

2 0 2 1 캡 스톤 디 자 인

딥러닝을 이용한 웹사이트의 광고 자동 가림 기술

시는 처음이죠

컴퓨터과학과

201710952 이세인

201610970 김경민

201710947 윤정연

CONTENTS



Contents 01

기획 의도 설명 및 팀원 소개



Contents 02

프로그램 소개



Contents 03

타 제품과의 비교



Contents 04

기대 효과



Contents 05

프로그램 상세 설명



Contents 06

발전 가능성

01

Contents 01

기획 의도 설명 및 팀원 소개





[스포츠조선 뉴스 웹페이지 캡처]

프로젝트 기획 의도

인터넷 뉴스 웹페이지에 광고가 너무 많아 불편함 야기.

이용자에게 광고를 보이지 않게 하고 뉴스 기사 이미지만 노출시켜야 한다.

딥러닝 기법을 사용하여 광고 이미지 자체를 가지고 어떤 광고인지 인식하고 분류하는 기능을 가진 프로그램을 제작.

팀원 소개 및 역할 분담

설계 & 데이터 수집

201710952 이세인

AI 모델 구축 & 실험

201610970 김경민

웹 연동 & 확장 프로그램 구현

201710947 윤정연



02

Contents 02

프로그램 소개

프로그램 구성

AI part

Web part

모델의 파라미터를 조정

모델이 좋은 성능을 가지도록 실험을 반복하며 layer, epoch, drop-out, learning-rate 등 파라미터를 조정한다.



크롬 확장 프로그램으로 구현

필요한 파일들을 로컬에 다운받은 뒤, 확장 프로그램을 찾아 로드하고 활성화시킨다.



데이터 수집

모델 학습

모델 완성

웹 서버와 연결

프로그램 이용 가능!



수집한 데이터를 모델에 입력

데이터를 내부적인 카테고리별로 알맞게 나누고, 딥러닝 모델을 구현하여 지도학습시킨다.



완성된 딥러닝 모델을 웹 서버에 배포

프로그램이 웹 페이지에서 동작하도록 하기 위해 모델을 웹 서버에 올린다.

개발 환경

딥러닝 모델 구현

- 파이썬 (tensorflow, keras, sklearn, numpy, ...)
- Google Colab
- Jupyter notebook

FrontEnd

- Java Script

BackEnd

- 파이썬
- flask
- heroku



A yellow square containing the letters "JS" in a bold, black, sans-serif font.

03

Contents 03

타 제품과의 비교



기존의 광고 차단 확장 프로그램들 on 크롬 웹 스토어



AdGuard 광고차단기

제공업체: adguard.com

★★★★★ 51,984 | 생산성 | 👤 사용자 10,000,000+명



Adblocker Unlimited - 광고 차단 및 안전한 검색

제공자: uAdblock Inc.

★★★★★ 120 | 생산성 | 👤 사용자 60,000+명



AdBlock Max - 광고 차단기

제공자: AdBlock Pro X

★★★★★ 4,443 | 생산성 | 👤 사용자 90,000+명



AdBlocker Ultimate

제공업체: <https://adblockultimate.net>

★★★★★ 45,459 | 생산성 | 👤 사용자 900,000+명

기존의 광고 차단 확장 프로그램들 on 크롬 웹 스토어

AdGuard AdBlocker가 어떻게 온라인 위협을 방어할까요?

현재 유해 웹사이트 목록을 2만 개가 넘게 보유하고 있습니다. AdGuard는 악성 프로그램을 퍼뜨리는 것으로 알려진 도메인을 차단하여 바이러스, 트로이 목마, 웜, 스파이웨어 및 애드웨어로부터 컴퓨터를 보호합니다. AdGuard는 실제로 바이러스 감염 위험을 효과적으로 낮추며, 잠재적인 공격 위험을 예방하기 위해, 유해한 웹 사이트에 대한 접속을 차단합니다.

Unlimited Adblock은 강력하지만 간단한 광고 차단 도구입니다. 당사의 알고리즘은 모든 광고, 팝업, 데이터 브로커, 비콘, 픽셀 추적기, 히트 맵 추적기와 같은 온라인 추적기의 모든 형식을 차단합니다.

향상된 무료 애드블로커. 모든 광고를 완전히 차단합니다. "용인할 만한" 광고 또는 허용된 광고주 리스트가 없고 사용자 추적과 악성 코드를 차단합니다.

