

파일 읽기

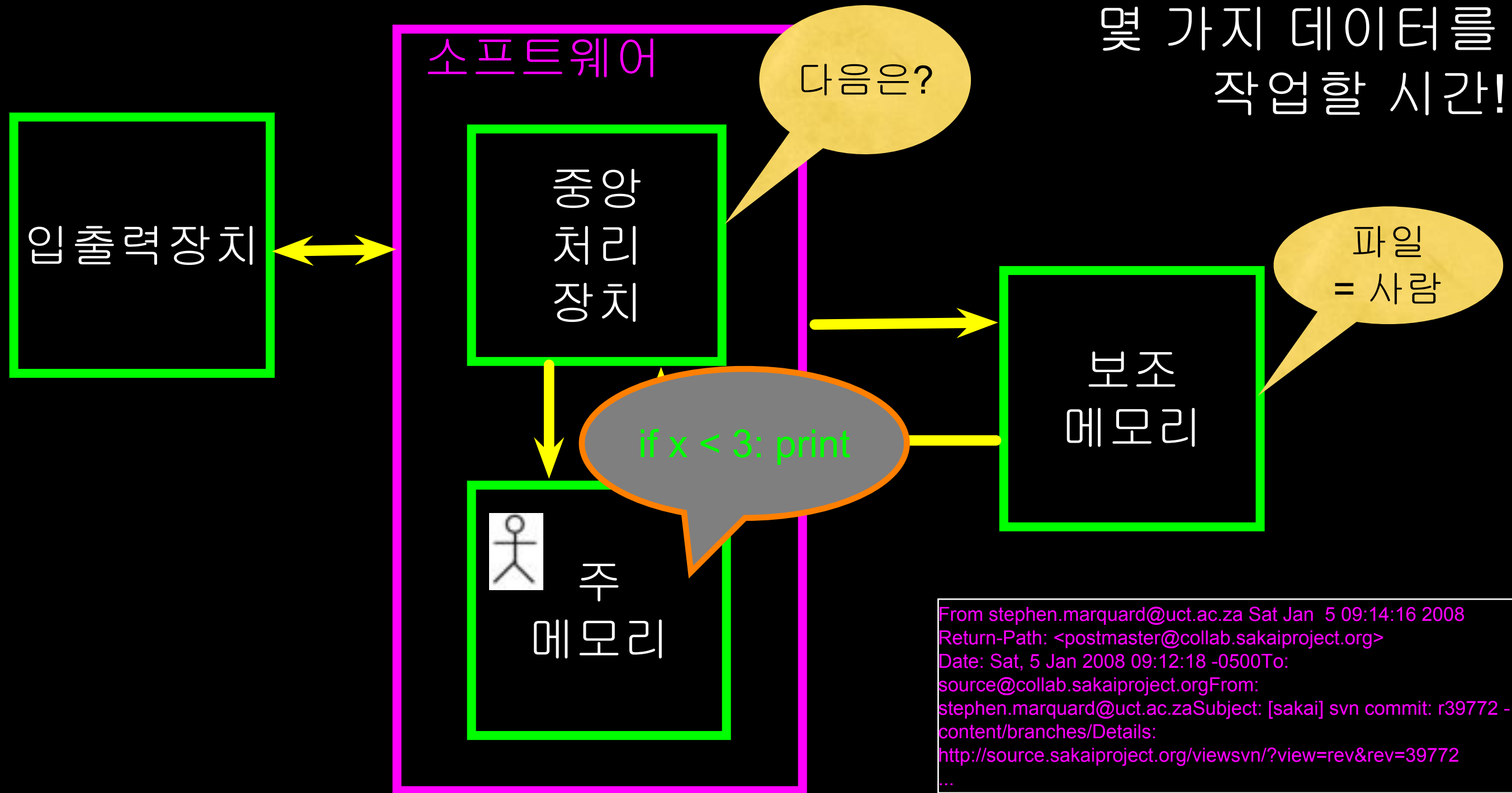
제7장



Python for Everybody
www.py4e.com



몇 가지 데이터를 찾아
작업할 시간!



