

# **비정형 빅데이터 분석의 응용과 실습**

**Week-05. Practice and Assignment**

**서중원 2020.10.17**

# 1. EasyOCR 활용하기

## Option1. Pytorch 재설치

```
{sys.executable} -m pip install torch==1.6.0+cpu torchvision==0.7.0+cpu -f https://  
download.pytorch.org/whl/torch\_stable.html
```

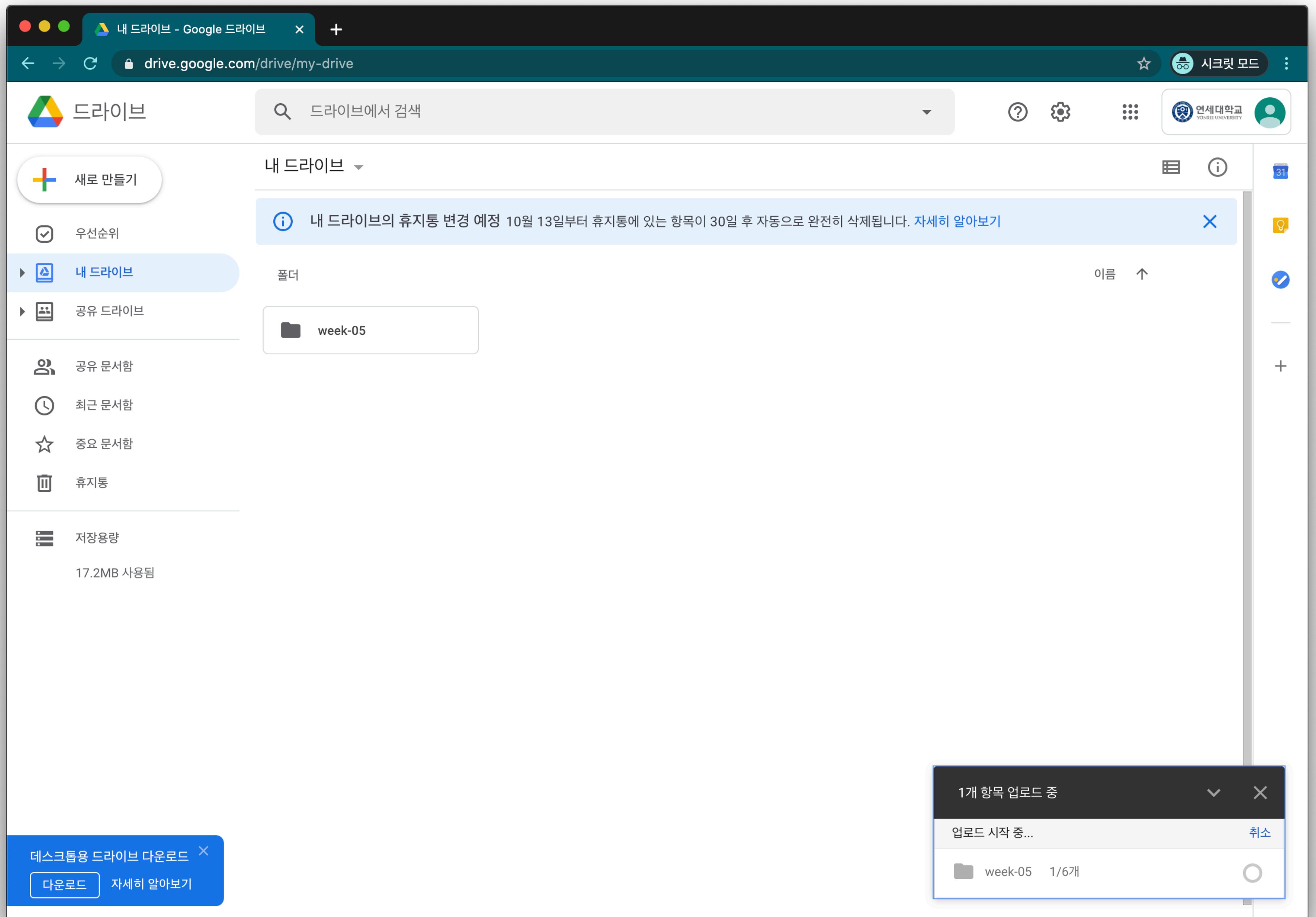
```
import sys  
{sys.executable} -m pip install torch==1.6.0+cpu torchvision==0.7.0+cpu -f https://download.pyt  
{sys.executable} -m pip install easyocr
```

