중간보고서

[문서 부제]

천승현

[연도]

목차

[1.개발동기 1](#_Toc7383784)

[1.1 동기 1](#_Toc7383785)

[1.2 개발제목 1](#_Toc7383786)

[2. FCMW개발내용 1](#_Toc7383787)

[2.1 패션 데이터 선정 1](#_Toc7383788)

[2.2 api 1](#_Toc7383789)

[2.3 FCMW 웹 사이트 1](#_Toc7383790)

[3. FCMW 세부내용 2](#_Toc7383791)

[3.1 일정표 2](#_Toc7383792)

[3.2 개발환경 2](#_Toc7383793)

[3.3 개발상황 2](#_Toc7383794)

[3.4 개발예정 2](#_Toc7383795)

[3.5 협업방식 2](#_Toc7383796)

[3.6 팀원 소개, 역할 및 소감 2](#_Toc7383797)

# 1.개발동기

## 1.1 동기

추천 시스템(recommend system)은 현대 상업시장에서 가장 강력하고 필수적인 전략 중 하나이다. 사용자(User)는 모든 아이템들을 다 둘러볼 시간이 없고, 자신에게 가장 필요로 하는 상품을 추천 해 주길 원한다. 추천 시스템을 적용한 웹 쇼핑사이트들은 자신들의 알고리즘을 이용하여 사용자들에게 적합한 아이템을 추천해준다. 사용자들이 기존에 구매했던 상품들의 속성이나, 평점을 분석하여 관련 있는 상품들을 추천 리스트로 작성해 보여준다.

<아마존 예시>

이러한 추천 시스템을 적용한 웹시스템을 개발하기 위해 빅데이터 분야의 요소를 접목시킨다. 빅데이터 수집단계를 통해 데이터를 수집하고, 분석단계를 통해 데이터를 파싱, 정규화 하며 활용단계에서 수집한 데이터들을 이용하기로 한다. 이렇게 빅데이터의 요소를 웹 사이트와 알고리즘을 구현하며 직접 학습해보기로 한다.

## 1.2 개발제목

20대 남녀를 위한 패션 코디 추천 웹 사이트 개발 (FCMW)

Fashion coordy   
for 20’s man & woman   
in Website

- FCMW

# 2. FCMW개발내용

## 2.1 패션 데이터 선정

빅데이터 수집

- 수작업

빅데이터 수집을 위한 방법으로 하나인 수작업으로 데이터를 수집하는 법은 FCMW에서 필요한 상의, 하의, 신발 등의 사진 등을 직접 검색 엔진, 여러 사이트에 접속하여 직접 사진을 다운로드 받아 데이터로써 변형 작업을 거치는 방법이다. 하지만 FCMW에 맞는 데이터를 수집을 위해서는 수작업은 불가능 할 것으로 판단하여 새로운 방법인 웹 크롤러라는 방식으로 대체한다..

- 웹 크롤러

웹 크롤러(Web Crawler)는 조직적, 자동화된 방법으로 월드 와이드 웹을 탐색하는 컴퓨터 프로 그램이다. 웹 크롤러가 하는 작업을 '웹 크롤링'(Web Crawling) 혹은 '스파이더링'(Spidering)이라 부른다. 검색 엔진과 같은 여러 사이트에서는 데이터의 최신 상태 유지를 위해 웹 크롤링한다. 웹 크롤러는 대체로 방문한 사이트의 모든 페이지의 복사본을 생성하는 데 사용되며, 검색 엔진은 이렇게 생성된 페이지를 보다 빠른 검색을 위해 인덱싱한다. 또한 크롤러는 링크 체크나 HTML 코드 검증과 같은 웹 사이트의 자동 유지 관리 작업을 위해 사용되기도 하며, 자동 이메일 수집과 같은 웹 페이지의 특정 형태의 정보를 수집하는 데도 사용된다. FCMW에서 사용하게 될 데이터 수집을 위해 구글에서 패션 코디와 관련 있는 데이터를 크롤링한다.

빅데이터 저장 및 분석

-ImageAI를 이용하여 분석한 정보를 DB에 저장

-SQLite3

DB에 저장하기 위해 사용할 SQLIite3는 별도의 서버 프로세스가 필요 없고 SQL질의 언어의 비표준 변형을 사용하여 데이터베이스의에 액세스할 수 있는 경량 디스크 기반 데이터베이스를 제공하는 C 라이브러리이다. 일부 응용 프로그램은 내부 데이터 저장을 위해 SQLite를 사용할 수 있습니다. SQLite를 사용하여 응용 프로그램을 프로토타입 한 다음 PostgreSQL 이나 Oracle과 같은 더 큰 데이터베이스로 코드를 이식할 수 있다.

https://www.sqlite.org/index.html

-ImageAI

DeepQuiest AI 사의 CEO인 Moses Olafenwa가 자신의 형제와 함께 개발한 통합 라이브러리로 자유명한 이미지 인식 network(SqueezzeNet 캘리포니아 버클리 대학, 스탠포드 대학, ResNet50 마이크로소프트, InceptionV3 구글, DenseNet121페이스북)들을 한 곳으로 묶어 약 1400만 이미지를 미리 트레이닝하여 사용할 수 있고 자신만의 이미지로 학습시켜 나만의 모델을 생성 가능 하게 한다.

http://imageai.org/

https://github.com/OlafenwaMoses/ImageAI

시각화

-ANACONDA PYTHON Ver 3.7

오픈 소스인 ANACONDA는 Python/R data science 그리고 머신 러닝을 윈도우나 리눅스 맥OS 등에서 가장 쉽게 사용할 수 있게 해주는 프로그램으로 데이터를 분석하여 시각화 하기 위해 사용한다.

https://www.anaconda.com/distribution/

-Django

장고는 파이썬으로 작성된 오픈 소스 웹 애플리케이션 프레임워크로 FCMW 웹 개발에 있어서 번거로운 요소들을 새로 개발할 필요 없이 내장된 기능만을 이용해 빠른 개발을 할 수 있기에 사용한다.

데이터 시각화는 FCMW 코디 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하고 전달되는 과정을 말한다. FCMW데이터 시각화의 목적은 최종 코디 결과 사진을 통해 명확하고 효과적으로 전달 하는 것이다.

<https://www.djangoproject.com/>

-사용하는 데이터

FCMW 웹 구현을 위해서는 20대 남녀를 위한 코디를 추천하기 위해 방대한 양의 종류별 의류의 사진, 신발 등의 사진을 필요로 한다. 의류에 대한 데이터로는 의류의 종류(ex.셔츠, 티셔츠, 후드티셔츠, 맨투맨, 블라우스, 니트, 가디건, 코트, 재킷) 이러한 카테고리 등으로 나누어 데이터를 수집한다. 수집한 데이터를 FCMW 웹 구현을 위해 사용한다.

