



Πανεπιστήμιο Αιγαίου  
Τμήμα Γεωγραφίας

# Βάσεις Γεωγραφικών Δεδομένων

Ενότητα 4: Αποθήκευση εγγραφών – Ευρετήρια

Μιχάλης Βαΐτης



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



# Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Αιγαίου**» έχει χρηματοδοτήσει μόνο τη 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.



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο

2



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ  
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

# ΟΡΙΣΜΟΙ

**Block**: ομάδα από bytes, σταθερού μήκους ανά λειτουργικό σύστημα και αποθηκευτικό μέσο

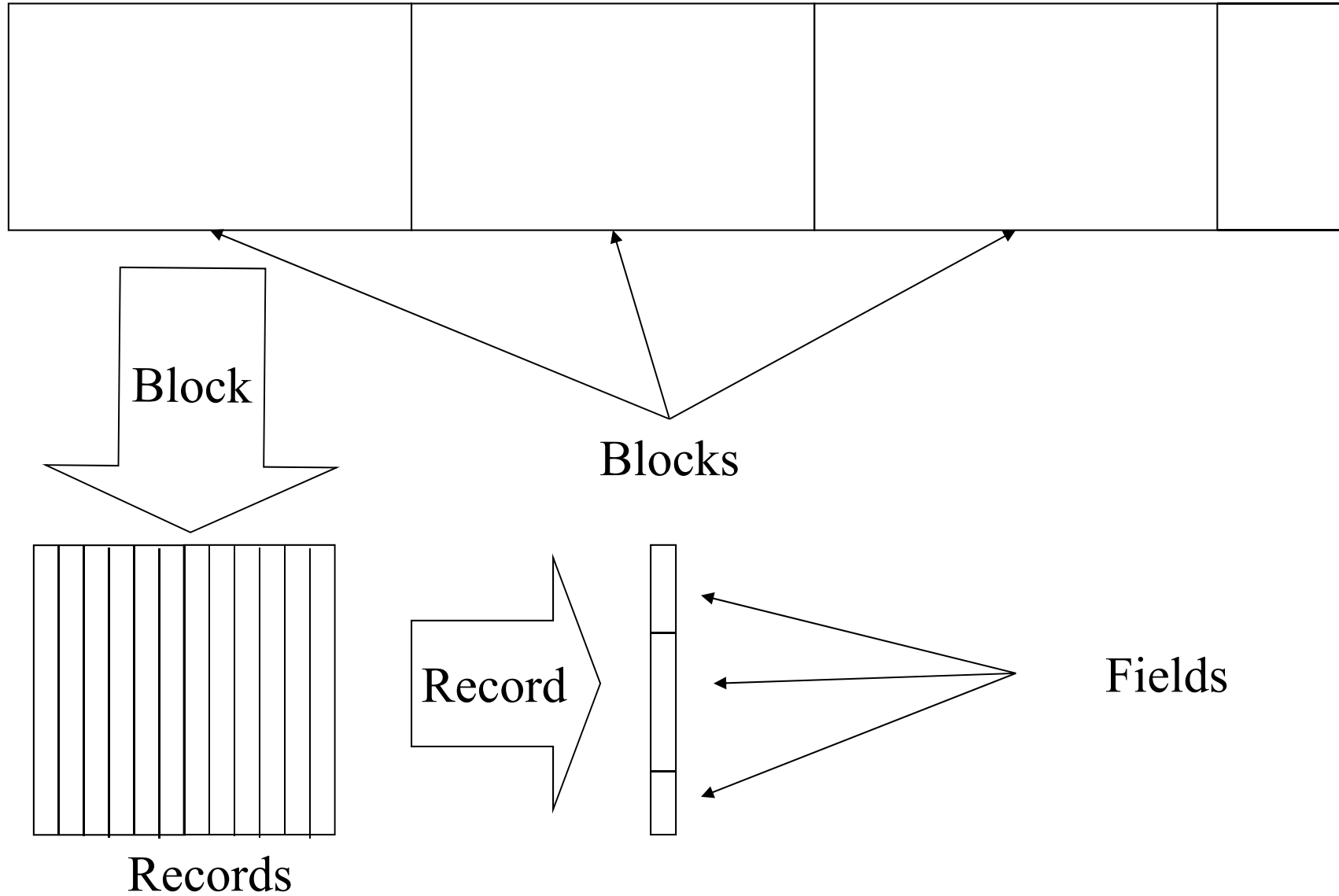
**Αρχείο (file)**: Σύνολο από blocks (ακέραιος αριθμός)

**Πεδίο (field)**: σύνολο από bytes που αντιστοιχεί σε ένα *τύπο δεδομένων (data type)*

**Τύπος εγγραφής (record type)**: διατεταγμένο σύνολο πεδίων

**Αρχείο εγγραφών (record file)**: σύνολο από εγγραφές.  
[Κάθε block του αρχείου περιλαμβάνει ακέραιο αριθμό εγγραφών]

# Αποθηκευτικό μέσο



## Αρχείο εγγραφών:

(ΑΜ ΕΠΩΝΥΜΟ ΟΝΟΜΑ ΕΤΟΣ-ΓΕΝ.)

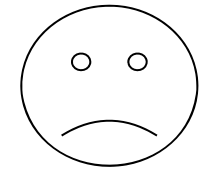
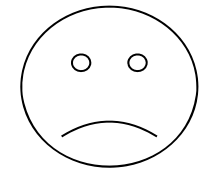
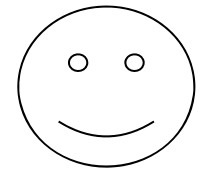
1	Παπαδόπουλος	Κων/νος	1970
3	Νικολάου	Γεώργιος	1971
6	Τζίμα	Μαρία	1969
2	Δημοπούλου	Ελένη	1972
5	Ανδρέου	Δήμητρα	1970
9	Μελάς	Νικόλαος	1969
8	Ιωάννου	Σοφία	1973
4	Παπαγεωργίου	Σοφοκλής	1971

B  
L  
O  
C  
K  
S

Στο παράδειγμα, κάθε εγγραφή αποτελείται από τέσσερα πεδία, κάθε πεδίο καταλαμβάνει έναν αριθμό από byte, κάθε block «χωράει» μέχρι τρεις εγγραφές και το αρχείο αποτελείται από τρία block.

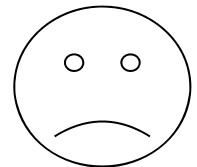
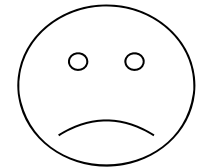
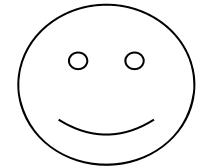
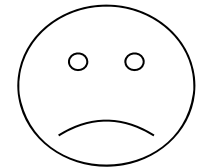
# Αρχεία σωρού ή μη-ταξινομημένα αρχεία (Heap or pile files)

- Εισαγωγή εγγραφής στο τέλος του αρχείου
- Γραμμική αναζήτηση  
(χρόνος για κλειδί =  $n/2$  (μ.ο.), για μη κλειδί =  $n$ )
- Σημάδι διαγραφής (deletion flag)  
Αναδιοργάνωση μετά από αριθμό διαγραφών



# Ταξινομημένα αρχεία [ως προς ένα πεδίο] (Sequential files)

- Πεδίο ταξινόμησης - κλειδί ταξινόμησης  
(ordering field - ordering key)
- Εισαγωγή στη σωστή θέση
- Δυαδική αναζήτηση για το πεδίο ταξινόμησης  
(χρόνος (μ.ο.) =  $\log_2 n$ )
- Γραμμική αναζήτηση για άλλο πεδίο
- Σημάδι διαγραφής (deletion flag)  
Αναδιοργάνωση μετά από αριθμό διαγραφών



