

# Leggere i files

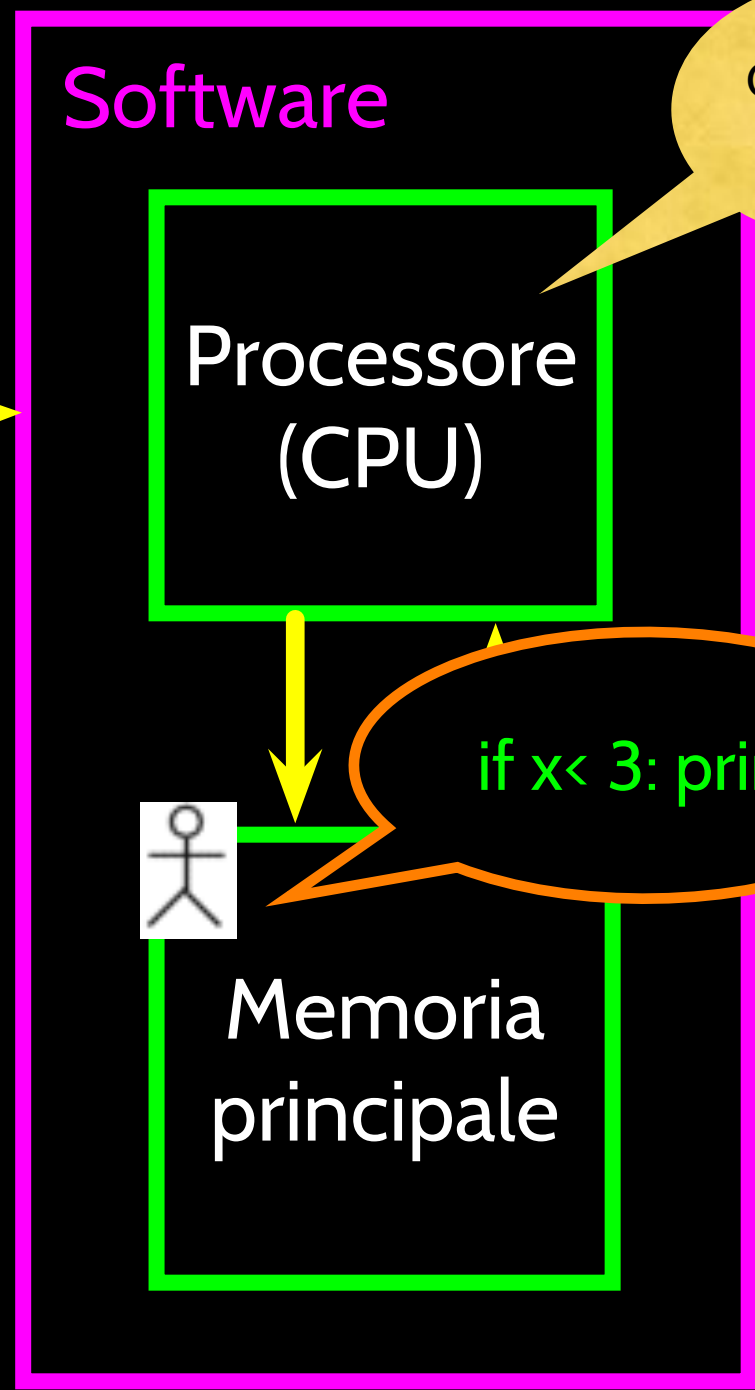
## Capitolo 7



Python for Informatics: Exploring Information  
[www.pythonlearn.com](http://www.pythonlearn.com)



Dispositivi di input e output



if x < 3: print

Cosa faccio ora?

I file siamo noi

Cerchiamo alcuni dati con cui giocare!

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  
...