- 데이터 크롤링 방법

FCMW 웹 구현을 위해 필요한 사진 데이터들을 가져오기 위하여 선택 한 웹 크롤링 방식을 사용하기 위해 ANACONDA Python Ver 3.7에서 코드를 작성해야 한다.

from google\_images\_download import google\_images\_download #importing the library  
  
response = google\_images\_download.googleimagesdownload() #class instantiation  
  
arguments = {"keywords":"Polar bears","limit":5,"output\_directory":'image',"print\_urls":True} #creating list of arguments  
paths = response.download(arguments) #passing the arguments to the function  
print(paths) #printing absolute paths of the downloaded images

구글에서 이미지를 받아오는 방법인 웹 크롤러를 사용하기 위해 파이썬 코드를 작성한다.

- 데이터 분석, 활용 방법

데이터 분석

웹 크롤러를 통해 구글에서 받아온 의류의 종류에 따라 크롤링 해온 비정형 데이터(unstructured data, 유동 데이터, 미리 정의된 데이터 모델이 없거나 미리 정의된 방식으로 정리되지 않은 정보를 말한다. Ex. 동영상, 사진, SNS의 텍스트) 이러한 비정형데이터를 사용하게 되면 프로그램을 사용하여 이해하는 것을 불가능하게 만든다. 이러한 이유 때문에 정형 데이터(기존 데이터베이스에 저장이 되는 ‘규격화’된 데이터 Ex. 엑셀파일, DB테이블)로 변환이 필요하다.

데이터 활용

데이터 분석 과정을 통해 얻은 FCMW 웹 구현을 위한 정현 데이터인 의류나 신발 데이터를 의류 중에서도 원피스, 셔츠, 팬츠, 티셔츠 등의 키워드를 통해 FCMW 웹에서의 사용자의 취향에 맞는 의류나 신발 등을 보여 줄 수 있다.

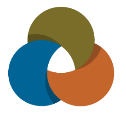
## 2.2 api

Api란

Application Programming Interface(응용 프로그래밍 인터페이스)로, 프로그램에서 사용할 수 있도록 운영 체제나 프로그래밍 언어가 제공하는 기능을 제어 할 수 있게 만든 인터페이스이다. 패션 이미지들은 비정형데이터로, 용량이 크기 때문에 그대로 사용하기에는 무리가 있으므로 이미지 분석 API를 이용하여 다루기 쉬운 데이터로 변환한다.

Google cloude vision api는

구글 포토에서 쓰이고 있는 서비스로, 머신 비전(Machine Vision) 즉, 머신 러닝(기계학습)을 이용한 이미지 인식 기술의 한 종류이며 구글에서 개발되었다. 또한, 누구나 API를 사용해 자신의 앱, 서비스, 하드웨어에 구글의 기술을 적용 가능하다. 클라우딩 기반으로 모바일이나 성능이 낮은 디바이스에서 구현 가능하며, 높은 수준의 인식률로 오류 비율이 사람이 분류했을 때와 비슷한 수준인 6.65% 으로 사람의 얼굴 이미지에서 감정 읽어내기, 브랜드,언어, 경치, 인공 구조물, 부적절 컨텐츠 인식 가능하다. 그러나 기본으로 제공되는 정도를 사용하면 유료로 변환이 된다는 점 그리고 기능을 구현하기 위한 함수 코드의 복잡성을 가지고 있으며 필요하지 않은 기능들이 많다는 단점을 가지고있다.

ImageAI 

DeepQuiest AI 사의 CEO인 Moses Olafenwa가 자신의 형제와 함께 개발한 통합 라이브러리 이다.유명한 이미지 인식 network들을 한곳으로 묶어 사용하기 쉽게 만들어 놓았다. ImageAI의 학습 모델은 캘리포니아 버클리 대학, 스탠포드 대학에서 사용한 SqueezzeNet, 마이크로소프트사에서 사용한 ResNet50, 구글에서 사용한 InceptionV3, 페이스북엣 사용하는 DenseNet121가 있다. 1400만 이미지를 미리 트레이닝 시킨 모델을 사용할 수 있으며, 자신만의 이미지로 학습시켜 나만의 모델을 생성가능하다.

