ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν Πανεπιστήμιον Αθηνών
ΙΔΡΥΘΕΝ ΤΟ 1837

Τμήμα Πληροφορικής και Τηλεπικοινωνιών • Εισαγωγή στον Προγραμματισμό

Περιεχόμενα

Κεφάλαιο 11: Ταξινόμηση και αναζήτηση

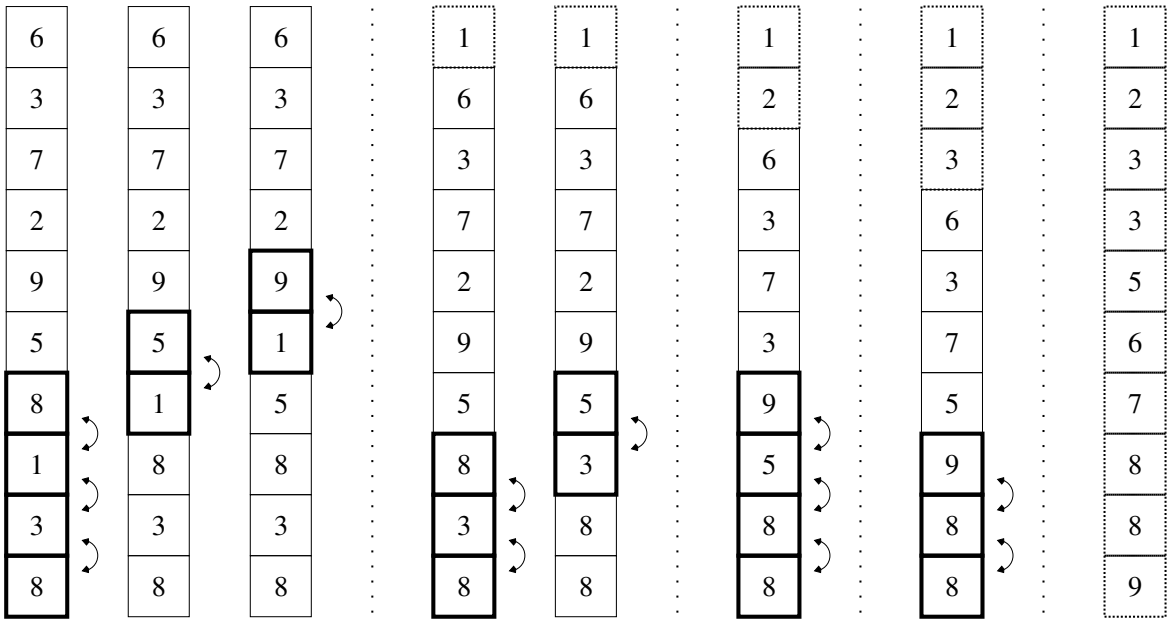
Ταξινόμηση πινάκων

- Υπάρχουν διάφορες μέθοδοι, άλλες λιγότερο, άλλες περισσότερο αποδοτικές, για την ταξινόμηση των στοιχείων ενός πίνακα.
- Η ταξινόμηση ενός πίνακα έχει νόημα όταν υπάρχει μία σχέση διάταξης στο σύνολο από το οποίο παίρνουν τιμές τα στοιχεία του πίνακα. Για αριθμούς (ακέραιους ή κινητής υποδιαστολής) έχουμε την αριθμητική διάταξη, για συμβολοσειρές την αλφαβητική. Για άλλους τύπους δεδομένων, πρέπει να έχει ορισθεί σαφώς η σχέση διάταξης. Για παράδειγμα, διάταξη μεταξύ δομών μπορεί να οριστεί με βάση τη διάταξη ως προς ένα μέλος-κλειδί της δομής (αριθμητικό ή συμβολοσειρά).
- Στη συνέχεια, οι αλγόριθμοι, τα παραδείγματα και τα προγράμματα που παρουσιάζονται στοχεύουν στην ταξινόμηση πινάκων, σε αύξουσα σειρά, με στοιχεία που είναι πραγματικοί αριθμοί διπλής ακρίβειας.
- Οι μέθοδοι ταξινόμησης μπορούν να προσαρμοσθούν εύκολα και για περιπτώσεις άλλων μονοδιάστατων δομών δεδομένων, για παράδειγμα συνδεδεμένων λιστών.
- Ένα θέμα που έχει γενικότερο ενδιαφέρον στους αλγορίθμους είναι ο χρόνος που χρειάζεται για την εκτέλεσή τους, ανάλογα με το μέγεθος της εισόδου τους. Αυτό

εκφράζεται με την πολυπλοκότητα χρόνου τους. Οι μέθοδοι ταξινόμησης έχουν διάφορες πολυπλοκότητες χρόνου, οι οποίες αντανακλούν την αποδοτικότητά τους.

