

# 스파르타온라인 파이썬 혼자놀이 패키지 - 2일차

## [수업 목표]

1. 파이썬 기초 문법을 안다.
2. 크롤링한 데이터를 엑셀 파일로 저장할 수 있다.
3. 파이썬으로 파일이 첨부된 이메일을 보낼 수 있다.

## [목차]

01. 오늘 배울 것
02. 기사 웹스크래핑(크롤링)하기
03. 엑셀 파일로 저장하기
04. 이메일 보내기
05. 끝 & 속제 설명

## 01. 오늘 배울 것

- 지난 시간에 배운 크롤링 기술을 이용하여 이번에는 기사들을 수집해보겠습니다!
- 수집한 기사들을 엑셀 파일로 정리해보겠습니다.
- 파이썬으로 이메일도 보낼 수 있어요.
- ▼ 이번 시간이 끝나면 완성하는 것

	A	B	C
1	제목	링크	신문사
2	추석 연휴 교통사고 일평균 481건..."연휴 전날이 가장 많아"	<a href="https://imnews.imbc.com/n">https://imnews.imbc.com/n</a>	MBC
3	정총리, 오후 대국민담화...추석 방역협조·개천절 집회자제 당부	<a href="http://yna.kr/AKR20200926">http://yna.kr/AKR20200926</a>	연합뉴스
4	소상공인 새희망자금 내일 오후 5시까지 신청해야 추석 전 수	<a href="https://www.ytn.co.kr/_ln/01">https://www.ytn.co.kr/_ln/01</a>	YTN
5	정 총리, 오늘 오후 4시30분 대국민 담화... 추석연휴 특별방역	<a href="http://news.heraldcorp.com">http://news.heraldcorp.com</a>	헤럴드경제
6	"음식점 방문, 안전한 패스 QR 출입증으로" 통신 3사 추석 캠페	<a href="http://news.tf.co.kr/read/ec">http://news.tf.co.kr/read/ec</a>	더팩트
7	SKT, 추석 연휴 네트워크 집중관리 돌입	<a href="https://zdnet.co.kr/view/?nc">https://zdnet.co.kr/view/?nc</a>	ZDNet
8	[속보]서울시, 추석연휴 특별방역기간 지정..."실외 일부 체육·문	<a href="http://news.khan.co.kr/kh_n">http://news.khan.co.kr/kh_n</a>	경향신문
9	내일부터 2주간 추석특별방역...코로나19 재확산 갈림길	<a href="http://www.wowtv.co.kr/Ne">http://www.wowtv.co.kr/Ne</a>	한국경제TV
10	대통령 추석 선물 담양 '대잎술'	<a href="https://www.ytn.co.kr/_ln/01">https://www.ytn.co.kr/_ln/01</a>	YTN
11	소상공인 새희망자금 내일 오후 5시까지 신청해야 추석 전 받	<a href="http://yna.kr/AKR20200926">http://yna.kr/AKR20200926</a>	연합뉴스

## 02. 기사 웹스크래핑(크롤링)하기

- ▼ 1) HTML 구조 파악하기
  - 먼저 네이버 검색 페이지의 HTML을 살펴봅시다!
- ▼ [코드스니펫] - 네이버 검색 URL

```
https://search.naver.com/search.naver?where=news&sm=tab_jum&query=추석
```

- URL을 살펴보면, **where=news** 는 뉴스 검색, **query=추석** 은 검색어입니다.

- ▼ 2) 기사 정보 스크래핑하기
  - 이제 본격적으로 기사를 스크래핑해봅시다
- ▼ [코드스니펫] - 기사 스크래핑 시작코드

```
from bs4 import BeautifulSoup
from selenium import webdriver
```

```

driver = webdriver.Chrome('chromedriver')

url = "https://search.naver.com/search.naver?where=news&sm=tab_jum&query=추석"

driver.get(url)
req = driver.page_source
soup = BeautifulSoup(req, 'html.parser')

#####
# 여기에 코드 적기!
#####

driver.quit()

```

- 먼저 제목을 스크래핑해볼까요? 우선 각 기사를 선택하고, 그 안의 기사 제목을 프린트해봅시다.

```

articles = soup.select("#main_pack > div.news.mynews.section._prs_nws > ul > li")
for article in articles:
    a_tag = article.select_one("dl > dt > a")
    print(a_tag.text)

