



ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

AGRICULTURAL UNIVERSITY OF ATHENS

# Εισαγωγή στην Επιστήμη των Υπολογιστών

## Ενότητα 6:

Αρχεία – Δομές Αρχείων, 2ΔΩ

Τμήμα: Αγροτικής Οικονομίας & Ανάπτυξης

Διδάσκων: Θεόδωρος Τσιλιγκιρίδης



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης





# Μαθησιακοί Στόχοι

Με την ολοκλήρωση της ενότητας ο φοιτητής/τρια θα έχει αποκτήσει γνώσεις για:

- Μέθοδοι Προσπέλασης σε αρχεία - Ορισμοί
- Ακολουθιακά (σειριακά) αρχεία
- Ευρετηριασμένα αρχεία (αρχεία με δείκτες)
- Κατακερματισμένα αρχεία
- Αρχεία κειμένου
- Δυαδικά αρχεία

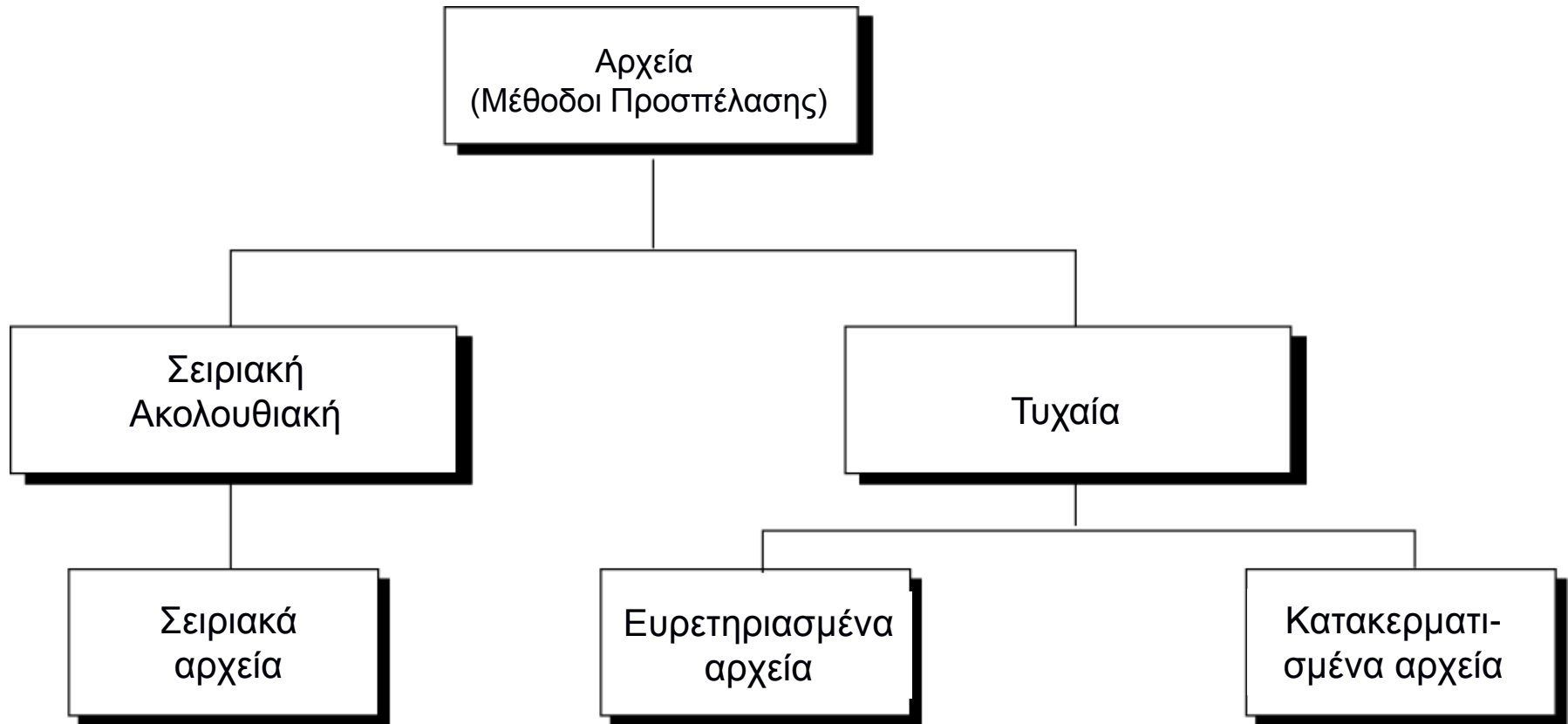


# Λέξεις Κλειδιά

- Ταξινόμηση
- Ακολουθιακά (σειριακά) αρχεία
- Ευρετηριασμένα αρχεία (αρχεία με δείκτες)
- Κατακερματισμένα αρχεία
- Σύγκρουση
- Δυαδικά αρχεία



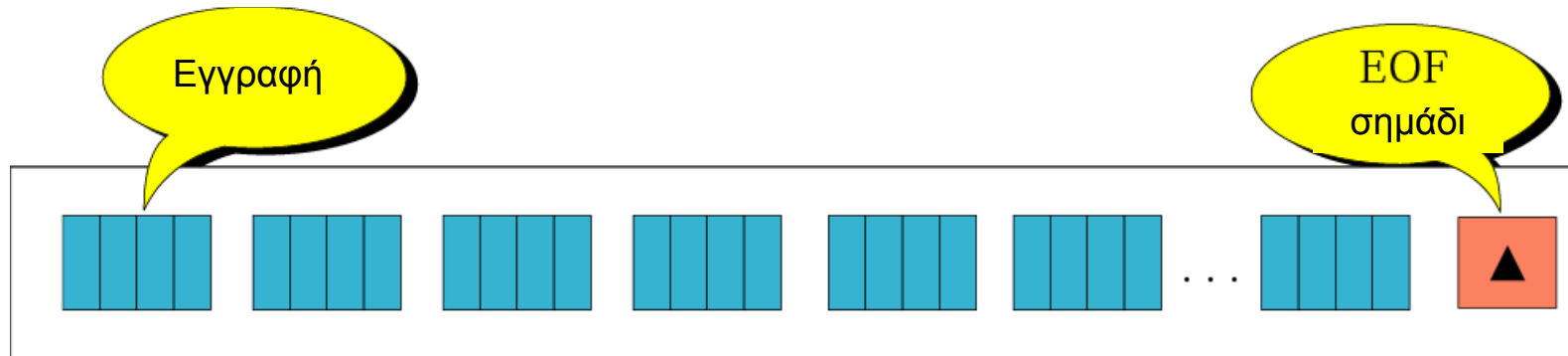
# Ταξινόμηση των Δομών Αρχείων







# Ακολουθιακά (Σειριακά) Αρχεία



**Επεξεργασία εγγραφών σε ακολουθιακό αρχείο**

```
While Not EOF
```

```
{
```

```
    Διάβασε την επόμενη εγγραφή
```

```
    Επεξεργασία της εγγραφής
```

```
}
```



# Ενημέρωση Ακολουθιακού Αρχείου 1/3

Αρχεία του προγράμματος ενημέρωσης

- **Νέο πρωτεύον αρχείο** (new master file)

Πρόκειται για ένα νέο μόνιμο αρχείο που περιέχει τα πλέον πρόσφατα δεδομένα

- **Παλιό πρωτεύον αρχείο** (old master file)

Πρόκειται για το μόνιμο αρχείο που πρέπει να ενημερωθεί. Μετά την ενημέρωση φυλάσσεται για αναφορά.



# Ενημέρωση Ακολουθιακού Αρχείου 2/3

- **Αρχείο κινήσεων** (transaction file)

Πρόκειται για το αρχείο που περιέχει τις αλλαγές που πρέπει να γίνουν στο παλιό πρωτεύον αρχείο. Οι αλλαγές μπορεί να είναι:

- (Π): Προσθήκη κινήσεων (Προσδιορίζει τις νέες εγγραφές που πρέπει να προστεθούν στο παλαιό πρωτεύον αρχείο).
- (Δ): Διαγραφή κινήσεων (Προσδιορίζει τις εγγραφές που πρέπει να διαγραφούν από το παλαιό πρωτεύον αρχείο).



