Vyhľadávanie

Vyhľadávanie

Vstupy:

- N prvkov identifikovateľných kľúčom
- kľúč K, ktorý charakterizuje prvok, ktorý chceme nájsť

Výstupy:

- Úspech : prvok sa v zadanej množine našiel
- Neúspech: prvok sa nenašiel

Rozdelenie vyhľadávaní

- · vnútorné, vonkajšie
- statické, dynamické
- · uporiadaná, neusporiadaná postupnosť
- lineárne, binárne, pomocou rozptylovej funkcie

Jednoduché vyhľadávanie

Výstup: pozícia nájdeného prvku v poli (ak sa prvok nenájde, tak sa vráti 0)

```
Procedure SEARCH(POST:pole, K: klūč, var KDE: integer)
VAR NAJDENY := BOOLEAN;
I : 1. VELKOST;
BEGIN
NAJDENY := FALSE;
I := 1;
WHILE((I <= VELKOST) AND (NOT NAJDENY)) DO
IF K = POST[I].KLUC THEN
BEGIN
KDE := I;
NAJDENY := TRUE;
END
ELSE
I ++;
IF NOT NAJDENY THEN
KDE := 0;
END:
```

Jednoduché vyhľadávanie - lineárne

```
// vracia true akk existuje integer I také,
// že arr[i] == target a 0 <= i < arr.length
boolean linearSearch(int[] arr, int target)
{
    int i = 0;
    while (i < arr.length) {
        if (arr[i] == target) {
            return true;
        } // if
        ++i;
    }
    return false;
```

Jednoduché vyhľadávanie - lineárne

```
type IntList is array(1..Max) of Integer;
function Search(List: IntList; Key: Integer)
return Integer
begin
for I in List.Range loop
    if List(I) = Key then
    return I;
    end if;
end loop;
return 0;
end Search;

popodmienka:
(List (I) = Key) \vee
((I = 0) \wedge \forall \forall I(1 \leq J \leq Max | (List (J) \neq Key))
```

Jednoduché vyhľadávanie – lineárne so zarážkou

```
type IntList is array(0..Max) of Integer;
```

```
function Search(List: IntList; Key: Integer) return Integer

I: Integer: := Max; begin
List(0) := Key; while List(1) \neq Key loop
I:= I - 1; end loop; return I; end Search; invariant cyklu:

\forall J(I < J \le Max \mid List (J) \ne Key)
```

Binárne vyhľadávanie

- · Vstupom je usporiadané pole.
- Algoritmus porovná prvok nachádzajúci sa v strede poľa so zadaným kľučom. Ak sa zhoduje, vráti jeho pozíciu. Ak sa nezhoduje, rozdelí pole na 2 polovice a pokračuje rovnakým hľadaním v tej polovici v ktorej sa prvok nachádza.

```
· Rekurzívna verzia algoritmu:
```

```
BinarySearch(A[0..N-1], value, low, high)

{
    if (high < low)
        return not_found
    mid = (low + high) / 2
    if (A[mid] > value)
        return BinarySearch(A, value, low, mid-1)
    else if (A[mid] < value)
    return BinarySearch(A, value, mid+1, high)
    else
    return mid
}
```

Binárne vyhľadávanie s cyklom

```
function Search(List: IntList; Key: Integer)
return Integer
    Low: Integer := List'First;
    High: Integer := List'Last;
    Mid: Integer;
    begin
    loop
    Mid := (Low + High) / 2;
    if Key = A[Mid] then return Mid;
        elsir Key < A[Mid] then High := Mid - 1;
    else Low := Mid + 1;
    end if;
    if Low > High then return 0;
end loop;
end Search;
```

Binárne vyhľadávanie

Nájdi prvok 92

```
11 14 16 27 31 35 39 39 43 49 <sup>56</sup> 50 64 72 74 80 85 92 92 97 97 97
```

Usporadúvanie

Usporadúvanie

```
    Postupnosť: a<sub>1</sub>, a<sub>2</sub>, a<sub>3</sub>...a<sub>n</sub>
    Položka:

            a<sub>i</sub>
            Kľúč k<sub>i</sub> položky a<sub>i</sub> k(a<sub>i</sub>)

    Usporiadanie: binárna relácia
    Úlohou je preusporiadať všetky položky postupnosti tak, aby platilo:
    K1 <= K2 <= K3 <= ... <= Kn</li>
```

