

Ανάγνωση Αρχείων

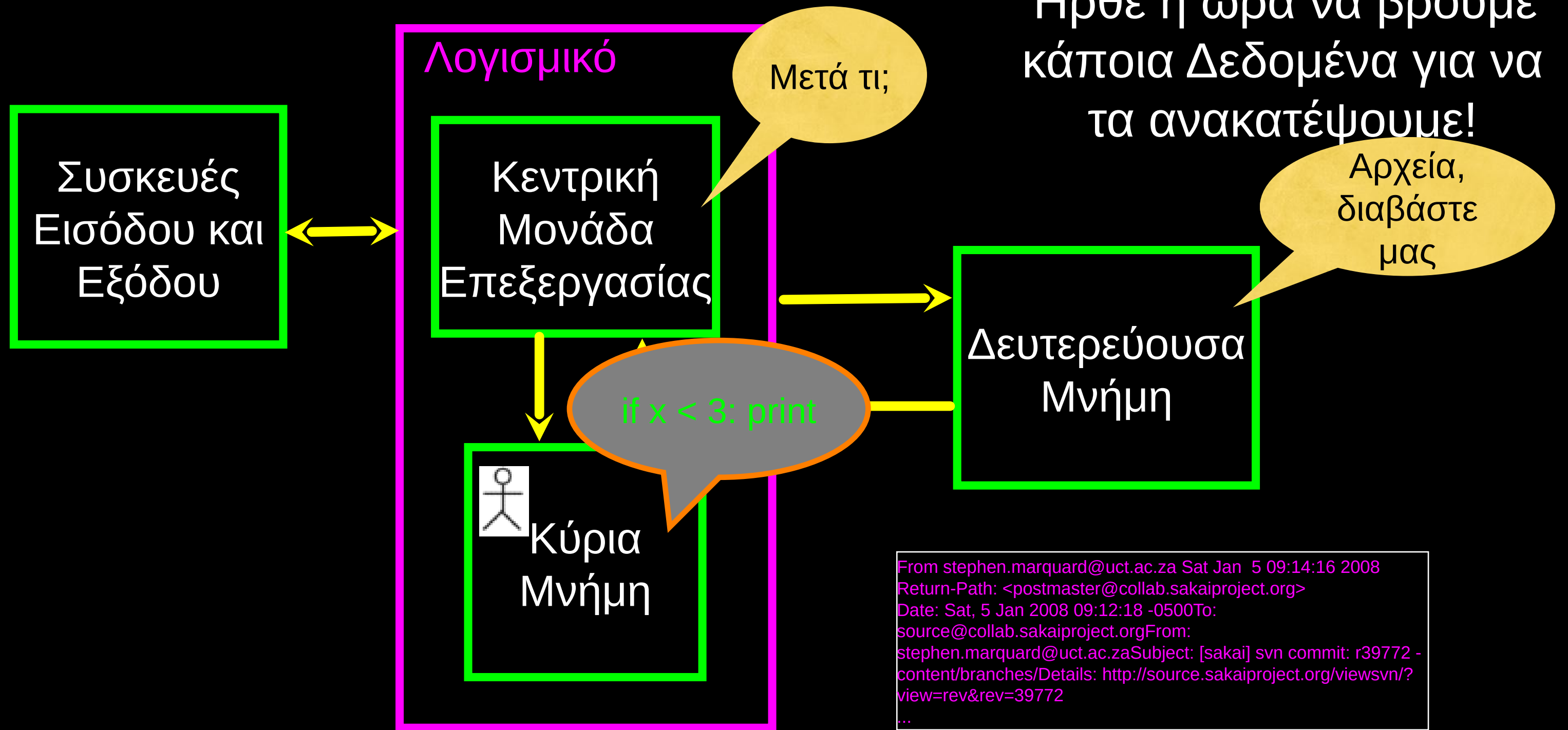
Κεφάλαιο 7



Python για Όλους
www.py4e.com



Έρθε η ώρα να βρούμε
κάποια Δεδομένα για να
τα ανακατέψουμε!