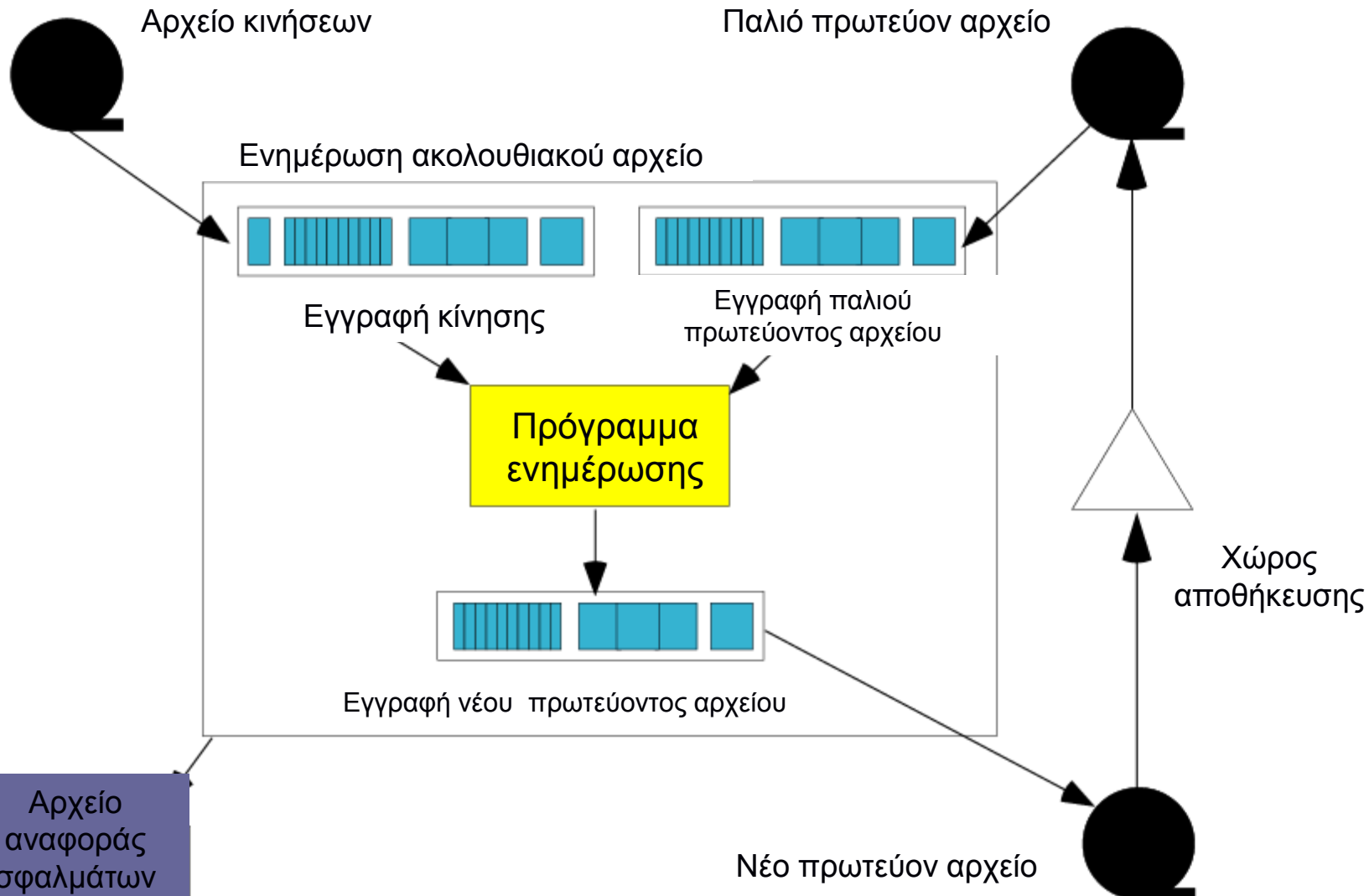


# Ενημέρωση Ακολουθιακού Αρχείου 3/3

- (M): Μεταβολή κινήσεων (Προσδιορίζει τις εγγραφές του παλαιού πρωτεύοντος αρχείου που πρέπει να διορθωθούν).
- **Αρχείο αναφοράς σφαλμάτων (error report file)**  
Πρόκειται για το αρχείο που περιέχει τη λίστα των σφαλμάτων που παρουσιάζονται κατά τη διαδικασία ενημέρωσης.



# Ενημέρωση Ακολουθιακού Αρχείου 3/3





# Ενημέρωση Ακολουθιακού Αρχείου 3/3

Πρόγραμμα ενημέρωσης:

Η ενημέρωση ενός ακολουθιακού αρχείου γίνεται περιοδικά, μέσω του προγράμματος ενημέρωσης

Η διαδικασία της ενημέρωσης περιλαμβάνει τα ακόλουθα βήματα:

1. Όλα τα αρχεία ταξινομούνται με το ίδιο πεδίο-κλειδί.
2. Αν το κλειδί του αρχείου κινήσεων είναι 'μικρότερο' από το κλειδί του παλαιού πρωτεύοντος αρχείου, η κίνηση προστίθεται στο νέο πρωτεύον αρχείο.



# Ενημέρωση Ακολουθιακού Αρχείου 3/3

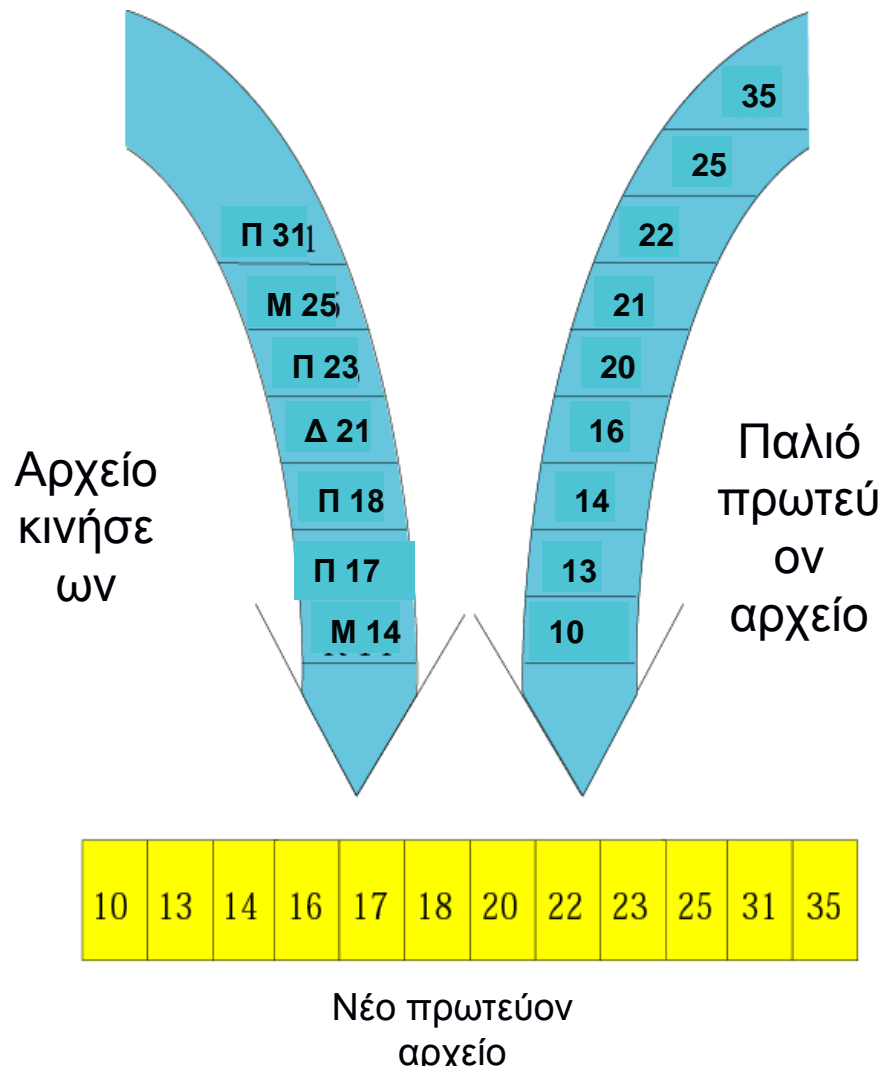
3. Αν το κλειδί του αρχείου κινήσεων είναι 'ίσο' με το κλειδί του παλαιού πρωτεύοντος αρχείου, τότε:
  - Σε περίπτωση που η κίνηση είναι μεταβολή (Μ), αλλάζουν τα περιεχόμενα των δεδομένων στο πρωτεύον αρχείο.
  - Σε περίπτωση που η κίνηση είναι διαγραφή (Δ), διαγράφονται τα περιεχόμενα των δεδομένων από το πρωτεύον αρχείο.
4. Αν το κλειδί του αρχείου κινήσεων είναι 'μεγαλύτερο' από το κλειδί του παλαιού πρωτεύοντος αρχείου, η εγγραφή του παλαιού πρωτεύοντος αρχείου γράφεται στο νέο πρωτεύον αρχείο. Σε αυτή την περίπτωση ο κωδικός κίνησης πρέπει να είναι προσθήκη (Π)



