

Comma Separate Value

- CSV, 필드를 쉼표(,)로 구분한 텍스트 파일
- 엑셀 양식의 데이터를 프로그램에 상관없이 쓰기 위한 데이터 형식이라고 생각하면 쉬움
- 탭(TSV), 빈칸(SSV) 등으로 구분해서 만들기도 함
- 통칭하여 character-separated values (CSV) 부름
- 엑셀에서는 "다름 이름 저장" 기능으로 사용 가능

엑셀로 CSV 파일 만들기

- 1) 파일 다운로드 https://goo.gl/lQvsQF
- 2) 파일 열기
- 3) 파일 → 다른 이름으로 저장
 - → CSV(쉼표로 분리) 선택 후 → 파일명 입력
- 4) 엑셀 종료 후 Notepad로 파일 열어보기

엑셀로 CSV 파일 만들기

데이터가 "," 로 나눠져 있음

```
- - X
                       korea floating population data.csv - 메모장
파일이 관심이 서식이 보기에 도움함이 가장 전기에 도움함이 가장 전기에 되었다. 이 자신에 가장 전기에 가장 전기에
                                                                                                                                   서식(<u>○</u>) 보기(<u>V</u>) 도움말(<u>H</u>)
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            , 여 자10대 , 여 자20대 , 여 자30대 , 여 자40대 , 여 자50대
4, 68, 50, 31 , 4, 37, 64, 44, 26
                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         ,2,24,68,50,31,4,37,64,44,26
,19,44,28,33,21,14,56,49,43,18
,13,33,34,61,55,13,32,29,28,12
,23,33,32,547,129,12,39,13,46,4
4시 서구 월평동,맑음,4,22,12,6,5,24,20,10,6
4시 서구 월평동,맑음,4,22,12,22,12,6,30,27,7,3
4시 서구 월평동,흐림,3,8,5,8,11,2,8,9,4,3
```

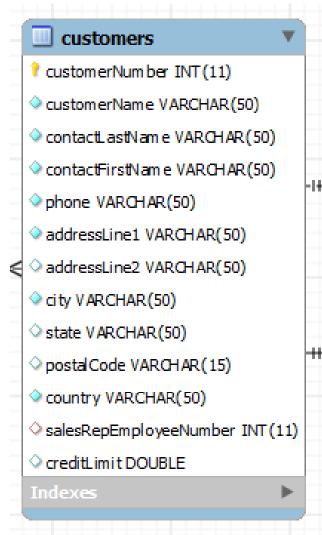
CSV 파일 다루기

CSV 파일 읽기/쓰기

- Text 파일 형태로 데이터 처리 예제
- 예제 데이터: customer.csv (https://goo.gl/FlsMqB)
- 일반적 textfile을 처리하듯 파일을 읽어온 후, 한 줄 한 줄씩 데이터를 처리함

http://goo.gl/1JdKst

CSV 파일 예제



해당데이터는 고객의 FullName, 성, 이름, 전화번호, 국가 등 기록 "", 로 분리

```
customers.csv - 메모장
| Castomerkunder.customerkane.contactLastName.contactFirstName.phone.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.addressLinel.a
                           파일(E) 편집(E) 서식(Q) 보기(V) 도움말(H)
                            customerNumber,customerName,contactLastName,contactFirstName,phone,addressLine1,addressLine2,city,state,postalCode,country,salesRepEmployeeNumber,creditLimit
```

CSV 파일 읽기 예제

```
data header = [] #data의 필드값을 저장하는 list
customer list = [] #cutomer 개별 List를 저장하는 List
with open ("customers.csv") as customer data: #customer.csv 파일을 customer data 객체에 저장
   while 1:
      data = customer_data.readline() #customer.csv에 한줄씩 data 변수에 저장
      if not data: break #데이터가 없을 때. Loop 종료
      if line counter==0: #첫번째 데이터는 데이터의 필드
         data_header = data.split(",") #데이터의 필드는 data_header List에 저장, 데이터 저장시 ","로 분리
      else:
         customer list.append(data.split(",")) #일반 데이터는 customer list 객체에 저장, 데이터 저장시 ","로 분리
      line counter += 1
print("Header:\t". data header) #데이터 필드 값 출력
for i in range(0,10): #데이터 출력 (샘플 10개만)
   print ("Data",i,":\t\t\t\",customer_list[i])
print (len(customer list)) #전체 데이터 크기 출력
```

