

Nome do Arquivo
text_handler.py
Documentação
<pre>#1[#1 TITULO: TEXTHANDLER #1 AUTOR: EDUARDO RIBEIRO SILVA DE OLIVEIRA #1 DATA: 07/10/2024 #1 VERSAO: 1 #1 FINALIDADE: REALIZAR EXTRAÇÃO E FILTRAGEM DE TEXTO BRUTO A PARTIR DE CONTEÚDO HTML #1 ENTRADAS: NOME DO DIRETÓRIO (NO CONSTRUTOR), CONTEÚDO HTML (NA FUNÇÃO EXTRACT RAW TEXT) #1 SAIDAS: TEXTO LIMPO EM MINÚSCULAS EXTRAÍDO DO CONTEÚDO HTML #1 ROTINAS CHAMADAS: EXTRACT_RAW_TEXT, _FILTER_PORTUGUESE_TEXT #1] #1[#1 ROTINA: __INIT__ #1 FINALIDADE: INICIALIZA A CLASSE TEXTHANDLER E CONFIGURA O DIRETÓRIO #1 ENTRADAS: NOME DO DIRETÓRIO (STRING) #1 DEPENDENCIAS: LOGGINGHANDLER #1 CHAMADO POR: TEXTHANDLER #1 CHAMA: LOGGINGHANDLER.__INIT__ #1] #1[#1 ROTINA: EXTRACT_RAW_TEXT #1 FINALIDADE: EXTRAI O TEXTO BRUTO DO CONTEÚDO HTML E O FILTRA PARA REMOVER CARACTERES INDESEJADOS #1 ENTRADAS: CONTEÚDO HTML (STRING) #1 DEPENDENCIAS: BEAUTIFULSOUP, RE #1 CHAMADO POR: USUÁRIO #1 CHAMA: _FILTER_PORTUGUESE_TEXT #1] #1[#1 ROTINA: _FILTER_PORTUGUESE_TEXT #1 FINALIDADE: FILTRA O TEXTO PARA REMOVER CARACTERES QUE NÃO SÃO DO PORTUGUÊS #1 ENTRADAS: TEXTO BRUTO (STRING) #1 DEPENDENCIAS: RE #1 CHAMADO POR: EXTRACT_RAW_TEXT #1 CHAMA: NENHUMA #1]</pre>

Nome do Arquivo
legislacao_prefeitura_sp_gov_br.py
Documentação
<pre>#1[#1 TITULO: LEGISLACAOPREFEITURASPGOVBR SCRAPPER #1 AUTOR: EDUARDO RIBEIRO SILVA DE OLIVEIRA #1 DATA: 07/10/2024 #1 VERSAO: 1 #1 FINALIDADE: REALIZAR O SCRAPING DE LEGISLAÇÕES NO SITE DA PREFEITURA DE SÃO PAULO COM BASE EM DIVERSOS ASSUNTOS #1 ENTRADAS: NOME DO DIRETÓRIO (NO CONSTRUTOR), ASSUNTO (NA FUNÇÃO SCRAPE) #1 SAIDAS: DADOS PROCESSADOS, LINKS EXTRAÍDOS, LEGISLAÇÕES VISITADAS #1 ROTINAS CHAMADAS: SCRAPE, BUILD_URL, PARSE_CONTENT, EXTRACT_LINKS, SCRAPE_LINKS, SCRAPE_ALL_SUBJECTS #1] #1[#1 ROTINA: __INIT__ #1 FINALIDADE: INICIALIZA A CLASSE LEGISLACAOPREFEITURASPGOVBR SCRAPPER E CONFIGURA AS VARIÁVEIS INICIAIS E O DIRETÓRIO DE SALVAMENTO #1 ENTRADAS: NENHUMA #1 DEPENDENCIAS: LOGGINGHANDLER, SCRAPINGHANDLER #1 CHAMADO POR: LEGISLACAOPREFEITURASPGOVBR SCRAPPER #1 CHAMA: SCRAPINGHANDLER.__INIT__ #1] #1[#1 ROTINA: SCRAPE #1 FINALIDADE: REALIZAR O SCRAPING DE LEGISLAÇÕES COM BASE NO ASSUNTO ESPECIFICADO</pre>

```
#1 ENTRADAS: ASSUNTO (STRING)
#1 DEPENDENCIAS: TIME, REQUESTINGHANDLER
#1 CHAMADO POR: USUÁRIO, SCRAPE ALL SUBJECTS
