

코알라 유니브_week4

발표자 김채은

<stage1. 파이썬으로 파일 만들기>

<파일 오픈 형식>

w(write) 새로 파일 쓸 때

R(read) 기존 파일의 데이터 읽을 때

a(append) 기존 파일에 데이터 추가하고 싶을 때

```
f = open("text.txt", "w")
```

```
f.write("Nice to meet you\n")
```

```
f.write("You can call me Amber")
```

```
f.close()
```

Print와 달리 write는 줄바꿈을 하지 않으므로
줄바꿈을 원한다면, 개행문자 추가

<stage2. 데이터 csv파일로 저장하기>

■ csv(comma separated value)란? 콤마로 구분되어지는 값들

행은 줄바꿈으로, 열은 콤마로 처리

예시)

카카오, 노랑
네이버, 초록
삼성, 파랑
sk, 빨강

company.csv

카카오	노랑
네이버	초록
삼성	파랑
sk	빨강

<stage2. 데이터 csv파일로 저장하기>

<코드>

```
f=open("company.csv", "w")

name = ["카카오", "네이버", "삼성", "sk"]
color = ["노랑", "초록", "파랑", "빨강"]

for i in range(len(name)):
    f.write(name[i] + ',' + color[i] + '\n')

f.close()
```

Q. 만약 파일내 데이터가 깨진다면 ?

A. F=open("hitsong.csv", "w", encoding="UTF-8") 추가

<stage3. 데이터 엑셀파일로 저장하기>

<엑셀파일 이해>

- workbook: 전체 작업 공간(엑셀파일 자체)
- Worksheet: 공간 내에 큰 표
- Cell: 한 칸

Q. 정확한 셀번호를 모르 땐?

A. `Sheet.cell(row=1, column=1).value = "Hi"`

openpyxl 설치하기



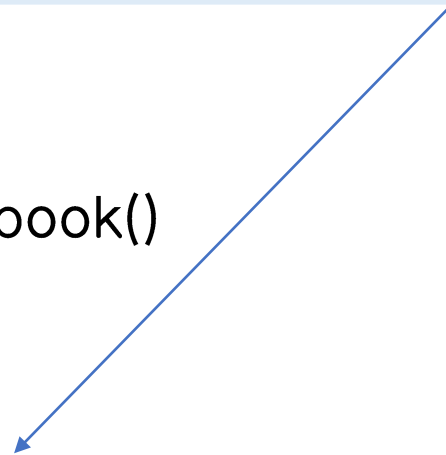
```
import openpyxl
```

```
wb = openpyxl.workbook()
```

```
sheet = wb.active
```

```
sheet['A1'] = "Hi"
```

```
wb.save("test.xlsx")
```



<stage3. 데이터 엑셀파일로 저장하기>

■ 가장 마지막 행에 데이터 입력하기

```
Sheet.append(["지은", "지혜", "예빈", "성민", "유진", "채은"])
```

■ 시트 관련 openpyxl 함수


워크북.create_sheet(시트이름) : 새로운 시트 만들기

워크북[시트이름] : 특정 시트 선택하기

시트.title(시트이름) : 시트 이름 바꾸기

<stage3. 데이터 엑셀파일로 저장하기>

데이터를 수집해서 파일에 저장하는 경우,
header로 첫행에 들어갈 데이터 분류명을 써줌



```
import openpyxl  
wb = openpyxl.workbook()  
sheet.append(["제목", "언론사"])
```

데이터 수집 코드

```
wb.save("navernews.xlsx")
```

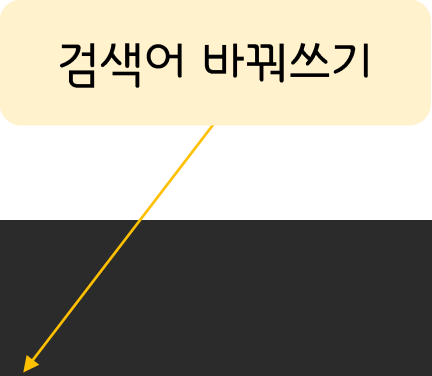
<stage4. 엑셀파일 저장하기(심화)>

■ 뉴스기사 여러 페이지의 데이터 수집

```
keyword = input("검색어를 입력해주세요: ")
```

```
page = 1
for page in range(1, 100, 10): #page에 1부터 100까지 10의 간격으로 반복해줘
    raw = requests.get
    ("https://search.naver.com/search.naver?where=news&sm=tab_jum&query="+keyword+"&start="+str(page),
     headers={"User-Agent":"Mozilla/5.0"})
    html = BeautifulSoup(raw.text, 'html.parser')
```

검색어 바꿔쓰기



여러 페이지 데이터 수집



<stage4. 엑셀파일 저장하기(심화)>

Q. keyword를 바꾸면 이전에 검색했던 데이터가 사라집니다.
여러 keyword에 대한 데이터를 모두 수집하고 싶습니다!

A. Try와 except를 사용하세요.

없으면 하나 만들어

이전의 workbook에
시도해봐!

```
try:
    wb = openpyxl.load_workbook("navernews.xlsx")
    sheet = wb.active
    print("불러오기 완료")
except:
    wb = openpyxl.Workbook()
    sheet = wb.active
    sheet.append(["제목", "언론사"])
    print("새로운 파일을 만들었습니다.")
```

<저의 질문입니다>

Q. 헤더에 `sheet.append(["제목", "언론사"])` 쓰는 건 이해가 되는데,
데이터 채워넣은 후에 `sheet.append([title,source])`는 왜 쓰는 지 궁금합니다.