파일 처리

텍스트 파일은 일련의 줄이 나열된 것으로 여길 수 있음

From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008

Return-Path: <postmaster@collab.sakaiproject.org>

Date: Sat, 5 Jan 2008 09:12:18 -0500

To: source@collab.sakaiproject.org

From: stephen.marquard@uct.ac.za

Subject: [sakai] svn commit: r39772 - content/branches/

Details: <http://source.sakaiproject.org/viewsvn/?view=rev&rev=39772>

<http://www.py4e.com/code/mbox-short.txt>

파일 열기

- 파일의 내용을 읽기 전에 파이썬에게 작업할 파일과 파일로 어떤 작업을 할지 알려줘야 함
- `open()` 함수가 해당 기능을 수행함
- `open()`이 “파일 핸들”을 반환 - 파일에 대한 작업을 수행하기 위해 사용하는 변수
- 워드 프로세서에서 “File -> Open” 과 비슷함

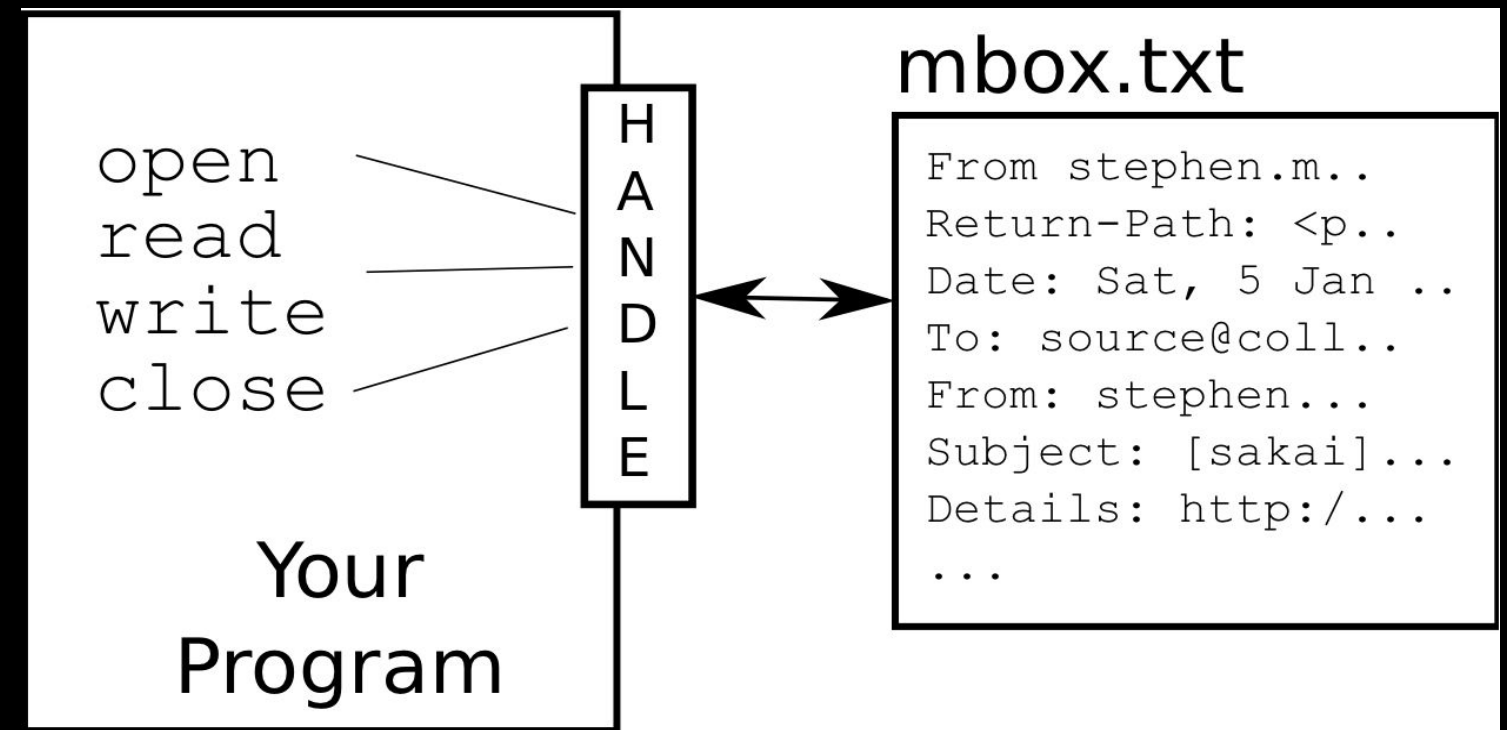
open() 사용

```
fhand = open('mbox.txt', 'r')
```