#1 CHAMA: BUILD_URL, MAKE_REQUEST (REQUESTINGHANDLER),
PROCESS_RESPONSE, PARSE_CONTENT, EXTRACT_LINKS, SCRAPE_LINKS
#1]
#1[
#1 ROTINA: BUILD_URL
#1 FINALIDADE: CONSTRUIR A URL PARA REALIZAR A REQUISIÇÃO HTTP
BASEADO NO ASSUNTO E NO NÚMERO DA PÁGINA
#1 ENTRADAS: ASSUNTO (STRING), NÚMERO DA PÁGINA (INT)
#1 DEPENDENCIAS: NENHUMA
#1 CHAMADO POR: SCRAPE
#1 CHAMA: NENHUMA
#1]
#1[
#1 ROTINA: PARSE_CONTENT
#1 FINALIDADE: ANALISAR O CONTEÚDO HTML OBTIDO NA REQUISIÇÃO E
TRANSFORMÁ-LO EM UM OBJETO BEAUTIFULSOUP
#1 ENTRADAS: CONTEÚDO HTML (STRING)
#1 DEPENDENCIAS: BEAUTIFULSOUP
#1 CHAMADO POR: SCRAPE
#1 CHAMA: NENHUMA
#1]
#1[
#1 ROTINA: EXTRACT_LINKS
#1 FINALIDADE: EXTRAIR LINKS RELEVANTES DO CONTEÚDO HTML ANALISADO
#1 ENTRADAS: CONTEÚDO HTML ANALISADO (OBJETO BEAUTIFULSOUP)
#1 DEPENDENCIAS: BEAUTIFULSOUP
#1 CHAMADO POR: SCRAPE
#1 CHAMA: NENHUMA
#1]
#1[
#1 ROTINA: SCRAPE_LINKS
#1 FINALIDADE: REALIZAR O SCRAPING DE CADA LINK EXTRAÍDO, SEGUINDO OS
LINKS E PROCESSANDO AS RESPOSTAS
#1 ENTRADAS: LISTA DE LINKS (LISTA DE STRINGS)
#1 DEPENDENCIAS: TIME, REQUESTINGHANDLER
#1 CHAMADO POR: SCRAPE
#1 CHAMA: MAKE_REQUEST (REQUESTINGHANDLER), PROCESS_RESPONSE
#1]
#1[
#1 ROTINA: SCRAPE_ALL_SUBJECTS
#1 FINALIDADE: REALIZAR O SCRAPING DE TODOS OS ASSUNTOS DEFINIDOS
SIMULTANEAMENTE UTILIZANDO MÚLTIPLAS THREADS
#1 ENTRADAS: NENHUMA (UTILIZA OS ASSUNTOS DA VARIÁVEL DE CLASSE
SUBJECTS)
#1 DEPENDENCIAS: THREADPOOLEXECUTOR
#1 CHAMADO POR: USUÁRIO
#1 CHAMA: SCRAPE (PARA CADA ASSUNTO)
#1]
```

Nome do Arquivo
main.py
Documentação
#1[#1 TITULO: MAIN #1 AUTOR: EDUARDO RIBEIRO SILVA DE OLIVEIRA #1 DATA: 07/10/2024 #1 VERSAO: 1 #1 FINALIDADE: INICIALIZAR O HANDLER DE DATAFRAMES E EXECUTAR SUAS FUNÇÕES #1 ENTRADAS: NENHUMA #1 SAIDAS: EXECUÇÃO DAS FUNÇÕES DO DATAFRAMEHANDLER #1 ROTINAS CHAMADAS: EXECUTE (DATAFRAMEHANDLER) #1] #1[#1 ROTINA: MAIN #1 FINALIDADE: CRIA UMA INSTÂNCIA DO DATAFRAMEHANDLER E EXECUTA SUAS FUNÇÕES #1 ENTRADAS: NENHUMA #1 DEPENDENCIAS: DATAFRAMEHANDLER #1 CHAMADO POR: SCRIPT PRINCIPAL #1 CHAMA: EXECUTE (DATAFRAMEHANDLER) #1]

Nome do Arquivo
scraping_handler.py
Documentação
