

1 Relatório de entrega

1.1 Tipos de dados

Neste trabalho decidimos utilizar listas de listas simplesmente ligadas. Ou seja, temos uma lista principal em que cada nó tem a apontadores para 1 o próximo nó da lista principal e para uma lista inteira secundária. Para ser de fácil entendimento depois da leitura do relatório a organização das nossas listas de listas foram organizadas, poderá analisar a figura 1 para conseguir ver um esquema que teria 4 países com diferente numero de semanas entre eles.

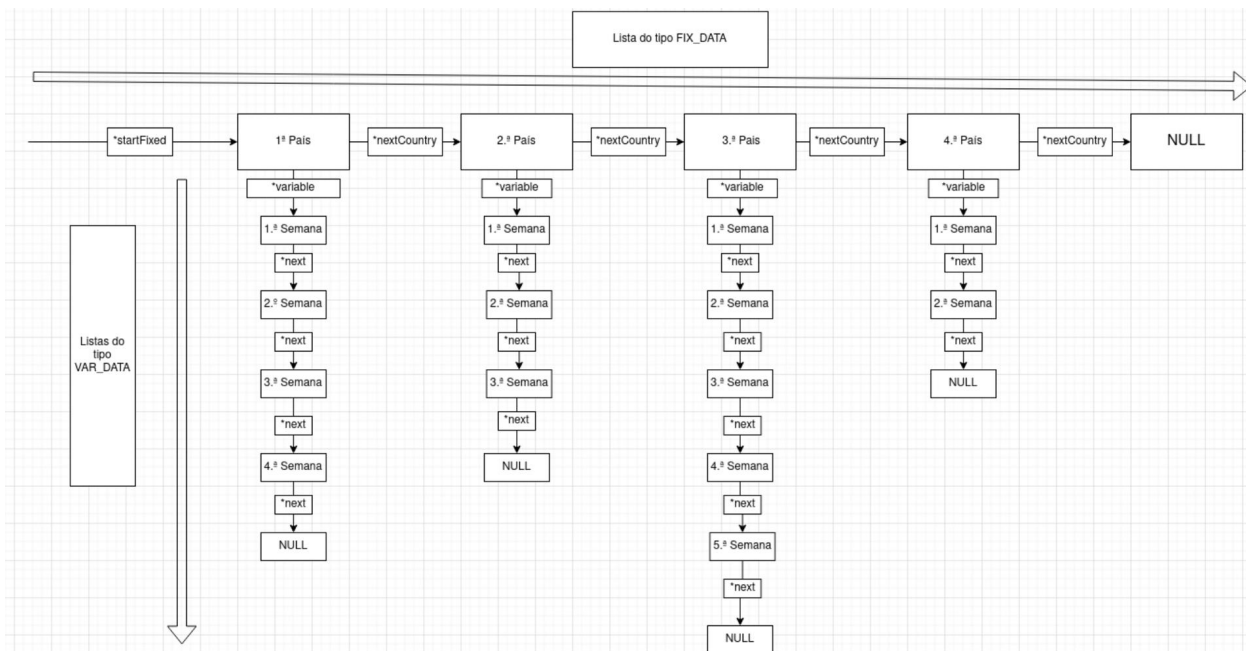


Figura 1: Estrutura das nossas listas

Em relação aos dados que cada estrutura guarda é algo bastante explicito. As listas do tipo FIX_DATA guardam as informações sobre um determinado país que nunca vão variar, esses dados são:

1. Nome do país
2. Código de identificação do país
3. Continente em que o país se encontra
4. População

Já a lista VAR_DATA guarda os dados que podem variar semanalmente dentro do país, esses dados são:

1. Indicador do que a semana se refere (Casos ou mortes)
2. Contagem semanal
3. Data - ano e semana em valores separado
4. Racio dos ultimos 14 dias relativo ao que o indicador tem (casos ou moretes)
5. Contagem total do que o indicador tem (casos ou mortes)