## Option2. Google Colab 활용

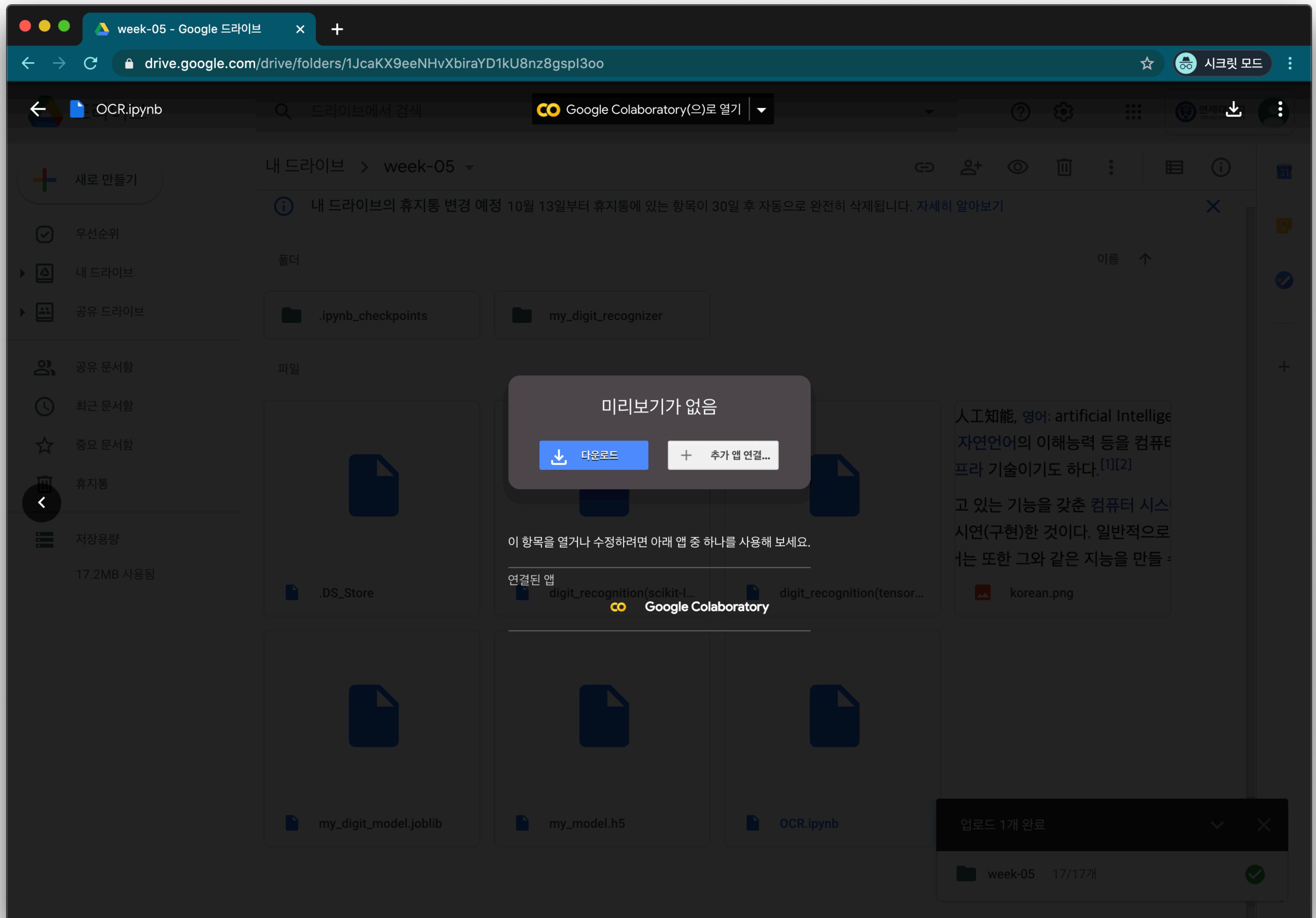
<https://drive.google.com/drive/my-drive> 접속