<pre>#1[#1 TITULO: SCRAPINGHANDLER #1 AUTOR: EDUARDO RIBEIRO SILVA DE OLIVEIRA #1 DATA: 07/10/2024 #1 VERSAO: 1 #1 FINALIDADE: REALIZAR O PROCESSO DE SCRAPING, EXTRAÇÃO DE TEXTO, LEMATIZAÇÃO E SALVAMENTO DOS DADOS EM ARQUIVOS JSON #1 ENTRADAS: NOME DO DIRETÓRIO (NO CONSTRUTOR), DICIONÁRIO DATA_DICT (NA FUNÇÃO PROCESS_RESPONSE) #1 SAIDAS: DICIONÁRIO PROCESSADO COM CAMPOS EXTRAÍDOS E ARQUIVO JSON COM AS INFORMAÇÕES SALVAS #1 ROTINAS CHAMADAS: PROCESS_RESPONSE, SAVE_REQUEST_INFO (SAVINGHANDLER), EXTRACT_RAW_TEXT (TEXTHANDLER), LEMMATIZE_AND_EXTRACT_TOPONYMS (TOPONYMHANDLER) #1] #1[#1 ROTINA: __INIT__ #1 FINALIDADE: INICIALIZA A CLASSE SCRAPINGHANDLER E CONFIGURA AS DEPENDÊNCIAS NECESSÁRIAS #1 ENTRADAS: NOME DO DIRETÓRIO (STRING) #1 DEPENDENCIAS: REQUESTINGHANDLER, SAVINGHANDLER, TEXTHANDLER, TOPONYMHANDLER, LOGGINGHANDLER #1 CHAMADO POR: SCRAPINGHANDLER #1 CHAMA: LOGGINGHANDLER.__INIT__, REQUESTINGHANDLER, SAVINGHANDLER, TEXTHANDLER, TOPONYMHANDLER #1] #1[#1 ROTINA: PROCESS_RESPONSE #1 FINALIDADE: PROCESSA O DICIONÁRIO DE RESPOSTA, EXTRAINDO TEXTO BRUTO, LEMATIZANDO E EXTRAINDO TOPÔNIMOS, E SALVA AS INFORMAÇÕES #1 ENTRADAS: DICIONÁRIO DATA_DICT (COM CAMPOS COMO RESPONSE_STRING) #1 DEPENDENCIAS: TEXTHANDLER, TOPONYMHANDLER, SAVINGHANDLER #1 CHAMADO POR: USUÁRIO #1 CHAMA: EXTRACT_RAW_TEXT (TEXTHANDLER), LEMMATIZE_AND_EXTRACT_TOPONYMS (TOPONYMHANDLER), SAVE_REQUEST_INFO (SAVINGHANDLER) #1]</pre>

Nome do Arquivo
saving_handler.py
Documentação
<pre>#1[#1 TITULO: SAVINGHANDLER #1 AUTOR: EDUARDO RIBEIRO SILVA DE OLIVEIRA #1 DATA: 07/10/2024 #1 VERSAO: 1 #1 FINALIDADE: GERENCIAR A CRIAÇÃO DE DIRETÓRIOS E O SALVAMENTO DE INFORMAÇÕES DE REQUISIÇÕES EM ARQUIVOS JSON #1 ENTRADAS: NOME DO DIRETÓRIO (NO CONSTRUTOR), DICIONÁRIO REQUEST_INFO (NA FUNÇÃO SAVE_REQUEST_INFO) #1 SAIDAS: CAMINHO DO ARQUIVO JSON GERADO COM AS INFORMAÇÕES DA REQUISIÇÃO #1 ROTINAS CHAMADAS: _CREATE_DIRECTORY, SAVE_REQUEST_INFO #1] #1[#1 ROTINA: __INIT__ #1 FINALIDADE: INICIALIZA A CLASSE SAVINGHANDLER E CRIA O DIRETÓRIO DE SALVAMENTO SE NÃO EXISTIR #1 ENTRADAS: NOME DO DIRETÓRIO (STRING) #1 DEPENDENCIAS: LOGGINGHANDLER, OS #1 CHAMADO POR: SAVINGHANDLER #1 CHAMA: LOGGINGHANDLER.__INIT__, _CREATE_DIRECTORY #1] #1[#1 ROTINA: _CREATE_DIRECTORY #1 FINALIDADE: VERIFICA SE O DIRETÓRIO EXISTE E O CRIA SE NECESSÁRIO #1 ENTRADAS: NENHUMA #1 DEPENDENCIAS: OS</pre>

```
#1 CHAMADO POR: __INIT__
#1 CHAMA: OS.MAKEDIRS
#1]
#1[
#1 ROTINA: SAVE_REQUEST_INFO
#1 FINALIDADE: SALVA AS INFORMAÇÕES DA REQUISIÇÃO EM UM ARQUIVO JSON
DENTRO DO DIRETÓRIO ESPECIFICADO
