

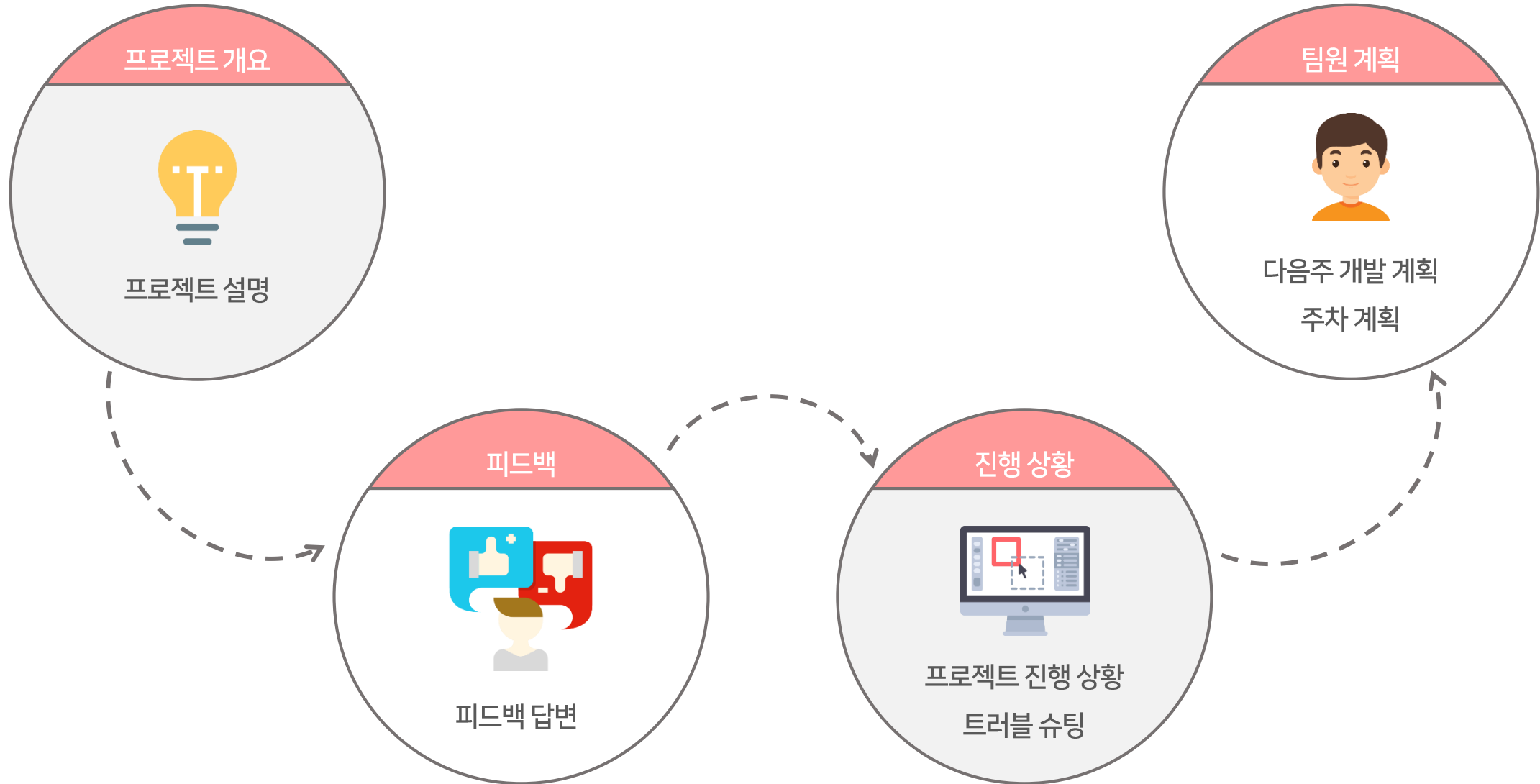


# 커뮤니티 상 거래를 위한 게시글 제목 변경 웹 확장 프로그램



팀명 : 제뭇  
박시현 + 박진영 + 이민희

# TABLE OF CONTENTS

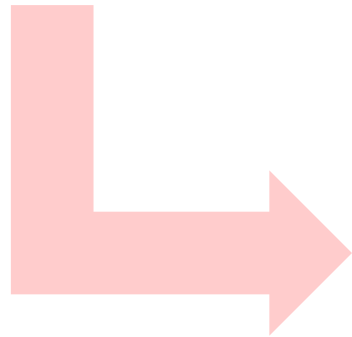


## 프로젝트 설명

♥ 무료배송 ♥ 워싱 면 라이더 자켓 판매 🧑🏻 N

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

\*\* ♠️ ♠️ 셀린느 지고트 레니본 시스템 CK 존스뉴욕 아디다스 모르간 MNG ♠️ ♠️ 4  
판매 🧑🏻 [2] N



워싱 면 라이더 자켓 | 15000 | 사용감 유 | 2주

필립 림 자켓 | 30000 | 새상품 | 0회

존스뉴욕 | 43000 | 최상 | 착용만

## 저번 주 피드백

---

1. 키워드 선택 시 같은 카테고리 안에서만 선택되게

➡ 반영 완료

2. 자연어 처리 알고리즘 및 모델 설정

➡ 뒷 부분에서 설명

## 프로젝트 진행 상황

박시현

6개의 카테고리에 대해 각 1만개씩 데이터셋 구축  
서버에 데이터베이스적용 연습

박진영

피드백 반영 프론트엔드 수정  
텍스트 대치 관련 조사 및 구현

이민희

모델 사용해서 유사도 가장 높은 토큰 반환  
사전에 존재하지 않는 경우 처리

# 모델 학습 위한 데이터셋 확보

여성상의.txt - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말