The screenshot shows the Google Drive web interface at <https://drive.google.com/drive/my-drive>. A prominent message in the center says "모든 파일을 한곳에 편리하게 보관". On the left, there's a sidebar with links like "새로 만들기", "내 드라이브", "공유 드라이브", and "휴지통". A message at the top right says "내 드라이브의 휴지통 변경 예정 10월 13일부터 휴지통에 있는 항목이 30일 후 자동으로 완전히 삭제됩니다. 자세히 알아보기". Below the message, there are sections for "Google 문서, 스프레드시트, 프레젠테이션 등의 Google 편집기" and "Microsoft Office 파일 등 수백 가지 형식의 파일". At the bottom, there's a download link for "데스크톱용 드라이브 다운로드".

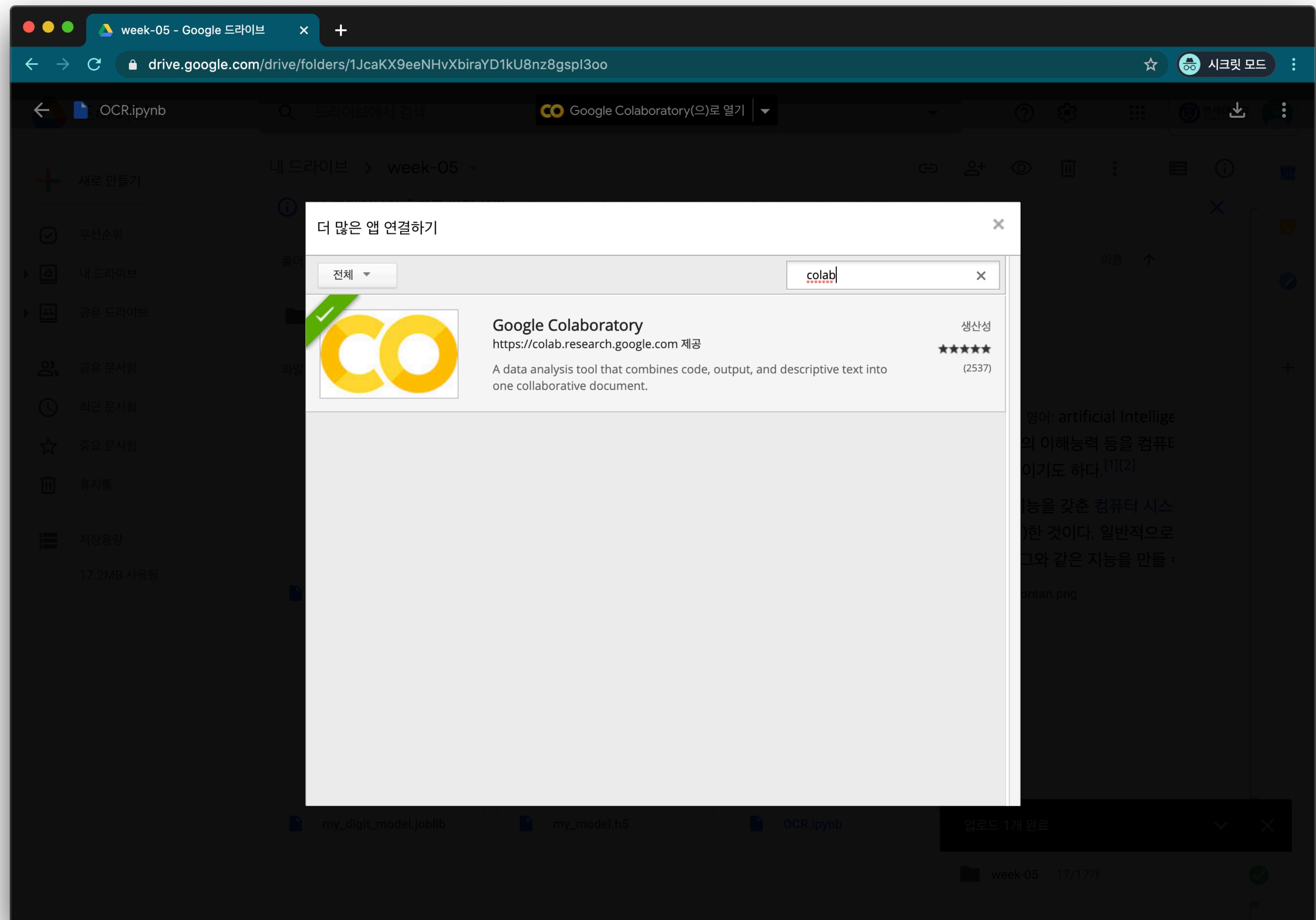
## Practice의 week-05 업로드!



