

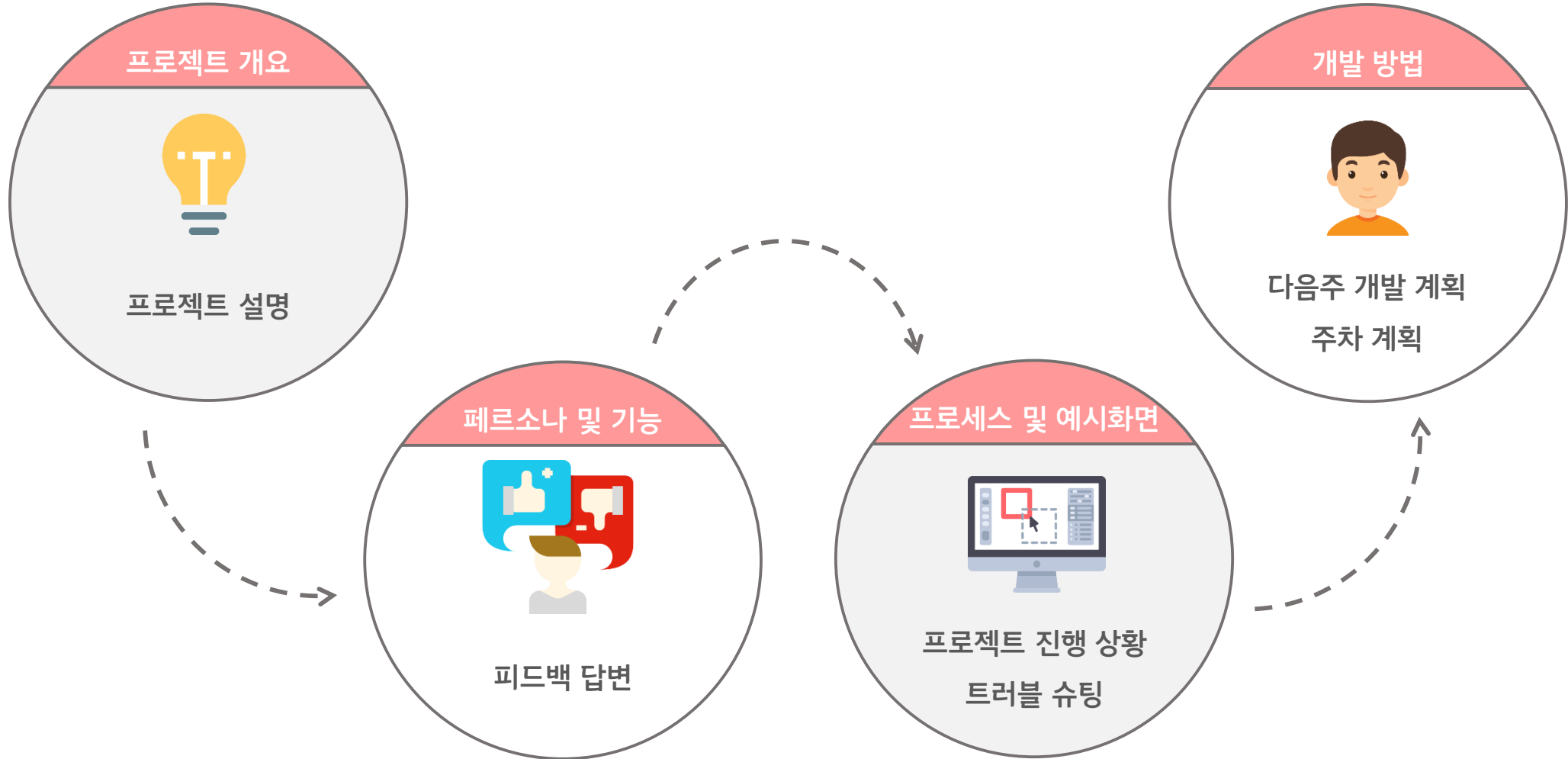


# 커뮤니티 상 거래를 위한 게시글 제목 변경 웹 확장 프로그램



팀명 : 제뭉  
박시현 + 박진영 + 이민희

# TABLE OF CONTENTS

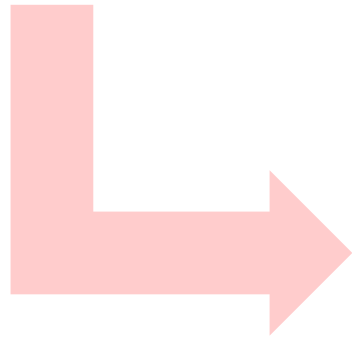


## 프로젝트 설명

♥ 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

우분투 서버와 크롤링 프로그램 연결 설정  
통신 모듈 구현

박진영

웹 스토리지 구현  
서버 통신 구현 시작

이민희

서버 모듈 연결 / 모델 파라미터 변경  
상세 코드 구현



## 현재 확보한 데이터셋

 KT.txt

 LGU.txt

 SKT.txt

 노트북.txt

 여성상의.txt

 여성신발.txt

 여성하의.txt

 여성하의2.txt

 태블릿PC.txt

 남성상의.txt

 남성상의2.txt

 남성신발.txt

 남성하의.txt

 남성하의2.txt

## 데이터 수집에 많은 시간 소요

10000개 데이터 수집 소요 시간 : 평균 4-5시간

일부 게시글 접근 불가 및 크롤링 불가능

게시물 삭제, 회원 탈퇴, 로딩시간 오류 등으로 추측



접근 불가능한 게시물을 건너 뛰고 크롤링 재진행

## 텍스트 대치 구현의 어려움

chrome.storage.local API



Promise, async, await 함수



window.localStorage  
( localStorage.setItem  
localStorage.getItem 함수 )



# 실행 화면

DevTools - cafe.naver.com/joongqonara?iframe\_url=/ArticleSearchList.nhn%3Fsearch.clubid=10050146%26search.searchdate=all%26...

Key	Value
54912086	도라ops 빨간원피스 인더유니 원피스