- Για να εκφράσουμε την πολυπλοκότητα ενός αλγορίθμου, συνήθως χρησιμοποιούμε τον συμβολισμό O. Όταν η είσοδος ενός αλγορίθμου έχει μέγεθος n (για παράδειγμα το πλήθος των στοιχείων ενός πίνακα που θέλουμε να ταξινομήσουμε), λέγοντας ότι η πολυπλοκότητα χρόνου του αλγορίθμου είναι $O(f(n))$, εννοούμε ότι ο χρόνος εκτέλεσής του, $T(n)$, δεν έχει μεγαλύτερο ρυθμό αύξησης από αυτόν της συνάρτησης $f(n)$, όσο αυξάνει το n. Δηλαδή υπάρχουν C και n_0 τέτοια ώστε $T(n) < C \cdot f(n)$ για κάθε $n > n_0$.
- Για παράδειγμα, η πολυπλοκότητα του αλγορίθμου για την εύρεση της μέσης τιμής n αριθμών είναι $O(n)$, ενώ η πολυπλοκότητα του αλγορίθμου κατασκευής ενός μαγικού τετραγώνου $n \times n$ είναι $O(n^2)$.
- Σε ορισμένες περιπτώσεις, η χρονική απόδοση ενός αλγορίθμου για είσοδο μεγέθους n εξαρτάται και από το ποια είναι η συγκεκριμένη είσοδος. Στις περιπτώσεις αυτές, μπορούμε να αναφερόμαστε στην πολυπλοκότητα χείριστης περίπτωσης (για την πιο "δύσκολη" είσοδο) ή στη μέση πολυπλοκότητα (μέση τιμή απόδοσης για όλες τις πιθανές εισόδους).
- *Η μέθοδος της φυσαλίδας*

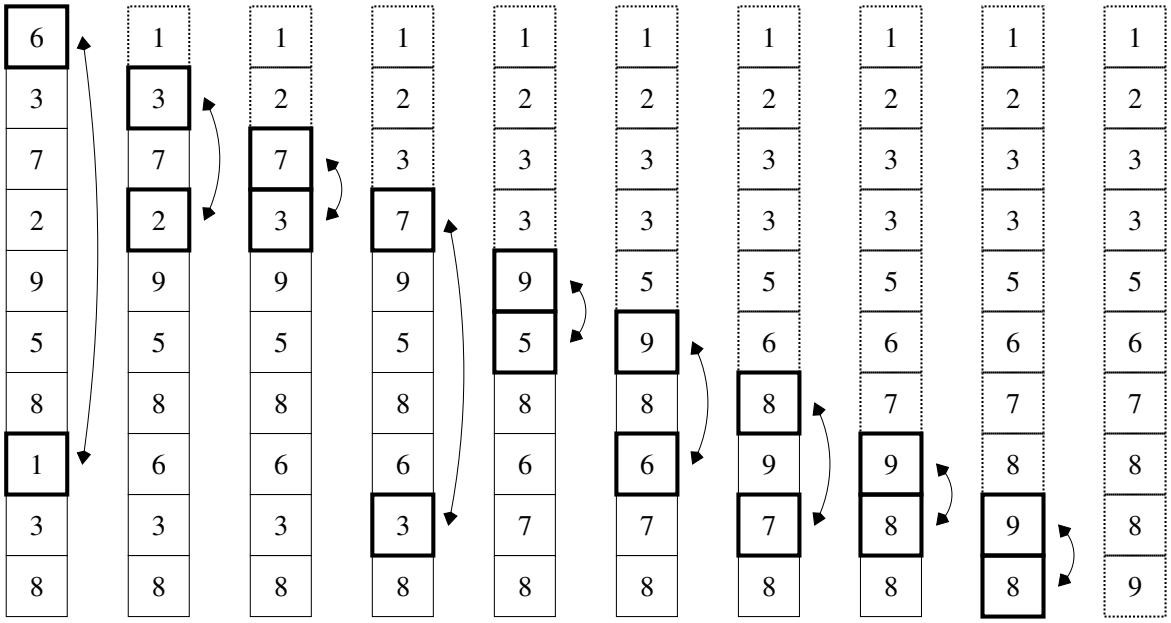
  Σύγκρινε ζευγάρια διαδοχικών στοιχείων, από κάτω προς τα επάνω, και όταν βρίσκεις δύο που δεν είναι στη σωστή σειρά, αντιμετάθεσέ τα. Μετά από το πρώτο πέρασμα, πρώτο στοιχείο θα είναι το μικρότερο όλων. Κάνε και δεύτερο πέρασμα, για τα στοιχεία από το δεύτερο και μετά, οπότε έτσι θα έλθει και το δεύτερο κατά σειρά στη θέση του. Μετά από $n-1$ περάσματα συνολικά, ο πίνακας θα έχει ταξινομηθεί.



