

Information Retrieval



Gabriel Richter
Junior Nogueira

PROBLEMA DE NEGÓCIO

Os dados são gerados em todo instante e atualmente, isso acontece com as leis, todo dia existem entidades do governo desenvolvendo e validando novas leis a todo momento, porém, muitas dessas novas leis desenvolvidas não possuem um acesso fácil ou essas informações estão desenvolvidas em locais não confiáveis com difícil acesso a algumas instituições interessadas requerentes de informações como escolas, empreendimentos, entre outros terceiros que muitas das vezes não possuem profissionais capacitados para fazer a coleta, transformação e visualização desses dados.

O desafio é o fácil acesso a essas informações importantes, nesse sentido existem algumas implicações, são elas.

1. Necessidade de consultar diversas fontes que podem não ser confiáveis ou não trazer as informações necessárias levando desgaste e perda de tempo.
2. Falha na leitura ou interpretação dessas leis específicas.
3. Trabalho improdutivo devido a vários fatores já citados acima.

Referências Possíveis Referências.

Constituição Federal:

https://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm

Lei de Diretrizes e Bases da Educação Nacional:

https://www.planalto.gov.br/ccivil_03/leis/L9394.htm

Parâmetros Curriculares da Educação Física:

<http://portal.mec.gov.br/seb/arquivos/pdf/livro07.pdf>

Regulamentação da Educação Física:

<https://www.confef.org.br/confef/>

Estatuto da Criança e do Idoso:

https://www.planalto.gov.br/ccivil_03/leis/L8069.htm

Dicionários Jurídicos:

1. Dicionario_Juridico_-_J._M._Othon_Sidou_-_2016
2. http://www.integrawebsites.com.br/versao_1/arquivos/d8545a815ba082afcb4d6d067b471373.pdf

Como que você faria para ajudar essas instituições que necessitam desses dados utilizando os conhecimentos que você já sabe a fim de desenvolver uma solução de ponta a ponta para conseguir desenvolver e concluir essa demanda?

PROPOSTA DE SOLUÇÃO

Um processo de ETL que vai fazer a coleta dos dados das instituições citadas e de n documentos a fim de melhorar a base de dados com as devidas informações necessárias para fazer as devidas análises dessas leis;

1. CLT: Consolidação das Leis de Trabalho.
2. CF: Constituição Federal.
3. CC: Código Civil.

Após essa coleta, já com os dados limpos, apresentar esses dados em uma ferramenta visual que o usuário vai poder fazer a pesquisa da lei específica para mais detalhes.

Exemplo de Registro já transformado:

Código Civil: Art.117: Salvo se o permitir a lei ou o representado, é anulável o negócio jurídico que o representante, no seu interesse ou por conta de outrem, celebrar consigo mesmo.

Exemplo da Base de dados:

row_id	tipo_lei	sigla	titulo	paragrafo	artigo	artigo_index	body	link	scrapy_datetime
0	Constituição Federal de 1988 - CF88	CF88	TÍTULO I - Dos Princípios Fundamentais		Art. 1º	Art. 1º	A República ... fundamentos:	https://www.planalto.gov.br...	15/07/2022 15:07
1	Constituição Federal de 1988 - CF88	CF88	TÍTULO I - Dos Princípios Fundamentais		Art. 1º	Art. 1º I	A Soberania;	https://www.planalto.gov.br...	16/07/2022 15:07
2	Constituição Federal de 1988 - CF88	CF88	TÍTULO I - Dos Princípios Fundamentais		Art. 1º	Art. 1º II	A Cidadania;	https://www.planalto.gov.br...	17/07/2022 15:07
3	Constituição Federal de 1988 - CF88	CF88	TÍTULO I - Dos Princípios Fundamentais		Art. 1º	Art. 1º III	A Dignidade;	https://www.planalto.gov.br...	18/07/2022 15:07
4	Constituição Federal de 1988 - CF88	CF88	TÍTULO I - Dos Princípios Fundamentais		Art. 1º	Art. 1º IV	Os Valores Sociais do Trabalho;	https://www.planalto.gov.br...	19/07/2022 15:07
5	Constituição Federal de 1988 - CF88	CF88	TÍTULO I - Dos Princípios Fundamentais		Art. 1º	Art. 1º V	O Pluralismo Político;	https://www.planalto.gov.br...	20/07/2022 15:07
6	Constituição Federal de 1988 - CF88	CF88	TÍTULO I - Dos Princípios Fundamentais		Art. 1º	Art. 1º PU	Todo poder emana do povo...	https://www.planalto.gov.br...	21/07/2022 15:07

Ferramentas Utilizadas

1. Python ferramenta de desenvolvimento;
2. Webscrapping para coleta de dados;
3. NLP para processamento;
4. ElasticSearch para coleta de dados;
5. Streamlit para data viz;

Primeira Proposta Análise (Gabriel Richter 15/07/2022)