AdBlocker Ultimate은 모든 광고를 차단하고 원하는 콘텐츠에만 집중할 수 있게 할 목적으로 만들어졌습니다.

광고 차단

- Facebook, YouTube 및 기타 웹 사이트의 모든 광고를 차단하십시오.
- 성가신 팝업 차단, 자동 재생 동영상 광고 제거 및 독창적인 오디오 광고 차단

은밀한

- 수많은 추적자가 개인 정보를 침해하는 것을 방지합니다.

안전한

- 바이러스 및 맬웨어 차단 (광고에 숨겨져 있음)

빠른

- Content Blocking Extensions라는 새로운 기술을 사용하면 브라우저 속도를 늦추지 않고도 작업을 수행할 수 있습니다.
- 성능에 영향을주는 느리게로드되는 광고 및 팝업 제거
- 검색 속도가 2 배 이상 빨라졌습니다.

화이트리스트

- 허용된 사이트를 허용하여 좋아하는 사이트를 지원하십시오.

최신 정보

- 끊임없이 변화하는 웹에 따라 차단 목록을 업데이트할 수 있습니다.

기존의 광고 차단 확장 프로그램들 on 크롬 웹 스토어

기존 프로그램들의 광고 차단 원리는 각각 보유한 유해 웹사이트 목록에 기재된 도메인을 무조건 차단하는 것!

이용자에게 노출되는 도메인의 목록인 '화이트 리스트' 기능을 제공하는 프로그램들도 있지만, 이용자가 직접 특정 사이트들을 등록해야 함.

이용자가 직접 자신이 원하는 몇몇 사이트만을 차단 해제하는 것은 번거로운 일이다.

또한 이러한 기능을 사용한다고 하더라도

이용자가 알지 못하는, 잠재적으로 유용한 광고는 차단되므로 접할 수 없기 때문에

이용자는 어떤 광고의 유용성을 판단할 기회를 받지 못하고, 자신에게 유용한 광고조차 보지 못한다.



