

오픈소스 SW 중간 보고서

인공지능 기술을 활용한 허위매물 필터링
시스템



SW 융합대학

모바일시스템공학과 32193869

장수윤, 32192969 이강훈

목차

제 1 장 개발 기술의 개요

1. 개발 배경
2. 개발 목적
3. 개발 현황 및 필요성
4. 기술 개발 성과 및 기대효과
5. 개발 문제점 및 해결책

제 2 장 개발목표 및 개발 내용

1. 개발 환경 및 구축 방안
2. 프로젝트에 사용된 기술
3. 프로그램 구성도

제 3 장 세부 개발 내용 및 방법

1. 기술개발 제공 기능
2. 기술개발 내용 및 결과
3. 기술 어려움

제 4 장 Reference

1. 역할 분배
2. 참고문헌
3. GitHub 링크
4. 후기

Project 명 : 인공지능 기술을 활용한 이미지(당근마켓 판매사진) 인식으로
허위매물 필터링 시스템

제 1 장 개발 기술 개요

1. 개발 배경

최근 코로나로 인해 O2O(online to offline) 서비스로 매물을 구매하는 추세이다. 펜데믹으로 인한 경제 침체로 소비심리가 위축되어 소비 트렌드가 변화하였는데, 합리적인 가격 뿐만 아니라 자원 재사용 선호, 중고제품에 대한 긍정적인 인식 전환으로 중고마켓에 대한 사용률이 급증하고 있다. 최근 중고거래 앱 이용자를 조사해보면, 당근마켓, 번개장터, 중고나라 순으로 많이 사용되고 있으며 당근마켓 1인당 월 평균 이용시간은 268 분을 사용한 것으로 나타난다. 이렇게 거래량과 사용자가 많아지면서 당근마켓 뿐만 아니라 대부분의 중고 물품 거래 사이트에서 허위 매물 증가하고 있다. 또한 중고마켓의 허위매물 뿐만 아니라 직방, 다방 등의 부동산 거래 어플리케이션에서도 허위매물 게시물이 증가하는 추세이다. 최근에는 다른 게시물의 이미지를 도용하여 본인의 게시물에 업로드 하는 사기가 증가하고 있다.

2. 개발 목적

본 필터링 시스템을 이용하여 허위매물 필터링 시스템의 검증, 처리 지연 시간, 비용 감축을 하고, 소비자의 피해를 최소화하는 것이 주요 목적이다. 또한 사기 게시물 및 스팸 게시물을 차단하여 소비자들의 안전하고 평화로운 거래를 희망하며 좀 더 나은 사회적 환경을 마련하기 위한 사례가 되었으면 한다.

3. 개발 현황 및 필요성

경기도는 2020 년 6 월 경기도 지사 SNS 에 접수된 제보에 따라 온라인 중고차 매매 사이트를 조사한 결과 매물 95%가 허위매물로 드러났다. 온라인 중고차 매물 뿐만 아니라 부동산 매 물 사기도 현저히 지속되고 있다. 현재는 플랫폼 내 개인 간 거래 환경 개선을 위해 지속적으로 모니터링을 강화하며 사기 거래 사전 차단에 주력하고 있다. 하지만 제대로 된 허위 매 물 필터링 시스템이 존재하는 것이 아니라 직접 모니터링을 사기 거래 같은 게시물을 차단하는 것임으로 인력이 너무 낭비되고 있다는 것이다. 본 서비스를 이용하여 인공지능이 스스로 모니터링하여 허위 매물 게시물들을

필터링하여 사용자들에게 위험 알림을 보내줌으로서 스스로 판단을 할 수 있도록 도와주고자 개발하게 되었다.

4. 기술 개발 성과 및 기대효과

소비자 10 명 가운데 8 명은 중고 거래를 경험을 해보았다는 통계가 있다. 그만큼 중고 물품을 구매하는 것에 대한 소비자들의 거부감이 줄어들고 있다는 뜻이고, 그에 따라 앞으로 중고거래 시장은 점점 더 커져갈 것이고 플랫폼 역시 다양해져가고 있다. 하지만 정작 중고거래 플랫폼을 신뢰하는 사용자는 28% 밖에 되지 않는다는 통계가 있다. 번개장터, 당근마켓 같은 새로 떠오르는 중고 거래 플랫폼들 역시 사기에 대한 내용을 사전에 방지하기 위해서 노력을 하고 있다.

5. 개발 문제점 및 해결책

요즘 들어 더 지능적으로 스팸 광고성 글을 작성하기가 쉽기 때문에, 더욱더 신경을 써서 이를 구분할 수 있는 알고리즘을 작성하는 것이 중요하다. 사기 수법에 대한 해답이 나오면, 그에 대응하는 사기 수법은 더 다양화 될 수 있기 때문에, 사기 수법을 예상하여 예방하는 것이 중요하다.

웹 크롤링/스크래핑 차단 방지

robots.txt 라는 텍스트에 접근을 해서, 접근해야하는 경로를 미리 체크할 수 있다. 크롤러에게 URL 의 어느 부분에 액세스가 가능한지에 대한 여부를 알려주기 때문에, 일종의 웹사이트 내에서의 가이드 역할을 한다고 볼 수 있다. robots.txt 는 사이트의 루트 폴더 혹은 메인 디렉토리에 위치해야하며, 대부분 텍스트 파일로 설정이 되어있다.

제 2 장 사업 목표 및 개발 내용

1. 개발 환경 및 구축 방안

- **HTML, CSS**

구글 크롬 익스텐션이라고 해서 특별한 기술을 사용하는 것이 아니라, 결국에는 웹의 연장선상으로 볼 수 있었다. 예전에 웹을 공부했을 때 사용했던 HTML 과 CSS 의 문법을 그대로 가져와서 확장 프로그램의 기본적인 틀을 잡는데 사용했다.

- **manifest.json**

구글 크롬 확장 프로그램에서 가장 중요한 파일 중 하나라고 할 수 있는 manifest.json 파일은 확장자에서도 알 수 있듯이 json 포맷 파일로, 웹 익스텐션이 반드시 포함을 하고 있어야 하는 파일이다. 이 파일에는 확장 프로그램의 이름, 버전과 같은 정보들 기본 정보들을 명시해준다. 뿐만 아니라 프로그램의 기능, 기본 스크립트, 내용 스크립트, 브라우저 활동 등을 명시한다.

2. 프로젝트에 사용되는 기술

2-1) Selenium

셀레니움은 프로그램을 통해서 브라우저 자동화를 구현할 수 있도록 하게 해주는 툴이다. 셀 레니움은 테스트들을 수행하기 위한 다양한 툴들과 DSL (Domain Specific Language, 도메인 특화 언어) 의 집약체이다. 이 툴은 모든 브라우저에 대해서 사용을 할 수 있는데, MacOS 나 윈도우 같은 주요한 OS 에 대해서 항상 동작하고, 파이썬, 자바, C# 등 다양한 언어를 통해서 작동한다. 우리는 이 작업을 위해 파이썬으로 작업을 하기로 결정했다.