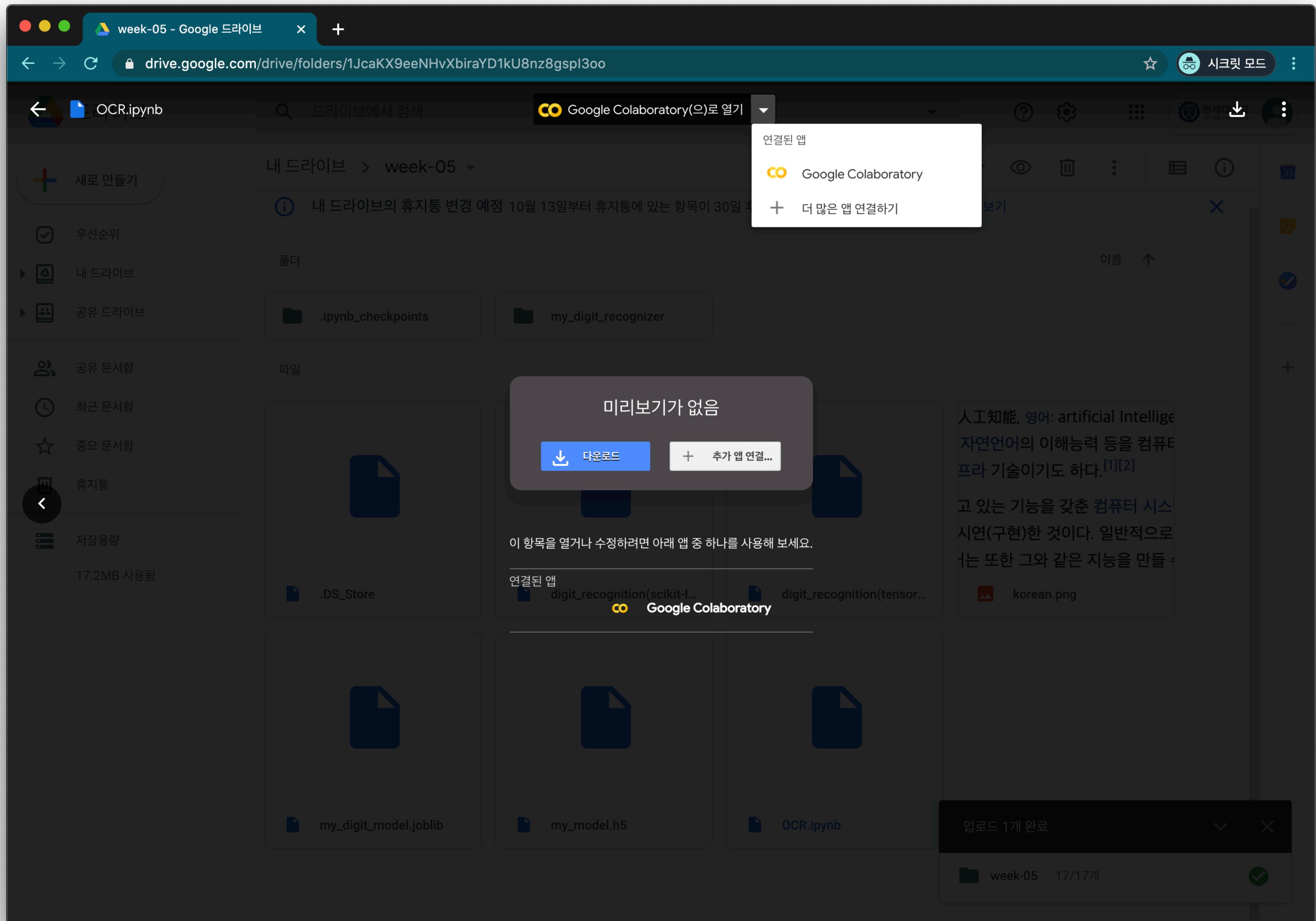
OCR.ipynb 클릭후 추가 앱 연결 클릭!