546926413	TASSE TASSE 빨간원피스 (백화점 브랜드)
542716194	[새상품] h&m 레드원피스 빨간원피스 에이치앤엠
526446408	다홍원피스, 빨간원피스, 레드원피스
523635707	유이니 빨간 원피스(신상- 새거) 현재 사이트 판매중.
520970984	W컨셉 lookast 빨간 원피스 팝니다. (새 상품)
518993083	보세 빨간원피스
516581663	빨간 원피스 만원에 팔아요
515476511	소개팅에 성공했던 빨간원피스 팝니다 ㅎㅎ
514007547	빨간 원피스
506060821	보세 데일리 빨간 원피스팔아요
506055089	보세 빨간 원피스 팔아요
504248041	Banana republic 빨간 원피스
501790859	뷔스티에 빨간 원피스 운포8천원
591452350	McGINN 빨간 원피스 팝니다
Is	NIYXEP37D2FF2
654912086_changed	제목a
646926413_changed	제목b
642716194_changed	제목c
626446408_changed	제목b
623635707_changed	제목e
620970984_changed	제목f
618993083_changed	제목g
616581663_changed	제목h
615476511_changed	제목i
614007547_changed	제목j
606060821_changed	제목k
606055089_changed	제목l
604248041_changed	제목m
601790859_changed	제목n
591452350_changed	제목o

654912086

도라ops 빨간원피스 인더유니 원피스

646926413

TASSE TASSE 빨간원피스 (백화점 브랜드)

642716194

[새상품] h&amp;m 레드원피스 빨간원피스 에이치앤엠

626446408

다홍원피스, 빨간원피스, 레드원피스

623635707

유이니 빨간 원피스(신상- 새거) 현재 사이트 판매중.