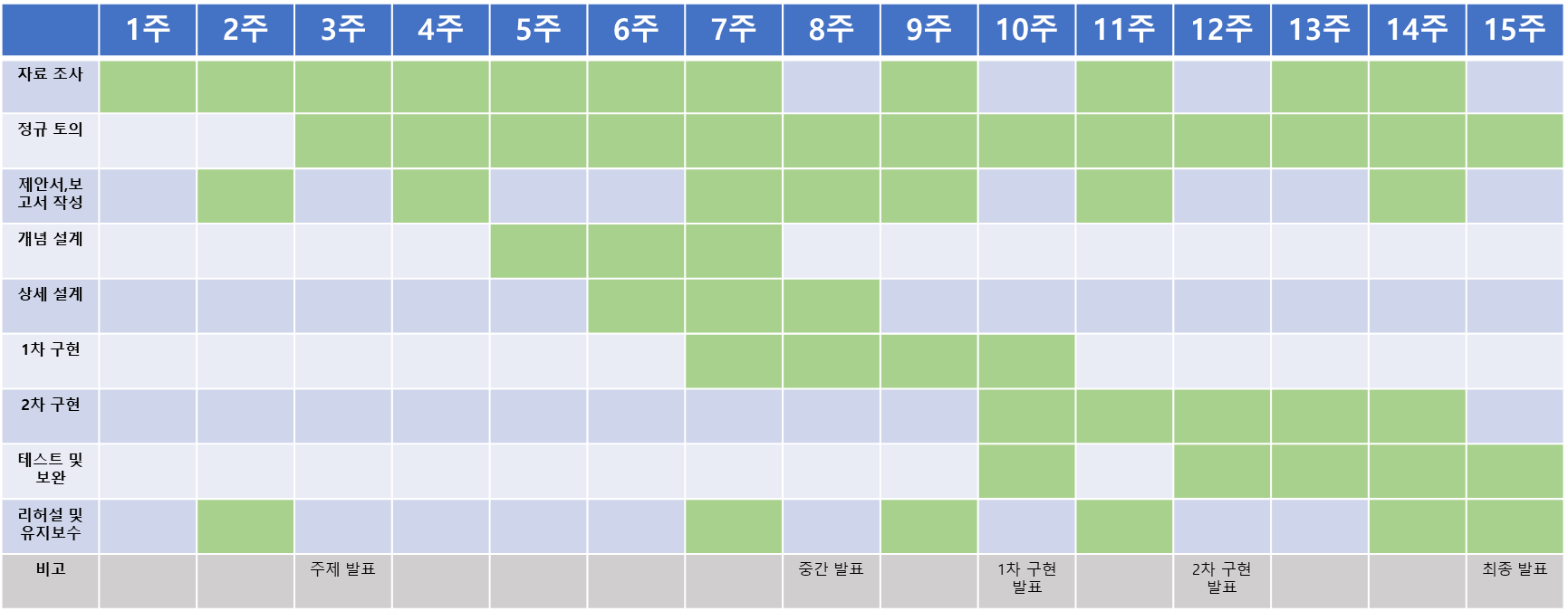
## 2.3 FCMW 웹 사이트

- django를 이용한 웹사이트 개발

# 3. FCMW 세부내용

## 3.1 일정표

- 일정표



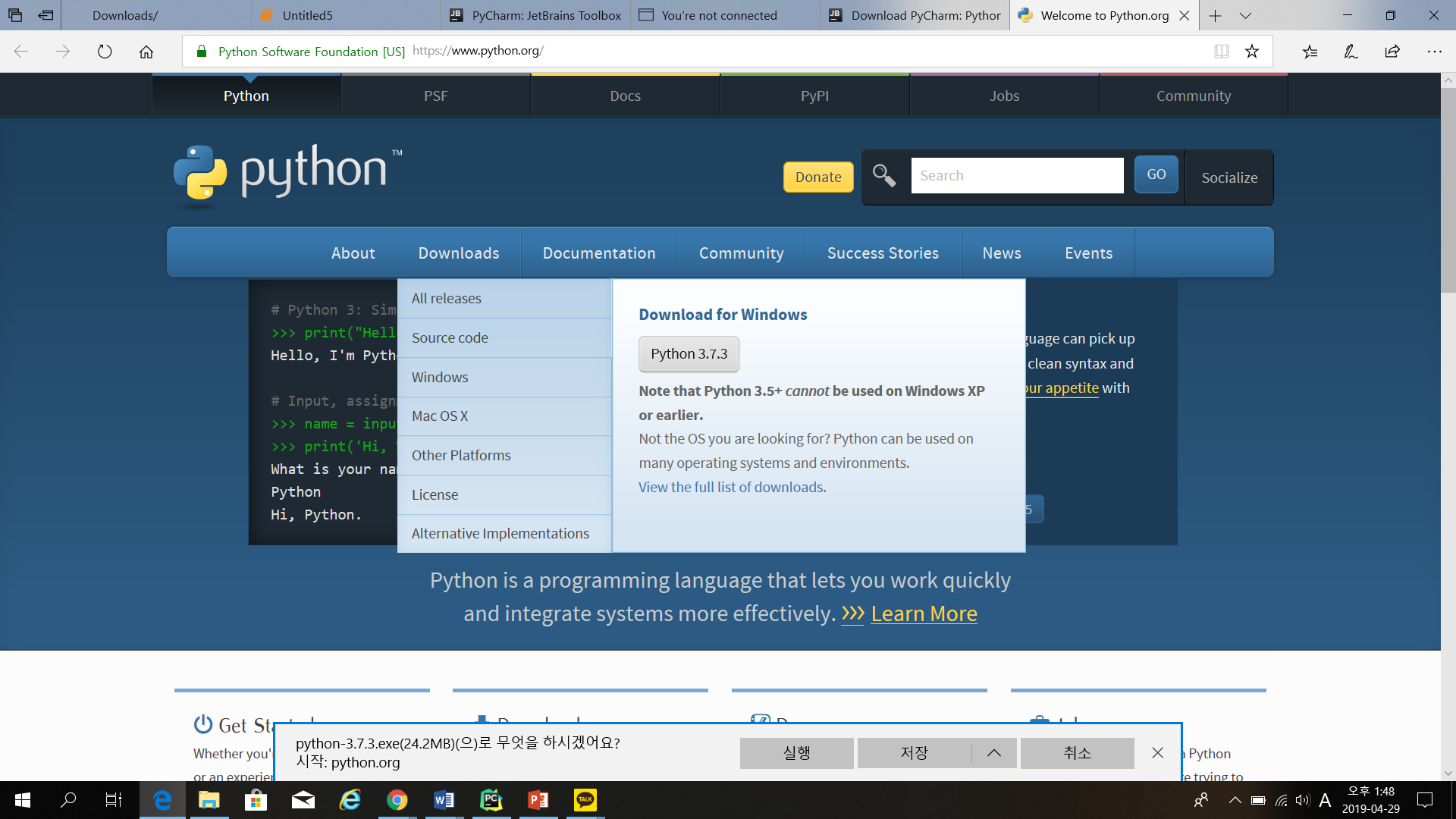
## 3.2 개발환경



- 파이썬 설치

파이썬 -> 플랫폼이 독립적이고 인터프리터식으로 만들어져 있으며, 객체지향적, 동적 타이핑 대화형 언어로 사용된다.데이터분석에 특화된 언어이며 ,장고를 분석해 웹사이트를 만들 때도 파이썬을 사용하기 때문에 설치를 해준다.