CSV 파일 쓰기 예제

```
line counter = 0
data_header = []
employee = []
customer_USA_only_list = []
customer = None
with open ("customers.csv", "r") as customer_data:
   while 1:
       data = customer_data.readline()
       if not data:
           break
       if line_counter==0:
           data header = data.split(",")
       else:
           customer = data.split(",")
           if customer[10].upper() == "USA": #customer 데이터의 offset 10번째 값
              customer_USA_only_list.append(customer) #즉 country 필드가 "USA" 것만
       line counter+=1
                           #sutomer USA only list에 저장
print ("Header :\footnote{\text{W}t}", data_header)
for i in range(0,10):
   print (len(customer USA only list))
with open ("customers_USA_only.csv", "w") as customer_USA_only_csv:
   for customer in customer_USA_only_list:
       customer_USA_only_csv.write(",".join(customer).strip('\n')+"\n")
       #cutomer_USA_only_list 객체에 있는 데이터를 customers_USA_only.csv 파일에 쓰기
```

CSV 객체 사용하기

CSV 객체로 CSV 처리

- Text 파일 형태로 데이터 처리시 문장 내에 들어가 있는 "," 등에 대해 전처리 과정이 필요
- 파이썬에서는 간단히 CSV파일을 처리하기 위해 csv 객체를 제공함
- 예제 데이터: korea_floating_population_data.csv (from http://www.data.go.kr)
- 예제 데이터는 국내 주요 상권의 유동인구 현황 정보
- 한글로 되어 있어 한글 처리가 필요

CSV 객체로 CSV 처리

- 시간대/조사일자/행정구역/날씨 등을 기준으로 연령별/성별 유동인가 해당 지역에 몇 명인가 표시

https://goo.gl/ujHuXQ

CSV 객체 활용

Attribute	Default	Meaning
delimiter	ı	글자를 나누는 기준
lineterminator	₩r₩n	줄 바꿈 기준
quotechar	11	문자열을 둘러싸는 신호 문자
quoting	QUOTE_MINIMAL	데이터 나누는 기준이 quotechar에 의해 둘러싸인 레벨

https://pymotw.com/2/csv/

CSV 객체 활용

유동 인구 데이터 중 성남 데이터만 수집

```
import csv
seoung nam data = []
header = []
rownum = 0
with open("korea_floating_population_data.csv", "r", encoding="cp949") as p_file:
   csv_data = csv.reader(p_file) #csv 객체를 이용해서 csv_data 읽기
   for row in csv data: #읽어온 데이터를 한 줄씩 처리
      if rownum == 0:
          header = row #첫 번째 줄은 데이터 필드로 따로 저장
      location = row[7]
      #"행정구역"필드 데이터 추출, 한글 처리로 유니코드 데이터를 cp949로 변환
      if location.find(u"성남시") != -1:
          seoung_nam_data.append(row)
      #" 행정구역" 데이터에 성남시가 들어가 있으면 seoung nam data List에 추가
      rownum +=1
with open("seoung_nam_floating_population_data.csv", "w", encoding="utf8") as s_p_file:
— writer = csv.writer(s p file, delimiter='\t', quotechar="'", quoting=csv.QUOTE ALL)
   # csv.writer를 사용해서 csv 파일 만들기 delimiter 필드 구분자
   # quotechar는 필드 각 데이터는 묶는 문자, quoting는 묶는 범위
   writer.writerow(header)
                             #제목 필드 파일에 쓰기
   for row in seoung_nam_data:
      writer.writerow(row)
                           #seoung nam data에 있는 정보 list에 쓰기
```



Human knowledge belongs to the world.



게임을 만들었는데 HACK쓰는 애들 때문에 망했어요...

어떻게 잡을 수 있을까?

일단은 기록부터!!

