

**Faculdade de Tecnologia de São José dos Campos**

**FATEC - Professor Jessen Vidal**

Ângelo Lima

Caique Fernandes

José Henrique dos Santos

Larissa Takahashi

Matheus Henrique Rothstein

Natália dos Reis Neves

Renato Passos

Sandro Toline

Tábatha Fróes

**Sistema de Identificação Única**

**FATEC - UOL**

**São José dos Campos**

**2021**

Controle de Versões			
Versão	Data	Autor	Notas da Revisão
1.0	19/09/2021	Natália Neves e Larissa Miho	Elaboração Inicial
1.1	28/11/2021	Natália Neves	Atualização das Informações

## ● Objetivo deste Documento

Descrever de forma clara, os objetivos do desenvolvimento, etapas, restrições e premissas. Além de documentar o funcionamento de todos os segmentos do software.

## ● Situação Problema Atual

O objetivo do projeto é implementar um sistema que possibilita a identificação, documentação e o controle dos diversos usuários únicos em seu sistema de e-mail BOL.

## ● Critérios de Sucesso

Desenvolver no período de 3.5 meses, um sistema que possibilita a identificação, documentação, e o controle dos diversos usuários únicos em seu sistema de e-mail BOL, entregando todos os requisitos e produtos previamente aprovados.

## ● Premissas

1. Todos os desenvolvedores precisarão de pelo menos uma máquina para a execução do projeto e para participar das reuniões online;
2. Todos os desenvolvedores deverão ter os softwares de desenvolvimento em suas máquinas para criar o projeto e rodar o mesmo;
3. Criar um ambiente para o armazenamento do banco de dados;

## ● Restrições

1. Orçamento de R\$ 9.250 por sprint;
2. O projeto deverá acontecer num período de 3 meses;
3. Início do projeto está programado para 08/09/2021;

- **Produtos e Principais Requisitos**

**Documentação do projeto**

- Organização GitHub
- Gif ou vídeo de funcionamento do projeto
- Documentação Geral (Engloba todos os requisitos e documentações do projeto)
- Termo de abertura do projeto
- Declaração do escopo

**Configuração e Modelagem do Banco de Dados**

- Diagrama do Banco de Dados
- Configuração do Banco
- Hospedagem do Banco
- Modelagem das Tabelas

**Dicionário de Dados**

- Documento com tabelas, dados e descrições gerais.

**Conexão com o Banco de Dados**

- Conexão entre o Banco e o sistema funcional, automatizando as informações.

**Captação dos Dados de Entropia do Usuário**

- Captação dos dados de entropia do usuário, dentre eles o dispositivo, IP, localização e outros parâmetros de validação.

**Criação do Hash para Usuários Únicos**

- Criação do Hash de validação do usuário único através dos dados de entropia que foram captados na tela de cadastro.

**Cadastro dos Usuários no sistema de e-mail**

- Realizar o Cadastro do Usuário contendo os campos (Nome, E-mail, senha, telefone e dados de entropia)

**Score de Pontuação dos Usuários**

- Score gerado na tela de cadastro que serve como parâmetro para treinamento da Inteligência Artificial. O funcionamento do Score será de 0 á 1, sendo cada vez mais próximo do 1 a chance do usuário ser um usuário único.

**Tela de Cadastro**

- Criação da Tela de Cadastro
- Criação dos Endpoints para passagem dos dados (Nome, E-mail, senha e dados de entropia)
- Política em caso de negação do compartilhamento de dados, em afirmativo, bloquear o cadastro.

### Método de Exibição dos Usuários com ID Único

- Criar método de listagem com todos os usuários que possuem um ID único no sistema.

### Tela para Exibição dos Usuários com ID Único

- Criar tela de listagem com todos os usuários que possuem um ID único no sistema.

### Criação da Inteligência Artificial

- Criação da Inteligência Artificial para realizar a automação do filtro pelos usuários únicos que se cadastrarem no sistema.

### Treinamento da Inteligência Artificial

- Treinamento da Inteligência Artificial para realizar a automação do filtro pelos usuários únicos que se cadastrarem no sistema.

