

Especificação funcional e técnica do Music Bot

Este documento fornece uma visão funcional e técnica das atividades necessárias para construir o Music Bot.

Sumário

Objetivo	1
Origem das informações	1
Fluxo das informações.....	1
Detalhamento das atividades.....	2
Ambiente de desenvolvimento	2
Função para limpeza de textos	2
Solicite o nome do artista	3
Pesquise o nome do artista na API do Vagalume.....	3
Preparação dos nomes dos artistas para confirmação do usuário	3
Pesquisando as músicas mais famosas do artista	3
Salvando um arquivo json com as letras das músicas.....	4
Exibindo o word cloud ao usuário	4
Buscando as notas musicais mais utilizadas pelo artista	4
Music Bot Management	4

Objetivo

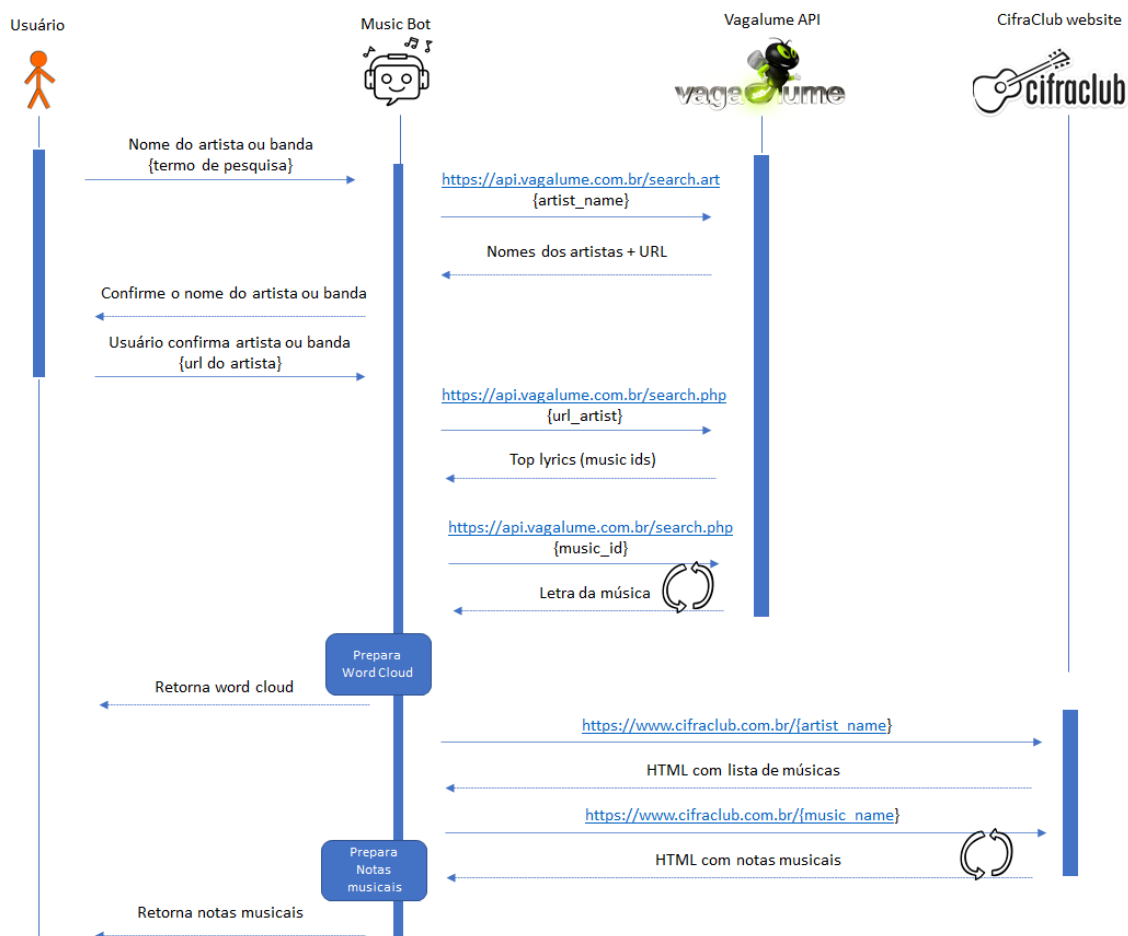
Criar um notebook interativo, capaz de capturar um nome de artista ou banda **digitado pelo usuário** e apresentar uma **nuvem de palavras** com os termos mais utilizados nas principais letras de músicas deste artista. Adicionalmente, o notebook também deve apresentar as **10 notas musicais mais utilizadas** nos acordes musicais do artista ou banda.

Origem das informações

As letras de músicas do artista ou banda serão buscadas na **API do Vagalume** (www.vagalume.com.br). As informações sobre os endpoints específicos serão fornecidas no decorrer da documentação. Já as informações sobre notas musicais, serão capturadas por meio de web scraping no **site do CifraClub** (<https://www.cifraclub.com.br/>).

Fluxo das informações

Com o objeto de facilitar a especificação técnica de cada atividade do projeto, o diagrama abaixo fornece uma visão geral da troca de informações entre usuário, notebook, API do Vagalume e site do Cifra Club. Use este diagrama apenas para entender, de forma geral, *como* e *quando* cada tipo de informação será capturada.



Perceba que teremos que fazer web scraping de várias páginas no site do CifraClub. Isso porque precisamos acessar URLs com o seguinte formato https://www.cifraclub.com.br/{music_name}.

Detalhamento das atividades

A seguir, serão detalhadas todas as atividades de desenvolvimento. Note que as atividades seguem uma sequência cronológica de acordo com a lógica de todo o processo.