## Colab 검색후 연결/설치!



다시 OCR.ipynb 쪽으로 와서, Google Colaboratory로 열기!



The screenshot shows a Google Colaboratory notebook titled "OCR.ipynb". The notebook interface includes a toolbar at the top with file operations like "파일", "수정", "보기", "삽입", "런타임", "도구", and "도움말". Below the toolbar are buttons for "댓글", "공유", and "설정". A status bar at the bottom right shows "연결" and "수정 가능".

The main content area displays a section titled "비정형 빅데이터 응용과 실습 - Week 05" which contains a subsection "Easy OCR 라이브러리 활용" with a link to <https://github.com/JaidedAI/EasyOCR>.

Below this, there is a section titled "1. EasyOCR 설치" containing a code cell:

```
[ ]
```

```
[ ] import torch
import torchvision
print(torch.__version__)
print(torchvision.__version__)
```

Output from the code cell:

```
1.6.0
0.7.0
```

At the bottom, there is a command cell:

```
▶ import sys
!{sys.executable} -m pip install easyocr
```

Output from the command cell:

```
Requirement already satisfied: easyocr in /opt/anaconda3/lib/python3.7/site-packages (1.1.9)
Requirement already satisfied: Pillow in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (7.0.0)
Requirement already satisfied: scikit-image in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (0.16.2)
Requirement already satisfied: opencv-python in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (4.2.0.34)
Requirement already satisfied: torchvision>=0.5 in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (0.7.0)
Requirement already satisfied: numpy in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (1.18.5)
Requirement already satisfied: python-bidi in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (0.4.2)
Requirement already satisfied: torch in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (1.6.0)
Requirement already satisfied: scipy in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (1.4.1)
Requirement already satisfied: imageio>=2.3.0 in /opt/anaconda3/lib/python3.7/site-packages (from scikit-image->easyocr) (2.6.1)
Requirement already satisfied: matplotlib!=3.0.0,>=2.0.0 in /opt/anaconda3/lib/python3.7/site-packages (from scikit-image->easyocr) (3.1.3)
Requirement already satisfied: PyWavelets>=0.4.0 in /opt/anaconda3/lib/python3.7/site-packages (from scikit-image->easyocr) (1.1.1)
Requirement already satisfied: networkx>=2.0 in /opt/anaconda3/lib/python3.7/site-packages (from scikit-image->easyocr) (2.4)
```

좌측에 폴더 클릭후, local에 있는 korean.png 업로드!

The screenshot shows a Google Colab interface with a notebook titled "OCR.ipynb". The left sidebar displays a file tree with a "sample\_data" folder selected. The main area contains a section titled "비정형 빅데이터 응용과 실습 - Week 05" under "Easy OCR 라이브러리 활용". It includes a link to the repository: <https://github.com/JaicedAI/EasyOCR>. Below this, a section titled "1. EasyOCR 설치" is shown, containing Python code for installing the library:

```
[ ]  
[ ] import torch  
import torchvision  
print(torch.__version__)  
print(torchvision.__version__)  
  
1.6.0  
0.7.0  
  
[ ]  
import sys  
!{sys.executable} -m pip install easyocr
```

The output of the command shows that all dependencies are already satisfied, including easyocr, Pillow, scikit-image, opencv-python, torchvision, numpy, python-bidi, torch, scipy, imageio, matplotlib, PyWavelets, and networkx.

