

Visualização de Dados Municipais da COVID-19 em Haskell

Rafael Calpena Rodrigues

5 de Dezembro de 2020

1 Introdução

Este projeto final foi criado para o curso de Paradigmas de Programação da UFABC, com o intuito de unificar a visualização de dados da COVID-19 em nível municipal para todos os estados do Brasil. O programa foi escrito na linguagem Haskell e é responsável por agrupar dados e gerar tabelas e gráficos dos dados provenientes do site Brasil.IO.

2 Descrição

A execução do programa é separada em duas etapas:

- Preparação: Download e indexação dos dados. Inicialmente, pede-se ao usuário que defina uma pasta para o armazenamento das informações. Serão realizados downloads do dataset completo do Brasil.IO, arquivo JSON de municípios vizinhos e arquivo CSV de correspondências IBGE ID e nome de municípios. Caso os arquivos já tenham sido baixados anteriormente essa etapa será cancelada. Para atualizá-los é necessário apagar o conteúdo do armazenamento ou selecionar uma outra pasta no início da execução.

Após a conclusão dos downloads, o programa executa a descompressão do arquivo gzip do dataset, e em seguida a sua leitura e agrupamento de linhas do CSV em um Map no seguinte formato: (id-ibge, linhas). Cada chave do Map será salva em um json separado, gerando no total aproximadamente 5570 arquivos. A pré-indexação nos permite acessar os dados de um município com rapidez em futuras execuções do programa, similar a técnica de caching, porém utilizando o disco rígido/SSD do usuário.

- Consulta: A partir do arquivo de nomes de municípios o programa cria um FuzzySet, permitindo a pesquisa de um município pelo seu nome aproximado. O nome mais semelhante à entrada será automaticamente selecionado, além de exibir no máximo outras 4 recomendações. Caso o

usuário desejar, também é possível realizar a consulta utilizando o IBGE ID. Ao encontrar o município, o programa exibirá uma tabela contendo data, casos e óbitos totais, novos casos e óbitos, população, média móvel de casos e óbitos (7 dias) e casos e óbitos por 100 mil habitantes. As últimas 4 colunas são computadas no algoritmo por não estarem presentes no csv ou por conterem dados inválidos que impossibilitariam a execução do programa. Os municípios vizinhos e seus respectivos códigos IBGE são exibidos como sugestões após a tabela. Finalmente, o programa volta para a tela de inserção do nome/código do município em um loop infinito.

Outra funcionalidade presente no projeto é a exibição dos gráficos com os mesmos dados das tabelas, possibilitando também a comparação entre municípios vizinhos nos gráficos de casos e óbitos por 100 mil habitantes. Ao contrário dos arquivos pré-indexados, os gráficos são gerados de forma “preguiçosa” (*lazy*), porém permanecem disponíveis para uso futuro dentro da pasta output, junto com os arquivos JSON.

3 Link do Vídeo

<https://www.youtube.com/watch?v=AEvItHQYOMs>

4 Utilização do Programa

Para utilizar o programa, basta ter o Haskell instalado na máquina e digitar “stack run” no terminal.

5 Dificuldades Durante o Desenvolvimento

- 3 Fontes de dados foram necessárias para relacionar o código IBGE (o Brasil IO não inclui o estado do município) e houve dificuldade para lidar com o *parsing* de arquivos JSON e CSV.
- O algoritmo de pré-indexação exige bastante memória e pode travar o computador durante vários minutos. Na primeira versão do algoritmo de agrupamento o computador não conseguiu completar a tarefa mesmo após 8h. A solução foi utilizar Maps para reduzir o tempo a aproximadamente 20 minutos. O tempo de gravação de arquivos JSON também é imprevisível/inconsistente.
- Documentação sem muitos exemplos para as bibliotecas de download, gráficos e tabelas do Haskell

6 Curiosidades

- Existem apenas 2 municípios brasileiros sem municípios vizinhos: Fernando de Noronha/PE e Ilhabela/SP
- O município com o maior número de municípios vizinhos é São Paulo/SP (24); O segundo maior é Caratinga/MG (19)
- Este trabalho se tornará um *plugin* do CoronavirusBrBot, no Telegram.

7 Recursos Utilizados

W. Cota, “Monitoring the number of COVID-19 cases and deaths in brazil at municipal and federative units level”, SciELOPreprints:362 (2020), 10.1590/sci-elopreprints.362

<https://github.com/wcota/covid19br>

Dataset do Brasil.IO

https://brasil.io/dataset/covid19/caso_full/

Arquivo de municípios vizinhos (originalmente cedido pelo IBGE)

<https://github.com/rafaelcalpena/municipios-vizinhos>