04

Contents 04

기대 효과

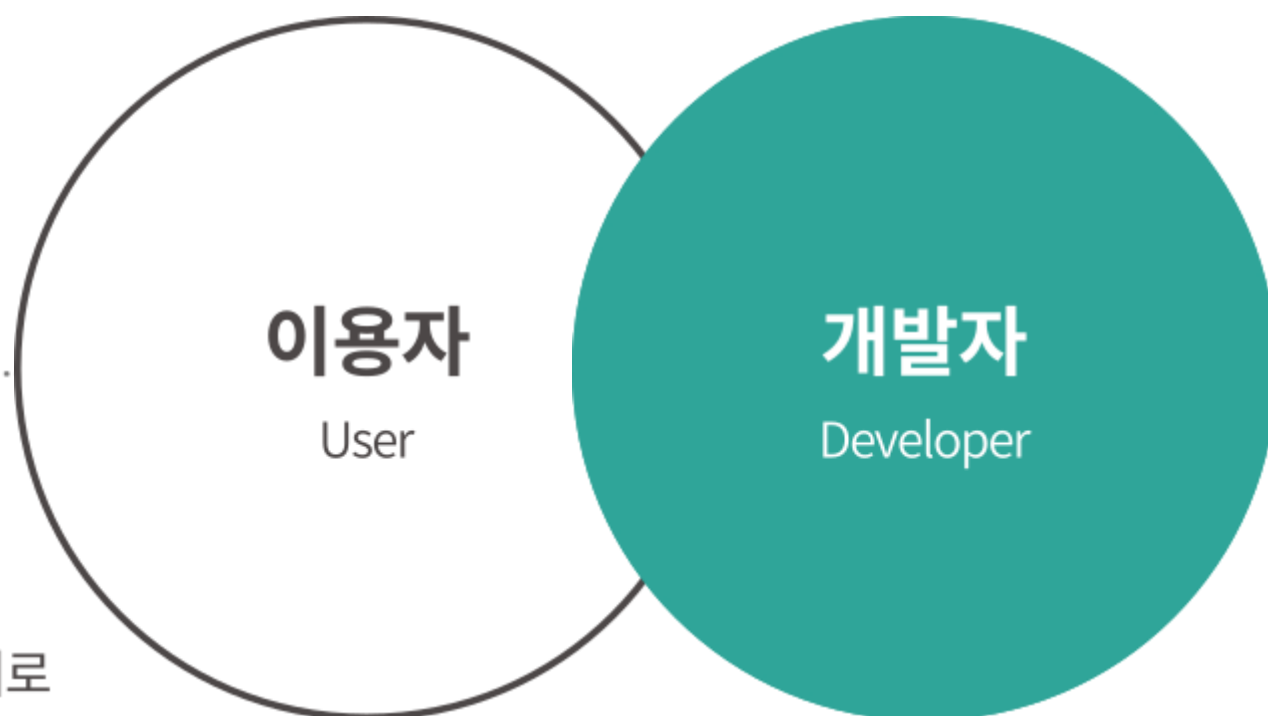
제품을 사용함으로써 얻는 효과

Advantage 01

불필요한 광고 이미지가 제거되어
뉴스 기사 읽기에만 집중할 수 있음

Advantage 02

광고 사이트의 llist가 변경되어도
일일이 갱신할 필요 없음



Advantage 03

각 이미지들을 내부에서 알맞은 카테고리
분류하도록 설계되었으므로, 데이터가 더 쌓이면
광고 이미지가 어떤 광고인지 유의미한 정확도로
판단하는 것이 가능하고, 카테고리에 따라 선택적
으로 광고를 차단할 수 있다.

Advantage 04

딥러닝 기법을 사용하였기 때문에
일정량 이상의 데이터를 확보하여
분류 모델에 학습시켜 놓으면
새로운 데이터를 모델이 자체적으로
잘 처리하므로 유지 및 보수가 간편

05

Contents 05

프로그램 상세 설명

- 프로그램 구현 과정
- 프로그램 실행



프로그램 구현 과정

AI part

- 자료 수집
- 수집된 데이터 분류
- 딥러닝 모델 구현
- 지도 학습
- 세부 요소 값 조정을 통한 성능 개선
- 성능 평가

Web part

- FrontEnd
- BackEnd

AI part

자료 수집

스포츠조선 뉴스 웹페이지에 뜨는
광고 이미지들과 기사 사진들
(이하 비광고 이미지들)을
수집한다.



[광고 이미지 예시]



[비광고 이미지 예시]



수집된 데이터 분류

수집한 이미지 데이터들을
시각적 특징에 따라 아래와 같이
비광고 또는 9개의 광고 카테고리
중 하나로 분류한다.

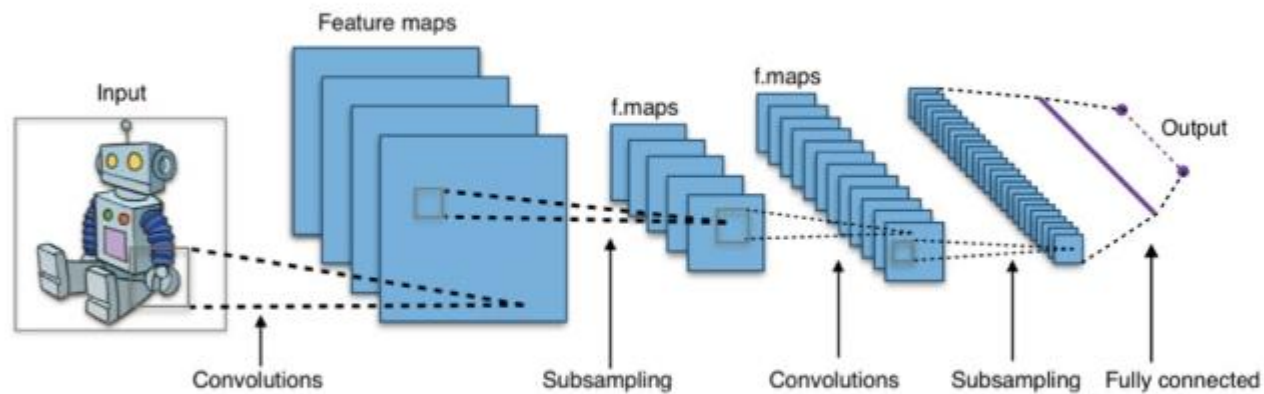
non_ad

- bill
- coin
- general_product
- graph
- human
- lotto_N
- lotto_P
- microbe
- photoshop

딥러닝 모델 구현

파이썬을 사용하여 여러 개의 layer로 이루어진 딥러닝 (CNN) 모델을 구현한다.

