

Tema Nr. 5: Căutarea în tabele de dispersie

Adresare deschisă, verificare pătratică

Timp alocat: 2 ore

Implementare

Se cere implementarea **corectă** și **eficientă** a operațiilor de *inserare* și *căutare* într-o tabelă de dispersie ce folosește *adresarea deschisă* cu *verificare pătratică*.

Informații utile și pseudo-cod găsiți în notițele de curs sau în carte (Cormen), în secțiunea *11.4 Open addressing*.

Noțiunile închis/deschis (closed/open) specifică dacă se obligă folosirea unei poziții sau structuri de date.

Hashing (se referă la tabela de dispersie (hash table))

- Open Hashing
 - Pe o anumită poziție se pot stoca mai multe elemente (ex: chaining)
- Closed Hashing
 - Se poate stoca doar un singur element pe o anumita pozitie (ex. linear/quadratic probing)

Addressing (se referă la poziția finală a unui element față de poziția inițială)

- Open Addressing (Adresare Deschisă)
 - Adresa finală (poziția finală) nu este complet determinată de către codul hash. Poziția depinde și de elementele care sunt deja în tabelă. (ex: linear/quadratic probing - verificare liniară/pătratică)
- Closed Addressing (Adresare Închisă)
 - Adresa finală este întotdeauna determinată the codul hash (poziția inițială calculată) și nu exista probing (ex: chaining)

Pentru această temă tabela de dispersie va avea o structură similară cu cea de mai jos:

```
typedef struct { int id; char name[30]; } Entry;
```

unde *adresa finală* în tabelă va fi calculată aplicând funcția de hashing pe câmpul *id* din structură. Câmpul *name* va fi folosit doar pentru demonstrarea corectitudinii operațiilor de căutare și ștergere și nu este necesară pentru evaluarea performanței (câmpul *name* va fi afișat în consolă dacă operația de căutare găsește *id*-ul respectiv, altfel afișați “negăsit”).

Cerințe minimele pentru notare

- Interpretați graficul și notați observațiile personale în antetul fișierului *main.cpp*, într-un comentariu bloc informativ.
- Pregătiți un exemplu pentru exemplificarea corectitudinii fiecărui algoritm implementat.
- Nu preluăm teme care nu sunt indentate și care nu sunt organizate în funcții (de exemplu nu preluăm teme unde tot codul este pus în main)
- ***Punctajele din barem se dau pentru rezolvarea corectă și completă a cerinței, calitatea interpretărilor din comentariul bloc și răspunsul corect dat de voi la întrebările puse de către profesor.***

Cerințe

1. Implementarea operației de inserare și căutare folosind structura de date cerută + *demo (dimensiune 10)* (5p)

Corectitudinea algoritmilor va trebui demonstrată pe date de intrare de dimensiuni mici (ex. 10).

2. Evaluarea operației de căutare pentru un singur factor de umplere 95% (2p)

! Înainte de a începe să lucrați pe partea de evaluare, asigurați-vă că aveți un **algoritm corect!**

Se cere evaluarea operației de *căutare* în tabele de dispersie cu adresare deschisă și verificare pătratică, în cazul **mediu statistic** (nu uitați să repetați măsurătorile de 5 ori). Pentru a obține evaluarea, trebuie să:

1. Alegeți N , dimensiunea tabelului, un număr prim în jur de 10000 (e.g. 9973, sau 10007);
2. Pentru fiecare din următoarele valori pentru factorul de umplere $\alpha = 0.95$:
 - a. Inserați în tabela n elemente aleator, astfel încât să ajungeți la valoarea lui α ($\alpha = n/N$)
 - b. Căutați aleator, în fiecare caz, m elemente ($m \sim 3000$), astfel încât aproximativ jumătate din elemente să fie găsite, iar restul să nu fie găsite (în tabelă). Asigurați-vă că elementele găsite sunt generate uniform, i.e. să căutați elemente care au fost introduse la momente diferite, cu probabilitate egală (există mai multe moduri în care se poate garanta acest lucru)
 - c. Numărați operațiile efectuate de procedura de căutare (i.e. numărul de celule accesate)
 - d. Atenție la valorile pe care le căutați, să fie într-o ordine aleatoare din cele introduse. Dacă sunt primele 1500 adăugate, implicit efortul mediu găsite va fi 1.

3. Generați un tabel de forma:

Factor de umplere	Efort mediu <i>gasite</i>	Efort maxim <i>gasite</i>	Efort mediu <i>ne-gasite</i>	Efort maxim <i>ne-gasite</i>
0.95

$Efort\ mediu = efort_total / nr_elemente$

$Efort\ maxim = număr\ maxim\ de\ accese\ efectuat\ de\ o\ operație\ de\ căutare$

3. Completarea evaluării pentru toți factorii de umplere (2p)

Respectând cerințele de la punctul 2 cu $\alpha \in \{0.8, 0.85, 0.9, 0.95, 0.99\}$, generați un tabel de forma:

Factor de umplere	Efort mediu <i>gasite</i>	Efort maxim <i>gasite</i>	Efort mediu <i>ne-gasite</i>	Efort maxim <i>ne-gasite</i>
0.8				
0.85				
...	

4. Implementare operației de ștergere într-o tabelă de dispersie, *demo* (dimensiune 10) și evaluarea operației de căutare după ștergerea unor elemente (1p)

Pentru evaluarea operației de de căutare după ștergere, umpleți tabela de dispersie până la factor de umplere 0.99. Stergeți elemente din tabela până ajungeți la factor umplere 0.8, după care căutați m elemente aleator ($m \sim 3000$), astfel încat aproximativ jumătate din elemente sa fie *găsite*, iar restul sa nu fie găsite (în tabela). Măsurați efortul necesar pentru *căutare* și adăugați în tabelul generat anterior.