

1º Projecto Prático

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Algoritmos e Estruturas de Dados**

2016/2017

João Afonso Póvoa Marques – 2015227041

José Miguel Saraiva Monteiro – 2015235572

Leonardo Machado Alves Vieira – 2015236155

**Índice**

[**Índice** 2](#_Toc478235854)

[**1 – Introdução** 2](#_Toc478235855)

[**2 – Estruturas** 3](#_Toc478235856)

[Estrutura 1: 3](#_Toc478235857)

[Estrutura 2: 4](#_Toc478235858)

[Estrutura 3: 4](#_Toc478235859)

[**3 - Operações Implementadas** 5](#_Toc478235860)

[Estrutura 1: 5](#_Toc478235861)

[Estrutura 2: 5](#_Toc478235862)

[Estrutura 3: 6](#_Toc478235863)

[**4 – Casos de Teste** 6](#_Toc478235864)

[Estrutura 1: 6](#_Toc478235865)

[Estrutura 2: 6](#_Toc478235866)

[Estrutura 3: 6](#_Toc478235867)

[**5 – Comparações** 6](#_Toc478235868)

[**6 – Conclusões** 6](#_Toc478235869)

# **1 – Introdução**

Neste projeto foi nos proposta a tarefa de criar 3 estruturas de dados suportadas em abstrações diferentes que possam armazenar e manipular a informação disponibilizada. No contexto deste projeto, informação era fornecida através do ficheiro dados*.csv* sobre a evolução do acesso a redes elétricas pela população mundial desde 1960 até 2016.

Sobre todas estas estruturas foram definidas e implementadas funções de pesquisa, inserção, edição e remoção para ambos anos e países, que foram adaptadas e melhoradas de acordo com a estrutura e que podem ser facilmente utilizadas através de um interface funcional (igual para todas as estruturas) que é apresentado ao utilizador ao iniciar o programa.

Também foram criadas em todas as estruturas, com uma vertente de teste, funções auxiliares para realização de *benchmarks* em que são feitos testes extremos e contabilizado o tempo de resposta para cada um.

Todo o projecto foi realizado em python3.

As estruturas implementadas foram as seguintes:

* Listas Duplamente Ligadas de Listas Duplamente Ligadas
* Stacks de Stacks
* Árvores AVL de Árvores AVL

# **2 – Estruturas**

## Estrutura 1:

Stacks de Stacks

[Implementação base retirada das estruturas apresentadas nas aulas de AED]

A primeira estrutura implementada foi um stack de stacks. Esta estrutura foi escolhida devido à sua facilidade de implementação, e também pelo facto de já ter existido contacto com estas estruturas. As principais desvantagens desta estrutura são a pesquisa e o acesso aos seus valores, operações que possuem uma complexidade de Θ(n) (embora *n* seja minimizado quando trabalhando com dados ordenados), mas que em operações de adição e remoção a complexidade é meramente Θ(1), o que se torna num ponto a favor desta estrutura.

No contexto deste problema, para facilitar a pesquisa em anos, todas as stacks de anos estão ordenadas por ano (2000 está mais acima na pilha do que 2016) isto é, não é necessário percorrer a stack inteira se o valor estiver logo no inicio, melhorando assim a performance. O mesmo não se verifica no stack dos países, o que poderá diminuir a performance do programa. É importante também referir que

Poderia no entanto, ter sido implementado algum tipo de estrutura auxiliar para melhorar o desempenho,

## Estrutura 2:

Listas Duplamente Ligadas de Listas Duplamente Ligadas

[Implementação base retirada das estruturas apresentadas nas aulas de AED]

Uma das estruturas implementadas foi uma Lista Duplamente Ligada de Listas Duplamente Ligadas. Esta estrutura foi escolhida por diversas razões, sendo uma delas a relativa facilidade de implementação e o facto de já termos tido contacto com este tipo de estruturas previamente. Para além disso foi também tido em conta a complexidade *BigO* das várias funções a implementar sobre esta estrutura. Assim sendo, o facto de esta ter uma complexidade de Θ(1) para as operações (isoladas) de inserção e remoção foi bastante apelativo. Isso faz com que as operações referidas sejam uma vantagem desta estrutura.