620970984

W컨셉 lookast 빨간 원피스 팝니다. (새 상품)

618993083

보세 빨간원피스

616581663

빨간 원피스 만원에 팔아요

615476511

소개팅에 성공했던 빨간원피스 팝니다 ㅎㅎ

614007547

빨간 원피스

606060821

보세 데일리 빨간 원피스팔아요

606055089

보세 빨간 원피스 팔아요

604248041

Banana republic 빨간 원피스

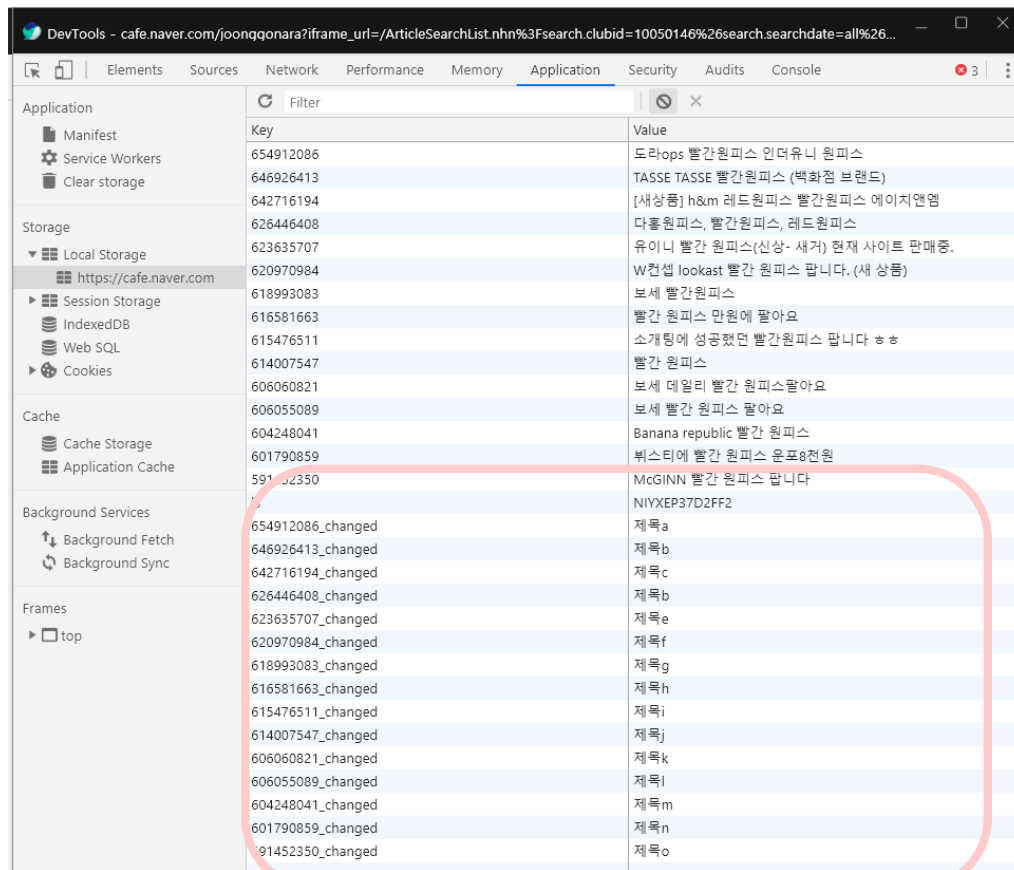
601790859

뷔스티에 빨간 원피스 운포8천원

591452350

McGINN 빨간 원피스 팝니다

# 실행 화면



Key	Value
654912086	도라ops 빨간원피스 인더유니 원피스
646926413	TASSE TASSE 빨간원피스 (백화점 브랜드)
642716194	[새상품] h&m 레드원피스 빨간원피스 에이치앤엠
626446408	다홍원피스, 빨간원피스, 레드원피스
623635707	유이니 빨간 원피스(신상- 새거) 현재 사이트 판매중.
620970984	W컨셉 lookast 빨간 원피스 팝니다. (새 상품)
618993083	보세 빨간원피스
616581663	빨간 원피스 만원에 팔아요
615476511	소개팅에 성공했던 빨간원피스 팝니다 ㅎㅎ
614007547	빨간 원피스
606060821	보세 데일리 빨간 원피스팔아요
606055089	보세 빨간 원피스 팔아요
604248041	Banana republic 빨간 원피스
601790859	뷔스티에 빨간 원피스 운포8천원
591452350	McGINN 빨간 원피스 팝니다
654912086_changed	제목a
646926413_changed	제목b
642716194_changed	제목c
626446408_changed	제목b
623635707_changed	제목e
620970984_changed	제목f
618993083_changed	제목g
616581663_changed	제목h
615476511_changed	제목i
614007547_changed	제목j
606060821_changed	제목k
606055089_changed	제목l
604248041_changed	제목m
601790859_changed	제목n
91452350_changed	제목o