# Ενημέρωση Ακολουθιακού Αρχείου 3/3

- Διαδικασία ενημέρωσης
  - Π: Προσθήκη
  - Δ: Διαγραφή
  - Μ: Μεταβολή

**Η διαδικασία αρχίζει με τη σύγκριση των κλειδιών των πρώτων εγγραφών**





# Διαδικασία Συγχώνευσης Δύο Ακολουθιακών Αρχείων

## Αλγόριθμος της διαδικασίας συγχώνευσης

Αρχεία εισόδου: InputFileA, InputFileB

Αρχείο εξόδου: OutputFile

Procedure MergeFiles (InputFileA, InputFileB, OutputFile)

If (και τα δύο αρχεία εισόδου είναι στο EOF)

    then (Τέλος, με το OutputFile κενό)

If (το InputFileA δεν είναι στο EOF)

    then (Δήλωσε την πρώτη του εγγραφή σαν την τρέχουσα εγγραφή)

If (το InputFileB δεν είναι στο EOF)

    then (Δήλωσε την πρώτη του εγγραφή σαν την τρέχουσα εγγραφή)

While (κανένα από τα αρχεία εισόδου δεν είναι EOF) do

    {Βάλε την τρέχουσα εγγραφή με τη 'μικρότερη' τιμή στο πεδίο κλειδιού στο OutputFile

If (αυτή η τρέχουσα εγγραφή είναι η τελευταία εγγραφή του αντίστοιχου αρχείου εισόδου της)

    then (δήλωσε ότι το αρχείο εισόδου βρίσκεται στο EOF)

    else (δήλωσε ως τρέχουσα εγγραφή του συγκεκριμένου αρχείου εισόδου την επόμενη εγγραφή σε αυτό)

}

Ξεκινώντας από την τρέχουσα εγγραφή του αρχείου εισόδου που δεν είναι στο EOF αντέγραψε τις υπόλοιπες εγγραφές στο OutputFile

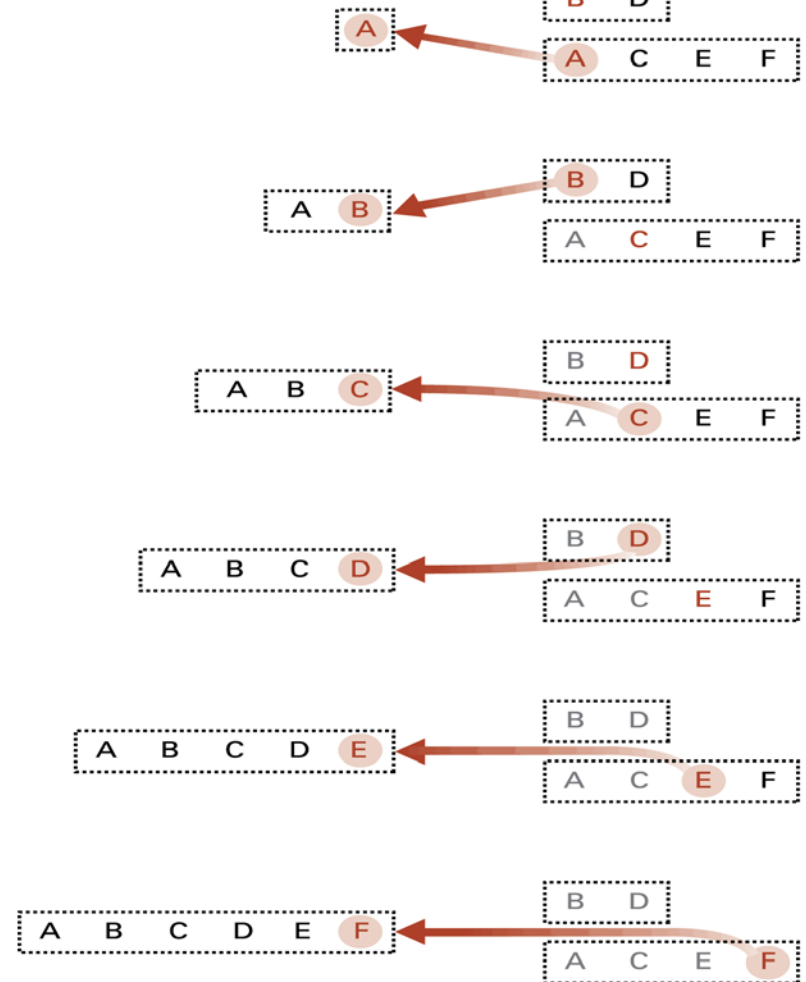


# Παράδειγμα: Εφαρμογή του Αλγόριθμου Συγχώνευσης Δύο Ακολουθιακών Αρχείων

- Τα γράμματα χρησιμοποιούνται για την αναπαράσταση ολόκληρων εγγραφών.  
Το συγκεκριμένο γράμμα δείχνει την τιμή του πεδίου-κλειδιού της εγγραφής..

Αρχείο Εξόδου

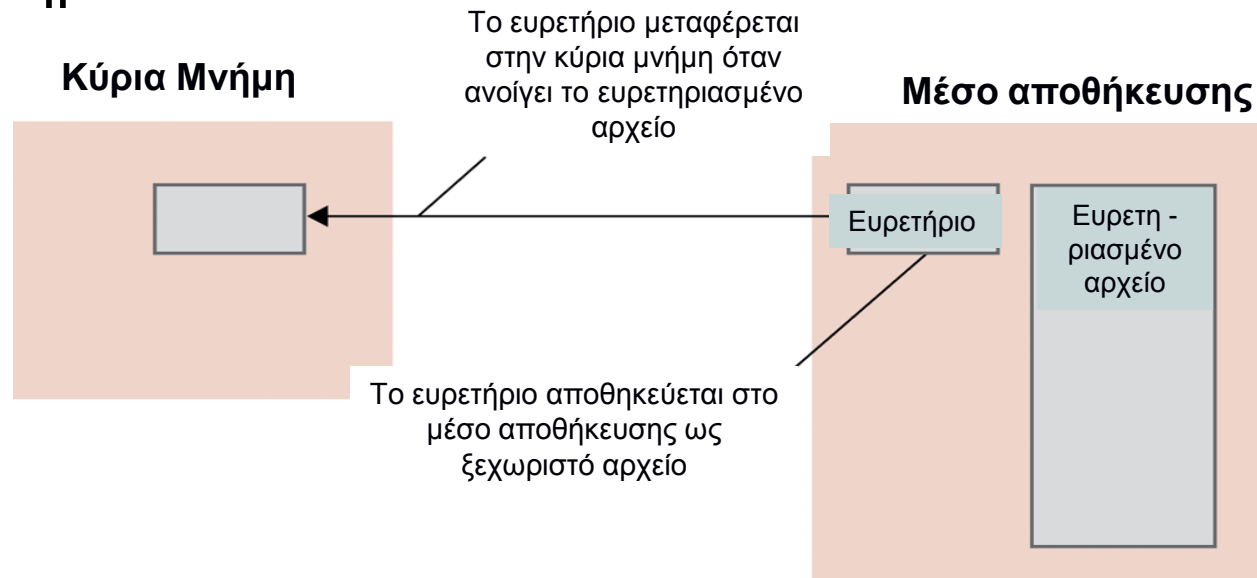
Αρχείο Εισόδου





# Ευρετηριασμένα Αρχεία (Αρχεία με Δείκτες) 1/3

- Κάθε ευρετηριασμένο αρχείο αποτελείται από
  - Ένα αρχείο δεδομένων (ακολουθιακό αρχείο) και
  - Ένα ευρετήριο (αρχείο με δύο πεδία - το κλειδί του ακολουθιακού αρχείου και τη διεύθυνση της αντίστοιχης εγγραφής).
- Η αντιστοίχιση των κλειδιών στις διευθύνσεις γίνεται μέσω του ευρετηρίου