πολυπλοκότητα $O(n^2)$ **Για** i = 1, 2, ..., $n-1$
  Για $j = n-1, n-2, ..., i$
  Αν $x_{j-1} > x_j$ **τότε**
  Αντιμετάθεσε x_{j-1} και x_j

- *Η μέθοδος της επιλογής*

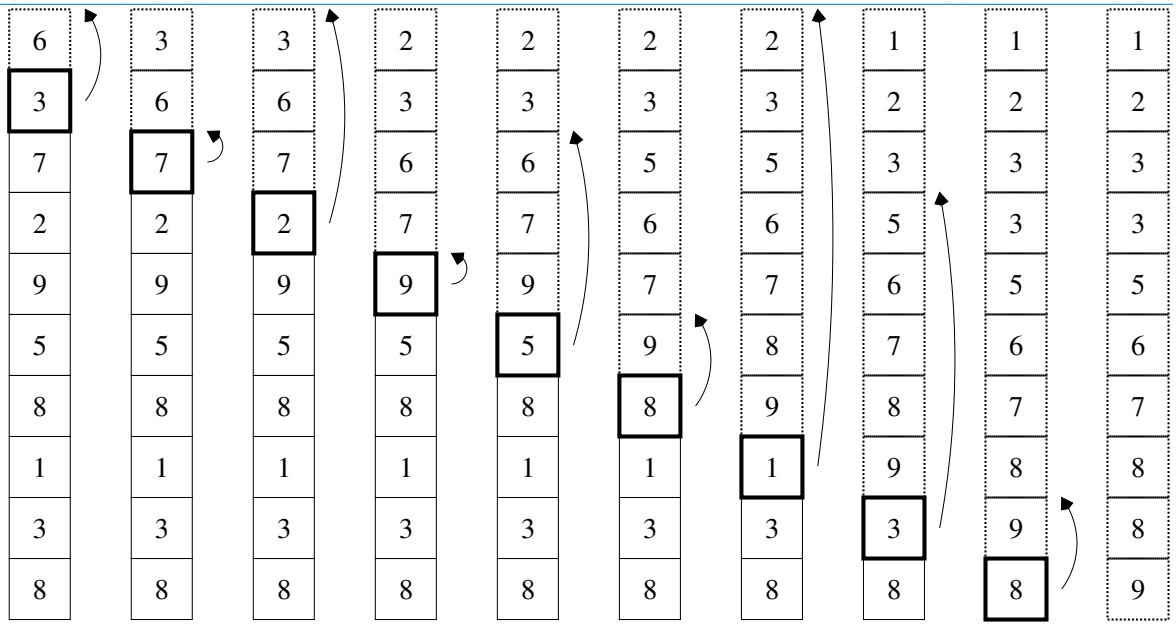
  Βρες το μικρότερο από όλα τα στοιχεία και αντιμετάθεσέ το με το πρώτο. Βρες το μικρότερο από τα υπόλοιπα και αντιμετάθεσέ το με το δεύτερο. Μετά από $n-1$ επιλογές και αντιμεταθέσεις συνολικά, ο πίνακας θα έχει ταξινομηθεί.



πολυπλοκότητα $O(n^2)$ **Για** i = 1, 2, ..., $n-1$
  $min = i-1$
  Για $j = i, i+1, ..., n-1$
  Αν $x_j < x_{min}$ **τότε**
  $min = j$
  Αντιμετάθεσε x_{i-1} και x_{min}

- *Η μέθοδος της εισαγωγής*

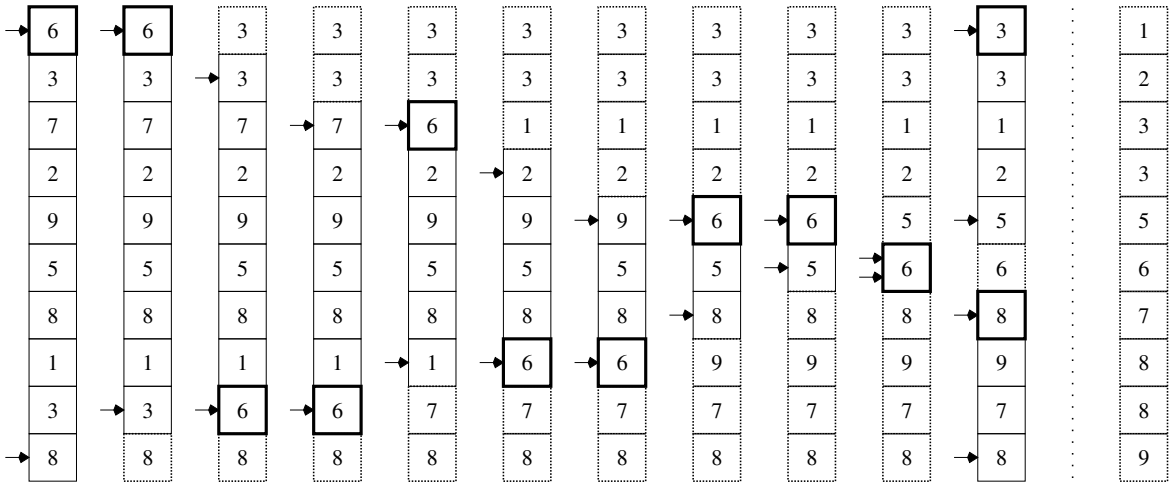
  Τοποθέτησε το δεύτερο στοιχείο πριν ή μετά το πρώτο ώστε να είναι στη σωστή σειρά. Τοποθέτησε το τρίτο στη σωστή θέση στα ήδη ταξινομημένα πρώτο και δεύτερο. Μετά τοποθέτησε το τέταρτο στα τρία πρώτα. Μετά από $n-1$ εισαγωγές συνολικά, ο πίνακας θα έχει ταξινομηθεί.



