



• U • C •

FCTUC

FACULDADE DE CIÊNCIAS  
E TECNOLOGIA  
UNIVERSIDADE DE COIMBRA

# AED

Ano letivo: 2016/2017  
RELATÓRIO DO PROJETO 1

**Gonçalo Amaral** 2015249122

**Jorge Martins** 2015199193

**Cristóvão Gouveia** 2015260507

## **Introdução e objetivos:**

No passado dia 17 de fevereiro de 2017 foi-nos proposto pela cadeira de AED que criássemos três tipos de estruturas diferentes. Essas estruturas teriam toda a mesma informação relativa há evolução do acesso a redes elétricas pela população mundial. Os dados foram oferecidos pelo Banco Mundial que é uma organização internacional que tem como objetivos principais acabar com a pobreza extrema e promover uma prosperidade partilhada.

O principal objetivo deste projeto seria que as três estruturas criadas fossem o mais otimizadas possível para o problema proposto. Para a realização do projeto utilizamos a linguagem de programação Python. Todas as estruturas foram criadas pelo grupo com base nos exemplos disponibilizados.

Estruturas escolhidas:

- Lista circular duplamente ligada
- Avl de países
- Dupla Avl (países e anos)

Iremos descrever a implementação de cada estrutura em mais pormenor posteriormente neste documento.

Para todas as estruturas criadas foram implementadas as funcionalidades listadas:

- Inserção (pais, sigla, ano, percentagem)
- Procura da percentagem de um ano e pais específicos
- Procura da percentagem de um ano e sigla específicos
- Procura das percentagens de um ano específico
- Procura das percentagens de um país específico
- Procura das percentagens de uma sigla específica
- Edição da percentagem de um ano e pais específicos
- Edição da percentagem de um ano e sigla específicos
- Remoção da percentagem de um ano e pais específicos
- Remoção da percentagem de um ano e sigla específicos
- Remoção de um ano específico
- Remoção um país específico
- Remoção de uma sigla específica

## Descrição detalhada da implementação das estruturas

### Lista circular duplamente ligada

Como o grande objetivo do projeto era criar estruturas o mais otimizadas possível, não poderíamos usar uma simples lista ligada para os países e cada nó dessa lista com outra lista ligada de anos. Por esta razão decidimos usar uma função hash para converter todos os nomes em números para uma rápida comparação. Esta função hash (predefinida na linguagem, injetiva) recorre ao algoritmo sha1 que gera um número em hexadecimal baseado no objeto passado que é convertido para inteiro pela função `hashCode` (implementada pelo grupo). Convertemos a lista ligada de países numa lista duplamente ligada circular e um header que tem como função ser o intermediário na comunicação com a estrutura. Este header contém a mediana atual dos números hash da lista de países, permitindo-nos escolher se percorremos a lista do fim para o início, se o número hash do país a procurar for maior que a mediana, ou de forma natural, do início para o fim se for menor ou igual à mediana (complexidade temporal desce de  $O(n)$  para  $O(n/2)$ , pois apenas temos que percorrer metade da lista). Apesar de termos melhorado a pesquisa por país, a pesquisa por sigla não é a melhor, e por falta de análise matemática aliado de uma má gestão de tempo, criamos um tradutor de siglas para países sendo este um AVL tendo como valor de comparação o valor hash da sigla, acabámos por piorar a implementação original que seria  $O(n)$  para  $O(\log n * n/2)$ , que apenas é menor que para  $n$  menor a 4. Finalmente em vez de usarmos uma lista ligada para os anos, implementamos uma AVL (complexidade temporal desta subestrutura desce de  $O(n)$  para  $O(\log n)$ )

Listamos as complexidades temporais após as mudanças da estrutura em todas as operações, sendo  $n$ , o número de países e  $m$  o número de anos, na página seguinte

- Inserção (pais, sigla, ano, percentagem)
  1. Original:  $O(n*m)$
  2. Alterada:  $O((n/2)*(\log m))$
- Procura da percentagem de um ano e pais específicos
  1. Original:  $O(n*m)$
  2. Alterada:  $O((n/2)*(\log m))$
- Procura da percentagem de um ano e sigla específicos
  1. Original:  $O(n*m)$
  2. Alterada:  $O((\log n)*(n/2)*(\log m))$
- Procura das percentagens de um ano específico
  1. Original:  $n * O(m)$
  2. Alterada:  $n * O(\log m)$
- Procura das percentagens de um país específico
  1. Original:  $O(n) * m$
  2. Alterada:  $O(n/2) * m$
- Procura das percentagens de uma sigla específica
  1. Original:  $O(n) * m$
  2. Alterada:  $O((\log n)*(n/2)) * m$
- Edição da percentagem de um ano e pais específicos
  1. Original:  $O(n*m)$
  2. Alterada:  $O((n/2)*(\log m))$
- Edição da percentagem de um ano e sigla específicos
  1. Original:  $O(n*m)$
  2. Alterada:  $O((\log n)*(n/2)*(\log m))$