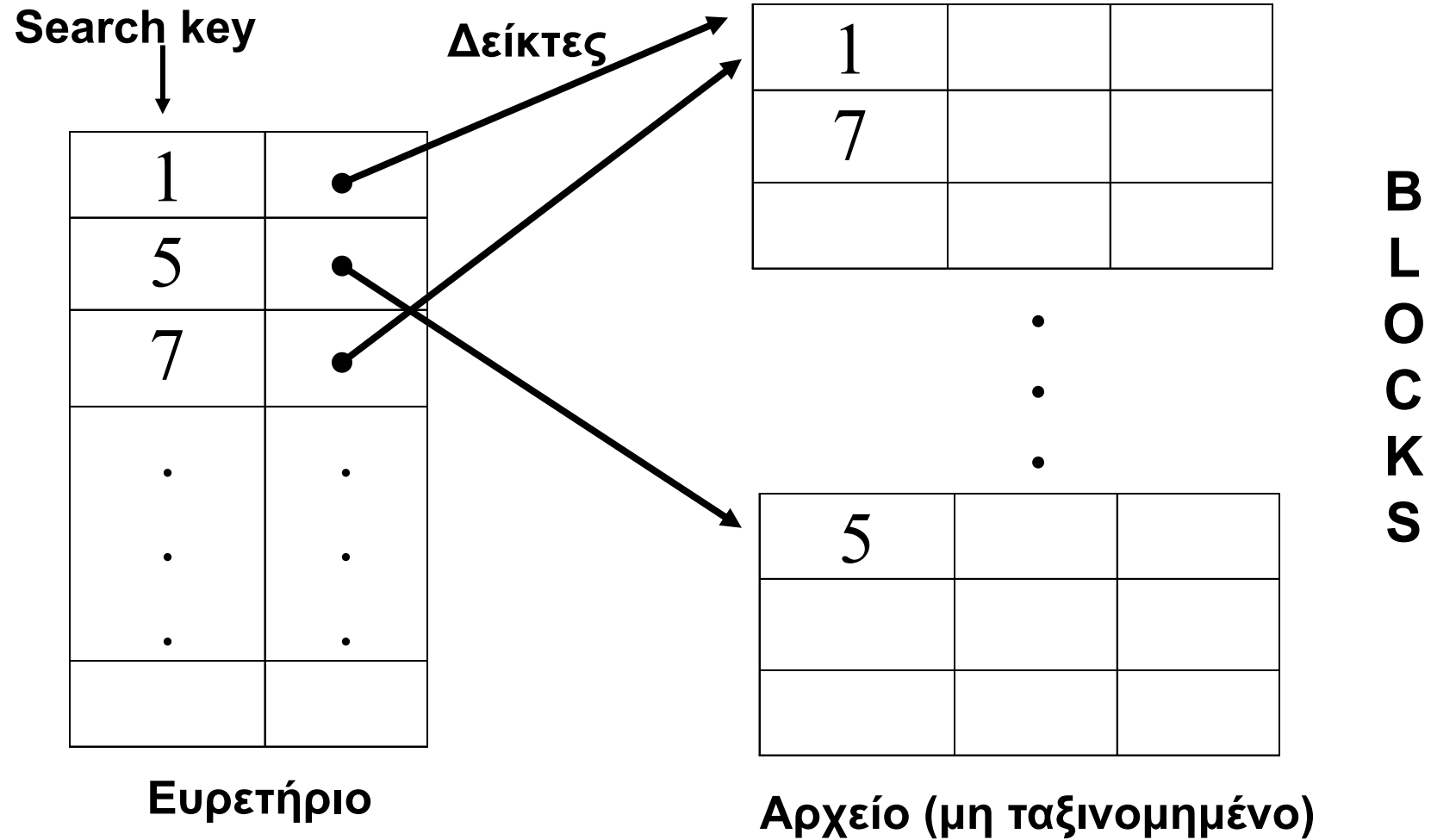
# Ευρετήριο (index)

Επιπρόσθετη δομή για την αποδοτικότερη εύρεση εγγραφών σε αρχείο. Το ευρετήριο είναι και αυτό αρχείο.

- Κάθε ευρετήριο βασίζεται σε ένα πεδίο του αρχείου εγγραφών, το οποίο καλείται *κλειδί αναζήτησης* (*search key*)
- Συνήθως το κλειδί αναζήτησης είναι το πρωτεύων κλειδί (primary key) του αρχείου.



# Γενική αρχιτεκτονική ευρετηρίων



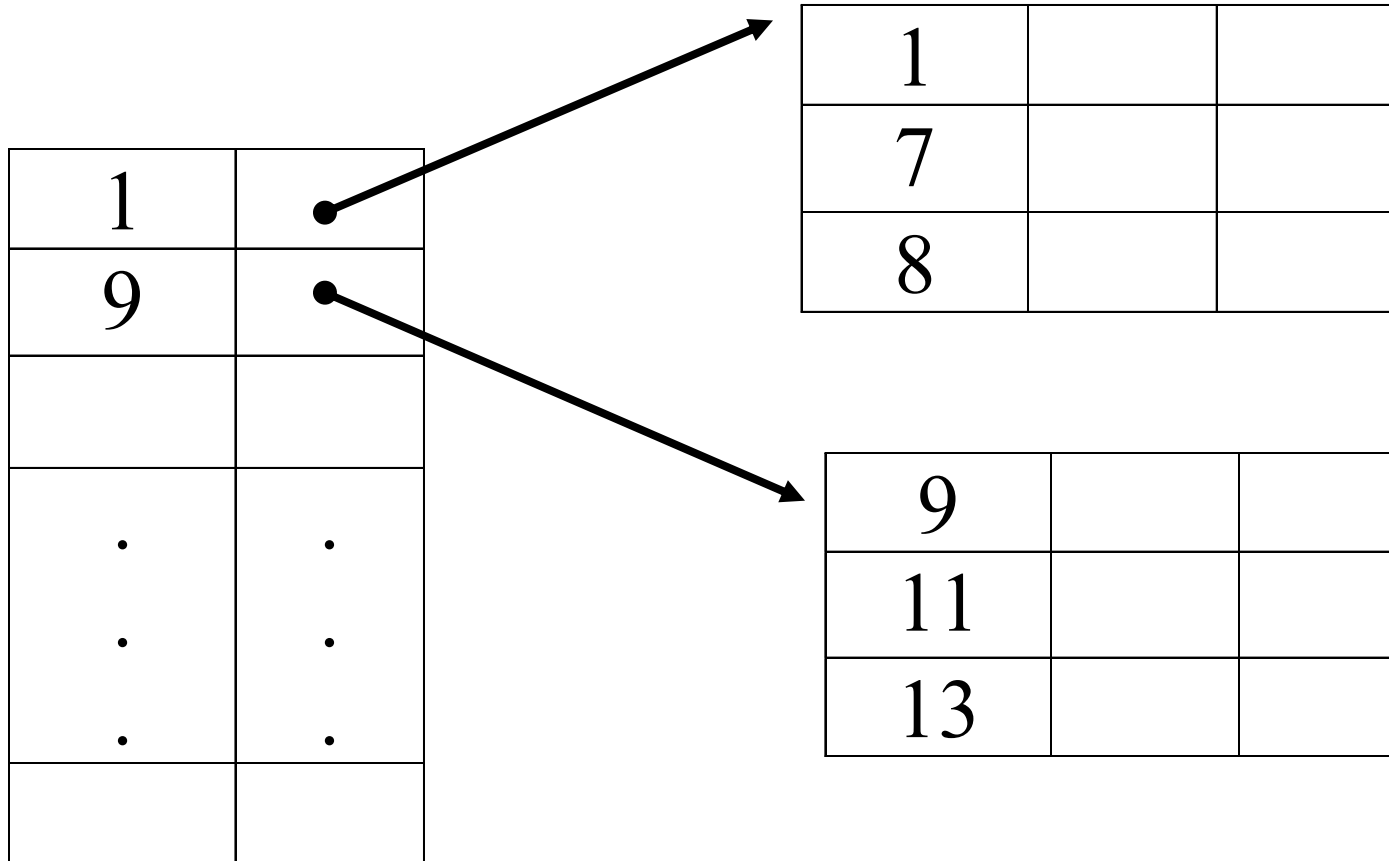
# Κατηγορίες ευρετηρίων

- **Πυκνά (dense):** Καταχωρείται μια εγγραφή στο ευρετήριο για κάθε εγγραφή του αρχείου [πλεονέκτημα: γρήγορη προσπέλαση των εγγραφών του αρχείου]
- **Αραιά (sparse):** Καταχωρείται μια εγγραφή στο ευρετήριο ανά ομάδα εγγραφών στο αρχείο (συνήθως ανά block) [πλεονεκτήματα: λιγότερος χώρος, λιγότερη επιβάρυνση στις τροποποιήσεις]

# Είδη ευρετηρίων

- **Πρωτεύων ευρετήριο (primary index)**. Το search key είναι το ordering key του αρχείου (για ταξινομημένα αρχεία)
- **Ευρετήριο συστάδας (clustering index)**. Το search key είναι το ordering field του αρχείου (για ταξινομημένα αρχεία)
- **Δευτερεύων ευρετήριο (secondary index)**. Το search key δεν αποτελεί ordering field του αρχείου (δεν υπάρχει ταξινόμηση)

