

GP

PITCH Workshop de Lean Manufacturing

JUSTIFICATIVAS
Passado

Necessidade de coletar dados dos boletins gerados pelo ONS, pois as informações para elaboração de relatórios mais específicos que requerem o cruzamento de informações, através da construção de modelos de aprendizagem de máquina (Machine Learning) ou mineração de dados (Data Mining), ainda não é possível.



PRODUTO

Robô Web Crawling para extração e construção de cluster de dados do ONS, que serão usados para consultas via API RESTful.

STAKEHOLDERS
EXTERNOS
& Fatores externos

- Engenheiros Eletricistas;
- Engenheiros de Dados;
- Cientistas de Dados; e
- Universidades (para fins de pesquisa acadêmica).



PREMISSAS

- Possibilidade de realizar operações de Web Crawling ao domínio do ONS;
- Possibilidade de realizar consultas aos boletins diários do ONS;
- Imutabilidade dos padrões de relatórios gerados pelo Produto; e
- Imutabilidade dos padrões de estrutura dos dados relacionais extraídos pelo robô Web Crawling.



RISCOS

- Complexidade na implementação dos dados relacionais aos clusters devido as alterações estruturais dos boletins gerados quanto sua padronização e nomenclaturas oficiais do ONS; e
- Complexidade na implementação de robôs Web Crawling para extração de dados dos boletins diários do ONS devido a mecanismos anti-crawling utilizados como padrões de segurança cibernética.



OBJ SMART

Desenvolver uma API de arquitetura RESTful para consulta de dados dos boletins do ONS.



REQUISITOS

- Acesso aos endpoints da API RESTful;
- Consultas aos cluster de dados gerados pela extração de dados dos boletins diários do ONS;
- Ativação de rotinas para acionamento do robô de Web Crawling; e
- Geração de novos clusters de dados para cada ano e contexto de boletim diário do ONS.



EQUIPE

Nádio Pontes.

GRUPO DE
ENTREGAS

- Análise dos principais dados a serem estruturados no cluster segundo os boletins gerados e padronizados pelo ONS;
- Criação das bases de dados relacionais SQL para cada cluster;
- Construção do robô de extração de dados via Web Crawling ao domínio de boletins público do NOS através de técnicas Web Scraping; e
- Construção da API RESTful para consultas aos dados do cluster.



LINHA DO TEMPO

BENEFÍCIOS
Futuro

Facilidade na coleta de dados para geração de modelos de aprendizagem de máquina (Machine Learning) e algoritmos de mineração de dados (Data Mining) através de consultas nos clusters de dados do ONS através dessa API RESTful.



RESTRIÇÕES

Este Produto, não contempla: implementação de modelos de aprendizagem de máquina (Machine Learning); operações com mineração de dados (Data Mining) para correlação de dados; operações com sistemas distribuídos para Data Warehousing; e autenticações por OAuth, Bearer Token ou qualquer outra tecnologia de validação para consultas à API RESTful.



CUSTOS

Não existem custos.