

Δέντρα Πολλαπλής Διακλάδωσης και (a, b)-Δέντρα

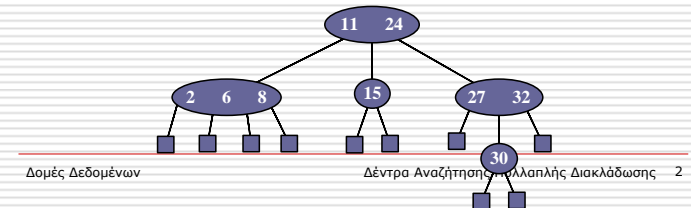
Δημήτρης Φωτάκης

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

Πανεπιστήμιο Αιγαίου

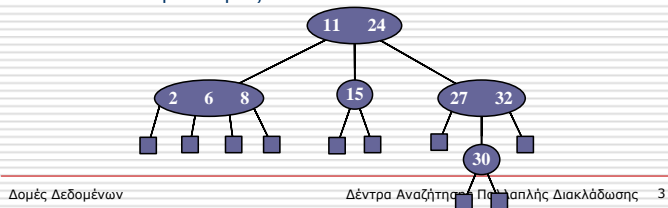
Δέντρα Αναζήτησης Πολλαπλής Διακλάδωσης (ΔΑΠΔ)

- ΔΑΠΔ ή Δέντρα m -δρόμων: ΣΔ Βάσεων Δεδομένων.
- Γενίκευση των Δυαδικών Δέντρων Αναζήτησης:
 - Δέντρο με ρίζα και εσωτερικοί κόμβοι έχουν τουλάχιστον 2 και το πολύ m παιδιά.
 - Εσωτερικός κόμβος με d παιδιά περιέχει $d - 1$ στοιχεία ταξινομημένα σε αύξουσα σειρά.
 - Εξωτερικοί κόμβοι: NULL-δείκτες (φύλλα).



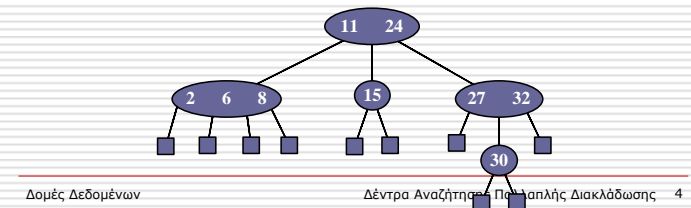
Ιδιότητα ΔΑΠΔ

- Κόμβος με d παιδιά και $d - 1$ στοιχεία:
 - Στοιχεία 1^{ου} υποδέντρου \leq 1^ο στοιχείο.
 - $k - 1$ στοιχείο \leq στοιχεία k -οστού υποδέντρου \leq k -οστό στοιχείο, $k = 2, \dots, d - 1$.
 - $d - 1$ στοιχείο \leq στοιχεία d -οστού υποδέντρου.
- Στοιχεία $k, k+1$ ορίζουν διάστημα για στοιχεία k υποδέντρου.
 - Γενίκευση ιδιότητας ΔΔΑ.



Χρησιμότητα

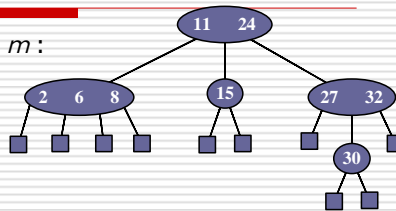
- Πολλά παιδιά : μικρότερο ύψος για ίδια στοιχεία.
 - Λειτουργίες επισκέπτονται μικρό αριθμό κόμβων.
 - Βάσεις Δεδομένων (DBMSs): κόμβοι στο σκληρό δίσκο. Προσπέλαση κόμβου : ανάγνωση από δίσκο (αργή!)
 - «Μεγάλοι» κόμβοι (πολλά παιδιά) ώστε λειτουργίες με μικρό αριθμό I/O's.



Αναπαράσταση

- Βαθμός διακλάδωσης $\leq m$:

```
typedef struct _mtnode {
    struct _mtnode *par;
    struct _mtnode *ch[m];
    int key[m-1];
    int ne; // #στοιχείων
} mtnode;
mtnode root = NULL;
```



- Λειτουργίες:

- Αναζήτηση, εισαγωγή, διαγραφή
- Διελεύσεις: inOrder(), ...
- Μετρήσεις: height(), size()
- Μέγιστο, ελάχιστο, προηγούμενος, επόμενος, ...