Usporadúvanie

- Nech K je nekonečná lineárne usporiadaná množina (ďalej skratka LUM), t.j.
- množina, na ktorej je definovaná relácia < taká, že sú splnené tieto 2 podmienky:
 - zákon trichotómie
 - Pre ľubovoľné dva prvky K1, K2 ∈ K platí práve jedna z relácií : K1 < K2, K1 = K2. K1 > K2.
 - tranzitívny zákon
 - Pre ľubovoľné tri prvky K1, K2, K3 ∈ K platí: Ak K1 < K2 a K2 < K3, tak K1 < K3.
- Pre ľubovoľné K1, K2 ∈ K budeme písať, že K1 ≤ K2 akk K1 < K2 alebo K1 = K2.

Stabilný algoritmus

- · Usporadúvací algoritmus je stabilný, ak vždy zachová originálne poradie elementov s rovnakými kľúčmi
- · Ak elementy s rovnakými kľúčmi sú neodlíšiteľné, tak nie je potrebné sa zaoberať stabilitou algoritmu (napr. ak kľúčom je samotný element)
- · Zachovať originálne poradie elementov je dôležité napr. pri viacnásobnom usporiadaní - najprv podľa priezviska a potom podľa mena.

- Príklad dvojice (kľúč, element):
- (4,5)(2,7)(2,3)(5,6)
- Dve možné usporiadania:
 - (2, 7) (2, 3) (4, 5) (5, 6) zachované poradie elementov s kľúčmi 2 stabilné usporiadanie

Stabilný algoritmus

- (2, 3) (2, 7) (4, 5) (5, 6) zmenené poradie elementov s kľúčmi 2 nestabilné
- Príklad na viacnásobné usporiadanie dvojice (kľúč 1, kľúč 2):
- (4,5)(2,7)(2,3)(4,6)
- Usporiadanie najprv podľa kľúča 2, potom podľa kľúča 1:
 - (2, 3) (4, 5) (4, 6) (2, 7) podľa kľúča 2
 (2, 3) (2, 7) (4, 5) (4, 6) podľa kľúča 1
- Usporiadanie najprv podľa kľúča 1, potom podľa kľúča 2:

 - (2, 7) (2, 3) (4, 5) (4, 6) podľa kľúča 1
 (2, 3) (4, 5) (4, 6) (2, 7) podľa kľúča 2 narušené poradie
- Pre zachovanie stability viacnásobného usporadúvania je potrebné usporadúvať postupne podľa kľúčov so zvyšujúcou sa prioritou.

Usporadúvanie

- Nech D je nekonečná množina a nech je na nej definovaná funkcia $k : D \rightarrow K$.
- Definícia (problém usporadúvania). Nech je daná postupnosť a1, ..., an, kde $n \in N$; ai \in D. Potom všeobecný problém usporadúvania tkvie v určení takej permutácie π čísel 1,...,n, že platí $k(a_{\pi(1)}) \le k(a_{\pi(2)}) \le ... \le k(a_{\pi(n)})$
- D množina údajov
- k funkcia usporiadania, daná zvyčajne explicitne pre každý prvok ako jeho zložka tzv. kľúč.
- Definícia. Permutácia n-prvkovej množiny je ľubovoľné usporiadanie prvkov tejto množiny do postupnosti.

Stabilný algoritmus

- Každý nestabilný algoritmus sa dá implementovať ako stabilný tým, že sa zapamätá originálne poradie elementov a pri zhodných kľúčoch sa berie do úvahy toto poradie
- Viacnásobné usporiadanie je možné obísť vytvorením jedného kľúča, ktorý je zložený z Viacnásobné možné obísť sekundárného, átď. primárneho, usporiadania
 - Takéto úpravy nestabilných algoritmov majú negatívny vplyv na výpočtovú zložitosť.