[55-66, 세일]핑크 케이프 울 캐시미어 코트 만원  
넘 귀여운 핑크색 케이프 코트예요!케이프 분리 가능합니다!! 사랑스러워용울과  
[55-66, 세일]핑크 케이프 울 캐시미어 코트 만원  
발렌키 트레이닝복 여성 집업 상의 새상품  
발렌키 트레이닝복 여성상의 집업입니다90사이즈 입니다 요즘 입기 딱 좋아요 옷감도 좋고 길어도 너무 짧지 않고 절개선이 만  
발렌키 트레이닝복 여성 집업 상의 새상품  
뉴발란스 윈터트랙 폴라텍 플리스자켓  
작년 구매 사이트 품질제품입니다1회 착용 가볍고 정말 따뜻해요최저가 반가격에 올려요 010-7사사일-6339  
뉴발란스 윈터트랙 폴라텍 플리스자켓  
임부복 모직 원피스  
모100 임부복 원피스입니다.두세번 입어서 상태 매우 좋아요.재질이 모라서 고급스러워요.01071팔오칠393 으로 문자주세요.  
임부복 모직 원피스

한 게시물 -> 제목/본문/제목  
순서로 데이터셋 확보

## 모델 학습 위한 데이터셋 확보

[여성상의, 여성하의, 여성신발, 남성상의, 남성하의, 남성신발]

6개 카테고리 별로 각각 10000개씩

한 페이지에서 랜덤하게 10개 선택 -> 1000페이지

## 게시물 다양성 부족

최대 목록 페이지 = 1000페이지

1000번째 페이지 게시물들의 작성일은 2019.11.01 (여성 상의 게시판 기준)

의류의 경우 계절별 의류 종류가 달라 문제



해결법:???



# 확장 앱 실행 화면

한 카테고리의 키워드 선택 시  
다른 카테고리 선택 불가

제목

공통/선택된 카테고리 안 키워드에 대해서만 선택 가능합니다.