```

- 기사 URL과 신문사도 같이 수집해볼까요?

```

url = a_tag["href"]
title = a_tag.text
comp = article.select_one("dd.txt_inline > span._sp_each_source").text.split(" ")[0].replace('언론사', '')
print(title, url, comp)

```

#### ▼ 기사 스크래핑 완성 코드

```

from bs4 import BeautifulSoup
from selenium import webdriver

driver = webdriver.Chrome('chromedriver')

url = "https://search.naver.com/search.naver?where=news&sm=tab_jum&query=추석"

driver.get(url)
req = driver.page_source
soup = BeautifulSoup(req, 'html.parser')

#####
# 여기에 코드 적기!
#####

articles = soup.select('#main_pack > div.news.mynews.section._prs_nws > ul > li')

for article in articles:
    a_tag = article.select_one('dl > dt > a')

    title = a_tag.text
    url = a_tag['href']
    comp = article.select_one('dd.txt_inline > span._sp_each_source').text.split(' ')[0].replace('언론사', '')
    print(title, url, comp)

driver.quit()

```

## 03. 엑셀 파일로 저장하기

### ▼ 4) 패키지 설치하기

- 파이썬으로 엑셀 파일을 읽고 쓸 때는 `openpyxl` 패키지를 이용하면 편합니다! Settings/Preferences > Project Interpreter 에서 +를 눌러 패키지를 검색해 설치해주세요.

### ▼ 5) 엑셀파일로 저장하기

- 이제 `openpyxl`을 임포트하고 파이썬에서 엑셀파일을 만들어봅시다.

#### ▼ [코드스니펫] - openpyxl 이용하기

```
from openpyxl import Workbook

wb = Workbook()
ws1 = wb.active
ws1.title = "articles"
ws1.append(["제목", "링크", "신문사"])

wb.save(filename='articles.xlsx')
```

- 엑셀 파일을 열어볼까요? `articles` 라는 워크시트에 `제목`, `링크`, `신문사` 가 1열 각 셀에 들어갔죠?
- 이 코드를 응용해서 우리가 스크래핑한 정보를 엑셀파일로 저장해봅시다. 아래처럼 나오면 성공!

	A	B	C
1	제목	링크	신문사
2	추석 연휴 교통사고 일평균 481건..."연휴 전날이 가장 많아"	https://imnews.imbc.com/n	MBC
3	정총리, 오후 대국민담화...추석 방역협조·개천절 집회자제 당부	http://yna.kr/AKR20200926	연합뉴스
4	소상공인 새희망자금 내일 오후 5시까지 신청해야 추석 전 수	https://www.ytn.co.kr/_ln/0	YTN
5	정 총리, 오늘 오후 4시30분 대국민 담화... 추석연휴 특별방역	http://news.heraldcorp.com	헤럴드경제
6	"음식점 방문, 안전한 패스 QR 출입증으로" 통신 3사 추석 캠페	http://news.tf.co.kr/read/ec	더팩트
7	SKT, 추석 연휴 네트워크 집중관리 돌입	https://zdnet.co.kr/view/?nc	ZDNet
8	[속보]서울시, 추석연휴 특별방역기간 지정..."실외 일부 체육·문	http://news.khan.co.kr/kh_n	경향신문
9	내일부터 2주간 추석특별방역...코로나19 재확산 갈림길	http://www.wowtv.co.kr/Ne	한국경제TV
10	대통령 추석 선물 담양 '대잎술'	https://www.ytn.co.kr/_ln/0	YTN
11	소상공인 새희망자금 내일 오후 5시까지 신청해야 추석 전 받	http://yna.kr/AKR20200926	연합뉴스