#1 ENTRADAS: DICIONÁRIO REQUEST_INFO (CONTENDO DADOS DA REQUISIÇÃO)
#1 DEPENDENCIAS: JSON, OS, DATETIME
#1 CHAMADO POR: USUÁRIO
#1 CHAMA: NENHUMA
#1]
```

Nome do Arquivo
dataframe_handler.py
Documentação
<pre>#1[#1 TITULO: DATAFRAMEHANDLER #1 AUTOR: EDUARDO RIBEIRO SILVA DE OLIVEIRA #1 DATA: 07/10/2024 #1 VERSAO: 1 #1 FINALIDADE: MANIPULAR E PROCESSAR ARQUIVOS JSON, TRANSFORMANDO-OS EM DATAFRAMES E EXTRAINDO INFORMAÇÕES ÚTEIS. #1 ENTRADAS: NENHUMA #1 SAIDAS: ARQUIVOS CSV GERADOS A PARTIR DE DATAFRAMES #1 ROTINAS CHAMADAS: LOAD_JSON_FILES, _PROCESS_AND_SAVE_DATAFRAME, EXTRACT_URL_INFORMATION, _EXTRACT_INSTITUTION_NAME, _EXTRACT_SUBJECT, EXECUTE #1] #1[#1 ROTINA: __INIT__ #1 FINALIDADE: INICIALIZA AS VARIÁVEIS DA CLASSE DATAFRAMEHANDLER E CHAMA O CONSTRUTOR DA CLASSE PAI. #1 ENTRADAS: NENHUMA #1 DEPENDENCIAS: LOGGINGHANDLER, OS #1 CHAMADO POR: DATAFRAMEHANDLER #1 CHAMA: LOGGINGHANDLER.__INIT__ #1] #1[#1 ROTINA: LOAD_JSON_FILES #1 FINALIDADE: CARREGA TODOS OS ARQUIVOS JSON DO DIRETÓRIO E OS PROCESSA. #1 ENTRADAS: NENHUMA #1 DEPENDENCIAS: OS, JSON, TQDM, RE #1 CHAMADO POR: EXECUTE #1 CHAMA: _PROCESS_AND_SAVE_DATAFRAME #1] #1[#1 ROTINA: _PROCESS_AND_SAVE_DATAFRAME #1 FINALIDADE: PROCESSA OS DADOS E SALVA EM ARQUIVOS CSV NO DIRETÓRIO CORRESPONDENTE. #1 ENTRADAS: LISTA DE DICIONÁRIOS COM OS DADOS JSON E O CAMINHO DO DIRETÓRIO #1 DEPENDENCIAS: PANDAS, OS #1 CHAMADO POR: LOAD_JSON_FILES #1 CHAMA: EXTRACT_URL_INFORMATION #1] #1[#1 ROTINA: EXTRACT_URL_INFORMATION #1 FINALIDADE: EXTRAI INFORMAÇÕES RELEVANTES DAS URLS PRESENTES NO DATAFRAME. #1 ENTRADAS: DATAFRAME COM A COLUNA 'URL' #1 DEPENDENCIAS: RE #1 CHAMADO POR: _PROCESS_AND_SAVE_DATAFRAME #1 CHAMA: _EXTRACT_INSTITUTION_NAME, _EXTRACT_SUBJECT #1] #1[#1 ROTINA: _EXTRACT_INSTITUTION_NAME #1 FINALIDADE: IDENTIFICA O NOME DA INSTITUIÇÃO BASEADO NO DOMÍNIO DA URL. #1 ENTRADAS: BASE_URL (STRING) #1 DEPENDENCIAS: RE #1 CHAMADO POR: EXTRACT_URL_INFORMATION #1 CHAMA: NENHUMA #1] #1[#1 ROTINA: _EXTRACT_SUBJECT</pre>

```
#1 FINALIDADE: EXTRAÍ O ASSUNTO DA URL COM BASE NO NOME DA INSTITUIÇÃO.
#1 ENTRADAS: URL (STRING), NOME DA INSTITUIÇÃO (STRING)
#1 DEPENDENCIAS: RE
#1 CHAMADO POR: EXTRACT_URL_INFORMATION
#1 CHAMA: NENHUMA
#1]
#1[
#1 ROTINA: EXECUTE
#1 FINALIDADE: EXECUTA A ROTINA PRINCIPAL DE CARREGAMENTO E PROCESSAMENTO DOS ARQUIVOS JSON.
