

Читання файлів

Розділ 7

Python для всіх
www.py4e.com





Обробка файлів

Текстовий файл можна розглядати як послідовність рядків

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>

Відкриття файлу

- Перш ніж ми зможемо прочитати вміст файлу, ми повинні повідомити Python, з яким файлом ми збираємося працювати і що ми будемо з ним робити
- Це робиться за допомогою функції `open()`
- `open()` повертає «обробник файлу» — змінну, яка використовується для виконання операцій над файлом
- Схоже на «File -> Open» у Word

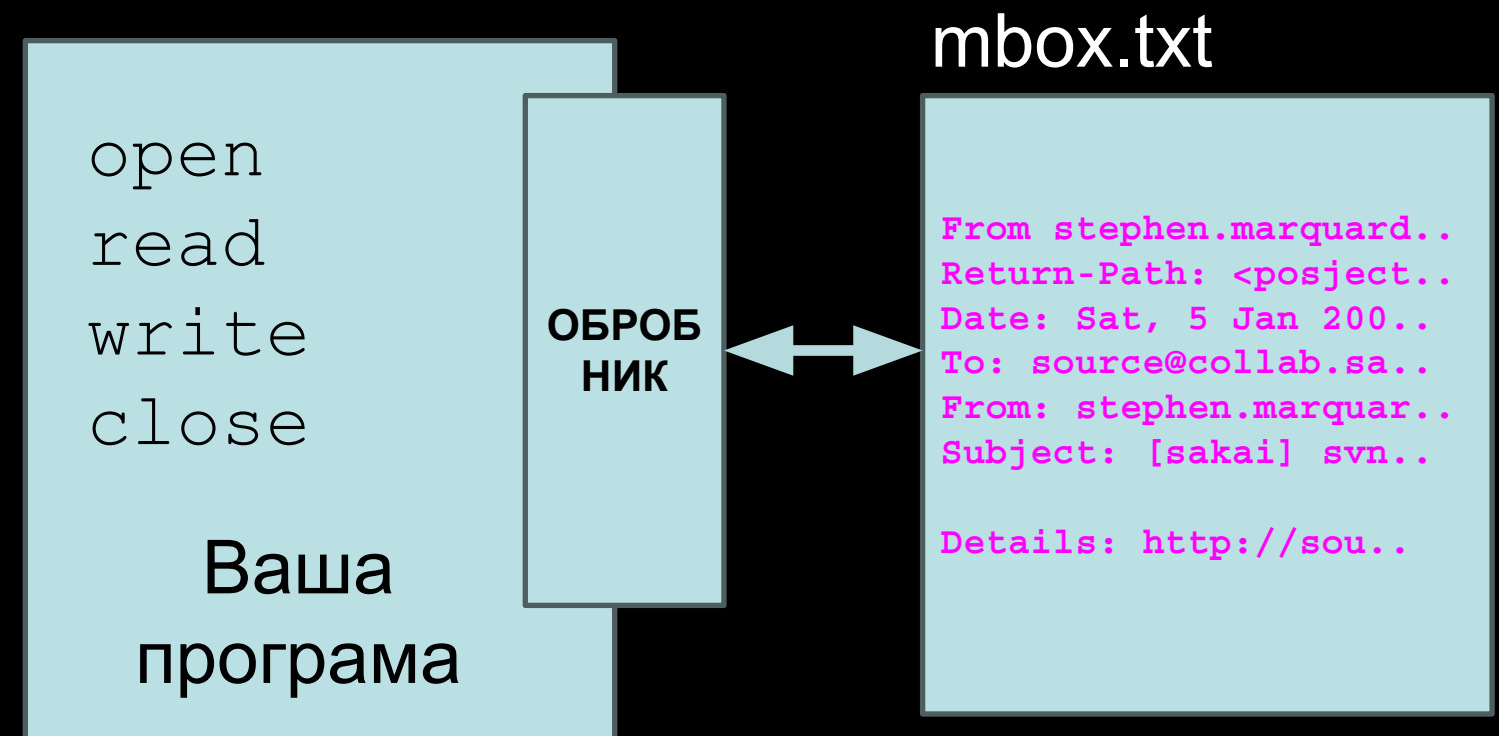
Використання open()

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

- `handle = open(filename, mode)`
- повертає обробник - змінну `handle` для маніпулювання файлом
- `filename` має бути рядком
- `mode` не є обов'язковим і може бути 'r', якщо ми плануємо читати файл, і 'w', якщо ми плануємо записувати до файлу

Що таке обробник?

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



Коли файли відсутні

```
>>> 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'
```

Символ `newline` (перехід на новий рядок)

- Ми використовуємо спеціальний символ, який називається «перехід на новий рядок», щоб вказати, коли рядок закінчується
- У рядках ми його подаємо як `\n`
- `Newline` (перехід на новий рядок) – це один символ, а не два

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


Обробка файлів

Текстовий файл можна розглядати як послідовність рядків

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>

Обробка файлів

Текстовий файл містить **переходи на новий рядок** в кінці кожного рядка

```
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
```

Читання файлів у Python

Обробник файлу як послідовність

- Обробник файлу, відкритий для читання, можна розглядати як **послідовність** рядків, де кожен рядок у файлі є рядком у послідовності.
- Можна використовувати інструкцію **for** для ітерацій по **послідовності**
- Пам'ятайте, що **послідовність** — це впорядкована множина

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

Підрахунок рядків у файлі

- Відкрийте **файл** у режимі читання
- Використовуйте цикл **for** для читання кожного рядка
- **Підрахуйте** рядки і виведіть кількість рядків

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

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

Читання *всього* файлу

Ми можемо прочитати
весь файл (разом з
переходами на новий
рядок) в один рядок

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

Пошук у файлі

Ми можемо додати інструкцію **if** в наш цикл **for**, щоб виводити тільки ті рядки, які відповідають певним критеріям

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

ООЙЙ!