#### ▼ 완성 코드

```
from bs4 import BeautifulSoup
from selenium import webdriver

from openpyxl import Workbook

driver = webdriver.Chrome('chromedriver')

url = "https://search.naver.com/search.naver?where=news&sm=tab_jum&query=추석"

driver.get(url)
req = driver.page_source
soup = BeautifulSoup(req, 'html.parser')

wb = Workbook()
ws1 = wb.active
ws1.title = "articles"
ws1.append(["제목", "링크", "신문사"])

articles = soup.select('#main_pack > div.news.mynews.section._prs_nws > ul > li')

for article in articles:
    a_tag = article.select_one('dl > dt > a')

    title = a_tag.text
    url = a_tag['href']
    comp = article.select_one('dd.txt_inline > span._sp_each_source').text.split(' ')[0].replace('연론사','')
    ws1.append([title, url, comp])

driver.quit()
wb.save(filename='articles.xlsx')
```

## 04. 이메일 보내기

### ▼ 6) SMTP란?

- 파이썬에서는 `smtplib` 패키지를 이용하여 메일 서버에 접속할 수 있습니다.
- 기본 내장되어 있는거라, 설치할 필요는 없어요!

## ▼ 7) 이메일 일단 한 번 보내보기



아래 코드는 지메일을 기준으로 설명드릴게요! 보내는 사람 이메일 주소는 꼭 지메일 계정(@gmail.com)으로 해주세요

▼ 우선 아래 코드를 이용해서 이메일을 보내보도록 하겠습니다.

### ▼ [코드스니펫] - 이메일 보내기 시작코드

```
import smtplib
from email.mime.multipart import MIMEMultipart
from email.mime.base import MIMEBase
from email.mime.text import MIMEText
from email import encoders

# 보내는 사람 정보
me = "보내는사람@gmail.com"
my_password = "비밀번호"

# 로그인하기
s = smtplib.SMTP_SSL('smtp.gmail.com')
s.login(me, my_password)

# 받는 사람 정보
you = "받는사람@아무_도메인"

# 메일 기본 정보 설정
msg = MIMEMultipart('alternative')
msg['Subject'] = "제목"
msg['From'] = me
msg['To'] = you

# 메일 내용 쓰기
content = "메일 내용"
part2 = MIMEText(content, 'plain')
msg.attach(part2)

# 메일 보내고 서버 끄기
s.sendmail(me, you, msg.as_string())
s.quit()
```

▼ 앓? 🤖 혹시 에러⚠️가 나나요? 지메일 보안 설정에 따라 여러 가능성이 있습니다! 당황하지 말고 차근차근 아래 목록을 확인해보세요.

#### 1. 이메일 주소 / 비밀번호가 틀린 경우

```
smtplib.SMTPAuthenticationError: (535, b'5.7.8 Username and Password not accepted. Learn more at\n5.7.8 https://su
```

당연하지만 내 이메일 주소나 비밀번호를 잘못 입력한 경우 제대로 로그인을 할 수 없겠죠? 정보를 다시 한 번 확인해주세요!

#### 2. 2-Step-Verification이 켜져있는 경우

```
smtplib.SMTPAuthenticationError: (534, b'5.7.9 Application-specific password required. Learn more at\n5.7.9 https:
```

2-Step-Verification (2단계 인증) 설정이 되어있을 때 위와 같은 메시지가 종종 나타납니다. 에러메시지의 URL을 클릭하여 단계를 따라가거나 아래 도움말을 보시고 2단계 인증을 풀어주세요.

### ▼ [코드스니펫] - 2단계 인증 해제

```
https://myaccount.google.com/signinoptions/two-step-verification
```

#### 3. '보안 수준이 낮은 앱의 액세스'가 사용 중지된 경우

```
smtplib.SMTPAuthenticationError: (535, b'5.7.8 Username and Password not accepted. Learn more at\n5.7.8 https://su
```

이메일 주소 / 비밀번호를 틀렸을 때와 같은 에러 메시지가 나오는데, 링크를 따라가보시면 다른 이메일 플랫폼을 통해 Gmail에 로그인할 수 없을 때의 해결법이 나옵니다. 여기서 '보안 수준이 낮은 앱이 계정에 액세스하도록 허용'을 따라가서 풀어주시면 됩니다.