654912086_changed	제목a
646926413_changed	제목b
642716194_changed	제목c
626446408_changed	제목b
623635707_changed	제목e
620970984_changed	제목f
618993083_changed	제목g
616581663_changed	제목h
615476511_changed	제목i
614007547_changed	제목j
606060821_changed	제목k
606055089_changed	제목l
604248041_changed	제목m
601790859_changed	제목n
591452350_changed	제목o

## 실행 화면

**카페정보** 나의활동

메니저 **qdillion**  
since 2003.12.10.  
카페소개

17,928,622  
★ 즐겨찾는 멤버 3,684,337명  
계시판 구독수 2,046,033회  
우리카페업 수 140,280회

2017년 대표 인기카페

카페 글쓰기

카페 채팅

검색

★ 즐겨찾는 게시판

전체글보기 142,303,276

중고나라★새소식

중고나라★이벤트

당첨자발표/종료

공지사항

이용정책

중고나라 이야기

물거래 후기(칭찬)

어려운 이웃 사연

중고나라 행복나눔

제품 무료나눔

마음껏 기부나눔

전체기간 여성

판매중

제목

사용기간 ×

공통

상품명 브랜드 가격 상태

색상 거래방식

의류

착용횟수 사이즈

IT/전자기기

용량 사양 사용기간

변경 전

최신순 제목만 보기 15개씩

제목	작성자	작성일	조회
도라ops 빨간원피스 인더유니 원피스 [판매]	자입지	2019.10.19.	18
TASSE TASSE 빨간원피스 (백화점 브랜드) [판매]	positiveart	2019.10.01.	11
[새상품] h&m 레드원피스 빨간원피스 에이치앤엠 [판매]	한옥부부	2019.09.21.	36
다홍원피스, 빨간원피스, 레드원피스 [판매]	SONGEI0901	2019.08.11.	19
유이니 빨간 원피스(신상- 새거) 현재 사이트 판매중. [판매]	궁디애기	2019.08.04.	55
W컨셉 lookast 빨간 원피스 팝니다. (새 상품) [판매] 안전	jinhee0448	2019.07.28.	86
새 빨간원피스 [판매]	갓요정	2019.07.23.	18
빨간 원피스 만원에 팔아요 [판매]	울날날	2019.07.17.	27
소개팅에 성공했던 빨간원피스 팝니다 ㅎㅎ [판매]	슈링스탈샤링스탈	2019.07.14.	37
빨간 원피스 [판매]	love is rich	2019.07.10.	18
보세 데일리 빨간 원피스팔아요 [판매] 안전	p37625	2019.06.20.	22
보세 빨간 원피스 팔아요 [판매] 안전	p37625	2019.06.20.	12
Banana republic 빨간 원피스 [판매]	아자화팅ㅋ	2019.06.16.	43
뷔스티에 빨간 원피스 운포8천원 [판매]	씨니아1	2019.06.09.	15
McGINN 빨간 원피스 팝니다 [판매] 안전	DJ DDong	2019.05.13.	24