설치 방법은 파이썬 다운로드링크는 <https://www.python.org/> 이고,링크에 들어가

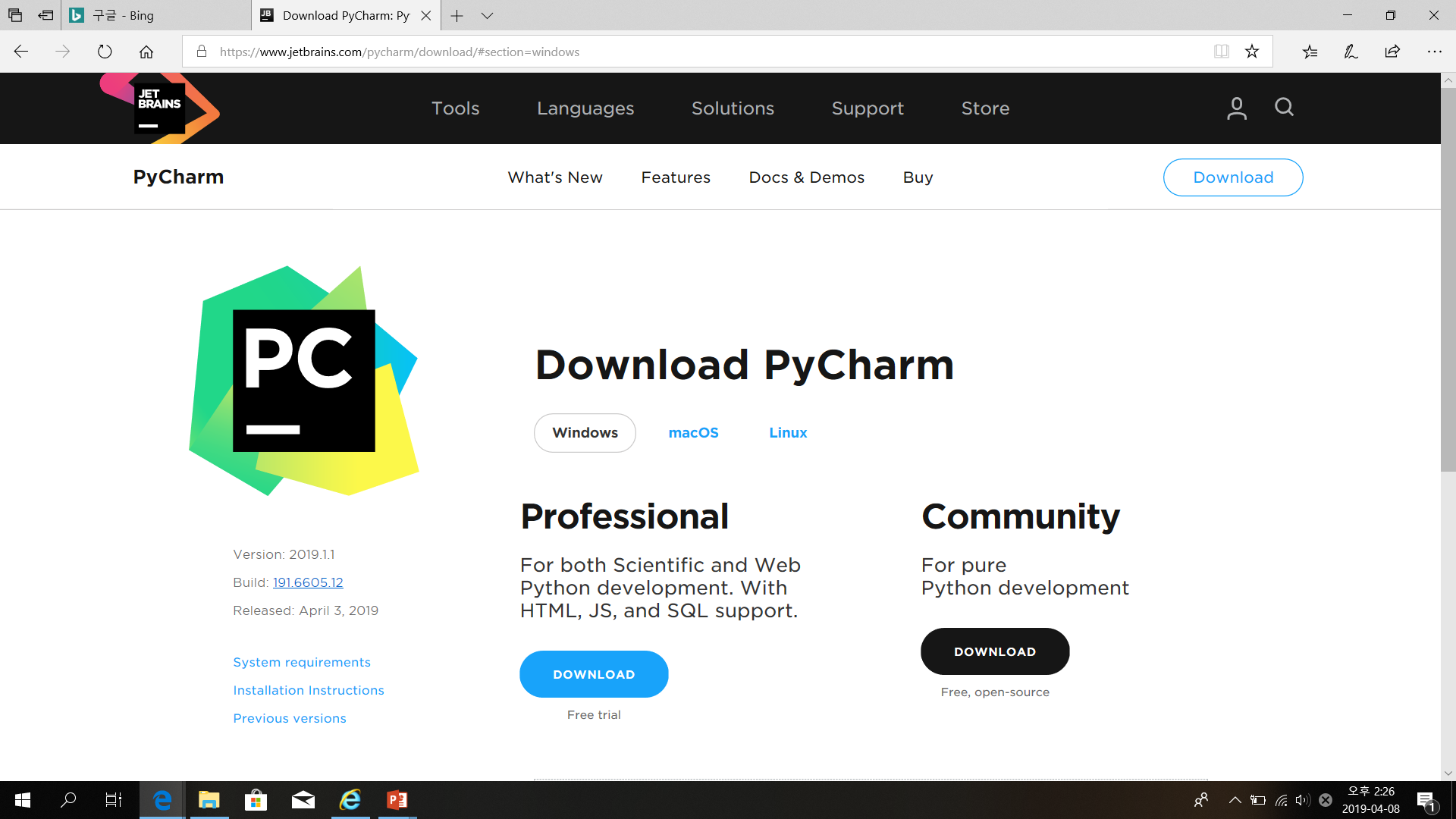


위 그림과 같은 페이지에 접속된다.접속후 downloads카테고리를 선택한후 자신의 환경을 선택하여 python 3.7.3 을 클릭하여 파이썬을 실행시켜준다.

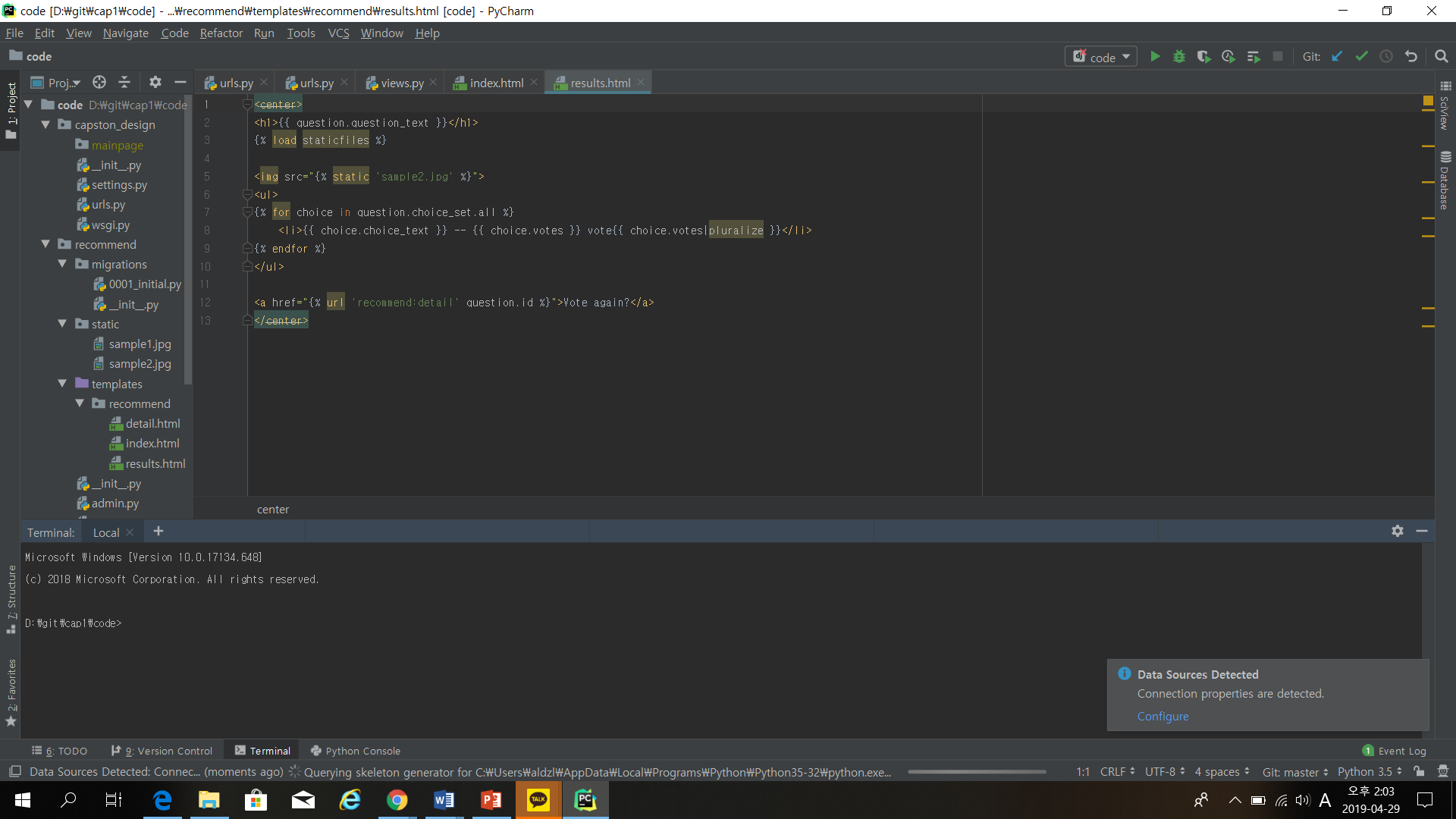
- 파이참 설치 

파이참은 Python 개발에 필요한 모든 도구를 제공해 주는 역할을 한다.파이참은 다운로드시 별도의 제약 없으며 다운로드후 ->power shell 에서 장고 설치하기위해 pip install Django 를 입력하며 개발환경을 만들어준다.

다운로드 방법은 [https://www.jetbrains.com/pycharm/download/#section=windows](https://www.jetbrains.com/pycharm/download/) 링크에 접속하여 화면에 보이는Download를 클릭하여 Community Download를 눌러준다. Professional Download 를 누를시 시간이흐른뒤 유료화 되기 때문에 community를 다운받아준다.