#### ▼ [코드스니펫] - 보안 수준이 낮은 앱 해제하기

```
https://myaccount.google.com/lesssecureapps
```

#### 4. 받는 사람 이메일 주소를 틀린 경우

😞 에러는 나지 않지만 이메일이 오지 않아요

#### ▼ 8) 이메일 보내기

##### ▼ 여러 사람에게 각각 이메일 보내기

💡 파이썬을 이용하면 여러 사람에게 각자 이메일을 보낼 수 있습니다! 업무자동화에도 응용할 수 있겠죠 😊

- 반복문을 이용하면 쉽게 여러 사람에게 이메일을 보낼 수 있습니다. 받는 사람 주소를 리스트에 넣어 반복문을 돌려봅시다.

##### ▼ [코드스니펫] - 여러 사람에게 이메일 보내기 시작코드

```
import smtplib
from email.mime.multipart import MIMEMultipart
from email.mime.base import MIMEBase
from email.mime.text import MIMEText
from email import encoders

# 보내는 사람 정보
me = "보내는사람@gmail.com"
my_password = "비밀번호"

# 로그인하기
s = smtplib.SMTP_SSL('smtp.gmail.com')
s.login(me, my_password)

# 받는 사람 정보
email_list = ["이메일1", "이메일2"]

for you in email_list:
    # 메일 기본 정보 설정
    msg = MIMEMultipart('alternative')
    msg['Subject'] = "제목"
    msg['From'] = me
    msg['To'] = you

    # 메일 내용 쓰기
    content = "메일 내용"
    part2 = MIMEText(content, 'plain')
    msg.attach(part2)

    # 메일 보내기
    s.sendmail(me, you, msg.as_string())

# 다 끝나고 닫기
s.quit()
```

##### ▼ 파일 첨부하기

- 메일에 파일을 첨부하기 위해서는 파일 내용을 컴퓨터가 이해할 수 있는 이진수로 바꿔주어야합니다. 여러 파일을 첨부할 때에는 똑같은 코드를 반복하여 각각 `msg.attach()` 해주면 됩니다.

##### ▼ [코드스니펫] - 파일 첨부하기

```
part = MIMEBase('application', "octet-stream")
with open("articles.xlsx", 'rb') as file:
```

```
part.set_payload(file.read())
encoders.encode_base64(part)
part.add_header('Content-Disposition', "attachment", filename="추석기사.xlsx")
msg.attach(part)
```



내 컴퓨터에서의 파일명과 실제로 보여지는 첨부 파일명은 달라도 된답니다!

#### ▼ 완성코드

```
import smtplib
from email.mime.multipart import MIMEMultipart
from email.mime.base import MIMEBase
from email.mime.text import MIMEText
from email import encoders

# 보내는 사람 정보
me = "보내는사람@gmail.com"
my_password = "비밀번호"

# 로그인하기
s = smtplib.SMTP_SSL('smtp.gmail.com')
s.login(me, my_password)

# 받는 사람 정보
email_list = ["이메일1", "이메일2"]

for you in email_list:
    # 메일 기본 정보 설정
    msg = MIMEMultipart('alternative')
    msg['Subject'] = "제목"
    msg['From'] = me
    msg['To'] = you

    # 메일 내용 쓰기
    content = "메일 내용"
    part2 = MIMEText(content, 'plain')
    msg.attach(part2)

    part = MIMEBase('application', "octet-stream")
    with open("articles.xlsx", 'rb') as file:
        part.set_payload(file.read())
    encoders.encode_base64(part)
    part.add_header('Content-Disposition', "attachment", filename="추석기사.xlsx")
    msg.attach(part)

    # 메일 보내기
    s.sendmail(me, you, msg.as_string())

# 다 끝나고 닫기
s.quit()
```

## 05. 끝 & 숙제 설명



썸네일을 포함해 기사 크롤링해서 자신에게 메일 보내기

- 이번에는 기사를 크롤링할 때 썸네일도 같이 받아와볼까요? 썸네일을 선택하려면 어떤 선택자를 써야할지 생각해 보세요. 수업시간에 썼던 코드를 활용하세요!

▼ 엑셀 파일을 아래처럼 만들면 성공!