## • Resultados das Entregas

Sprint 1	Sprint 2	Sprint 3	Sprint 4
Captação dos dados de entropia do usuário em todos os navegadores	Validações e regras de negócio aplicadas nos dados de entropia	Comparações de informações entre os dados de entropia e o hash gerado pelo sistema	Criação do método de exibir usuário com ID único com IA. A IA vai aprender a perceber comportamentos incomuns de usuários e dar uma lista desses usuários
Criação da tela de cadastro para obter os dados de entropia do usuário	Criação e configuração inicial da Inteligência Artificial	Treinamento da Inteligência Artificial utilizando o dataset gerado pela comparação de dados de entropia e hash	Exibição geral dos usuários únicos no sistema recebidos pela inteligência artificial
Criação e modelagem do banco de dados com as entidades necessárias	Integração entre o front-end, back-end e banco de dados	Alteração da tela de exibição dos usuários do sistema, para exibição apenas dos usuários únicos	Automatização da inteligência artificial
Planejamento da Inteligência Artificial e configurações iniciais	Criação da tela de exibição dos usuários do sistema		
Criação dos métodos de cadastro e captação dos dados no back-end			

- **Banco de Dados**

A aplicação inicial do banco de dados foi para realizar os testes de cadastro do usuário e seu funcionamento. Seguem abaixo algumas explicações e demonstrações do funcionamento:

## **Instalação**

- 1) Realize a instalação do DBeaver na sua máquina  
(<https://dbeaver.io/download>)
- 2) Realizar conexão com o banco de dados através dos dados disponibilizados abaixo:

Novo Banco de dados:

Host: ec2-54-146-82-179.compute-1.amazonaws.com

Database: d5gtucomvgir2l

Username : westwjbipxacva

Port: 5432

Password: 1d73874df84b39cab4f5d5c8f44514b0ef2244cd7bd7bc0596f323f0da2c1924

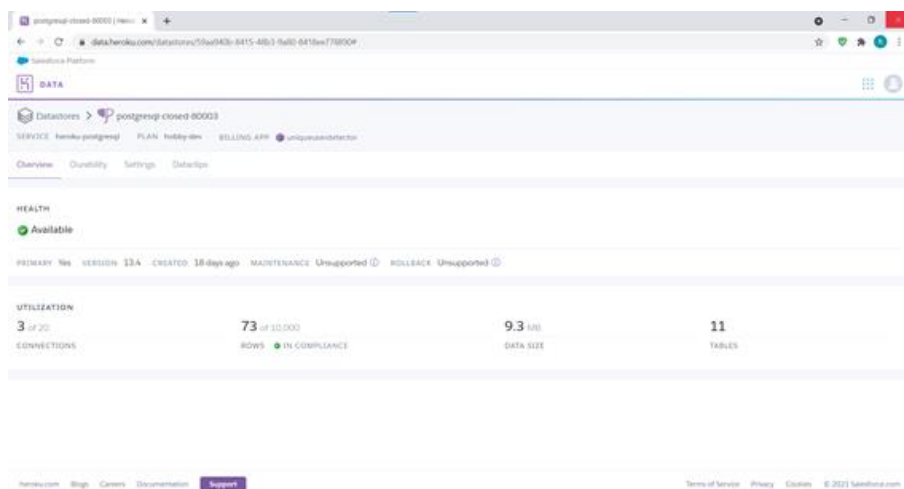
URI:

postgres://westwjbipxacva:1d73874df84b39cab4f5d5c8f44514b0ef2244cd7bd7bc0596f323f0da2c1924@ec2-54-146-82-179.compute-1.amazonaws.com:5432/d5gtucomvgir2l

Heroku CLI: heroku pg:psql postgresql-rugged-23386 --app uniqueuser (editado)

26 de novembro de 2021

- 3) Após a conexão, os registros estarão separados na interface por tabelas para a visualização.



*Figura 1 - Plataforma em Cloud, Heroku*

Utilizamos o Heroku, plataforma em cloud, para hospedar nosso banco de dados.