# Elaborazione di file

- Si può pensare ad un file di testo come ad una sequenza di linee

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/?](http://source.sakaiproject.org/viewsvn/?view=rev&rev=39772)

[view=rev&rev=39772](http://source.sakaiproject.org/viewsvn/?view=rev&rev=39772)

<http://www.py4inf.com/code/mbox-short.txt>

# Aprire un file

- Prima di poter leggere il contenuto di un file, dobbiamo dire a Python qual'è il file su cui lavoreremo e cosa faremo con questo file
- Per fare ciò useremo la funzione `open()`
- `open()` restituisce un “descrittore di file (file handle)” - una variabile usata per effettuare operazioni sul file stesso
- Simile a “File -> Apri” in qualsiasi programma di scrittura

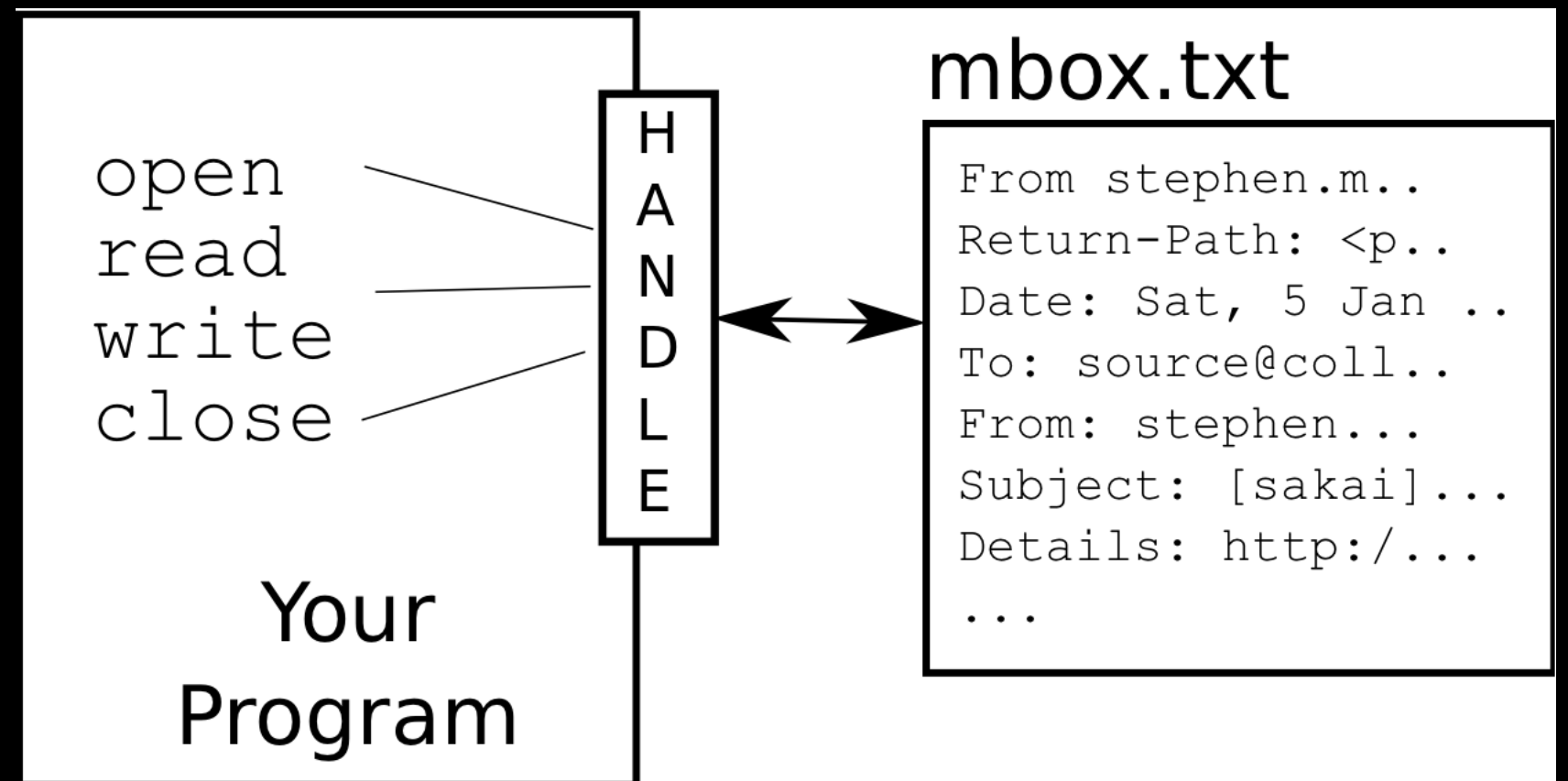
# Uso di open()