CNN



Convolution Neural Network

Source: Wikipedia

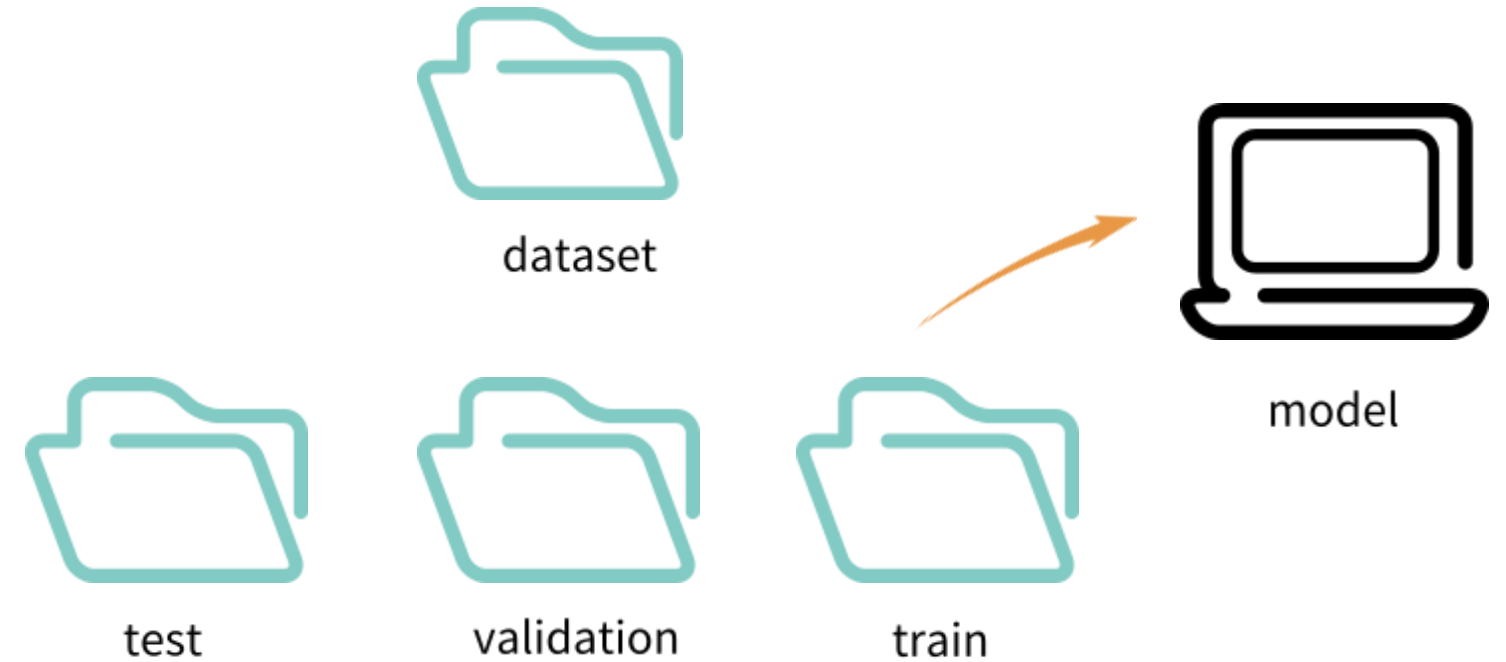
[CNN 구조 예시 이미지]

출처 : <https://medium.com/analytics-vidhya/>

image-classification-a-comparison-of-dnn-cnn-and-transfer-learning-approach-704535beca25