Επεξεργασία Αρχείων

Ένα αρχείο κειμένου μπορεί να θεωρηθεί ως μια ακολουθία γραμμών

```
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>

Άνοιγμα Αρχείου

- Πριν μπορέσουμε να διαβάσουμε τα περιεχόμενα του αρχείου, πρέπει να πούμε στην Python με ποιο αρχείο πρόκειται να εργαστούμε και τι θα κάνουμε με το αρχείο αυτό
- Αυτό γίνεται με τη συνάρτηση `open()`
- Η `open()` επιστρέφει μια «**λαβή αρχείου**» - μια μεταβλητή που χρησιμοποιείται για την εκτέλεση λειτουργιών στο αρχείο
- Παρόμοιο με το «Αρχείο -> Άνοιγμα» ενός Επεξεργαστή Κειμένου

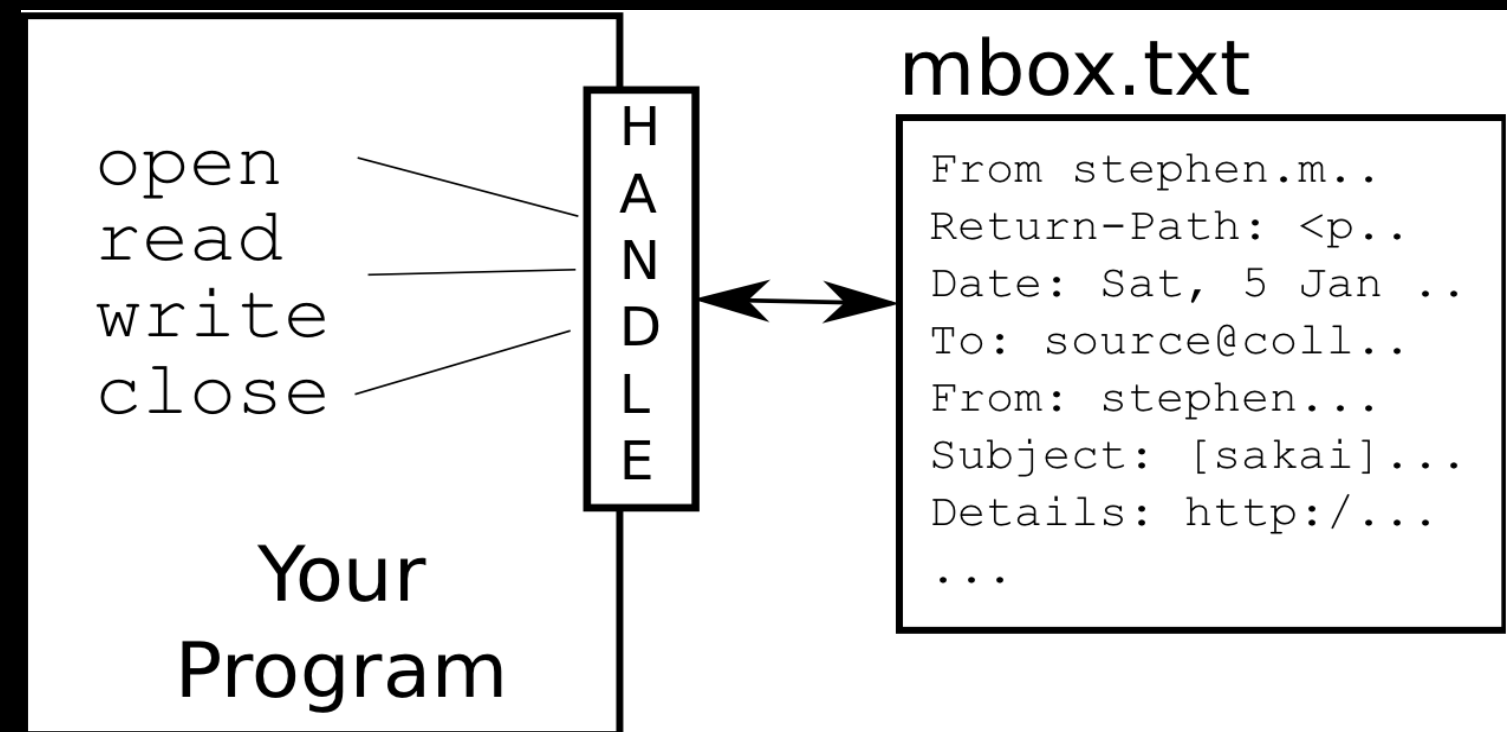
Χρήση της `open()`

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

- `handle = open(όνομα_αρχείου, επιλογή)`
- επιστρέφει μια λαβή που χρησιμοποιείται για να χειριστούμε το αρχείο
- Το `όνομα_αρχείου` είναι μια συμβολοσειρά
- Η επιλογή είναι προαιρετική και πρέπει να είναι «r» εάν σχεδιάζουμε να διαβάσουμε το αρχείο και «w» εάν πρόκειται να γράψουμε στο αρχείο

Τι είναι η Λαβή (Handle);

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



Όταν Λείπουν τα Αρχεία