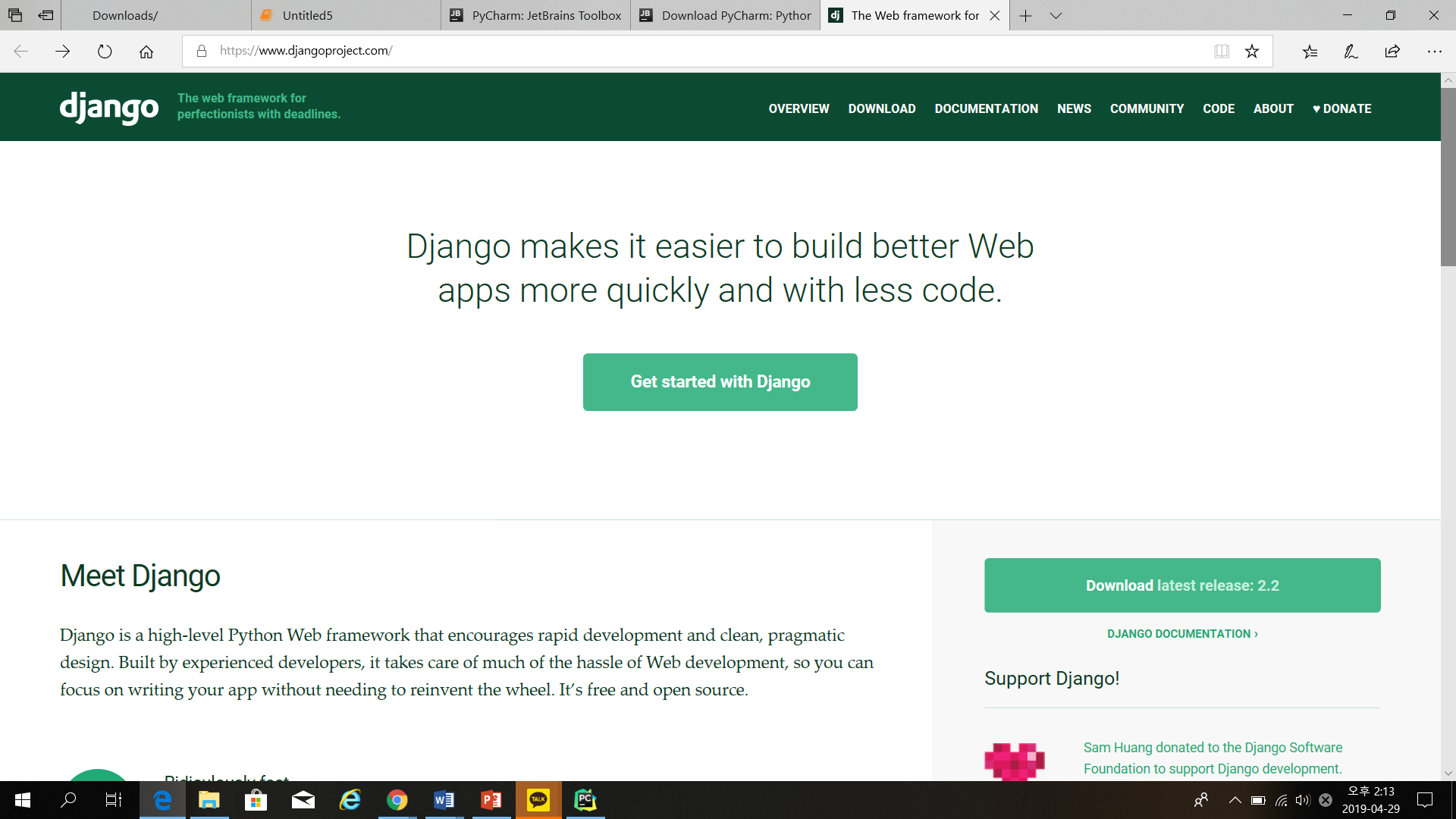


다운로드가 완료되고 실행시키면 pycharm 실행화면이 나온다.



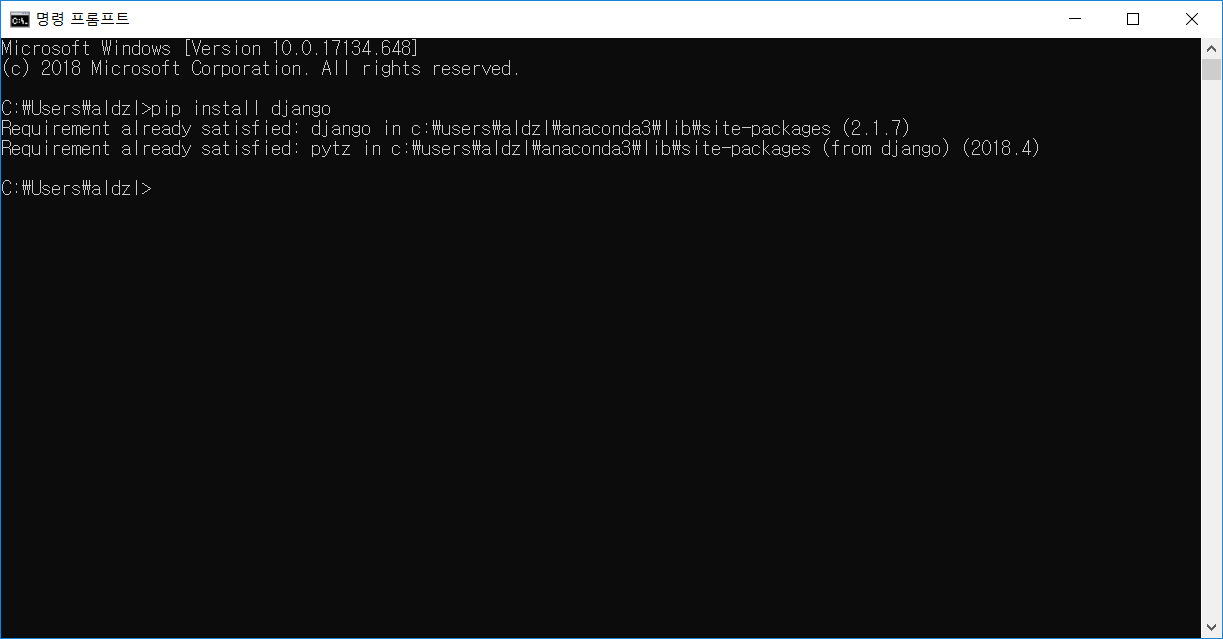
- 장고 

장고는 파이썬 웹 프레임워크로 파이썬을 이용하여 웹페이지를 쉽게 구성하기 위해 사용하였다.