De forma a ser mais fácil ao utilizador do código ter maior facilidade em entender a organização dos dados, a figura 2 encontra-se neste relatório de modo a que o utilizador possa ter conseguido analisar todas os documentos de input e saber o que vai para as estruturas dos dados fixos e as estruturas dos dados variáveis

country	,country_code	,continent	,population	,indicator	,weekly_count	,year_week	,rate_14_day	,cumulative_count
Angola	,AGO	,Africa	,32866268	,cases	,330	,2021-12	,2.078	,22063
Angola	,AGO	,Africa	,32866268	,cases	,568	,2021-13	,2.732	,22631
Angola	,AGO	,Africa	,32866268	,cases	,700	,2021-14	,3.858	,23331
Angola	,AGO	,Africa	,32866268	,cases	,1058	,2021-15	,5.349	,24389
Angola	,AGO	,Africa	,32866268	,deaths	,0	,2020-12	,	,0
Angola	,AGO	,Africa	,32866268	,deaths	,2	,2020-13	,0.061	,2
Angola	,AGO	,Africa	,32866268	,deaths	,0	,2020-14	,0.061	,2
Angola	,AGO	,Africa	,32866268	,deaths	,0	,2020-15	,0	,2

Dados Fixos - FIX_DATA
Dados Variáveis - VAR_DATA

Figura 2: Organização dos dados pelas listas respetivas

Como as nossas listas apenas são simplesmente ligadas o sentido de ligação das listas é única, e é do pointer inicial para o final da lista, acabamos por salvar 8bytes no tamanho de cada estrutura, isto foi feito de modo a salvar máximo espaço que não seja necessário ter alocados, algo que nos deixa ter um maior conjunto de dados de leitura. Para compensar esta falta de ligação temos sempre de utilizar dois pointer auxiliares que andam pela lista de modo que nunca se percam, ou saltem elementos da lista.

1.2 Organização dos dados

A organização dos dados é feita através do uso de um bubble sort para listas genéricas, que vai ajustando valor a valor de modo que depois de realizar todas as iterações o programa esteja completamente organizado. Sendo que a última passagem que é realizada é apenas para confirmação de que os dados realmente foram bem organizados, e que não haverá nenhum problema com a organização das listas. Apesar de o bubble sort não se o sorter mais eficiente possível para organizar listas de qualquer tipo, escolhemos utilizar, pois, é de mais fácil implementação, e como o volume de dados que tínhamos de organizar, não era grande o suficiente para mostrar grande diferença de tempo na organização dos ficheiros.

O bubble sort que foi utilizado mais do que uma vez com diversas condições de troca, garantindo sempre que no final os resultados pedidos pelo utilizador na linha de comandos foram cumpridos e que não houve erros de escolha nem de ordenação dos dados apresentados.

1.3 Escrita em binário

Quando o utilizador pede para haver a escrita do output num ficheiro do tipo .dat, não havia muitas indicações do como o fazer, e por isso, neste trabalho foi realizado de forma a produzir um ficheiro com poucas linhas, ao contrário do que aconteceu com o .csv, em que em todas as linhas era necessário escrever o o país que estava a ser representado, nos .dat, apenas é escrito uma vez os dados fixos de cada país, e uma vez cada 1 dos dados variáveis, ficando no final com muito menos dados escritos para o ficheiro, o que acaba por ajudar no tamanho total do output. Antes da escrita de qualquer tipo de informações num ficheiro binário, existe uma contagem de quantos países vão ser lidos no total, depois da contagem é colocado um inteiro no ficheiro binário com o número total de ficheiros que deve ser lido. O passo seguinte é escrever os dados fixos do primeiro país, seguindo-se a seguir a contagem de quantos dados variáveis existem para o país que está a ser escrito. A seguir à escrita de todos os dados variáveis, repete-se o processo começando não mesmo no início, mas na escrita dos dados fixos do país seguinte.

1.4 Particularidades das funções

Neste trabalho foi decidido que em vez de se ter uma função para cada uma das funções, foi decidido que se iria fazer menos funções com uma dimensão maior, e é enviado o parâmetro de seleção que é pedido. Ou seja, para fazer a seleção da semana com mais infetado, mortos, rácios de mortos e rácios de infetados, é utilizada a mesma função, mas são enviados dois parâmetros, um que é o indicador pretendido, "cases" ou "deaths", e o modo de seleção, se vai ser com o weeklycount ou com os rácios, de modo que não se tenha de repetir uma função 4 vezes, apenas tem de haver uma repetição dupla.