# 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	3
Gráficos	
Histogramas	7
Influência nas árvores com zip codes limitados	8
Código implementado	9
Output	13
Conclusão	17
Bibliografia	17

#### 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.

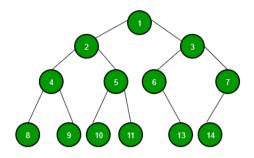
Para tal, será necessário compreender a estrutura de dados da **árvore binária** e os seus respectivos métodos.

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**

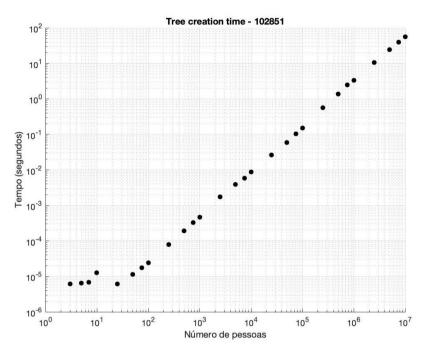
Uma **árvore binária** é uma estrutura dinâmica composta de *nodes*. Cada *node* de informação contém:

- 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.

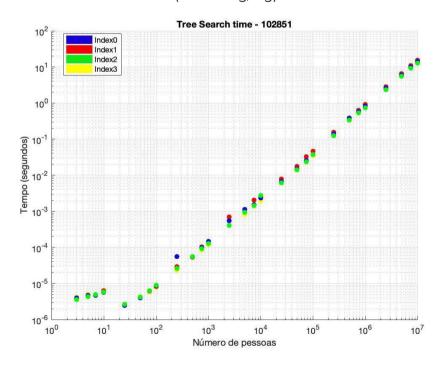


## Gráficos

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

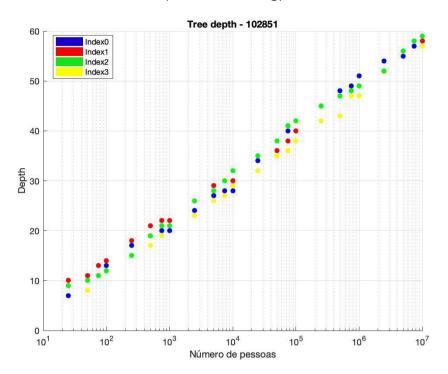


102851 - Relação Tree Search Time / Number of people (escala: log/log)

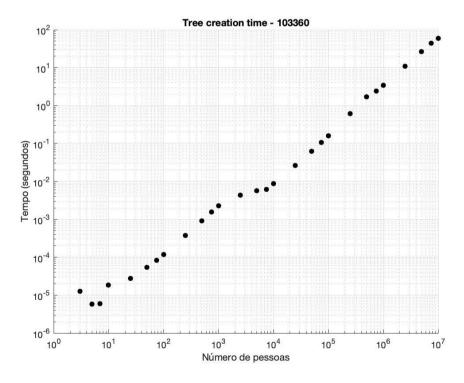


102851 - Relação Tree Depth / Number of people

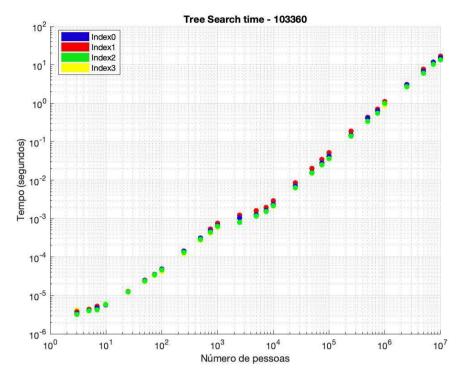
### (escala: linear/log)



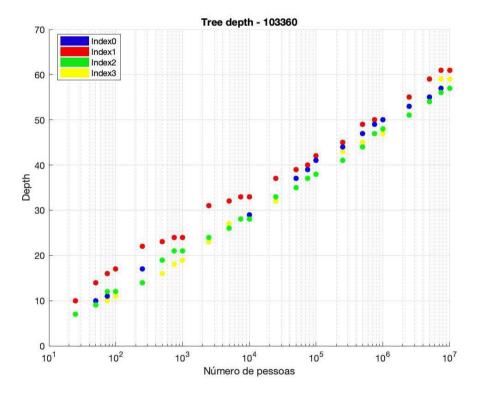
103360 - Relação Tree Creation Time / Number of people (escala: log/log)