```
>>> 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'
```

Ο Χαρακτήρας **νέαγραμμή**

- Χρησιμοποιούμε έναν ειδικό χαρακτήρα που ονομάζεται «**νέαγραμμή**» για να υποδείξουμε πότε τελειώνει μια γραμμή
- Το συμβολίζουμε με **\n** σε συμβολοσειρές
- Η **νέαγραμμή** εξακολουθεί να είναι ένας χαρακτήρας - όχι δύο

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

Επεξεργασία Αρχείων

Ένα αρχείο κειμένου μπορεί να θεωρηθεί ως μια ακολουθία γραμμών

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>

Επεξεργασία Αρχείων

Ένα αρχείο κειμένου έχει **νέες γραμμές** στο τέλος κάθε γραμμής

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

Ανάγνωση Αρχείων στην Python

Χειρισμός Αρχείου ως Ακολουθία

- Ένα ανοιχτό **αρχείο** μπορεί να **αντιμετωπιστεί** ως μια **ακολουθία** συμβολοσειρών όπου κάθε γραμμή στο αρχείο είναι μια συμβολοσειρά στην ακολουθία
- Μπορούμε να χρησιμοποιήσουμε την εντολή **for** για να προσπελάσουμε μια **ακολουθία**
- Θυμηθείτε - μια **ακολουθία** είναι ένα διατεταγμένο σύνολο

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

Μέτρηση Γραμμών ενός Αρχείου

- Ανοίξτε ένα **αρχείο** μόνο για ανάγνωση
- Χρησιμοποιήστε μια επανάληψη **for** για να διαβάσετε κάθε γραμμή
- **Μετρήστε** τις γραμμές και εκτυπώστε το πλήθος των γραμμών

```
fhand = open('mbox.txt')
πλήθος = 0
for line in fhand:
    πλήθος = πλήθος + 1
print('Πλήθος Γραμμών:', πλήθος)
```

```
$ python open.py
Πλήθος Γραμμών: 132045
```

Διάβασμα *Ολόκληρου* Αρχείου

Μπορούμε να
διαβάσουμε ολόκληρο
το αρχείο (νέες γραμμές
και όλα) σε μια
συμβολοσειρά

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

Αναζήτηση σε Αρχείο

Μπορούμε να βάλουμε μια δήλωση `if` στον βρόχο `for` για εκτύπωση μόνο των γραμμών που πληρούν ορισμένα κριτήρια

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

ΟΥΠΣ!

Τι κάνουν όλες αυτές οι
κενές γραμμές εδώ;

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

From: `louis@media.berkeley.edu`

From: `zqian@umich.edu`

From: `rjlowe@iupui.edu`

...

ΟΥΠΣ!

Τι κάνουν όλες αυτές οι κενές γραμμές εδώ;

- Κάθε γραμμή του αρχείου έχει έναν χαρακτήρα **νέαγραμμή** στο τέλος
- Η εντολή **print** προσθέτει μια **νέαγραμμή** σε κάθε γραμμή

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

Αναζήτηση σε Αρχείο (ενσωματωμένη)

- Μπορούμε να αφαιρέσουμε τους λευκούς χαρακτήρες από τη δεξιά πλευρά της συμβολοσειράς χρησιμοποιώντας το `rstrip()` από τη βιβλιοθήκη συμβολοσειρών
- Η νέα γραμμή θεωρείται «λευκός χώρος» και **αφαιρείται**

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

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

Παράλειψη με `continue`

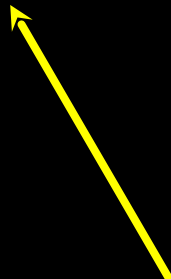
Μπορούμε εύκολα να παραλείψουμε μια γραμμή χρησιμοποιώντας τη δήλωση `continue`

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

Χρήση του `in` για επιλογή Γραμμών

Μπορούμε να αναζητήσουμε μια συμβολοσειρά οπουδήποτε **μέσα** σε μια **γραμμή** θέτοντάς την ως κριτήριο επιλογής μας

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



