

Ciência da Computação / Sistemas de Informação

Hinário Digital: uma base de dados para memória e análise da música evangélica brasileira do século XX até os dias atuais

Dênis de Souza Cordeiro - 8º módulo de Sistemas de Informação, UFLA, bolsista UFLA.

Esther Silva Magalhães - 4º módulo de Ciência da Computação, bolsista PIBIC/FAPEMIG

Eric Fernandes de Mello Araújo - Professor do Departamento de Ciência da Computação, UFLA - eric@ufla.br. Orientador(a) - Orientador(a)

Resumo

O uso de dados extraídos da internet e das plataformas de streaming por meio do uso de APIs (Interface de Programação de Aplicação) em tarefas de aprendizado de máquina e processamento de linguagem natural é útil para o estudo de vários fenômenos em nossa sociedade e em vários contextos. A construção de bases de dados para esses fins é uma tarefa muitas vezes árdua, que demanda o tratamento dos dados, bem como sua análise detalhada para preparação em caso do treinamento de modelos para predição ou busca por padrões a partir dos dados obtidos. Este trabalho consiste na construção de uma base de dados que contenha informações sobre a música evangélica no Brasil desde o Século XX até os dias atuais. A criação dessa base de dados servirá como um repositório histórico e também como um conjunto de dados (dataset) para a análise dos movimentos musicais brasileiros no contexto evangélico. Para realizar esse projeto, foi conduzido um levantamento de sites e plataformas de streaming que contivessem metadados sobre músicas, incluindo informações como nome da música, artista, álbum e ano de lançamento, além de outras características referentes à música, como ritmo, tempo de duração, etc. O Spotify foi escolhido devido à sua praticidade e à quantidade substancial de dados disponíveis. Posteriormente, foram selecionadas as ferramentas para a coleta de dados, optando-se pelo uso da linguagem Python em conjunto com a biblioteca Pandas para a manipulação dos dados e a biblioteca Spotipy para a interação com a API do Spotify. Além disso, as letras das músicas foram coletadas por meio da API Vagalume, que fornece acesso a uma ampla gama de letras de músicas. Após a coleta, foi realizada uma limpeza nos dados para remover caracteres especiais, normalizar o texto e eliminar letras que não estavam completas ou continham informações irrelevantes para a análise pretendida. Os dados foram então armazenados em arquivos CSV e serão utilizados para construir um banco de dados abrangente. Espera-se que este dataset forneça a matéria-prima necessária para estudos avançados, como o treinamento de modelos de aprendizado de máquina. Além disso, está previsto incluir as letras das músicas no banco de dados com o objetivo de facilitar estudos que envolvam a análise das letras.

Palavras-Chave: banco de dados, música evangélica, aprendizagem de máquina.

Instituição de Fomento: UFLA

Link do pitch: https://www.youtube.com/watch?v=0FvR_heeczQ&t=54s