



PESC

Programa de Engenharia
de Sistemas e Computação

Uma abordagem NoSQL no auxílio a unificação e estudo dos registros de pessoas desaparecidas

Aluno: Raul Sena Ferreira

Professor: Alexandre de Assis Bento Lima



Sumário

- Introdução
- Motivação
- Objetivo
- Tecnologias Utilizadas
- Metodologia
- Resultados
- Trabalhos Futuros
- Referências



Introdução

Segundo estimativa do Governo Federal, 40 mil crianças desaparecem todo ano no Brasil

Acredita-se que o número seja ainda maior, pois um inúmeros casos não são registrados e nem investigados

Segundo o SEDH, cerca de 15% dos desaparecimentos não são resolvidos por um longo período de tempo ou jamais são solucionados



Introdução

Objetivos da Rede Nacional de Identificação e Localização de Crianças e Adolescentes Desaparecidos (ReDESAP):

- Criar um cadastro unificado de dados e casos notificados
- Articular serviços especializados de atendimento público
- Promover o compartilhamento de informações sobre o fenômeno entre os diversos atores da Rede
- Coordenar um esforço coletivo em âmbito nacional



Motivação

Existe muita informação e ao mesmo tempo, nenhuma:

- Muitos sites com dados de crianças desaparecidas, buscar em qual?
- Muita informação repetida e desatualizada;
- Cadastro Nacional de Crianças e Adolescentes Desaparecidos: 369 casos registrados!
- Estatísticas oficiais desatualizadas e não condizentes com a realidade



Motivação

Dados não estão disponibilizados de forma organizada:

- Não existe uma interface de consulta que permita extrair maior informação sobre os dados
- Componente geográfica dos dados não é explorada
- Não existe uma interface que permita que as pessoas possam acessar e analisar, de forma simples, os dados



Objetivo

O trabalho se divide em três partes:

- Modelo de unificação das diferentes bases de dados existentes
- Modelo de interface de consulta (espacial e textual) e análise dos dados já coletados
- Disponibilização dos dados coletados através de API



Tecnologias - Bancos de Dados

ETL:



mongoDB

{ name: mongo, type: DB }

Consulta:



PostgreSQL



Cache:



redis



Tecnologias - Aplicação

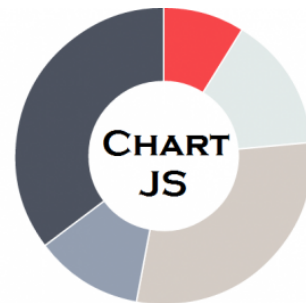
Servidor:



python



Interface:



