

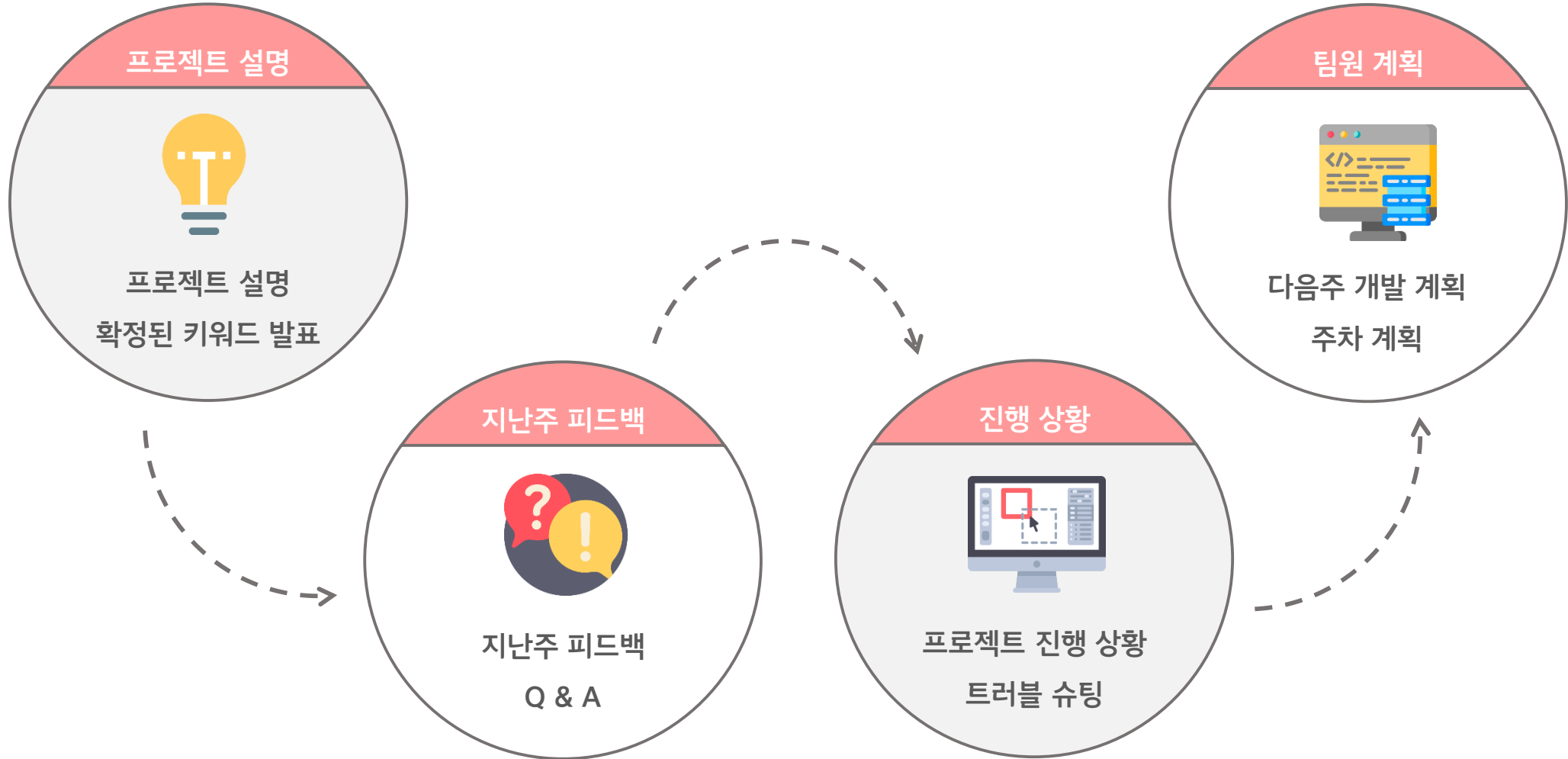


# 커뮤니티 상 거래를 위한 게시글 제목 변경 웹 확장 프로그램



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

# 제공할 키워드 목록

## - 기본 -

이름 / 가격 / 브랜드 / 상태 / 거래 방식 / 색상

## - 의류 -

착용횟수 / 사이즈

## - IT 기기 -

용량 / 사용 기간 / 사양 ( CPU | Graphic card | 크기 )

## 지난주 Feedback 및 Q&A

Q. 모바일 웹 페이지를 크롤링하는 건 어떨까요?

A. 단순히 링크 앞에 m.을 붙여서는 현재 페이지와 같은 모바일 웹 페이지로 이동 할 수가 없음

Q. 확장 프로그램 사용자가 파이썬 모듈을 사용하기 위해

다른 서버에서 처리 할 http 인터페이스가 필요할 것 같습니다!