#1 ENTRADAS: NENHUMA
#1 DEPENDENCIAS: LOGGINGHANDLER
#1 CHAMADO POR: USUÁRIO
#1 CHAMA: LOAD_JSON_FILES
#1]
```

Nome do Arquivo
leis_municipais_camara_sp.py
Documentação
<pre>#1[#1 TITULO: LEISMUNICIPAISCAMARASP SCRAPPER #1 AUTOR: EDUARDO RIBEIRO SILVA DE OLIVEIRA #1 DATA: 07/10/2024 #1 VERSAO: 1 #1 FINALIDADE: REALIZAR O SCRAPING DE DECRETOS NO SITE LEISMUNICIPAIS.COM.BR REFERENTE À CÂMARA DE SÃO PAULO, COM BASE EM DIVERSOS ASSUNTOS #1 ENTRADAS: NOME DO DIRETÓRIO (NO CONSTRUTOR), ASSUNTO (NA FUNÇÃO SCRAPE) #1 SAIDAS: DADOS PROCESSADOS, LINKS EXTRAÍDOS, DECRETOS VISITADOS #1 ROTINAS CHAMADAS: SCRAPE, BUILD_URL, PARSE_CONTENT, EXTRACT_LINKS, SCRAPE_LINKS, SCRAPE_ALL_SUBJECTS #1] #1[#1 ROTINA: __INIT__ #1 FINALIDADE: INICIALIZA A CLASSE LEISMUNICIPAISCAMARASP SCRAPPER E CONFIGURA AS VARIÁVEIS INICIAIS E O DIRETÓRIO DE SALVAMENTO #1 ENTRADAS: NENHUMA #1 DEPENDENCIAS: LOGGINGHANDLER, SCRAPINGHANDLER #1 CHAMADO POR: LEISMUNICIPAISCAMARASP SCRAPPER #1 CHAMA: SCRAPINGHANDLER.__INIT__ #1] #1[#1 ROTINA: SCRAPE #1 FINALIDADE: REALIZAR O SCRAPING DE DECRETOS COM BASE NO ASSUNTO ESPECIFICADO #1 ENTRADAS: ASSUNTO (STRING) #1 DEPENDENCIAS: TIME, REQUESTINGHANDLER #1 CHAMADO POR: USUÁRIO, SCRAPE_ALL_SUBJECTS #1 CHAMA: BUILD_URL, MAKE_REQUEST (REQUESTINGHANDLER), PROCESS_RESPONSE, PARSE_CONTENT, EXTRACT_LINKS, SCRAPE_LINKS #1] #1[#1 ROTINA: BUILD_URL #1 FINALIDADE: CONSTRUIR A URL PARA REALIZAR A REQUISIÇÃO HTTP BASEADO NO ASSUNTO E NO NÚMERO DA PÁGINA #1 ENTRADAS: ASSUNTO (STRING), NÚMERO DA PÁGINA (INT) #1 DEPENDENCIAS: NENHUMA #1 CHAMADO POR: SCRAPE #1 CHAMA: NENHUMA #1] #1[#1 ROTINA: PARSE_CONTENT #1 FINALIDADE: ANALISAR O CONTEÚDO HTML OBTIDO NA REQUISIÇÃO E TRANSFORMÁ-LO EM UM OBJETO BEAUTIFULSOUP #1 ENTRADAS: CONTEÚDO HTML (STRING) #1 DEPENDENCIAS: BEAUTIFULSOUP #1 CHAMADO POR: SCRAPE #1 CHAMA: NENHUMA #1] #1[#1 ROTINA: EXTRACT_LINKS #1 FINALIDADE: EXTRAIR LINKS RELEVANTES DO CONTEÚDO HTML ANALISADO #1 ENTRADAS: CONTEÚDO HTML ANALISADO (OBJETO BEAUTIFULSOUP) #1 DEPENDENCIAS: BEAUTIFULSOUP #1 CHAMADO POR: SCRAPE</pre>

```
#1 CHAMA: NENHUMA
#1]
#1[
#1 ROTINA: SCRAPE_LINKS
#1 FINALIDADE: REALIZAR O SCRAPING DE CADA LINK EXTRAÍDO, SEGUINDO OS
LINKS E PROCESSANDO AS RESPOSTAS
#1 ENTRADAS: LISTA DE LINKS (LISTA DE STRINGS)
#1 DEPENDENCIAS: TIME, REQUESTINGHANDLER
#1 CHAMADO POR: SCRAPE
#1 CHAMA: MAKE_REQUEST (REQUESTINGHANDLER), PROCESS_RESPONSE
#1]
#1[