전체 실행!

The screenshot shows a Google Colab notebook titled 'OCR.ipynb'. The notebook interface includes a top bar with tabs for 'week-05 - Google 드라이브' and 'OCR.ipynb - Colaboratory', a toolbar with file operations, and a status bar indicating '시크릿 모드' (Secret Mode). The main area contains a sidebar with a '파일' (File) section showing a folder structure with 'sample\_data' and 'korean.png'. The main content area has tabs for '+ 코드' (Code) and '+ 텍스트' (Text), with the Code tab selected. A section titled '비정형 빅데이터 응용과 실습 - Week 05' is expanded, containing a heading 'Easy OCR 라이브러리 활용' and a bullet point with a link: '• <https://github.com/JaidedAI/EasyOCR>'. Another section titled '1. EasyOCR 설치' is also expanded. In the code editor, there are several code snippets and their outputs:

```
[ ]  
[ ] import torch  
import torchvision  
print(torch.__version__)  
print(torchvision.__version__)  
  
1.6.0  
0.7.0  
  
[ ]  
import sys  
!{sys.executable} -m pip install easyocr  
  
Requirement already satisfied: easyocr in /opt/anaconda3/lib/python3.7/site-packages (1.1.9)  
Requirement already satisfied: Pillow in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (7.0.0)  
Requirement already satisfied: scikit-image in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (0.16.1)  
Requirement already satisfied: opencv-python in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (4.1.2.30)  
Requirement already satisfied: torchvision>=0.5 in /opt/anaconda3/lib/python3.7/site-packages (from easyocr)  
Requirement already satisfied: numpy in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (1.18.5)  
Requirement already satisfied: python-bidi in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (0.4.0)  
Requirement already satisfied: torch in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (1.6.0)  
Requirement already satisfied: scipy in /opt/anaconda3/lib/python3.7/site-packages (from easyocr) (1.4.1)  
Requirement already satisfied: imageio>=2.3.0 in /opt/anaconda3/lib/python3.7/site-packages (from scikit-image)  
Requirement already satisfied: matplotlib!=3.0.0,>=2.0.0 in /opt/anaconda3/lib/python3.7/site-packages (from scikit-image)  
Requirement already satisfied: PyWavelets>=0.4.0 in /opt/anaconda3/lib/python3.7/site-packages (from scikit-image)  
Requirement already satisfied: networkx>=2.0 in /opt/anaconda3/lib/python3.7/site-packages (from scikit-image)
```

성공!

The screenshot shows a Google Colab notebook titled 'OCR.ipynb' running on a Google Drive. The notebook contains the following code:

```
[3] import easyocr
reader = easyocr.Reader(['ko', 'en']) # need to run only once to load model into memory

CUDA not available - defaulting to CPU. Note: This module is much faster with a GPU.
Downloading detection model, please wait. This may take several minutes depending upon your network connection.
Downloading recognition model, please wait. This may take several minutes depending upon your network connect:
```

**4. OCR 진행**

```
[4] result = reader.readtext('korean.png')

▶ for i in result:
    print(i[0])
    print(i[1])
```

Output of the code:

```
[[[0, 13], [1674, 13], [1674, 88], [0, 88]]
인공지능( 인공지능, 영어: artificial Intelligence)은 인간의 학습 능력, 추론 능력,
[[0, 101], [1634, 101], [1634, 176], [0, 176]]
지각 능력, 자연 언어의 이해 능력을 컴퓨터 프로그램으로 실현한 기술이다.
[[0, 179], [799, 179], [799, 263], [0, 263]]
하나의 인프라 기술이기도 하다.[1][2]
[[0, 306], [1678, 306], [1678, 379], [0, 379]]
지능을 갖고 있는 기능을 갖춘 컴퓨터 시스템이며, 인간의 지능을 기계 등에 인
[[0, 397], [1657, 397], [1657, 465], [0, 465]]
공적으로 시연(구현)한 것이다. 일반적으로 범용 컴퓨터에 적용한다고 가정한
[[0, 482], [1662, 482], [1662, 556], [0, 556]]
다. 이용어는 또한 그와 같은 지능을 만들 수 있는 방법론이나 실현 가능성 등
[[0, 559], [1087, 559], [1087, 643], [0, 643]]
을 연구하는 과학 분야를 지칭하기도 한다.[3][4][5]
```