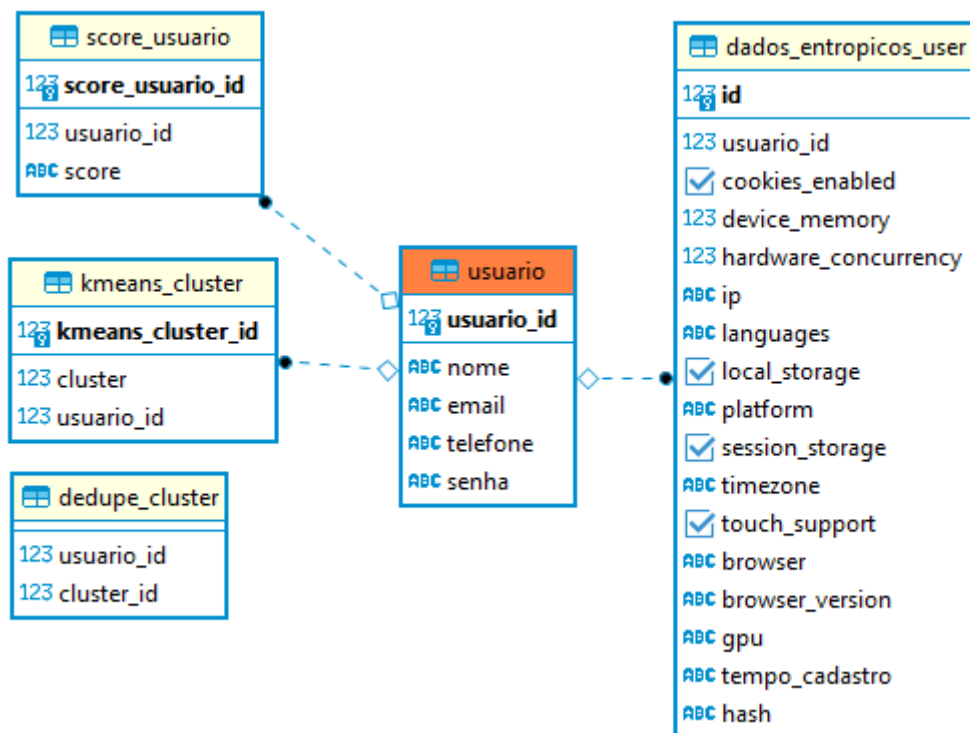


Figura 2 - Diagramação do Banco de Dados

Diagramação Final do Banco de Dados. As tabelas são as atuais presentes no banco de dados com todas as informações diagramadas acima.

## ● Back-End

A aplicação inicial do back-end foi para realizar os testes de cadastro do usuário e seu funcionamento, além da criação de rotas e integração entre front-end e banco de dados. Através do sistema desenvolvido, realizamos os filtros e validações pelo back-end da aplicação. Seguem abaixo algumas explicações e demonstrações do funcionamento:

### Instalação

- 1) Baixar ou atualizar o java para a versão 11

- 2) Baixar e abrir a aplicação do software no editor de código da sua preferência.
- 3) Instalar o Loombok
- 4) Abrir o arquivo "*BackJavaApplication.java*"
- 5) Clicar no botão Debug e Run para iniciar a aplicação.

## **Explicação Básica sobre o desenvolvimento Back-End**

### **Classe Domain**

Adaptação dos dados recebidos pelo banco de dados para serem utilizados pelo back-end.

### **Classe DTO**

Realiza a serialização dos dados. Utilizado para filtrar os dados que são transacionados entre o Banco de dados, Back-end e front-end para serem utilizados na programação back-end.

### **Interface Repositories**

Interface responsável pela conversação com as tabelas do banco de dados. Aonde as query's e requisições são solicitadas.

### **Controlador - Resources**

Controlador REST. Onde são realizadas as requisições e respostas.

### **Classe Services**

Classe com as regras de negócio que serão utilizadas e aplicadas na classe de domínio.

### **Util**

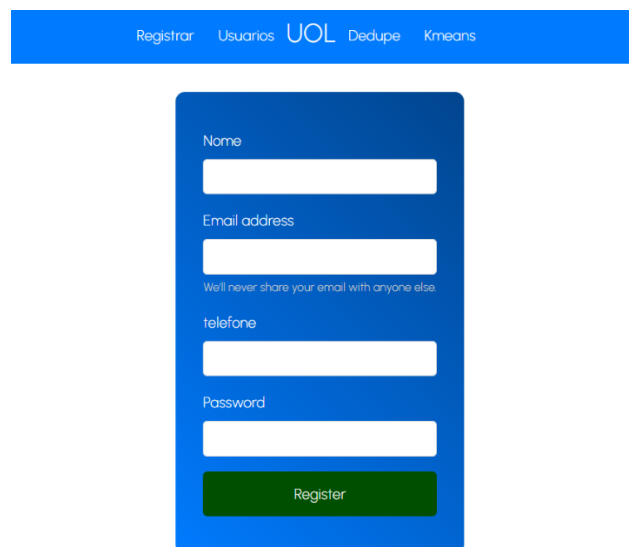
Encriptar e Decriptar os dados utilizando o algoritmo de criptografia de mensagens - MD5. O MD5 é uma função hash amplamente utilizada com a finalidade de produzir um valor de hash de 128 bits expresso em 32 caracteres.

## Desenvolvimento do Score

O score realiza a captação de todos os dados entrópicos do usuário. Conforme os cadastros são realizados, a aplicação compara os dados do usuário que foi cadastrado naquele momento com os demais que já estão no banco de dados. A pontuação do Score é inicialmente setada como 1 e conforme existem semelhanças com os dados entrópicos o valor vai decaindo até negativar.

