Algoritmos e Estruturas de Dados

Multi-ordered Trees

Licenciatura em Engenharia Informática

Leonardo dos Santos Flórido - 103360 - 50% Gabriel Hall Abreu - 102851 - 50%

Índice

Introdução	3
Binary Trees	
Gráficos	4
Histogramas	
Influência nas árvores com zip codes limitados	
Código implementado	9
Output	13
Conclusão	17
Bibliografia	

Introdução

No âmbito desta unidade curricular pretendemos com este projeto aprofundar os nossos conhecimentos sobre a linguagem C, aproveitando também para pôr em prática novos métodos necessários para a resolução deste problema.

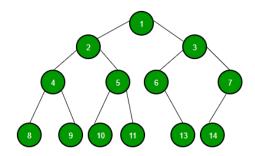
Para tal, será necessário compreender a estrutura de dados da **árvore binária** e os seus respectivos métodos.

Em particular, neste trabalho será necessário compreender como guardar e processar dados diferentes de forma que se tenha acesso aos mesmos usando uma de várias possíveis chaves.

Binary Trees

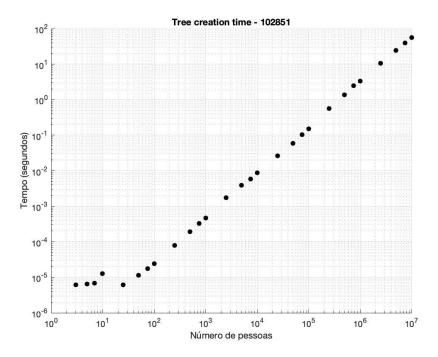
Uma **árvore binária** é uma estrutura dinâmica composta de *nodes*. Cada *node* de informação contém:

- A própria informação (item de dados).
- Um ponteiro para o *node* à esquerda; numa árvore binária ordenada os itens de dados estão guardados neste lado são todos eles mais pequenos do que os dados guardados no *node*.
- Um ponteiro para o *node* à direita; numa árvore binária ordenada os itens de dados estão guardados neste lado são todos eles maiores do que os dados guardados no *node*.
- Opcionalmente, um ponteiro para o node pai.

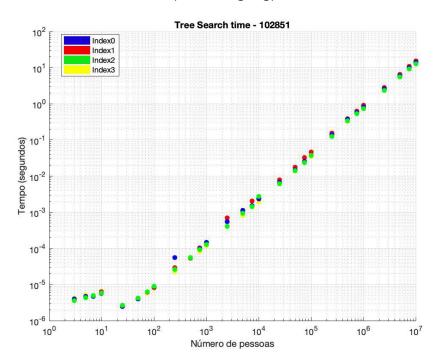


Gráficos

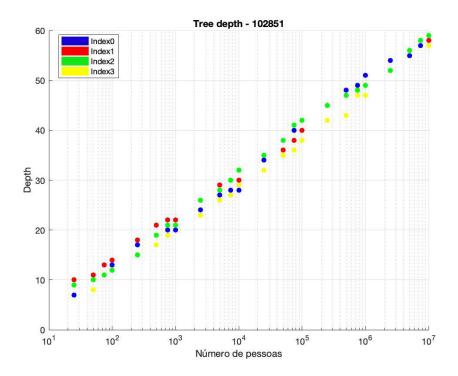
102851 - Relação Tree Creation Time / Number of people (escala: log/log)



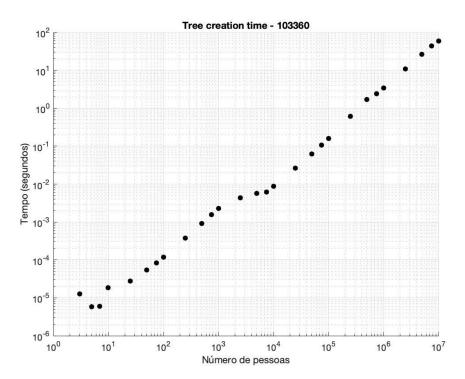
102851 - Relação Tree Search Time / Number of people (escala: log/log)