A. Django 사용 예정

## 지난주 Feedback 및 Q&A

Q. 의류 / IT기기 외에도 제공 할 키워드 종류가 적어 보이는  
도서 카테고리도 제공하는 게 어떨까요?

A. 도서명, 영화명, 음악명, TV 프로그램 명 등과 같은  
제목 개체명(Title Named Entity)인식의 어려움

1. 제목 개체명은 단어 명사구 문장 등 형태가 매우 다양
2. 제목 개체명은 다른 종류의 개체명이 될 수 있음
3. 제목 개체명은 내부에 특징이 없음

## 프로젝트 진행 상황

박시현

카테고리, 항목 별 데이터셋 구축 / 로그인 정보 가져오기 시도  
게시글 내용 형식 파악 / 파싱 전처리 작업

박진영

웹 프론트엔드 개발  
화면 구성 구현 / 키워드 클릭 반응 구현

이민희

자연어 처리 라이브러리 학습 / 개발 방향 설정  
단어 전처리 및 토큰 추출

## 게시글 크롤링 구현

제목

가격

SYSTEM 시스템 프리넥 니트 탑, 90 사이즈 메인 제품 팝니다 | 155,000원 | <p>SYSTE  
! 프리넥 니트 탑, 90 사이즈 SY1I4KTO157W IN 팝니다.<br>18년도 SS 제품으로

내용

각 카테고리 별로 45개의  
게시물 제목, 가격, 내용 크롤링



## 게시글 크롤링 구현

45개 게시물 크롤링 시간

```
--- 10.986102819442749 seconds ---
```

페이지 하나 ( 15개 게시물 ) 걸리는 시간 = 3.6 초

## 크롤링 시 로그인 보안관련 사항 문제

중고나라 카페 가입자만 볼 수 있는 게시물에 접근하기 위해서  
사용자가 네이버에 로그인이 되어있는지, 카페에 가입이 되어있는지 확인 필요

동적 크롤링 라이브러리인 Selenium으로 직접 로그인 하는 방법도 있지만,  
BeautifulSoup보다 최대 20배 느려서 Selenium사용하지 않고 진행하는 방향으로 고려 중

## 확장 앱 실행 화면

확장 앱 실행 시  
기본 화면

제목

금품

상품명 브랜드 가격 상태

색상 거래발식

의류

착용횟수 사이즈

IT/전자기기

용량 사양 사용기간

변경

커서 올리면  
+ 버튼으로 변경

제목

금품

+ 브랜드 가격 상태

색상 거래발식

의류

착용횟수 사이즈

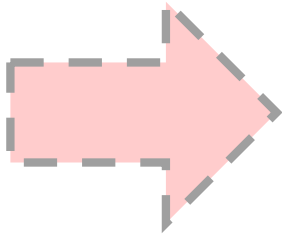
IT/전자기기

용량 사양 사용기간

변경

## 콘텐츠 보안 정책으로 인한 인라인 코드 비허용

버튼 클릭 시 수행 할  
onClick() 메소드를  
인라인 코드로 구현 할 수 없음



인라인 코드 대신  
addEventListener 사용

The mockup shows a form with the following sections:

- 제목 (Title):** A text input field.
- 상품 (Product):** Buttons for '+', '브랜드' (Brand), '가격' (Price), '상태' (Status), '색상' (Color), and '거래방식' (Transaction Method).
- 의류 (Clothing):** Buttons for '착용횟수' (Wearing Frequency) and '사이즈' (Size).
- IT/전자기기 (IT/Electronics):** Buttons for '용량' (Capacity), '시각' (View), and '사용기간' (Usage Period).
- 변경 (Change):** A teal button at the bottom right.