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



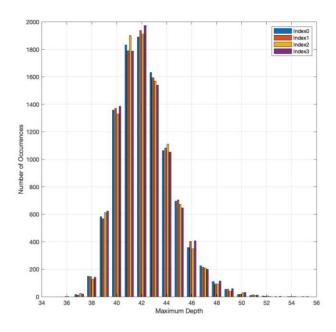
103360 - Relação Tree Depth / Number of people (escala: linear/log)



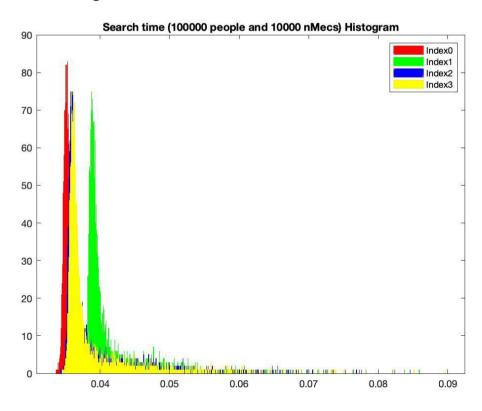
Histogramas

Para cada histograma foram geradas 100000 pessoas e 10000 nMecs.

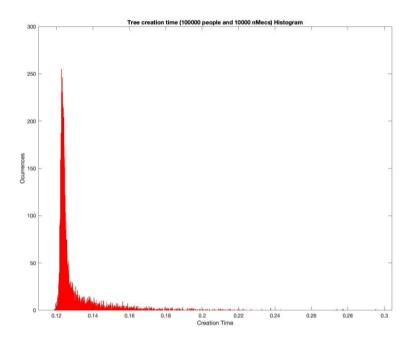
Histograma - número de ocorrências / tree depth



Histograma - número de ocorrências / tree search time



Histograma - número de ocorrências / tree creation 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.

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)

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()'.

```
//
// tree insertion routine (place your code here)
//
// tree insert(tree_node_t **roots, tree_node_t *node, int main_index)

/* If the tree is empty, the tree will be to equal the node and will not return any value */
if (roots[main_index] == NULL)

/* If the tree is empty, the tree will be to equal the node and will not return any value */
if (roots[main_index] == NULL)

/* roots[main_index] = node;
return;

/* Otherwise, recur down the tree */
if (compare_tree_nodes(roots[main_index], node, main_index) > 0)
tree_insert(roots[main_index] -> left, node, main_index);
else if (compare_tree_nodes(roots[main_index], node, main_index) < 0)
tree_insert(roots[main_index] -> right, node, main_index);
else

/* fprintf(stderr, "ERROR generating tree, two equal people!\n");
fprintf(stderr, "Please chose another student number.\n");
exit(1);
}

/* Printf(stderr, "Please chose another student number.\n");
exit(1);
}
```

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
      // tree depdth
99
100
101
      int tree depth(tree node t *root, int main index)
102
103
          if (root == NULL)
104
105
              return 0;
          int l = tree depth(root->left[main index], main index);
107
108
          int r = tree_depth(root->right[main_index], main_index);
109
110
          if (l > r)
111
              return l + 1;
112
113
          return r + 1;
114
```

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

```
116
     // list, i,e, traverse the tree (place your code here)
117
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("
127
                         name ----- %s\n", root->name);
                         zip code ----- %s\n", root->zip code);
             printf("
128
                         telephone number --- %s\n", root->telephone_number);
129
             printf("
                         cc ----- %s\n", root->cc);
             printf("
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.

```
135
     // list the people with a given zip code
136
138
139
     int c2 = 1; // global variable
     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++);
                 printf("
                            name ----- %s\n", root->name);
                 printf("
                            zip code ----- %s\n", root->zip_code);
                 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);
156
                 findZipCode(root->right[1], zip code);
158
```