**5. 텍스트만 추출 할 경우**

```
reader.readtext('korean.png', detail = 0)
```

## 2. Educate 프로그램으로 AWS 인스턴스 생성

- EasyOCR이 요구하는 RAM이, 현재 저희가 사용하는 t2.micro 보다 높음
  - 더 좋은 성능의 인스턴스를 사용할 것입니다.
- 하지만 더 좋은 인스턴스는 무료 플랜에 들어있지 않음
  - AWS Educate 계정을 사용하여 무료로 사용하도록 하겠습니다!
- 주의!
  - 일반적으로 접속하는 AWS 계정과 Educate 계정은 다른 것입니다!
  - 혹시 기존에 사용하시던 계정에 추가적인 인스턴스나 서비스를 런칭하시면 비용이 나올 수도 있으니 꼭 우측 상단에 아이디가 어떻게 되어있는지 확인을 해주세요!

The screenshot shows a web browser window with the URL <https://www.awseducate.com/signin> in the address bar. The page itself is a login form for AWS Educate. At the top center is the AWS Educate logo, which consists of the word "aws" in black lowercase letters followed by "educate" in a larger, stylized orange font where the letter "e" has a graduation cap icon on top. Below the logo are two input fields: one for "email" and one for "password". To the right of the password field is a blue "Sign In" button with white text. Underneath the "Sign In" button are two links: "Forgot password?" and "Not an AWS Educate member? [Apply now.](#)". At the bottom of the page, there is a footer bar containing the text "AWS Educate는 학생들이 졸업하기 전 실제 클라우드 기술을 배울 수 있는 Amazon의 프로그램입니다. 학생과 교육자에게 클라우드 관련 학습 속도를 높이는 데 필요한 리소스를 제공합니다." and "Terms and Conditions | © 2020, (c)2019, Amazon Web Services, Inc. 또는 자회사. All rights reserved."

The screenshot shows a web browser window for the AWS Educate student dashboard at [awseducate.com/student/s/](https://awseducate.com/student/s/). The dashboard features a navigation bar with links for Portfolio, Career Pathways, Badges, Jobs, AWS Account (which is highlighted with a red box), and Logout. Below the navigation bar, there's a user profile section for Jungwon Seo, showing Consecutive Days: 2, Pathways Completed: 0, and Badges Earned: 0. A language preference dropdown is set to English. The main content area includes a welcome message about cloud jobs, a call-to-action to begin the journey, and a reminder to watch a missed webinar. On the right, there's a 'Suggested Jobs' sidebar listing three job opportunities: Software Engineer at Mitel Networks, Software Development Engineer Intern - Summer 2021 (Canada) at Amazon, Inc., and Software Development Engineer - Intern at Amazon, Inc.

aws<sup>!</sup> educate

Jungwon Seo

Consecutive Days: 2

Pathways Completed: 0

Badges Earned: 0

Preferred Language: English

AWS Account

Logout

Cloud technology is everywhere, creating over 18 million cloud jobs worldwide (source: Wanted Analytics). AWS Educate introduces you to lucrative cloud-enabled careers through more than 25 learning pathways, each with content from industry professionals, learning activities and labs, opportunities to earn AWS Educate Badges and Certificates of Completion, and access to the AWS Educate Job Board. Coupled with courses at your school or through online providers, AWS Educate puts you on the pathway to your dream job in the clouds.

Begin your journey today!

If you missed out the "Optimizing your AWS Educate Profile to Help You Find a Cloud Career" webinar and Q&A session, watch it [here](#)!

Suggested Jobs

Software Engineer  
Mitel Networks  
[more about this opportunity](#)

Software Development Engineer Intern - Summer 2021 (Canada)  
Amazon, Inc.  
[more about this opportunity](#)

Software Development Engineer - Intern  
Amazon, Inc.  
[more about this opportunity](#)

Cloud Consulting Intern  
Amazon, Inc.  
[more about this opportunity](#)

AWS Account

awseducate.com/student/s/awssite

Portfolio Career Pathways Badges Jobs AWS Account Logout

## AWS Educate Starter Account

Your cloud journey has only just begun. Use your AWS Educate Starter Account to access the AWS Console and resources, and start building in the cloud!

