Κατακερματισμός

Ο κατακερματισμός γίνεται χρησιμοποιώντας την συνάρτηση κατακερματισμού που είναι η **θέση = τιμή mod N**, όπου N το μέγεθος του πίνακα.

Ετσι για παράδειγμα, μια τιμή 24 σε έναν πίνακα 22 θέσεων , θα πάρει την θέση 24 mod 22=2.

Αυτός ο τρόπος καρακερματισμού όμως μπορεί να δημιουργήσει κάποια πορβλήματα.
Αν υποθέσουμ ε ότι στον παραπάνω πίνακα , μετά το 24, έρχεται προς αποθήκευση η τιμή 2, αυτή θα έπρεπε να πάρει την θέση 2 mod 22=2, θέση όμως που είναι ήδη πιασμένη από την τιμή 24.
Αυτού του είδους τα προβλήματα ονομάζονται συγκρούσεις και υπάρχουν αρκετοί τρόποι για να ξεπεραστούν.

Ο πιο εύκολος είναι η **γραμμική δοκιμή**. Στην ουσία , σε αυτόν τον τρόπο, η κάθε τιμή , όταν η θέση που της αντιστοιχεί , είναι πιασμένη, αποθηκεύεται στην αμέσως επόμενη ελεύθερη θέση.
Ετσι στο παράδειγμά μας , η τιμή 2 θα αποθηκευθεί στην θέση 3 που είναι η αμέσως επόμενη ελεύθερη από την θέση 2.

Ο τύπος της είναι ο εξής.

Θέση = (τιμή mod N + i) mod N όπου **i** είναι οι φορές που έχει δημιουργηθεί σύγκρουση.

Ετσι για την τιμή 24 μιας και ακόμα δεν είχε υπάρξει σύγκρουση το i παίρνει την τιμή 0 και συνεπώς ο τύπος γίνεται Θέση = (τιμή mod N) mod N και από την αριθμητική των υπολοίπων γνωρίζουμε ότι όταν υπάρχουν 2 mod στην σειρά , το ένα είναι σαν να μην υπάρχει και συνεπώς καταλή γουμε στην αρχική συνάρτηση κατακερματισμού.

Ο δεύτερος τρόπος είναι η **τετραγωνική δοκιμή**. Σε αυτόν τον τρόπο, η κάθε τιμή εξετάζει όχι τις διαδοχικές θέσεις για να δει ποια είναι ελεύθερη, αλλά αυτές που βρίσκονται στα διαδοχικά τετράγωνα.
Ο τύπος της είναι **Θέση = (τιμή mod N + i^2) mod N.**

Για την πρώτη σύγκρουση η θέση αποθήκευσης είναι η ίδια αφού 1^2=1.

Αν υποθέσουμε όμως ότι έρχεται η τιμή 46 να μπει στον πίνακα, εφαρμόζοντας την γραμμική δοκιμή, ο τύπος θα έδινε για i=0 , θέση =2 που είναι πιασμένη από την τιμή 24, συνεχίζοντας για i=1 , θέση =3 που είναι πιασμένη από την τιμή 2 και τέλος **για i=2, θέση = 4** όπου και αποθηκεύεται η τιμή 46 μιας και η θέση 4 είναι ελεύθερη.

Εφαρμόζοντας όμως την τετραγωνική δοκιμή , για την ίδια περίπτωση, θα είχαμε πάλι για i=0, θέση =2, πιασμένη, για i=1, θέση =3 , πιασμένη , ενώ για **i=2, θέση = 6**, ελεύθερη άρα αποθήκευση.

Ο τρίτος τρόπος για την επίλυση συγκρούσεων είναι ο **διπλός κατακερματισμός**.

Σε αυτόν δίνεται μια δεύτερη συνάρτηση κατακερματισμού h2(x) η οποία έχει την μορφή **h2(x)=1+ x mod K** όπου K ένας πρώτος αριθμός και χ η τιμή κάθε φορά που θέλουμε να αποθηκεύσουμε.

Ετσι με διπλό κατακερματισμό, ο τύπος γίνεται **θέση = (τιμή mod N + i * (1+ τιμή mod K)) mod N**, όπου και πάλι το i εκφράζει τον αριθμό των συγκρούσεων που έχουν γίνει ήδη.

Στο παράδειγμά μας, η τιμή 24, μιας και δεν υπάρχει σύγκρουση και το i=0, θα πάρει πάλι την θέση 2.
Αν πάρουμε K =7 (απλά επιλέγουμε έναν τυχαίο πρώτο αριθμό για να συνεχιστεί το παράδειγμα)
η τιμή 2 ,αφού έχουμε **μία** σύγκρουση μιας και αρχικά εξετάζεται η θέση 2 που είναι πιασμένη, για **i=1** λοιπόν, θα πάρει την **θέση =(2 mod 22 + (1 +2 mod 7))mod 22= (2 + 3)mod 22=5**.

Ενώ η τιμή 46 ,μιας και θα έχουν γίνει ήδη **2** συγκρούσεις , θα πάρει την **θέση = (46 mod 22 + 2*(1+46 mod 7)) mod 22=**
(2 + 2* (1 +4))mod 22=(2+2*5)mod 22=12.

Αυτά σε γενικές γραμμές για κατακερματισμό σε πίνακα.