장고 다운로드를 위해 <https://www.djangoproject.com/> 링크를 타고 들어가시면, 

위와 같은 화면을 볼수 있다.

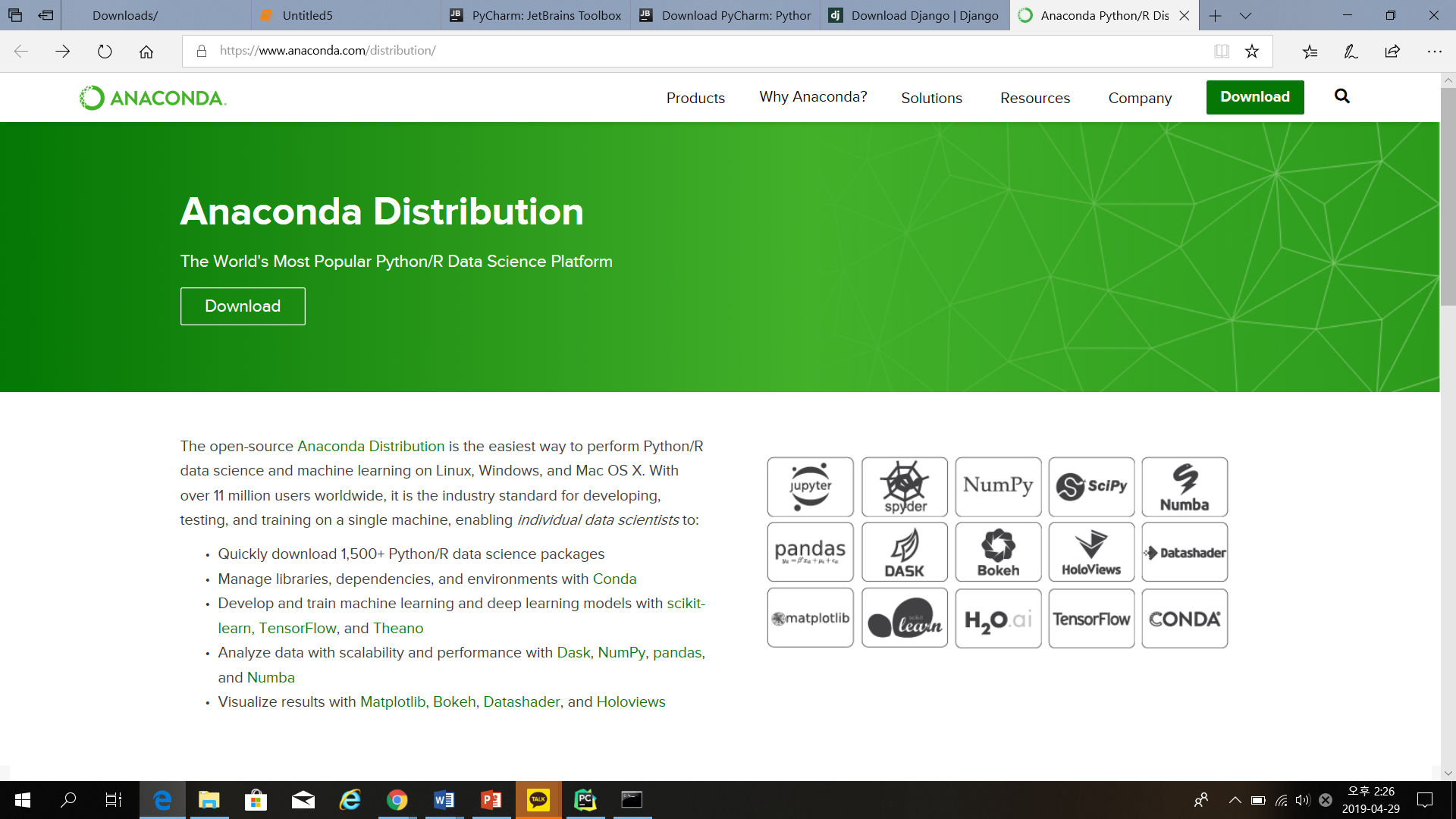
get started with Django를 클릭한 후, Download version 2.2를 클릭하여 다운받아 준다. 또다른 방법으로는 cmd창 에서 사진과 같이 >pip install django명령어를 입력해주면 다운로드 홈페이지에 접속하지 않아도 바로 설치하여 사용 가능하다.