셀레니움을 파이썬으로 작업할 경우에는 여러가지 장점이 있다. 우선, 셀레니움 파이썬은 오픈 소스이고, 대부분의 웹 브라우저에 대해서 이식성이 매우 뛰어나다. 파이썬이라는 언어 자체가 갖는 장점이긴 한데, 이해하기가 매우 쉽고, 또 실행시키기가 다른 언어들에 비해 훨씬 수월하다. 셀레니움은 작업을 자동화할 수 있게 해주기 때문에, 사람이 일일이 해야하는 일을 줄여주기 때문에, 작업을 수행하는데 있어서 총 비용을 줄여주기도 한다. 비즈니스적으로 보더라도 클라이언트 입장에서는 시간과 비용을 동시에 줄일 수 있기 때문에 매우 유용하게 사용된다. 또한 테스터들에게도 부담을 덜어주기도 한다.

셀레니움을 이용하기 위해서는 웹 엘리먼트(요소)와 웹 드라이버의 개념에 대해서 알아야한다.

- 웹 엘리먼트

Web 은 하나의 문서이다. 우리가 보는 모든 페이지들은 문서의 집합이라고 볼 수 있다. 각각의 문서는 HTML 의 문법을 따라서 작성이 되는데, 모든 문서는 엘리먼트와 태그로 이루어져 있다. 엘리먼트는 우리말로 직역하자면 요소라고 할 수 있는데, 요소와 태그는 종종 혼동되어 사용되기도 한다. HTML 문서는 태그를 포함하지만, 요소를 포함하지는 않는다. 요소는 구문 분석 단계 이후에 태그들로부터 생성되는 것이다.

- 웹 드라이버

웹드라이버는 셀레니움에서 사용하는 용어로, 사용되는 메소드와 클래스들의 부모 클래스이자 메소드이다. 웹드라이버가 존재해야 웹페이지의 다양한 요소들의 작동을 수행할 수 있게 된다.

크롬 드라이버는 구글 크롬을 사용하게 해주는 셀레니움 웹드라이버가 사용할 수 있도록 만든, 분리되어 있는 실행가능한 단독 서버이다. 다시 한번 설명을 하자면, 웹 드라이버는 웹 어플리케이션의 테스트를 자동화하여 사용할 수 있도록 해주는 API 들의 모임을 가리킨다.

	A	B	C	D	E	F	G
1	index	date	price	keyword			
2	1	22-4-11	50000	갤럭시s6엣지 32기가 팝니다 직거래 갤럭시s6엣지 32기가 품명 갤럭시s6엣지제품 단품상태 사진참고용량 32기가선			
3	2	22-4-11	60000	갤럭시s6 폴박스로 팝니다~~~ 새상품급 실기스조차 없어요 직접거래 구매한지 1개월만에 해외 발령나서 갔다가 0			
4	3	22-4-12	50000	sk g5 티탄 폰 단품판매합니다 직접거래 sk g5 티탄 폰 단품판매합니다 올 5월17일까지 as가능하며 폰단품이나 상			
5	4	22-4-12	60000	sony 엑스페리아 c5 ultra e5506 16gb 미사용 새제품 팝니다 11 소니 엑스페리아 c5 울트라sony xperia c5 ultra e550			
6	5	22-4-12	70000	통신사skt, a710s 상태좋은 갤럭시a7 팔아봅니다~~블랙,핑크 a710s 갤럭시 a7 2016년형 sm-a710s 기기 외관상E			
7	6	22-4-13	60000	갤럭시s6 32gb 블랙 a급팝니다			
8	7	22-4-14	80000	나이키, 38000원, 팝니다, 팔아요, 직거래, 택배, 가능, 정상, 동작			
9	8	22-4-14	60000	갤럭시s6 블랙 32gb 팝니다통신사 유플러스이면 3사유심 모두사용가능합니다 유심 확인했습니다			
10	9	22-4-14	50000	갤럭시 노트2 배터리 포함 a급 5만원 판매 1 홍대 직거래액정에 금이 갔지만 사용하는데 지장없어요불편하시면 필			
11	10	22-4-15	50000	인천 g4 배터리커버 스카이블루 민트색 새거 팝니다 1 5만 g4배터리커버 스카이블루 g4 배터리 커버 스카이블루 색			
12	11	22-4-15	60000	팝니다 미국 유심 심플모바일 t-mobile망 마이크로 유심 2개 개당 1 5만,무료배송 1 중고나라 공식 앱 다운받기 편			
13	12	22-4-15	55000	팝니다 미국 유심 심플모바일 t-mobile망 마이크로 유심 2개 개당 1 5만,무료배송 1 중고나라 공식 앱 다운받기 편			
14	13	22-4-16	60000	lgu+ 피쳐폰 폴더 슬라이드 2g폰 3g폰 정리합니다 피쳐폰 lgu+ 피쳐폰 폴더 슬라이드 2g폰 3g폰 정리합니다 lgu+ 피			
15	14	22-4-16	50000	skt 2g폰 정리합니다 폴더 슬라이드 01x 011용 휴대폰 skt2g폰 정리합니다 폴더 슬라이드 01x 011용▶소액이니 기			
16	15	22-4-16	30000	skt kt 3g 유심방식 피쳐폰 정리합니다 폴더폰 슬라이드폰 skt kt 3g 유심방식 피쳐폰 skt kt 3g 유심방식 피쳐폰 정			
17	16	22-4-17	50000	갤럭시 s4 e300i 부품 팝니다			
18	17	22-4-17	80000	아이폰6 16gb 실버 공기계 팝니다			
19	18	22-4-18	70000	스니커즈, 거래, 직거래, 팝니다			
20	19	22-4-18	50000	풀세트 이구요6번째 사진 보시면 아시겠지만 오른쪽 옆면에 살짝 긁힘 있어요~나머지는 상태 정			
21	20	22-4-18	60000	아이폰6 16gb gold 아이폰6 16기가 폰은 사진에서 보신 것과 같이 깨끗한 폰입니다 기스는 충전단자쪽에 케이블 연			
22	21	22-4-18	90000	갤럭시s6 32gb 골드팝니다			
23	22	22-4-19	60000	팝니다 미국 유심 심플모바일 t-mobile망 마이크로 유심 2개 개당 1 5만,무료배송 1 중고나라 공식 앱 다운받기 편			
24	23	22-4-20	80000	팝니다 미국 유심 심플모바일 t-mobile망 마이크로 유심 2개 개당 1 5만,무료배송 1 중고나라 공식 앱 다운받기 편			
25	24	22-4-20	60000	팝니다 미국 유심 심플모바일 t-mobile망 마이크로 유심 2개 개당 1 5만,무료배송 1 중고나라 공식 앱 다운받기 편			
26	25	22-4-20	50000	팝니다 미국 유심 심플모바일 t-mobile망 마이크로 유심 2개 개당 1 5만,무료배송 1 중고나라 공식 앱 다운받기 편			