Δομές Δεδομένων

Δέντρα Αναζήτησης Πολλαπλής Διακλάδωσης 5

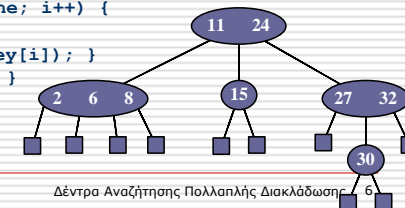
Inorder

- Ενδο-διατεταγμένη (inorder) διέλευση:

- 1^ο υποδέντρο - 1^ο στοιχείο - 2^ο υποδέντρο - 2^ο στοιχείο - ... - (d-1)-οστό στοιχείο - d-υποδέντρο.
- k-οστό στοιχείο εξετάζεται **μετά** από (k-1)-οστό υποδέντρο και **πριν** από κόμβους k-οστού υποδέντρο.

```
void inOrder(mtnode *x) {
    if (x == NULL) return;
    for (int i = 0; i < x->ne; i++) {
        inOrder(x->ch[i]);
        printf("%d ", x->key[i]);
        inOrder(x->ch[x->ne]);
    }
}
```

inOrder(root);

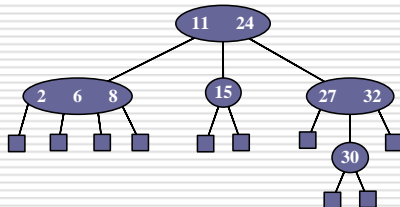


Δομές Δεδομένων

Δέντρα Αναζήτησης Πολλαπλής Διακλάδωσης

Inorder

- Τυπώνει στοιχεία σε **αύξουσα σειρά**.
- Γραμμικός χρόνος εκτέλεσης : $\Theta(n)$
 - Εύκολη απόδειξη με επαγωγή.

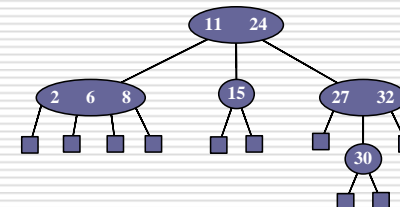


Δομές Δεδομένων

Δέντρα Αναζήτησης Πολλαπλής Διακλάδωσης 7

Άλλες Λειτουργίες

- Ελάχιστο στοιχείο: αριστερότατο
- Μέγιστο στοιχείο: δεξιότατο.
- Ύψος: 1+μέγιστο ύψος υποδέντρων.
- Μέγεθος: διέλευση και μέτρηση.
- Επόμενος / προηγούμενος ;

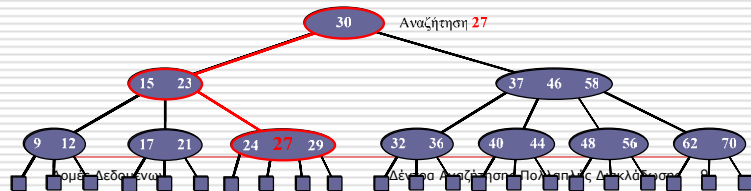


Δομές Δεδομένων

Δέντρα Αναζήτησης Πολλαπλής Διακλάδωσης 8

Αναζήτηση

- Ακολουθούμε κατάλληλο μονοπάτι από ρίζα προς φύλλα:
 - Σε κάθε κόμβο u , διαλέγουμε υποδέντρο:
 - Αν $x == u \rightarrow key[k]$, **εύρεση**
 - Αν $x < u \rightarrow key[k]$, ακολουθώ $u \rightarrow ch[k]$
 - Αν $x > u \rightarrow key[k]$, $k = k+1$
 - Αν $k == u \rightarrow ne$, ακολουθώ $u \rightarrow ch[u \rightarrow ne]$