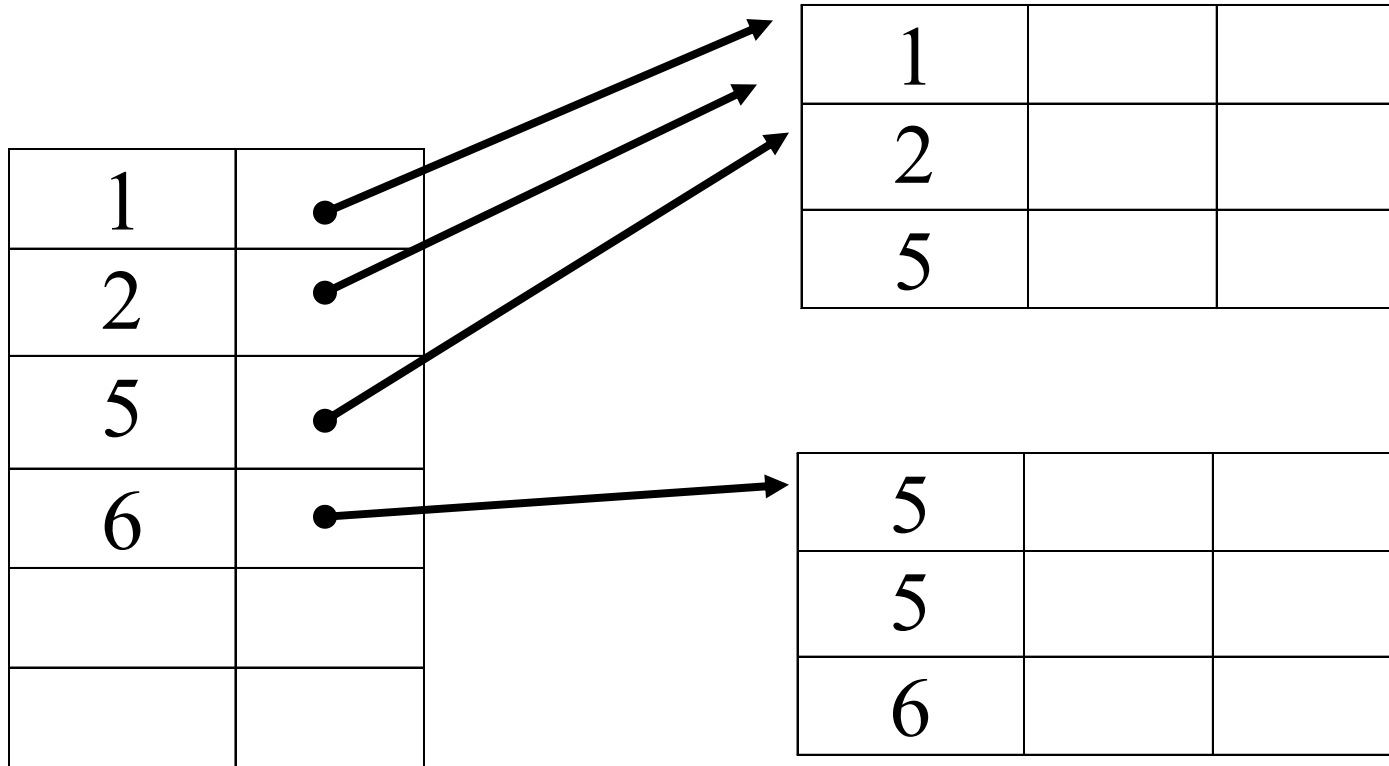
# Primary index



Ευρετήριο (sparse)

Αρχείο (ταξινομημένο - key)

# Clustering index

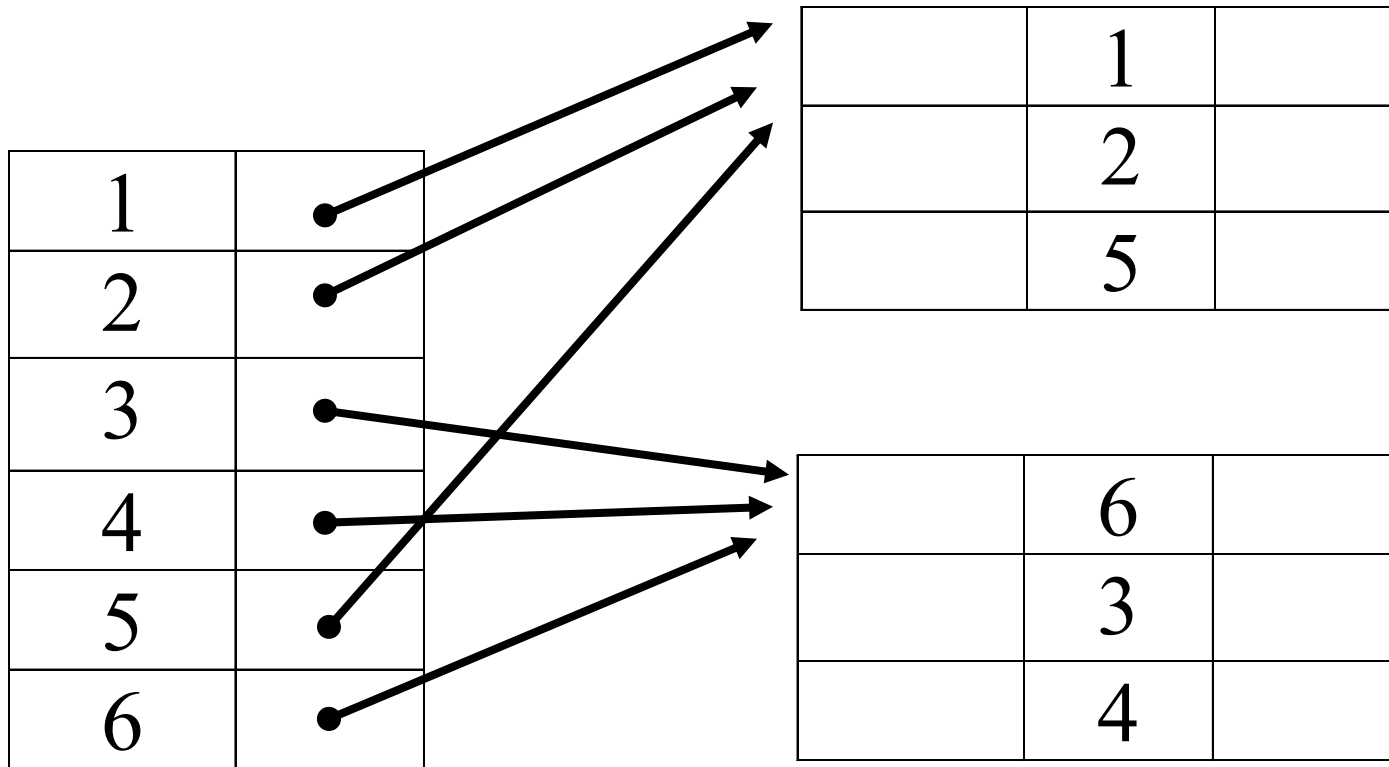


Ευρετήριο (sparse)

Αρχείο (ταξινομημένο - not key)

# Secondary index

Search key = candidate key του αρχείου

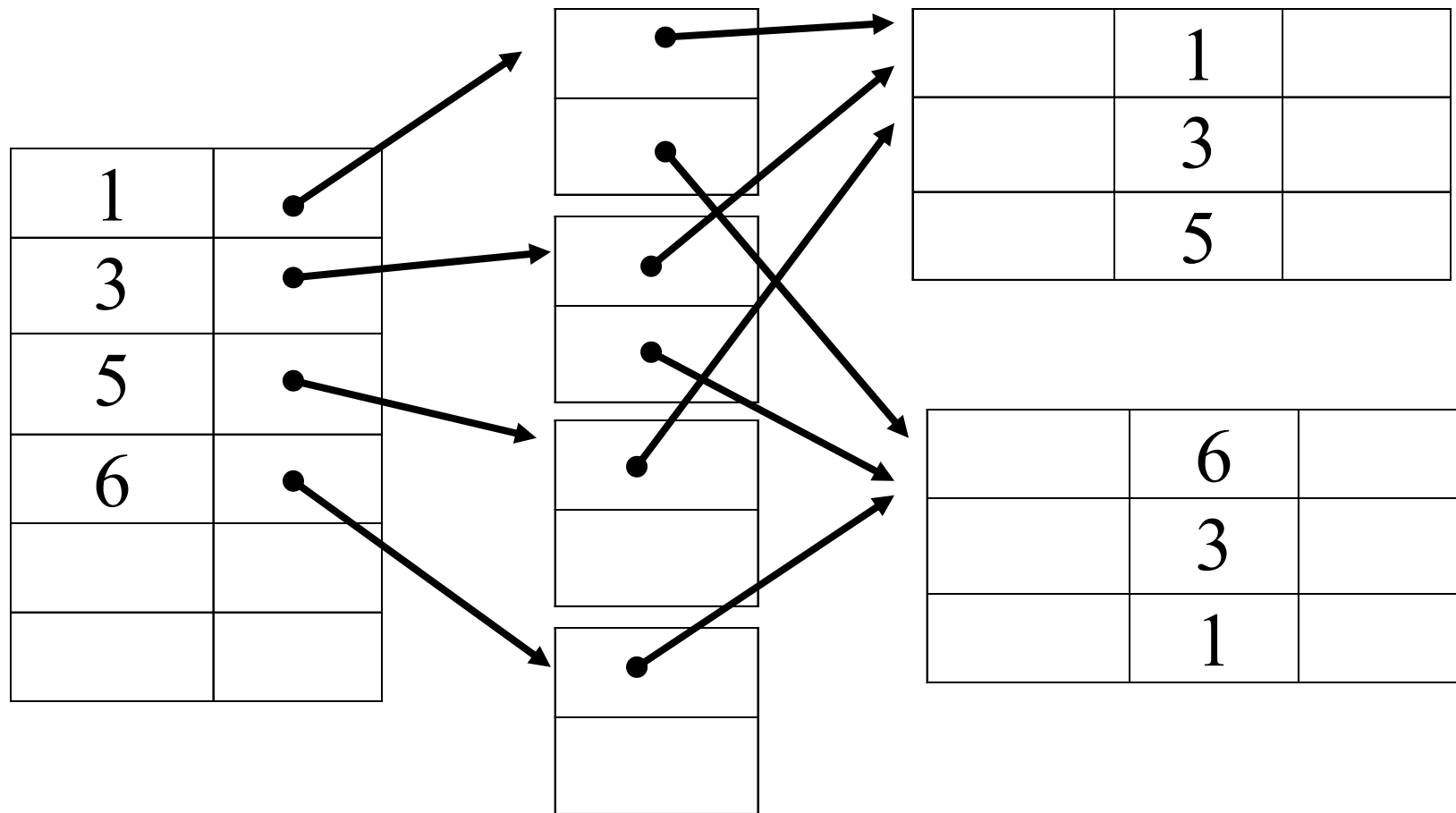


Ευρετήριο (dense)