	A	B	C	D
1	제목	링크	신문사	썸네일
2	[속보]서울시, 추석연휴 특별방역기간...	<a href="http://news.khan.co.kr/kh_news/kh">http://news.khan.co.kr/kh_news/kh</a>	경향신문	<a href="https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews">https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews</a>
3	추석 연휴 교통사고 일평균 481건...	<a href="https://imnews.imbc.com/news/20">https://imnews.imbc.com/news/20</a>	MBC	<a href="https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews">https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews</a>
4	정총리, 오후 대국민담화...추석 방역협	<a href="http://yna.kr/AKR2020092605310">http://yna.kr/AKR2020092605310</a>	연합뉴스	<a href="https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews">https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews</a>
5	소상공인 새희망자금 내일 오후 5시까지	<a href="https://www.ytn.co.kr/_ln/0102_20">https://www.ytn.co.kr/_ln/0102_20</a>	YTN	<a href="https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews">https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews</a>
6	정 총리, 오늘 오후 4시30분 대국민 담	<a href="http://news.heraldcorp.com/view">http://news.heraldcorp.com/view</a>	헤럴드경제	<a href="https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews">https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews</a>
7	추석 연휴 전날 교통사고 사상자 1200	<a href="http://www.fnnews.com/news/20">http://www.fnnews.com/news/20</a>	파이낸셜뉴	<a href="https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews">https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews</a>
8	SKT, 추석 연휴 통신 서비스 집중관리	<a href="https://biz.chosun.com/site/data/1">https://biz.chosun.com/site/data/1</a>	조선비즈	<a href="https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews">https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews</a>
9	SKT, 추석 연휴 네트워크 집중관리 돌	<a href="https://zdnet.co.kr/view/?no=202">https://zdnet.co.kr/view/?no=202</a>	ZDNet	<a href="https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews">https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews</a>
10	"음식점 방문, 안전한 패스 QR 출력중"	<a href="http://news.tf.co.kr/read/economy">http://news.tf.co.kr/read/economy</a>	더팩트	<a href="https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews">https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews</a>
11	광주 거리두기 2단계 내달 11일까지...	<a href="http://yna.kr/AKR2020092703625">http://yna.kr/AKR2020092703625</a>	연합뉴스	<a href="https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews">https://search.pstatic.net/common/?src=https%3A%2F%2Fimgnews</a>
12				
13				

#### ▼ [코드스니펫] - 기사 크롤링해서 저장하기 시작 코드

```
from bs4 import BeautifulSoup
from selenium import webdriver

driver = webdriver.Chrome('chromedriver')

from openpyxl import Workbook

wb = Workbook()
ws1 = wb.active
ws1.title = "articles"
ws1.append(["제목", "링크", "신문사", "썸네일"])

url = f"https://search.naver.com/search.naver?&where=news&query=추석"

driver.get(url)
req = driver.page_source
soup = BeautifulSoup(req, 'html.parser')

#####
# 각 요소 크롤링해서 엑셀에 붙여넣기
#####

wb.save(filename='articles.xlsx')
driver.quit()
```

- 한 걸음 더! (소곤소곤) 크롤링한 정보를 내 메일로 한번 더 보내보세요! 연습, 또 연습!

#### ▼ [코드스니펫] - 파일을 첨부한 이메일 보내기 시작 코드

```
import smtplib
from email.mime.multipart import MIMEMultipart
from email.mime.base import MIMEBase
from email.mime.text import MIMEText
from email import encoders

# 보내는 사람 정보
me = "보내는사람@gmail.com"
my_password = "비밀번호"

# 로그인하기
s = smtplib.SMTP_SSL('smtp.gmail.com')
s.login(me, my_password)

# 받는 사람 정보
you = "받는사람@아무_도메인"

# 메일 기본 정보 설정
msg = MIMEMultipart('alternative')
msg['Subject'] = "제목"
msg['From'] = me
msg['To'] = you

# 메일 내용 쓰기
content = "메일 내용"
part2 = MIMEText(content, 'plain')
msg.attach(part2)

# 파일 첨부하기
part = MIMEBase('application', "octet-stream")
with open("파일 이름", 'rb') as file:
```

```
part.set_payload(file.read())
encoders.encode_base64(part)
part.add_header('Content-Disposition', "attachment", filename="첨부파일 이름")
msg.attach(part)

# 메일 보내고 서버 끄기
s.sendmail(me, you, msg.as_string())
s.quit()
```