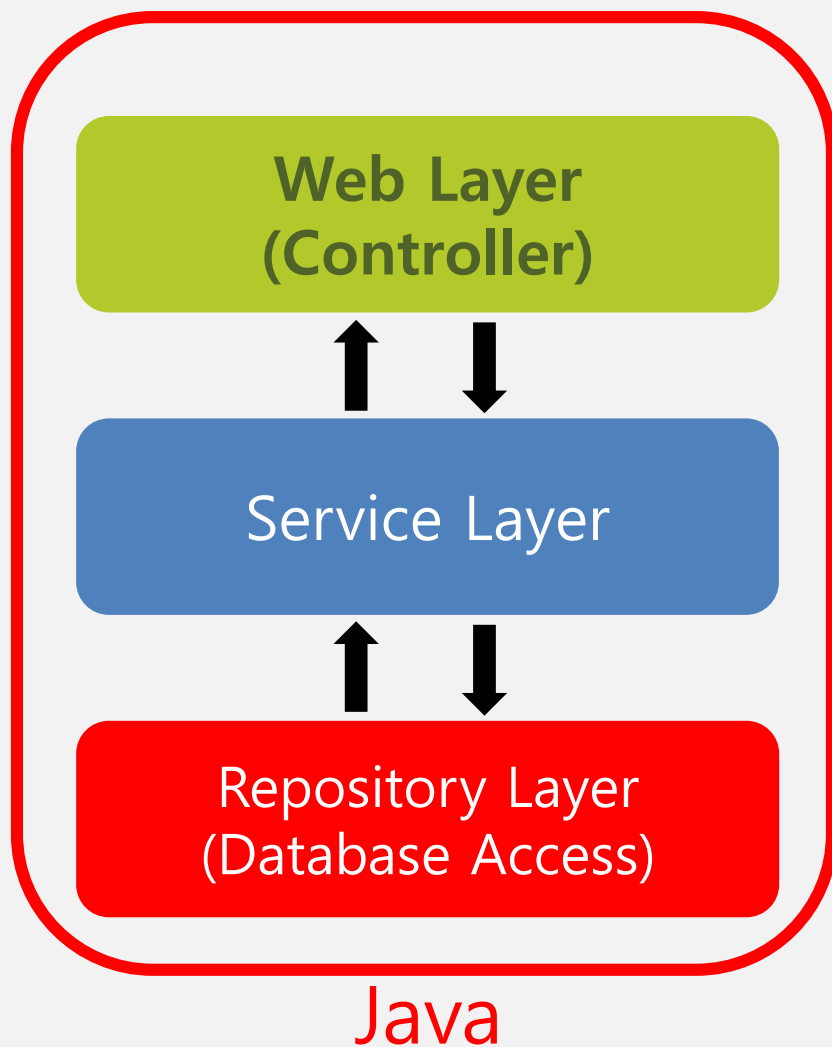


정상
0

Part 1 딥러닝과 웹 서비스

- 딥러닝이 욕설이라 판단한 댓글은 즉시 차단
- 유저 신고를 통해 신조어 학습

Part 1 Spring

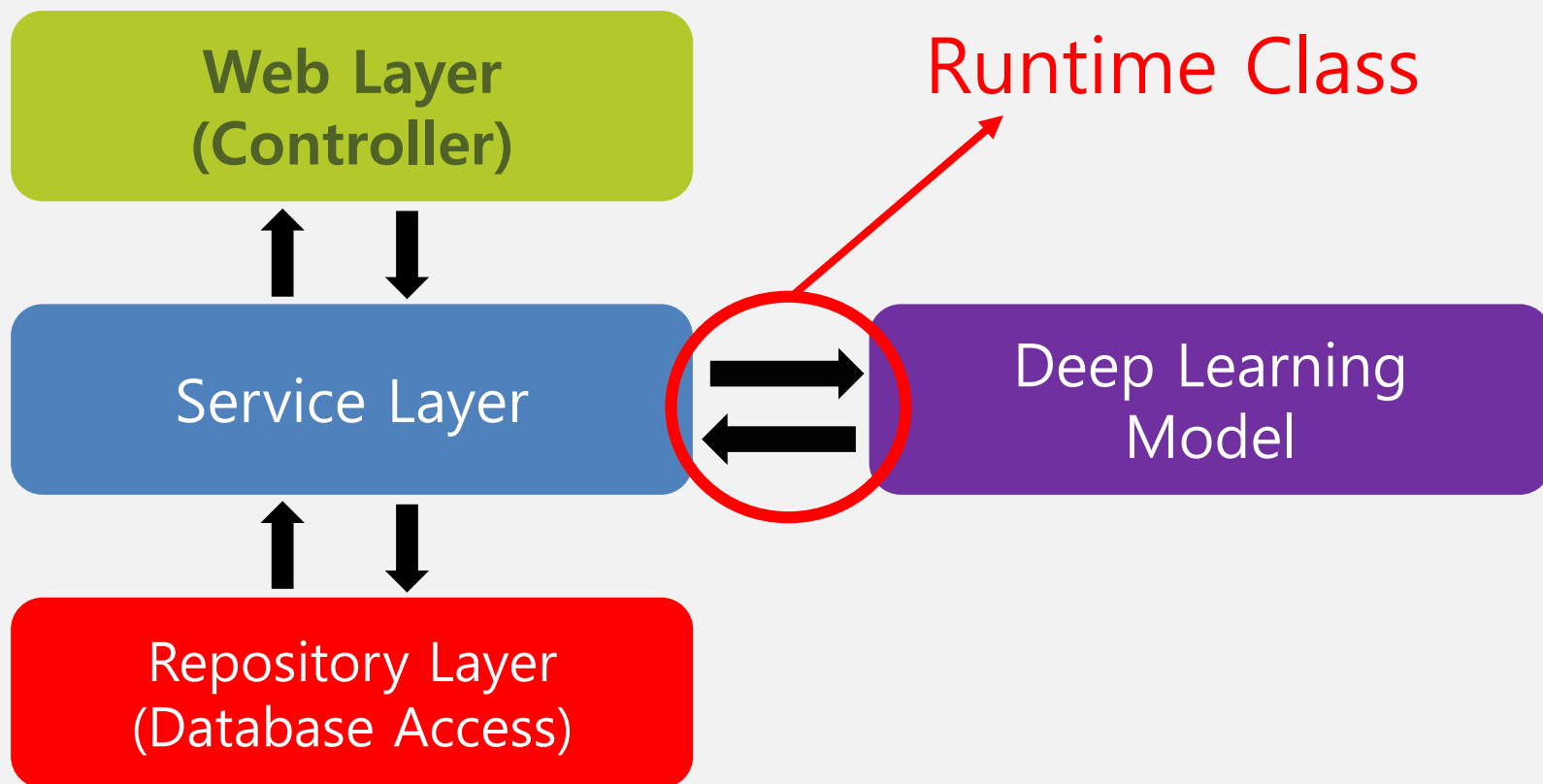


Python

Part 2 Java With Python

- Jython
JVM환경에서 구동되는 Python
- Java RunTime Class
OS 기반의 프로그램 실행

Part 2 Spring



육설 인사이드

관리자 페이지

번호	제목	작성자	날짜
1	"Pawn" 허원석 LCK 서머 분석데스크 합류	너닭은	06-09
2	포킹챔의 모렐vs라바돈vs공허 효율	김히로	06-09
3	타게팅으로 착각하기 쉬운 스킬	엄준식김형섭	06-09
4	펄스건임무 3주차짬사서 지금까지 된거 다했는데	Kimrie	06-09

글쓰기

감사합니다

이후 PPT에는 영상의 캡처본이 있습니다.

Part 2 웹 서비스

작성자	내용	날짜	신고
○○	잭스q 신짜오e	06-09	신고
쇼지군	좋은정보 감사합니다	06-09	신고
○○	알아서 뭐해 A1324	06-09	신고

○○

어쩌라고 [REDACTED] 같은게

취소

등록

Part 2 웹 서비스

작성자	내용	날짜	신고
○○	잭스q 신짜오e	06-09	신고
쇼지군	좋은정보 감사합니다	06-09	신고
○○	알아서 뭐해 A132w	06-09	신고
○○	이제라고 #1월 뽀얀살은제	06-09	신고

Part 2 웹 서비스

작성자	내용	날짜	신고
○○	잭스q 신짜오e	06-09	신고
쇼지군	좋은정보 감사합니다	06-09	신고