확인

제목

사이즈 x

공통

상품명

브랜드

가격

상태

색상

거래방식

의류

착용횟수

사이즈

IT/전자기기

용량

사양

사용기간

변경

# 텍스트 대치 구현

## JS 파일 삽입 방법

1. Content\_script 이용
2. Whale(chrome).tabs.executeScript 이용



- DOM에 접근 가능 (이벤트 핸들러 설정, DOM 탐색, 변형 가능)
- XMLHttpRequest / fetch API로 다른 도메인의 페이지와 자료 송수신 가능

# 텍스트 대치 구현

## <사용 API>

<a href="#">whale.tabs</a>	탭 · 창	새로운 탭을 생성하거나 이미 생성된 탭을 제어할 수 있습니다.	없음
<a href="#">whale.runtime</a>	확장앱	백그라운드 페이지 검색, 매니페스트 확인 및 확장앱 수명주기에 관한 이벤트 수신, 메시지 교환 등의 기능을 제공합니다.	없음
<a href="#">whale.extension</a>	확장앱	서로 다른 확장앱 사이에 메시지를 교환하거나, 현재 확장앱에 관한 정보를 얻을 수 있습니다.	없음

## 콘텐츠 스크립트 제약

- 콘텐츠 스크립트는 삽입된 웹페이지 / 확장앱 페이지 / 백그라운드 페이지 모두와 완전히 분리된 별도의 맥락에서 동작
- 브라우저 API 사용 제한
  - 콘텐츠 스크립트에서 사용할 수 없는 브라우저 API는 메시지 교환을 통해 간접적 이용

## 네이버 카페 구성 (iframe)

★

카페정보

나의활동

매니저

qdition

since 2003.12.10.

카페소개

소

17,902,577

초대하기

★

즐거찾는 멤버

3,684,031명

게시판

구독수

2,047,612회

우리카페앱

수

140,104회

NAVER

대표카페

2018년 대표 인기카페

카페 글쓰기

카페 채팅

검색

★

즐거찾는 게시판

전체글보기

141,064,486

중고나라★새소식

중고나라★이벤트

당첨자발표/종료

공지사항

이용정책

중고나라 이야기

클거래 후기(칭찬)

어려운 이웃 사연

여성상의

새글알림

제목

661024349 1회입고 보관한 엘리 타하리 꽃 카라 코트 회색

661024125 커비넷 무스탕 판매해요. 판매

661024070 해브어긋타임 후드 팝니다 (기모) 판매

661023986 임블리 원피스 새상품 판매

661023889 자라 한겨울 무스탕 5만원 판매

661023783 체크 롱자켓 시즌 오프 저렴하 판매

661023755 빈폴슬카라 오픈가디건 55 판매

661023734 임블리 리버시를 덤블롱 핑크베이지 판매

661023718 브럼핑 고급라인,노블레스 윈넨혼방 아우터(브로치세트) 판매

661023686 버쉬카 양털 누빔 청자켓 판매

661023644 COS PANELLED SLEEVELESS DRESS(34사이즈) 판매

661023618 arlat wool blend pile dress(34사이즈) 판매

661023613 판매

661023583 자기복스 5만원에 저림아 드려요 판매

661023196 로브로브 한해연 콜라보 구해요!!! 로브로브서울 lovlov 완료

<div id="group-area" class="skin-1080 fr"> == \$0

<div id="main-area" class="skin-1080 fr"> == \$0

<input id="gateOpen" type="hidden" value="1">

<input id="searchviewtype" type="hidden" value="title">

<iframe name="cafe\_main" id="cafe\_main" title="카페 메인" src="//cafe.naver.com/ArticleList.nhn?search.clubid=10050146&search.menuid=356&search.boardtype=L&search.totalCount=501&search.page=1000" width="860" height="100%" frameborder="0" scrolling="no" marginwidth="0" marginheight="0" allowtransparency="true" allowfullscreen style="height: 1283px;">

#document

<!doctype html>

<html lang="ko">

<head>...</head>

게시물 테이블이 iframe 안에 존재

→ Manifest에 "all\_frames":true, "run\_at": "document\_idle" 정의

## OOV(Out of Vocabulary) 문제

```
File "C:\nlp\env\lib\site-packages\gensim\models\keyedvect
raise KeyError("word '%s' not in vocabulary" % word)
KeyError: "word '18000원' not in vocabulary"
(env)
이민희@DESKTOP-OKC9JF6 MINGW64 /c/nlp
```