πολυπλοκότητα $O(n^2)$ **Για** i = 1, 2, ..., $n-1$
  $j = i-1$
  Ενόσω $j \geq 0$ **και** $x_j > x_{j+1}$
  Αντιμετάθεσε x_j και x_{j+1}
  $j = j-1$

- *Η γρήγορη μέθοδος*

  Με βάση το πρώτο στοιχείο σαν οδηγό, χώρισε τον πίνακα σε δύο άλλους, έναν που περιέχει στοιχεία μικρότερα ή ίσα με τον οδηγό και έναν με μεγαλύτερα ή ίσα. Ταξινόμησε τους δύο αυτούς πίνακες με την ίδια μέθοδο αναδρομικά (εφ' όσον έχουν τουλάχιστον δύο στοιχεία), οπότε το τελικό αποτέλεσμα είναι η παράθεση των ταξινομημένων αυτών πινάκων με ενδιάμεση παρεμβολή του οδηγού στοιχείου.



μέση πολυπλοκότητα $O(n \log n)$

Συνάρτηση $qs(x, up, down)$
  $start$ = up
  end = $down$
  Ενόσω $up < down$
  Ενόσω $x_{down} \geq x_{up}$ **και** $up < down$
  $down$ = $down-1$
  Αν $up \neq down$ **τότε**
  Αντιμετάθεσε x_{up} και x_{down}
  up = $up+1$
  Ενόσω $x_{up} \leq x_{down}$ **και** $up < down$
  up = $up+1$
  Αν $up \neq down$ **τότε**
  Αντιμετάθεσε x_{up} και x_{down}
  $down$ = $down-1$
  Αν $start < up-1$ **τότε**
  Κάλεσε $qs(x, start, up-1)$
  Αν $end > down+1$ **τότε**
  Κάλεσε $qs(x, down+1, end)$

Για να ταξινομήσουμε τον πίνακα x που έχει n στοιχεία, θα πρέπει να καλέσουμε $qs(x, 0, n-1)$.

- Άλλες μέθοδοι ταξινόμησης
  - Η μέθοδος του σωρού, $O(n \log n)$
  - Η μέθοδος της συγχώνευσης, $O(n \log n)$
  - Η μέθοδος του Shell, $O(n^2)$ (σελ. 94 [KR])