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 search time (10000000 people, index 3): 1.387e+01s
Tree depth for index 0: 59 (done in 5.332e-01s)
Tree depth for index 2: 57 (done in 5.934e-01s)
Tree depth for index 3: 57 (done in 5.988e-01s)
leonardodsf@leonardodsf-Creator-15M-A9SD:~/Documents/GitHub/LEI/AED/Multi-ordered_trees$ ./multi_ordered_tree 103360 4 -list0
Tree creation time (4 people): 1.535e-06s
Tree search time (4 people, index 0): 7.660e-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 search time (4 people, index 3): 7.400e-07s
Tree depth for index 0: 3 (done in 5.210e-07s)
Tree depth for index 2: 3 (done in 4.800e-07s)
Tree depth for index 3: 3 (done in 4.510e-07s)
List of people:
 List of people:
Person #1
          son #1
name ------- Ethelyn Barry
zip code ------ 95828 Sacramento (Sacramento county)
telephone number --- 9030 741 502
cc ------------ 76297810
  Person #2
          name ------ Lisa Hernandez
zip code ------ 11368 Corona (Queens county)
telephone number --- 8327 821 401
cc ------ 53708478
  Person #3
           son #3
name ------ Mark Jenkins
zip code ------ 60629 Chicago (Cook county)
telephone number --- 4438 898 422
cc ------ 32264209
  cc ---
Person #4
          son #2
name ------ Michael Ramirez
zip code ------ 33027 Hollywood (Broward county)
telephone number --- 9658 733 399
cc ------ 34477417
          oname ------ Mark Jenkins
zip code ------ 60629 Chicago (Cook county)
telephone number 4438 898 422
cc ------ 32264209
  Person #4
           son #4
name ------ Ethelyn Barry
Zip code ------ 95828 Sacramento (Sacramento county)
telephone number --- 9030 741 502
cc ------ 76297810
```

```
-A9SD:~/Documents/GitHub/LEI/AED/Multi-ordered trees$ ./multi ordered tree 103360 4 -list2
leonardodsf@leonardodsf-Creator-15M-A9SD:~/Docume
Tree creation time (4 people): 6.310e-06s
Tree search time (4 people, index 0): 5.813e-06s
Tree search time (4 people, index 1): 5.055e-06s
Tree search time (4 people, index 2): 4.614e-06s
Tree search time (4 people, index 3): 4.899e-06s
Tree depth for index 0: 3 (done in 3.150e-06s)
Tree depth for index 1: 4 (done in 3.043e-06s)
Tree depth for index 2: 3 (done in 6.882e-06s)
Tree depth for index 3: 3 (done in 2.491e-06s)
 List of people:
Person #1
 name ------ Lisa Hernandez
zip code ------ 11368 Corona (Queens county)
telephone number --- 8327 821 401
cc ------ 53708478
 Person #3
         name ------ Ethelyn Barry
zip code ----- 95828 Sacramento (Sacramento county)
telephone number --- 9030 741 502
cc ------ 76297810
 Person #4
          nom ## romane ----- Michael Ramirez
zip code ------ 33027 Hollywood (Broward county)
telephone number --- 9658 733 399
          name --
          cc ----- 34477417
leonardodsf@leonardodsf-Creator-15M-A9SD:~/Documents/GitHub/LEI/AED/Multi-ordered_trees$ ./multi_ordered_tree 103360 4 -list3
Tree creation time (4 people): 6.001e-06s
Tree search time (4 people, index 0): 4.489e-06s
Tree search time (4 people, index 1): 3.823e-06s
Tree search time (4 people, index 2): 3.683e-06s
Tree search time (4 people, index 3): 3.620e-06s
Tree search time (4 people, index 3): 3.620e-06s
Tree depth for index 0: 3 (done in 2.450e-06s)
Tree depth for index 1: 4 (done in 2.281e-06s)
Tree depth for index 2: 3 (done in 2.193e-06s)
Tree depth for index 3: 3 (done in 2.297e-06s)
List of people:
 List of people:
Person #1
name -----
          ion #1
name ------ Mark Jenkins
zip code ------ 60629 Chicago (Cook county)
telephone number --- 4438 898 422
cc ------- 32264209
         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
          name ----
          name ------ Ethelyn Barry
zip code ------ 95828 Sacramento (Sacramento county)
telephone number --- 9030 741 502
          cc ----- 76297810
       Betty Le
11206 Brooklyn (Kings county)
4569 063 154
72406663
          zip code -----
telephone number
           n #2
name -
                                               Jacalyn Kim
11206 Brooklyn (Kings county)
5122 296 196
31927034
          zip code -----
telephone number
           n #3
                                               Bobbie Cole
11206 Brooklyn (Kings county)
5833 122 918
76578885
          zip code -----
telephone number
           n #4
name -
                                               Andrew Murray
11206 Brooklyn (Kings county)
7475 441 241
01606069
          zip code -----
telephone number
           n #5
                                               Gene Ahmed
11206 Brooklyn (Kings county)
7505 344 933
07875628
          zip code -----
telephone number
```

- Student\_number = 102851 :