<셀레니움을 이용한 데이터 수집 샘플 .csv 파일>

2-2) Matplotlib

Matplotlib 는 파이썬에서 동적인, 애니메이션화된, 그리고 상호작용이 가능한 시각화를 가능하게 해주는 라이브러리이다. 이 라이브러리를 통해 다양한 시각화를 사용할 수 있다. 라인 플롯, 바 차트, 파이차트, 히스토그램, Box Plot, Scatter Plot 등을 비롯하여 다양한 차트와 플롯 스타일을 지원하며, Matplotlib.org 웹페이지에 다른 사용자들이 만들어놓은 또 다른 샘플 차트를 볼 수 있다. 보여지는 시각 효과와 레이아웃을 사용자가 직접 지정해서 수정할 수 있고, 다양한 파일 포맷으로 추출이 가능하다.

2-3) konlpy

NLP 는 텍스트에서 유의미한 정보를 분석하여 추출하는 기술 집합이다. NLP 는 컴퓨터와 인간 언어 사이의 상호 작용하는 기술로 인공지능의 핵심 기능 중 하나이다. 자연어처리가 되면, 컴퓨터는 처리된 정보를 바탕으로 내용 요약, 음성 인식, 번역, 감성 분석, 텍스트 분류 작업을 할 수 있다. 우리 일상에서 자주 쓰이는 애플의 Siri 나 구글 번역기가 NLP 의 대표적인 응용 사례이다. 한국어 텍스트를 분석할 때 가장 기본적으로 해야하는 것은 형태소 분석이다. 형태소를 나눈다면 크게 명사, 동사, 형용사, 부사, 조사, 접사, 관형사, 어미 등으로 나뉜다. 한국어를 처리 할때는 대체로 KoNLPy(코엔엘파이) 라이브러리를 많이 사용한다. KoNLPy 는 한국어 정보처리를 위한 파이썬 패키지이다. KoNLPy 는 파이썬 프로그래밍 언어로 사용할 수 있도록 만들어졌다. 파이썬이 간결한 문법구조, 강력한 스트링 연산 기능을 가지고 있을 뿐 아니라 크롤링, 웹프로그래밍, 그리고 데이터 분석을 수행할 수 있는 다양한 패키지를 사용할 수 있는 언어이기 때문이다. 우리는 이 패키지를 이용하여 형태소 분석을 한 뒤, 사기 게시글 의 대표적인 구문들을 등록해놓고 비교를 하며 허위 게시글인지 아닌지를 판별할 예정이다.

• 사기일 가능성이 높은 게시물의 특징

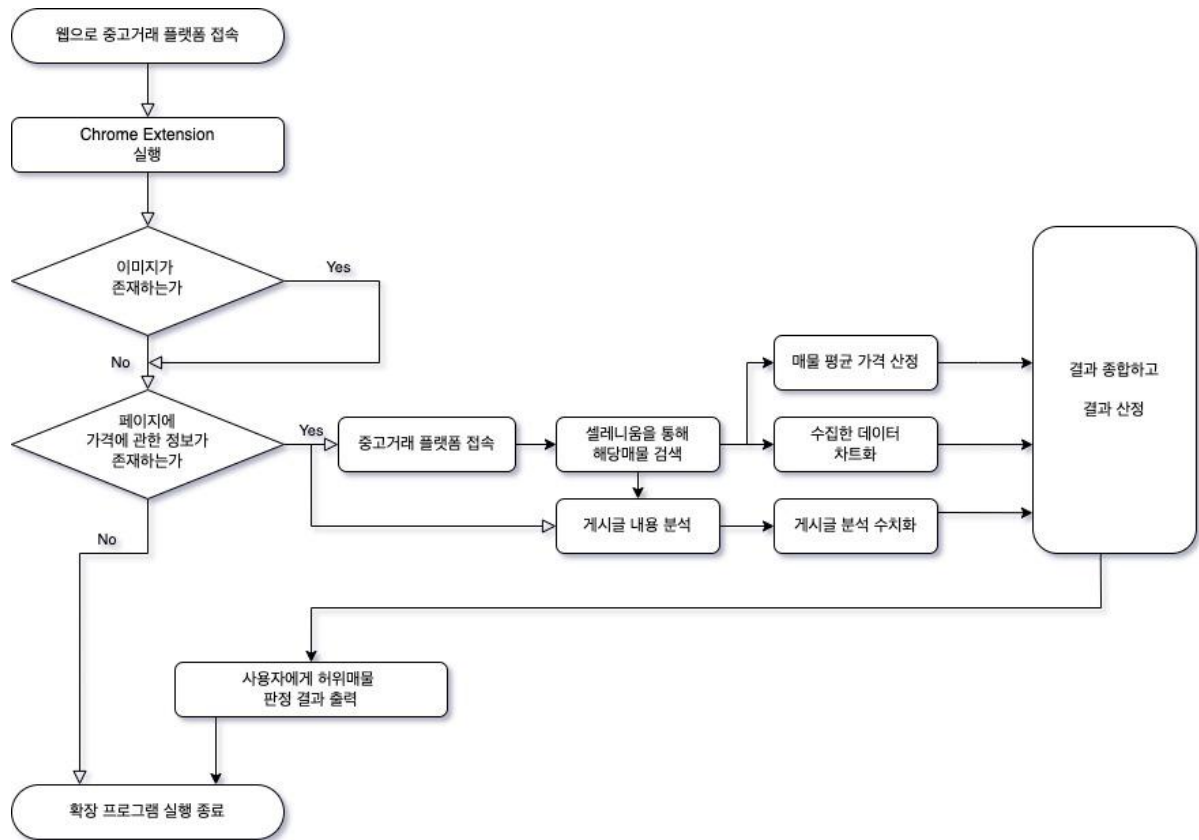
- 거래 내역이 없거나 최근 며칠 이내 뿐
- 전화번호 없고 카카오톡 메신저나 아이디로 연락 유도
- 직거래 내용은 없고, 오직 택배 거래만 요구할 경우
- 시세보다 20~30% 저렴할 경우
- 선불을 요구하거나 가상계좌를 통해 거래를 제안할 경우

- (커뮤니티일 경우) 아이디 실명과 예금주의 이름이 일치하지 않을 경우
- 사기꾼 조회 사이트(더치트, 노스캠, 경찰청 사이버안전국)검색에서 이름, 계좌, 전화번호 로 사기 신고 내역 조회
- 판매자가 결제하라고 보내준 안전거래 사이트가 네이버에서 검색이 되지 않을 경우
- 게시글 사진이 퍼온 사진일 경우



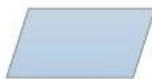


이중 저희는 몇가지의 사기 게시물 특징들을 추려서 본 서비스만의 채점표를 만들었다. 처음에는 8 가지의 특징들이 있었으나, 기술적인 부분의 문제가 있어서 4 가지의 특징들로만 채점을 하기로 했다. 9 점 만점 중 6 점 이상의 매물들은 허위 매물이라는 경고문을 띄울 예정이다.