Μέθοδοι ταξινόμησης

```c
/* File: sorting.c */
#include <stdio.h>
#include <stdlib.h>
#include <time.h>

void bubblesort(int, double *);
void selectsort(int, double *);
void insertsort(int, double *);
void quicksort(int, double *);
void quicksort_body(double *, int, int);
void swapd(double *, double *);

int main(int argc, char *argv[])
{ char method = 'b', *name; /* Default sorting method is bubblesort */
```

```
  int i, n = 10;                          /* Default array size is 10 */
  long seed;
  double *x, sttime, endtime;
  void (*fun)(int, double *);         /* Pointer to sorting function */
  seed = time(NULL); /* Get current time, in case seed is not given */
  if (argc > 1)                  /* First character of first argument */
    method = *argv[1];         /* denotes the employed sorting method */
  if (argc > 2)
    n = atoi(argv[2]);      /* Second argument is number of elements */
  if (argc > 3)                           /* Third argument is seed for */
    seed = atoi(argv[3]);                    /* random number generator */
  switch(method) {         /* Prepare calling the appropriate method */
    case 'b':
      fun = bubblesort; name = "bubblesort"; break;
    case 's':
      fun = selectsort; name = "selectsort"; break;
    case 'i':
      fun = insertsort; name = "insertsort"; break;
    case 'q':
      fun = quicksort; name = "quicksort"; break;
    default:
      printf("Sorry, no such method\n");
      return 1;
  }                                    /* Allocate memory for the array */
  if ((x = malloc(n * sizeof(double))) == NULL) {
    printf("Sorry, not enough memory\n");
    return 1; }
  srand((unsigned int) seed); /* Initialize random number generator */
  for (i=0 ; i < n ; i++) /* Generate double floating point numbers */
    x[i] = ((double) rand())/RAND_MAX;
  printf("Random numbers\n");
  for (i=0 ; i < n ; i++) {
    printf("%6.4f ", x[i]);                         /* Print them out */
    if (i%10 == 9)                                    /* 10 in a line */
      printf("\n"); }
  printf("\n");
  printf("Sorting by %s\n", name);
  sttime = ((double) clock())/CLOCKS_PER_SEC;         /* Get CPU time */
                                  /* consumed since start of program */
  (*fun)(n, x);                                    /* Call sorting method */
  endtime = ((double) clock())/CLOCKS_PER_SEC;     /* Again CPU time */
                               /* Difference endtime-sttime should be */
                                     /* CPU time consumed for sorting */
```

```
  for (i=0 ; i < n ; i++) {
    printf("%6.4f ", x[i]);              /* Print out sorted array */
    if (i%10 == 9)
      printf("\n"); }
  printf("\n");            /* Print out CPU time needed for sorting */
  printf("Time: %.2f secs\n", endtime-sttime);
  free(x); return 0;
}

void bubblesort(int n, double *x)
{ int i, j;
  for (i=1 ; i <= n-1 ; i++)          /* Bring appropriate element, */
                                 /* that is the bubble, to place i-1 */
    for (j=n-1 ; j >= i ; j--)
      if (x[j-1] > x[j])     /* Compare pairwise from bottom to top */
        swapd(&x[j-1], &x[j]);                /* and swap if needed */
}

void selectsort(int n, double *x)
{ int i, j, min;
  for (i=1 ; i <= n-1 ; i++) {
    min = i-1;            /* Let current minimum be the i-1 element */
    for (j=i ; j <= n-1 ; j++)
      if (x[j] < x[min])  /* Check if any element after i-1 is less */
        min = j; /* than so far minimum and make it the new minimum */
    swapd(&x[i-1], &x[min]); } /* Exchange minimum with i-1 element */
}
void insertsort(int n, double *x)
{ int i, j;
  for (i=1 ; i <= n-1 ; i++) {      /* Insert element at place i in */
                        /* its correct position from places 0 to i-1 */
    j = i-1;
    while (j >= 0 && x[j] > x[j+1]) {/* Move repeatedly the element */
      swapd(&x[j], &x[j+1]);                    /* until it reaches */
      j--; } }                             /* its correct position */
}

void quicksort(int n, double *x)
{ quicksort_body(x, 0, n-1);    /* Call recursive quicksort to sort */
}          /* elements of the array from position 0 to position n-1 */

void quicksort_body(double *x, int up, int down)
{ int start, end;
```

```
  start = up;              /* Save start position of small elements */
  end = down;                /* Save end position of large elements */
  while (up < down) {            /* Pivot element is at up position */
    while (x[down] >= x[up] && up < down)      /* Let down elements */
      down--;              /* larger than pivot stay where they are */
    if (up != down) {                    /* If pivot is not reached */