HTML



CSS





Metodologia

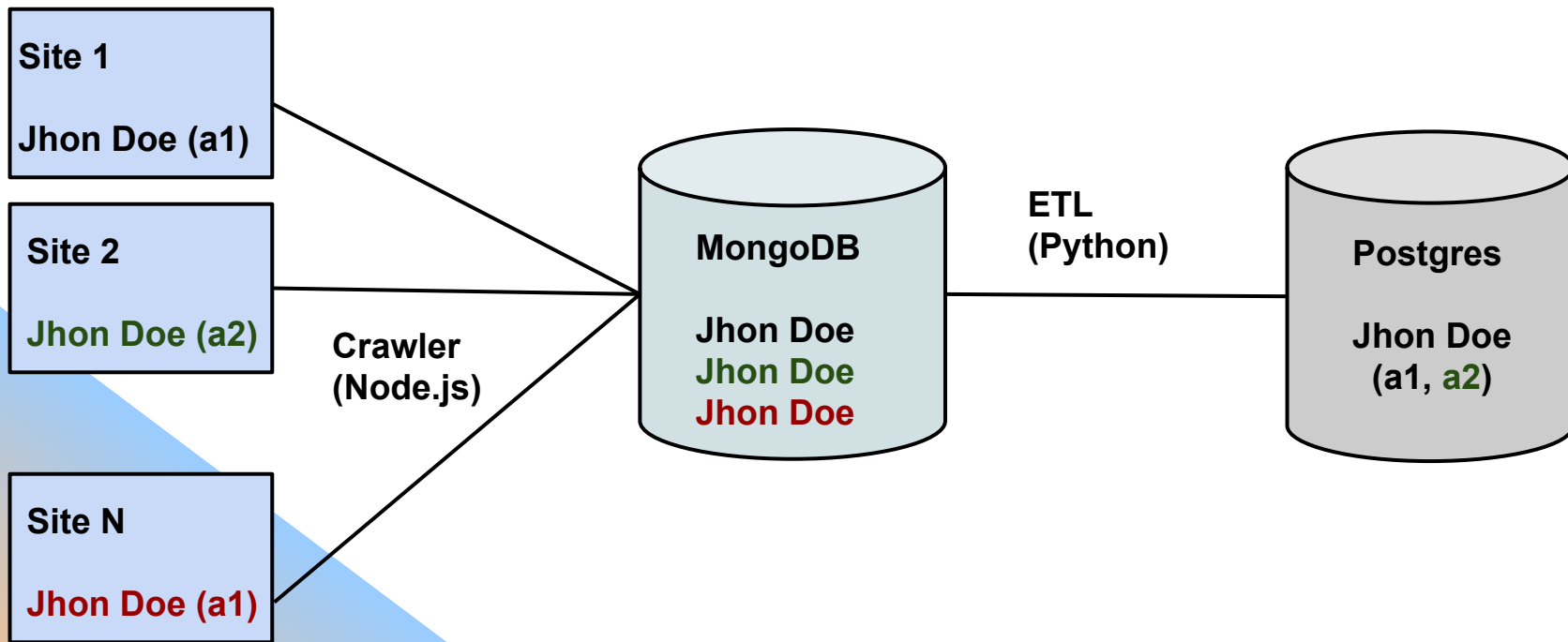
Modelo de unificação das bases:

- O *crawler* feito em Node.js varre os sites que possuem registros de pessoas desaparecidas
- As informações são guardadas no MongoDB onde cada site é tratado como uma coleção de documentos
- Em seguida, um processo de ETL, feito em python, extrai, limpa, unifica e salva os dados de forma estruturada no Postgres



Metodologia

Modelo de unificação das bases:





Metodologia

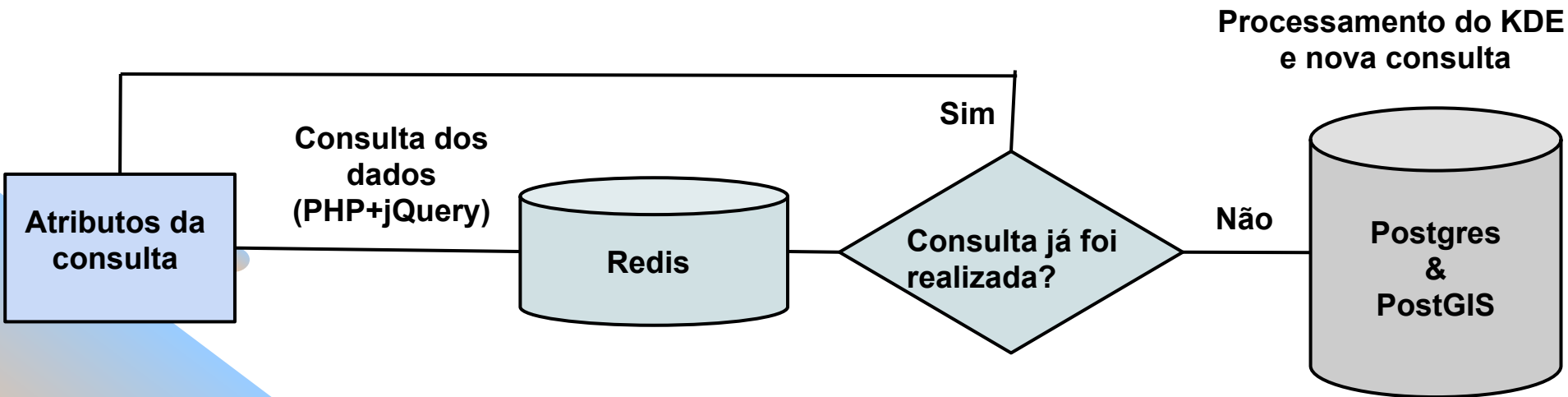
Modelo de interface de consulta:

- Interface da aplicação permite consultas por atributos como: Sexo, Situação, Cor, Nome, Estado
- Resultado da consulta vem em forma de:
 - Gráficos: Gênero, Raça e Desaparecido/Estado
 - Pontos geográficos e coloração dos estados baseado na probabilidade de ocorrer um novo caso
- O Redis atua como cache, salvando o hash da consulta como chave e seu resultado como valor



Metodologia

Modelo de interface de consulta:

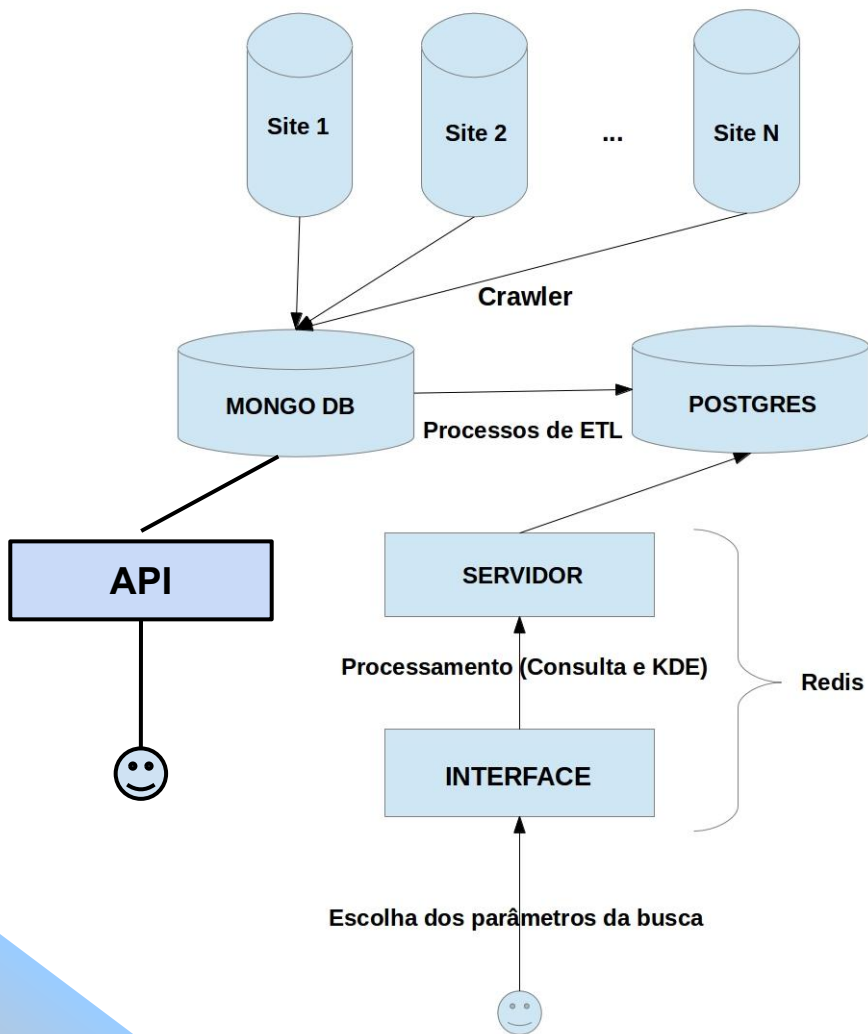




Metodologia

API de dados:

- Node.js + “Express” fornece os dados através de uma URL própria, onde ao receber uma requisição via parâmetro (GET) faz uma consulta no MongoDB e retorna o resultado da consulta em formato JSON





Resultados

Sites que o sistema já está “crawleando”:

- <http://www.desaparecidos.gov.br/>
- <http://portal.mj.gov.br/Desaparecidos/>
- <http://www.desaparecidos.rs.gov.br>
- <http://www.biamap.com.br/>

Fase final:

- <http://www.desaparecidosdobrasil.org>
- <http://www.desaparecidos.mg.gov.br>



Resultados - Links

A versão beta do sistema encontra-se funcional e pode ser acessado através do link:

<http://www.raulferreira.com.br/Projeto-Myosotis/>

A versão beta da API também está operante e pode ser acessado através dos links:

<http://www.raulferreira.com.br:4000/collections/siteX/all>

Onde o X deve ser alterado por 1, 2, 4 ou 6, representando os sites já crawleados



Resultados - Tela de Consulta

http://www....-Myosotis/ x +

www.raulferreira.com.br/Projeto-Myosotis/ Pesquisar

Myosotis GRÁFICOS API SOBRE T

+
-

ESCOLHA O FILTRO DE BUSCA

Você deseja buscar por...