Υπάρχει και άλλη μία μέθοδος που αφορά λίστες όπου κάθε θέση του πίνακα χωράει περισσότερες από μία τιμές. Οπως είναι λογικό σε αυτόν τον τρόπο, απλά δημιουργούμε λίστες των τιμών σε κάθε θέση και συνεπώς δεν έχουμε φαινόμενα συγκρούσεων μιας και "όλοι οι καλοί χωράνε".

Τέλος ένα ακόμα πράγμα που πρέπει να γνωρίζουμε είναι ο ανακατακερματισμός, ο οποίος χρησιμοποιείται ώστε οι πίνακες που θα αποθηκεύονται οι τιμές να μην είναι ούτε πολύ γεμάτοι αλλά ούτε και πολύ άδειοι, έτσι ώστε όταν γίνεται η αναζήτηση να είναι πιο αποδοτικοί.

Στον ανακατακερματισμό, ξεκινάμε με έναν μικρό πίνακα, και μόλις αυτός γεμίσει σε ένα ποσοστό, συνήθως ο μισός, συντελεστής φόρτου α=1/2, διπλασσιάζουμε τον πίνακα και ξαναβάζουμε πάλι από την αρχή τις ήδη υπάρχοντες τιμές που υπήρχαν στον μικρό πίνακα ,πριν συνεχίσουμε με τις υπόλοιπες.
Ετσι για παράδειγμα, αν έχουμε έναν πίνακα 4 θέσεων, μόλις μπουν σε αυτόν 2 τιμές, πρέπει να τον διπλασιάσουμε, να βάλουμε στον νέο πίνακα αυτές τις 2 τιμές ξανά, ανάλογα με το που πρέπει να μπουν σύμφωνα με το μέγεθος του νέου πίνακα (8), και αφού οι τιμές γίνουν 4 , ξαναδιπλασιάζουμε και ξαναβάζουμε Αυτές τις 4 τιμές πάλι στον νέο πίνακα 8 θέσεων. Αυτή η διαδικασία συνεχίζεται μέχρι να μπουν όλες οι τιμές που μας δίνονται.
Ας δούμε τώρα στο διπλανό μέρος της σελίδας, μια άσκηση με την εκφώνησή της και πώς λύνεται χρησιμοποιώντας αυτά που είπαμε.

ΘΕΜΑ Γ. Κατακερματισμός (Hashing) (1.5 μοναδα)
Τι μορφή θα έχει ένας πίνακας κατακερματισμού στον οποίο εισάγουμε τα στοιχεία 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 αν ο πίνακας ξεκινά με 4 στοιχεία και χρησιμοποιείται ανακατακερματισμός όταν α=1/2. Θεωρείστε ότι για την επίλυση συγκρούσεων χρησιμοποιείται γραμμική δοκιμή.

Αρχικά δημιουργούμε τον αρχικό πίνακα 4 στοιχείων που μας λέει η εκφώνηση.

Θέση	0	1	2	3
Τιμή				

Θέση	0	1	2	3
Τιμή		1		

Τιμή 1

1 mod 4 =1 άρα θέση 1

Θέση	0	1	2	3
Τιμή		1	2	

Τιμή 2

2 mod 4 =2 άρα θέση 2

Στο σημείο αυτό βλέπουμε ότι το α, ο συντελεστής φόρτου (το πόσο γεμάτος δηλαδή είναι ο πίνακας) έχει γίνει 1/2 μιας και οι μισές θέσεις του είναι πιασμένες.
Έτσι πρέπει να διπλασιάσουμε τον πίνακα ως εξής.

Θέση	0	1	2	3	4	5	6	7
Τιμή								

Στον καινούργιο αυτόν πίνακα πρέπει να ξαναβάλουμε από την αρχή τις τιμές που υπήρχαν στον μικρότερο, χρησιμοποιώντας την σειρά με την οποία βρίσκονται στον μικρότερο.

Θέση	0	1	2	3	4	5	6	7
Τιμή		1						

Τιμή 1

1 mod 8 =1 άρα θέση 1

Θέση	0	1	2	3	4	5	6	7
Τιμή		1	2					

Τιμή 2

2 mod 8 =2 άρα θέση 2

Και συνεχίζουμε με τις επόμενες τιμές που μας δίνονται.

Θέση	0	1	2	3	4	5	6	7
Τιμή		1	2	3				

Τιμή 3

3 mod 8 =3 άρα θέση 3

Θέση	0	1	2	3	4	5	6	7
Τιμή		1	2	3	4			

Τιμή 4

4 mod 8 =4 άρα θέση 4

Πάλι οι μισές θέσεις του πίνακα είναι πιασμένες, άρα διπλασιάζουμε.

Θέση	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Τιμή																

Με τον ίδιο τρόπο μιας και δεν υπάρχουν συγκρούσεις , για τιμές μέχρι την τιμή 8, ο πίνακας γίνεται

Θέση	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Τιμή		1	2	3	4	5	6	7	8							

Ξαναδιπλασιάζουμε μιας και γέμισαν οι μισές θέσεις.

Θέση	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Τιμή																																

Και μιας και δεν υπάρχουν και πάλι συγκρούσεις, ο τελικός πίνακας που ζητείται είναι ως εξής:

Θέση	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Τιμή		1	2	3	4	5	6	7	8	9	10																					

Αυτή η άσκηση είναι από τις ευκολότερες γιατί δν έχει συγκρούσεις.. Αλλά είναι λίγο χρονοβόρα γιατί πρέπει κάθε φορά να διπλασιάζεις τον πίνακα και να ξαναβάζεις τις τιμές.

Καλή Επιτυχία.