No entanto, como no contexto do problema não faz sentido haver duplicação de dados e faz sentido alguns valores estarem ordenados algumas desvantagens são criadas. Isto é, para cada operação de adição é necessário uma primeira pesquisa para garantir que o valor a adicionar não se encontra previamente na lista. Se o valor estiver de alguma forma ordenado, será escusado a pesquisa completa da lista, no entanto com valores desordenados terá sempre que iterar por todos os nós da lista à procura de um com o valor igual. Com isto, facilmente se entende que uma das desvantagens desta estrutura será a pesquisa, pois está subentendida em todas as outras operações e tem uma complexidade de Θ(n) (sendo o *n* minimizado quando operando sobre valores ordenados).

Poderiam ter sido implementadas estruturas auxiliares para melhorar tempos e complexidades de certas operações. Por exemplo, adicionar uma estrutura auxiliar para cada ano em que se iria guardar o par país, percentagem desse país nesse ano. Assim, seria possível obter a informação de todos os países para um ano de uma maneira muito mais “económica” e sem piorar a complexidade espacial pois toda a informação aparentemente duplicada iria apontar para o mesmo sitio da memória.

## Estrutura 3:

Árvore AVL de Árvores AVL

[Implementação base retirada das estruturas apresentadas nas aulas de AED]

A terceira estrutura a ser implementada foi uma árvore AVL de árvores AVL. A escolha desta estrutura teve como base a ideia de que seria uma das melhores implementações para os dados em questão. Esta estrutura possibilita-nos operações (isoladas) de inserção, remoção e pesquisa com uma complexidade Θ(log n). Além disto todas as adições/remoções não fazem com que a arvore perca o seu ordenamento, pois esta esta sempre organizada (por ordem alfabética e crescente de anos no nosso caso especifico) e sempre balanceada, o que facilita operações de pesquisa. Estas características de facilidade de acesso e baixa complexidade fazem com que esta estrutura tenha vantagens, num aspeto mais geral, às apresentadas anteriormente.

Como auxiliar a esta estrutura, para evitar ter de criar uma nova árvore organizada pela *tags* e para evitar ter de percorrer toda a árvore para encontrar uma *tag* especifica, foi criado ainda um dicionário com os nomes de todos os países (*key*) e a *tag*  correspondente (*value*). Como o acesso a este tipo de estruturas é de Θ(1) não irá afetar a complexidade da pesquisa.

# **3 - Operações Implementadas**

## Estrutura 1:

Stacks de Stacks

* Pesquisa:

Como foi referido previamente, a pesquisa é um dos pontos fracos desta estrutura de dados, tendo em conta que se a estrutura não estiver ordenada, a complexidade será sempre Θ(n), pois será necessário percorrer a estrutura toda até encontrar o valor desejado. No entanto, na pilha de anos, os valores encontram-se ordenados, o que irá fazer com que o *n* para esta stack seja reduzido. Ficamos assim com Θ(1) como limite inferior de complexidade e Θ(n) como limite superior de complexidade.

Foi também implementada uma função de pesquisa na stack dos anos, em que fornecendo como input dois valores, um mínimo e um máximo, irão ser impressos todos os anos com valores compreendidos entre o mínimo e o máximo. Esta função de pesquisa terá uma complexidade de Θ(n), pois será sempre necessário percorrer a stack inteira para saber quais os anos com valores compreendidos entre os que são dados à função.