Porovnávací algoritmus

- · usporadúvací algoritmus, ktorý prechádza vstupné kľúče a na základe operácie porovnávania rozhoduje, ktorý z dvoch elementov sa má v usporiadanom poli objaviť ako prvý.
- Operácia porovnávania musí mať tieto vlastnosti:
 - Ak $a \le b$ a $b \le c$, tak potom $a \le c$
 - Pre všetky a a b, buď $a \le b$ alebo $b \le a$
- Základným limitom je dolné ohraničenie počtu porovnávania Ω(n log n), ktoré je potrebné na usporiadanie postupností. Preto aj tie najlepšie algoritmy usporadúvania založené na porovnávaní, majú priemernú časovú zložitosť O (n log n) – na rozdiel od neporovnávacích algoritmov, kde sa môže dosiahnuť časová zložitosť aj O (n).

Výhody porovnávacích algoritmov

- Použiteľné pre rôzne dátové typy
- Jednoduchá implementácia porovnávania n-tíc v lexikografickej postupnosti
- Reverzná funkcia porovnávania = reverzne usporiadaná postupnosť

Najznámejšie algoritmy usporadúvania

- založené na porovnávaní:
 - · usporadúvanie výberom,
 - vkladaním,
 - výmenou,
 zlučovaním
 - quicksort.
 - · heapsort, ...
- · neporovnávacie algoritmy:
 - · radix sort,
 - counting sort,
 - · bucket sort, ...

Usporadúvanie vkladaním

- algoritmus, ktorý realizuje usporadúvanie priamym vkladaním.
- vychádza z predpokladu, že do už usporiadanej postupnosti sa na správne miesto vloží ďalší prvok.
- ak sa miesto, na ktoré sa nový prvok vkladá, zisťuje binárnym vyhľadávaním, hovoríme o usporadúvaní binárnym vkladaním

Usporiadanie vkladaním

- Príklad:
 - usporiadajte vkladaním pole 6 4 5 2 3 1 7

```
• 1. krok 6 | 452317 -> 46 | 52317
```

• 2. krok 46 | 52317 -> 456 | 2317

• 3. krok 456 | 2317 -> 2456 | 317 • 4. krok 2456 | 317 -> 23456 | 17

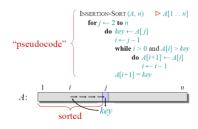
• 5. krok 23456 | 17 -> 123456 | 7

• 6. krok 123456 | **7** -> 123456**7** |

• 7. krok 1234567 -> 1234567

Usporiadanie vkladaním

- · Časová zložitosť závisí od vstupného poľa
 - pre takmer usporiadané vstupné pole algoritmus prebehne rýchlo – vtedy sa akoby vymieňali len dva susedné prvky (najpravejší z usporiadanej časti s najľavejším z neusporiadanej časti) a nedochádza tak k posunu ostatných prvkov.



Usporadúvanie vkladaním

```
Procedure InsertionSort(var A: pole)
Var i, j, x : Integer;
  for i:= 2 to Dlzka(A) do begin
     x := A[i];
     a[0] := x;
     j := i - 1;
     while x < A[j] do
                a[j+1] := a[j];
                j := j - 1;
          end;
     a[j+1] := x;
```

Insertion Sort

```
INSERTION-SORT(A)
1. for j = 2 to length[A]
2.
       do key \leftarrow A[i]
          //insert A[j] to sorted sequence A[1..j-1]
3.
4.
          while i > 0 and A[i] > key
               do A[i+1] \leftarrow A[i] //move A[i] one position right
7.
          A[i+1] \leftarrow key
```

správnosť algoritmu Insertion Sort

- · invariant cyklu
 - na začiatku každej iterácie obsahuje podpole A[1..j-1] pôvodné hodnoty z A[1..j-1] ale v usporiadanom poradí.
- dôkaz:
 - inicializácia : *j*=2, A[1..*j*-1]=A[1..1]=A[1], je usporiadané.
 - udržiavanie: každý krok cyklu udržiava platnosť invariantu.
 - ukončenie: j=n+1, takže A[1..j-1]=A[1..n] je usporiadané.