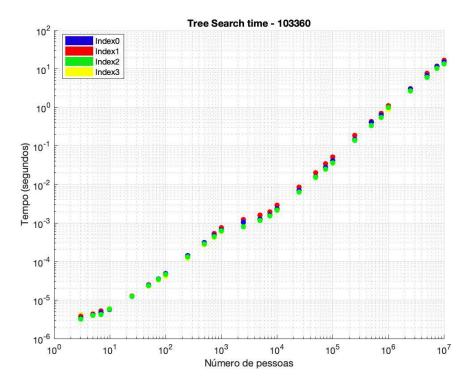
102851 - Relação Tree Depth / Number of people (escala: linear/log)



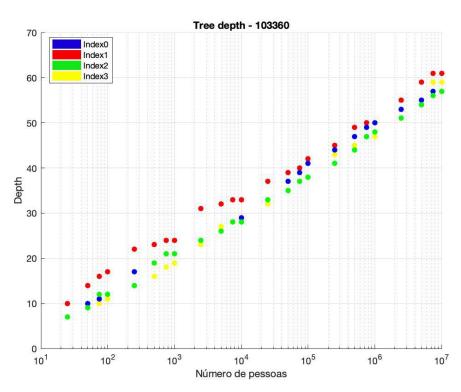
103360 - Relação Tree Creation Time / Number of people (escala: log/log)



103360 - Relação Tree Search Time / Number of people (escala: log/log)



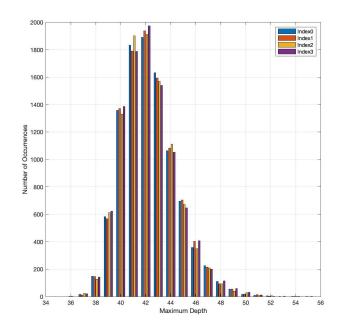
103360 - Relação Tree Depth / Number of people (escala: linear/log)



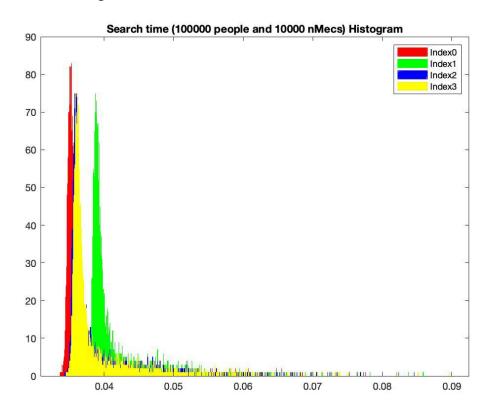
Histogramas

Para cada histograma foram geradas 100000 pessoas e 10000 nMecs.

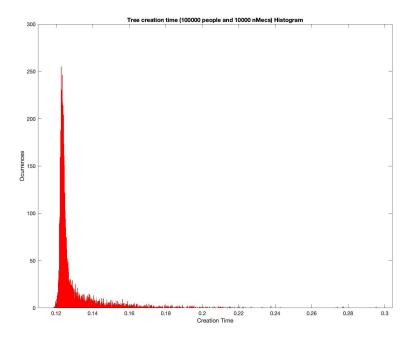
Histograma - número de ocorrências / tree depth



Histograma - número de ocorrências / tree search time







Influência nas árvores com zip codes limitados

"Does that influence much the execution times for the construction and searches for the tree ordered by the zip codes?"

Resposta:

São disponibilizados 500 zip codes diferentes, ao criar uma árvore (*root*) com mais de 500 pessoas, vão existir pessoas com o mesmo zip code.

Na criação da árvore (root) ordenada por zip code, que vai ser feita pela a função 'tree_insert()' recorrendo a função 'compare_tree_nodes()', a função 'compare_tree_nodes()' em certos casos vai comparar pessoas com o mesmo zip code. Para saber a ordem certa de inserção dessas pessoas na árvore (root), terá de comparar o próximo parâmetro das pessoas (no nosso caso vai comparar os cc). Com isto, poderão existir casos em que a função terá de fazer 2 comparações para saber a ordem correta para inserir na árvore (root), o que aumenta o tempo de criação da árvore. O mesmo raciocínio aplica-se ao search time (search time é o tempo que demora a função 'find()' a correr). A função 'find()' recorre também à função 'compare_tree_nodes()', como em alguns casos têm de fazer 2 comparações, o search time vai aumentar.