번호	사기 게시물 특징	점수
1	시세보다 20~30% 저렴하거나 고가인 경우	3 점
2	택배거래와 직거래 중 오직 택배 거래만 요구할 경우	2 점
3	거래 내역이 없거나 최근 며칠 이내 뿐	1 점
4	리뷰 개수가 현저히 적을 경우 (10 개 이하)	2 점
	총점	9 점

3. 프로그램 순서도



<Chrome Web Extension Flowchart>

Symbol	Name	Function
	Start/end	An oval represents a start or end point
	Arrows	A line is a connector that shows relationships between the representative shapes
	Input/Output	A parallelogram represents input or output
	Process	A rectangle represents a process
	Decision	A diamond indicates a decision

<Flowchart Symbol>

제 3 장 세부 개발 내용 및 방법

1. 기술개발 제공 기능

1.1 네이버쇼핑 사이트 크롤링

Selenium 과 BeautifulSoup 패키지를 활용하여 네이버쇼핑에 있는 “아이폰 13”의 매물 정보를 불러왔다. 현재 불러온 정보는 제품명, 가격, 리뷰 개수, 제품 등록한 날짜, 거래 방식, 그리고 제품의 url 정보이다. 아래 사진들은 정보들 중 가격에 대한 기본적인 알고리즘 코드와 코드에 대한 출력값들이다.

```
metadata2 = soup.find_all('div', class_='basicList_price_area__1UXXR')[i]
price = metadata2.find('span', class_='price_num__2WUXn')
if price != None :
    price = price.text
    print("<가격> : ", price)                # 가격
elif price == None :
    price = metadata2.strong.get_text()
    print("<가격> : ", price)
```