지도 학습

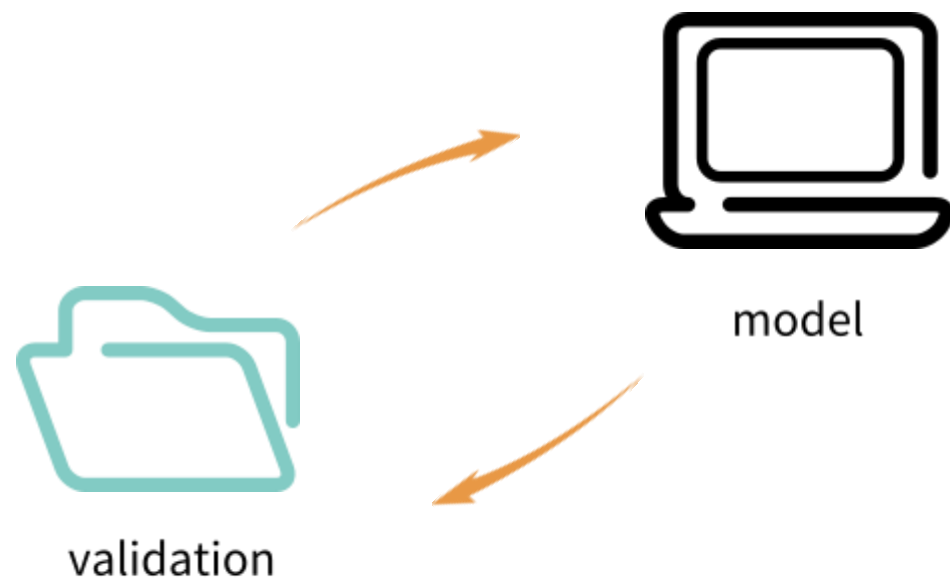
각 카테고리별로 동일한 개수의 이미지를 random하게 선택하여 구성된 이미지 데이터셋을 test, train, validation용으로 분할한다. Train 데이터셋의 이미지들을 모델에게 반복적으로 입력하여 학습시킨다.



AI part

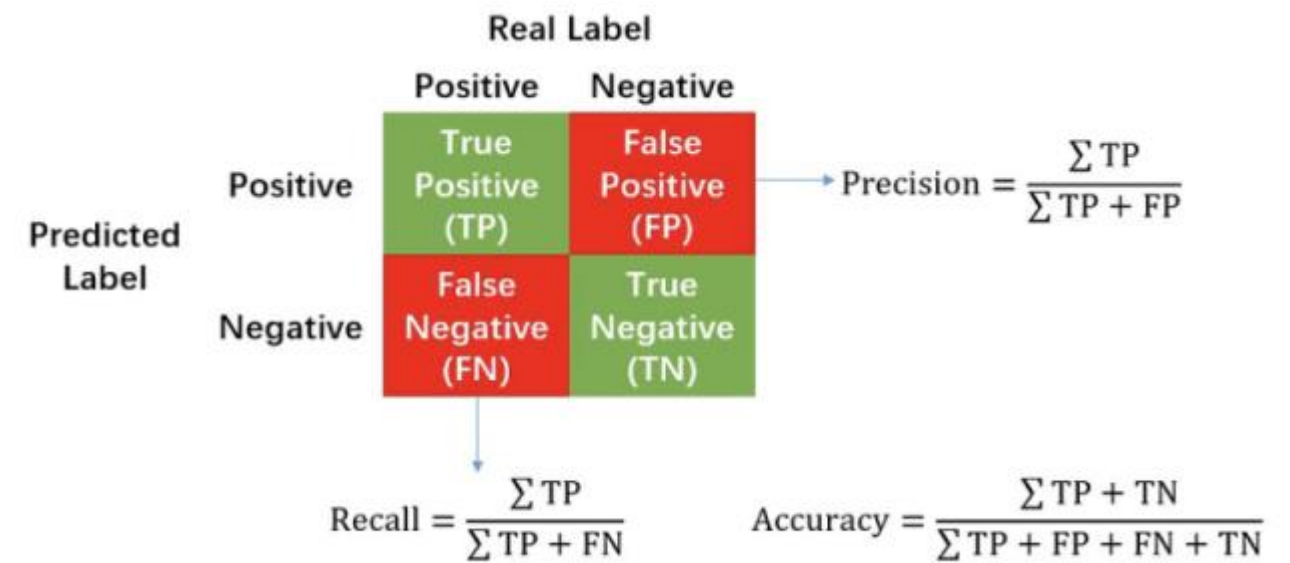
세부 요소 값 조정

모델이 더 좋은 성능을 낼 수 있도록, 학습이 완료된 모델에 validation 데이터셋의 이미지들을 입력하고 출력된 결과에 따라 모델에 포함된 세부 요소들의 값을 조정한다.



성능 평가

모델이 완성되었으면 test 데이터셋의 이미지들을 입력하여 최종적으로 모델의 성능을 평가한다. 이때 정확도 (Accuracy), Precision, Recall, f1 등의 다양한 지표가 사용될 수 있다.



Web part

FrontEnd

Chrome Extension을 이용해 스포츠조선의 뉴스 페이지에서 Content Script를 실행한다.