Código implementado

Dentro do ficheiro random_data.c desenvolvemos a função 'random_cc()' que devolve um número de cartão de cidadão aleatório a ser utilizado posteriormente como quarto campo da estrutura tree_node_t.

Em termos de estrutura, temos *tree_node_t*, no qual define os *nodes* das *Binary Trees*. Acrescentamos o *cc* à estrutura, na qual se define o número de cartão de cidadão.

```
// AED, January 2022
// AED, January 2022
// Solution of the second practical assignement (multi-ordered tree)
//
// Place your student numbers and names here
//
// Place your student numbers and names here
//
//
// #include <stdio.h>
//
// #include <stdib.h>
//
// #include "AED_2021_A02.h"
//
//
// #include "AED_2021_A02.h"
//
// we want to maintain three ordered trees (using the same nodes!), so we need four left and four right pointers
// yo, when inserting a new node we need to do it four times (one for each index), so we will end upo with 4 three roots
//
//
// typedef struct tree_node_s
// typedef struct tree_node_s
// char name[MAX_NAME_SIZE + 1]; // index 0 data item
// char zip_code[MAX_ZIP_CODE_SIZE + 1]; // index 1 data item
// char ctelephone_number[MAX_TIRLEPHONE_NUMBER_SIZE + 1]; // index 2 data item
// char ctelephone_number[MAX_TIRLEPHONE_NUMBER_SIZE + 1]; // index 2 data item
// index 1 data item
// index 2 data item
// index 3 data item
// index 5 data item
// index 6 data item
// index 7 data item
// index 8 data item
// index 9 data item
// index 1 data item
// index 1 data item
// index 2 data item
// index 3 data item
// index 6 data item
// index 7 data item
// index 6 data item
// index 7 data item
// index 6 data item
// index 7 data item
// index 8 data item
// index 9 data item
// index 1 data item
// index 1 data item
// index 2 data item
// index 3 data item
// index 6 data item
// index 7 data item
// index 8 data item
// index 9 data item
// index 9 data item
// index 1 data item
// index 1 data item
// index 1 data item
// index 2 data item
// index 3 data item
// index 6 data item
// index 9 data item
/
```

Usámos a função 'compare_tree_nodes()' fornecida pelo o docente para equiparar os nodes, o que será útil para criar as *Binary Trees* de uma forma ordenada.

Como acrescentamos um quarto parâmetro, foi necessário alterar a função.

```
// the node comparison function (do not change this)
// the node comparison function (do not change this)
// 
int compare_tree_nodes(tree_node_t *node1, tree_node_t *node2, int main_idx)

{
    int i, c;

for (i = 0; i < 4; i++)
    {
        if (main_idx == 0)
        | c = strcmp(node1->name, node2->name);
        else if (main_idx == 1)
        | c = strcmp(node1->zip_code, node2->zip_code);
        else if (main_idx == 2)
        | c = strcmp(node1->telephone_number, node2->telephone_number);
        else if (main_idx == 3)
        | c = strcmp(node1->cc, node2->cc);
        if (c != 0)
        | return c;
        main_idx = (main_idx == 3) ? 0 : main_idx + 1; // advance to the next index
}
return 0;
}
```

Seguidamente temos a função 'tree_insert()', adaptada da contida nas Lecture notes, que passando como argumentos os roots, o node e o main_index vai inserir os nodes nos roots de forma ordenada recorrendo a função 'compare_tree_nodes()'.

A função 'find()', recorrendo à função 'compare_tree_nodes()', percorre toda a árvore (root) e será posteriormente utilizada para calcular o 'tree search time' (o tempo de execução da função será o 'tree search time').