Figure 3.1.1

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:24:
page = 0 크롤링 완료
<제품명> : Apple 아이폰 13 256GB [자급제]
<가격> : 1,130,000원
<리뷰> : 497
<날짜> : 리뷰별점 4.9497등록일 2021.09.찜하기366정보 수정요청
['택배거래']
<url> : https://cr.shopping.naver.com/adcr.nhn?x=DeJht020VA7kQeK
=====
<제품명> : Apple 아이폰 13 프로 256GB [자급제]
<가격> : 1,394,770원
<리뷰> : 2,588
<날짜> : 리뷰별점 4.92,588등록일 2021.09.찜하기483정보 수정요청
['택배거래, 직거래']
<url> : https://cr.shopping.naver.com/adcr.nhn?x=L0BhuRLrBy5Gbyq
=====
<제품명> : Apple 아이폰 13 프로 128GB [자급제]
<가격> : 1,243,000원
<리뷰> : 2,287
<날짜> : 리뷰별점 4.92,287등록일 2021.09.찜하기383정보 수정요청
['직거래']
<url> : https://cr.shopping.naver.com/adcr.nhn?x=jcgpul0gNzWchlU
```

Figure 3.1.2

Workbook 이라는 패키지를 이용하여 google colab 안에서 제품의 정보를 크롤링하고 xlsx 파일로 저장하였다. 이 패키지를 이용하면 곧바로 코랩 안에서 download 가 가능해서 쉽게 연동이 가능하였다. 아래 사진은 코랩에 저장되어 있는 xlsx 저장 위치와 xlsx 파일을 열었을 때 보여지는 데이터 값들이다.

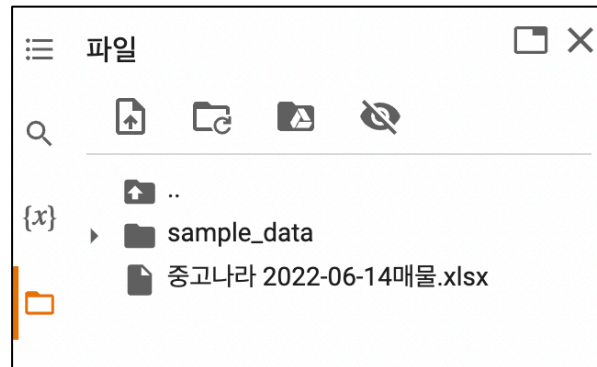


Figure 3.1.3

	A		B	C	D	E	F	G	H	I
1	제목		가격	리뷰개수	날짜	거래 종류	url			
2	Apple	아이폰 13 256GB	1,130,000	497	2021. 9.	택배거래	https://cr.shopping.naver.com/adcr.nhn?			
3	Apple	아이폰 13 프로 1,28GB	1,394,770	2,588	2021. 9.	택배거래, 직거래	https://cr.shopping.naver.com/adcr.nhn?			
4	Apple	아이폰 13 프로 1,28GB	1,243,000	2,287	2022.1.	직거래	https://cr.shopping.naver.com/adcr.nhn?			
5	Apple	아이폰 13 미니 128GB	920,530	1,241	2022. 3.	택배거래	https://cr.shopping.naver.com/adcr.nhn?			
6	Apple	아이폰 13 128GB	1,055,870	1,131	2021. 10.	택배거래, 직거래	https://cr.shopping.naver.com/adcr.nhn?			
7	Apple	아이폰 13 미니 128GB	1,046,260	753	2021. 8.	택배거래, 직거래	https://cr.shopping.naver.com/adcr.nhn?			
8	Apple	아이폰 13 프로 1,28GB	1,555,690	519	2021. 10.	택배거래	https://cr.shopping.naver.com/adcr.nhn?			
9	Apple	아이폰 13 프로 1,28GB	1,648,700	180	2021. 09.	택배거래, 직거래	https://cr.shopping.naver.com/adcr.nhn?			
10	Apple	아이폰 13 프로 1,28GB	1,455,000	313	2021. 10.	택배거래, 직거래	https://cr.shopping.naver.com/adcr.nhn?			
11	Apple	아이폰 13 프로 1,28GB	1,805,000	103	2021. 11.	택배거래	https://cr.shopping.naver.com/adcr.nhn?			
12	Apple	아이폰 13 미니 128GB	1,289,000	55	2021. 10.	택배거래	https://cr.shopping.naver.com/adcr.nhn?			