      swapd(&x[up], &x[down]);   /* echange it with smaller element */
      up++;     /* Pivot is at down position, move up a bit further */
    }
    while (x[up] <= x[down] && up < down)        /* Let up elements */
    up++;                  /* smaller than pivot stay where they are */
    if (up != down) {                    /* If pivot is not reached */
      swapd(&x[up], &x[down]);   /* exchange it with larger element */
      down--;   /* Pivot is at up position, move down a bit further */
    } }       /* Now up = down is the position of the pivot element */
  if (start < up-1) /* Is there at least one element left of pivot? */
    quicksort_body(x, start, up-1); /* Quick(sort) smaller elements */
  if (end > down+1)/* Is there at least one element right of pivot? */
    quicksort_body(x, down+1, end);  /* Quick(sort) larger elements */
}

void swapd(double *a, double *b)       /* Just exchange two doubles */
{ double temp;
  temp = *a;
  *a = *b;
  *b = temp; }
```

```
% gcc -o sorting sorting.c
% ./sorting b 30
Random numbers
0.1914 0.8545 0.6266 0.4347 0.0597 0.1278 0.1717 0.9190 0.8703 0.5835
0.5544 0.2181 0.4099 0.3580 0.7294 0.7789 0.1550 0.3378 0.4779 0.8070
0.2298 0.0863 0.6197 0.7538 0.2502 0.4336 0.5880 0.5354 0.6012 0.6745

Sorting by bubblesort
0.0597 0.0863 0.1278 0.1550 0.1717 0.1914 0.2181 0.2298 0.2502 0.3378
0.3580 0.4099 0.4336 0.4347 0.4779 0.5354 0.5544 0.5835 0.5880 0.6012
0.6197 0.6266 0.6745 0.7294 0.7538 0.7789 0.8070 0.8545 0.8703 0.9190

Time: 0.00 secs
% ./sorting s 30
Random numbers
0.1520 0.1810 0.6908 0.9386 0.9103 0.1460 0.6984 0.2948 0.5047 0.7014
0.2916 0.7622 0.9437 0.1535 0.0158 0.2708 0.6973 0.9403 0.3101 0.8563
```

```
0.0342 0.6943 0.6948 0.5267 0.4925 0.9227 0.8265 0.4250 0.9225 0.0146

Sorting by selectsort
0.0146 0.0158 0.0342 0.1460 0.1520 0.1535 0.1810 0.2708 0.2916 0.2948
0.3101 0.4250 0.4925 0.5047 0.5267 0.6908 0.6943 0.6948 0.6973 0.6984
0.7014 0.7622 0.8265 0.8563 0.9103 0.9225 0.9227 0.9386 0.9403 0.9437

Time: 0.00 secs
% ./sorting i 30
Random numbers
0.1593 0.5741 0.8110 0.1777 0.4501 0.6941 0.7238 0.8614 0.1461 0.7922
0.8593 0.8994 0.7509 0.0171 0.4619 0.9863 0.9694 0.7962 0.1611 0.9896
0.5480 0.8230 0.6144 0.6332 0.5996 0.2770 0.3051 0.2945 0.5684 0.7232

Sorting by insertsort
0.0171 0.1461 0.1593 0.1611 0.1777 0.2770 0.2945 0.3051 0.4501 0.4619
0.5480 0.5684 0.5741 0.5996 0.6144 0.6332 0.6941 0.7232 0.7238 0.7509
0.7922 0.7962 0.8110 0.8230 0.8593 0.8614 0.8994 0.9694 0.9863 0.9896

Time: 0.00 secs
% ./sorting q 30
Random numbers
0.1687 0.4653 0.4296 0.9194 0.4879 0.2398 0.2521 0.4263 0.7844 0.3863
0.9253 0.5324 0.5617 0.8795 0.9027 0.7057 0.7409 0.1454 0.0162 0.1232
0.5537 0.9557 0.0353 0.7299 0.2103 0.1340 0.7720 0.6671 0.2188 0.4183

Sorting by quicksort
0.0162 0.0353 0.1232 0.1340 0.1454 0.1687 0.2103 0.2188 0.2398 0.2521
0.3863 0.4183 0.4263 0.4296 0.4653 0.4879 0.5324 0.5537 0.5617 0.6671
0.7057 0.7299 0.7409 0.7720 0.7844 0.8795 0.9027 0.9194 0.9253 0.9557

Time: 0.00 secs
%

% hostname
linux29
% ./sorting b 20000 2016 | tail -4
0.9987 0.9987 0.9987 0.9987 0.9987 0.9988 0.9988 0.9989 0.9991 0.9991
0.9992 0.9992 0.9992 0.9992 0.9993 0.9993 0.9993 0.9995 0.9996 1.0000

Time: 2.07 secs
% ./sorting b 50000 2016 | tail -4
0.9994 0.9994 0.9994 0.9994 0.9994 0.9994 0.9994 0.9995 0.9995 0.9995
0.9995 0.9995 0.9996 0.9996 0.9997 0.9997 0.9998 0.9999 0.9999 1.0000
```

```
Time: 12.87 secs
% ./sorting s 20000 2016 | tail -4
0.9987 0.9987 0.9987 0.9987 0.9987 0.9988 0.9988 0.9989 0.9991 0.9991
0.9992 0.9992 0.9992 0.9992 0.9993 0.9993 0.9993 0.9995 0.9996 1.0000

Time: 0.78 secs
% ./sorting s 50000 2016 | tail -4
0.9994 0.9994 0.9994 0.9994 0.9994 0.9994 0.9994 0.9995 0.9995 0.9995
0.9995 0.9995 0.9996 0.9996 0.9997 0.9997 0.9998 0.9999 0.9999 1.0000

Time: 4.85 secs
% ./sorting i 20000 2016 | tail -4
0.9987 0.9987 0.9987 0.9987 0.9987 0.9988 0.9988 0.9989 0.9991 0.9991
0.9992 0.9992 0.9992 0.9992 0.9993 0.9993 0.9993 0.9995 0.9996 1.0000

Time: 1.02 secs
% ./sorting i 50000 2016 | tail -4
0.9994 0.9994 0.9994 0.9994 0.9994 0.9994 0.9994 0.9995 0.9995 0.9995
0.9995 0.9995 0.9996 0.9996 0.9997 0.9997 0.9998 0.9999 0.9999 1.0000

Time: 6.36 secs
% ./sorting q 20000 2016 | tail -4
0.9987 0.9987 0.9987 0.9987 0.9987 0.9988 0.9988 0.9989 0.9991 0.9991
0.9992 0.9992 0.9992 0.9992 0.9993 0.9993 0.9993 0.9995 0.9996 1.0000

Time: 0.01 secs
% ./sorting q 50000 2016 | tail -4
0.9994 0.9994 0.9994 0.9994 0.9994 0.9994 0.9994 0.9995 0.9995 0.9995
0.9995 0.9995 0.9996 0.9996 0.9997 0.9997 0.9998 0.9999 0.9999 1.0000

Time: 0.01 secs
% ./sorting q 3000000 2016 | tail -4
1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000
1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000 1.0000

Time: 0.84 secs
%
```