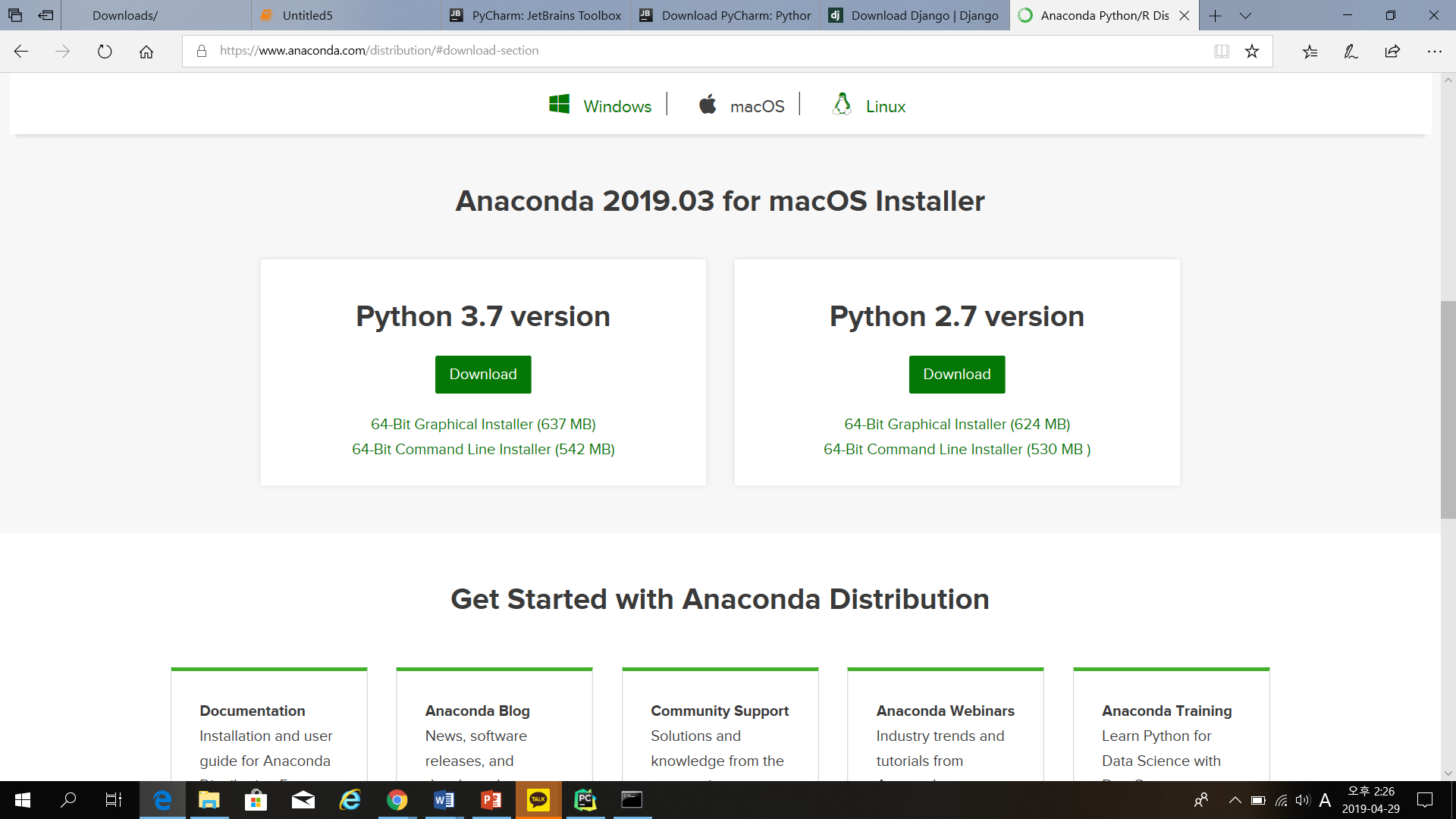


이미 설치되어 있으면 cmd창에 Requirement already satisfied 라는 결과가 나오게 된다.

- 아나콘다 네비게이터 ,

**Anaconda** 는 데이터 분석용으로 사용되며, 패키지관리를 단순화하고 과학 컴퓨팅을위한 Python및 R 프로그래밍 언어를 위한 무료 및 오픈소스 배포판이다. **아나콘다 설치를 위해** [**https://www.anaconda.com/distribution/**](https://www.anaconda.com/distribution/) **링크에 클릭해주면,** 위와 같은 페이지에 접속이 된다.



여기서 Download버튼을 클릭 하고 각자 필요한 version을 클릭하여 다운로드하여 설치해준다.