Αρχείο (μη ταξινομημένο ως προς το candidate key)

# Secondary index

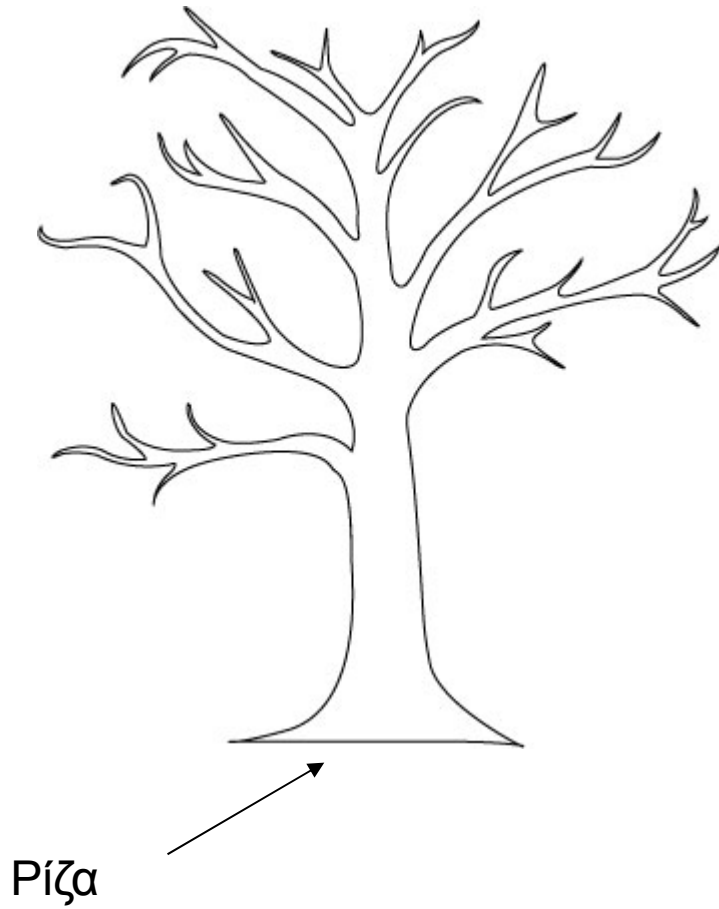
Search key = οποιοδήποτε πεδίο του αρχείου (not key)



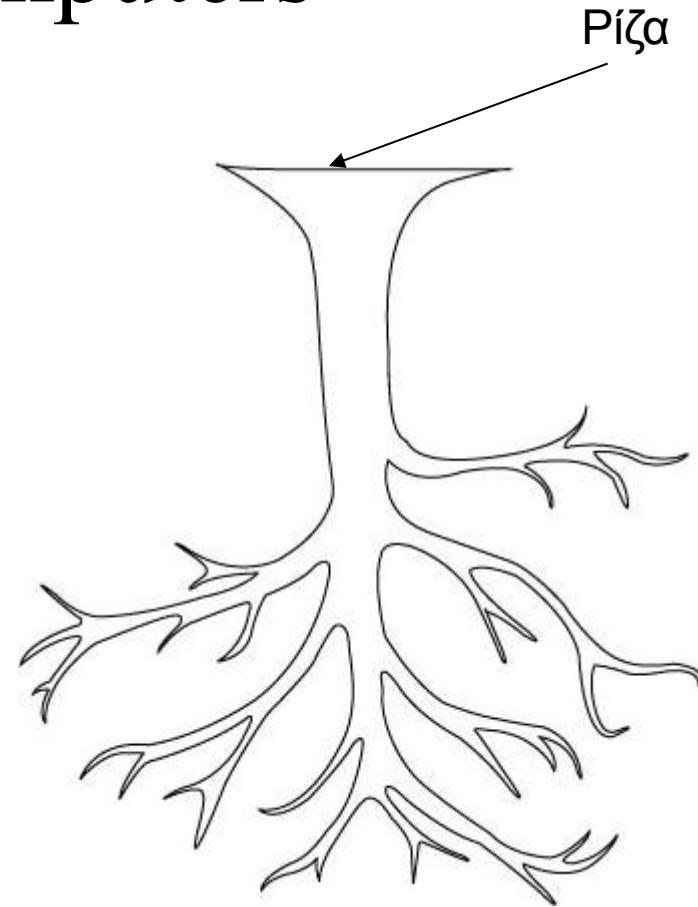
Ευρετήριο (sparse)

Αρχείο (μη ταξινομημένο)

# Tree in nature

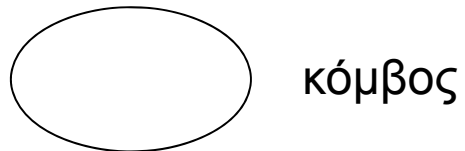
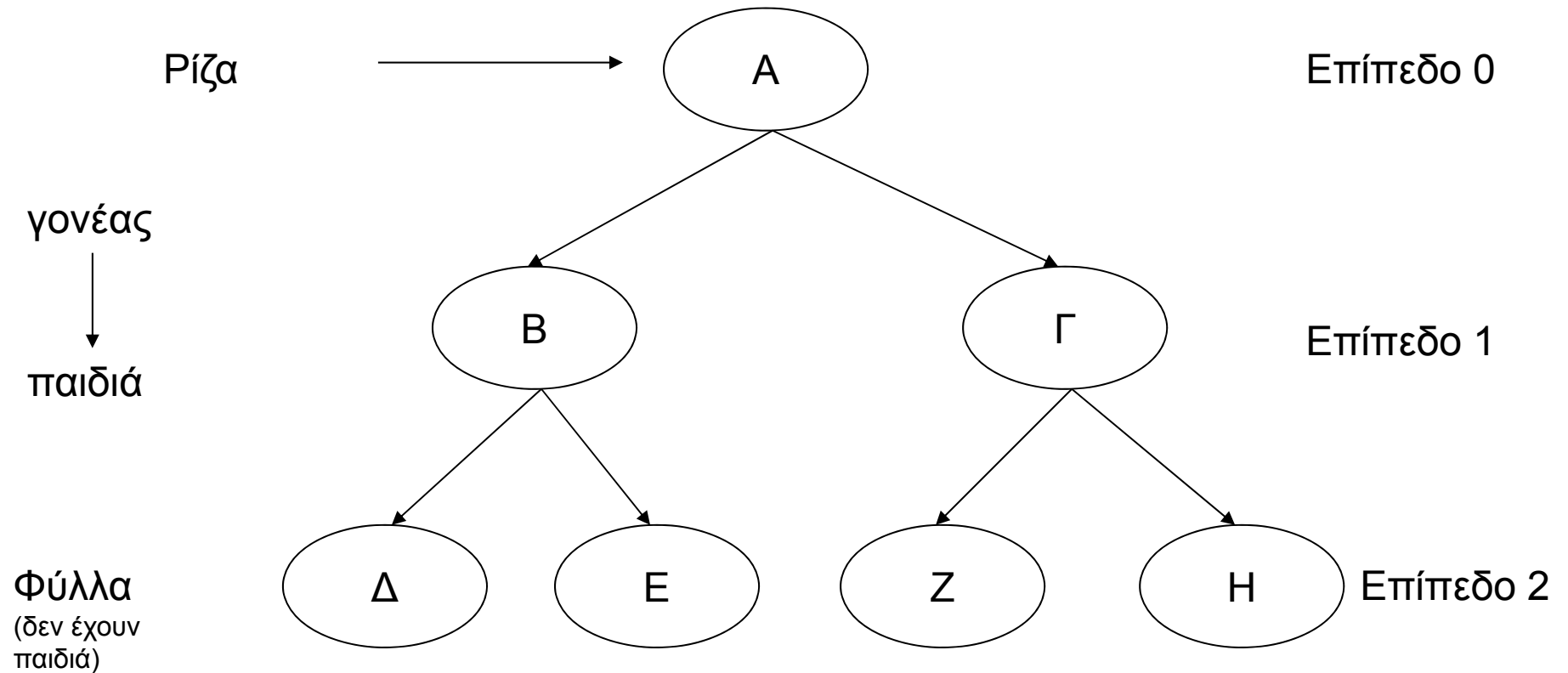