A função 'tree_depth()' calcula a profundidade máxima da árvore (root). Se a profundidade do lado esquerdo da root for maior do que a do lado direito será retornada a profundidade do lado esquerdo, caso contrário será retornada a profundidade do lado direito.

```
98
99
      // tree depdth
100
101
      int tree depth(tree node t *root, int main index)
102
103
          if (root == NULL)
              return 0;
105
106
          int l = tree depth(root->left[main index], main index);
107
          int r = tree depth(root->right[main index], main index);
109
110
          if (l > r)
111
              return l + 1;
112
          return r + 1;
113
114
```

A função 'list()' lista ordenadamente todos os *nodes* pertencentes à árvore (root) de acordo com o index escolhido pelo utilizador.

```
117
     // list, i,e, traverse the tree (place your code here)
118
119
120
     int c1 = 1; // global variable
     void list(tree node t *root, int main index)
121
122
         if (root != NULL)
123
124
             list(root->left[main index], main index);
125
126
             printf("Person #%d\n", c1++);
             printf("
                        name ----- %s\n", root->name);
127
                        zip code ----- %s\n", root->zip_code);
             printf("
128
             printf("
                        telephone number --- %s\n", root->telephone number);
129
             printf("
                        cc -----%s\n", root->cc);
130
             list(root->right[main index], main index);
131
132
133
```

A função 'findZipCode()', dado um zip code, percorre toda a árvore (root) e quando encontra um node com o zip code desejado lista-o. Isto acontece até não existirem mais nodes a comparar.

```
// list the people with a given zip code
136
137
138
     int c2 = 1; // global variable
139
     void findZipCode(tree_node_t *root, char *zip_code)
         if (root != NULL)
             if (strcmp(root->zip code, zip code) == 0)
                 findZipCode(root->left[1], zip code);
                 printf("Person #%d\n", c2++);
                            name ----- %s\n", root->name);
                 printf("
                            zip code ----- %s\n", root->zip_code);
                 printf("
                 printf("
                            telephone number --- %s\n", root->telephone_number);
                 printf("
                            cc ----- %s\n", root->cc);
                 findZipCode(root->right[1], zip_code);
             else
                 findZipCode(root->left[1], zip code);
                 findZipCode(root->right[1], zip_code);
```