Para além da função de pesquisa de valores num certo intervalo, foi também implementada uma função que pesquisa se existe informação de um determinado ano, em todos os países. Como para saber se existe esta informação para cada país existente é necessário percorrer a stack dos países, e posteriormente a stack dos anos à procura do valor, a complexidade desta função será horrivel, sendo esta igual a Θ(n²) num limite superior, ou então Θ(n) num limite inferior. Note-se que estes valores de complexidade poderiam ser melhorados com a utilização de uma estrutura auxiliar com todos os valores para cada ano, de todos os países (caso a informação exista). Esta estrutura seria um “espelho” da estrutura implementada.

Na stack de países foi apenas implementada uma função de pesquisa de complexidade Θ(n), que desencadeia a maior parte de métodos, quer para adicionar, pesquisar, remover ou até imprimir informação.

* Inserção:

Para esta estrutura foram implementados dois tipos de inserção. Para inserção de novos elementos à stack de países (desordenada), a função limita-se a adicionar as informações que o utilizador fornece (nome do país e código do mesmo, neste caso) e adicionar ao topo da pilha. É de notar que esta implementação não verifica se há duplicação de dados. Tal operação irá ter uma complexidade Θ(1).

A outra implementação realizada, neste caso para a stack de anos, é uma implementação que preza manter a informação ordenada e não duplicada, isto é, cada vez que se adiciona um novo elemento, acontece uma pesquisa prévia à stack existente, que primeiramente serve para adicionar o novo valor na posição pretendida, e, caso este já exista, para não permitir a sua duplicação. A complexidade de tal operação, será Θ(1) num caso ótimo, e Θ(n) no pior caso possível, em que teremos que adicionar o valor no fim.

* Edição:

Nesta estrutura, temos duas possibilidades de edição. Editar o ano de um país, mantendo o seu valor, ou editar o valor de um ano de um país, mantendo o ano.

Para a edição da percentagem, é simplesmente feita uma pesquisa na stack de países, à procura do país, e posteriormente executada uma pesquisa na stack dos anos, até encontrar o ano do qual se pretende editar o valor. A complexidade de tal operação é Θ(n²) no pior cenário, tendo em conta que a função para pesquisar o ano encontra-se dentro do ciclo de pesquisa do país.

Para editar o ano, o cenário é bastante parecido, primeiro existe um ciclo que procura o país pretendido, depois disso chama a função para editar o valor do ano pretendido, que irá chamar a função para adicionar um novo ano. Sendo assim, a complexidade da função será Θ(n³) no pior cenário possível, tornando esta estrutura possivelmete na pior para esta operação, devido à sua complexidade.

* Remoção:

A implementação desta operação foi bastante semelhante tanto para a stack de países, como para a stack de anos. Tal como as outras operações, esta função requer uma pesquisa prévia pelo elemento que pretende ser removido, operação tal que na pior das hipoteses terá uma complexidade de Θ(n). No caso dos países, no melhor cenário possível, iremos ter uma complexidade de Θ(1), caso o elemento que se pretenda remover seja logo o primeiro. No caso dos anos, no melhor cenário possível iremos ter uma complexidade de Θ(n), enquanto que na pior das hipóteses seria Θ(n²) pois teria que percorrer o stack dos anos à procura do elemento desejado, posteriormente a ter percorrido a stack dos países.

## Estrutura 2:

Listas Duplamente Ligadas de Listas Duplamente Ligadas

* Pesquisa:

Como referido anteriormente a pesquisa é um dos

pontos fracos desta estrutura.

Num caso base, a pesquisa teria sempre uma complexidade de Θ(n), pois teria de correr sempre toda a lista à procura do nó desejado. No entanto, na lista de anos, os anos encontram-se ordenados, o que faz com que para esta lista, o *n* seja reduzido. Para este caso temos num limite inferior uma complexidade de pesquisa de Θ(1) e num limite superior a complexidade base de Θ(n).

Foi implementada uma função de pesquisa na lista dos anos, que permite a pesquisa de anos dado um limite inferior e superior de percentagens. Esta função tem uma complexidade de Θ(n) pois esta vai correr toda a lista imprimindo apenas os anos cujo valor esteja dentro dos limites.