Ambiente de desenvolvimento

Siga as instruções do link a seguir para configurar o ambiente de desenvolvimento e cadastrar uma API Key: <https://github.com/weslleymoura/music-bot/blob/master/README.md>

Função para limpeza de textos

Crie uma função que recebe um texto e um código de idioma. Esta função deve retornar o texto após realizar as seguintes etapas de limpeza:

- Remover qualquer pontuação existente
- Remover tokens não alfabéticos
- Remover palavras com apenas um caractere
- Remover stop words

Atenção: remova stop words de acordo com o idioma informado no parâmetro de entrada da função. Dentro da função, use o seguinte mapeamento de linguagens {1: 'portuguese', 2: 'english', 3: 'spanish'}

Solicite o nome do artista

Solicite o nome do artista ou banda ao usuário e grave a informação em uma variável.

Pesquise o nome do artista na API do Vagalume

Faça uma chamada ao endpoint <https://api.vagalume.com.br/search.art> e informe a seguinte query string {"q":name_artist, "limit": "10"}, onde nome_artist é o nome do artista ou banda capturado na etapa anterior.

Preparação dos nomes dos artistas para confirmação do usuário

A pesquisa anterior retorna, entre outras informações, o **nome e a URL das bandas** que fazem match com o termo da pesquisa. Use estas informações para preparar os dados de confirmação ao usuário. O objetivo desta tarefa é possibilitar que o usuário nos informe a URL que devemos utilizar deste ponto em diante. Use a função `ipywidgets.interact` para apresentar a informação ao usuário e solicitar a confirmação da pesquisa.

Pesquisando as músicas mais famosas do artista

Neste ponto o usuário já confirmou a URL do artista que ele deseja consultar. Utilize esta URL para buscar os dados do artista da seguinte forma https://www.vagalume.com.br/{url_artista}/index.js

As músicas principais do artista estão disponíveis no objeto['artist']['toplyrics']['item']. Para buscar a letra de cada uma dessas músicas, basta acessar o endpoint <https://api.vagalume.com.br/search.php> e informar a seguinte query string {"musid": <<id_da_música>>, "apikey": <<sua_api_key>>}.

Grave as letras de cada música dentro de um dicionário. Este dicionário deve possuir as seguintes chaves:

- letra: a letra original da música (['mus'][0]['text'])
- letra_clean: a letra da música, após passar pela função de limpeza de dados
- url_letra: a URL da letra, retornada pela API (['mus'][0]['url'])
- letra_traduzida: a letra traduzida da música, retornada pela API (['mus'][0]['translate'][0]['text'])
- url_letra_traduzida: a URL da letra traduzida, retornada pela API (['mus'][0]['translate'][0]['url'])

Note: este é o único endpoint que exige o uso de uma API Key.

Cenários que exceção: considere o tratamento dos seguintes cenários de exceção:

- A API não retornou nenhuma música do artista informado
Neste caso, informe ao usuário que o assistente não conhece nenhuma música do artista
- Houve um erro ao buscar a letra da música
Existem mecanismos de segurança na API do Vagalume contra robôs. Caso a API identifique várias chamadas idênticas com origem do mesmo endereço IP, um erro será retornado. Nestes casos, informe ao usuário que não foi possível encontrar a letra da música e solicite que retorne mais tarde OU informe outra banda.

Salvando um arquivo json com as letras das músicas

Crie um arquivo chamado `catalogo_musicas.json` com os dados armazenados em `['artist']['toplyrics']['item']`.

Exibindo o word cloud ao usuário

Crie um word cloud usando as letras de músicas com o texto tratado (`letra_clean`).

Buscando as notas musicais mais utilizadas pelo artista

Usaremos outro conjunto de músicas para tabular as notas musicais mais utilizadas pelo artista. Isso é necessário porque as informações de notas musicais estão disponíveis apenas no site do CifraClub. As etapas necessárias para esta tarefa são:

1. Acessar a página principal da banda do site CifraClub

https://www.cifraclub.com.br/{url_do_artista}. Note que estamos reutilizando a URL do artista já definida anteriormente para encontrar a página do CifraClub. Esta foi uma estratégia adotada neste projeto, já que, após realizar uma análise inicial, verifiquei que as páginas do CifraClub seguem a mesma nomenclatura das URLs retornadas pela API do Vagalume!

2. Recolher os links das músicas disponíveis

A lista de músicas está disponível da tag `<a>`, classe `art_music-link`

3. Acessar cada link de música encontrado

Esta informação está disponível em <https://www.cifraclub.com.br/{nome-da-musica}>, onde o `{nome-da-musica}` foi encontrado na etapa anterior.

4. Extrair somente as notas musicais de cada música

As notas musicais estão disponíveis na tag `<div>`, classe `songContent`.

5. Criar um contador para tabular todas as notas musicais utilizadas em todas as músicas

Crie um dicionário com o formato `{nota-musical: quantidade de ocorrências}`

6. Exibir as notas mais usadas pelo artista

Mostre as 10 notas musicais com mais ocorrências ao longo de todas as letras de músicas.

Music Bot Management

O objetivo desta feature é gravar todas as execuções do notebook em um banco de dados do SQL Lite.

Crie uma tabela com o nome `historico_pesquisas` com as seguintes colunas:

Nome da coluna	Tipo de dados	Valor
<code>data_requisicao</code>	<code>date</code>	<code>date('now')</code>
<code>artista</code>	<code>text</code>	Salve o nome do artista
<code>url_vagalume</code>	<code>text</code>	Salve a URL do artista no site do Vagalume

<i>url_cifraclub</i>	text	Salve a URL do artista do site do CifraClub
<i>qtd_musicas_vagalume</i>	int	Número de músicas retornadas pela API do Vagalume (top lyrics)
<i>qtd_musicas_cifraclub</i>	int	Número de músicas do site do CifraClub usadas para tabular as notas musicais