analýza algoritmu Insertion Sort

```
INSERTION-SORT(A)
                                                                       cena počet opakovaní
       for j = 2 to length[A]
                                                                        C_1
2.
          do key \leftarrow A[j]
3.
            //insert A[j] to sorted sequence A[1..j-1]
                                                                                    n-1
4.
            i \leftarrow j-1
                                                                         c_4
                                                                                    n-1
                                                                                    \begin{array}{l} \sum_{j=2}^{n} t_{j} \\ \sum_{j=2}^{n} (t_{j}-1) \\ \sum_{j=2}^{n} (t_{j}-1) \end{array}
            while i >0 and A[i]>key
                                                                         c_5
6.
                do A[i+1] \leftarrow A[i]
                   i \leftarrow i-1
            A[i+1] \leftarrow key
                                                                                    n-1
(t_j udáva, koľkokrát sa vykoná test cyklu while v riadku 5 pre danú hodnotu j)
Celková cena T(n) = suma cena × počet opakovaní pre každý riadok
                  =c_1n+c_2(n-1)+c_4(n-1)+c_5\sum_{j=2}^nt_j+c_6\sum_{j=2}^n(t_{j}-1)+c_7\sum_{j=2}^n(t_{j}-1)+c_8(n-1)
```

analýza algoritmu Insertion Sort

- · cena v najlepšom prípade: prvky sú už usporiadané
 - t=1, a riadky 6 a 7 sa vykonajú 0 krát
 - $T(n) = c_1 n + c_2 (n-1) + c_4 (n-1) + c_5 (n-1) + c_8 (n-1)$ $=(c_1+c_2+c_4+c_5+c_8)n-(c_2+c_4+c_5+c_8)=cn+c'$
- · cena v najhoršom prípade: prvky sú už usporiadané, ale v

 - $\begin{array}{ll} SO \ \sum_{j \in 2^n} t_j = \sum_{j \in 2^n} j = n(n+1)/2 \cdot 1, \ a \ \sum_{j \in 2^n} (t_j 1) = \sum_{j \in 2^n} (j-1) = n(n-1)/2, \ a \\ \ \Pi(n) = c_1 n + c_2(n-1) + c_4(n-1) + c_5(n(n+1)/2 \cdot 1) + + c_6(n(n-1)/2 \cdot 1) + c_5(n(n-1)/2) + c_6(n-1) = (|c_1 + c_2 + c_3 + c_3)/2 \cdot |c_2 + c_3 + c_3$
- · cena v priemernom prípade: čísla sú náhodné
 - v priemere, $t_j = j/2$. T(n) bude aj v priemernom prípade stále rádu n^2 , rovnako ako v najhoršom prípade.

časová zložitosť algoritmu Insertion Sort

- · časová zložitosť:
 - v najlepšom prípade: T(n) = Θ(n),
 - v najhoršom prípade: T(n) = Θ(n²),
 - v priemernom prípade: T(n) = Θ(n²)
- · je to rýchly algoritmus?
 - pre malé n celkom prijateľný
 - pre veľké n vonkoncom nie.

Usporadúvanie výmenou - bublinkové (bubble)

- bubble sort usporadúva prvky priamou výmenou
- je implementačne jednoduchý, ale neefektívny
- pri usporadúvaní porovnáva dva susedné prvky a ak nie sú v správnom poradí, vymenia sa
- procedúra sa opakuje, až kým nie sú potrebné žiadne výmeny

Usporadúvanie výmenou - bublinkové (bubble)

Príklad:

```
- Usporiadajte bublinkovou výmenou pole 5 1 4 2 8

• 1. fáza

» 51 4 2 8 -> 15 4 2 8

» 15 4 2 8 -> 145 2 8

» 145 28 -> 14 2 5 8

» 1 4 2 5 8 -> 1 4 2 5 8

» 1 4 2 5 8 -> 1 2 4 5 8

» 1 4 2 5 8 -> 1 2 4 5 8

» 1 2 4 5 8 -> 1 2 4 5 8

» 1 2 4 5 8 -> 1 2 4 5 8

» 1 2 4 5 8 -> 1 2 4 5 8

» 1 2 4 5 8 -> 1 2 4 5 8
```

Usporadúvanie výmenou - bublinkové (bubble)