○○	알아서 뭐해 +132w	06-09	신고
○○	이제라고 +1할 땀안팔아	06-09	신고

LL

야발 이걸 몰랐네...

취소

등록

Part 2 웹 서비스

작성자	내용	날짜	신고
○○	잭스q 신짜오e	06-09	신고
쇼지군	좋은정보 감사합니다	06-09	신고
○○	알아서 뭐해 A 132w	06-09	신고
○○	이제라고 #1할 땀은얼만큼	06-09	신고
LL	야발 이걸 몰랐네...	06-09	신고

Part 2 웹 서비스

작성자	내용	날짜	신고
○○	잭스q 신짜오e	06-09	신고
쇼지군	좋은정보 감사합니다	06-09	신고
○○	알아서 뭐해 +132w	06-09	신고
○○	이해하고 +1할 땜에...	06-09	신고
LL	야발 이걸 몰랐네...	06-09	신고

어떤 단어를 신고하시겠습니까?

야발

확인

취소

Part 2 웹 서비스

관리자 페이지

댓글	신고 내용	확인	취소
야발 이걸 몰랐네...	야발	확인	취소
			돌아가기

Part 2 웹 서비스

작성자	내용	날짜	신고
○○	잭스q 신짜오e	06-09	신고
쇼지군	좋은정보 감사합니다	06-09	신고
○○	알아서 뭐해 +132w	06-09	신고
○○	어짜고 +1일 뽀글글은게	06-09	신고
LL	야발 이걸 몰랐네..	06-09	신고

BB

야발 이걸 자기만 알고있었네;;;

취소

등록

Part 2 웹 서비스

작성자	내용	날짜	신고
○○	잭스q 신짜오e	06-09	신고
쇼지군	좋은정보 감사합니다	06-09	신고
○○	알아서 뭐해 A 132w	06-09	신고
○○	이제라고 #1발 뽕샐만세	06-09	신고
LL	이발 이글 올랐네...	06-09	신고
ㅂㅂ	이발 이글 차가만 올고있었네...	06-09	신고