Output

- Student_number = 103360 :

```
leonardodsf@leonardodsf-Creator-15M-A9SD:~/Documents/GitHub/LEI/AED/Multi-ordered_trees$ ./multi_ordered_tree 103360 100000000
Tree creation time (10000000 people): 5.682e+01s
Tree search time (10000000 people, index 0): 1.366e+01s
Tree search time (10000000 people, index 1): 1.576e+01s
Tree search time (10000000 people, index 2): 1.509e+01s
Tree search time (10000000 people, index 3): 1.387e+01s
Tree depth for index 0: 59 (done in 5.332e-01s)
Tree depth for index 1: 61 (done in 5.414e-01s)
Tree depth for index 2: 57 (done in 5.934e-01s)
Tree depth for index 3: 57 (done in 5.988e-01s)
                                                                                                    -15M-A9SD:~/Documents/GitHub/LEI/AED/Multi-ordered trees$ ./multi ordered tree 103360 4 -list0
 leonardodsf@Leonardodsf-Creator-15M-A9SD:~/Docume
Tree creation time (4 people): 1.535e-06s
Tree search time (4 people, index 0): 8.300e-07s
Tree search time (4 people, index 1): 7.660e-07s
Tree search time (4 people, index 2): 7.030e-07s
Tree search time (4 people, index 3): 7.400e-07s
Tree depth for index 0: 3 (done in 5.210e-07s)
Tree depth for index 1: 4 (done in 4.800e-07s)
Tree depth for index 2: 3 (done in 4.510e-07s)
Tree depth for index 3: 3 (done in 4.810e-07s)
List of people:
 name ------- Lisa Hernandez
zip code ------- 11368 Corona (Queens county)
telephone number --- 8327 821 401
cc ------- 53708478
Person #3
 name ------ Mark Jenkins
zip code ------- 60629 Chicago (Cook county)
telephone number --- 4438 898 422
cc ------- 32264209
Person #4
name ------- Michael Ramirez
             name ----- Michael Ramirez
zip code ------ 33027 Hollywood (Broward county)
telephone number --- 9658 733 399
cc ------ 34477417
                                                                                                              -A9SD:~/Documents/GitHub/LEI/AED/Multi-ordered trees$ ./multi ordered tree 103360 4 -list1
 leonardodsf@Leonardodsf-Creator-15M-A950:~/Docume
Tree creation time (4 people): 6.297e-06s
Tree search time (4 people, index 0): 4.281e-06s
Tree search time (4 people, index 1): 3.495e-06s
Tree search time (4 people, index 2): 3.235e-06s
Tree search time (4 people, index 3): 3.467e-06s
Tree depth for index 0: 3 (done in 2.440e-06s)
Tree depth for index 1: 4 (done in 2.369e-06s)
Tree depth for index 2: 3 (done in 2.208e-06s)
Tree depth for index 3: 3 (done in 2.233e-06s)
 Tree depth for index 3: 3 (done in 2.233e-06s)
List of people:
Person #1
name -------- Lisa Hernandez
zip code ------ 11368 Corona (Queens county)
telephone number -- 8327 821 401
cc -------- 53708478
Person #2
Michael Pamiroz
            name ----- Michael Ramirez
zip code ----- 33027 Hollywood (Broward county)
telephone number --- 9658 733 399
cc ----- 34477417
              name ----- Mark Jenkins
              zip code ------ 66629 Chicago (Cook county)
telephone number --- 4438 898 422
cc ------ 32264209
   Person #4
            name ----- Ethelyn Barry
zip code ----- 95828 Sacramento (Sacramento county)
telephone number --- 9030 741 502
                                                             ---- 76297810
```

```
son #1
name ------- Mark Jenkins
zip code ------ 60629 Chicago (Cook county)
telephone number --- 4438 898 422
cc ------- 32264209
 Person #2
      name ----
                           ----- Lisa Hernandez
      zip code ------ 11368 Corona (Queens county)
telephone number --- 8327 821 401
cc ------ 53708478
name ----- Michael Ramirez
zip code ------ 33027 Hollywood (Broward county)
telephone number --- 9658 733 399
cc ------ 34477417
zip code ------ 60629 Chicago (Cook county)
telephone number --- 4438 898 422
 cc ----- 32264209
Person #2
Person #2
name ------ Michael Ramirez
zip code ------ 33027 Hollywood (Broward county)
telephone number --- 9658 733 399
cc ------ 34477417
Person #3
     son #3"
name ------- Lisa Hernandez
zip code ------ 11368 Corona (Queens county)
telephone number --- 8327 821 401
cc ------ 53708478
 Person #4
      nom #4
name ------ Ethelyn Barry
Zip code ------ 95828 Sacramento (Sacramento county)
telephone number --- 9030 741 502
cc ------ 76297810
 Documents/GitHub/LEI/AED/Multi-ordered_trees$ ./multi_ordered_tree 103360 1500 -find "11206 Brooklyn (Kings county)"
     cc ---
Person #2
     Non #2 Jacalyn Kim
name ------ Jacalyn Kim
Zip code ------ 11206 Brooklyn (Kings county)
telephone number -- 5122 296 196
CC ------ 31927034
  cc ---
erson #3
name -
                           Bobbie Cole
11206 Brooklyn (Kings county)
5833 122 918
76578885
     zip code -----
telephone number ---
  cc ---
erson #4
     name ------ Andrew Murray
zip code ------ 11206 Brooklyn (Kings county)
telephone number -- 7475 441 241
cc ------- 01606069
  cc ---
erson #5
name ·
     name -----zip code -----telephone number ---
                           Gene Ahmed
11206 Brooklyn (Kings county)
7505 344 933
07875628
```

