

## LISTA OBRIGATÓRIA PARCIAL DE PROJETO E ANÁLISE DE ALGORITMO

### SEMANA 5

**Esta lista deve usar os algoritmos da Semana 5: Heaps e Heapsort, *k-way-merge* e ordenação por contagem.**

Para os problemas a seguir faça:

- Descreva a ideia de sua solução;
  - Elabore o algoritmo em pseudo-linguagem;
  - Calcule a complexidade do algoritmo para o pior caso;
  - Implemente o algoritmo e teste-o.
- Dado um vetor de  $n$  inteiros distintos em ordem arbitrária, elabore um algoritmo para encontrar o  $k$ -ésimo menor elemento do vetor. Vimos este problema quando estudamos Ordem Estatística, mas agora vocês são obrigados a usar o conteúdo da Semana 5 para resolver o problema. A complexidade da sua solução deve ser  $O(n \log k)$ . Você pode usar um espaço adicional de tamanho  $k$ .
  - Considere que você tem  $n$  fios para conectar, de tamanhos variados. Quando dois fios são conectados, eles passam a ser considerados como um único fio. A realização de uma conexão custa um valor proporcional à soma dos tamanhos dos fios que foram conectados. Por exemplo, para conectar um fio de tamanho 3 com outro de tamanho 5, o custo é  $8x$ , onde  $x$  é o custo base de conexão. Elabore um algoritmo para determinar o custo mínimo para realizar as conexões de  $n$  fios, gerando um único fio conectado.

Para o vetor de fios com tamanhos  $[2, 4, 8, 4, 5, 3]$  o custo mínimo para gerar um único fio seria  $65x$ . Este custo é obtido conectando-se os fios de tamanhos 2 e 3, consumindo  $5x$  e gerando-se um fio de tamanho 5. Em seguida, junta-se os dois fios de tamanho 4, consumindo-se  $8x$  e depois os dois fios de tamanho 5 (original e conectado), consumindo-se  $10x$ . Neste ponto temos dois fios de tamanho 8 (original e conectado), que serão unidos, consumindo-se  $16x$ . Finalmente, este fio seria unido com o de tamanho 10 (conectado), consumindo  $26x$ . No total, seria gasto  $(5+8+10+16+26)x = 65x$ .

- Considere um vetor de palavras extraídas de um texto. Elabore um algoritmo *in loco* para identificar as duas palavras que mais aparecem no texto. Soluções quadráticas não serão aceitas.
- Considere a construção de um índice remissivo de um livro. Tópicos e conceitos importantes do livro constam do índice na forma

Palavra pag, pag, pag, ..., pag

em que palavra refere-se ao tópico ou conceito e pag são números de páginas do livro nas quais o tópico ou conceito aparece. A quantidade de páginas em que as palavras aparecem no livro é variável. Além disso, em um índice remissivo as palavras aparecem uma única vez e estão ordenadas em ordem alfabética.

Elabore um algoritmo para dado um vetor de (palavra, pag), representando todos os termos que irão aparecer no índice, gere o índice remissivo. Na lista de entrada a palavra pode ocorrer mais de uma vez, em ordem arbitrária, mas no índice remissivo isto não é permitido. Para evitar problemas com acentuação, suponha que as palavras estão sem acentuação, na língua inglesa. Soluções quadráticas **não serão** aceitas.

Por exemplo:

Entrada: [(sorting, 70), (heapsort, 200), (heapsort, 100), (sorting, 190), (tree, 80), (sorting, 90), (heap, 60), (list, 30), (tree, 50)]

Saída:

Índice Remissivo

```
heap      60
heapsort  100, 200
list      30
sorting   70, 90, 190
tree      50, 80
```

Nas duas questões a seguir considere o seguinte cenário. Suponha que cada um dos 75 municípios do estado de Sergipe mantém um cadastro de pessoas que foram vacinadas para COVID, organizado em ordem crescente de CPF (somente números). Além do CPF, este cadastro possui outros dados dos cidadãos, incluindo nome, data de nascimento, município que reside, estado, vacina e data das doses. O total de registros, considerando todos os cadastros dos vários municípios, é de  $n$  cidadãos, sendo  $n$  aproximadamente 2 milhões. Suponha que você tenha memória primária para armazenar todos esses dados na íntegra. Além disso, possui disponível um espaço adicional que comporta mais 150 registros desse tipo apenas. A secretaria de saúde do estado contratou vários estagiários da UFS para automatizar algumas tarefas sobre esses dados e você foi um dos contratados. O chefe do setor de Tecnologia da Informação lhe designou as seguintes tarefas:

5. Elaborar um algoritmo em  $O(n \log m)$  para gerar um cadastro único dos cidadãos vacinados do estado, ordenados pelo CPF, a partir dos cadastros ordenados gerados pelos  $m$  municípios do estado.
6. Elaborar um programa para determinar quantos cidadãos vacinados existem por faixas de idade, a partir do cadastro único de vacinados gerado em (5). O gestor irá interagir com o seu programa. Na primeira vez que ele usar o sistema, informando uma faixa de idade (valor da idade menor e da idade maior da faixa) você deve gastar  $O(n)$  para gerar a resposta. No entanto, a partir da segunda consulta, quando uma nova faixa for solicitada, você deve ser capaz de responder em  $O(1)$ .