#1 ROTINA: SCRAPE_ALL_SUBJECTS
#1 FINALIDADE: REALIZAR O SCRAPING DE TODOS OS ASSUNTOS DEFINIDOS
SIMULTANEAMENTE UTILIZANDO MÚTIPLAS THREADS
#1 ENTRADAS: NENHUMA (UTILIZA OS ASSUNTOS DA VARIÁVEL DE CLASSE
SUBJECTS)
#1 DEPENDENCIAS: THREADPOOLEXECUTOR
#1 CHAMADO POR: USUÁRIO
#1 CHAMA: SCRAPE (PARA CADA ASSUNTO)
#1]
```

Nome do Arquivo
logging_handler.py
Documentação
<pre>#1[#1 TITULO: LOGGING HANDLER #1 AUTOR: EDUARDO RIBEIRO SILVA DE OLIVEIRA #1 DATA: 07/10/2024 #1 VERSAO: 1 #1 FINALIDADE: GERENCIA O LOGGING PARA UMA CLASSE, PERMITINDO A CRIACAO DE LOGS E FORMATO ESPECIFICO, E LOGAR EXECUCAO DE METODOS COM PARAMETROS E RESULTADOS #1 ENTRADAS: NOME DO DIRETORIO PARA ARMAZENAR OS LOGS, NOME DA CLASSE, NOME DO METODO, OPCOES PARA LOGAR PARAMETROS E RESULTADOS #1 SAIDAS: LOGS EM ARQUIVO DE TEXTO #1 ROTINAS CHAMADAS: __INIT__, LOG_METHOD #1 DEPENDENCIAS: LOGGING, OS #1] #1[#1 ROTINA: __INIT__ #1 FINALIDADE: INICIALIZA O MANIPULADOR DE LOGS E CONFIGURA O FORMATO DE LOGS PARA A CLASSE LOGGINGHANDLER #1 ENTRADAS: NOME DO DIRETORIO PARA ARMAZENAR OS LOGS #1 DEPENDENCIAS: LOGGING, OS #1 CHAMADO POR: LOGGINGHANDLER #1 CHAMA: OS.MAKEDIRS (SE O DIRETORIO NAO EXISTIR), LOGGING.GETLOGGER, LOGGING.FILEHANDLER, LOGGING.FORMATTER #1] #1[#1 ROTINA: LOG_METHOD #1 FINALIDADE: DECORADOR PARA LOGAR A EXECUCAO DE METODOS, COM OPCOES PARA EXIBIR PARAMETROS E RESULTADOS #1 ENTRADAS: NOME DA CLASSE, NOME DO METODO, OPCOES DE EXIBIR PARAMETROS E EXIBIR RESULTADO #1 DEPENDENCIAS: LOGGING #1 CHAMADO POR: LOGGINGHANDLER #1 CHAMA: FUNC (METODO DECORADO) #1]</pre>

Nome do Arquivo
requesting_handler.py
Documentação
<pre>#1[#1 TITULO: REQUESTINGHANDLER #1 AUTOR: EDUARDO RIBEIRO SILVA DE OLIVEIRA #1 DATA: 07/10/2024 #1 VERSAO: 1 #1 FINALIDADE: REALIZAR REQUISIÇÕES HTTP (GET E POST) E LOGAR O TEMPO DE RESPOSTA E OUTRAS INFORMAÇÕES RELEVANTES</pre>

```
#1 ENTRADAS: NOME DO DIRETÓRIO (NO CONSTRUTOR), MÉTODO HTTP, URL, KWARGS
(DADOS ADICIONAIS OPCIONAIS)
#1 SAIDAS: DICIONÁRIO CONTENDO DADOS DA REQUISIÇÃO E RESPOSTA, INCLUINDO
TEMPO, TAMANHO E ERROS (SE HOVER)
#1 ROTINAS CHAMADAS: SET_METHOD_MAPPING, _REQUEST_WRAPPER, MAKE_REQUEST
#1]

#1[
#1 ROTINA: __INIT__
#1 FINALIDADE: INICIALIZA A CLASSE REQUESTINGHANDLER E CONFIGURA O
MAPEAMENTO DE MÉTODOS HTTP
#1 ENTRADAS: NOME DO DIRETÓRIO (STRING)
#1 DEPENDENCIAS: LOGGINGHANDLER
#1 CHAMADO POR: REQUESTINGHANDLER
#1 CHAMA: LOGGINGHANDLER.__INIT__, SET_METHOD_MAPPING