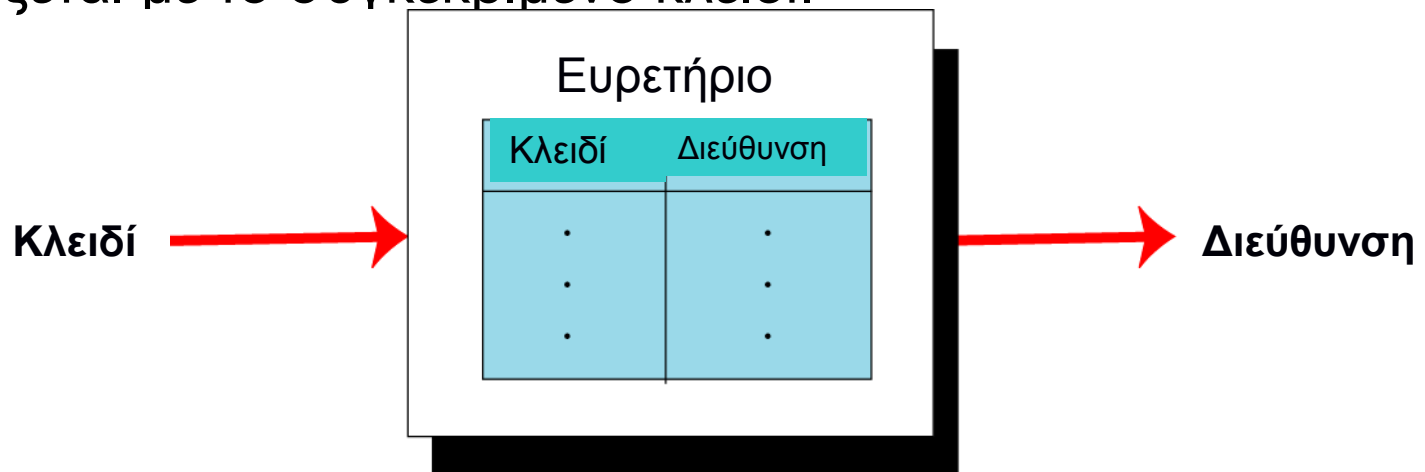






# Ευρετηριασμένα Αρχεία (Αρχεία με Δείκτες) 2/3

- **Ευρετήριο:** Λίστα με όλες τις τιμές κλειδιά που είναι αποθηκευμένα σε αυτό, μαζί με τις καταχωρίσεις που προσδιορίζουν τις θέσεις (διευθύνσεις) όπου είναι αποθηκευμένη η εγγραφή που περιέχει το κάθε κλειδί. Για να βρεθεί μια συγκεκριμένη εγγραφή βρίσκουμε το προσδιοριστικό κλειδί στο ευρετήριο και μετά ανακτούμε το τμήμα πληροφοριών που είναι αποθηκευμένο στη θέση (διεύθυνση) η οποία σχετίζεται με το συγκεκριμένο κλειδί.



**Αντιστοίχιση σε ένα αρχείο με δείκτες**



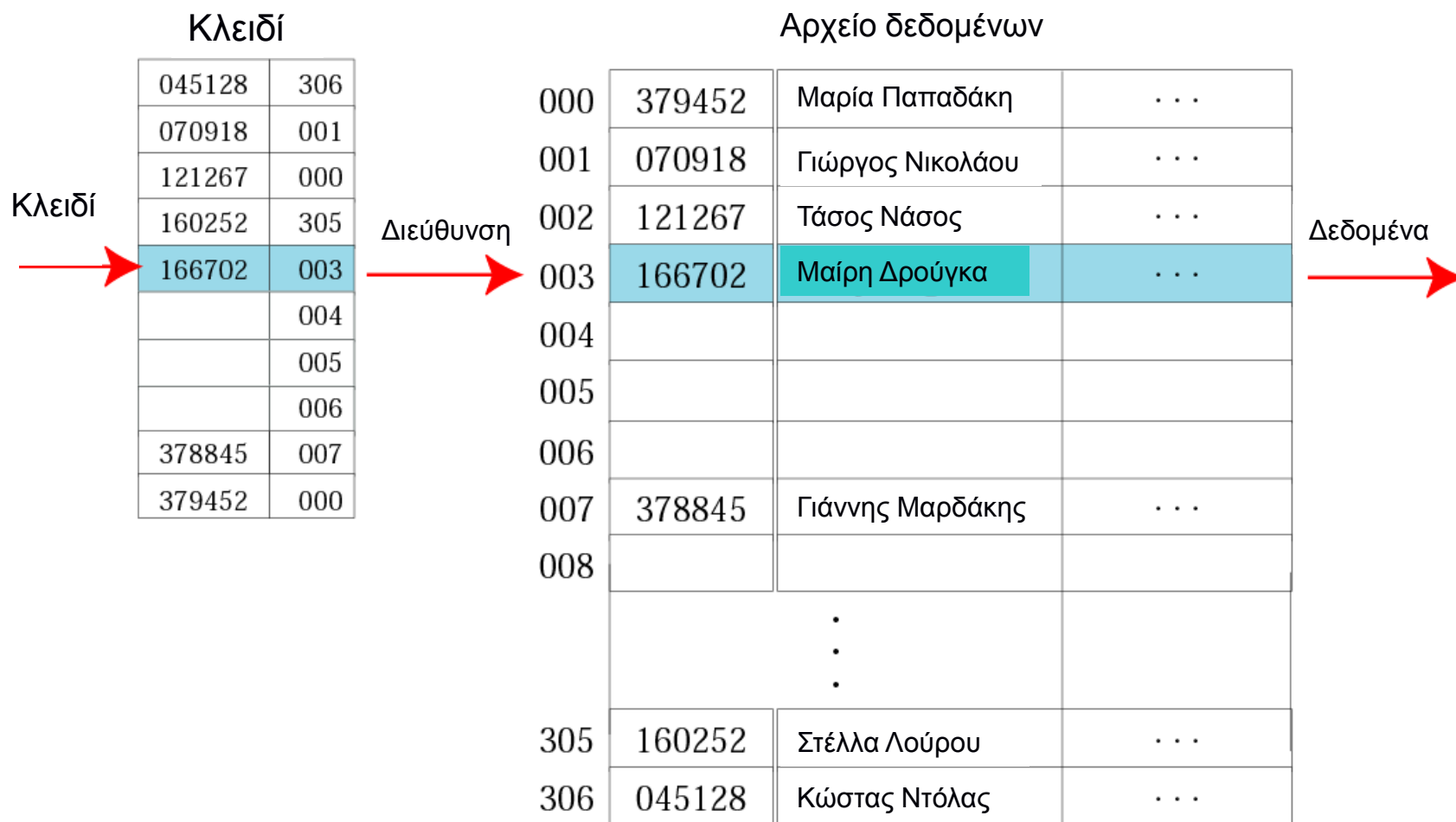
# Ευρετηριασμένα Αρχεία (Αρχεία με Δείκτες) 3/3

- Αλγόριθμος προσπέλασης εγγραφής σε ευρετηριακό αρχείο
  1. Το αρχείο του ευρετηρίου φορτώνεται στην κύρια μνήμη
  2. Ερευνώνται οι καταχωρίσεις με κάποιον αποδοτικό αλγόριθμο αναζήτησης, όπως η δυαδική αναζήτηση, για να βρεθεί το επιθυμητό πεδίο.
  3. Ανακτάται η διεύθυνση της εγγραφής
  4. Χρησιμοποιείται η διεύθυνση για να ανακτηθεί η εγγραφή δεδομένων και να μεταβιβαστεί στο χρήστη.
- Αντεστραμμένα αρχεία:

Τύπος ευρετηριασμένου αρχείου που επιτρέπει την παρουσία πολλών ευρετηρίων, το καθένα από τα οποία έχει διαφορετικό κλειδί.