- 핸들 = open(파일명, 모드)
- 파일을 조작하는데 쓰는 핸들을 반환함
- 파일명에는 문자열이 들어감
- 모드에 매개 변수를 넣는 것은 선택 사항이며
파일을 읽으려면 'r'을, 파일에 쓰려면 'w'를 입력

핸들이란?

```
>>> fhand = open('mbox.txt')
>>> print(fhand)
<_io.TextIOWrapper name='mbox.txt' mode='r' encoding='UTF-8'>
```



존재하지 않는 파일인 경우

```
>>> fhand = open('stuff.txt')
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
FileNotFoundError: [Errno 2] No such file or
directory: 'stuff.txt'
```

개행 문자

- 각 줄이 끝날 때 이를 알리기 위해 “개행 문자”라고 부르는 특수한 문자를 사용
- 문자열에서는 `\n`으로 표현함
- 개행 문자는 한 글자
 - 두 글자가 아님

```
>>> stuff = 'Hello\nWorld!'
>>> stuff
'Hello\nWorld!'
>>> print(stuff)
Hello
World!
>>> stuff = 'X\nY'
>>> print(stuff)
X
Y
>>> len(stuff)
3
```


파일 처리

텍스트 파일은 일련의 줄이 나열된 것으로 여길 수 있음

From stephen.marquard@uct.ac.za Sat Jan 5 09:14:16 2008

Return-Path: <postmaster@collab.sakaiproject.org>

Date: Sat, 5 Jan 2008 09:12:18 -0500

To: source@collab.sakaiproject.org

From: stephen.marquard@uct.ac.za

Subject: [sakai] svn commit: r39772 - content/branches/

Details: <http://source.sakaiproject.org/viewsvn/?view=rev&rev=39772>

파일 처리

텍스트 파일은 각 줄 끝에 개행 문자가 있음

```
From stephen.marquard@uct.ac.za Sat Jan  5 09:14:16 2008\nReturn-Path: <postmaster@collab.sakaiproject.org>\nDate: Sat, 5 Jan 2008 09:12:18 -0500\nTo: source@collab.sakaiproject.org\nFrom: stephen.marquard@uct.ac.za\nSubject: [sakai] svn commit: r39772 - content/branches/\n\nDetails: http://source.sakaiproject.org/viewsvn/?view=rev&rev=39772\n
```

파이썬에서 파일 읽기

시퀀스로써 파일 핸들

- 읽기용으로 열린 파일 핸들은 파일의 각 줄에 대한 문자열의 시퀀스로 볼 수 있음
- **for** 문을 사용하여 시퀀스를 반복하여 돌 수 있음
- 시퀀스는 정렬된 집합임을 기억하기

```
xfile = open('mbox.txt')  
for cheese in xfile:  
    print(cheese)
```

파일의 줄 세기

- 파일을 읽기 모드로 열기
- **for** 문을 이용하여 각 줄을 읽기
- 줄을 세고 줄의 수를 출력하기

```
fhand = open('mbox.txt')
count = 0
for line in fhand:
    count = count + 1
print('Line Count:', count)
```

```
$ python open.py
Line Count: 132045
```

파일 *전체*를 읽기

파일 전체(개행 문자 포함 전체)를
하나의 문자열로 읽을 수 있음

```
>>> fhand = open('mbox-short.txt')
>>> inp = fhand.read()
>>> print(len(inp))
94626
>>> print(inp[:20])
From stephen.marquar
```

파일 내용 탐색

for 문에 **if** 문을 넣어서
특정 조건을 만족하는 줄만 출력

```
fhand = open('mbox-short.txt')
for line in fhand:
    if line.startswith('From:') :
        print(line)
```

이런!

줄과 줄 사이의 빈 줄은
무엇일까?