#1]

#1[
#1 ROTINA: SET_METHOD_MAPPING
#1 FINALIDADE: DEFINE O MAPEAMENTO DOS MÉTODOS HTTP (GET E POST)
#1 ENTRADAS: NENHUMA
#1 DEPENDENCIAS: REQUESTS
#1 CHAMADO POR: __INIT__
#1 CHAMA: NENHUMA
#1]

#1[
#1 ROTINA: _REQUEST_WRAPPER
#1 FINALIDADE: ENVOLVE A EXECUÇÃO DA REQUISIÇÃO HTTP E COLETA DADOS
COMO TEMPO DE RESPOSTA E TAMANHO DA RESPOSTA
#1 ENTRADAS: FUNÇÃO DO MÉTODO HTTP, URL, DICIONÁRIO REQUEST_INFO,
KWARGS OPCIONAIS
#1 DEPENDENCIAS: TIME, REQUESTS
#1 CHAMADO POR: MAKE_REQUEST
#1 CHAMA: MÉTODOS HTTP (GET OU POST), RAISE_FOR_STATUS (REQUESTS)
#1]

#1[
#1 ROTINA: MAKE_REQUEST
#1 FINALIDADE: FAZ A REQUISIÇÃO HTTP USANDO O MÉTODO ESPECIFICADO
(GET OU POST) E RETORNA AS INFORMAÇÕES DA REQUISIÇÃO
#1 ENTRADAS: MÉTODO (GET OU POST), URL, KWARGS OPCIONAIS
#1 DEPENDENCIAS: DATETIME, REQUESTS, TIME
#1 CHAMADO POR: USUÁRIO
#1 CHAMA: _REQUEST_WRAPPER
#1]
```

Nome do Arquivo
toponym_handler.py
Documentação
<pre>#1[#1 TITULO: TOPONYMHANDLER #1 AUTOR: EDUARDO RIBEIRO SILVA DE OLIVEIRA #1 DATA: 07/10/2024 #1 VERSAO: 1 #1 FINALIDADE: LEMATIZAR TEXTOS E EXTRAIR TOPÔNIMOS (NOMES DE LOCAIS) USANDO PADRÕES DEFINIDOS #1 ENTRADAS: NOME DO DIRETÓRIO (NO CONSTRUTOR), TEXTO (NA FUNÇÃO LEMMATIZE_AND_EXTRACT_TOPONYMS) #1 SAIDAS: TEXTO LEMATIZADO E LISTA DE TOPÔNIMOS EXTRAÍDOS #1 ROTINAS CHAMADAS: _ADD_PATTERNS, LEMMATIZE_AND_EXTRACT_TOPONYMS #1] #1[#1 ROTINA: __INIT__ #1 FINALIDADE: INICIALIZA A CLASSE TOPONYMHANDLER E CONFIGURA O PROCESSAMENTO DE LINGUAGEM NATURAL E PADRÕES DE TOPÔNIMOS #1 ENTRADAS: NOME DO DIRETÓRIO (STRING) #1 DEPENDENCIAS: SPACY, MATCHER, LOGGINGHANDLER #1 CHAMADO POR: TOPONYMHANDLER #1 CHAMA: LOGGINGHANDLER.__INIT__, SPACY.LOAD, MATCHER #1] #1[#1 ROTINA: _ADD_PATTERNS #1 FINALIDADE: ADICIONA PADRÕES DE TOPÔNIMOS AO MATCHER USANDO NOMES PRÓPRIOS E PALAVRAS-CHAVE RELACIONADAS A LOCAIS #1 ENTRADAS: NENHUMA #1 DEPENDENCIAS: SPACY, RE #1 CHAMADO POR: __INIT__ #1 CHAMA: MATCHER.ADD #1] #1[</pre>

<pre>#1 ROTINA: LEMMATIZE AND EXTRACT TOPONYMS #1 FINALIDADE: LEMATIZA O TEXTO E EXTRAÍ OS TOPÔNIMOS USANDO O MATCHER #1 ENTRADAS: TEXTO (STRING) #1 DEPENDENCIAS: SPACY, MATCHER #1 CHAMADO POR: USUÁRIO #1 CHAMA: MATCHER, NLP #1]</pre>
--

Nome do Arquivo
alesp.py
Documentação