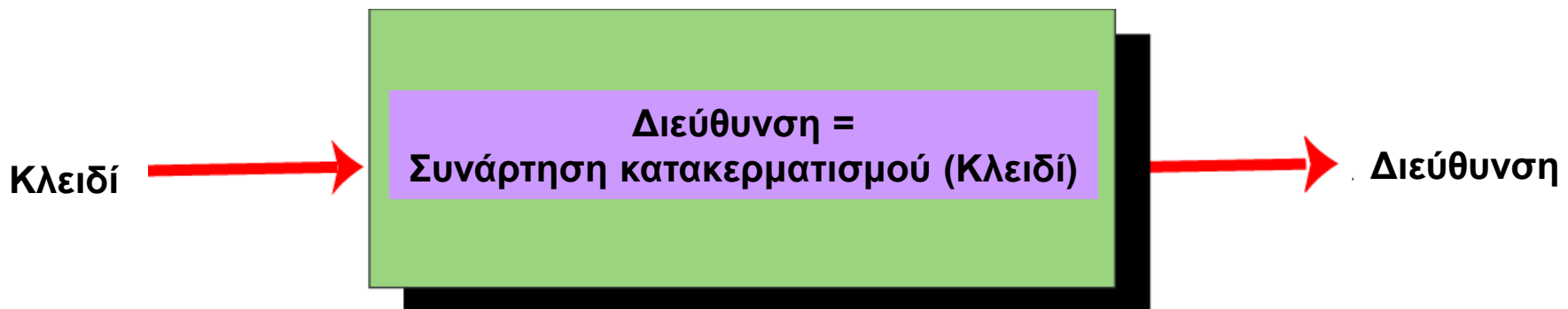
# Λογική Αναπαράσταση Ευρητηριασμένου Αρχείου





# Κατακερματισμένα Αρχεία

- Κατακερματισμός (hashing) είναι μία τεχνική που εντοπίζει μια εγγραφή ενός αρχείου απευθείας από το κλειδί της αντιστοιχίζοντας το με τη διεύθυνση της εγγραφής μέσω κάποιας συνάρτησης.
- Το αρχείο που εφαρμόζει τη μέθοδο κατακερματισμού λέγεται **κατακερματισμένο αρχείο**.
- Η συνάρτηση που καθορίζει τη διεύθυνση της εγγραφής από το κλειδί της λέγεται **συνάρτηση κατακερματισμού**.
- Δεν απαιτείται ευρετήριο – με αποτέλεσμα η μέθοδος να είναι πιο αποτελεσματική από αυτήν που χρησιμοποιεί ευρετήριο αντί για συνάρτηση.

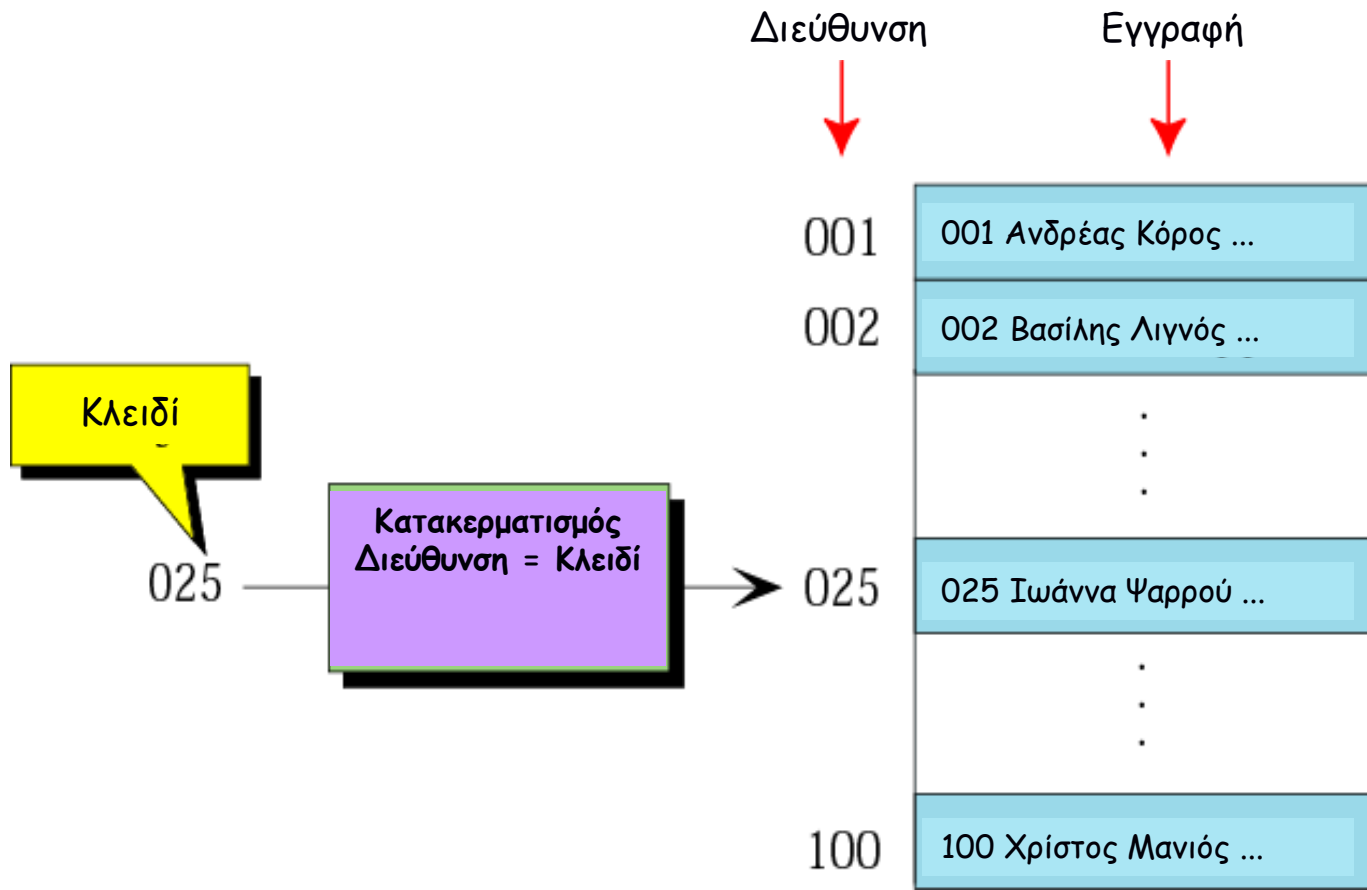


Αντιστοίχιση σε ένα κατακερματισμένο αρχείο



# Μέθοδος Άμεσου Κατακερματισμού 1/2

- Άμεση μέθοδος κατακερματισμού: Το κλειδί αποτελεί και τη διεύθυνση.