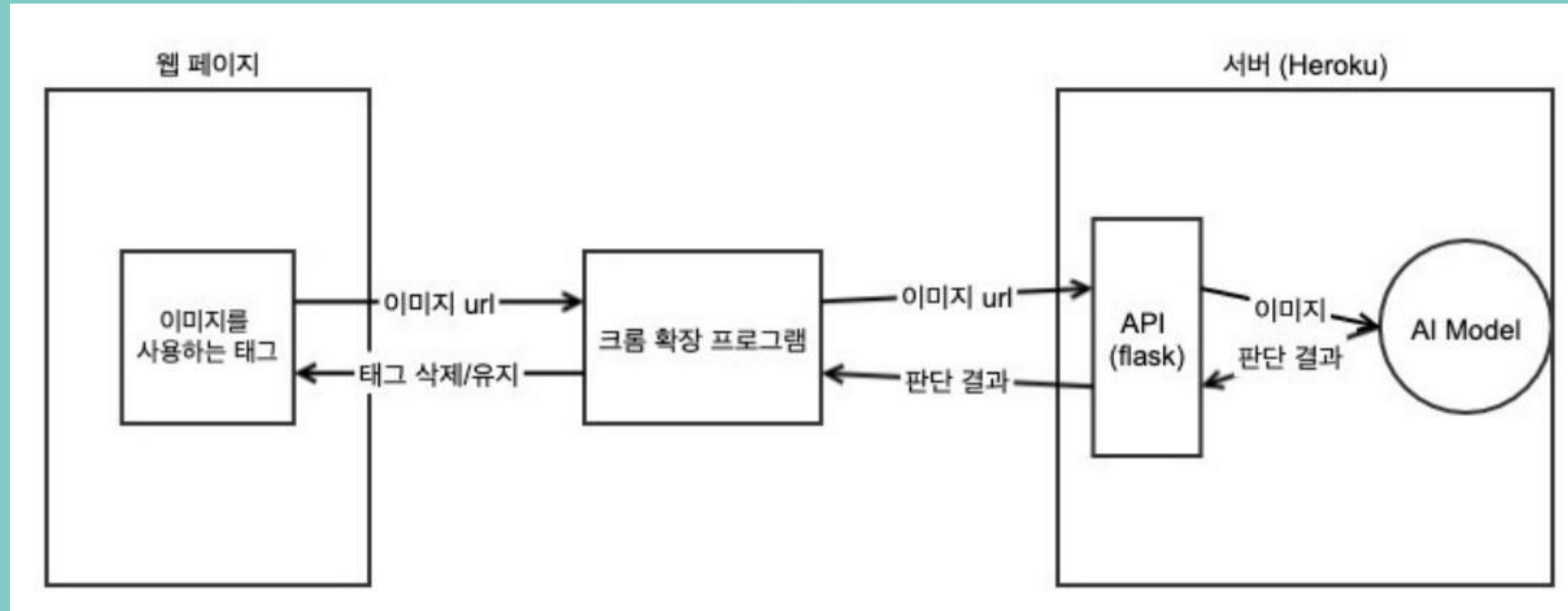
img 태그와 같은 태그에서 사용하는 이미지의 url을 수집하여 모델로 전달한다.

모델의 분류 결과에 따라 해당 태그를 제거하거나 유지한다.

BackEnd

flask를 이용해서 api를 구현하고, heroku를 이용해 배포한다.

Chrome Extension으로부터 전달받은 이미지의 url을 열어서 모델에 입력하고 분류 결과를 받아 Extension에 전달한다.



프로그램 실행

프로그램 설치 및 실행 방법

01

필요한 파일들을 로컬 디스크에 다운 받고
구글 크롬 브라우저를 켜 후 우측 상단에서 확장
프로그램 아이콘을 찾아 클릭한다.

03

확장 프로그램을 찾아 로드하고 활성화시킨다.