## NER 구조 설명

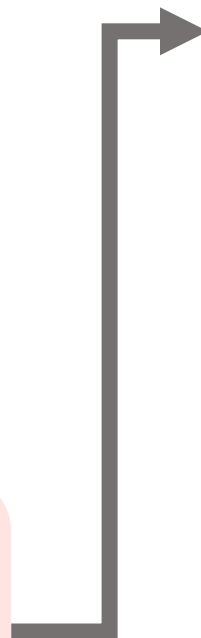
tokenizing



일정 빈도 이상 나오는  
단어 사전 구축



text CNN (1 X n window)  
가운데 단어를 기준으로 앞 뒤 단어의  
co occurrence 계산



Word2vec 모델 Most\_similar 사용해  
개체 인식을 하고 싶은 종류의  
seed\_word 생성



Seed word 는 1  
아닌 단어는 0으로 라벨링



LogisticRegression을 통해  
Window classifier

# NER 튜닝

tokenizing

Word2vec 모델 Most\_similar 사용해  
개체 인식을 하고 싶은 종류의  
단어 생성

일정 빈도 이상 나  
단어 사전 구축

1. Twitter -> Mecab

주요 키워드가 명사일 것이라 판단하고,  
세세한 분석보단 속도가 빠른 Twitter 사용

1  
라벨링

-> NER 은 세세한 형태소 분석이  
필수라는 것을 조언으로 들음

text CNN (1 X n wi  
가운데 단어를 기준으로 앞  
co occurrence 계산

을 통해  
Window classifier

## NER 튜닝

4. 모델, seed\_word 수 변경

Word2vec 모델 Most\_similar 사용해  
개체 인식을 하고 싶은 종류의  
seed\_word 생성

일정 빈도 이상 나오는  
단어 사전 구축

2. 빈도 수 변경

Seed  
아닌 단어

5. 기준 값, 사전에  
추가 할 단어 수 변경

text CNN (1 X n window)  
가운데 단어를 기준으로 앞 뒤 단어의

3. Window 크기 변경

LogisticRegression을 통해  
Window classifier

## NER 튜닝

tokenizing



일정 비도 이상 나오느



Word2vec 모델 Most\_similar 사용해  
개체 인식을 하고 싶은 종류의  
seed\_word 생성



Seed word 는 1  
인 단어는 0으로 라벨링



LogisticRegression을 통해  
Window classifier

6. seed\_word 후처리  
추출된 사전 경향을 보고  
원치 않는 seed가 있을 경우  
negative seed를 구해서  
Seed에서 제외

text C  
가운데 단어  
co

## 이름, 브랜드 사전 처리

특징  
(재질/무늬)



옷  
카테고리

브랜드

이름

브랜드

Ex) 가죽 자켓  
기모 후드티  
과배기 니트  
스트라이프 셔츠

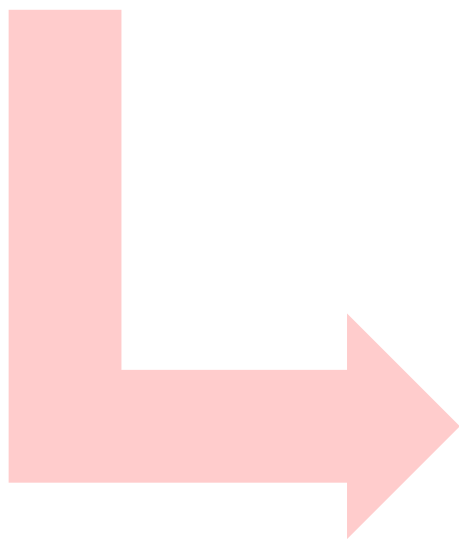
Ex) 나이키  
데상트  
파타고니아



## 정확도 문제

클래식 블루종 모직 자켓

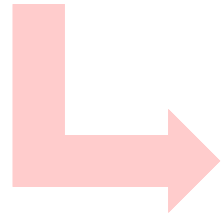
명품답게 따뜻한 모직과 화려한 체크 안감이 돋보이는 쉼터



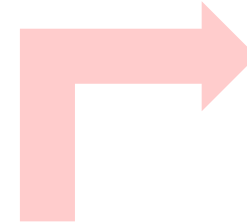
```
[ '티기모 사이즈', '?', '택배',  
[ '패딩 코트', '?', '택배', 5  
[ ' 사이즈', '?', '?', 110000  
[ '티후드 사이즈', '?', '택배',  
[ '모직체크 자켓', '?', '택배'  
(env)
```

## Negative seed\_word

| 구요 김수현 코트로 유명했었네요 95입  
| 싱글 롱코트 95 / 김수현 니트



( '캐시미어 /Noun', 0.354702353477478),  
( '울 /Noun', 0.3544993996620178),  
( '천연 /Noun', 0.3522285223007202),  
( '세련되다 /Adjective', 0.34476912021636963),  
( '소가 /Noun', 0.3404116928577423),  
( '김수현 /Noun', 0.33747971057891846),  
( '아콘 /Noun', 0.3374019265174866),  
( '털 /Noun', 0.3336946666240692),  
( '캐시미어 /Noun', 0.3317150000000000)