Αναζήτηση σε πίνακες

- Πολλές φορές έχουμε ένα πίνακα από στοιχεία (αριθμούς, συμβολοσειρές, οτιδήποτε) και θέλουμε να δούμε αν ένα συγκεκριμένο στοιχείο, ίδιου τύπου με αυτά του πίνακα, ανήκει

στον πίνακα. Η διαδικασία αυτή λέγεται αναζήτηση.

- Ο απλούστερος τρόπος να κάνουμε αναζήτηση σ' ένα πίνακα είναι με τη λεγόμενη σειριακή αναζήτηση. Διατρέχουμε τα στοιχεία του πίνακα, ένα προς ένα, μέχρι να βρούμε το στοιχείο που ψάχνουμε, αν το βρούμε τελικά.
- Αν ο πίνακας είναι μεγάλος, η σειριακή αναζήτηση μπορεί να είναι πολύ χρονοβόρα. Αν όμως ταξινομήσουμε τον πίνακα, έστω σε αύξουσα σειρά, μπορούμε να εφαρμόσουμε μία πολύ ταχύτερη μέθοδο, τη δυαδική αναζήτηση.
- Με τη δυαδική αναζήτηση, συγκρίνουμε το στοιχείο που ψάχνουμε με το μεσαίο (ή ένα από τα δύο μεσαία) στοιχείο του πίνακα. Αν συμπίπτει, σταματάμε την αναζήτηση αφού το στοιχείο βρέθηκε. Αν όχι, στην περίπτωση που το στοιχείο προηγείται του μεσαίου, ψάχνουμε να το βρούμε στο πρώτο μισό του πίνακα. Αν έπεται του μεσαίου, το ψάχνουμε στο δεύτερο μισό. Η αναζήτηση στον κατάλληλο υποπίνακα γίνεται με τον ίδιο τρόπο, συγκρίνοντας το στοιχείο με το μεσαίο του υποπίνακα και η διαδικασία συνεχίζει με αυτόν τον τρόπο, μέχρις ότου είτε να βρεθεί το στοιχείο είτε να φτάσουμε σε κενό υποπίνακα.
- Ποιες είναι οι πολυπλοκότητες των μεθόδων; [1]

Μέθοδοι αναζήτησης

```c
/* File: searching.c */
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <time.h>

void heapsort(int, char **);
void heapify(char **, int, int);
int seqsearch(char *, int, char **);
int binsearch(char *, int, char **);
void swapwords(char **, char **);

int main(int argc, char *argv[])
{ int k = 0;                    /* Counter of words that will be read */
  int nmax = 1000;        /* Default maximum number of words to read */
  double sttime, endtime;
  char search = 's';                /* Default is sequential search */
  char **words, *arg, buf[81];
  while (--argc) {
    arg = *++argv;
    if (!strcmp(arg, "-max")) {               /* Get maximum number */
      if (argc > 1 && --argc)                   /* of words to read */
        nmax = atoi(*++argv); }
    else if (!strcmp(arg, "-seq"))      /* Select sequential search */
```