- Student_number = 102851 :

```
eonardodsf@leonardodsf-Creator-<mark>15M-A9SD:</mark>~/Documents/GitHub/LEI/AED/Multi-ordered_trees$ ./multi_ordered_tree 102851 10000000
Leonardodsf@leonardodsf-Creator-15M-A9SD:-/Documents/Git
Tree creation time (100000000 people): 5.949e+01s
Tree search time (100000000 people, index 0): 1.478e+01s
Tree search time (10000000 people, index 1): 1.797e+01s
Tree search time (100000000 people, index 2): 1.619e+01s
Tree search time (100000000 people, index 3): 1.458e+01s
Tree depth for index 0: 57 (done in 5.584e-01s)
Tree depth for index 1: 58 (done in 5.327e-01s)
Tree depth for index 3: 59 (done in 6.338e-01s)
Tree depth for index 3: 59 (done in 6.124e-01s)
    eonardodsf@leonardodsf-Creator-15M-A95D:-/Documents/GitHub/LEI/AED/Multi-ordered trees$ ./multi ordered tree 102851 4 -list0
leonardodsf@leonardodsf-Creator-15M-A95D:~/Docume
Tree creation time (4 people): 1.633e-06s
Tree search time (4 people, index 0): 9.430e-07s
Tree search time (4 people, index 1): 7.860e-07s
Tree search time (4 people, index 2): 7.530e-07s
Tree search time (4 people, index 3): 7.360e-07s
Tree depth for index 0: 4 (done in 5.440e-07s)
Tree depth for index 1: 4 (done in 6.000e-07s)
Tree depth for index 2: 4 (done in 5.020e-07s)
Tree depth for index 3: 3 (done in 8.220e-07s)
  List of people:
Person #1
 name ------ Mary Paul
zip code ----- 37211 Nashville (Davidson county)
telephone number --- 3146 767 353
cc ------ 78085756
         name ------ Rita Hensley
zip code ------ 89031 North Las Vegas (Clark county)
telephone number --- 9413 320 621
cc -------- 87162912
 Person #4
          son #4
name ------- Shannon Christian
zip code ------ 90201 Bell (Los Angeles county)
telephone number --- 6189 678 649
cc ------ 48368067
                                       nardodsf-Creator-15M-A9SD:~/Documents/GitHub/LEI/AED/Multi-ordered_trees$ ./multi_ordered_tree 102851 4 -list1
leonardodsf@leonardodsf-Creator-15M-A9SD:~/Docume
Tree creation time (4 people): 1.493e-05s
Tree search time (4 people, index 0): 4.491e-06s
Tree search time (4 people, index 1): 3.892e-06s
Tree search time (4 people, index 2): 3.802e-06s
Tree search time (4 people, index 3): 3.759e-06s
Tree depth for index 0: 4 (done in 2.730e-06s)
Tree depth for index 1: 4 (done in 2.605e-06s)
Tree depth for index 2: 4 (done in 2.586e-06s)
Tree depth for index 3: 3 (done in 2.491e-06s)
List of people:
  List of people:
Person #1
         son #1
name ------ Mary Paul
zip code ------ 37211 Nashville (Davidson county)
telephone number --- 3146 767 353
cc ------ 78085756
          name ----- Shannon Christian
           zip code ------ 90201 Bell (Los Angeles county)
telephone number --- 6189 678 649
cc ----- 48368067
           zip code ------ 92804 Anaheim (Orange county)
telephone number --- 1963 704 614
                         ----- 07000733
```

```
name ------ Mary Paul
Zip code ------ 37211 Nashville (Davidson county)
telephone number --- 3146 767 353
CC ------ 78085756
  name ----
cc ---
Person #3
  son #3
name ------- Shannon Christian
zip code ------ 90201 Bell (Los Angeles county)
telephone number --- 6189 678 649
cc ------ 48368067
  Person #2
   son #2
name ------ Shannon Christian
zip code ----- 90201 Bell (Los Angeles county)
telephone number --- 6189 678 649
cc ------ 48368067
```

```
| Recommended Fig Howard odds f. Creatur. 1587.458D. __Nocuments / Github/LET/AED/Multi-ordered_trees ./multi_ordered_tree 102851 1500 -find "11206 Brooklyn (Kings county)" Tree creation time (1580 people) is .800=-845
Tree search time (1580 people, index 0): 2.701=-045
Tree search time (1580 people, index 2): 2.700=-045
Tree depth for Index 0: 21 (done in 1.700=-055)
Tree depth for Index 0: 21 (done in 1.700=-055)
Tree depth for Index 2: 22 (done in 1.700=-055)
Tree depth for Index 2: 22 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree depth for Index 2: 23 (done in 1.700=-055)
Tree Index 2: 23 (done in 1.700=-055)
Tree Index 2: 23 (done in 1.700=-055)
```

Conclusão

Em suma, este trabalho ajudou-nos a uma melhor percepção de métodos de criação de código e estruturas de dados, bem como a implementação das Binary Trees. Enriquecemos também os nossos conhecimentos da linguagem de programação C, na gestão de tempo e no trabalho em equipa.

Bibliografia

- Lecture notes da disciplina
- https://www.geeksforgeeks.org/tree-sort/