



ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ

HASHING

ΣΑΛΤΟΓΙΑΝΝΗ ΑΘΑΝΑΣΙΑ

saltogiann@ceid.upatras.gr

ΔΟΜΕΣ ΔΕΔΟΜΕΝΩΝ

ΤΟ ΠΡΟΒΛΗΜΑ ΤΟΥ ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΥ

- Θέλουμε τα δεδομένα που διαθέτουμε να μπορούν να αποθηκευτούν σε κάποιο πίνακα ή άλλη δομή και να χωράνε στη μνήμη του υπολογιστή που χρησιμοποιούμε.
- Ένα προς ένα αντιστοίχιση στοιχείων - **Απευθείας διευθυνσιοδότηση**
- Η διαθέσιμη μνήμη του υπολογιστή δεν μπορεί να ανταποκριθεί στον τεράστιο όγκο δεδομένων.

ΛΥΣΗ: Η μέθοδος του κατακερματισμού

Προσπαθεί να δώσει λύση στο πρόβλημα έλλειψης θέσεων σε μία δομή ή στη μνήμη του υπολογιστή, αντιστοιχίζοντας θέσεις της δομής/μνήμης στα δεδομένα με βάση κάποια συνάρτηση, ώστε να αποφεύγονται οι συγκρούσεις.

ΟΡΙΣΜΟΣ

Έστω:

- **σύμπαν** $U = \{0, 1, \dots, N-1\}$, $|U| = N$
- **πίνακας κατακερματισμού** $T[0, 1, \dots, m-1]$, $|T| = m$
- **συνάρτηση κατακερματισμού** $h: U \rightarrow [0, 1, \dots, m-1]$, $|h| = m$

Το στοιχείο $x \in U$ θα αποθηκευτεί στην θέση $T[h(x)]$

ΕΙΔΗ ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΎ

■ Κατακερματισμός με αλυσίδες

- Σε κάθε θέση του πίνακα υπάρχουν συνδεδεμένες αλυσίδες και μπορούν να επεκταθούν δυναμικά
- Τεχνική κλειστής διεύθυνσης

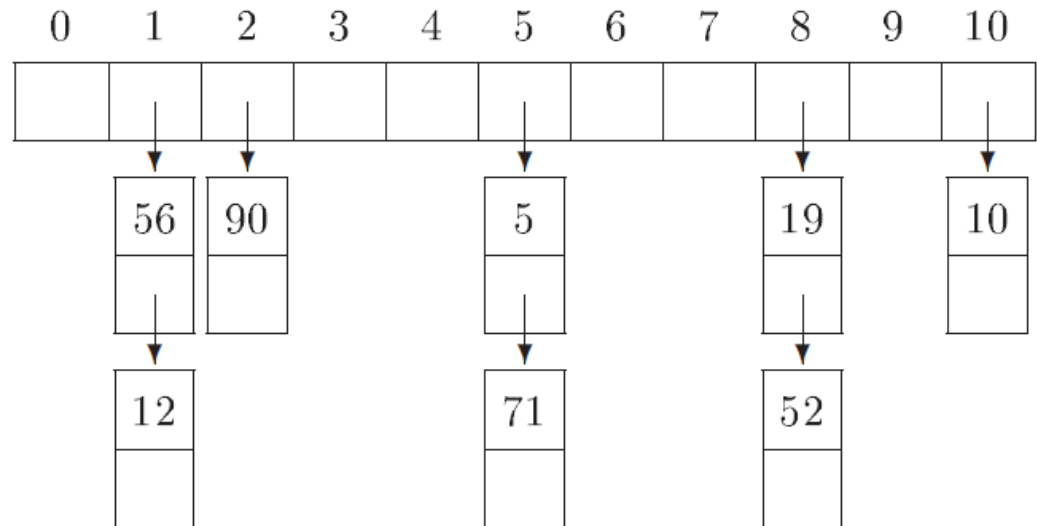
■ Κλειστός κατακερματισμός

- Δεν χρησιμοποιεί δείκτες για το χειρισμό των πινάκων
- Τεχνική ανοικτής διεύθυνσης
- Γραμμική Δοκιμή, Τετραγωνική Δοκιμή, Διπλός κατακερματισμός

ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ ΜΕ ΑΛΥΣΙΔΕΣ

- Στις αλυσίδες, υπάρχει μία δομή record όπου το πρώτο πεδίο χρησιμοποιείται για την αποθήκευση των δεδομένων ενώ το δεύτερο πεδίο, αποτελεί δείκτη προς το επόμενο ζευγάρι της αλυσίδας.