Para além disso, foi também implementado uma pesquisa *inversa* à primeira referida. Isto é, o utilizador dá *input* de um ano, e é lhe apresentado o valor de todos os países para esse ano. Esta função é obviamente má a nivel de complexidade, pois como tem de correr sempre a lista de países e dentro de cada nó, ainda correr a lista de anos, terá uma complexidade de Θ(n2) num limite superior e O(n) num limite inferior. Note-se que isto poderia ser melhorada

com uma estrutura auxiliar com as percentragens para todos os paises num certo ano. Isto seria no fundo uma estrutra *“espelho”* da que foi de facto implementada.

Dentro da lista de países foi apenas implementada uma pesquisa simples através de uma função Θ(n) que corre a lista toda à procura do nó correto que depois é retornado para ser usado noutras funções.

* Inserção:

Também já foi referida a vantagem da inserção neste tipo de estruturas.

Foram implementadas dois tipos de inserções. Para a inserção na lista (desordenada) de países, recorre-se primeiramente a uma pesquisa pelo nome e *tag* do país a inserir, para garantir que o país não está previamente na lista evitando assim duplicação de informação. Depois desta pesquisa, e em caso negativo de duplicação, procede-se à inserção do novo nó no final da list. Esta parte da operação é realizada com uma complexidade de Θ(1), visto que a pesquisa inicial retorna o ultimo nó da lista, de maneira a evitar precorrê-la novamente.

Será importante notar que quando é adicionado um novo país, a lista de anos encontra-se a *None.*

A outra implementação foi feita sobre a lista (ordenada) de anos de um país. À semelhança da anterior, também é necessário uma pesquisa prévia (que também retorna um nó) para garantir que não existe duplicação de informação. Porém, esta é uma inserção ordenada, o que implica que o nó retornado nem sempre é o ultimo, mas sim, caso exista, o primeiro com valor maior do que o nó a adicionar. No melhor caso esta inserção é ótima, com uma complexidade de Θ(1). Nos restantes dos casos será de uma complexidade de Θ(n), aumentado o *n* até ao pior caso, que é ter de percorrer toda a lista.

* Edição:

Existem sobre esta estrutura duas possibilidades de edição. Edição de um ano de um país, e de uma percentagem de um ano de um país.

Para a edição da percentagem, simplesmente faz-se uma pesquisa pelo país e pelo ano (2\*Θ(n)), e caso estes existam é feito uma mudança do valor da variável respetiva, sendo isto uma operação Θ(1).

A edição do ano em si já é mais complexa pois não se pode desprezar a ordem da lista de anos e também porque não se pode criar duplicação de anos. Assim sendo, para evitarmos isto tudo, a operação realiza-se nos seguintes paços. Primeiro procurar o país e garantir que ano a alterar existe e que o novo ano para o qual vai ser alterado não existe (3\*Θ(n)). Depois disto, caso “passe” o primeiro passo, iremos guardar numa variável auxiliar a percentagem do ano e remover esse ano. Finalmente é adicionado um novo nó com a percentagem antiga e com o novo ano. Isto é feito para garantir que o nó fica ordenado pois a função de inserção, como foi dito anteriormente, certifica-se disso. Assim conclui-se que esta operação tem uma complexidade de Θ(n) e é bastante dispendiosa.

* Remoção:

A implementação da operação de remoção foi bastante semelhante para a lista de anos e para a lista de países.

Tal como todas as outras, esta operação tambem requer uma pesquisa prévia para encontrar e garantir que existe o ano / país. Logo aqui limitamos a complexidade a (Θ(n)). De seguida, de acordo com a posição do nó pedido, são executadas uma serie de opreções Θ(1) para garantir que todas as ligações *next* e *prev* são bem feitas. Feito isto todas as variáveis do nó de um ano / país são colocadas a *None*.