김우빈 /Noun  
해비 /Noun  
솔라 경량패딩 /Noun  
커클랜드 /Noun  
김영주 /Noun  
구스패딩 /Noun  
에스크 /Noun

## Negative seed\_word

#seed\_words 타입은 set이야 집합!!

```
pseed_words = {word for word, _ in word2vec.wv.most_similar('나이키/Noun', topn=50)}  
pseed_words.update({word for word, _ in word2vec.wv.most_similar('데상트/Noun', topn=50)})  
pseed_words.update({word for word, _ in word2vec.wv.most_similar('파타고니아/Noun', topn=50)})
```

```
nseed_words = {word for word, _ in word2vec.wv.most_similar('김수현/Noun', topn=20)}  
nseed_words.update({word for word, _ in word2vec.wv.most_similar('한정판/Noun', topn=20)})  
nseed_words.update({word for word, _ in word2vec.wv.most_similar('농구/Noun', topn=20)})  
nseed_words.update({word for word, _ in word2vec.wv.most_similar(' ', topn=20)})  
#합집합을 만들어 주는 느낌 입니다
```

```
seed_words = pseed_words - nseed_words
```

# Negative seed\_word

```
word2vec.wv.most_similar('김수현/Noun', topn=20): [('스타디움/Noun', 0.65822434425354)
```

```
> 00: ('스타디움/Noun', 0.65822434425354)
> 01: ('항공/Noun', 0.6506903767585754)
> 02: ('마테호른/Noun', 0.6207840442657471)
> 03: ('체스터필드/Noun', 0.6093805432319641)
> 04: ('스위브/Noun', 0.5974478721618652)
> 05: ('아우토반/Noun', 0.5919034481048584)
> 06: ('너클/Noun', 0.582004189491272)
> 07: ('몬스터/Noun', 0.5797088146209717)
> 08: ('젠틀/Noun', 0.5777378082275391)
> 09: ('탕/Noun', 0.5726131796836853)
> 10: ('라반/Noun', 0.5712717771530151)
> 11: ('타루/Noun', 0.5707287192344666)
> 12: ('꿍다/Verb', 0.5639631748199463)
> 13: ('잠바/Noun', 0.5606921911239624)
> 14: ('워모/Noun', 0.5523853898048401)
> 15: ('바치/Noun', 0.5511715412139893)
> 16: ('알랭/Noun', 0.5455688238143921)
> 17: ('레나/Noun', 0.5450652837753296)
```

```
> 02: ('레어/Noun', 0.5501296520233154)
> 03: ('미드/Noun', 0.48607710003852844)
> 04: ('르브론/Noun', 0.48117443919181824)
> 05: ('신상/Noun', 0.4805793762207031)
> 06: ('트레일/Noun', 0.47512999176979065)
> 07: ('트랙/Noun', 0.46846067905426025)
> 08: ('레귤러/Noun', 0.4649038314819336)
> 09: ('콜라보/Noun', 0.4570951759815216)
> 10: ('된장/Noun', 0.45484811067581177)
> 11: ('후디/Noun', 0.45190557837486267)
> 12: ('에어포스/Noun', 0.4406619668006897)
> 13: ('잉글랜드/Noun', 0.4318622946739197)
> 14: ('카모/Noun', 0.4300890266895294)
> 15: ('복가ㄱ/Noun', 0.42952263355255127)
> 16: ('아노락/Noun', 0.42876505851745605)
> 17: ('뱀피/Noun', 0.4254387617111206)
> 18: ('볼트/Noun', 0.42372432351112366)
> 19: ('허라취/Noun', 0.41345322132110596)
__len__: 20
```

## 한계점

거/미둠/로버트/타테오시안/클립온  
에파라아여/돌체가바나/샌드피퍼/  
른/보스턴백/여행가방/옷장정리/했

word2vec.wv.most\_similar('바나나/Noun', topn=20): [('야크/Noun', 0.576083242893219)