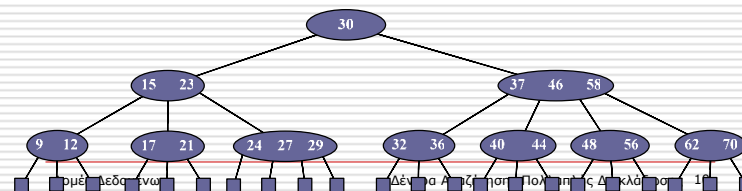


Αναζήτηση

```

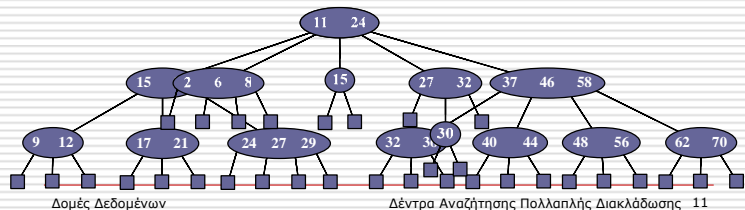
tmnode *mSearch(int x, tmnode *u) {
    if (u == NULL) return(NULL);
    for (int i = 0; (i < u->ne) && (x > u->key[i]); i++);
    if (i == u->ne) return(mSearch(x, u->ch[u->ne]));
    if (x == u->key[i]) return(u);
    else return(mSearch(x, u->ch[i]));
}
    
```

- Ορθότητα : ιδιότητα ΔΑΠΔ
- Χρόνος : $O(d \times \text{ύψος})$. Μπορεί να γίνει $O(\log d \times \text{ύψος})$.



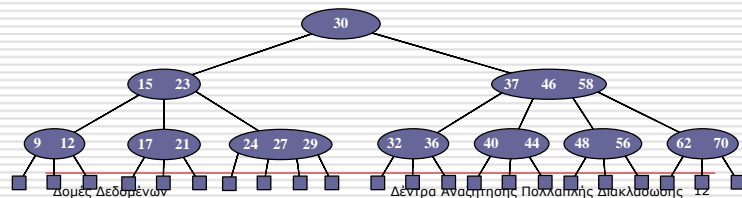
(a, b)-Δέντρα

- Ζυγισμένη εκδοχή ΔΑΠΔ.
- Ακέραιοι $a \geq 2$ και $b \geq 2a$. (a, b)-δέντρο:
 - Εσωτερικοί κόμβοι έχουν $\leq b$ παιδιά.
 - Ρίζα έχει τουλάχιστον 2 παιδιά.
 - Εσωτερικοί κόμβοι (εκτός ρίζας) τουλάχιστον a παιδιά.
 - Πλήρης ζύγισση: Εξωτερικοί κόμβοι (NULL-φύλλα) ίδιο βάθος.



(a, b)-Δέντρα

- Β-δέντρο τάξης b : $a = \lceil \frac{b}{2} \rceil$, π.χ. (2, 4), (4, 7), κοκ.
- B*-δέντρο τάξης b : $a = \lceil \frac{2b}{3} \rceil$, π.χ. (3, 4), (5, 7), κοκ.
- (2, 4)-δέντρα: απλό παράδειγμα Β-δέντρων.
- Συνήθως μεγαλύτερες τιμές για a και b .
- Διάφορες μορφές (a, b)-δέντρων σε όλα τα DBMSs !



Ύψος (a, b)-Δέντρων

- Ελάχιστο ύψος h ενός (a, b) -Δέντρου με n στοιχεία:
 - «Μέγιστη» διακλάδωση.
 - Όλοι οι εσωτερικοί κόμβοι είναι πλήρεις (b παιδιά και $b - 1$ στοιχεία ο καθένας).

$$(b-1) + b(b-1) + b^2(b-1) + \dots + b^{h-1}(b-1) = (b-1) \frac{b^h - 1}{b-1}$$

ρίζα επ. 1 επ. 2 ... επ. $h-1$

$$n \leq b^h - 1 \Rightarrow h \geq \log_b(n+1)$$

Δομές Δεδομένων

Δέντρα Αναζήτησης Πολλαπλής Διακλάδωσης 13

Ύψος (a, b)-Δέντρων

- Μέγιστο ύψος h ενός (a, b) -Δέντρου με n στοιχεία:
 - «Ελάχιστη» διακλάδωση.
 - Ρίζα έχει 2 παιδιά και 1 στοιχείο.
 - Λοιποί εσωτερικοί κόμβοι έχουν a παιδιά και $a - 1$ στοιχεία ο καθένας.

$$1 + 2(a-1) + 2a(a-1) + \dots + 2a^{h-2}(a-1) = 1 + 2(a-1) \frac{a^{h-1} - 1}{a-1}$$

ρίζα επ. 1 επ. 2 ... $h-1$ επ.