## Estrutura 3:

Árvore AVL de Árvores AVL

* Pesquisa:

A pesquisa numa AVL pode ser feita tal como seria numa árvore binária não balanceada. No nosso caso em especifico pesquisa pelo nome de um país, a complexidade desta pesquisa seria Θ(log n).

Em adição à pesquisa pelo nome de um país foi ainda adicionada um método de pesquisa por anos. Este método na AVL de anos continuaria a ter complexidade Θ(log n), mas como para a pesquisa em todos em países por esse ano teríamos de percorrer toda a árvore, operação de complexidade Θ(n), a complexidade total da pesquisa de um ano em todos os países será Θ(n\*log n). Esta acaba por ser a operação de maior complexidade da estrutura da forma como está implementada e usada.

Como referido a cima as AVLs de anos tem apenas implementada uma pesquisa por anos de complexidade Θ(log n)

Outra opção de pesquisa implementada é a pesquisa, dentro de um país, de valores entre um certo limite. Esta opção foi implementada de forma semelhante à pesquisa de um ano em todos os países. Será então feita uma pesquisa pelo país desejado e após isso terá de se percorrer toda a árvore de anos. Tal como a pesquisa de um ano em todos os países a complexidade deste método será também Θ(n\*log n).

* Inserção:

O método de inserção utilizado é uma adaptação do código fornecido pelo professor nas aulas adaptado aos novos nós que serão inseridos na árvore. Como tal mantemos a vantagem de todas as inserções serem automaticamente ordenadas, devido a natureza balanceada da AVL. Mais uma vez voltamos a ter a complexidade característica das AVLs de Θ(log n).

* Edição:

Para esta estrutura poderiam ser implementados dois métodos para realizar a edição. Um deles passaria por pesquisar na árvore pelo nó que se desejava alterar e alterar um valor, embora a complexidade desta operação fosse a mesma das pesquisas, Θ(log n), poderíamos perder uma propriedade bastante útil desta estrutura que seria a sua organização, afetando assim pesquisas futuras e podendo mesmo resultar numa pesquisa falhada (pelo método implementado).

Para evitar isto e sem sacrificar a complexidade o que foi implementado é um processo de pesquisa, após termos encontrado o nó que se deseja alterar irá ser feita um copia da informação antiga para um novo nó com, juntamente com a nova informação a ser adicionada. O nó antigo será removido da árvore e o novo será inserido. Assim a árvore permanecerá ordenada e a inserção irá ser balanceada automaticamente. Estas três operações (pesquisa, remoção e adição) por serem chamadas de forma isolada vão permanecer com a complexidade Θ(log n).

* Remoção:

A operação de remoção tem a mesma base para países e anos, apenas com ligeiras diferenças em termos de código. Basta apenas fornecer o ano/país que se deseja remover da árvore e o método irá pesquisar pelo nó desejado remove-lo da árvore e prosseguir ao rebalanceamento da árvore tudo isto com a mesma complexidade de todas as outras operações isoladas, Θ(log n).

# **4 – Casos de Teste**

De modo a testar as nossas estruturas de modo válido e eficiente, foram adicionadas funções de *benchmark* a cada uma das estruturas de dados.

A funçoes de teste implementadas foram: inserção, pesquisa e remoção de anos e países. Isto é feito de forma a que todos os países / anos adicionados são posteriormente pesquisados e finalmente removidos

Estas funções foram implementadas de maneira semelhante em todas a estruturas, no entanto existem alguma diferenças. Nas listas duplamente ligadas testou-se a inserção na lista (ordenada) de anos em diferentes sitios da lista.

**Os resultados foram os seguintes:**

## Estrutura 1:

Stacks de Stacks

## Estrutura 2:

Listas Duplamente Ligadas de Listas Duplamente Ligadas

## Estrutura 3:

Árvore AVL de Árvores AVL

# **5 – Comparações**

# **6 – Conclusões**