Sexo: ☐ Feminino ☐ Masculino ☐ Ambos

Situação: ☐ Desaparecidos ☐ Encontrados ☐ Todos

Cor de pele: ☐ Branca ☐ Preta ☐ Parda ☐ Amarela ☐ Indígena ☐ Todas

Nome:

Estado:

Cancelar BUSCAR



Resultados - Mapa de Resultados

www.raulferreira.com.br/Projeto-Myosotis/

Pesquisar



Myosotis

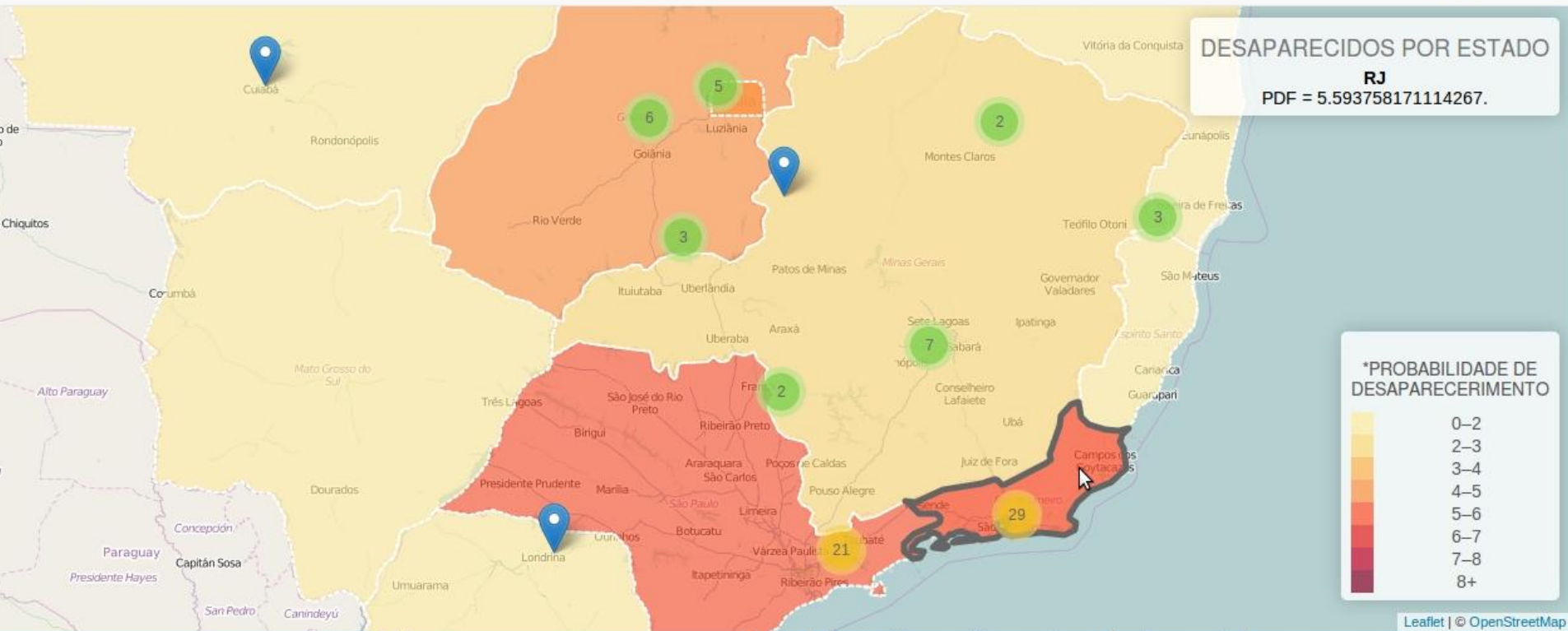
GRÁFICOS

API

SOBRE

TERMOS DE USO

CONTATO





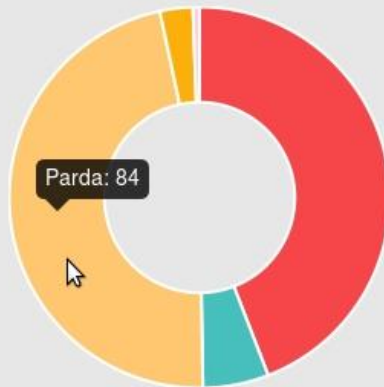
Resultados - Gráficos dos Resultados



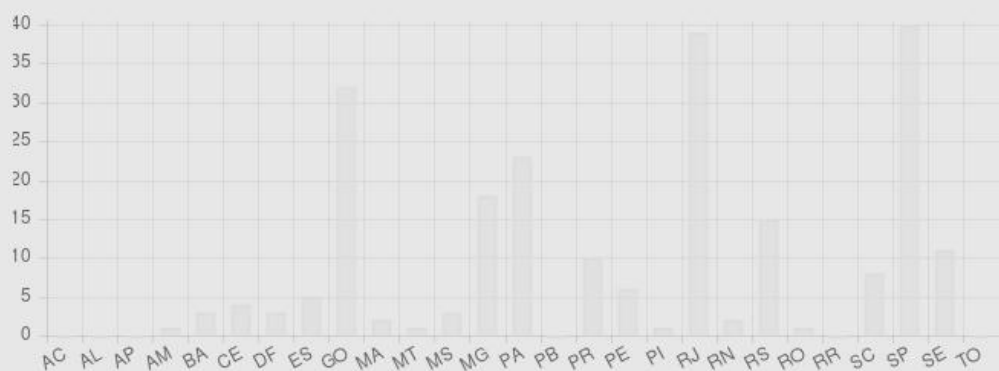
GÊNERO



RAÇA



DESAPARECIDO/ESTADO





Resultados - Registros

Quantidade de registros maior do que nos sites buscados:

- 1794 Registros distintos (4 sites)
- 1600 Registros com coordenadas geográficas
- 1023 Registros de pessoas ainda desaparecidas

A informação de vários registros foram beneficiados com a agregação da informação:

Ex: joao rafael santos kovalski

- Bairro, Data de Nascimento e Geolocalização



Resultados - Código

Todo o sistema faz uso de ferramentas e bibliotecas de código livre.

O sistema inteiro está disponível sob a licença APACHE de código livre no link:

<https://github.com/raulsenaferreira/Projeto-Myosotis>