# Μέθοδος Άμεσου Κατακερματισμού 2/2

- **Παράδειγμα:** Εφαρμογή της άμεσης μεθόδου σε κωδικούς υπαλλήλων ( $\leq 100$ ) κάποιας εταιρείας.
- Κάθε εγγραφή υπαλλήλου έχει κλειδί τον κωδικό του υπαλλήλου.
- Η εγγραφή κάθε υπαλλήλου έχει διεύθυνση το κλειδί της εγγραφής, δηλαδή τον κωδικό του υπαλλήλου

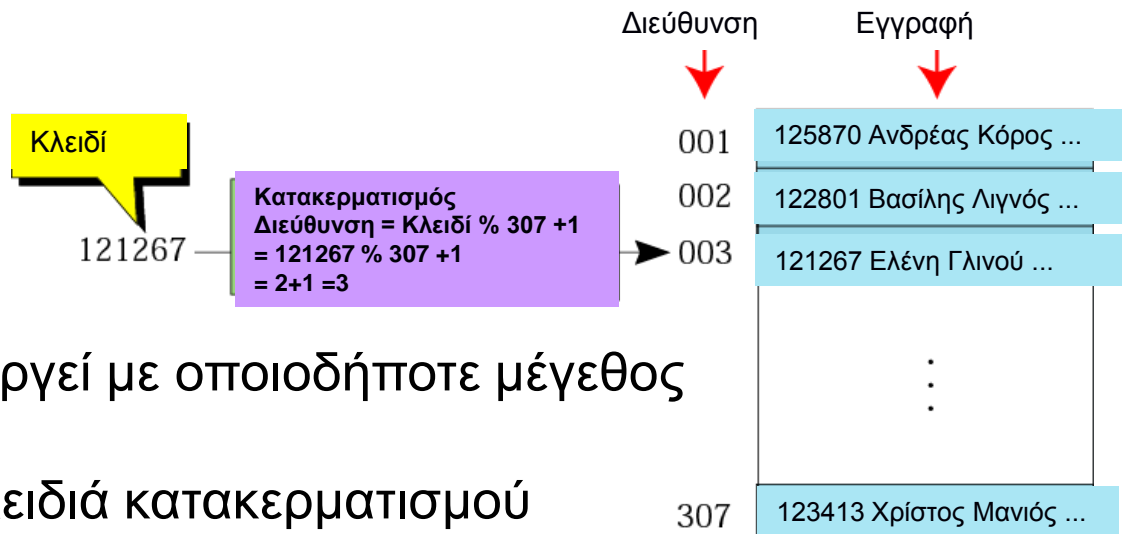


# Μέθοδος Κατακερματισμού Υπολοίπου Διαίρεσης 1/2

- Μέθοδος κατακερματισμού υπολοίπου διαίρεσης (διαίρεσης Modulo): Η διεύθυνση της εγγραφής προσδιορίζεται από το υπόλοιπο της διαίρεσης του κλειδιού της εγγραφής με το μέγεθος της λίστας των εγγραφών του αρχείου (πλήθος εγγραφών) συν ένα.
- Συνάρτηση: Διεύθυνση = Κλειδί % μέγεθος λίστας % +1



# Μέθοδος Κατακερματισμού Υπολοίπου Διαίρεσης 2/2



- Η συνάρτηση λειτουργεί με οποιοδήποτε μέγεθος λίστας
- Πρόβλημα: Δύο κλειδιά κατακερματισμού υπολοίπου διαίρεσης οδηγούν στο ίδιο υπόλοιπο και επομένως στην ίδια διεύθυνση.
- Αντιμετώπιση: Επιλέγουμε, όπου αυτό είναι δυνατόν, το μέγεθος λίστας να είναι πρώτος αριθμός.
- Παράδειγμα: Αν θέλουμε να διαθέσουμε χώρο για 300 εγγραφές (υπαλλήλους) επιλέγουμε μέγεθος λίστας 307 εγγραφές – ο αμέσως μεγαλύτερος του 300 πρώτος αριθμός.

$$\begin{array}{r} 395 \\ \hline 307(121267 \\ 121265 \\ \hline 2 \end{array}$$





# Άλλες Μέθοδοι Κατακερματισμού

- Μέθοδος εξαγωγής ψηφίων: Τα επιλεγμένα ψηφία εξάγονται από το κλειδί και χρησιμοποιούνται ως διεύθυνση της εγγραφής.
- Μέθοδος μέσου τετραγώνου
- Μέθοδος αναδίπλωσης
- Περιστροφική μέθοδος
- Ψευδοτυχαία μέθοδος
- (Άσκηση: Να γραφεί ο αλγόριθμος κάθε μία από τις ανωτέρω μεθόδους).



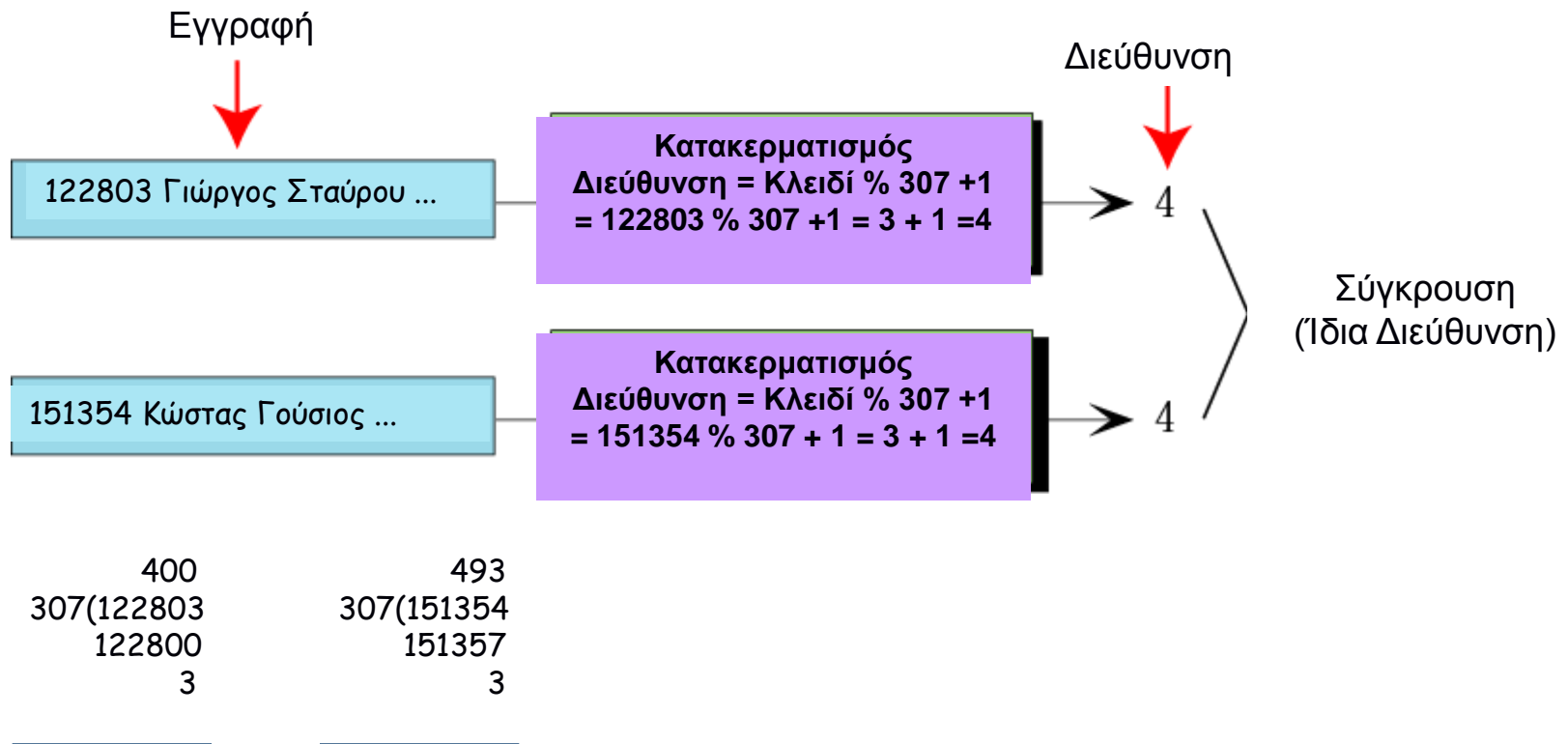
# Σύγκρουση 1/2

Σύγκρουση έχουμε όταν η μέθοδος κατακερματισμού παράγει μία διεύθυνση η οποία είναι ήδη κατειλημμένη.

- Τα κλειδιά που οδηγούν στην ίδια διεύθυνση λέγονται συνώνυμα.
- Η διεύθυνση που παράγεται από τον αλγόριθμο κατακερματισμού λέγεται οικεία διεύθυνση.
- Το τμήμα του αρχείου που περιέχει όλες τις οικείες διευθύνσεις λέγεται κύρια περιοχή.
- Επίλυση: Όταν δύο κλειδιά συγκρούονται σε μία οικεία διεύθυνση η σύγκρουση πρέπει να επιλυθεί με την τοποθέτηση της μιας εγγραφής (κλειδί και των δεδομένων της εγγραφής που αντιστοιχεί) σε άλλη θέση.