# Tree in computers





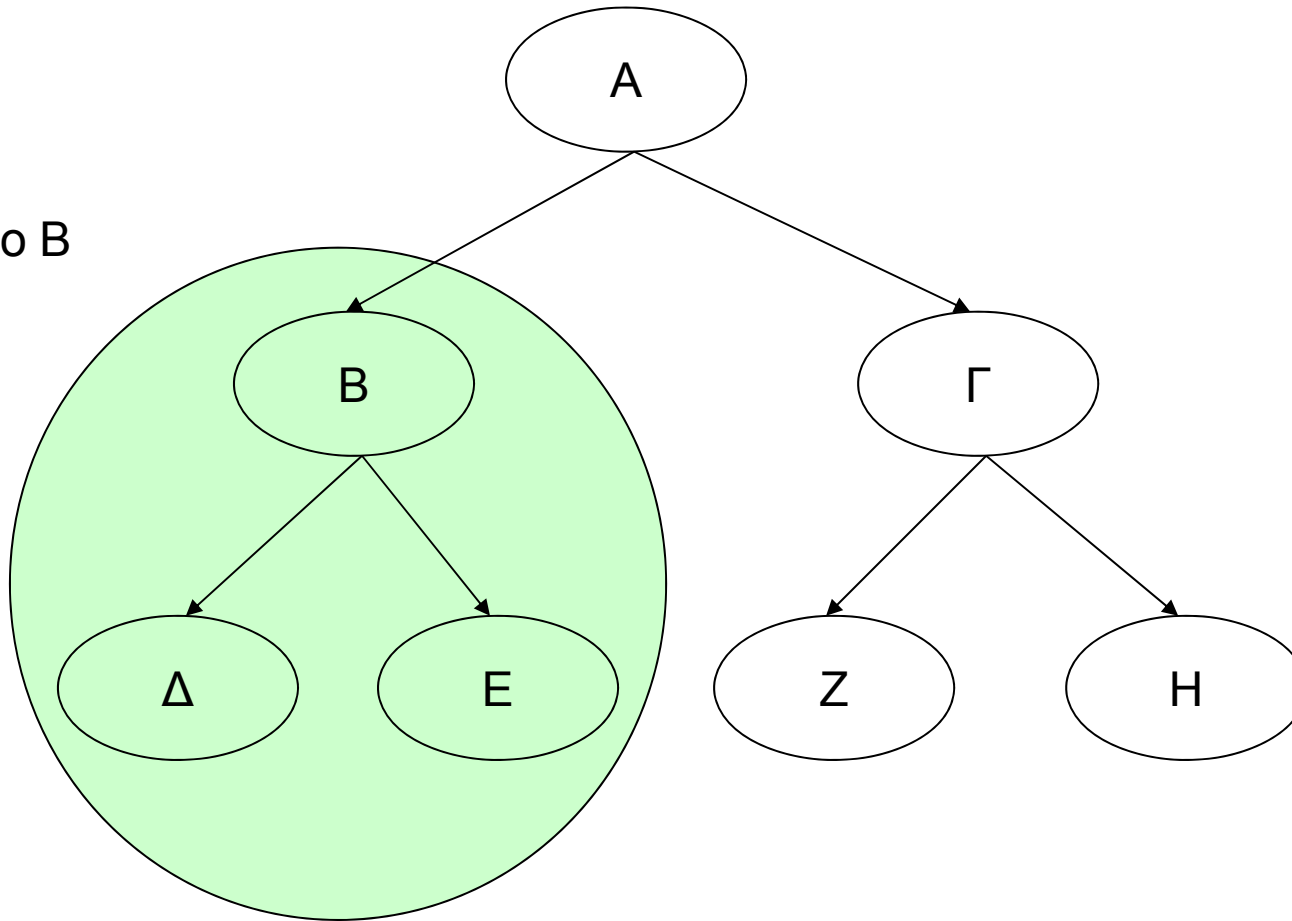
# Δέντρο



Κάθε κόμβος έχει ακριβώς ένα γονέα. Η ρίζα δεν έχει γονέα.

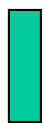
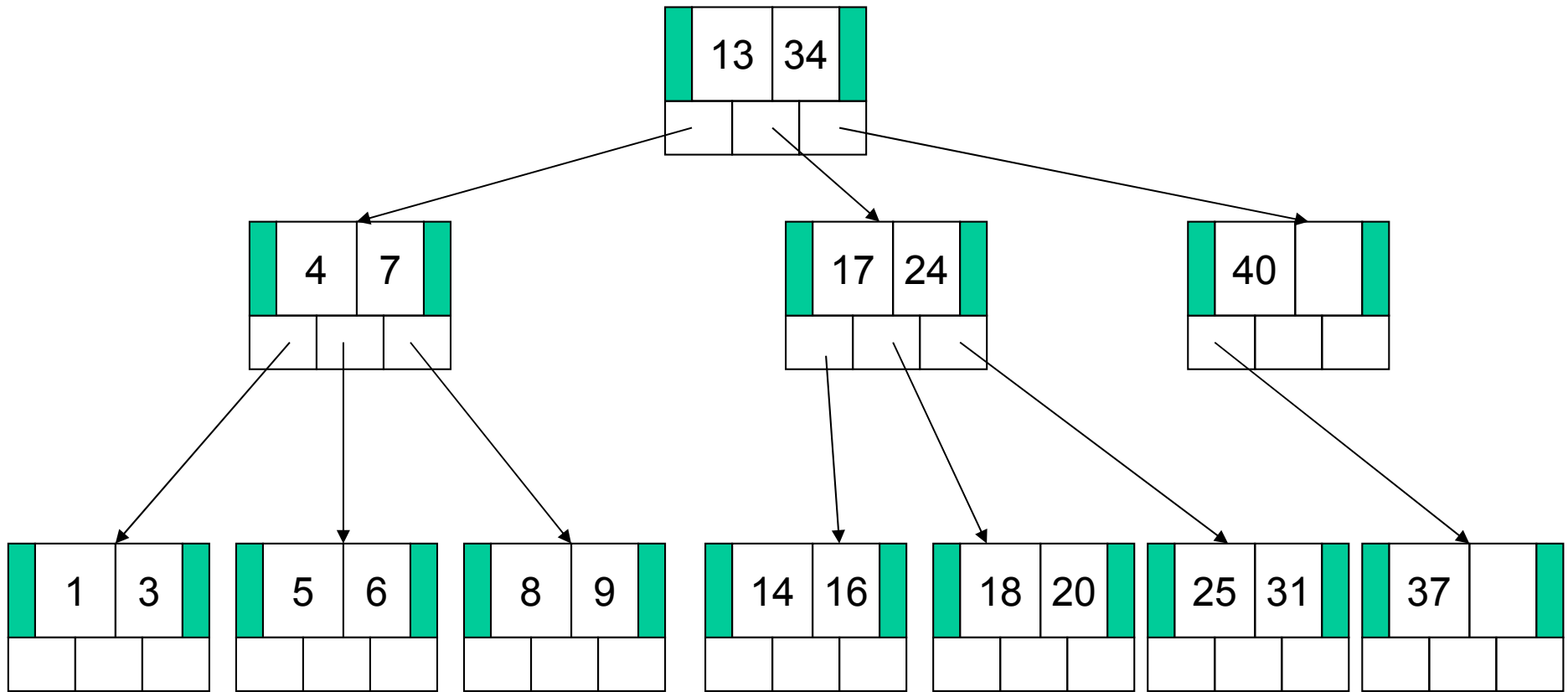
# Δέντρο

Υποδέντρο Β



**Ευρετήρια βασισμένα σε δέντρα**

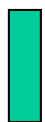
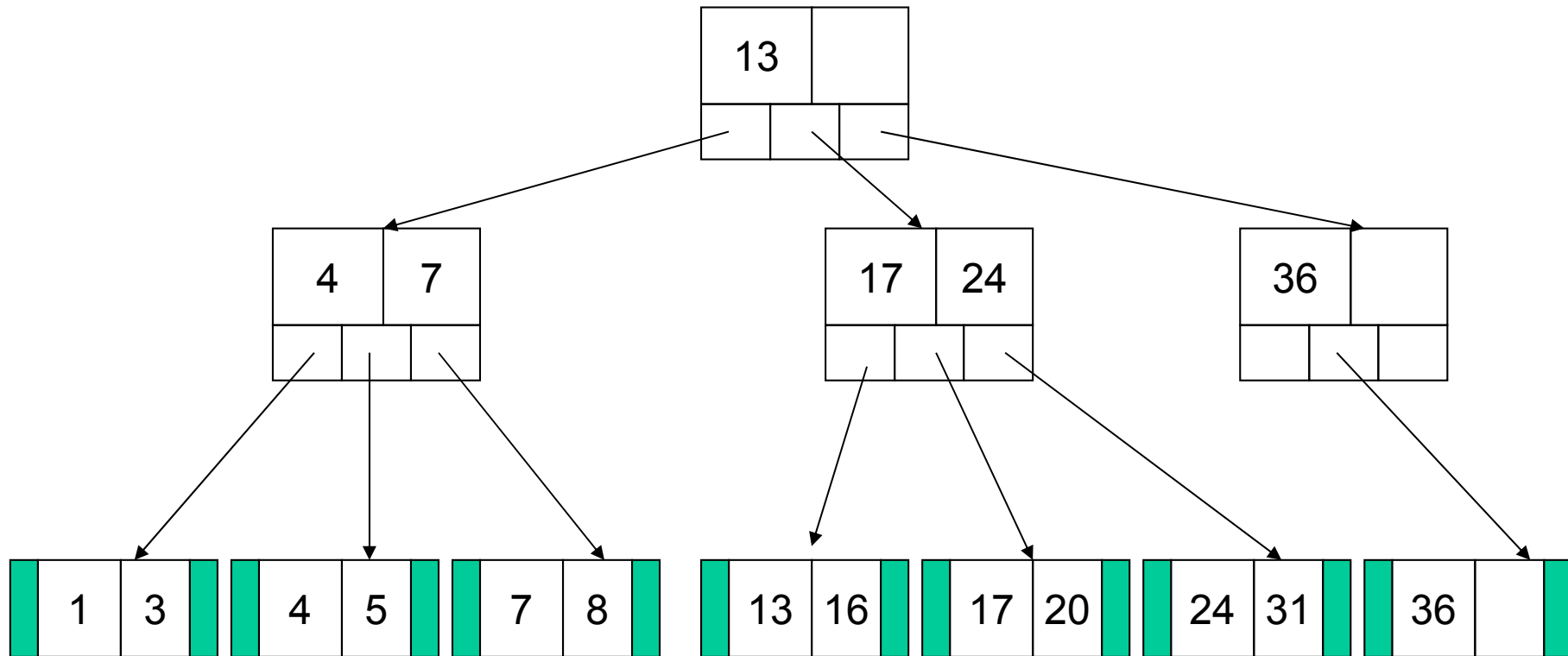
# B-tree