[1] $O(n)$ και $O(\log n)$, για τη σειριακή και τη δυαδική αναζήτηση, αντίστοιχα.

```
      search = 's';
    else if (!strcmp(arg, "-bin"))          /* Select binary search */
      search = 'b';
    else if (!strcmp(arg, "-words")) {  /* Give words to search for */
      argc--;
      break; } }
  argv++;    /* Allocate memory to store addresses of words to read */
  if ((words = malloc(nmax * sizeof(char *))) == NULL) {
    fprintf(stderr, "Not enough memory\n");
    return 1; }
  while (k < nmax && scanf("%80s", buf) != EOF) {
              /* Read words until EOF or maximum number reached and */
                                  /* allocate memory to store them */
    if ((words[k] = malloc((strlen(buf)+1) * sizeof(char))) == NULL) {
      fprintf(stderr, "Not enough memory\n");
      return 2; }
    strcpy(words[k++], buf); }                    /* Store word read */
  if (search == 'b')                    /* If binary search selected */
    heapsort(k, words);       /* sort words via the heapsort method */
  sttime = ((double) clock())/CLOCKS_PER_SEC;  /* Search start time */
  while (argc--) {
    arg = *argv++;
    switch (search) {
      case 's':                          /* The sequential search case */
        printf("%sfound %s\n",
          seqsearch(arg, k, words) ? "    " : "not ", arg);
        break;
      case 'b':                             /* The binary search case */
        printf("%sfound %s\n",
          binsearch(arg, k, words) ? "    " : "not ", arg);
        break; } }
  endtime = ((double) clock())/CLOCKS_PER_SEC;   /* Search end time */
  printf("Searching time is %.2f seconds\n", endtime-sttime);
  return 0; }

void heapsort(int n, char **x)
{ int i;
  for (i=(n/2)-1 ; i >= 0 ; i--)       /* Transform array to a heap */
/* A heap is an implicit binary tree where each node is not smaller */
    heapify(x, i, n-1);             /* than its immediate successors */
  for (i=n-1 ; i >= 1 ; i--) {/* Move heap root to bottom rightmost */
    swapwords(&x[0], &x[i]);    /* position not already settled and */
    heapify(x, 0, i-1); }/*transform to a heap the rest of the tree */
```

```
}
void heapify(char **x, int root, int bottom)
{ int maxchild; /* Transform to a heap the subtree starting at root */
                                          /* up to element bottom */
  while (2*root < bottom) {         /* Do we have still work to do? */
    if (2*root+1 == bottom)    /* If left child is last to consider */
      maxchild = 2*root+1;             /* this is the maximum child */
    else if (strcmp(x[2*root+1], x[2*root+2]) > 0)     /* Otherwise */
      maxchild = 2*root+1;           /* select maximum between left */
    else
      maxchild = 2*root+2;               /* and right child of root */
    if (strcmp(x[maxchild], x[root]) > 0) {/* Compare maximum child */
      swapwords(&x[maxchild], &x[root]);      /* with root and swap */
      root = maxchild; } /* accordingly, defining also the new root */
    else
      break; } }                            /* OK, we made our heap */

int seqsearch(char *w, int n, char **x)
{ int i;
  for (i=0 ; i < n ; i++)
    if (!strcmp(w, x[i]))       /* So simple, what to comment here! */
      return 1;
  return 0; }

int binsearch(char *w, int n, char **x)
{ int cond, low, high, mid;
  low = 0;                              /* Lower limit for searching */
  high = n-1;                           /* Upper limit for searching */
  while (low <= high) {          /* Do we have space for searching? */
    mid = (low+high)/2;         /* Medium element of search interval */
                          /* to compare with word we are looking for */
    if ((cond = strcmp(w, x[mid])) < 0)   /* Compare medium to word */
      high = mid-1;       /* Not found, word might be at first half */
    else if (cond > 0)
      low = mid+1;       /* Not found, word might be at second half */
    else return 1; }                                 /* We found it! */
  return 0; }                             /* Sorry, word not found */

void swapwords(char **w1, char **w2)
{ char *temp;  /* The well-known swap function for the strings case */
  temp = *w1;
  *w1 = *w2;
  *w2 = temp; }
```

```
% gcc -o searching searching.c
% cat test_words.txt
Hello there! How are you?
Hello!!! I am fine. What about you? Are you OK?
Yes, I am fine. Thank you.
% ./searching -seq -max 50 -words you word fine I < test_words.txt
    found you
not found word
not found fine
    found I
Searching time is 0.00 seconds
% ./searching -bin -max 50 -words you word fine I < test_words.txt
    found you
not found word
not found fine
    found I
Searching time is 0.00 seconds
% ./searching -bin -max 12 -words you word fine I < test_words.txt
not found you
not found word
not found fine
    found I
Searching time is 0.00 seconds
% wc -w /usr/share/dict/words
234937 /usr/share/dict/words
% ./searching -seq -max 240000 \
?       -words `cat KRExcerpt.txt` < /usr/share/dict/words | tail
    found well
    found to
    found write
    found major
not found programs
    found in
    found many
    found different
not found domains.
Searching time is 0.08 seconds
% ./searching -bin -max 240000 \
?       -words `cat KRExcerpt.txt` < /usr/share/dict/words | tail
    found well
    found to
    found write
    found major
```

```
not found programs
    found in
    found many
    found different
not found domains.
Searching time is 0.00 seconds
%

% hostname
linux29
% wc -w big.txt
1095695 big.txt
% ./searching -seq -max 1100000 \
?       -words `cat KRExcerpt.txt` < big.txt | tail -6
    found programs
    found in
    found many
    found different
not found domains.
Searching time is 0.14 seconds
% ./searching -bin -max 1100000 \
?       -words `cat KRExcerpt.txt` < big.txt | tail -6
    found programs
    found in
    found many
    found different
not found domains.
Searching time is 0.00 seconds
% ./searching -bin -max 1100000 \
?       -words `head -180000 /usr/share/dict/words` < big.txt \
?                                           | grep 'not found' | wc -l
167331
% ./searching -bin -max 1100000 \
?       -words `head -180000 /usr/share/dict/words` < big.txt \
?                                           | grep '    found' | wc -l
12669
% ./searching -bin -max 1100000 \
?       -words `head -20000 /usr/share/dict/words` < big.txt \
?                                           | grep 'Searching time'
Searching time is 0.01 seconds
% ./searching -seq -max 1100000 \
?       -words `head -20000 /usr/share/dict/words` < big.txt \
?                                           | grep 'Searching time'
Searching time is 209.13 seconds
```

%