## 확장 앱 실행 화면

+ 버튼을  
클릭하면

확장 앱 목록

제목

상품명 x

상품명

브랜드

+

상태

색상

거래발식

의류

착용횟수

사이즈

IT/전자기기

용량

사양

사용기간

변경

상단 토큰 리스트에  
추가가 됨

확장 앱 목록

제목

상품명 x

가격 x

상품명

브랜드

+

상태

색상

거래발식

의류

착용횟수

사이즈

IT/전자기기

용량

사양

사용기간

변경

## 확장 앱 실행 화면

선택된 키워드에  
커서 올리면 색 변환

제목

상품명 x 가격 x

공통

상품명 브랜드 가격 상태

색상 거래발식

의류

착용횟수 사이즈

IT/전자기기

용량 사양 사용기간

변경

변경 버튼에  
커서를 올렸을 시

확장 앱 영역

제목

상품명 x 가격 x

공통

상품명 브랜드 가격 상태

색상 거래발식

의류

착용횟수 사이즈

IT/전자기기

용량 사양 사용기간

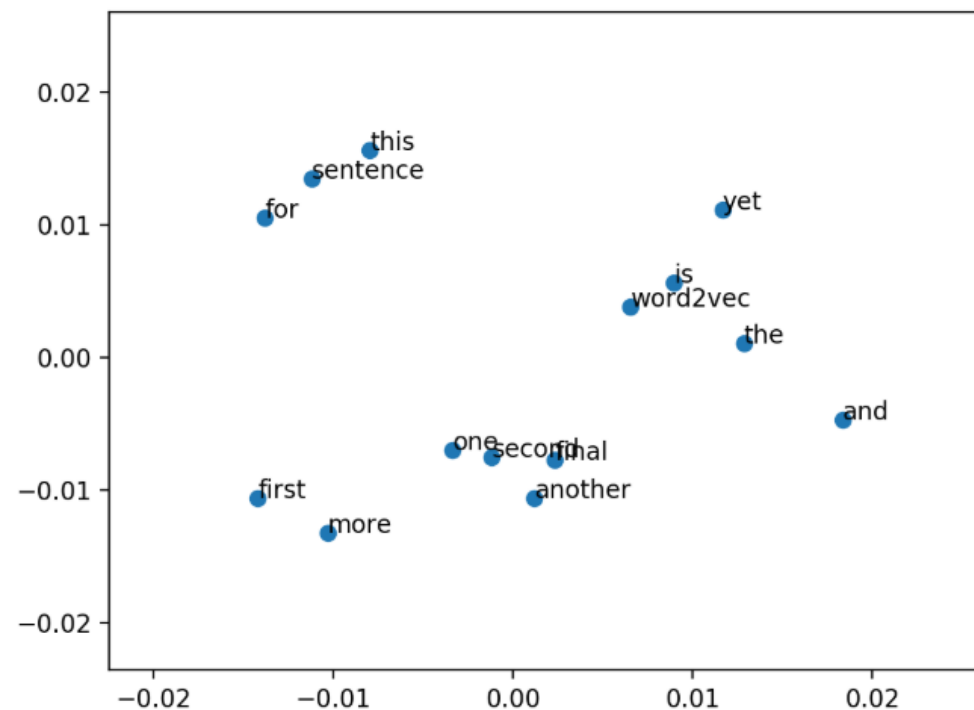
변경

# 단어 벡터 표현

컴퓨터가 인지할 수 있게 단어를 수치적인 방식인  
‘ 벡터 ’ 로 표현

one-hot encoding X

단어 자체가 가지는 의미 자체를  
다차원 공간에서 벡터화 하는 방식



# Word2Vec

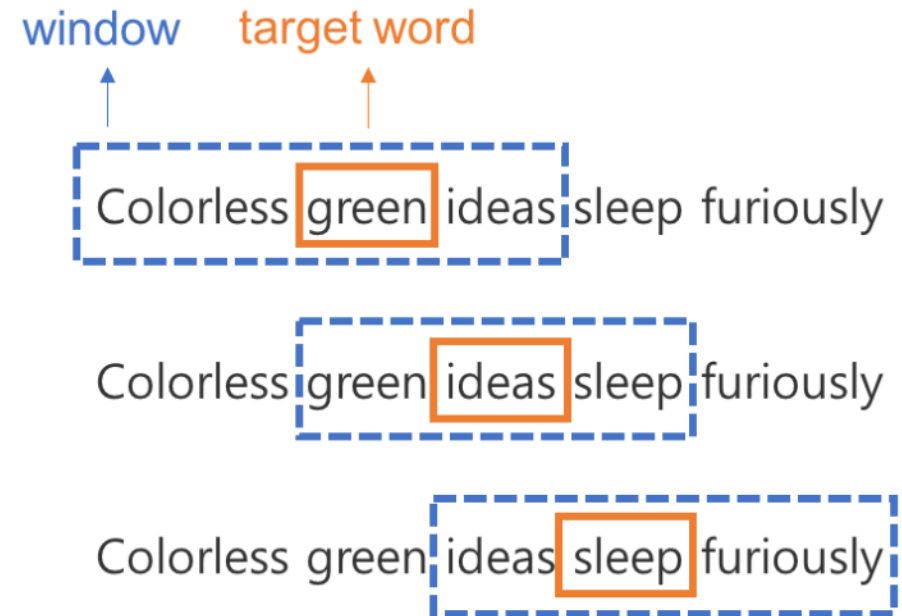
단어 주변을 보고, 단어를 예측 맥락을 예측

-> 예측 모델을 학습하면서 단어를 어떻게 표현할 지 배우고  
비슷한 단어가 비슷한 벡터로 표현하게 됨

1. 주변에 있는 단어들로 중앙에 위치한 단어를 예측
  2. 중앙 단어를 이용해서 주변 단어 예측
- + 임의로 만든 틀린 데이터를 사용해 학습