Δείκτης στο block που περιέχει την εγγραφή

Όλοι οι κόμβοι έχουν την ίδια δομή. Κάθε πληροφορία αποθηκεύεται μία μόνο φορά (είτε σε εσωτερικό κόμβο είτε σε φύλλο).

# B+tree

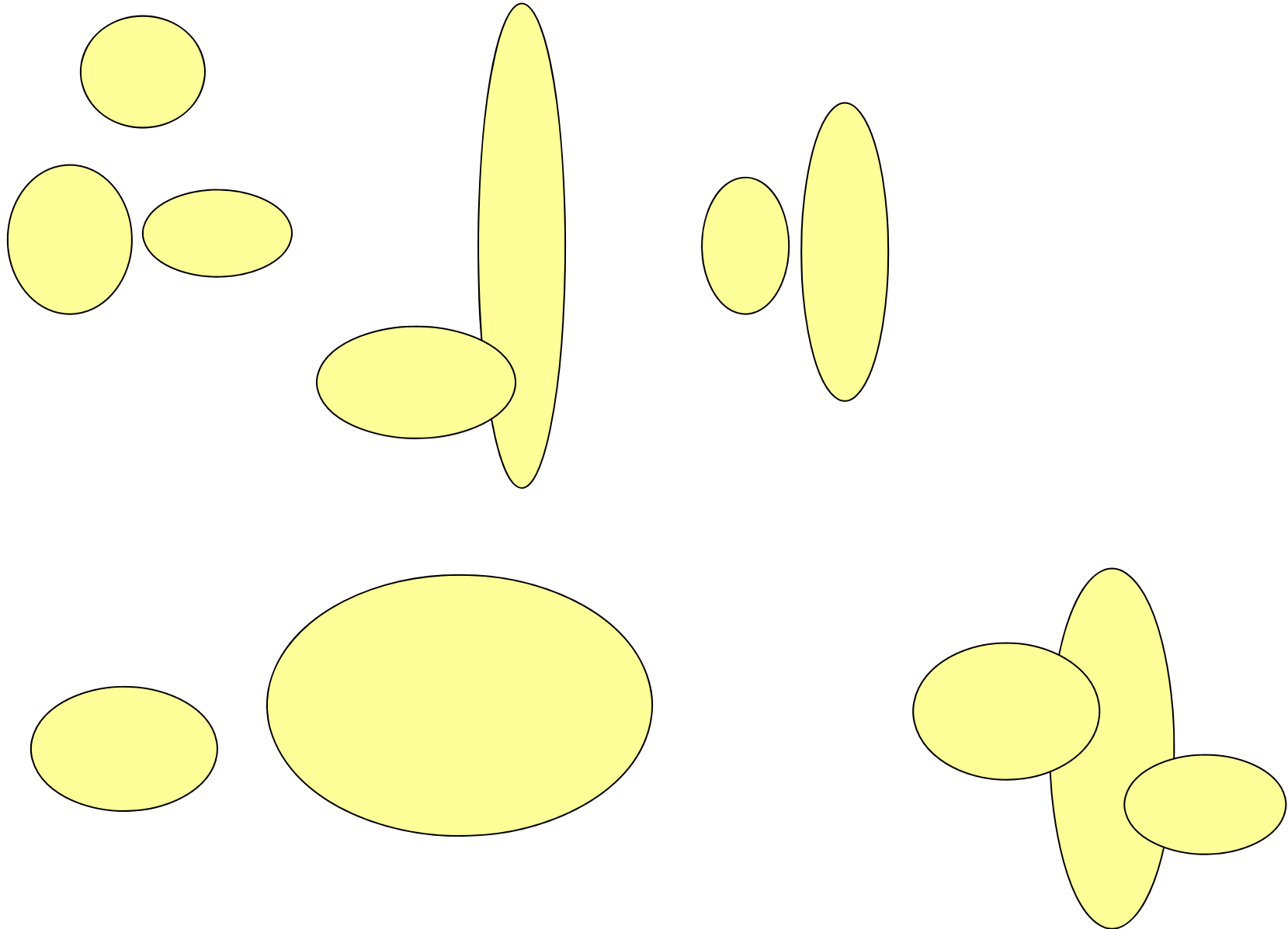


Δείκτης στο block που περιέχει την εγγραφή

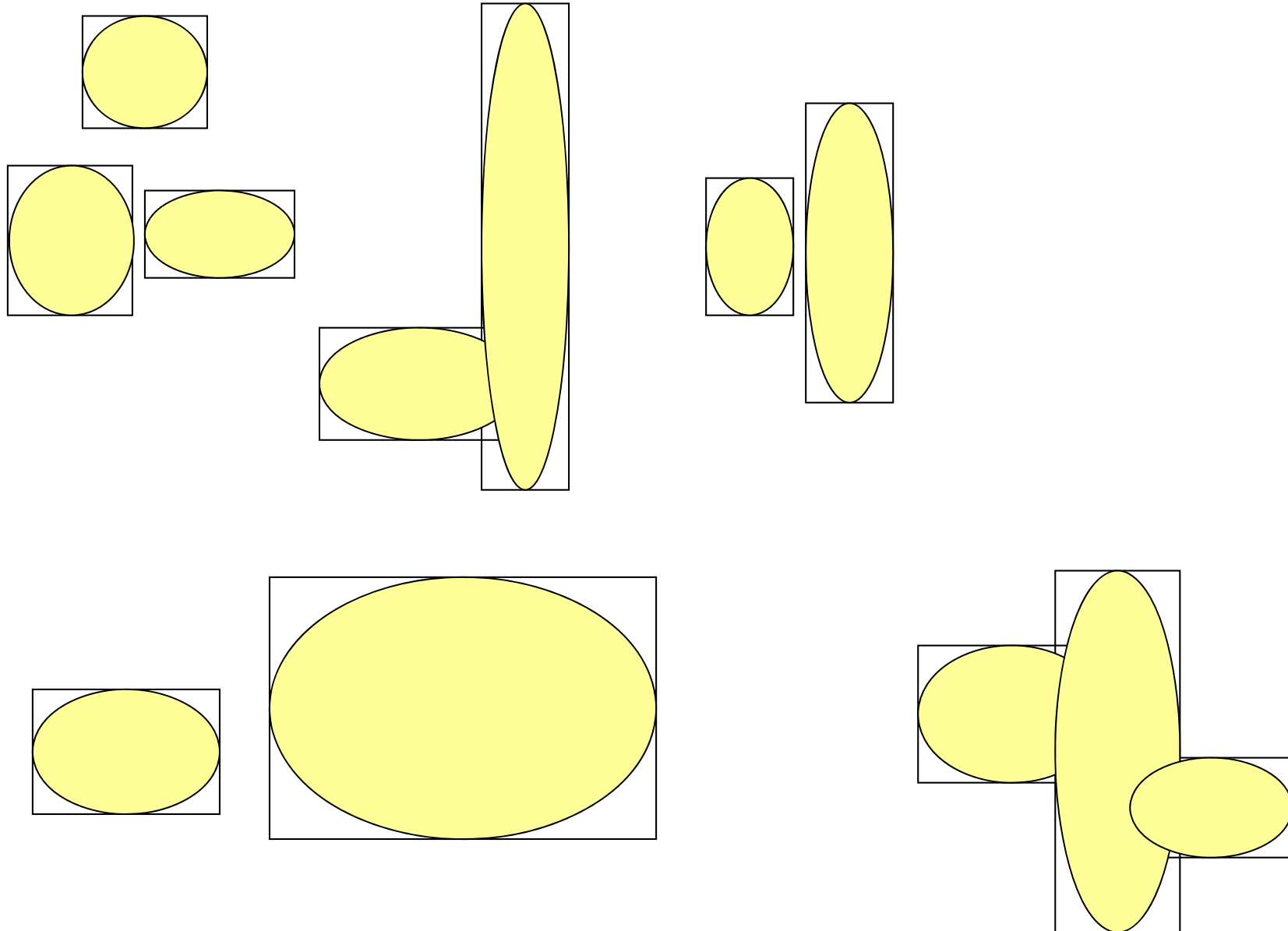
Μόνο τα φύλλα έχουν δείκτες.  
Κάθε πληροφορία αποθηκεύεται οπωσδήποτε σε ένα φύλλο. Κάποιες πληροφορίες αποθηκεύονται και σε εσωτερικό κόμβο.