13	Apple	아이폰 13 프로 1,28GB	2,070,320	53	2021. 09.	택배거래, 직거래	https://cr.shopping.naver.com/adcr.nhn?			
14	Apple	아이폰 13 프로 1,28GB	1,902,610	43	2022. 01.	직거래	https://cr.shopping.naver.com/adcr.nhn?			
15	Apple	아이폰 13 512GB	1,329,000	50	2021. 09.	직거래	https://cr.shopping.naver.com/adcr.nhn?			
16	Apple	아이폰 13 프로 1,28GB	611,100	1	2021. 09.	택배거래, 직거래	https://cr.shopping.naver.com/adcr.nhn?			
17	Apple	아이폰 13 미니 128GB	187,100	1	2021. 07.	직거래	https://cr.shopping.naver.com/adcr.nhn?			

Figure 3.1.4

이 데이터 값들을 통해서 우리는 konlpy 와 pandas 등의 파이썬 패키지를 이용하여 게시물의 정보들을 분석하였고, 위에서 본 서비스의 채점표를 이용하여 각각의 기준에 따라 점수를 매기고 합산한 값을 출력하였다. 아래 맨 마지막 열 "sum"이 우리의 최종 값이 된다.

	제목	가격	리뷰개수	날짜	거래 종류	url	price	review	send_box	date	sum
0	Apple 아이폰 13 256GB [자급제]	1130000	497	2021. 9.	택배거래	https://cr.shopping.naver.com/adcr.nhn?x=BXk8y...	3	0	2	0	5
1	Apple 아이폰 13 프로 256GB [자급제]	1394770	2588	2021.9.	택배거래, 직거래	https://cr.shopping.naver.com/adcr.nhn?x=NqhJU...	3	0	0	0	3
2	Apple 아이폰 13 프로 128GB [자급제]	1243000	2287	2022.1.	직거래	https://cr.shopping.naver.com/adcr.nhn?x=pihKE...	3	0	0	0	3
3	Apple 아이폰 13 미니 128GB [자급제]	920530	1241	2022.3.	택배거래	https://cr.shopping.naver.com/adcr.nhn?x=EukM3...	3	0	2	0	5
4	Apple 아이폰 13 128GB [자급제]	1055870	1131	2021.10.	택배거래, 직거래	https://cr.shopping.naver.com/adcr.nhn?x=yOIRq...	3	0	0	0	3
...
105	Apple 아이폰 13 Pro 자급제 - 알파인그린	1618536	0	2021.07.	택배거래, 직거래	https://cr.shopping.naver.com/adcr.nhn?x=ILfVV...	0	2	0	0	2
106	Apple 아이폰 13 mini 자급제 - 미드나이트	1269616	3	2020.08.	직거래	https://cr.shopping.naver.com/adcr.nhn?x=HO8mP...	3	2	0	1	6
107	Apple 아이폰 13 mini 자급제 - 레드	2796759	10	2019.12.	택배거래	https://cr.shopping.naver.com/adcr.nhn?x=gzf0v...	3	2	2	1	8
108	Apple 아이폰 13 Pro 자급제 - 실버	1343455	21	2021.12.	택배거래, 직거래	https://cr.shopping.naver.com/adcr.nhn?x=YpTfw...	3	0	0	0	3
109	Apple 아이폰 13 자급제 - 미드나이트	1503863	236	2022.04.	택배거래, 직거래	https://cr.shopping.naver.com/adcr.nhn?x=OUxul...	3	0	0	0	3

110 rows × 11 columns

Figure 3.1.5

2. 기술개발 내용 및 결과

2.1 Designing User Interface

당초에는 한 페이지에 전부 원하는 기능들을 보여지게 하려고 했었지만, 구현하고 보여져야 하는 기능들이 많아짐에 따라서 Top Navigation 바(bar)를 두고, 여러 개의 페이지로 나눠서 나타나도록 했다. 네비게이션 바에는 다양한 메뉴들이 있는데, 아래와 같다. (+ Figure 2.1.3 참고)