From: `stephen.marquard@uct.ac.za`

From: `louis@media.berkeley.edu`

From: `zqian@umich.edu`

From: `rjlowe@iupui.edu`

...

이런!

줄과 줄 사이의 빈 줄은
무엇일까?

- 파일의 각 줄은 끝에
개행 문자를 갖고 있음
- **print** 구문이 각 줄에
개행 문자를 추가로 더함

```
From: stephen.marquard@uct.ac.za\n\nFrom: louis@media.berkeley.edu\n\nFrom: zqian@umich.edu\n\nFrom: rjlowe@iupui.edu\n\n...
```

파일 내용 탐색 (전형적인 방법)

- 문자열 라이브러리에 있는 `rstrip()`을 이용하여
오른쪽에서부터
공백을 지울 수 있음
- 개행 문자는 “공백”으로
취급되어 제거됨

```
fhand = open('mbox-short.txt')
for line in fhand:
    line = line.rstrip()
    if line.startswith('From:') :
        print(line)
```

```
From: stephen.marquard@uct.ac.za
From: louis@media.berkeley.edu
From: zqian@umich.edu
From: rjlowe@iupui.edu
....
```

continue로 넘어가기

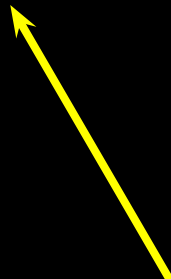
continue 문을 사용하여
편리하게 줄을 넘길 수
있음

```
fhand = open('mbox-short.txt')
for line in fhand:
    line = line.rstrip()
    if not line.startswith('From:') :
        continue
    print(line)
```

줄을 선택하기 위해 in 사용하기

설정한 선택 기준에 맞춰 줄 안
어디에서든 문자열을 찾을 수
있음

```
fhand = open('mbox-short.txt')
for line in fhand:
    line = line.rstrip()
    if not '@uct.ac.za' in line :
        continue
    print(line)
```



```
From stephen.marquard@uct.ac.za Sat Jan  5 09:14:16 2008
X-Authentication-Warning: set sender to stephen.marquard@uct.ac.za using -f
From: stephen.marquard@uct.ac.za
Author: stephen.marquard@uct.ac.za
From david.horwitz@uct.ac.za Fri Jan  4 07:02:32 2008
X-Authentication-Warning: set sender to david.horwitz@uct.ac.za using -f...
```

```
fname = input('Enter the file name: ')
fhand = open(fname)
count = 0
for line in fhand:
    if line.startswith('Subject:') :
        count = count + 1
print('There were', count, 'subject lines in', fname)
```

파일명에 대한 프롬프트

Enter the file name: mbox.txt

There were 1797 subject lines in mbox.txt

Enter the file name: mbox-short.txt

There were 27 subject lines in mbox-short.txt

부적절한 파일명

```
fname = input('Enter the file name: ')
try:
    fhand = open(fname)
except:
    print('File cannot be opened:', fname)
    quit()

count = 0
for line in fhand:
    if line.startswith('Subject:') :
        count = count + 1
print('There were', count, 'subject lines in', fname)
```

Enter the file name: mbox.txt

There were 1797 subject lines in mbox.txt

Enter the file name: na na boo boo

File cannot be opened: na na boo boo

요약

- 보조 저장소
- 파일 열기 - 파일 핸들
- 파일 구조 - 개행 문자
- 루프를 이용하여
한 줄씩 한 줄씩 파일 읽기
- 줄 찾기
- 파일명 읽기
- 잘못 입력된 파일
다루기



Acknowledgements / Contributions



These slides are Copyright 2010- Charles R. Severance (www.dr-chuck.com) of the University of Michigan School of Information and open.umich.edu and made available under a Creative Commons Attribution 4.0 License. Please maintain this last slide in all copies of the document to comply with the attribution requirements of the license. If you make a change, feel free to add your name and organization to the list of contributors on this page as you republish the materials.

Initial Development: Charles Severance, University of Michigan School of Information

Contributor:

- Seung-June Lee (plusjune@gmail.com)
- Connect Foundation

Translator:

- Hyunjun Kim
- Jeungmin Oh (tangza@gmail.com)