Contribuições são bem vindas!



Trabalhos Futuros

O que ainda falta?

- Extrair os dados dos sites restantes
- Extrair mais informações dos dados (ex: bairros)
- Incluir séries históricas
- Fomentar a contribuição à ferramenta
- Tornar sua visibilidade maior
- Incluir mais tipos de gráficos e cruzamento de dados de outras bases



Referências

<http://www.desaparecidos.gov.br/index.php/statistics>

<http://www.mshp.dps.missouri.gov/CJ51/map.jsp>

http://s3.amazonaws.com/info-mongodb-com/MongoDB_BI_Analytics.pdf

● <http://apublica.org/2014/03/rio-de-janeiro-mapa-dos-desaparecimentos/>

<http://www.desaparecidosdobrasil.org/>

<http://www.caminhodevolta.fm.usp.br/index.php>



Referências

http://www.ssp.sp.gov.br/servicos/pessoas_desaparecidas.aspx

<http://www.mpsp.mp.br/portal/page/portal/Desaparecidos>

<http://www.sicride.pr.gov.br/modules/noticias/index.php>

http://www2.camara.leg.br/documentos-e-pesquisa/publicacoes/estnottec/areas-da-conle/tema21/H-Coord_Legislativa-Setex-Internet-2008_1686.pdf

<http://iidac.org/iidac/rede-nacional-de-identificacao-e-localizacao-de-criancas-e-adolescentes-desaparecidos?area=2>