

Граф сайта Росстата

Лицензия: CC BY

Доступ: Доступен всем пользователям

Дата обновления: 27.11.2020

Теги:

Росстат, граф

Описание

Набор данных содержит сведения о веб-страницах в структуре сайта Росстата (rosstat.gov.ru) и их содержимое на 25 сентября 2020 года в виде графа. В наборе хранятся данные о 104711 наблюдениях по 14 атрибутам.

Граф сайта Росстата обеспечивает охват всех веб-страниц, на которые можно попасть, навигируясь по сайту. Он освобождает пользователя от необходимости знать порядок переходов по ссылкам, чтобы попасть на тот или иной узел. Организация графа в виде реляционной базы позволяет использовать средства и методы баз данных для доступа к информации.

Характеристики датасета

Использование

Лицензия: [CC BY](#)

Доступ: Доступен всем пользователям

Предоставление данных

Формат:

POSTGRESQL

Характеристики набора

Временной период: 25.09.2020

Количество наблюдений: 104711

Обновления

Размещено: 27.11.2020

Обновлено: 27.11.2020

Частота обновления: Ежегодно

Дата следующего обновления: не указано

Происхождение

Источники данных: Федеральная служба государственной статистики

Ответственные за набор данных:

Структура набора данных

граф сайта (graph)

Атрибут	Описание	Количество пропусков (NaN) (шт/%)	Единица измерения	ОКЕИ	Формат
hash	хэш-функция содержимого узла	0			varchar
href	значение атрибута <href> гиперссылки, по которой осуществляется переход на узел	0			text
timestamp	время сбора узла графа	0			timestamp
id	уникальный идентификатор узла в графе сайта rosstat.gov.ru	0			bigint
rootname	наименование корневого узла	0			varchar
level	уровень графа, т.е. удаленность веб-страницы от корневого узла, выраженная в количестве переходов по гиперссылкам	0			integer
name	заголовок веб-страницы	0			text
path	абсолютный URL-адрес веб-страницы	0			text
redirect	URL-адрес, с которого произошло перенаправление	0			text
parent	URL-адрес родительского узла	0			text
document	HTML-код веб-страницы в узле	0			text
file	содержимое документа в узле в двоичном коде	0			bytea
type	формат документа в узле в виде MIME-типа	0			varchar
done	техническое поле с отметкой о пересборе узла	0			bool

Ссылка на открытый репозиторий GitHub

https://github.com/CAG-ru/rosstat_graph_crawler

Внесенные в набор данных изменения по сравнению с источниками

Данные опубликованы в том виде, в котором они представлены в первоисточнике.
Изменения не вносились.