

Database model documentation

Table of contents

1. Model details	3
2. Tables	4
1.1. Table agendador	4
1.2. Table api	4
1.3. Table apidados	5
1.4. Table formato	6
1.5. Table jornalista	6
1.6. Table noticia	6
1.7. Table portal	7
1.8. Table sinonimo	8
1.9. Table tag	9
1.10. Table api_tag	9
1.11. Table api_dados_tag	10
1.12. Table portal_tag	10
1.13. Table noticia_tag	10
3. References	12
2.1. Reference FK_0	12
2.2. Reference FK_2	12
2.3. Reference FK_4	12
2.4. Reference FK_5	12
2.5. Reference FK_7	12
2.6. Reference FK_8	12
2.7. Reference FK_9	12
2.8. Reference FK_10	12
2.9. Reference FK_11	13
2.10. Reference FK_12	13
2.11. Reference FK_13	13
2.12. Reference FK_14	13
2.13. Reference FK_16	13
2.14. Reference FK_18	13
2.15. Reference FK_21	13

1. Model details

Model name:

sprint4

Version:

2.4

Database engine:

MySQL

Description:

2. Tables

2.1. Table agendador

Description:

Armazena informações sobre o agendamento do período de scraping feito nos portais e APIs, podendo ser diário, semanal ou mensal

2.1.1. Columns

Column name	Type	Properties	Description
id	int	PK	Identificador único da tabela, chave primária, auto-incremento
tipo	varchar(255)	null	Tipo de agendador, valor textual (Diário, Semanal e Mensal)
quantidade	int	null	Quantidade associada ao agendador, valor inteiro (1 para Diário, 7 para Semanal e 30 para Mensal)

2.2. Table api

Description:

Armazena informações sobre as APIs, como nome, descrição, URL, e dados de criação e atualização

2.2.1. Columns

Column name	Type	Properties	Description
id	int	PK	Identificador único da tabela, chave primária, auto-incremento
nome	varchar(255)	null	Nome da API, valor textual, existe validação que impede valores repetidos
descricao	varchar(1000)	null	Descrição escrita pelo usuário sobre a API e os dados que ela retorna, valor textual
url	varchar(255)	null	URL da API, valor textual, existe validação que impede URLs repetidas

datacriacao	date	null	Data de criação da API. Como a API não tem data de publicação (na maioria das vezes), utilizamos a data de salvamento dos dados da API no banco
ultimaatualizacao	date	null	Data da última atualização da API, atualizada automaticamente no banco quando feito o scraping dos dados novos, respeitando o agendamento
agendador_id	int	null	Chave estrangeira que referencia o agendador (id), valor inteiro
formato_id	int	null	Chave estrangeira que referencia o formato da API (id), valor inteiro
ativo	boolean	null	Indicador booleano (verdadeiro/falso) que indica se a API está ativa (0 para desativo e 1 para ativo)

2.3. Table apidados

Description:

Armazena dados coletados de uma API específica, por meio da requisição e scraping

2.3.1. Columns

Column name	Type	Properties	Description
id	int	PK	Identificador único da tabela, chave primária, auto-incremento
conteudo	text	null	Conteúdo dos dados coletados, valor textual de tamanho variável
descricao	varchar(255)	null	Descrição opcional cadastrada pelo usuário, valor textual
datacoleta	date	null	Data da coleta dos dados da API, pois normalmente não possuem uma data de publicação, usada para o filtro por data

api_id	int		Chave estrangeira que referencia os dados à API que pertencem, valor inteiro (não nulo).
--------	-----	--	--

2.4. Table formato

Description:

Armazena informações sobre o formato das APIs, para formatar corretamente os dados armazenados

2.4.1. Columns

Column name	Type	Properties	Description
id	int	PK	Identificador único da tabela, chave primária, auto-incremento
nome	varchar(255)	null	Nome do formato, podendo ser CSV, JSON ou XML, valor textual

2.5. Table jornalista

2.5.1. Columns

Column name	Type	Properties	Description
id	int	PK	
nome	varchar(255)	null	

2.5.2. Alternate keys

Key name	Columns	Description
AK_0	nome	

2.6. Table noticia

Description:

Armazena informações sobre notícias publicadas por meio de web scraping, com associação a portais e jornalistas

2.6.1. Columns

Column name	Type	Properties	Description
id	int	PK	Identificador único da tabela, chave primária, auto-incremento

titulo	varchar(255)	null	Título da notícia, valor textual
data	date	null	Data de publicação da notícia
conteudo	varchar(20000)	null	Conteúdo da notícia, valor textual de até 20.000 caracteres
href	varchar(255)	null	Link único para a notícia, valor textual
portal_id	int	null	Chave estrangeira que referencia o portal onde a notícia foi publicada, valor inteiro
jornalista_id	int	null	Chave estrangeira que referencia o jornalista associado à notícia, valor inteiro

2.6.2. Alternate keys

Key name	Columns	Description
AK_1	href	

2.7. Table portal

Description:

Armazena informações sobre os portais de notícias, incluindo suas informações de URL e seletores de dados, status, agendamento e datas de cadastro e atualização

2.7.1. Columns

Column name	Type	Properties	Description
id	int	PK	Identificador único da tabela, chave primária, auto-incremento
nome	varchar(255)	null	Nome do portal cadastrado pelo usuário, valor textual
url	varchar(255)	null	URL do portal cadastrado pelo cliente, valor textual
seletorjornalista	varchar(254)	null	Seletor HTML do portal, inspecionado pelo usuário, referente ao jornalista da notícia, valor textual

seletordatapublicacao	varchar(254)	null	Seletor HTML do portal, inspecionado pelo usuário, referente a data de publicação da notícia, valor textual
seletorconteudo	varchar(254)	null	Seletor HTML do portal, inspecionado pelo usuário, referente ao conteúdo da notícia, valor textual
seletortitulo	varchar(254)	null	Seletor HTML do portal, inspecionado pelo usuário, referente ao título da notícia, valor textual
seletorcaminhonoticia	varchar(254)	null	Seletor HTML do portal, inspecionado pelo usuário, referente ao caminho para as notícias na página inicial do portal, valor textual
ultimaatualizacao	date	null	Data da última atualização do portal, atualizada automaticamente no banco quando feito o scraping das notícias novas, respeitando o agendamento
datacriacao	date	null	Data de criação do portal, inserida no banco automaticamente no momento do cadastro
agendador_id	int	null	Chave estrangeira que referencia o agendador, valor inteiro
ativo	boolean	null	Indicador booleano (verdadeiro/falso) que indica se o portal está ativo (0 para desativo e 1 para ativo)
hasscrapedtoday	boolean	null	Indicador booleano que define se foi realizado o scraping no portal no dia atual

2.8. Table sinonimo

Description:

Armazena sinônimos de tags

2.8.1. Columns

Column name	Type	Properties	Description
id	bigint	PK	Identificador único da tabela, chave primária, auto-incremento
nome	varchar(46)	null	Nome do sinônimo, valor textual
tag_id	int		Chave estrangeira que referencia a tag à qual o sinônimo está associado, valor inteiro (não nulo)

2.8.2. Alternate keys

Key name	Columns	Description
AK_2	nome, tag_id	

2.9. Table tag

Description:

Armazena tags associadas a APIs, notícias e portais

2.9.1. Columns

Column name	Type	Properties	Description
id	int	PK	Identificador único da tabela, chave primária, auto-incremento
nome	varchar(46)	null	Nome da tag, valor textual (máximo de 46 caracteres), deve ser único

2.9.2. Alternate keys

Key name	Columns	Description
AK_3	nome	

2.10. Table api_tag

Description:

Relaciona APIs com Tags

2.10.1. Columns

Column name	Type	Properties	Description
api_id	int	PK	Chave estrangeira que referencia a API
tag_id	int	PK	Chave estrangeira que referencia a Tag

2.11. Table api_dados_tag

Description:

Relaciona os dados de API com as tags

2.11.1. Columns

Column name	Type	Properties	Description
api_dados_id	int	PK	Chave estrangeira que referencia os dados da API
tag_id	int	PK	Chave estrangeira que referencia a Tag

2.12. Table portal_tag

Description:

Relaciona Portais com Tags

2.12.1. Columns

Column name	Type	Properties	Description
portal_id	int	PK	Chave estrangeira que referencia o Portal
tag_id	int	PK	Chave estrangeira que referencia a Tag

2.13. Table noticia_tag

Description:

Relaciona Notícias com Tags

2.13.1. Columns

Column name	Type	Properties	Description
noticia_id	int	PK	Chave estrangeira que referencia a Notícia
tag_id	int	PK	Chave estrangeira que referencia a Tag

3. References

3.1. Reference FK_0

agendador	0..*	api
id	<->	agendador_id

3.2. Reference FK_2

api	0..*	apidados
id	<->	api_id

3.3. Reference FK_4

jornalista	0..*	noticia
id	<->	jornalista_id

3.4. Reference FK_5

agendador	0..*	portal
id	<->	agendador_id

3.5. Reference FK_7

api	0..*	api_tag
id	<->	api_id

3.6. Reference FK_8

tag	0..*	api_tag
id	<->	tag_id

3.7. Reference FK_9

apidados	0..*	api_dados_tag
id	<->	api_dados_id

3.8. Reference FK_10

tag	0..*	api_dados_tag
id	<->	tag_id

3.9. Reference FK_11

portal	0..*	portal_tag
id	<->	portal_id

3.10. Reference FK_12

tag	0..*	portal_tag
id	<->	tag_id

3.11. Reference FK_13

noticia	0..*	noticia_tag
id	<->	noticia_id

3.12. Reference FK_14

tag	0..*	noticia_tag
id	<->	tag_id

3.13. Reference FK_16

formato	0..*	api
id	<->	formato_id

3.14. Reference FK_18

portal	0..*	noticia
id	<->	portal_id

3.15. Reference FK_21

tag	0..*	sinonimo
id	<->	tag_id