- **Front-End**

A aplicação do front-end foi a nossa principal proposta de valor inicial, onde nela coletamos os dados de entropia do usuário, geramos o hash de identificação única junto ao score do usuário. A tela protótipo inicial é a base apenas da coleta dos dados. Seguem abaixo algumas explicações e demonstrações do funcionamento:



*Figura 3 - Interface Inicial de Cadastro do Sistema*



Registrar Usuarios UOL Dedupe Kmeans			
Usuários			
Nome	Email	Hash	Score
serafina fina rodrigues	fininha@bol.com.br	Igd7abb2e497c9f5a00ff94da5dc9e2	0.56600
Ângelo Victor de Lima	angelovictor2001@bol.com.br	IadI54I48eac5h2035mbf9Ih8Ie8I	0.86550
Ângelo Victor de Lima	angelovictor2001@bol.com.br	IadI54I48eac5h2035mbf9Ih8Ie8I	0.86550
Ângelo Victor de Lima	angelovictor2001@bol.com.br	IadI54I48eac5h2035mbf9Ih8Ie8I	0.72650
Alice	Alice@gmail.com	e09c80c42fda55f9d992e59cab3307d	0.71750
Alice	Alice@gmail.com	e09c80c42fda55f9d992e59cab3307d	0.62500
Alice	Alice@gmail.com	e09c80c42fda55f9d992e59cab3307d	0.71750
Laura	Laura@gmail.com	eIffaffe9c3c80If2f8c3fbe7cb032cb2	0.75550
Laura	Laura@gmail.com	eIffaffe9c3c80If2f8c3fbe7cb032cb2	0.50300
Laura	Laura@gmail.com	eIffaffe9c3c80If2f8c3fbe7cb032cb2	0.75550
Isabella	Isabella@gmail.com	4Ifcba09f2bdcdf3I5ba4II9dc7978dd	0.71750

Figura 4 – Tela de exibição dos usuários únicos identificados pela Inteligência Artificial

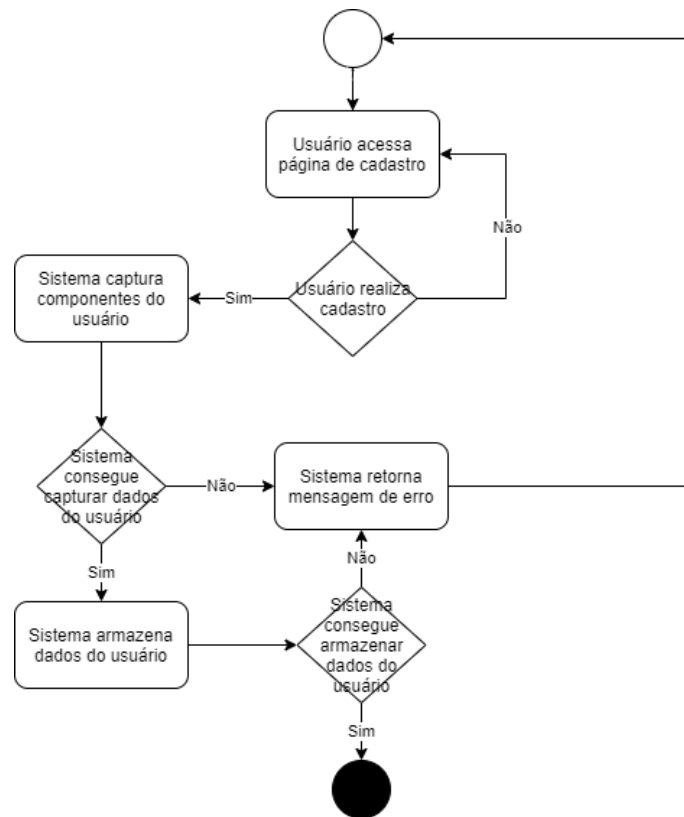
Registrar Usuarios UOL Dedupe Kmeans		
Usuários		
Nome	Email	Cluster
Script java 7	scriptjava@bol.com.br	0
Java script 11	javascript@bol.com.br	0
Script java	scriptjava@bol.com.br	0
Script java 5	scriptjava@bol.com.br	0
Script java 2	scriptjava@bol.com.br	0
Script java 6	scriptjava@bol.com.br	0
Script java 3	scriptjava@bol.com.br	0
Script java 4	scriptjava@bol.com.br	0
Java script 9	javascript@bol.com.br	1
Script java 8	scriptjava@bol.com.br	1
Java script 10	javascript@bol.com.br	1

Figura 5 - Tela de exibição dos usuários únicos identificados pela Inteligência Artificial através do Dedupe.