```
leonardodsf@leonardodsf-Creator-15M-A9SD:~/Documents/Git
Tree creation time (10000000 people): 5.949e+01s
Tree search time (10000000 people, index 0): 1.478e+01s
Tree search time (10000000 people, index 1): 1.797e+01s
Tree search time (10000000 people, index 2): 1.619e+01s
Tree search time (10000000 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)
                                                                                                       /Documents/GitHub/LEI/AED/Multi-ordered_trees$ ./multi_ordered_tree 102851 10000000
 leonardodsf@Leonardodsf-Creator-15M-A950:~/Documents/GitHub/LEI/AED/Multi-ordered_trees$ ./multi_ordered_tree 102851 4 -list0
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 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:
   List of people:
Person #1
          son #1
name ------- Clair Rodriguez
zip code ------ 92804 Anaheim (Orange county)
telephone number --- 1963 704 614
cc ------ 07000733
   Person #2
          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
         name ------- Shannon Christian
zip code ------ 90201 Bell (Los Angeles county)
telephone number --- 6189 678 649
cc ------ 48368067
son #1
name ------ Mary Paul
zip code ------ 37211 Nashville (Davidson county)
telephone number --- 3146 767 353
cc ------- 78085756
   Person #2
          Person #3
          on wo
name ------- Shannon Christian
zip code ------ 90201 Bell (Los Angeles county)
telephone number --- 6189 678 649
cc ------ 48368067
   Person #4
           name ----
                                               ----- Clair Rodriguez
           Zip code ----- 92804 Anaheim (Orange county)
telephone number --- 1963 704 614
cc ----- 07000733
```

```
leonardodsf@leonardodsf-Creator-15M-A9SD:~/Docume
Tree creation time (4 people): 7.051e-06s
Tree search time (4 people, index 0): 5.183e-06s
Tree search time (4 people, index 1): 4.269e-06s
Tree search time (4 people, index 2): 4.242e-06s
Tree search time (4 people, index 3): 4.103e-06s
Tree depth for index 0: 4 (done in 2.684e-06s)
Tree depth for index 1: 4 (done in 2.605e-06s)
Tree depth for index 2: 4 (done in 2.576e-06s)
Tree depth for index 3: 3 (done in 2.526e-06s)
List of people:
                                                                   A9SD:~/Documents/GitHub/LEI/AED/Multi-ordered trees$ ./multi ordered tree 102851 4 -list2
 Person #2
      son #2
name ------ Mary Paul
zip code ------ 37211 Nashville (Davidson county)
telephone number --- 3146 767 353
cc ------ 78085756
 Person #3
       son #3
name ------ Shannon Christian
zip code ------ 90201 Bell (Los Angeles county)
telephone number --- 6189 678 649
cc ------ 48368067
  Person #4
      son #4
name ------- Rita Hensley
zip code ------ 89031 North Las Vegas (Clark county)
telephone number --- 9413 320 621
cc ------- 87162912
name ----
                                      ---- Shannon Christian
        zip code ----- 90201 Bell (Los Angeles county)
telephone number --- 6189 678 649
cc ------ 48368067
 cc ---
Person #3
        name
 rame ------- Mary Paut
zip code ------- 37211 Nashville (Davidson county)
telephone number --- 3146 767 353
cc -------------- 78085756
Person #4
        name ------ Rita Hensley
zip code ------ 89031 North Las Vegas (Clark county)
telephone number --- 9413 320 621
cc ------- 87162912
```

#### 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/