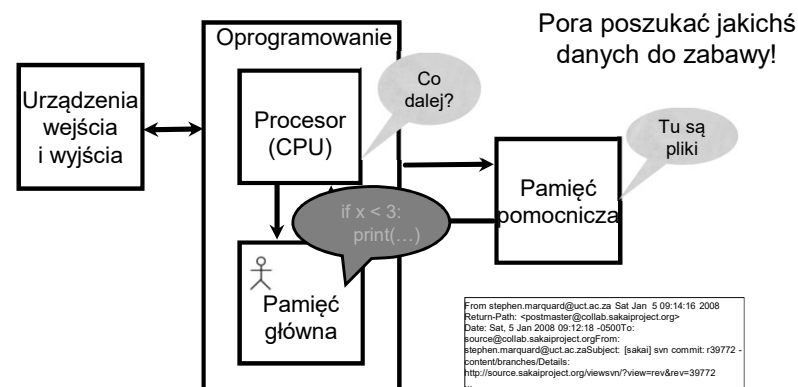


Czytanie plików

Rozdział 7



Python dla wszystkich
www.py4e.pl



Przetwarzanie pliku

Plik tekstowy może być uważany za sekwencję linii.

```
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
```

<https://py4e.pl/code3/mbox-short.txt>

Otwieranie pliku

- Zanim wczytamy zawartość pliku, musimy powiedzieć Pythonowi, z którym plikiem chcemy pracować i co chcemy z nim robić
- Do tego służy funkcja `open()`
- `open()` zwraca "uchwyt pliku" – zmienną służącą do wykonywania operacji na pliku
- To podobne do "Plik -> Otwórz" w edytorze tekstu

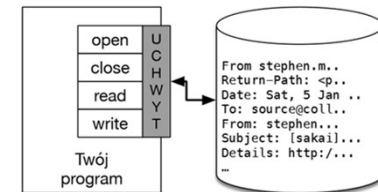
Używanie open()

- `uchwyt = otwórz(nazwa pliku, tryb)`
- zwraca uchwyt do obsługi pliku
- nazwa pliku jest ciągiem znaków
- wybór trybu jest opcjonalny i należy ustawić 'r', jeśli chcemy czytać (read) plik, i 'w' – jeśli chcemy do niego zapisywać (write)

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

Czym jest uchwyt?

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



Nie można odnaleźć pliku

```
>>> 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'
```

Znak końca linii

- Aby zaznaczyć, gdzie kończy się linia, używamy specjalnego "znaku"
- W ciągach jest reprezentowany przez `\n`
- Znak końca linii to w rzeczywistości jeden znak – nie dwa

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

Przetwarzanie pliku

Plik tekstowy może być uważany za sekwencję linii.

```
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
```

Czytanie plików w Pythonie

Przetwarzanie pliku

Na końcu każdej linii w pliku tekstowym znajduje się znak końca linii.

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

Uchwyty pliku jako sekwencja

- uchwyt pliku otwartego do czytania można traktować jak sekwencję ciągów znaków, w której każda linia pliku jest jednym ciągiem
- Możemy użyć instrukcji for, żeby przejść przez sekwencje
- Pamiętaj, że sekwencja jest uporządkowanym zbiorem

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

Liczenie linii w pliku

- Otwórz plik w trybie do odczytu
- Użyj pętli for, żeby wczytać każdą linię
- Zlicz linie i wypisz ich liczbę

```
fhand = open('mbox-short.txt')
count = 0
for line in fhand:
    count = count + 1
print('liczba linii:', count)
```

```
$ python3 open.py
liczba linii: 1910
```

Wczytywanie *całego* pliku

Możemy wczytać cały plik (razem ze znakami końca linii) do jednego ciągu znaków

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

Przeszukiwanie pliku

Możemy umieścić instrukcję if wewnątrz pętli for, żeby wypisać tylko linie spełniające pewne kryteria

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

Ups!

Co tu robią te wszystkie puste linie?

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

Ups!

Co tu robią te wszystkie puste linie?

- Każda linia w pliku kończy się znakiem końca linii
- Funkcja print() dodaje znak końca linii do każdej linii

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

Przeszukiwanie pliku (poprawione)

- Możemy usunąć białe znaki z prawej strony ciągu, używając rstrip() z biblioteki ciągów
- Znak końca linii jest “białym znakiem” i zostanie usunięty

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

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

Pomijanie za pomocą continue

Możemy wygodnie pominąć linię, używając instrukcji continue

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

Korzystanie z in do wybierania linii

Możemy użyć wyszukiwania ciągu gdziekolwiek w linii jako kryterium wyboru

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

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

```
fname = input('Podaj nazwę pliku: ')
fhand = open(fname)
count = 0
for line in fhand:
    if line.startswith('Subject:'):
        count = count + 1
print('Mamy', count, 'linii z tematem w pliku', fname)
```

Pytanie o nazwę pliku

Podaj nazwę pliku: mbox.txt
Mamy 1797 linii z tematem wiadomości w pliku mbox.txt

Podaj nazwę pliku: mbox-short.txt
Mamy 27 linii z tematem wiadomości w pliku mbox-short.txt

Błędne nazwy pliku

```
fname = input('Podaj nazwę pliku: ')
try:
    fhand = open(fname)
except:
    print('Nie można otworzyć pliku:', fname)
    quit()

count = 0
for line in fhand:
    if line.startswith('Subject:'):
        count = count + 1
print('Mamy', count, 'linii z tematem w pliku', fname)
```

Podaj nazwę pliku: mbox.txt
Mamy 1797 linii z tematem wiadomości w pliku mbox.txt

Podaj nazwę pliku: trele morele
Nie można otworzyć pliku: trele morele

Podsumowanie

- Pamięć pomocnicza
- Wyszukiwanie linii
- Otwieranie pliku – uchwyt pliku
- Czytanie nazw plików
- Struktura pliku – znak końca linii
- Radzenie sobie z błędnymi nazwami plików
- Czytanie pliku linia po linii z użyciem pętli for



Podziękowania dla współpracowników



Copyright slajdów 2010 - Charles R. Severance
(www.dr-chuck.com) University of Michigan School of Information
i open.umich.edu dostępne na licencji Creative Commons
Attribution 4.0. Aby zachować zgodność z wymaganiami licencji
należy pozostawić ten slajd na końcu każdej kopii tego
dokumentu. Po dokonaniu zmian, przy ponownej publikacji tych
materiałów można dodać swoje nazwisko i nazwę organizacji do
listy współpracowników

Autorstwo pierwszej wersji: Charles Severance,
University of Michigan School of Information

Polska wersja powstała z inicjatywy Wydziału Matematyki
i Informatyki Uniwersytetu im. Adama Mickiewicza w Poznaniu

Tłumaczenie: Agata i Krzysztof Wierzbiccy, EnglishT.eu

... wstaw tu nowych współpracowników i tłumaczy