AWS Educate Starter Account

Your account has an estimated **97** credits remaining and access will end on **Sep 18, 2021**.

Note: Clicking this button will take you to a third party site managed by Vocareum, Inc. ("Third Party Servicer"). In addition to the AWS Educate terms of service, your use of the AWS Educate Starter Account is governed by the Third Party Servicer's terms, including its Privacy Policy. AWS assumes no responsibility or liability and makes no representations or warranties regarding services provided by a Third Party Servicer.



The screenshot shows a web browser window with the title bar "AWS Account" and "Workbench". The address bar displays the URL "labs.vocareum.com/main/main.php?m=editor&nav=1&asnid=14334&stepid=14335". The main content area is divided into two sections: "Welcome to your AWS Educate Account" on the left and "Your AWS Account Status" on the right.

**Welcome to your AWS Educate Account**

AWS Educate provides you with access to a wide variety of AWS Services for you to get your hands on and build on AWS! To get started, click on the AWS Console button to log in to your AWS console.

Please read the FAQ below to help you get started on your Starter Account.

- What are the list of services supported?
- What regions are supported with Starter Accounts or Classroom Accounts?
- I can't start any resources. What happened?
- Can I create users within my Starter or Classroom Account for others to access?
- Can I create my own IAM policy within Starter Account or Classroom?
- Can I use marketplace software with my Starter Account or Classrooms?
- Are there any restrictions on AWS services in my AWS Educate Account?
- Are FPGA Instances Supported?

**Your AWS Account Status**

**Active**  
full access ([jungwons@yonsei.ac.kr](mailto:jungwons@yonsei.ac.kr))

**\$97.53**  
remaining credits (estimated)

**2:60**  
session time

[Account Details](#) [AWS Console](#)

Please use AWS Educate Account responsibly. Remember to shut down your instances when not in use to make the best use of your credits. And, don't forget to logout once you are done with your work!

우측 상단에 유저 이름과 Region이 이런식으로 나오면 정상적으로 Educate계정으로 접속이 된것입니다!

The screenshot shows the AWS Management Console homepage. At the top, there is a navigation bar with tabs for 'AWS Account', 'Workbench', and 'AWS Management Console'. Below the navigation bar, the URL is 'console.aws.amazon.com/console/home?region=us-east-1#'. The top right corner displays the user's profile information: 'vocstartsoft/user968505=jungwons@yonsei.ac.kr @ 3723-1404-2704' with a red box around it, followed by 'N. Virginia' and 'Support'. The main title 'AWS Management Console' is centered above the content area.

**AWS services**

**Find Services**  
You can enter names, keywords or acronyms.

**Recently visited services**

EC2	DynamoDB	Systems Manager
RDS	Billing	Lambda
CloudWatch	ElastiCache	Secrets Manager
EFS	S3	CloudFront
AWS AppConfig	IAM	AWS Cost Explorer

**All services**

**Build a solution**  
Get started with simple wizards and automated workflows.

**Launch a virtual machine** With EC2    **Build a web app** With Elastic Beanstalk    **Build using virtual servers** With Lightsail

**Stay connected to your AWS resources on-the-go**

Download the AWS Console Mobile App to your iOS or Android mobile device. [Learn more](#)

**Explore AWS**

**Move to Managed File Storage**  
Reduce complexity, overhead, and cost by moving to fully managed storage. [Learn more](#)

**Free Digital Training**  
Get access to 350+ self-paced online courses covering AWS products and services. [Learn more](#)

**Amazon SageMaker Autopilot**  
Get hands-on with AutoML. [Learn more](#)

**RDS Read Replicas**  
Achieve scale and low-latency for read-heavy workloads with RDS Read Replicas. [Learn more](#)

Feedback English (US) © 2008 - 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

**E.O.D**