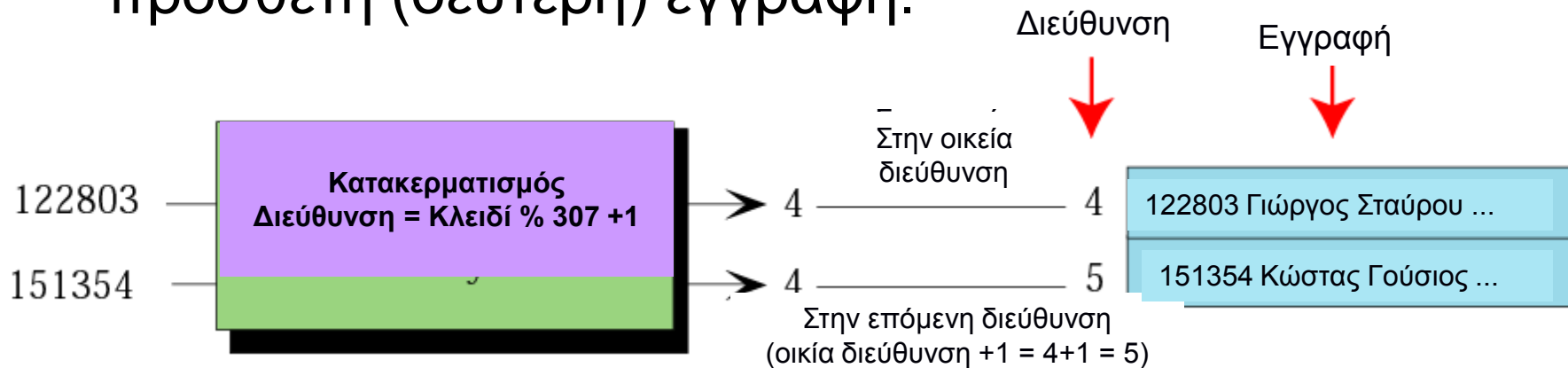
# Σύγκρουση 2/2





# Σύγκρουση: Επίλυση Ανοιχτής Διευθυνσιοδότησης

- Επίλυση ανοιχτής διευθυνσιοδότησης: Όταν συμβαίνει μια σύγκρουση οι διευθύνσεις της κύριας περιοχής αναζητούν μία μη κατειλημμένη (ανοιχτή) διεύθυνση για να τη διαθέσουν στην πρόσθετη (δεύτερη) εγγραφή.



Συνάρτηση Κατακερματισμού: Διεύθυνση = Κλειδί % 307 + 1

Οικεία Διεύθυνση:  $122803 \% 307 + 1 = 3 + 1 = 4$

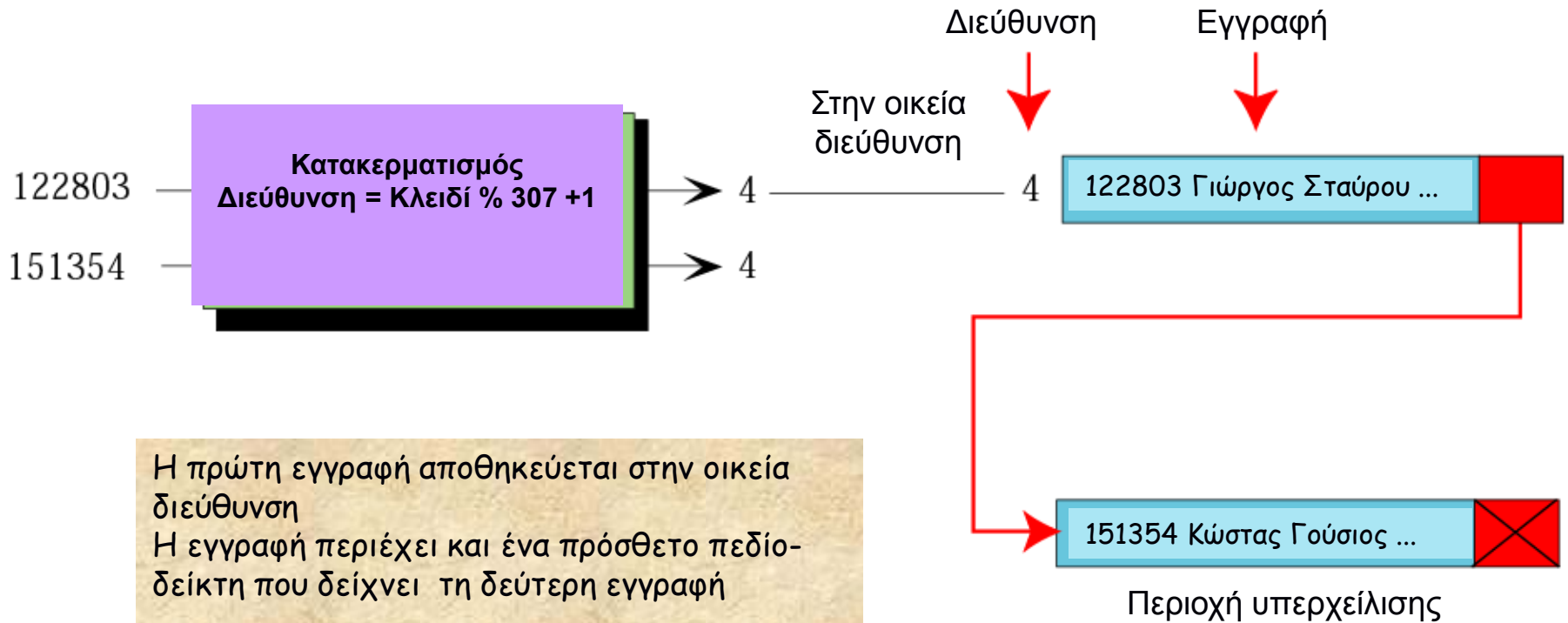
$151354 \% 307 + 1 = 3 + 1 = 4$

Επόμενη Διεύθυνση: Οικεία διεύθυνση + 1 = 4 + 1 = 5

Δεν επιλύεται πλήρως το πρόβλημα αφού κάθε σύγκρουση αυξάνει την πιθανότητα μελλοντικών συγκρούσεων



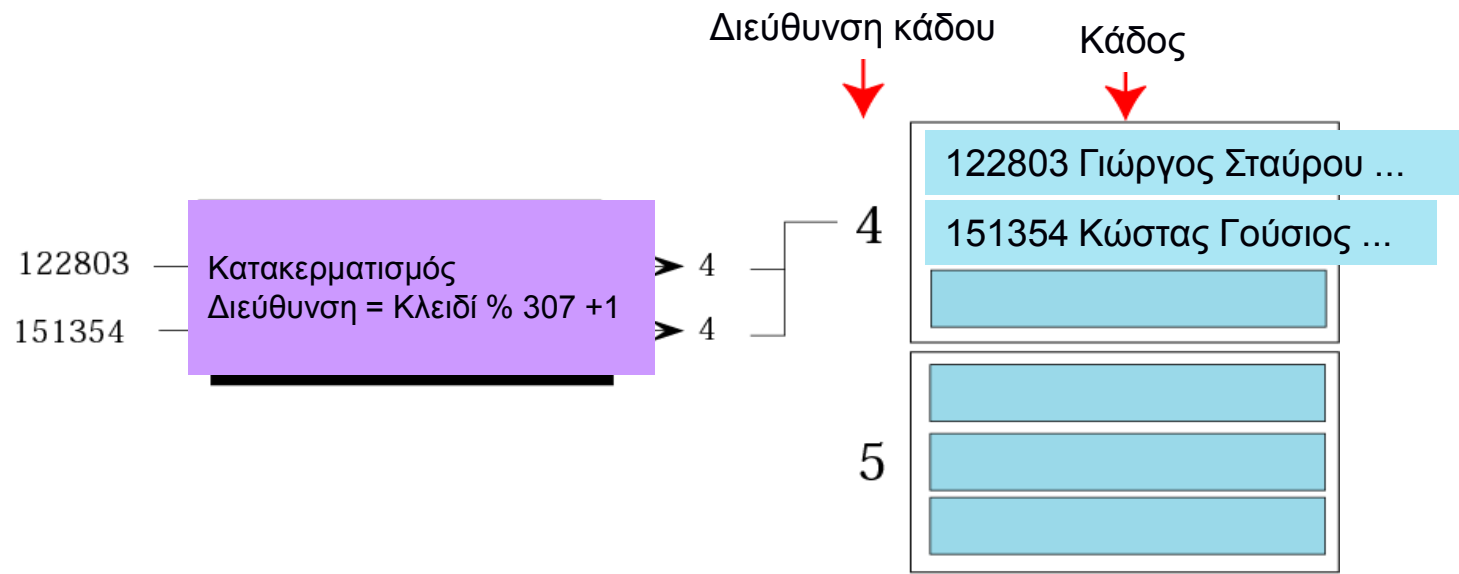
# Σύγκρουση: Επίλυση Συνδεδεμένης Λίστας