$$n \geq 2a^{h-1} - 1 \Rightarrow h \leq \log_a\left(\frac{n+1}{2}\right) + 1$$

Δομές Δεδομένων

Δέντρα Αναζήτησης Πολλαπλής Διακλάδωσης 14

Ύψος (a, b)-Δέντρων

- (a, b) -Δέντρο με ύψος h και n στοιχεία:

$$2a^{h-1} - 1 \leq n \leq b^h - 1$$

$$\log_b(n+1) \leq h \leq \log_a\left(\frac{n+1}{2}\right) + 1$$

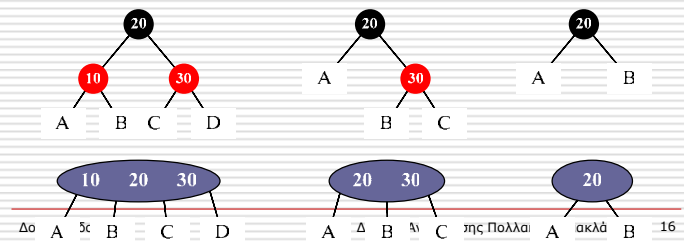
- Αναζήτηση προσπελαύνει $\leq \log_a\left(\frac{n+1}{2}\right) + 1$ κόμβους
 - Μεγαλώνει a , μειώνονται προσπελάσεις κόμβων:
 - Π.χ. $n = 2 \cdot 10^{12} = 2 \cdot 100^6$ και $a = 100$, $\log_a\left(\frac{n}{2}\right) = 6$

Δομές Δεδομένων

Δέντρα Αναζήτησης Πολλαπλής Διακλάδωσης 15

ΚΜ-Δέντρα και (2, 4)-Δέντρα

- ΚΜ-δέντρο: μαύροι κόμβοι «απορροφούν» κόκκινα παιδιά.
 - Ιδιότητα ΔΑΠΔ (λόγω ιδιότητας ΔΔΑ)
 - Εσωτερικοί κόμβοι ≥ 2 παιδιά (δυαδικό) και ≤ 4 παιδιά (όχι διαδοχικοί κόκκινοι κόμβοι).
 - Εξωτερικοί κόμβοι στο ίδιο ύψος (πλήρης ζύγιση με μαύρο ύψος)

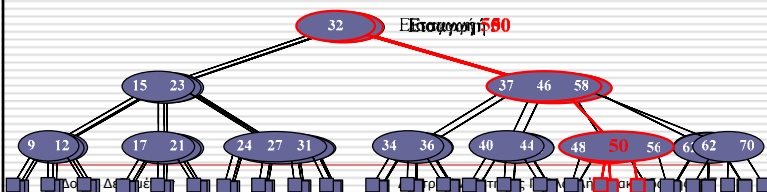


Δομ. Πολλ. Διακλάδωσης 16

Δομ. Πολλ. Διακλάδωσης 16

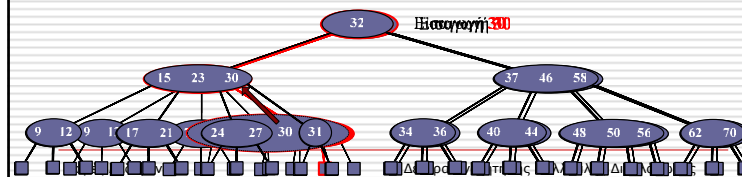
Εισαγωγή

- Στον πατέρα NULL-φύλλου όπου καταλήγει αναζήτηση.
 - Νέο στοιχείο στην αντίστοιχη θέση NULL-φύλλου όπου κατέληξε η (αποτυχημένη) αναζήτηση.
 - Στοιχεία σε αύξουσα σειρά και διατήρηση ιδιότητας ΔΑΠΔ.
 - Παραμένει ίδιο ύψος.
 - Αν κόμβος έχει $\leq b$ παιδιά, όχι επαναζυγιστικές πράξεις.



Εισαγωγή με Διάσπαση

- Αν κόμβος έχει $b + 1$ παιδιά, υπερχειλίση.
 - Δομική επαναζυγιστική πράξη **διάσπαση**.
 - Κόμβος με υπερχειλίση \Rightarrow 2 κόμβοι.
αριστερός: $1, \dots, \lfloor (b+1)/2 \rfloor - 1$ δεξιός: $\lfloor (b+1)/2 \rfloor + 1, \dots, b$
 - Μεσαίο $\lfloor (b+1)/2 \rfloor$ -στοιχείο «ανεβαίνει» στον πατέρα ως στοιχείο διαχωρισμού δύο υποδέντρων.



