

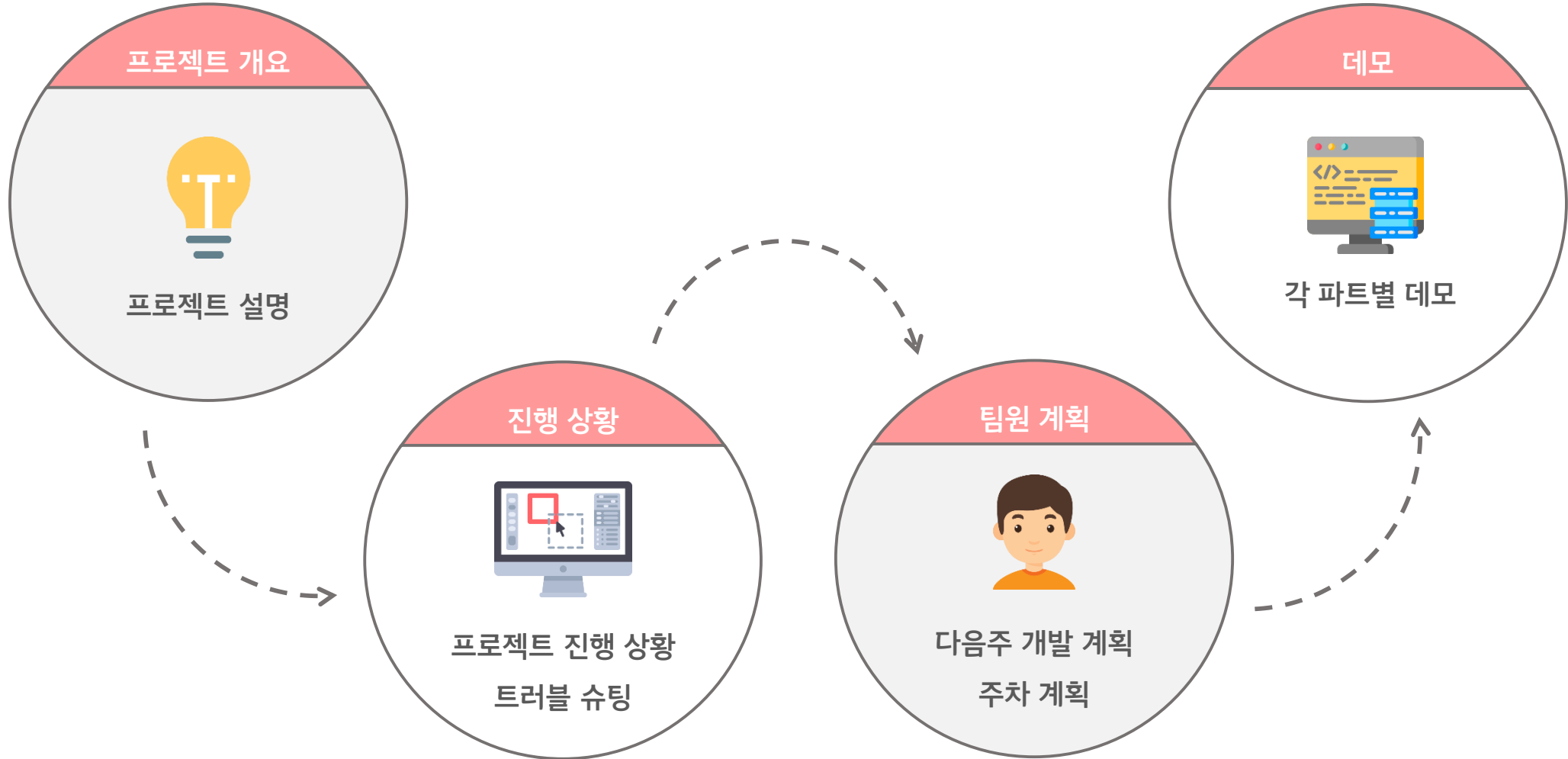


커뮤니티 상 거래를 위한 게시글 제목 변경 웹 확장 프로그램



팀명 : 제뭉
박시현 + 박진영 + 이민희

TABLE OF CONTENTS

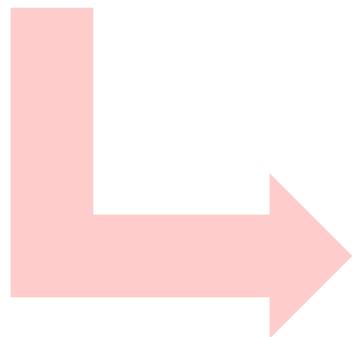


프로젝트 설명

♥ 무료배송 ♥ 워싱 면 라이더 자켓 판매 🧑 🇺🇸

니트, 필립 림 자켓, 양가죽 패딩코트(BCBG, Part or lady, 마이클 코어) 판매 🧑 🇺🇸

** ♠️ ♠️ 셀린느 지고트 레니본 시스템 CK 존스뉴욕 아디다스 모르간 MNG ♠️ ♠️ 4
판매 🧑 [2] 🇺🇸



워싱 면 라이더 자켓 | 15000 | 사용감 유 | 2주

필립 림 자켓 | 30000 | 새상품 | 0회

존스뉴욕 | 43000 | 최상 | 착용만

프로젝트 진행 상황

박시현

서버, DB 환경 설정 (EC2 생성, RDS 생성)
크롤링 문제 해결 및 방향 재설정

박진영

프론트엔드 구현 완료
텍스트 대치 구현 테스트

이민희

데이터 문장 단위 분리 / 키워드 주변 단어 토큰 검색 구현
코스튬 사전 제공 라이브러리 사용 / 초기 모델 구축

서버 환경 설정 인스턴스 생성

인스턴스 | EC2 Management Co x

sa-east-1.console.aws.amazon.com/ec2/v2/home?region=sa-east-1#Instances:sort=instancetype

aws 서비스 리소스 그룹

인스턴스 시작 연결 작업

태그 및 속성별 필터 또는 키워드별 검색

| Name | 인스턴스 ID | 인스턴스 유형 | 가용 영역 | 인스턴스 상태 | 상태 검사 | 경보 상태 | 퍼블릭 DNS(IPv4) | IPv4 퍼블릭 IP | IPv6 IP |
|------|---------|----------|------------|---------|-------------|-------|--------------------------|--------------|---------|
| | | t2.micro | sa-east-1a | running | 2/2개 검사 ... | 없음 | ec2-18-229-82-71.sa-e... | 18.229.82.71 | - |

인스턴스: i-0c809100c4d96b387 탄력적 IP: 18.229.82.71