**Ευρετήρια για χωρικά δεδομένα**

# Χωρικά αντικείμενα

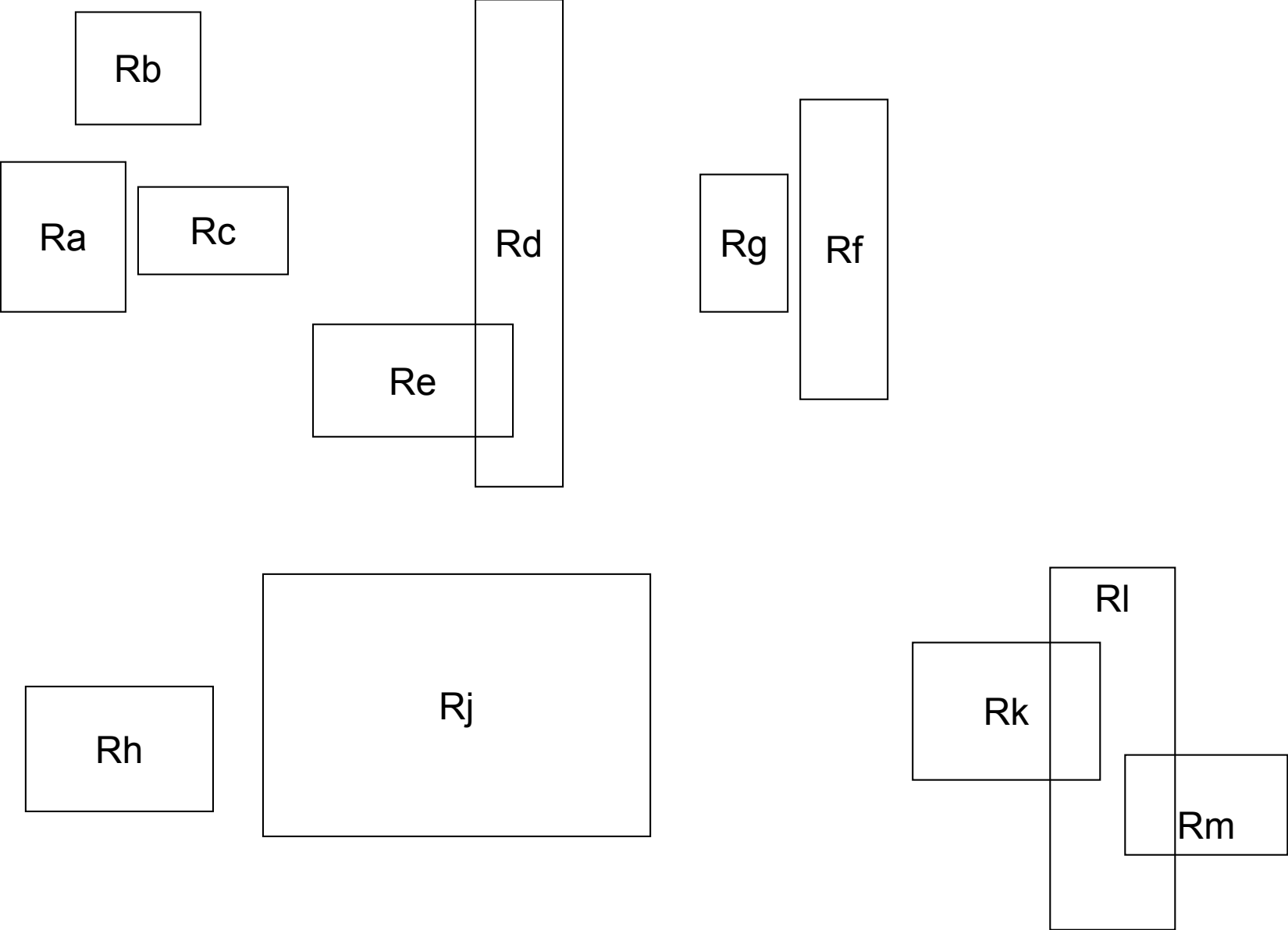


# Εύρεση του minimum bounding box για κάθε αντικείμενο

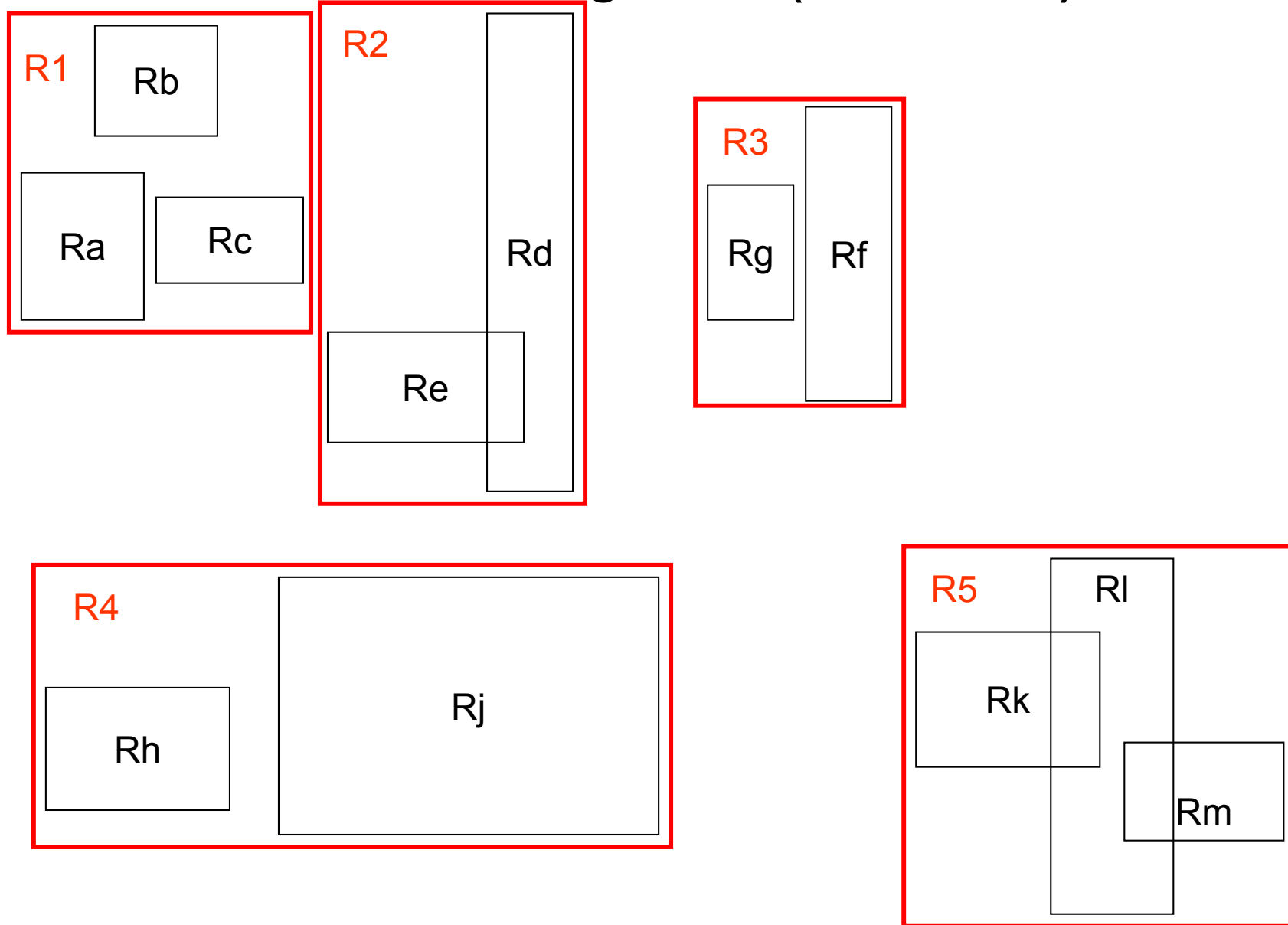




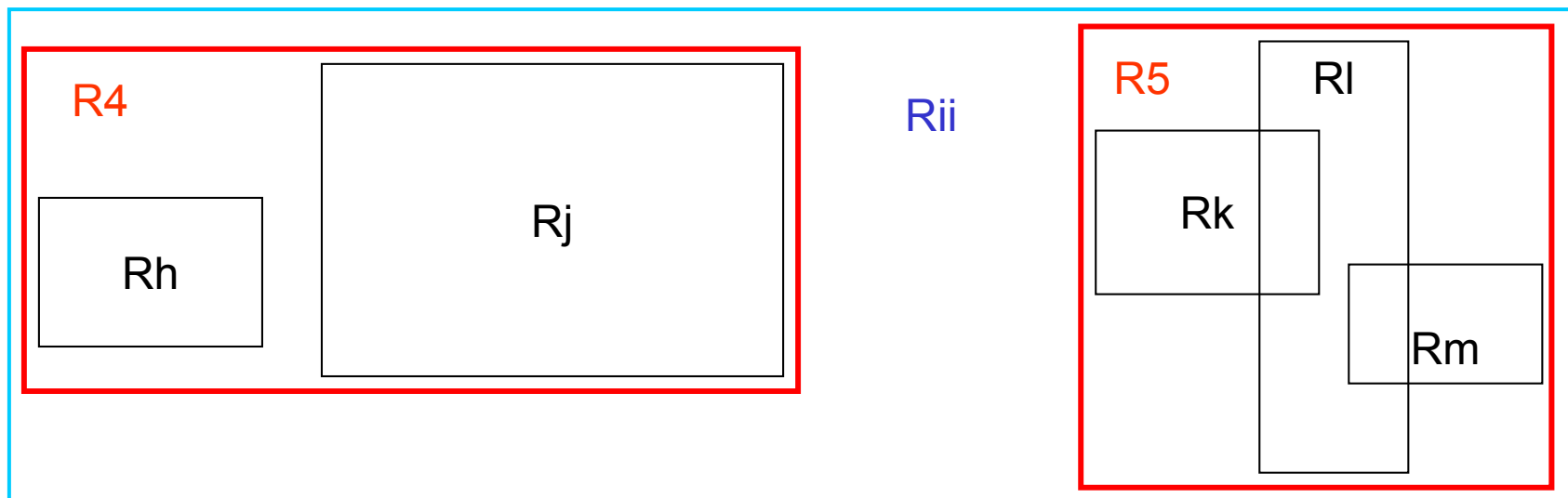
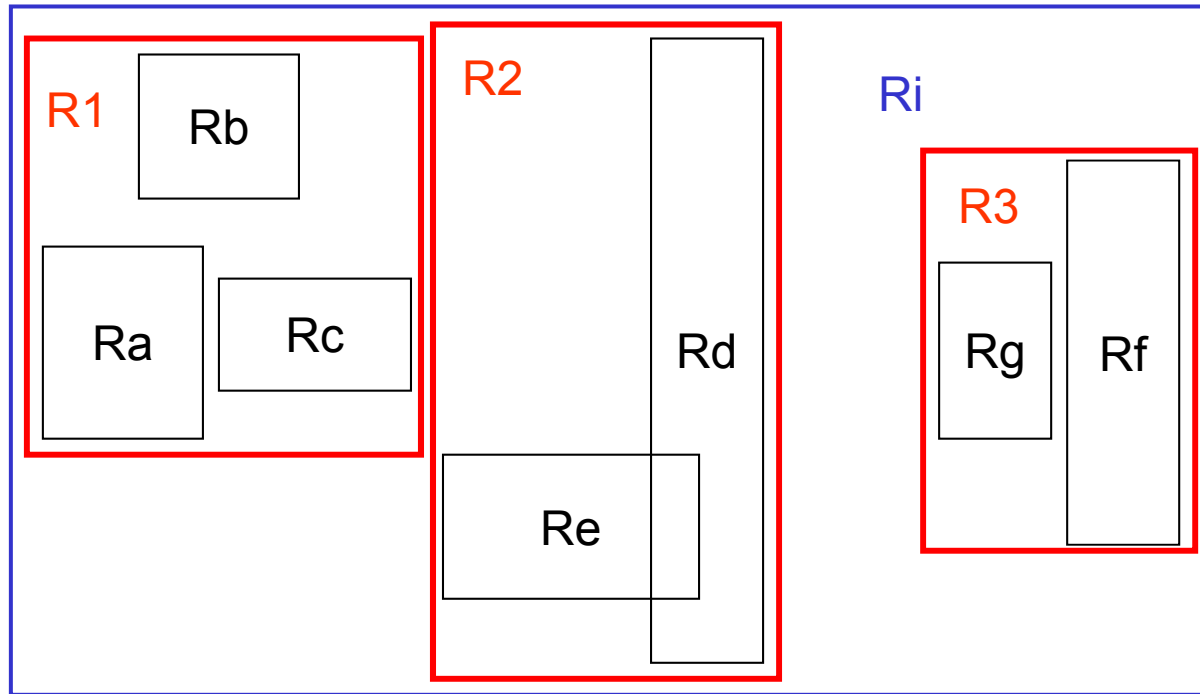
# Minimum bounding boxes



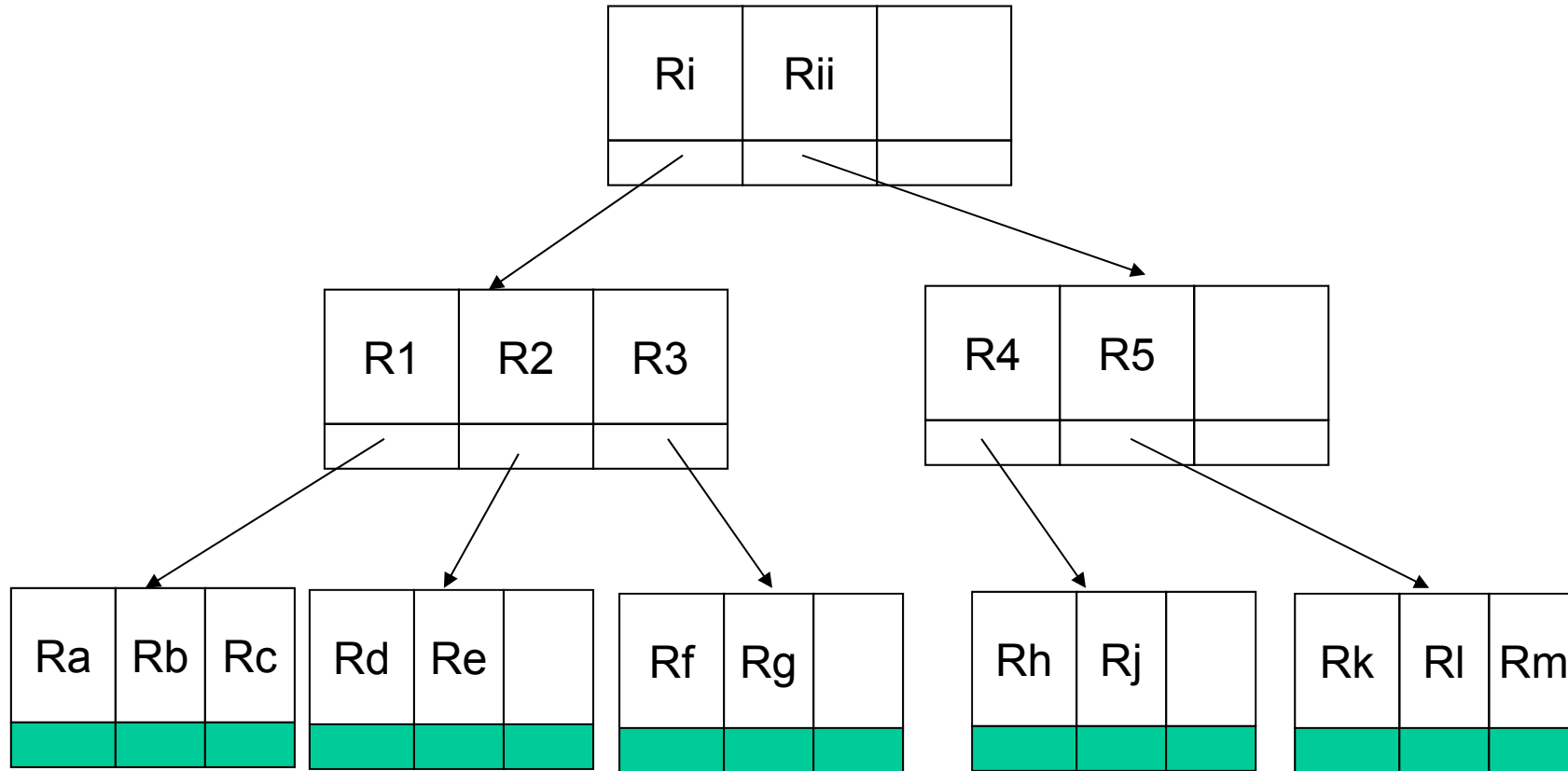
# Ομαδοποίηση γειτονικών minimum bounding boxes (1<sup>ο</sup> επίπεδο)



# Ομαδοποίηση (2<sup>ο</sup> επίπεδο)



# R-tree



Δείκτης στο block που περιέχει το χωρικό αντικείμενο

Η εύρεση του χωρικού αντικειμένου γίνεται μέσω της εύρεσης του mbb που αντιστοιχεί σε αυτό.

# Τέλος Ενότητας



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