- Remoção da percentagem de um ano e país específicos
  1. Original:  $O(n*m)$
  2. Alterada:  $O((n/2)*(log m))$
- Remoção da percentagem de um ano e sigla específicos
  1. Original:  $O(n*m)$
  2. Alterada:  $O((log n)*(n/2)*(log m))$
- Remoção de um ano específico
  1. Original:  $n * O(m)$
  2. Alterada:  $n * O(log m)$
- Remoção um país específico
  1. Original:  $O(n)$
  2. Alterada:  $O(n/2)$
- Remoção de uma sigla específica
  1. Original:  $O(n)$
  2. Alterada:  $O((log n)*(n/2))$

No início do planeamento das três estruturas de dados reconhecemos que esta iria ser a mais lenta de todas.

## Avl de países

Devido à eficiência natural de uma árvore binária de pesquisa balanceada, não realizamos muitas alterações à mesma, tendo obtido uma Avl de países em que o método de comparação seria através do número hash do país, tal como na estrutura referida anteriormente. Pelas mesmas razões anteriormente referidas usamos um tradutor de siglas para países.

Listamos as complexidades temporais após as mudanças da estrutura em todas as operações, sendo  $n$ , o número de países e  $m$  o número de anos. Reconhecemos a estrutura original a mesma implementação mas sem o tradutor de siglas para países, tendo cada nos o valor hash da sigla:

- Inserção (país, sigla, ano, percentagem)
  1. Original:  $O(\log n * \log m)$
  2. Alterada:  $O(\log n * \log m)$
- Procura da percentagem de um ano e país específicos
  1. Original:  $O(\log n * \log m)$
  2. Alterada:  $O(\log n * \log m)$
- Procura da percentagem de um ano e sigla específicos
  1. Original:  $O(n * \log m)$
  2. Alterada:  $O(\log n * \log n * \log m)$
- Procura das percentagens de um ano específico
  1. Original:  $n * O(\log m)$
  2. Alterada:  $n * O(\log m)$
- Procura das percentagens de um país específico
  1. Original:  $O(\log n) * m$
  2. Alterada:  $O(\log n) * m$
- Procura das percentagens de uma sigla específica
  1. Original:  $O(n) * m$
  2. Alterada:  $O((\log n) * (\log n) * (n/2)) * m$
- Edição da percentagem de um ano e país específicos
  1. Original:  $O(\log n * \log m)$
  2. Alterada:  $O(\log n * \log m)$

- Edição da percentagem de um ano e sigla específicos
  1. Original:  $O(n * \log m)$
  2. Alterada:  $O(\log n * \log n * \log m)$
- Remoção da percentagem de um ano e país específicos
  1. Original:  $O(\log n * \log m)$
  2. Alterada:  $O(\log n * \log m)$
- Remoção da percentagem de um ano e sigla específicos
  1. Original:  $O(n * \log m)$
  2. Alterada:  $O(\log n * \log n * \log m)$
- Remoção de um ano específico
  1. Original:  $n * O(\log m)$
  2. Alterada:  $n * O(\log m)$
- Remoção um país específico
  1. Original:  $O(\log n)$
  2. Alterada:  $O(\log n)$
- Remoção de uma sigla específica
  1. Original:  $O(n)$
  2. Alterada:  $O(\log n * \log n)$

No início do planeamento das três estruturas de dados reconhecemos que esta iria ser uma das mais rápidas

### Dupla Avl (países e anos)

Reparamos que nas outras implementações apesar de pouparmos a complexidade espacial, na pesquisa de um ano específico a complexidade temporal aumentava pois tinhas que percorrer cada país e ver se o ano que procuramos existe no mesmo. Motivados por este problema e por sugestão do

professor das PLs decidimos criar uma estrutura que sacrificasse complexidade espacial por complexidade temporal. A estrutura é composta por um header que tem a função de intermediário na interação com as duas estruturas extra, uma Avl de anos em que cada nó contém um Avl de países onde a comparação é feita e mais uma vez o número hash do país, em que cada nó desta estrutura por sua vez contém um nó com a percentagem em comum com a outra Avl, Avl dos países. A segunda Avl contém os países em que a comparação é feita através do número hash do país, e cada nó desta estrutura contém uma Avl com os anos em que por sua vez, cada nó tem um nó com a percentagem em comum com a outra Avl, a Avl dos anos. Teríamos como objetivo usar este nó comum para poupar um pouco de espaço de maneira a que as percentagens não estivessem duplicadas e a remoção de um ano e país específica apenas requerem-se percorrer uma AVL e a sua sub Avl. Também pretendíamos que este nó como percentagem contivesse referências para as Avls que o continham de maneira a que a remoção apenas necessitaria de percorrer um Avl e a sua sub Avl para remover um país ou um ano das duas Avls pois teríamos uma referência a outra Avl no nó. Para além desta funcionalidade, esta estrutura também permitiria fazer remoções rápidas, de complexidade temporal igual à estrutura anterior (Avl de países), num ambiente multithread, percorrendo cada thread a uma Avl. Nenhuma destas duas últimas funcionalidades foram implementadas. Acabamos por sacrificar complexidade temporal na inserção e remoção pois é necessário remover os dados de duas estruturas.