Negative sampling

<https://arxiv.org/pdf/1301.3781.pdf>





# 자연어 처리 개발 방향 설정의 어려움

단어 간의 관계 데이터 벡터의 유사성만으로는  
원하는 키워드의 추출 처리가 어려울 것



전처리 작업에 높은 가중치 -> 모델은 최소로 사용  
여러 모델 구축 · 비교 · 발전  
일반적이지 않은 경우는 배제

Ex) 여러 물건을 판매하는 한 게시글, 키워드로 추정되는 토큰이 너무 여러 개 나오는 게시글

# 자연어 처리 부분 계획

1. 토큰 리스트 중 사전과 일치하는 토큰이 존재하면 원하는 키워드를 반환. string.find 사용  
Ex) 택배 택배 운포 착불 -> 택배거래 / 유명한 브랜드 이름 -> 삼성, LG, 델
2. 규칙이 있는 키워드는 정규 표현식 라이브러리 re사용. findall로 각 토큰 리스트 반환  
Ex) ~색, ~번, ~년, ~GB
3. 키워드 주제 단어 주변 토큰 중 규칙이 있는 키워드 반환  
Ex) 55 사이즈입니다, size : M, 1년 전 구매
4. 머신러닝 모델을 사용해 키워드 주제에 적합한 예시 토큰과 비슷하다고 추정된 토큰 반환  
Ex) 아이보리 -> 블랙, 그린 / 버버리 -> 칼하트, 에잇세컨즈 => NER 구축

## 사전 일치 · 규칙 존재 키워드 판별

블랙 색상의 모직코트입니다 정가 6만 9천원인데 택포 1만 5천원에 팔아욤 사이즈 : M (55입으시면 잘 맞아요) 미착용 새상품입니다

블랙  
택배  
1만5천원  
사용안함

운포 4만원 상태좋아요 진한노랑색~ 흔치 않아요 작년에 구매해서 2~3번정도 입어서 새옷수준입니다 사이즈는 프리 사이즈 입니다

진한노랑색  
택배  
4만원  
2~3번

s사이즈 약95 착용 몇번안하고 드라이클리닝해서 상태는 거의 새옷이에요! 택포2만원 ?

택배  
택포2만원  
몇번

들어온 String을 토큰화 후, 사전 일치 키워드와 규칙 있는 키워드를 출력한 결과

## 다음주 개발 계획

박시현

로그인 정보 가져와서 모든 게시글에 대해 접근 권한 생성  
텍스트로 저장된 데이터를 list형태로 값 전달

박진영

웹 프론트엔드 구현 완료  
키워드 삭제 기능 / 변경 버튼 클릭 시 동작 구현 / 도움말 및 버튼 설명 추가

이민희

데이터 문장 단위로 쪼개기  
키워드 단어 주변의 토큰 검색 구현

## 팀원 별 각 주간 개발 계획

박시현	7주차	8주차	9주차	10주차	11주차	12주차	13주차	14주차	15주차
크롤링 · 전처리		중간 고사 + 중간 데모 준비						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

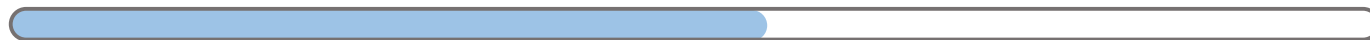
박진영	7주차	8주차	9주차	10주차	11주차	12주차	13주차	14주차	15주차
프론트엔드 화면 구현		중간 고사 + 중간 데모 준비						최종 데모	최종 리포트 제출
프론트엔드 동작 구현									
텍스트 대치 구현									
확장 앱 통신 구현									
웹 페이지 쿠키 구현									
테스트									

## 팀원 별 각 주간 개발 계획

이민희	7주차	8주차	9주차	10주차	11주차	12주차	13주차	14주차	15주차
문장 분리		중간 고사 + 중간 데모 준비						최종 데모	최종 리포트 제출
키워드 주변 토큰 처리									
키워드 사전 변경									
모델 구축, 파라미터 변경									
서버 모듈 연결									
성능 및 정확도 향상									



감사합니다!





## 전체 각 주간 개발 계획

	7주차	8주차	9주차	10주차	11주차	12주차	13주차	14주차	15주차
데이터 확보 · 전처리		중간 고사 + 중간 데모 준비						최종 데모	최종 리포트 제출
프론트 엔드									
서버 구축									
자연어 처리 알고리즘									
모듈 연결									
유닛 · 전체 테스트									