- > 00: ('야크/Noun', 0.576083242893219)
- > 01: ('얼반/Noun', 0.5752901434898376)
- > 02: ('엠폴햄/Noun', 0.5279896259307861)
- > 03: ('트리베카/Noun', 0.5134754180908203)
- > 04: ('지오다노/Noun', 0.5103771686553955)
- > 05: ('헤드/Noun', 0.4998543858528137)
- > 06: ('해드/Noun', 0.4951708912849426)
- > 07: ('캣반팔/Noun', 0.4883071184158325)
- > 08: ('테일러/Noun', 0.45113077759742737)
- > 09: ('파로스/Noun', 0.4449973702430725)
- > 10: ('머슬/Noun', 0.4445072114467621)
- > 11: ('크로커다일/Noun', 0.44205576181411743)
- > 12: ('스퀘어/Noun', 0.43898287415504456)
- > 13: ('사인/Noun', 0.4231773018836975)
- > 14: ('일포르소/Noun', 0.42303532361984253)

### 바나나 브랜드 옷

전체

이미지

뉴스

동영상

지도

더보기

설정

도구

검색결과 약 11,200,000개 (0.48초)

### 바나나 브랜드 옷 관련 이미지



→ 바나나 브랜드 옷에 대한



바나나리퍼블릭 L 셔츠  
에스피오나지 덱자켓

## 브랜드 다양성

```
#seed_words 타입은 set이야 집합
pseed_words = {word for word, _ in word2vec.wv.most_similar('택배', topn=50)}
pseed_words.update({word for word, _ in word2vec.wv.most_similar('패딩 코트', topn=50)})
pseed_words.update({word for word, _ in word2vec.wv.most_similar('사이즈', topn=10000)})
pseed_words.update({word for word, _ in word2vec.wv.most_similar('티후드 사이즈', topn=50)})
pseed_words.update({word for word, _ in word2vec.wv.most_similar('모직체크 자켓', topn=50)})
pseed_words.update({word for word, _ in word2vec.wv.most_similar('데상트/Noun', topn=50)})
pseed_words.update({word for word, _ in word2vec.wv.most_similar('파타고니아/Noun', topn=50)})

nseed_words = {word for word, _ in word2vec.wv.most_similar('김수현/Noun', topn=20)}
nseed_words.update({word for word, _ in word2vec.wv.most_similar('한정판/Noun', topn=20)})
nseed_words.update({word for word, _ in word2vec.wv.most_similar('농구/Noun', topn=20)})
nseed_words.update({word for word, _ in word2vec.wv.most_similar('호떡/Noun', topn=20)})
#합집합을 만들어 주는 느낌 입니다

seed_words = pseed_words - nseed_words
```