로그 남기기 - Logging

- 프로그램이 실행되는 동안 일어나는 정보를 기록을 남기기
- 유저의 접근, 프로그램의 Exception, 특정 함수의 사용
- Console 화면에 출력, 파일에 남기기, DB에 남기기 등등
- 기록된 로그를 분석하여 의미있는 결과를 도출 할 수 있음
- 실행시점에서 남겨야 하는 기록, 개발시점에서 남겨야하는 기록

print vs logging

- 기록을 print로 남기는 것도 가능함
- 그러나 Console 창에만 남기는 기록은 분석시 사용불가
- 때로는 레벨별(개발, 운영)로 기록을 남길 필요도 있음
- 모듈별로 별도의 logging을 남길필요도 있음
- 이러한 기능을 체계적으로 지원하는 모듈이 필요함

logging 모듈

- Python의 기본 Log 관리 모듈

```
import logging
logging.debug("틀렸잖아!")
logging.info("확인해")
logging.warning("조심해!")
logging.error("에러났어!!!")
logging.critical ("망했다...")
```

logging level

- 프로그램 진행 상황에 따라 다른 Level의 Log를 출력함
- 개발 시점, 운영 시점 마다 다른 Log가 남을 수 있도록 지원함
- DEBUG > INFO > WARNING > ERROR > Critical
- Log 관리시 가장 기본이 되는 설정 정보

logging level https://stackoverflow.com/questions/2031163/when-to-use-the-different-log-levels

Level	개요	예시
debug		- 다음 함수로 A 를 호출함 - 변수 A 를 무엇으로 변경함
info	처리가 진행되는 동안의 정보를 알림	- 서버가 시작되었음 - 서버가 종료됨 - 사용자 A가 프로그램에 접속함
warning	사용자가 잘못 입력한 정보나 처리는 가능하나 원래	- Str입력을 기대했으나, Int가 입력됨 → Str casting으로 처리함 - 함수에 argument로 이차원 리스트를 기대했으나 → 일차원 리스트가 들어옴, 이차원으로 변환후 처리
error	잘못된 처리로 인해 에러가 났으나, 프로그램은 동작 할 수 있음을 알림	- 파일에 기록을 해야하는데 파일이 없음> Exception 처리후 사용자에게 알림 - 외부서비스와 연결 불가
critical	잘못된 처리로 데이터 손실이나 더이상 프로그램이 동작할 수 없음을 알림	- 잘못된 접근으로 해당 파일이 삭제됨 - 사용자의 의한 강제 종료

logging level

import logging

```
logger = logging.getLogger("main")

stream_hander = logging.StreamHandler()

Logger의 output 방법 선언

logger.addHandler(stream_hander)

Logger의 output 등록
```

logger.setLevel(logging.DEBUG) logger.setLevel(logging.CRITICAL) logger.debug("틀렸잖아!") logger.info("확인해") logger.info("확인해") logger.warning("조심해!") logger.warning("조심해!") logger.error("에러났어!!!") logger.critical("망했다...")



Human knowledge belongs to the world.



실제 프로그램을 실행할 땐 여러 설정이 필요

데이터 파일 위치 파일 저장 장소 Operation Type 등

이러한 정보를 설정해줄 방법이 필요

- 1) Configparser 파일에
- 2) Argparser 실행시점에

Configparser

Configparser

- 프로그램의 실행 설정을 file에 저장함
- Section, Key, Value 값의 형태로 설정된 설정 파일을 사용
- 설정파일을 Dict Type으로 호출후 사용

Config file

[SectionOne]

Status: Single

Name: Derek

Value: Yes

Age: 30

Single: True

Section- 대괄호

속성 - Key : Value

[SectionTwo]
FavoriteColor = Green

[SectionThree]

FamilyName: Johnson

Configparser file

```
import configparser
config = configparser.ConfigParser()
config.sections()
config.read('example.cfg')
config.sections()
for key in config['SectionOne']:
    print(key)
config['SectionOne']["status"]
```

'example.cfg'