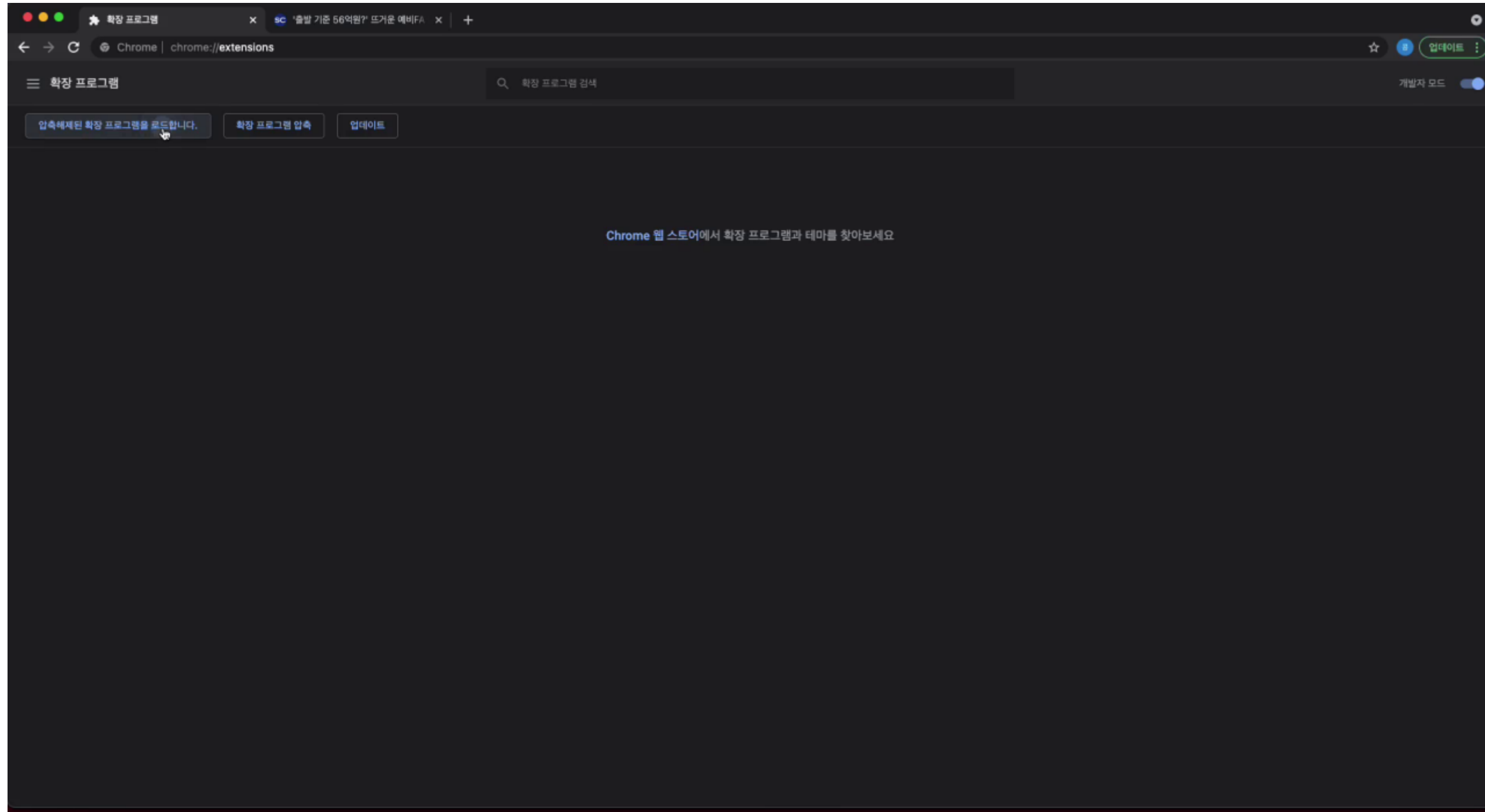
02

확장 프로그램 관리 페이지에서 '압축 해제된 확장
프로그램을 로드합니다' 메시지를 클릭한다.

04

스포츠 조선 사이트에 접속해 기사를 클릭하면
잠시 후 광고가 제거된 것을 확인할 수 있다.