Що тут роблять усі ці
порожні рядки?

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

From: `louis@media.berkeley.edu`

From: `zqian@umich.edu`

From: `rjlowe@iupui.edu`

...

ООЙЙ!

Що тут роблять усі ці порожні рядки?

- Кожен рядок з файлу містить **перехід на новий рядок** в кінці
- Інструкція **print** додає **перехід новий рядок** до кожного рядка

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

Пошук у файлі (виправлено)

- Ми можемо видалити пробіли з правого боку рядка за допомогою функції `rstrip()` з бібліотеки рядків
- Перехід на новий рядок вважається «недрукованим символом» і видаляється

```
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
....
```

Пропуск з інструкцією `continue`

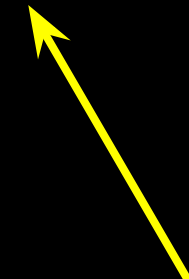
Ми можемо легко
пропустити рядок за
допомогою інструкції
`continue`

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

Використання `in` для відбору рядків

Ми можемо шукати рядок
у будь-якому місці рядка
`line` як наш критерій
відбору

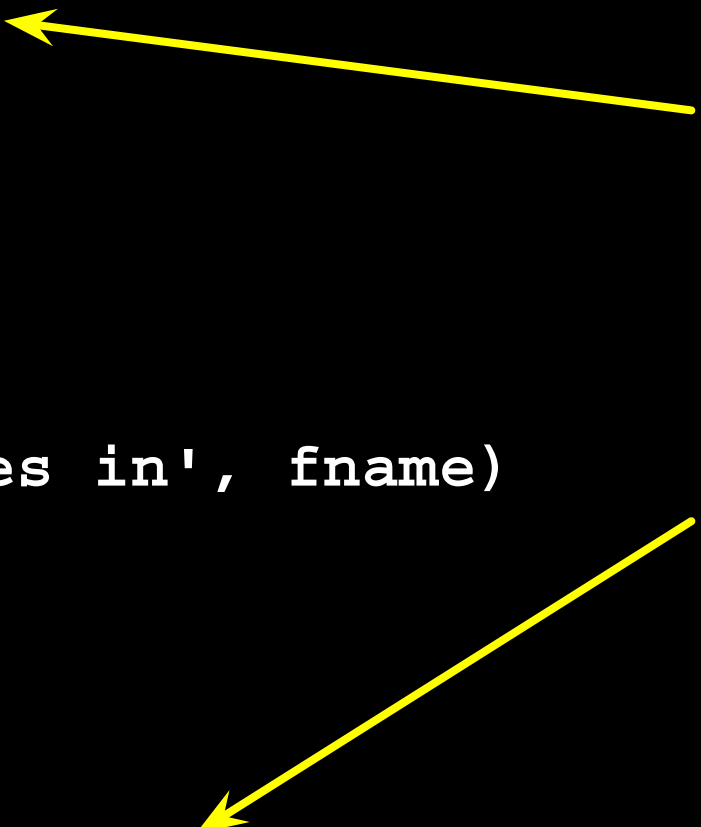
```
fhand = open('mbox-short.txt')
for line in fhand:
    line = line.rstrip()
    if not '@uct.ac.za' in line :
        continue
    print(line)
```



```
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...
```

Запит імені файлу

```
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)
```



Enter the file name : **mbox.txt**

У файлі mbox.txt було 1797 рядків теми

Введіть ім'я файлу : **mbox-short.txt**

У файлі mbox-short.txt було 27 рядків теми

Погані імена файлів

```
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)
```

Введіть ім'я файлу : mbox.txt

У файлі mbox.txt було 1797 рядків теми

Введіть ім'я файлу : na na boo boo

Неможливо відкрити файл: na na boo boo

Підсумки

- Постійна пам'ять
- Відкриття файлу – обробник файлу
- Структура файлу – символ переходу на новий рядок
- Порядкове читання файлу за допомогою циклу `for`
- Пошук рядків
- Читання імен файлів
- Поводження з пошкодженими файлами

Права власності / Застереження



Авторські права на ці слайди з 2010 року належать Чарльзу Северенсу (www.dr-chuck.com) зі Школи інформації Мічиганського університету та захищені ліцензією Creative Commons Attribution 4.0. Будь ласка, збережіть цей фінальний слайд у всіх копіях документа, щоб відповідати вимогам ліцензії

щодо посилань на джерела. При повторній публікації матеріалів, якщо щось зміните, додайте ім'я та організацію до переліку співавторів нижче.

Першоджерело: Чарльз Северенс, Школа інформації Мічиганського університету

Переклад: платформа Prometheus