<pre>#1[#1 TITULO: ALESP SCRAPPER #1 AUTOR: EDUARDO RIBEIRO SILVA DE OLIVEIRA #1 DATA: 07/10/2024 #1 VERSAO: 1 #1 FINALIDADE: REALIZAR O SCRAPING DE NORMAS DA ASSEMBLEIA LEGISLATIVA DO ESTADO DE SÃO PAULO (ALESP) COM BASE EM DETERMINADOS ASSUNTOS #1 ENTRADAS: NOME DO DIRETÓRIO, ASSUNTO (NA FUNÇÃO SCRAPE) #1 SAIDAS: DADOS DE RESPOSTA PROCESSADOS, URLS VISITADAS #1 ROTINAS CHAMADAS: SCRAPE, BUILD_URL, PARSE_CONTENT, EXTRACT_LINKS, SCRAPE_LINKS, SCRAPE_ALL_SUBJECTS #1] #1[#1 ROTINA: __INIT__ #1 FINALIDADE: INICIALIZA A CLASSE ALESP SCRAPPER E CONFIGURA AS VARIÁVEIS INICIAIS E O DIRETÓRIO DE SALVAMENTO #1 ENTRADAS: NENHUMA #1 DEPENDENCIAS: LOGGINGHANDLER, SCRAPINGHANDLER #1 CHAMADO POR: ALESP SCRAPPER #1 CHAMA: SCRAPINGHANDLER.__INIT__ #1] #1[#1 ROTINA: SCRAPE #1 FINALIDADE: REALIZAR O SCRAPING DE NORMAS COM BASE NO ASSUNTO ESPECIFICADO #1 ENTRADAS: ASSUNTO (STRING) #1 DEPENDENCIAS: URLLIB, TIME, REQUESTINGHANDLER #1 CHAMADO POR: USUÁRIO, SCRAPE_ALL_SUBJECTS #1 CHAMA: BUILD_URL, MAKE_REQUEST (REQUESTINGHANDLER), PROCESS_RESPONSE, PARSE_CONTENT, EXTRACT_LINKS, SCRAPE_LINKS #1] #1[#1 ROTINA: BUILD_URL #1 FINALIDADE: CONSTRUIR A URL PARA REALIZAR A REQUISIÇÃO HTTP BASEADO NO ASSUNTO E NO NÚMERO DA PÁGINA #1 ENTRADAS: NÚMERO DA PÁGINA (INT), ASSUNTO ESCAPADO (STRING), ASSUNTO DECODIFICADO (STRING) #1 DEPENDENCIAS: NENHUMA #1 CHAMADO POR: SCRAPE #1 CHAMA: NENHUMA #1] #1[#1 ROTINA: PARSE_CONTENT #1 FINALIDADE: ANALISAR O CONTEÚDO HTML OBTIDO NA REQUISIÇÃO E TRANSFORMÁ-LO EM UM OBJETO BEAUTIFULSOUP #1 ENTRADAS: CONTEÚDO HTML (STRING) #1 DEPENDENCIAS: BEAUTIFULSOUP #1 CHAMADO POR: SCRAPE #1 CHAMA: NENHUMA #1] #1[#1 ROTINA: EXTRACT_LINKS #1 FINALIDADE: EXTRAIR LINKS RELEVANTES DO CONTEÚDO HTML ANALISADO #1 ENTRADAS: CONTEÚDO HTML ANALISADO (OBJETO BEAUTIFULSOUP) #1 DEPENDENCIAS: BEAUTIFULSOUP #1 CHAMADO POR: SCRAPE #1 CHAMA: NENHUMA #1] #1[#1 ROTINA: SCRAPE_LINKS #1 FINALIDADE: REALIZAR O SCRAPING DE CADA LINK EXTRAÍDO, SEGUINDO OS LINKS E PROCESSANDO AS RESPOSTAS #1 ENTRADAS: LISTA DE LINKS (LISTA DE STRINGS) #1 DEPENDENCIAS: TIME, REQUESTINGHANDLER #1 CHAMADO POR: SCRAPE</pre>


```
#1 CHAMA: MAKE_REQUEST (REQUESTINGHANDLER), PROCESS_RESPONSE
#1]
#1[
#1 ROTINA: SCRAPE_ALL_SUBJECTS
#1 FINALIDADE: REALIZAR O SCRAPING DE TODOS OS ASSUNTOS DEFINIDOS NA
VARIÁVEL SUBJECTS UTILIZANDO MÚLTIPLAS THREADS
#1 ENTRADAS: NENHUMA (UTILIZA OS ASSUNTOS DA VARIÁVEL DE CLASSE
SUBJECTS)
#1 DEPENDENCIAS: THREADPOOLEXECUTOR
#1 CHAMADO POR: USUÁRIO, FUNÇÃO MAIN
#1 CHAMA: SCRAPE (PARA CADA ASSUNTO)
#1]
```

Nome do Arquivo
logs.txt
Documentação