- `handle = open(filename, mode)`      `fhand = open('mbox.txt', 'r')`
  - > restituisce un descrittore usato per manipolare il file
  - > 'filename' (nome del file) è una stringa
  - > 'mode' è facoltativo e dovrebbe essere 'r' se pensiamo di leggere il file (read) e 'w' se pensiamo di modificare il file (write)

<http://docs.python.org/lib/built-in-funcs.html>

# Cosa è un descrittore (handle)?

```
>>> fhand = open('mbox.txt')
>>> print fhand
<open file 'mbox.txt', mode 'r' at
0x1005088b0>
```



# Quando un file non esiste

```
>>> fhand = open('stuff.txt')
```

```
Traceback (most recent call last): File
```

```
"<stdin>", line 1, in <module>IOError:
```

```
[Errno 2] No such file or directory:
```

```
'stuff.txt'
```

# Il carattere 'nuovalinea'

- Usiamo un carattere speciale detto "nuovalinea" per indicare dove finisce una linea
- Si rappresenta come `\n` nelle stringhe
- **Nuovalinea** rimane un solo carattere, non due

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



# Elaborazione di file

- Si può pensare ad un file di testo come ad una sequenza di linee

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/?](http://source.sakaiproject.org/viewsvn/?view=rev&rev=39772)

[view=rev&rev=39772](http://source.sakaiproject.org/viewsvn/?view=rev&rev=39772)

# Elaborazione di file

- Un file di testo presenta un carattere 'nuovalinea' alla fine di ogni linea

```
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/\nDetails: http://source.sakaiproject.org/viewsvn/?view=rev&rev=39772\n
```

# Descrittore di file come sequenza

- Un **descrittore di file** aperto per sola lettura può essere visto come una **sequenza** di stringhe dove ogni linea nel file è una stringa della sequenza
- È possibile utilizzare l'istruzione **for** per scorrere lungo una **sequenza**
- Ricorda - una **sequenza** è un insieme ordinato

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

# Come contare linee in un file

- Apri un **file** in modalità sola lettura
- Usa un ciclo **for** per leggere ogni linea
- **Conta** le linee e mostra (print) il numero di linee contate

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

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

# Leggere il file `*intero*`

- È possibile `leggere` l'intero file (comprese nuove linee e tutto il resto) in una `singola stringa`.

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

# Cercare in un file

- È possibile inserire una condizione **if** nel nostro ciclo **for** per mostrare solo le linee che si adattano ad un criterio preciso

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

# OOPS!

Cosa sono tutte queste  
linee vuote?

From: [stephen.marquard@uct.ac.za](mailto:stephen.marquard@uct.ac.za)

From: [louis@media.berkeley.edu](mailto:louis@media.berkeley.edu)

From: [zqian@umich.edu](mailto:zqian@umich.edu)

From: [rjlowe@iupui.edu](mailto:rjlowe@iupui.edu)

...

# OOPS!

Cosa sono tutte queste  
linee vuote?

- Ciascuna **linea** del file ha un carattere **nuovalinea** alla fine
- Il comando **print** aggiunge una **nuovalinea** ad ogni linea

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



# Cercare in un file

- È possibile rimuovere gli spazi bianchi dalla parte destra della stringa usando `rstrip()` dalla libreria delle stringhe
- Il carattere nuovalinea è considerato uno "spazio bianco" e `rimosso`

```
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
....
```

# Omettere con continue

- È possibile omettere linee a convenienza usando il comando **continue**

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

# Uso di **in** per selezionare **linee**

- È possibile cercare sequenze precise di caratteri **in** una **linea** secondo i nostri criteri di ricerca

```
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](mailto: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 = raw_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
```

Richiedi il  
nome del file

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

# Bad File Names (nome del file errato)

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

# Sommario

- Memoria secondaria
- Cercare linee specifiche
- Aprire un file - file handle (descrittore di file)
- Leggere nomi di file
- Struttura del file - carattere nuovovalinea
- Trattare file corrotti
- Leggere un file linea per linea con un ciclo for



# Acknowledgements / Contributions



These slides are Copyright 2010- Charles R. Severance ([www.dr-chuck.com](http://www.dr-chuck.com)) of the University of Michigan School of Information and [open.umich.edu](http://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