### <해결방법>

1. 많은 데이터 셋 수집
2. Pretraining 벡터를 사용해서 fastText 적용



이름 , 브랜드



NER 모델구축



사전 일치 반환



가격, 색상, 착용횟수  
용량, 사용기간, 사양



re.findall

규칙 맞는 단어 반환

1개 반환

반환

여러 개 반환

모델 사용

0개 반환

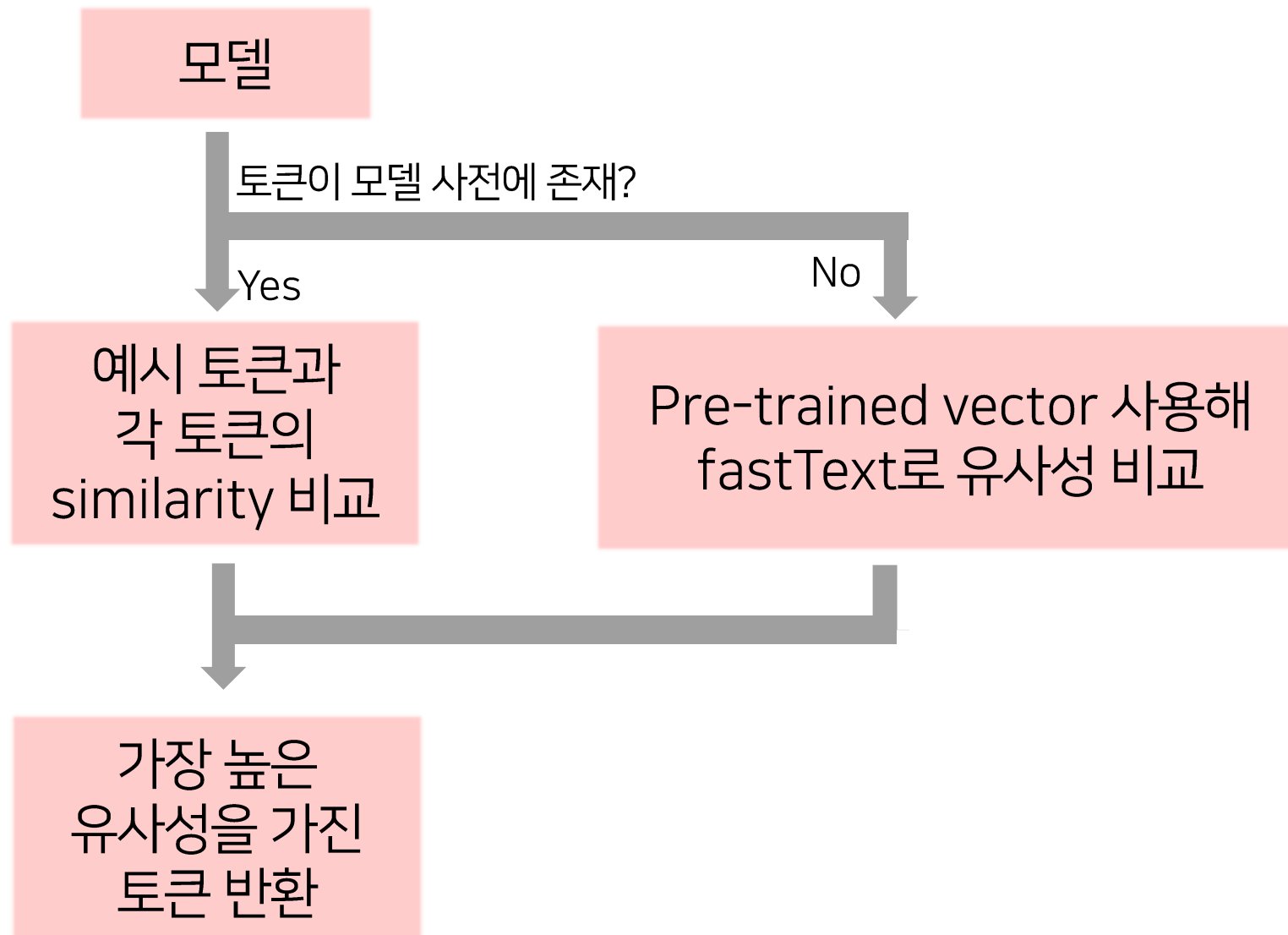
주제 단어 근처  
토큰 획득

← 사이즈, 상태

특징으로 전처리

모델 사용





## 유사도 처리 과정

✓ `string_list`: ['NAIN 나인 코트 (s) 팝니다.', '사진 색상은 노랑인데 실물은 연두...']