# Σύγκρουση: Επίλυση Κατακερματισμού Κάδου 1/3

- Ο χώρος αποθήκευσης διαιρείται τμήματα που ονομάζονται **κάδοι**.
- Κάθε κάδος είναι ένας κόμβος που μπορεί να συμπεριλάβει πολλές εγγραφές ταυτόχρονα.
- Κάθε εγγραφή έχει ένα πεδίο-κλειδί.
- Μία **συνάρτηση κατακερματισμού** υπολογίζει τον αριθμό του κάδου για κάθε κλειδί (τιμή του πεδίου-κλειδί) για κάθε εγγραφή.
- Κάθε εγγραφή αποθηκεύεται στον κάδο που αντιστοιχεί στον κατακερματισμό του κλειδιού της.
- Όταν τα δύο κλειδιά κατακερματισμού οδηγούν στην ίδια διεύθυνση κάδου έχουμε σύγκρουση.
  - Πρόκειται για μεγάλο πρόβλημα όταν ο πίνακας είναι γεμάτος πάνω από 75%
  - Λύση: Αύξηση του αριθμού των κάδων και επανακατακερματισμό όλων των δεδομένων.





# Σύγκριση: Επίλυση Κατακερματισμού Κάδου 2/3

**Παράδειγμα:** Κατακερματισμός της τιμής του πεδίου-κλειδιού 25X3Z σε ένα από τους 41 κάδους

Τιμή Πεδίου-Κλειδιού (κλειδί):

25X3Z

Αναπαράσταση ASCII II:

0011001000110101010110000011001101011010

Ισοδύναμη δεκαδική τιμή :

215,643,337,562

Υπόλοιπο μετά τη διαίρεση με 41:

3

Αριθμός κάδου:

3



# Σύγκριση: Επίλυση Κατακερματισμού Κάδου 3/3

## ● Κατακερματισμός κάδου

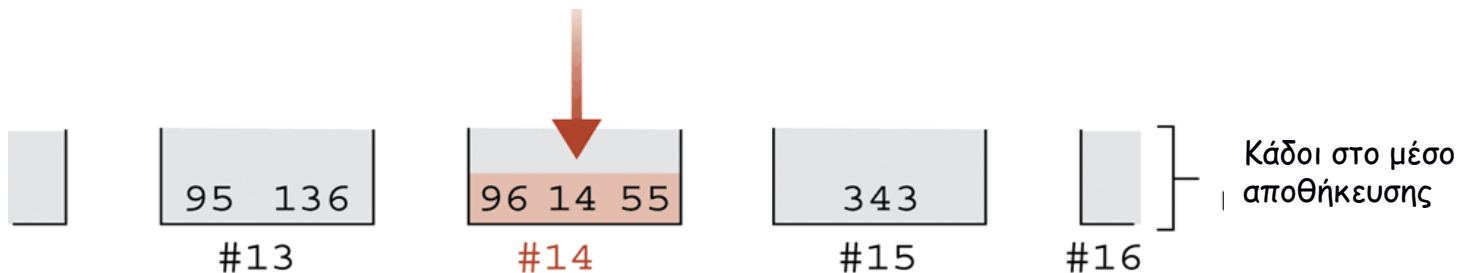
$$\begin{array}{r} 1 \\ 41 \overline{)55} \\ \underline{41} \\ 14 \end{array}$$

$$\begin{array}{r} 2 \\ 41 \overline{)96} \\ \underline{82} \\ 14 \end{array}$$

$$\begin{array}{r} 0 \\ 41 \overline{)14} \\ \underline{0} \\ 14 \end{array}$$

Υπόλοιπα

Κατά τη διαίρεση με το 41 οι τιμές πεδίου κλειδιού 15, 55 και 96 παράγουν ένα υπόλοιπο ίσο με 14. Έτσι αυτές οι εγγραφές αποθηκεύονται στον κάδο 41.







# Ερμηνεία Αρχείου Κειμένου και Δυαδικού Αρχείου

- Τα αρχεία κειμένου είναι αρχεία χαρακτήρων που δεν περιέχουν ακεραίους, αριθμούς κινητής υποδιαστολής, ή άλλες δομές δεδομένων που έχουν αποθηκευτεί στην εσωτερική μορφή του υπολογιστή.
- Τα δυαδικά αρχεία είναι μία συλλογή δεδομένων που έχουν αποθηκευτεί στην εσωτερική μορφή του υπολογιστή (π.χ. ακέραιοι, αριθμοί κινητής υποδιαστολής, χαρακτήρες, ή οποιαδήποτε άλλη μορφή δομημένων πληροφοριών).

Ερμηνεύεται ως αρχείο κειμένου

01000001 01000010



Δύο bytes αντιπροσωπεύουν δύο χαρακτήρες (το A και το B)

Ερμηνεύεται ως δυαδικό αρχείο

01000001 01000010



Δύο bytes αντιπροσωπεύουν έναν αριθμό (το 16706)



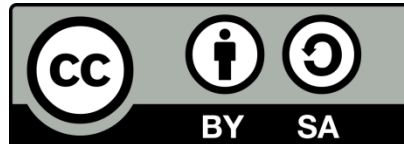
# Βιβλιογραφία

- Α. Σιδερίδης, *Εισαγωγή στην Επιστήμη των Υπολογιστών*, Αθήνα.
- J. Glenn Brookshearr, *Η επιστήμη των Υπολογιστών - Μια Ολοκληρωμένη Παρουσίαση*, Κλειδάριθμος.
- L. Goldschlager, A. M. Lister, *Εισαγωγή στη Σύγχρονη Επιστήμη των Υπολογιστών*, Δίαυλος
- Behrouz Forouzan, Firouz Mosharraf, *Εισαγωγή στην Επιστήμη των Υπολογιστών - 2η Αγγλική Έκδοση - Επιμέλεια: Γιώργος Στεφανίδης, Αλέξανδρος Χατζηγεωργίου* . Κλειδάριθμος.



# Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδεια χρήσης, η άδεια χρήσης αναφέρεται ρητώς.





# Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα Γεωπονικού Πανεπιστημίου Αθηνών**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.





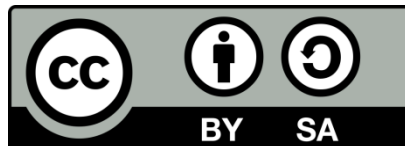
# Σημείωμα Αναφοράς

- Copyright Γεωπονικό Πανεπιστήμιο Αθηνών. Τμήμα Αγροτικής Οικονομίας και Ανάπτυξης, Θεόδωρος Τσιλιγκιρίδης, «Εισαγωγή στην Επιστήμη των Υπολογιστών». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση:  
<https://oceclass.aua.gr/courses/OCDAERD111/>



# Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων, π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



Η άδεια αυτή ανήκει στις άδειες που ακολουθούν τις προδιαγραφές του Ορισμού Ανοικτής Γνώσης [2], είναι ανοικτό πολιτιστικό έργο [3] και για το λόγο αυτό αποτελεί ανοικτό περιεχόμενο [4].

[1] <http://creativecommons.org/licenses/by-sa/4.0/>

[2] <http://opendefinition.org/okd/ellinika/>

[3] <http://freedomdefined.org/Definition/EI>

[4] <http://opendefinition.org/buttons/>



# Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων
- το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)

μαζί με τους συνοδευόμενους υπερσυνδέσμους.