Príklad:

```
- na začiatku sme mali pole 5 1 4 2 8
• 3. fáza

» 12 4 5 8 -> 12 4 5 8

» 12 4 5 8 -> 12 4 5 8

» 12 4 5 8 -> 12 4 5 8

» 12 4 5 8 -> 12 4 5 8
```

- na konci máme usporiadané pole 12458

Bublinkové usporadúvanie

```
Procedure BubbleSort(var A: pole)
Var i, j, t : integer;
Begin
  for i := Dlzka(A) downto 1 do
    for j := 1 to Dlzka(A)-1 do
        if A[j] > A[j+1] then
        begin
            T := A[j];
            A[j] := A[j+1];
            A[j+1] := T;
        end;
```

Bublinkové usporadúvanie

časová zložitosť bublinkového usporadúvania

- 1 prechod = presun najväčšieho prvku na koniec
- i-tý prechod: n-i+1 operácií
 - čac

$$(n-1) + (n-2) + ... + 1 = (n-1) n / 2 = O(n^2)$$

- ktorý prípad je najlepší?
- ktorý prípad je najhorší?

Usporadúvanie výberom

- algoritmus realizuje usporadúvanie priamym výberom
- vychádza z prepokladu, že najmenší prvok môžeme zaradiť priamo na začiatok vstupného poľa, najmenší prvok zo zvyšku poľa zase na jeho začiatok atď.
- podľa toho, či usporadúva prvky vzostupne/zostupne, sa môže označovať ako MinSort/MaxSort

Usporadúvanie výberom

Príklad:

- usporiadajte výberom pole 6 4 5 2 3 1 7

```
• 1. krok  | 6452317 -> 1 | 645237

• 2. krok  | 1 | 645237 -> 12 | 64537

• 3. krok  | 12 | 64537 -> 123 | 6457

• 4. krok  | 123 | 6457 -> 1234 | 657

• 5. krok  | 1234 | 657 -> 12345 | 67

• 6. krok  | 12345 | 67 -> 123456 | 7

• 7. krok  | 123456 | 7 -> 123456 | 7
```

 Miera usporiadanosti vstupného poľa nemá vplyv na časovú zložitosť – vždy sa vykoná maximálny počet krokov.

Usporadúvanie výberom

```
procedure SelectionSort(var A: pole)
var i, j, t : integer;
begin
  for i:= 1 to Dlzka(A) - 1 do
    for j:= Dlzka(A) downto i+1 do
        if A[i] > A[j] then
        begin
            T := A[i];
            A[i] := A[j];
            A[j] := T;
    end;
```

Donald Shell

1.3.1924-

1959 PhD Uni of Cincinnati

Shell, D.L. (1959). "A high-speed sorting procedure".
Communications of the ACM 2 (7): 30–32.



Shellovo usporadúvanie

- algoritmus, ktorý realizuje usporadúvanie priamym vkladaním so zmenšovaním prírastku
- je to vlastne zlepšenie usporadúvania vkladaním a bublinkového
- táto metóda je jedna z najrýchlejších pre usporiadanie menších postupností (menej ako 1000 prvkov)

Shellovo usporadúvanie

- Príklad, prírastky n = n/2, atď (pôvodný návrh Shella):
 - Usporiadajte pole 6 4 5 2 8 3 1 7 pomocou algoritmu shell sort, n = 8/2 = 4

```
    1. krok, krokovanie 4, vyznačené čísla sa usporiadajú vkladaním
        » 64 52 83 17 -> 64 52 83 17
        » 64 52 83 17 -> 63 52 84 17
        » 63 52 84 17 -> 63 52 84 57
        » 63 52 84 17 -> 63 12 84 57
        » 63 12 84 57 -> 63 12 84 57
        » 63 12 84 57 -> 13 52 64 87
        » 13 52 64 87 -> 12 53 64 87
        » 13 52 64 87 -> 12 53 64 87
    3. krok, krokovanie 1
        » 12 53 64 87 -> 12 34 56 7 8
```