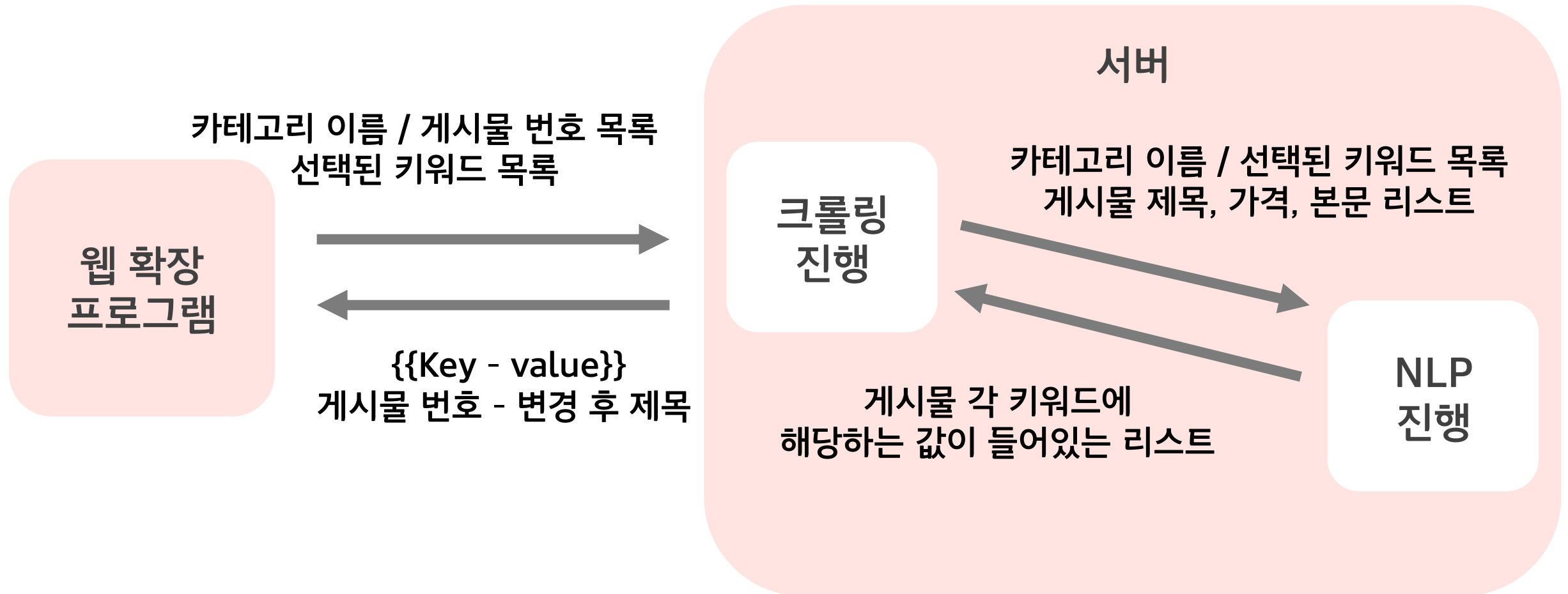
## 토큰 간 거리 비교

```
string_list: '디키즈 기모 후드티 사이즈 90 정가 1만 택포 8천 입니다'
> t: ('8천', 'Number')
> tempTokenList: [('1만', 'Number'), ('8천', 'Number')]
> tokenList: [('1만', 'Number'), ('8천', 'Number')]
```

```
find_keyword = tokenList[0][0]
elif len(tokenList) > 1:
    find_n_distance = getMeanDistance(docs_list, tokenList, ['정가', '구입'])
    #np 값이 가장 큰 값인 키워드 토큰 반환
    find_keyword = tokenList[find_n_distance.index(max(find_n_distance))][0]
else:
```

```
mean_dis_list: [1, 3]
```

## 전체적인 통신 흐름





## 다음주 개발 계획

박시현

서버 통신 구현  
통신 모듈 구현

박진영

서버와의 통신 구현  
변경 후 제목 마우스 오버 툴팁 기능 추가 작업

이민희

서버 모듈 연결 / 모델 파라미터 변경  
성능 및 정확도 향상

## 팀원 별 각 주간 개발 계획

박시현	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
주제 확정/제안서 작성								중간 고사 & 중간 데모 준비						최종 데모	최종 리포 트 제출
크롤링 · 전처리															
데이터셋 구축															
서버 구축															
통신 모듈 작업															
서버 통신 구현															
데모 준비 · 서비스 테스트															

## 팀원 별 각 주간 개발 계획

박진영	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
주제 확정/제안서작성								중간 고사 & 중간 데모 준비						최종 데모	최종 리포 트 제출
프론트엔드 화면 구현															
프론트엔드 동작 구현															
텍스트 대치 구현															
확장 앱 통신 구현															
웹 페이지 쿠키 구현															
테스트															

## 팀원 별 각 주간 개발 계획

이민희	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
주제 확정/제안서작성								중간 고사 & 중간 데모 준비						최종 데모	최종 리포트 제출
자연어 처리 공부															
라이브러리 사용 전처리															
알고리즘 구체화															
문장분리															
주변 토큰 처리															
키워드 사전 변경															
모델 구축, 파라미터 변경															
서버 모듈 연결															
성능 및 정확도 향상															

# 전체 각 주간 개발 계획

공통: 연두색 / 박시현:핑크색 / 박진영:하늘색 / 이민희:주황색

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
주제확정/제안서 작성								중간고사 & 중간데모 준비						최종데모	최종리포트 제출
크롤링/데이터 확보															
데이터 전처리															
프론트 엔드															
서버 구축															
텍스트 대치 구현															
자연어 처리 모델 구축															
웹 페이지 쿠키 구현															
모듈 연결															
유닛 · 전체 테스트															



감사합니다! QnA