설명

상태 검사

모니터링

태그

인스턴스 ID

인스턴스 상태

인스턴스 유형

탄력적 IP

퍼블릭 DNS(IPv4)

IPv4 퍼블릭 IP

IPv6 IP

프라이빗 DNS

인스턴스 ID

인스턴스 상태

인스턴스 유형

탄력적 IP

퍼블릭 DNS(IPv4)

IPv4 퍼블릭 IP

IPv6 IP

프라이빗 DNS

의견

한국어

© 2008 - 2019, Amazon Web Services, Inc. 또는 계열사. All rights reserved. 개인 정보 보호 정책 이용 약관

IP 생성

탄력적 IP | EC2 Management Console

sa-east-1.console.aws.amazon.com/ec2/v2/home?region=sa-east-1#Addresses:sort=tag:Name

aws 서비스 리소스 그룹

SiHyunPark 상파울루 지원

예약 인스턴스
전용 호스트
용량 예약

이미지
AMI
번들 작업

ELASTIC BLOCK STORE
볼륨
스냅샷
수명 주기 관리자

네트워크 및 보안
보안 그룹
탄력적 IP
배지 그룹
키 페어
네트워크 인터페이스

로드 밸런싱
로드밸런서
대상 그룹

AUTO SCALING
시작 구성
Auto Scaling 그룹

새 주소 할당 작업

태그 및 속성별 필터 또는 키워드별 검색

| Name | 탄력적 IP | 할당 ID | 인스턴스 | 프라이빗 IP 주소 | 범위 | 연결 ID | 네트워크 인터페이스 ID |
|------|--------------|----------------------------|---------------------|-------------|-----|----------------------------|-----------------------|
| | 18.229.82.71 | eipalloc-0ffea5a317504d... | i-0c809100c4d96b387 | 172.31.4.38 | vpc | eipassoc-03c6a047d328652eb | eni-00de5b138469e5e33 |

주소: 18.229.82.71

설명

태그

탄력적 IP

인스턴스

범위

할당 ID

프라이빗 IP 주소

연결 ID

18.229.82.71

i-0c809100c4d96b387

vpc

eipalloc-0ffea5a317504d...