Home / Safe / Chart / Link / About

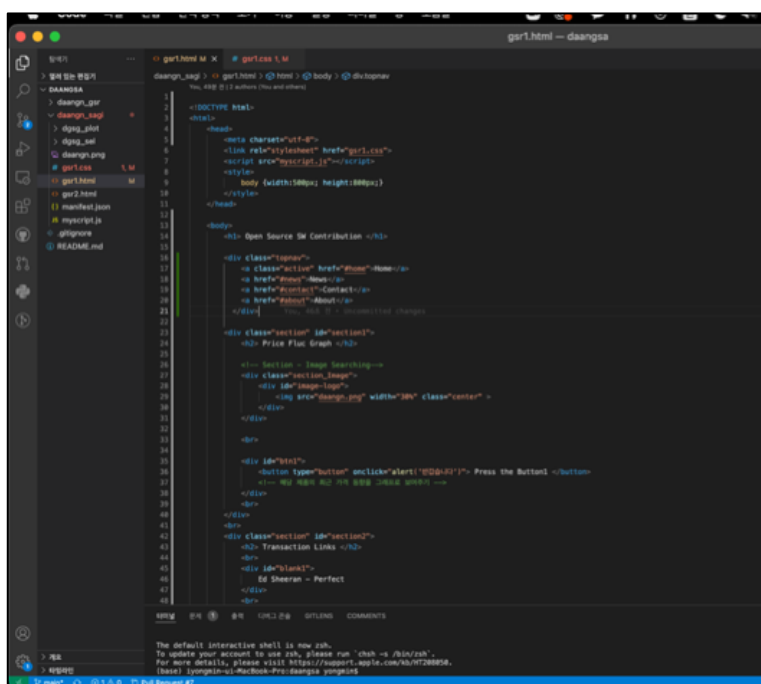
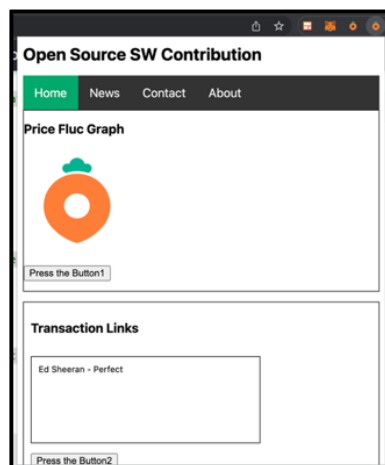


Figure 3.2.1.1 & 3.2.1.2

2.2 홈 화면 정의

1) 홈 화면에서는 기본적으로 Search 버튼을 추가했다.

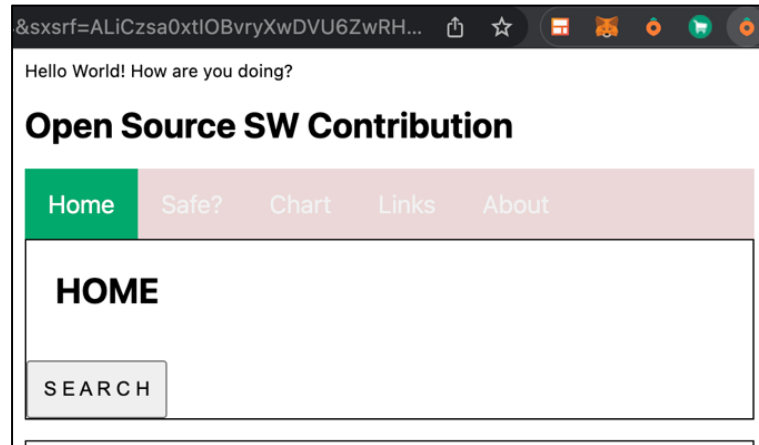


Figure 3.2.1.3

2) 두번째 페이지에서는 최종 결과를 보여주는 페이지인데, 여러 분석 요인들(가령, 중고 매물의 최근 일주일/한달 간의 동향, 중고거래 페이지의 키워드 분석, 사용자 정보 등)을 토대로 안전한 매물인지, 불안정한 매물인지를 판단한 결과를 신호등처럼 보여주게 된다.

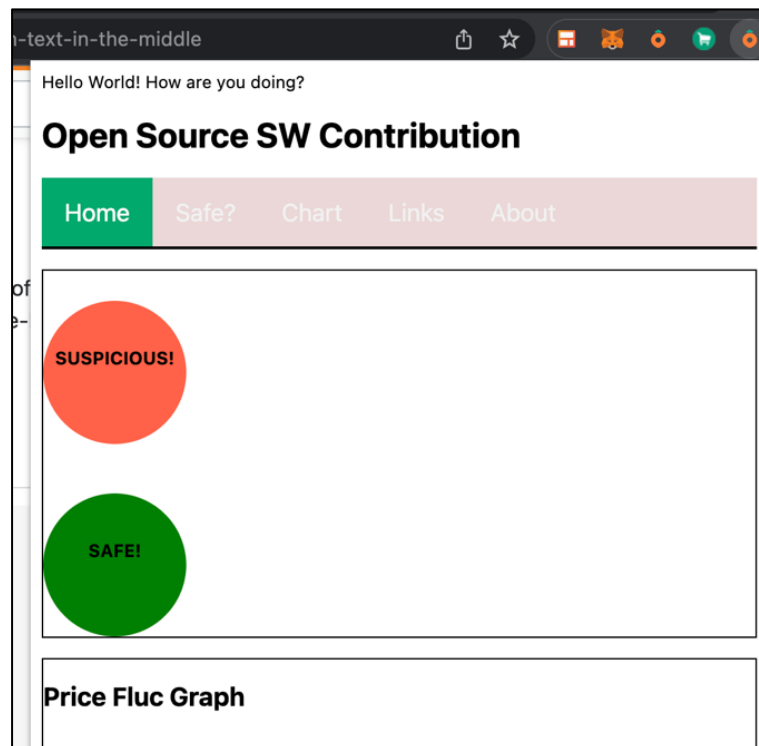


Figure 3.2.1.4

3) Chart 메뉴에서는 웹스크래핑 검색을 통해 얻은 최근 매물의 가격정보를 가져와서 차트로 보여주는 페이지가 될 것이다.

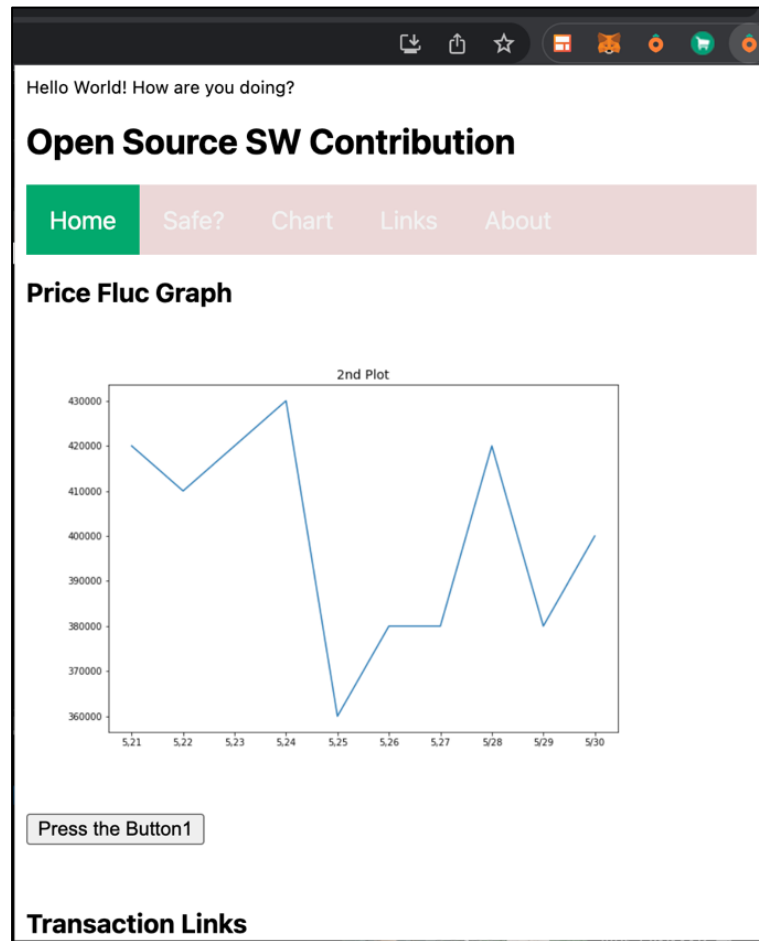


Figure 3.2.1.5

4) Links 메뉴는 매물들의 정보를 가져오게 된 페이지들의 링크들을 보여주게 될텐데, 일종의 정확한 판단 근거를 보여주는 셈이다.

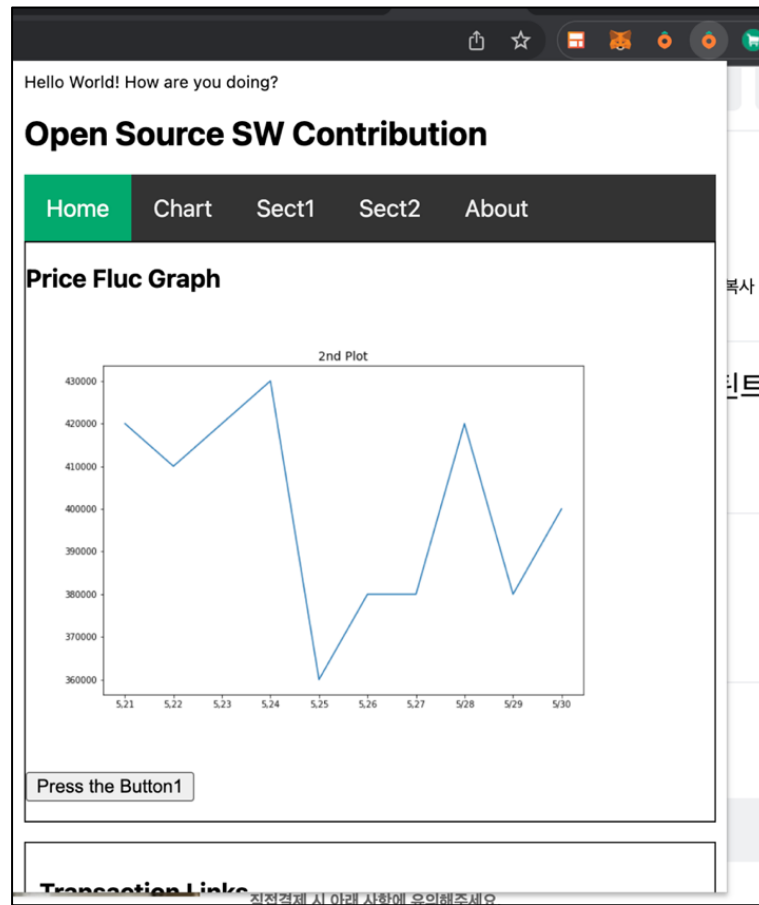


Figure 3.2.1.6

5) About 메뉴에서는 작성자에 대한 정보를 제공해주었다.

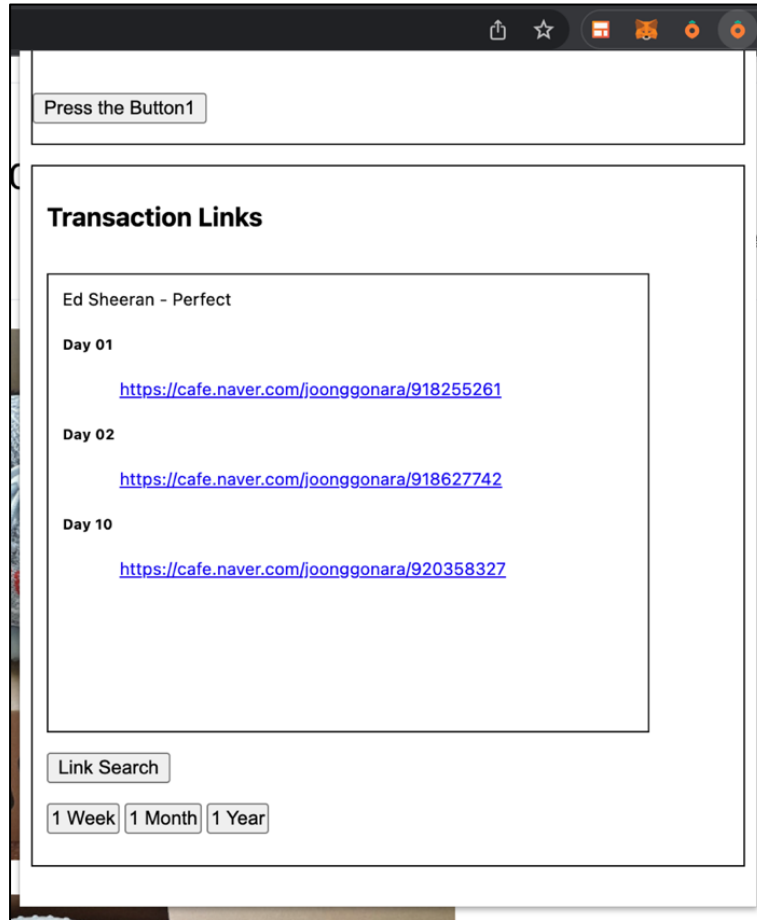


Figure 3.2.1.7

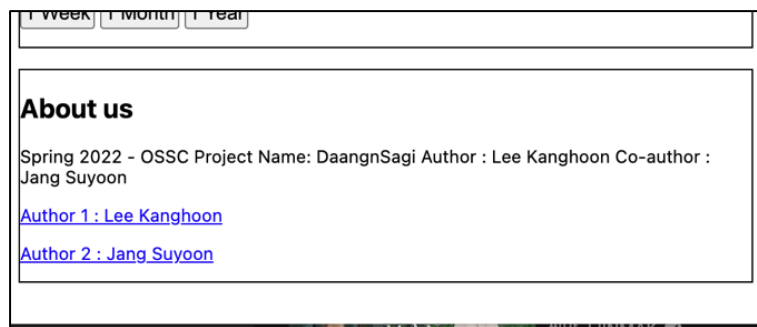


Figure 3.2.1.8

- 홈 화면은, 크롬 확장 프로그램을 실행시켰을 때 제일 먼저 보이는 화면으로, 기본적인 내용들이 들어갈 예정이다.
- Safe 메뉴는 제일 중요한 결과값을 보여주는 페이지인데, 여러 분석 요인들(가령, 중고 매물의 최근 일주일/한달간의 동향, 중고거래 페이지의 키워드 분석, 사용자 정보 등)을

토대로 안전한 매물인지, 불안정한 매물인지를 판단한 결과를 신호등처럼 보여주게 된다.

- Chart 메뉴에서는 웹스크래핑 검색을 통해 얻은 최근 매물의 가격정보를 가져와서 차트로 보여주는 페이지가 될 것이다.

- Links 메뉴는 매물들의 정보를 가져오게 된 페이지들의 링크들을 보여주게 될텐데, 일종의 정확한 판단 근거를 보여주는 셈이다.

- About 메뉴에서는 작성자에 대한 정보를 제공해주었다.

3. 기술 어려움

컴퓨터에게 이미지는 거대한 숫자 집합에 불과하다. GAN 알고리즘은 각 픽셀 값의 차이를 절대값을 이용해서 구현하는데, 쉽게 값을 구할 수 있다는 장점이 있지만 위상변화 등에 매우 취약하다는 단점이 존재합니다. 촬영의 각도에 따라(Viewpoint variation), 조명, 밝기 차이에 따라(illumination) 크게 달라진다. 프로젝트를 진행하면서 이미지 정확도 체크를 하는 것이 현실적으로 불가능하다는 것을 알게 되었다. 처음에는 GAN 알고리즘을 활용해서 매물에 대한 사진 정보들을 기준으로 유사한 사진이나 등록정보를 찾아낼 수 있는 방법을 구상하였다. 하지만 이것은 중고시장 매물에 대한 부분적인 허위 여부를 판별하거나 허위 정보일 가능성을 파악하는 수준이지, 실질적인 서비스 가능성 여부는 매우 낮은 것으로 파악된다.

GAN 모델 학습 중 생긴 문제점

Model collapsing

학습할 때 클래스가 여러개인 실제 데이터의 분포를 따라가지 않고, 특정 데이터에 강하게 몰린다. 이는 다양성을 잃어버리고 현상이다. 모델을 손실값만을 줄이려고 학습하기 때문에, 전체 데이터의 분포를 찾지 못하고 하나의 모델에 강하게 편향된다. 분류기보다 생성기를 학습시키는 것이 어려운데, 생성기에는 학습이 잘 안되어 있어 둘 사이의 힘의 균형이 깨지는 경우 GAN 학습이 더이상 진전될 수 없다. 초기에는 생성기와 분류기 간의 업데이트 횟수를 조절하거나 함수의 비율을 조절하는 등의 밸런스를 맞추기 위해 시도하였으나 부족하였다.