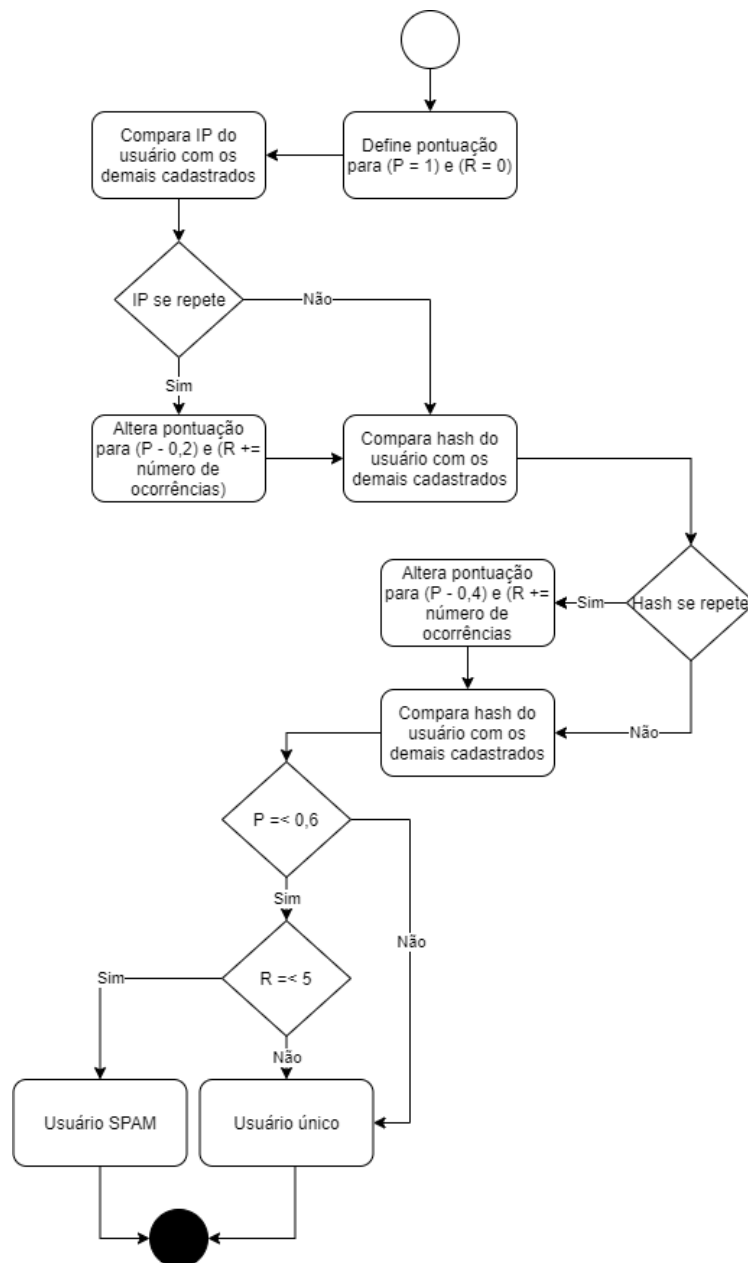


- **Inteligência Artificial**

## Diagramas de Utilização da IA



*Figura 1 - Diagrama de Funcionamento do Sistema*



*Figura 2 - Diagrama de Funcionamento da Inteligência Artificial*

## Dedupe

O Dedupe é responsável pela construção de clusters, ele realiza o agrupamento e separação dos usuários únicos ou não únicos.

Um usuário único por exemplo possui um cluster para si, ao contrário de um usuário identificado como não único que se agrupa em um cluster específico com mais de um usuário. Logicamente falando, os usuários únicos estarão presentes em apenas 1 cluster, fato que é contrário caso o usuário não seja

único, afinal ele estará presente em mais de um cluster com outros supostos usuários que possuíram aquelas determinadas semelhanças.

### **Script para Clusterização via Python**

Inicialmente a programação realiza a Normalização do dataset gerado pela aplicação para utilizar na aplicação do código e realizar a clusterização dos usuários recebidos pelo dataset.

Após a normalização eles são transcritos para o script afim de realizar a clusterização dos usuários únicos. O script analisa dado por dado baseado nas semelhanças entre as informações receptadas para realizar a divisão entre os clusters. Através do método **Elbow** ele realiza a divisão do dataset em 3 clusters, 0, 1 e 2

- **Links Úteis**

**GitHub:** <https://github.com/tabathafroes/uniqueUserDetector>