**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](#_Toc42335738)

[Origem das informações 1](#_Toc42335739)

[Fluxo das informações 1](#_Toc42335740)

[Detalhamento das atividades 2](#_Toc42335741)

[Ambiente de desenvolvimento 2](#_Toc42335742)

[Função para limpeza de textos 2](#_Toc42335743)

[Solicite o nome do artista 3](#_Toc42335744)

[Pesquise o nome do artista na API do Vagalume 3](#_Toc42335745)

[Preparação dos nomes dos artistas para confirmação do usuário 3](#_Toc42335746)

[Pesquisando as músicas mais famosas do artista 3](#_Toc42335747)

[Salvando um arquivo json com as letras das músicas 4](#_Toc42335748)

[Exibindo o word cloud ao usuário 4](#_Toc42335749)

[Buscando as notas musicais mais utilizadas pelo artista 4](#_Toc42335750)

[Music Bot Management 4](#_Toc42335751)

# Objetivo

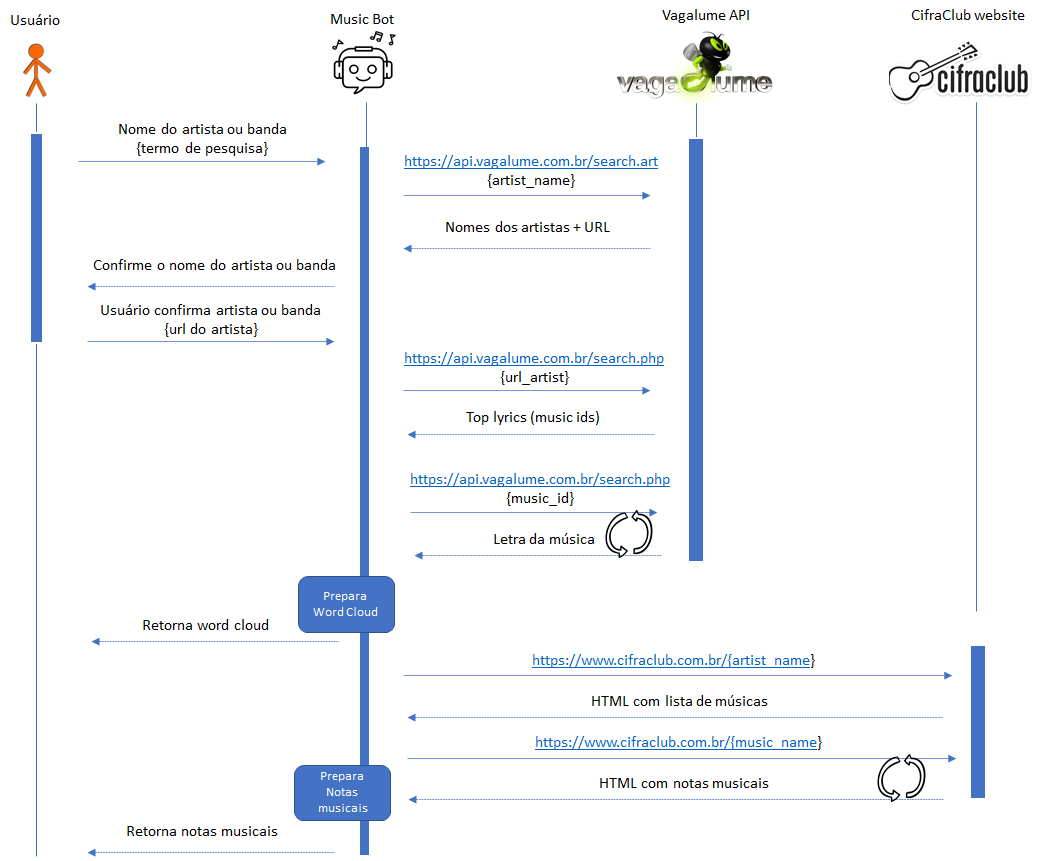
Criar um notebook interativo, capaz de capturar um nome de artista ou banda **digitado pelo usuário** e apresentar uma **nuvem de palavras** com os termos mais utilizados nas principais letras de músicas deste artista. Adicionalmente, o notebook também deve apresentar as **10 notas musicais mais utilizadas** 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](http://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}*](https://www.cifraclub.com.br/%7bmusic_name%7d)*.*

# 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 nomedo 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](https://www.vagalume.com.br/%7burl_artista%7d/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}*](https://www.cifraclub.com.br/%7burl_do_artista%7d)*. 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!*

1. Recolher os links das músicas disponíveis

*A lista de músicas está disponível da tag <a>, classe art\_music-link*

1. Acessar cada link de música encontrado

*Esta informação está disponível em* [*https://www.cifraclub.com.br/{nome-da-música}*](https://www.cifraclub.com.br/%7bnome-da-música%7d)*, onde o {nome-da-música} foi encontrado na etapa anterior.*

1. Extrair somente as notas musicais de cada música

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

1. 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}*

1. 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 |
| data\_requisicao | date | date('now') |
| artista | text | Salve o nome do artista |
| url\_vagalume | text | Salve a URL do artista no site do Vagalume |
| url\_cifraclub | text | Salve a URL do artista do site do CifraClub |
| qtd\_musicas\_vagalume | int | Número de músicas retornadas pela API do Vagalume (top lyrics) |
| qtd\_musicas\_cifraclub | int | Número de músicas do site do CifraClub usadas para tabular as notas musicais |