제 4 장 Reference

4. 역할 분배

이강훈

- 구글 크롬 익스텐션 프로그램 UI/UX 디자인
- 웹크롤링을 통한 가격정보 & 페이지 링크 정보 수집
- 크롤링 데이터 차트화 with Python

장수윤

- 셀레니움을 통한 웹크롤링
- 게시글 정보 추출하여 csv 화 하여 저장 (가격, 리뷰수, 사용자 정보, 제목, url)
- 게시글 내용 분석 및 수치화

5. 참고 자료

1) w3schools library

https://www.w3schools.com/howto/howto_js_topnav.asp

2) Plot CSV Data in Python

<https://plotly.com/python/plot-data-from-csv/>

3) 프로그래머 김플 스튜디오 - 파이썬 자동화 selenium 사용법 브라우저 매크로 프로그램 만들기

<https://youtu.be/ZFmTwbRQ0uc>

4) [StackOverflow] Save plot to image file instead of displaying it using Matplotlib

<https://stackoverflow.com/questions/9622163/save-plot-to-image-file-instead-of-displaying-it-using-matplotlib>

5)

<https://www.koreascience.or.kr/article/JAKO202129857861065.pdf>

6)

<https://github.com/jin05154/project->

[rum/tree/93557836807847a827555af49dcd84026eb82fb5](https://github.com/Kang-Hoon/daangsa)

7)

<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

6. Github 주소

<https://github.com/Kang-Hoon/daangsa>

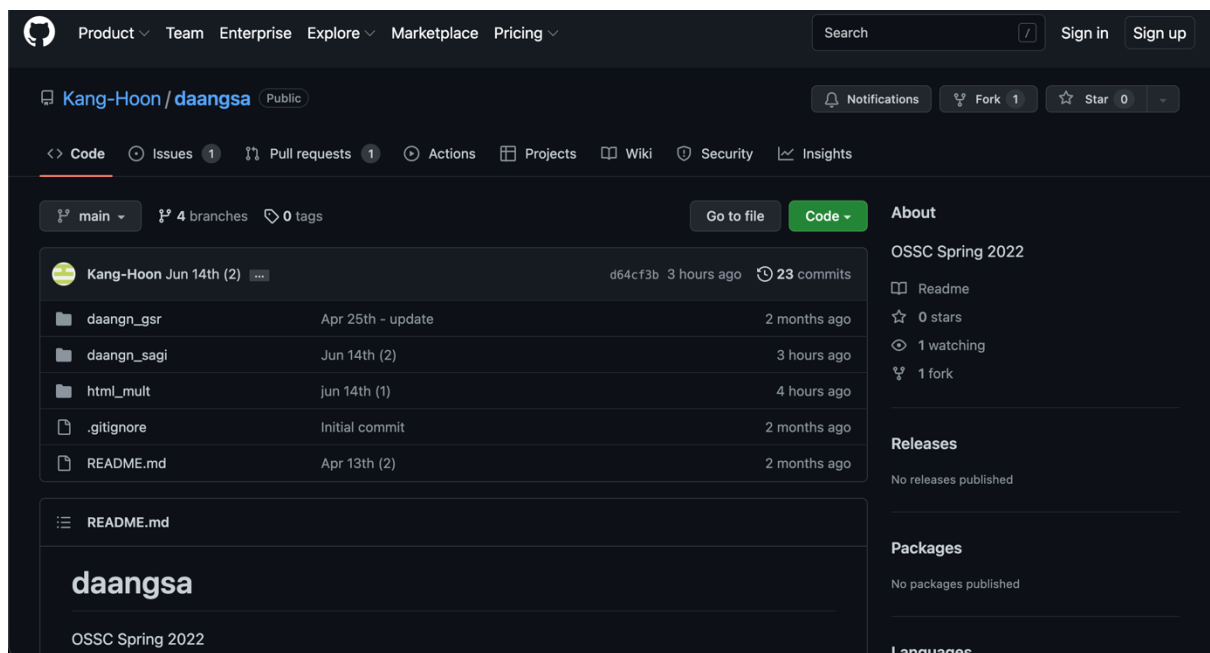


Figure 4.3

7. 후기

- 이강훈

이번 프로젝트를 수행하기 전까지는, 오픈소스에 대해서 전혀 알지 못했다. 오픈소스의 중요성에 대해서도, 왜 내가 작성한 소중한 코드를 남들에게 공개를 해야하는지, 그로 인해 파생될 수 있는 오픈소스의 효과에 대해서도 전혀 알지 못했다. 하지만 오픈소스는 이미 생각보다 많은 곳에서 사용이 되고 있었다. 유수의 세계 IT 대기업들은 자신들의 프로젝트를 Github 에 올리고 있고, 일반 사용자들에게 오픈하면서 더 나은 서비스의 개발을 기대하고 있다.

또한 구글 크롬 확장 프로그램, 파이썬을 이용한 웹크롤링 등 평소에 관심 가지고 공부해보고

싶었던 분야들에 대해서 공부해볼 수 있었던 좋은 기회가 되었다. 이번 프로젝트를 토대로 앞으로도 관련된 프로젝트를 더 많이 진행해서 프론트 개발을 경험하는데 좋은 초석을 다질 수 있도록 해야겠다.

- 장수윤

현재 우리 과에는 이렇게 팀으로 하는 프로젝트가 흔치 않은 기회라서 엄청난 기대를 가지고 프로젝트를 시작하였습니다. 중고나라 매물을 크롤링하여 매물의 기본적인 정보와 매물의 사진까지 불러와서 GAN 알고리즘을 이용하여 상품의 사진이 도용된 것인지 분석도 하고 게시물의 내용도 형태소 분석을 이용하여 저희만의 기준이 되는 채점표를 만들어서 기준치 이상의 점수를 얻는 게시물의 경우 허위매물 게시물로 판별하여 구글 크롬 익스텐션을 이용하여 경고 알림을 주는 시나리오까지 생각하였습니다. 하지만 막상 3 개월짜리 프로젝트에서 너무 많은 아이디어와 고난도의 기술을 활용하려고 해서 기술적인 문제가 많았습니다. 또한 중간에 배에 염증이 생겨 병원에 입원을 하여 개발을 하는데 차질이 있었습니다. 그래서 기대치만큼의 결과가 나오지 못하여 아쉬울 따름입니다. 이 프로젝트는 아이디어가 되게 좋다고 생각되어 이 "오픈소스기여" 수업이 끝나더라도 저희 팀원과 함께 추가적으로 개발의 완성도를 높이고 이전에 하지 못하였던 게시물 형태소 분석 및 이미지 전처리도 하여 포트폴리오에 활용할 예정입니다.