

DataSourceType

DATABASE WEB_SOURCE DataType

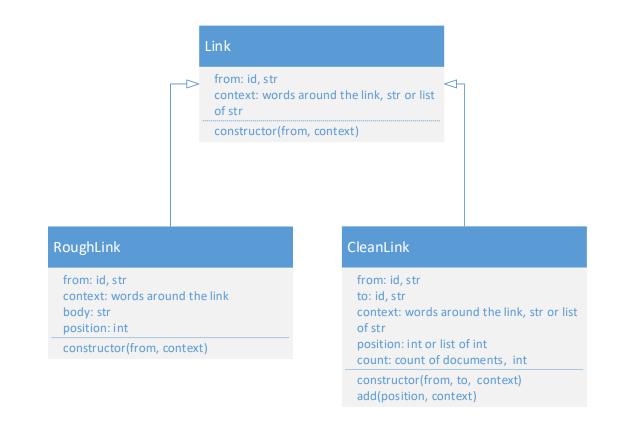
DOCUMENT_HEADER
DOCUMENT_TEXT

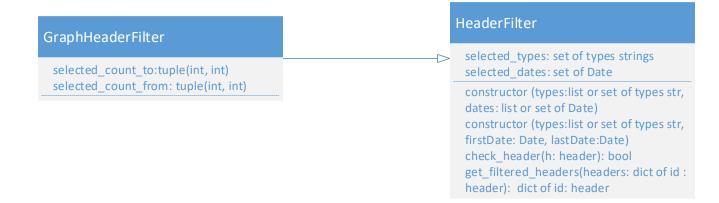
ataSour

data_source_name: str data_source_type: DataSourceType get_data(dataUid, dataType: DataType) get_all_data(dataType)

WebCrawler

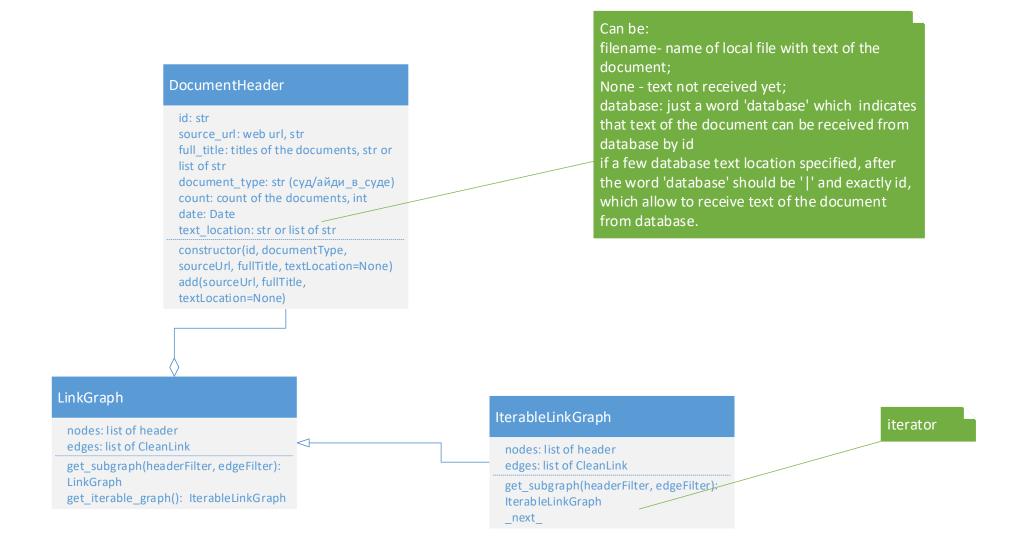
available_sources: set of DataSource
get_data_source(name): DataSource





LinkFinding

get_rough_links_for_multiple_document s(headers, webCrawler) get_rough_links(header, webCrawle)



GraphEdgeFilter

selected_filter_from: HeaderFilter
selected_filter_to: HeaderFilter
weight_diapason: tuple (min weight, max
weight)

check_edge(edge): bool
get_allowed_edges(edges)): list of
CleanLink
constructor(headerFilterFrom,
headerFilterTo, weights)

LinkAnalyzi

get_clean_links(roughLinks:
 {DocumentHeader:RoughLink}, headers):
 {DocumentHeader: CleanLink}
 get_links_graph(links): LinkGraph

ApiModule

process_period(firstDate, lastDate, headersFilters=None, edgeFilters=None, nodeFilters=None)

start_process_with(id, depth, headersFilters=None, edgeFilters=None, nodeFilters=None))