172.31.4.38

eipassoc-03c6a047d328652eb

의견 한국어

© 2008 - 2019, Amazon Web Services, Inc. 또는 계열사. All rights reserved. 개인 정보 보호 정책 이용 약관

DB 인스턴스 생성

The screenshot displays the AWS RDS console interface for a database instance named 'db-server'. The left sidebar contains navigation links for Dashboard, 데이터베이스 (Database), 성능 개선 도우미 (Performance Improvement Assistant), 스냅샷 (Snapshots), Automated backups, 예약 인스턴스 (Reserved Instances), 서브넷 그룹 (Subnet Groups), 파라미터 그룹 (Parameter Groups), 옵션 그룹 (Option Groups), 이벤트 (Events), 이벤트 구독 (Event Subscriptions), and Recommendations (1). The main content area shows the instance details for 'db-server'.

db-server [수정] [작업 ▼]

| 요약 | | | |
|---------------------|---------------|-----------------------|-----------------------|
| DB 식별자 db-server | CPU 1.50% | 정보 사용 가능 | 클래스 db.t2.micro |
| 역할 인스턴스 | 현재 활동 2 연결 | 엔진 MySQL Community | 리전 및 AZ sa-east-1a |

연결 & 보안 | 모니터링 | 로그 및 이벤트 | 구성 | 유지 관리 및 백업 | 태그

| 연결 & 보안 | | |
|---|--|--|
| 엔드포인트 및 포트 엔드포인트 db-server.czdh3lquhjqr.sa-east-1.rds.amazonaws.com 포트 3306 | 네트워킹 가용 영역 sa-east-1a VPC vpc-8ef4a9e9 서브넷 그룹 | 보안 VPC 보안 그룹 only-database (sg-0c6eab02c16f7f06a) (활성) 퍼블릭 액세스 가능성 예 |

© 2008 - 2019, Amazon Web Services, Inc. 또는 계열사. All rights reserved. 개인 정보 보호 정책 이용 약관

네이버 보안 정책

네이버의 소프트웨어 입력을 통한 입력을 막기 위한
Captcha 시스템



크롤링 시 네이버 로그인 불가



Selenium의 키 조합을 통해 captcha 탐지 우회

프론트엔드 구현 상황
데모 예정

제품

상품명 x

가격 x

상품명

브랜드

가격

상태

색상

거래방식

착용횟수

사이즈

IT/전자기기

용량

사양

사용기간

변경

jQuery 사용

드래그 앤 드롭 구현 시
처음 사용한 jQuery 구문에 많은 시간 소요



언어의 기초부터 공부하는 것의 중요함

Customized KoNLPy

```
def custom_konlpy():  
    twitter.add_dictionary('미착용', 'Noun')  
    twitter.add_dictionary('택포', 'Noun')
```

Input argument / result

제공할 키워드 목록을
String으로 받음

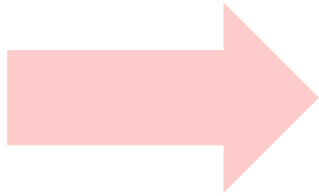
크롤링 한 정보들
각 문장당 list로 받음

```
input_args = ['color', 'size', ['아디다스 블랙 색상의 모직코트입니다 정가  
print(selected_keyword(input_args[:len(input_args)-1], input_args[len
```

목록에 맞는 키워드 반환 함수를
호출해 합쳐 키워드 리스트 반환

Sentence slice / tokenize

‘ 아디다스 레깅스입니다 택도 안 땀 새상품이에요 ’

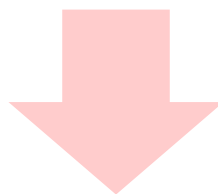


‘ 아디다스 레깅스입니다
택도 안 땀 새상품이에요 ’

명사 형용사 숫자 알파벳 동사 O
조사 부사 X

kkma 라이브러리 속도

문장 분리 함수를 제공하는 kkma 사용
하지만 kkma는 KoNLPy 라이브러리 들 중 가장 느린 속도



문장 분리 시 정확도와 총 걸린 시간을 비교하고 선택

사전과 일치하는 키워드

```
def getBrand(string_list):  
    #브랜드 반환  
    searchList = ['구찌', '나이키', '아디다스']  
  
    for searchToken in searchList:
```

```
def getMethod(string_list):  
    #거래방식 반환  
    flag = 0  
  
    searchList1 = ['택배', '택비', '택 포', '배송', '운송', '운포', '착불']  
    searchList2 = ['직거래']
```