```
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...
```

Προτροπή για Όνομα Αρχείου

```
ον_αρχείου = input('Δώστε το όνομα του αρχείου: ')\nfhand = open(ον_αρχείου)\nπλήθος = 0\nfor line in fhand:\n    if line.startswith('Subject:') :\n        πλήθος = πλήθος + 1\nprint('Βρέθηκαν', πλήθος, 'γραμμές subject στο', ον_αρχείου)
```

Δώστε το όνομα του αρχείου: mbox.txt
Βρέθηκαν 1797 γραμμές subject στο mbox.txt

Δώστε το όνομα του αρχείου : mbox-short.txt
Βρέθηκαν 27 γραμμές subject στο mbox-short.txt

Άκυρα Ονόματα Αρχείων

```
ον_αρχείου = input('Δώστε το όνομα του αρχείου: ')
try:
    fhand = open(ον_αρχείου)
except:
    print('Το αρχείο δεν μπορεί να ανοίξει:', ον_αρχείου)
    quit()

πλήθος = 0
for γραμμή in fhand:
    if γραμμή.startswith('Subject:') :
        πλήθος = πλήθος + 1
print('Βρέθηκαν', πλήθος, 'γραμμές subject στο', ον_αρχείου)
```

Δώστε το όνομα του αρχείου: **mbox.txt**
Βρέθηκαν 1797 γραμμές subject στο mbox.txt

Δώστε το όνομα του αρχείου: **na na boo boo**
Το αρχείο δεν μπορεί να ανοίξει: na na boo boo

Σύνοψη

- Δευτερεύον αποθηκευτικός χώρος
- Άνοιγμα αρχείου – λαβή αρχείου
- Δομή αρχείου – χαρακτήρας νέαγραμμή
- Διάβασμα αρχείου γραμμή – γραμμή με βρόχο for
- Αναζήτηση γραμμής
- Εισαγωγή ονόματος αρχείου
- Διαχείριση άκυρων ονομάτων αρχείων



Ευχαριστίες / Συνεισφορές



Αυτές οι διαφάνειες είναι Πνευματική ιδιοκτησία 2010- Charles R. Severance (www.dr-chuck.com) του University of Michigan School of Information και είναι διαθέσιμες υπό την άδεια Creative Commons Attribution 4.0. Παρακαλώ να διατηρήσετε αυτήν την τελευταία διαφάνεια σε όλα τα αντίγραφα του εγγράφου για να συμμορφωθείτε με τις απαιτήσεις απόδοσης της άδειας. Εάν κάνετε κάποια αλλαγή, μη διστάσετε να προσθέσετε το όνομα και τον οργανισμό σας στη λίστα των συντελεστών αυτής της σελίδας καθώς αναδημοσιεύετε το υλικό.

Συνέχεια...

Αρχική ανάπτυξη : Charles Severance, University of Michigan School of Information

Απόδοση στα Ελληνικά: Κιουρτίδου Δ. Κωνσταντία

... Εισαγάγετε νέους Μεταφραστές και άτομα που έχουν συνεισφέρει εδώ