Listamos as complexidades temporais após as mudanças da estrutura em todas as operações, sendo  $n$ , o número de países e  $m$  o número de anos. Neste caso não reconhecemos nenhuma estrutura original:

- Inserção (país, sigla, ano, percentagem)
  1.  $O(2 * \log n * \log m)$
- Procura da percentagem de um ano e país específicos
  1.  $O(\log n * \log m)$
- Procura da percentagem de um ano e sigla específicos
  1.  $O(\log n * \log n * \log m)$



- Procura das percentagens de um ano específico
  1.  $O(\log m) * n$
- Procura das percentagens de um país específico
  1.  $O(\log n) * m$
- Procura das percentagens de uma sigla específica
  1.  $O(\log n * \log n) * m$
- Edição da percentagem de um ano e país específicos
  1.  $O(\log n * \log m)$
- Edição da percentagem de um ano e sigla específicos
  1.  $O(\log m * \log n * \log n)$
- Remoção da percentagem de um ano e país específicos
  1.  $O(\log n * \log m)$
- Remoção da percentagem de um ano e sigla específicos
  1.  $O(\log n * \log n * \log m)$
- Remoção de um ano específico
  1.  $O(\log m + n * \log m)$
- Remoção um país específico
  1.  $O(\log n + m * \log n)$
- Remoção de uma sigla específica
  1.  $O(\log n * \log n + m * \log n * \log n)$

No início do planeamento das três estruturas de dados reconhecemos que esta teria grande potencial a nível de procuras e pior desempenho em inserções e remoções.

## Benchmarks

Para simplificação:

1. Lista circular duplamente ligada
2. Avl países
3. Dupla Avl (países e anos)

Os parâmetros usados nos benchmarks são iguais em todas as estruturas, incluído a máquina de testes e dia.

- Inserção (Portugal, PRT, 1990, 100)
  1. 0.000543067843418
  2. 0.000241929799753
  3. 0.000226506346803
- Procura da percentagem de um ano e país específicos (1990, Portugal)
  1. 0.00021462487592
  2. 2.47986990301e-05
  3. 2.62338973088e-05
- Procura da percentagem de um ano e sigla específicos (1990, PRT)
  1. 0.000266927664742
  2. 7.27174244567e-05
  3. 7.6481419865e-05
- Procura das percentagens de um ano específico (1990)
  1. 0.0021080659025
  2. 0.00267284183733
  3. 0.00186176019121

- Procura das percentagens de um país específico (Portugal)
  1. 0.000234525445441
  2. 3.98692375518e-05
  3. 4.6615370965e-05
  
- Procura das percentagens de uma sigla específica (PRT)
  1. 0.000288505187798
  2. 8.92243535355e-05
  3. 9.57347762676e-05
  
- Edição da percentagem de um ano e país específicos  
(1990, Portugal, 100)
  1. 0.000224479279309
  2. 3.40164811703e-05
  3. 3.06746632507e-05
  
- Edição da percentagem de um ano e sigla específicos  
(1990, PRT, 100)
  1. 0.000277155084929
  2. 8.26903902355e-05
  3. 7.94317545348e-05
  
- Remoção da percentagem de um ano e país específicos  
(1990, Portugal)
  1. 0.000222782888863
  2. 3.4092953357e-05
  3. 3.0976171067e-05
  
- Remoção da percentagem de um ano e sigla específicos  
(1990, PRT)
  1. 0.000274747320146
  2. 8.31622275089e-05
  3. 8.01462343409e-05

- Remoção de um ano específico (1990)
  1. 0.00383271284943
  2. 0.00450440094939
  3. 1.32255846324e-05
- Remoção um país específico (Portugal)
  1. 0.000568160203408
  2. 8.68025309576e-05
  3. 4.42356663173e-05
- Remoção de uma sigla específica (PRT)
  1. 0.000616760690302
  2. 0.000135148175259
  3. 9.18076328004e-05

## Conclusões

Grande parte dos resultados foram os esperados, apenas a inserção e a remoção não se comportaram como esperado. Esperávamos que a terceira estrutura fosse pior na remoção e na inserção que a segunda mas verificamos que em todos os casos isto não se verifica. Na remoção de um ano na primeira estrutura, era esperado que tivesse pontuações semelhantes à segunda lista, o que não ocorreu. Na remoção através de sigla a segunda estrutura obteve valores na ordem de grandeza da primeira estrutura.

Apesar de não ser o esperado concluímos que a terceira estrutura foi a mais rápida nos testes feitos e a pior seria a primeira, o que neste caso era esperado.