[SectionOne]
Status: Single
Name: Derek
Value: Yes

Age: 30

Single: True

[SectionTwo]
FavoriteColor = Green

[SectionThree] FamilyName: Johnson

Argparser

Argparser

- Console 창에서 프로그램 실행시 Setting 정보를 저장함
- 거의 모든 Console 기반 Python 프로그램 기본으로 제공
- 특수 모듈도 많이 존재하지만(TF), 일반적으로 argparse를 사용
- Command-Line Option 이라고 부름

```
Х

    C:₩WINDOWS₩system32₩cmd.exe

 :#workspace#introduction_to_pythoy_TEAMLAB_MOOC#code>python --v
Unknown option: --
usage: python [option] ... [-c cmd | -m mod | file | -] [arg] ...
Try `python -h' for more information.
F:#workspace#introduction_to_pythoy_TEAMLAB_MOOC#code>python -h
usage: python [option] ... [-c cmd | -m mod | file | <del>-] [arg] ..</del>.
Options and arguments (and corresponding environment variables):
       issue warnings about str(bytes_instance), str(bytearray_instance)
and comparing bytes/bytearray with str. (-bb: issue errors)
       : don't write .py[co] files on import; also PYTHONDONTWRITEBYTECODE=x
  cmd : program passed in as string (terminates option list)
       : debug output from parser; also PYTHONDEBUG=x
         ignore PYTHON* environment variables (such as PYTHONPATH)
       : print this help message and exit (also --help)
         inspect interactively after running script; forces a prompt even
         if stdin does not appear to be a terminal; also PYTHONINSPECT=x
       : isolate Python from the user's environment (implies -E and -s)
  mod : run library module as a script (terminates option list)
       : optimize generated bytecode slightly; also PYTHONOPTIMIZE=x
       : remove doc-strings in addition to the -0 optimizations
       : don't print version and copyright messages on interactive startup
       : don't add user site directory to sys.path; also PYTHONNOUSERSITE
        don't imply 'import site' on initialization
       : unbuffered binary stdout and stderr, stdin always buffered;
         also PYTHONUNBUFFERED=x
         see man page for details on internal buffering relating to '-u'
       : verbose (trace import statements); also PYTHONVERBOSE=x
         can be supplied multiple times to increase verbosity
```

Argparser

import argparse

```
parser = argparse.ArgumentParser(description='Sum two integers.')

짧은 이름 긴 이름 표시명 Help 설명 Type

parser.add_argument('-a', "--a_value", dest="A_value", help="A integers", type=int)

parser.add_argument('-b', "--b_value", dest="B_value", help="B integers", type=int)

args = parser.parse_args()

print(args)

print(args.a)

print(args.b)

print(args.a + args.b)
```



Human knowledge belongs to the world.



Logging formmater

- Log의 결과값의 format을 지정해줄 수 있음

```
formatter =
logging.Formatter('%(asctime)s %(levelname)s %(process)d %(message)s')
```

2018-01-18 22:47:04,385 ERROR 4410 ERROR occurred 2018-01-18 22:47:22,458 ERROR 4439 ERROR occurred 2018-01-18 22:47:22,458 INFO 4439 HERE WE ARE 2018-01-18 22:47:24,680 ERROR 4443 ERROR occurred 2018-01-18 22:47:24,681 INFO 4443 HERE WE ARE 2018-01-18 22:47:24,970 ERROR 4445 ERROR occurred 2018-01-18 22:47:24,970 INFO 4445 HERE WE ARE

Log config file

```
logging.config.fileConfig('logging.conf')
logger = logging.getLogger()
```

```
[loggers]
keys=root
[handlers]
keys=consoleHandler
[formatters]
keys=simpleFormatter
[logger_root]
level=DEBUG
handlers=consoleHandler
[handler_consoleHandler]
class=StreamHandler
level=DEBUG
formatter=simpleFormatter
args=(sys.stdout,)
```

Logging examples

```
logger.info('Open file {0}'.format("customers.csv",))
try:
  with open("customers.csv", "r") as customer_data:
    customer_reader = csv.reader(customer_data, delimiter=',', quotechar='"')
    for customer in customer reader:
      if customer[10].upper() == "USA": #customer 데이터의 offset 10번째 값
        logger.info('ID {0} added'.format(customer[0],))
        customer_USA_only_list.append(customer) #즉 country 필드가 "USA" 것만
except FileNotFoundError as e:
    logger.error('File NOT found {0}'.format(e,))
```



Human knowledge belongs to the world.