0: 'NAIN 나인 코트 (s) 팝니다.'

1: '사진 색상은 노랑인데 실물은 연두에 가까운 듯요 가벼워요.'

2: '한번 입은 상태 최상상이고 오버 핏으로 멋스러워요.'

3: '택포 55,000원 20000원에 사실 분 문자 주세요.'

4: '0104284공구사팔'

예시 문장

## 유사도 처리 과정

```
def getColor(string_list, docs_list):
```

```
    tokenList: [('사진', 'Noun'), ('노랑', 'Noun')]
```

근처 토큰

```
    if len(find_keyword) == 1 :
        find_keyword = find_keyword[0]
    elif len(find_keyword) > 1:
        find_keyword = getMostSimilarity(find_keyword, "빨강")
        #find_keyword = find_keyword[len(find_keyword)-1]
    else:
        find_keyword = '?'

        searchList = ["색상", "컬러"]
        #TODO 사전에 이런거 추가

        tokenList = getAroundToken(docs_list, searchList)

        find_keyword = getMostSimilarity(tokenList, "빨강")
```

## 유사도 처리 과정

#예시 토큰이랑 가장 비슷한 토큰 반환

```
def getMostSimilarity(token_list, ex_str):  
    ex_token = wordTokenize(ex_str)  
    maxSimilarity = 0  
    for t in token_list:  
        try:  
            tempSimilarity = model.similarity(ex_token, '/'.join(t))  
        except:  
            #내가 만든 모델에 없는 단어  
            try:  
                tempSimilarity = fastTextModel.similarity(ex_str, t[0])  
            except:  
                #더 큰 모델에 없는 단어  
                tempSimilarity = 0  
        if tempSimilarity > maxSimilarity:  
            find_keyword = t[0]  
            maxSimilarity = tempSimilarity
```

‘사진’이 직접 구현한 모델 사전에 X  
미리 학습된 사전에서 유사도 검색

## 유사도 처리 과정

```
ex_str: '빨강'
ex_token: '빨강/Noun'
maxSimilarity: 0
> t: ('사진', 'Noun')
> tempSimilarity: 0.14938183
> token_list: [('사진', 'Noun'), ('노랑', 'Noun')]
```

'사진'과 '빨강'의  
유사도

## 유사도 처리 과정

```
ex_str: '빨강'
ex_token: '빨강/Noun'
find_keyword: '사진'
> maxSimilarity: 0.14938183
> t: ('노랑', 'Noun')
> tempSimilarity: 0.79581606
> token_list: [('사진', 'Noun'), ('노랑', 'Noun')]
```

'노랑'과 '빨강'의  
유사도

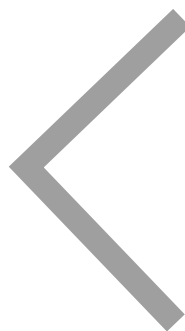
## 유사도 처리 과정

'사진', '빨강' 사이 유사도

```
> t: ('사진', 'Noun')  
> tempSimilarity: 0.14938183
```

'노랑', '빨강' 사이 유사도

```
> t: ('노랑', 'Noun')  
> tempSimilarity: 0.79581606
```



'노랑' 변환

## 다음주 개발 계획

**박시현**

웹일 브라우저와 서버 통신 구현 시작  
자연어처리 결과에 따른 데이터셋 수정 (필요시)

**박진영**

테스트 데이터로 텍스트 대치 구현 완료  
서버와의 통신 구현 시작

**이민희**

NER 모델 구축  
모델 파라미터 변경



## 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 리포 트 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								중간 고사 & 중간 데모 준비						최종 데모	최종 리포 트 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