## 3.3 개발상황

- 개발 현황, 실행화면, 코드

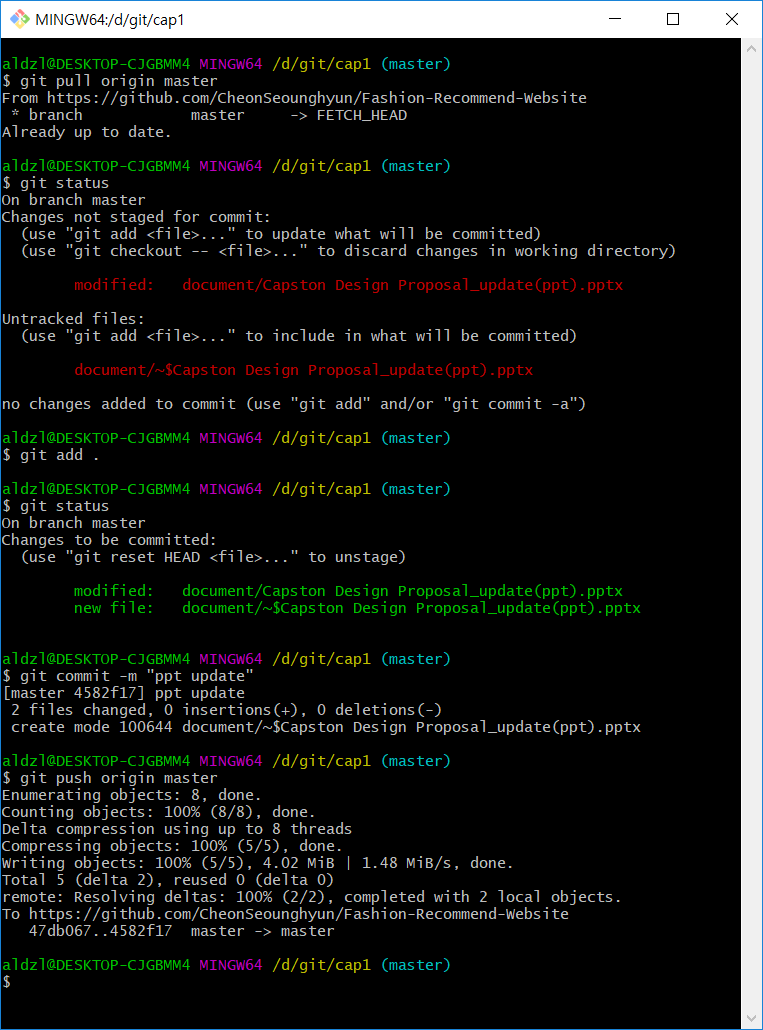
## 3.4 개발예정

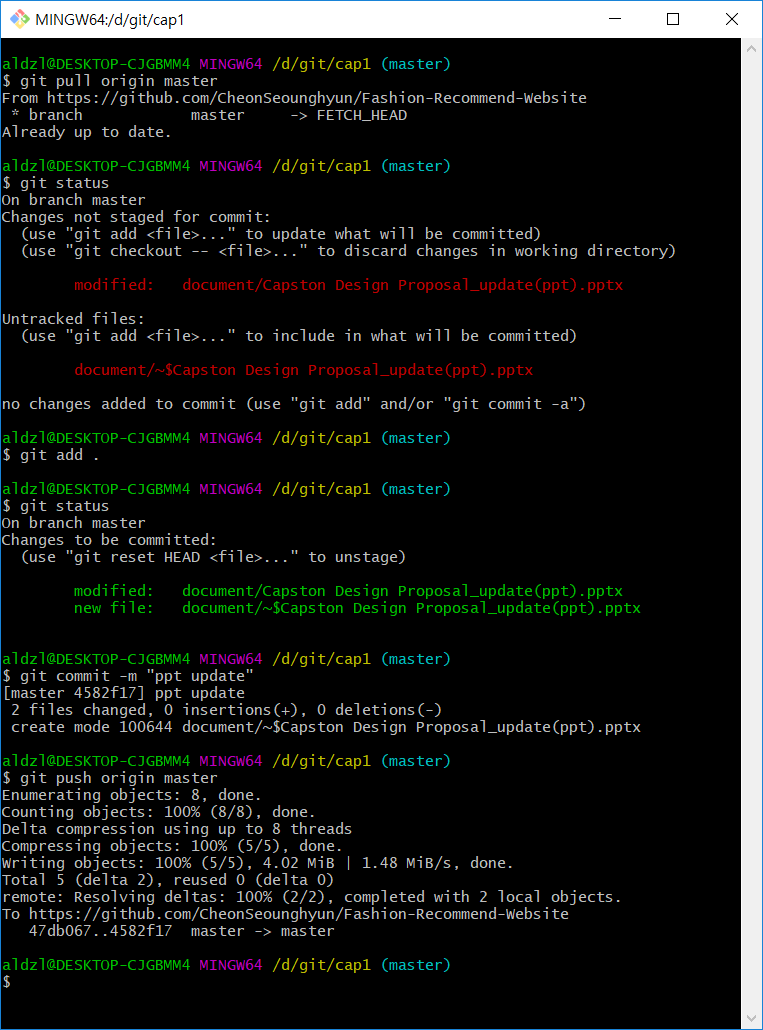
## 3.5 협업방식

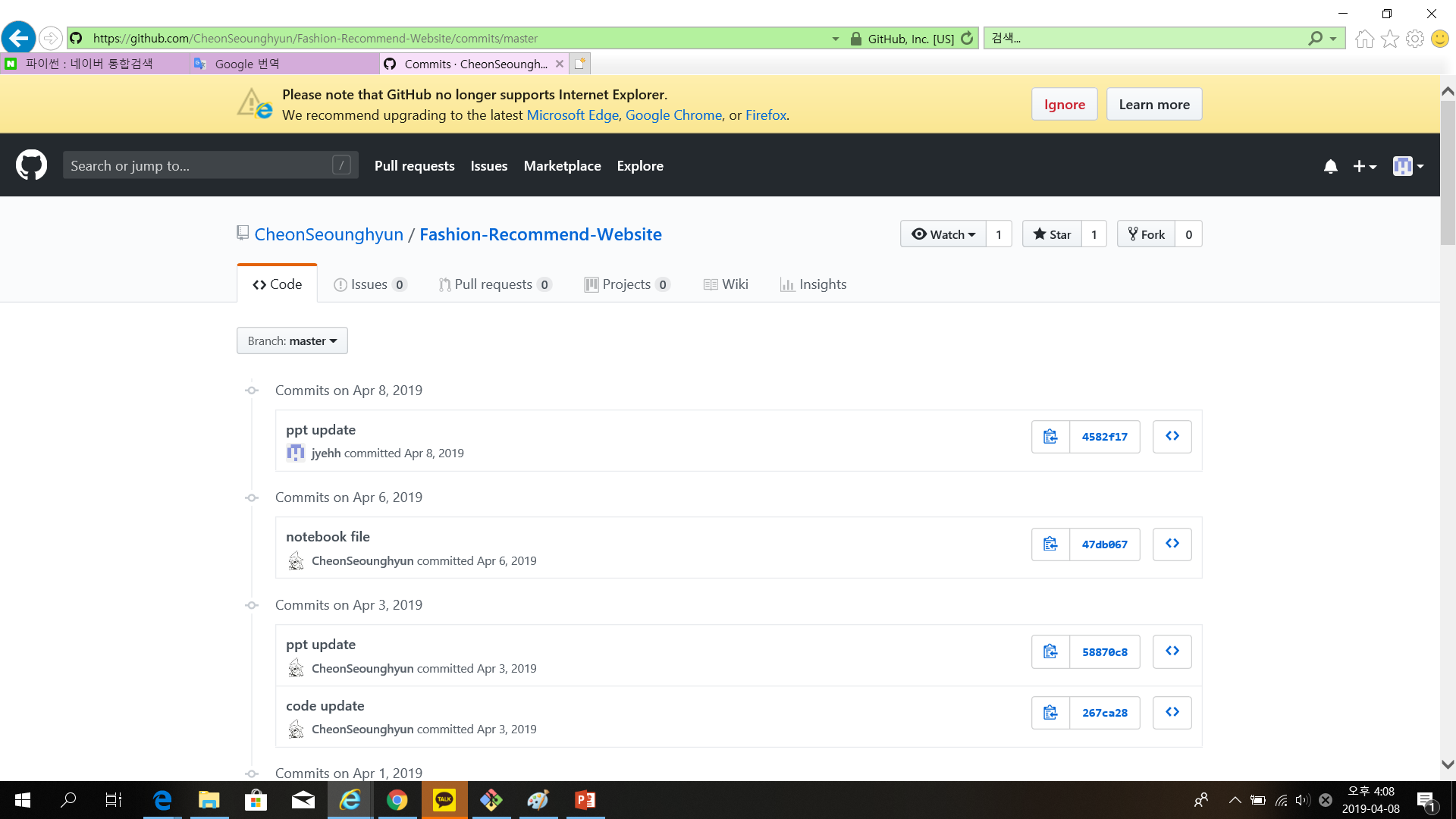
- 깃허브

‘깃’(Git)은 전문 호스팅 업체이다. 깃허브는 오픈소스 소프트웨어의 중심지(hub) 역할을 하면서 오픈소스 프로젝트가 널리 퍼지는 데 크게 기여하고 있다. 깃허브는 깃을 보다 편하게 이용할 수 있게 만든 호스팅 서비스다. 깃은 명령어를 입력하면서 이용해야 하지만 깃허브는 웹 그래픽 기반으로 깃을 이용할 수 있는 환경을 제공했다. 깃은 오픈소스 소프트웨어로 2005년에 개발된 분산형 버전관리 시스템(DVCS)이다. 포크는 내 계정으로 외부 프로젝트 코드 저장소를 그대로 복사해주는 것을 말한다. 오픈소스 프로젝트들은 수십명에서 수백명의 사람들이 사람들이 함께 소스코드를 고칠수있다. 위와 같은 이유로 깃허브를 사용하였다.

깃허브 사용시 알아야할 명령어







$git pull origin master는 원격 저장소에서 파일을 가져오는 명령어 로 파일 내용을 최신화 시킨다.여기서 origin 은 원격 저장소 이름으로 링크를 간략화 시킨것이고, master는 버전을 나타낸다. $git status는 파일들의 변경 상태를 확인해준다. 변경된 파일은 빨간색으로 나타나고, 변경사항이 없을 시 초록색으로 나타난다. $git add .에서 .은 디렉토리 전체를 의미 하며, 깃허브에 변경된 파일을 등록해준다. $git commit –m “----” 는 파일 업로드 명령어로, “ “ 안에 들어가는 이름으로 파일이 업로드 된다. $git push origin master라는 명령어로 원격 저장소인 깃허브 사이트에 전체적인 내용이 업로드된다.

프로젝트 저장소 링크 [https://github.com/CheonSeounghyun/Fashion-Recommend- Website](https://github.com/CheonSeounghyun/Fashion-Recommend-Website) 조원별 깃허브 id, 저장소 링크

<https://github.com/LeeYooKyum> LeeYooKyum , <https://github.com/choiyesom> choiyesom, <https://github.com/jyehh> jyehh

## 3.6 팀원 소개, 역할 및 소감

- 각자 소개와 역할, 소감 작성( 5줄 이상)