Εισαγωγή

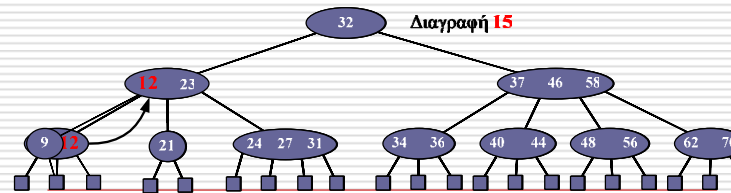
- Διάσπαση μπορεί να μην είναι τερματική:
 - Πιθανή υπερχειλίση του πατέρα.
 - Αν υπερχειλίσει η ρίζα, έχουμε νέα ρίζα με 2 παιδιά και αύξηση ύψους κατά 1.
- Χρόνος εισαγωγής: $O(\log n)$
 - Στην καλύτερη περίπτωση, καμία διάσπαση.
 - Στη χειρότερη περίπτωση, $O(\log n)$ διασπάσεις.

Δομές Δεδομένων

Δέντρα Αναζήτησης Πολλαπλής Διακλάδωσης 19

Διαγραφή

- Αν κόμβος περιέχει NULL-φύλλα, διαγραφή στοιχείου και NULL-φύλλου.
- Διαφορετικά, προηγούμενο (ή επόμενο) στοιχείο (στην inorder) παίρνει θέση διαγραφέντος.
 - Πραγματική απομάκρυνση συμβαίνει στο κατώτατο επίπεδο.
- Αν δεν προκύπτει έλλειμμα, όχι επαναζυγιστικές πράξεις.

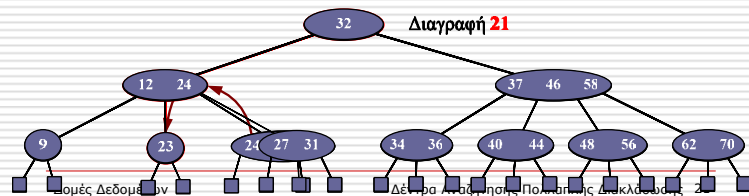


Δομές Δεδομένων

Δέντρα Αναζήτησης Πολλαπλής Διακλάδωσης 20

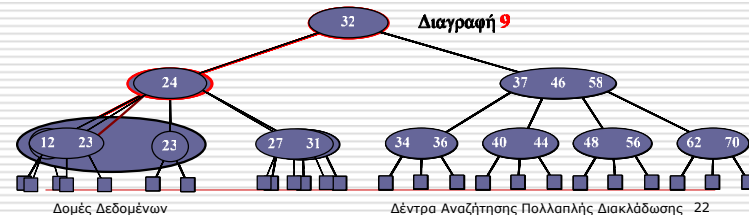
Διαγραφή με Δανεισμό

- Έλλειμμα (κόμβος με $< a$ παιδιά) \Rightarrow επαναζυγιστική πράξη.
- **Δανεισμός** (αντιστοιχεί σε περιστροφή):
 - Αν δεξιός ή αριστερός αδελφός (υπάρχει τουλάχιστον ένας) έχει $\geq a+1$ παιδιά:
 - Ακραίο στοιχείο «ανεβαίνει» σαν διαχωριστικό.
 - Διαχωριστικό συμπληρώνει ελλειμματικό κόμβο.
 - Τερματική πράξη.



Διαγραφή με Συγχώνευση

- Αν αδέρφια έχουν a παιδιά, **συγχώνευση** με αδελφό.
- Συγχώνευση (αντιστοιχεί σε αντιστροφή χρώματος):
 - Αδέρφια συγχωνεύονται σε έναν κόμβο.
 - Διαχωριστικό «κατεβαίνει» στο νέο κόμβο.
 - Στοιχεία νέου κόμβου: $a \leq 2(a-1) \leq b-2$
 $a-2$ (ελλειμματικός) + 1 (διαχωριστικό) + $a-1$ (αδελφός)



Διαγραφή

- Δανεισμός είναι **τερματική** πράξη.
- Συγχώνευση μπορεί να **μην είναι τερματική**:
 - Πιθανό έλλειμμα στον πατέρα.
 - Αν ρίζα ελλειμματική (1 παιδί), έχουμε νέα ρίζα και μείωση ύψους κατά 1.
- Χρόνος διαγραφής : $O(\log n)$
 - ≤ 1 δανεισμός και $O(\log n)$ διασπάσεις.
- «Προληπτική» επαναζύγιση στην «κάθοδο».
 - Επαναζύγιση σε ένα πέρασμα.
 - Σημαντικό όταν δομή σε σκληρό δίσκο.