- π.χ. για $h(x) = x \bmod m$



- Search(x)

- Insert(x)

- Delete(x)

ΚΛΕΙΣΤΟΣ ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ

- Αντιμετωπίζει τις συγκρούσεις χωρίς να χρησιμοποιεί επιπλέον χώρο
- Το $x \in U$ αντιστοιχίζεται σε μία ακολουθία θέσεων του πίνακα κατακερματισμού
- Πιο σύνθετη συνάρτηση κατακερματισμού
 - $h(x,0), \dots, h(x,m-1)$
 - $l = 0, \dots, m-1 \rightarrow$ προσπάθεια εύρεσης κατάλληλης θέσης

ΚΛΕΙΣΤΟΣ ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ

■ ΓΡΑΜΜΙΚΗ ΔΟΚΙΜΗ

- εξετάζονται διαδοχικές θέσεις του πίνακα για την ανεύρεση ελεύθερης θέσης χρησιμοποιώντας συνάρτηση κατακερματισμού που δέχεται δύο ορίσματα, την τιμή έναρξης και το βήμα
- $h(x, i) = [h_1(x) + i] \bmod x, i = 0, 1, \dots, m - 1$
- $h_1(x) = x \bmod m$

Παράδειγμα

- ❖ $m=11$
- ❖ $S = \{52, 12, 71, 56, 5, 10, 19, 90\}$

	0	1	2	3	4	5	6	7	8	9	10
(α)		12				71			52		
(β)		12	56			71			52		
(γ)		12	56			71	5		52		10
(δ)		12	56			71	5		52	19	10
(ε)		12	56	90		71	5		52	19	10

ΚΛΕΙΣΤΟΣ ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ

■ ΤΕΤΡΑΓΩΝΙΚΗ ΔΟΚΙΜΗ

➤ Η συνάρτηση κατακερματισμού είναι της μορφής:

$$(h_1(x) + c_1i^2 + c_2i^2) \bmod m$$

- Το i δηλώνει τις προσπάθειες για την εύρεση κενής θέσης και ισχύει $i = 0, 1, 2 \dots$ και οι c_1, c_2 είναι σταθερές.
- Εξετάζεται η πρώτη θέση για $i = 0$ και έπειτα αναζητούνται θέσεις σε αποστάσεις ανάλογες του τετραγώνου του i .
- Με αυτόν τον τρόπο, αποφεύγεται να καταλαμβάνονται συνεχόμενα μεγάλα τμήματα του πίνακα A στον οποίο εισάγονται τα κλειδιά.
- Μεγαλύτερη απόδοση από την γραμμική δοκιμή

ΚΛΕΙΣΤΟΣ ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ

ΤΕΤΡΑΓΩΝΙΚΗ ΔΟΚΙΜΗ

Παράδειγμα

- ❖ $m=11$
- ❖ $S = \{52, 12, 71, 56, 5, 10, 19, 90\}$
- ❖ $c_1 = 1, c_2 = 2.$

	0	1	2	3	4	5	6	7	8	9	10
(a)		12				71			52		
(b)		12			56	71			52		
(g)	5	12			56	71			52		10
(d)	5	12	90	19	56	71			52		10

ΚΛΕΙΣΤΟΣ ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ

■ ΔΙΠΛΟΣ ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ

- Πιο αποτελεσματική μέθοδος από τις άλλες 2.
- $h(x, i) = (h_1(x) + ih_2(x)) \bmod m$
- $h_1, h_2 \rightarrow$ κλασικές συναρτήσεις κατακερματισμού
- Αν προκύψει νέα σύγκρουση, τότε υπολογίζεται η επόμενη θέση σε ίση απόσταση από τη δεύτερη θέση σύγκρουσης.
- Στην περίπτωση που το πηλίκο που προκύπτει είναι ίσο με 0, εξισώνεται με 1 ώστε αν κάποια κλειδιά συγκρούονται στην ίδια αρχική θέση, να μην συγκρουστούν σε επόμενη θέση.
- Το μέγεθος του πίνακα να είναι πρώτος αριθμός, για να προσπελαστούν όλες οι θέσεις του πίνακα.

ΚΛΕΙΣΤΟΣ ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ

Παράδειγμα

Ας εισάγουμε στην παρακάτω δομή με $m=11$ και $n=8$, τα στοιχεία $S = \{52, 12, 71, 56, 5, 10, 19, 90\}$ με $h_1(x) = x \bmod m$ και $h_2(x) = (x \div m) \bmod m$.

	0	1	2	3	4	5	6	7	8	9	10
(a)		12				71			52		
(b)		12				71	56		52		
(g)		12				71	56	5	52		10
(d)		12	90			71	56	5	52	19	10

ΙΔΑΝΙΚΟΣ ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ

- **ΙΔΑΝΙΚΟΣ ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ – perfect hashing**
- Μπορούμε κατά τη σχεδίαση μίας δομής να επιλέξουμε συνάρτηση κατακερματισμού που να αποφεύγει τις συγκρούσεις.
- Μία συνάρτηση $h:U = [0,1,\dots,N-1] \rightarrow [0,1,\dots,m-1]$ είναι ιδανική συνάρτηση κατακερματισμού για το $S \subseteq [0,1,\dots,N-1]$ αν για όλα τα $x, y \in S$ ισχύει $h(x) \neq h(y)$.

ΠΑΓΚΟΣΜΙΟΣ ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ

■ ΠΑΓΚΟΣΜΙΟΣ ΚΑΤΑΚΕΡΜΑΤΙΣΜΟΣ – universal hashing

- Μπορούμε να επιλέξουμε για συνάρτηση κατακερματισμού μέσα από μία κλάση ομοιόμορφων συναρτήσεων.
- Μία συλλογή συναρτήσεων $H = \{h : U = [0, 1, \dots, N-1] \rightarrow [0, 1, \dots, m-1]\}$ καλείται c -universal με $c \in \mathbb{R}$ αν για κάθε $x, y \in U, x \neq y$, ισχύει:

$$|\{h : h \in H, h(x) = h(y)\}| \leq \frac{c|H|}{m}$$