Re Library

```
def getPrice(string_list):
    #가격 반환
    find_keyword = re.findall("\d*십*일*이*삼*사*오*육*칠*팔*구*십")

    if len(find_keyword) == 1 :
        find_keyword = find_keyword[0]
    elif len(find_keyword) > 1:
        find_keyword = '여러개야'
    else:
        find_keyword = '?'
```

```
def getColor(string_list):
    #색상 반환
    find_keyword = re.findall("\w+ ?색", "".join(string_list))
```


키워드 주제 단어 근처 등장 키워드

```
def getSize(docs_list):  
    #사이즈 반환  
    searchList = ["사이즈", "size"]  
    tokenList = []  
  
    tokenList = getAroundToken(docs_list, searchList)  
  
    for t in tokenList:  
        if t[1] == 'Alpha' or t[1] == 'Number':  
            token = t[0]  
            break  
  
    return token
```

주제 단어 근처의 토큰을 반환
그 후 특징으로 추출

사이즈 : FREE

90size 게스 후드티 정리합니다.
오염없고 상태좋아요-!

초기 모델 구조

```
from collections import namedtuple
from gensim.models import doc2vec
from gensim.models import word2vec
import gensim

num_features = 300      # Word vector dimensionality
min_word_count = 10     # Minimum word count
num_workers = 2         # Number of threads to run in parallel
context = 4             # Context window size
downsampling = 1e-3     # Downsample setting for frequent words

model = gensim.models.Word2Vec(train_data, workers=num_workers,
                               size=num_features, min_count = min_word_count,
                               window = context, sample = downsampling)
```

다음주 개발 계획

박시현

새로운 방법으로 크롤링 재진행 & 데이터 수집
통신 모듈 학습 및 서버 학습

박진영

테스트 데이터로 텍스트 대치 구현 완료
서버와의 통신 구현 시작

이민희

모델 파라미터 변경
문장 분리 용 / pos tag 분리 용 / 키워드 추출 용 사전 추가

팀원 별 각 주간 개발 계획

| 박시현 | 9주차 | 10주차 | 11주차 | 12주차 | 13주차 | 14주차 | 15주차 |
|-----------------|-----|------|------|------|------|----------|-----------------|
| 크롤링 · 전처리 | | | | | | 최종 데모 | 최종 리포트 제출 |
| 데이터셋 구축 | | | | | | | |
| 서버 구축 | | | | | | | |
| 통신 모듈 작업 | | | | | | | |
| 서버 통신 구현 | | | | | | | |
| 데모 준비 · 서비스 테스트 | | | | | | | |

팀원 별 각 주간 개발 계획

| 박진영 | 9주차 | 10주차 | 11주차 | 12주차 | 13주차 | 14주차 | 15주차 |
|-------------|-----|------|------|------|------|----------|-----------------|
| 프론트엔드 화면 구현 | | | | | | 최종 데모 | 최종 리포트 제출 |
| 프론트엔드 동작 구현 | | | | | | | |
| 텍스트 대치 구현 | | | | | | | |
| 확장 앱 통신 구현 | | | | | | | |
| 웹 페이지 쿠키 구현 | | | | | | | |
| 테스트 | | | | | | | |

팀원 별 각 주간 개발 계획

| 이민희 | 9주차 | 10주차 | 11주차 | 12주차 | 13주차 | 14주차 | 15주차 |
|----------------|-----|------|------|------|------|----------|-----------------|
| 문장 분리 | | | | | | 최종 데모 | 최종 리포트 제출 |
| 키워드 주변 토큰 처리 | | | | | | | |
| 키워드 사전 변경 | | | | | | | |
| 모델 구축, 파라미터 변경 | | | | | | | |
| 서버 모듈 연결 | | | | | | | |
| 성능 및 정확도 향상 | | | | | | | |

전체 각 주간 개발 계획

| | 9주차 | 10주차 | 11주차 | 12주차 | 13주차 | 14주차 | 15주차 |
|--------------|-----|------|------|------|------|----------|-----------------|
| 데이터 확보 · 전처리 | | | | | | 최종 데모 | 최종 리포트 제출 |
| 프론트 엔드 | | | | | | | |
| 서버 구축 | | | | | | | |
| 자연어 처리 알고리즘 | | | | | | | |
| 모듈 연결 | | | | | | | |
| 유닛 · 전체 테스트 | | | | | | | |



- 중간 데모 -