Shellovo usporadúvanie

```
procedure ShellSort(var f: pole)
var i, j, h, v, N : integer;
begin
N := Dlzka(f);
h := l;
repeat // priprav prirastky podla Knutha
h := (3 * h) + 1;
until h > N;

repeat
h := (h div 3);
for i:= (h + 1) to N do begin
v := f[i];
j := i;
while ((j > h) and (f[j-h] > v )) do begin
f[j] := f[j - h];
end;
end;
until h = 1;
end
```

Shellovo usporadúvanie

- používa postupnosť prírastkov h₁, h₂, ..., h_t
- môže byť ľubovoľná postupnosť, len musí byť
 h₁ = 1 a h₁ < h₂ < ... < h_t
- rôzne voľby postupnosti prírastkov vedú k rôzne efektívnym verziám algoritmu.

Shellovo usporadúvanie

```
urči postupnosť prírastkov h_1, h_2, ..., h_t urob t prechodov cez postupnosť prvkov prechod 1: h_t – usporiadaná postupnosť, t.j. pre \forall i \ a[i] \leq a[i+h_t] prechod 2: h_{t-1} – usporiadaná postupnosť, t.j. pre \forall i \ a[i] \leq a[i+h_{t-1}] ... prechod t: h_1 – usporiadaná postupnosť, t.j. pre \forall i \ a[i] \leq a[i+h_1]
```

 v každom prechode dosiahni h_k – usporiadanie pomocou usporadúvania vkladaním

Shellovo usporadúvanie

- po každom prechode pri nejakom prírastku h_k, pre všetky i máme a[i] ≤ a [i + h_k] t.j. všetky prvky umiestnené od seba o h_k miest sú usporiadané.
- posledný prechod sa robí s prírastkom h₁ = 1
 ⇒ pre ∀i a[i] ≤ a[i + 1]
 ⇒ postupnosť a[] je usporiadaná.

Shellovo usporadúvanie: prírastky

- Shell: \[\ln/2 \rd \, \ln/2^2 \rd \, \ln/2^3 \rd \, ..., 1 alebo (ešte horšie) postupnosť prírastkov mocniny 2: O(n²)
- Knuth: 1, 4, 13, 40, 121, 364,1093, 3280, 9841,... t.j.
 h₁ = 1, h_{i+1} = 3*h_i + 1
- Knuth: blízko O(n log² n) a O(n¹.25)
- Hibbard: 1, 3, 7,...2^{k-1}: O(n^{3/2})
- Sedgewick: 1, 8, 23, 77, 281, 1073, 4193, 16577..., (4ⁱ⁺¹ + 3·2ⁱ + 1) pre i > 0, má byť lepšia než Knuth
- Pratt: log²n prírastkov 2ⁱ3^j < n/2

Shellovo usporadúvanie

- najlepší prípad: postupnosť je už usporiadaná
 bude treba menej porovnaní
- najhorší prípad (pre postupnosť prírastkov podľa Pratta): O(n log² n)
- priemerný prípad (pre postupnosť prírastkov podľa Pratta): Θ(n log² n)

jedna varianta inšpirovaná Shellovým usporadúvaním: Shaker-sort*

```
h-utrasenie postupnosti a[1], ..., a[n]:
    porovnávanie a prípadná výmena týchto dvojíc prvkov v uvedenom poradí:
    (1; 1 + h), (2; 2 + h),..., (N - h - 1; N - 1), (N - h; N),
    (N - h - 1; N - 1), (N - h - 2; N - 2),..., (2; 2 + h); (1; 1 + h)
    všimnime si, že medzi prvkami každej dvojice je rovnaká vzdialenosť h.

usporadúvanie utrasením:
    urči postupnosť prírastkov h<sub>1</sub>, h<sub>2</sub>, ..., h<sub>1</sub>
    urob t prechodo vez postupnosť prvkov
    prechod : h<sub>2</sub> - utrasenie
    prechod : h<sub>2</sub> - utrasenie
    prechod : h<sub>2</sub> - utrasenie
    usporiadaj vkladaním // alebo opakuj 1-utrasenie dovtedy, kým nie je postupnosť usporiadaná
*Incerpi, J. and R. Sedgewick, Practical Variations on Shellsort, Inform. Process. Lett. 26 (1987), 37-43.
```