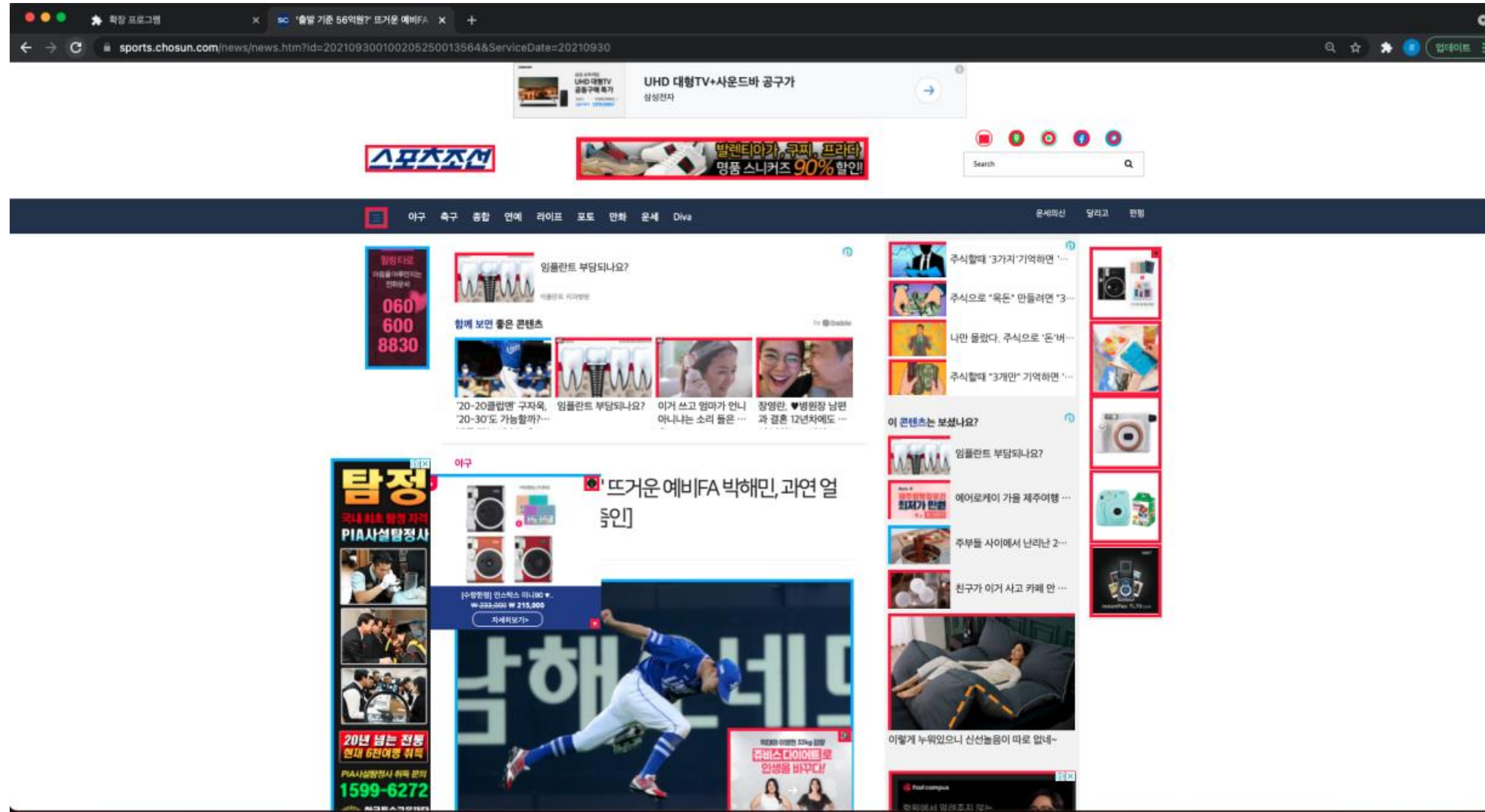
프로그램 실행 영상



스포츠조선 사이트 메인

<https://sports.chosun.com/>

프로그램 실행 결과



프로그램이 **광고**로 분류한 이미지는 **빨간색** 테두리로,
프로그램이 **비광고**로 분류한 이미지는 **파란색** 테두리로 표시하였다.



06

Contents 06

발전 가능성

제품의 발전 가능성

추후 여러 기능을 추가하여 제품의 발전과 성능 향상을 도모할 수 있다.



광고 카테고리별 분류

입력받은 이미지가 어떤 카테고리의 광고인지 판단하여, 이용자는 자신이 원하는 카테고리의 광고 이미지만 볼 수 있다.



텍스트 인식

텍스트가 포함된 이미지들에 있는 단어를 인식해 정확도 향상.



속도 향상

웹페이지에서 이미지가 제거되는 데 걸리는 시간 단축.



서비스 대상 플랫폼 추가

스포츠 조선 이외의 다른 뉴스 사이트에서도 서비스를 제공.

Q & A

감사합니다.