**Предметная область**: статьи и категории Википедии.

**Собранные данные**:

- информация о категориях Википедии (по 500 категорий на каждую букву английского алфавита, всего 13000 категорий). Информация включает в себя название категории, количество страниц в данной категории и количество файлов в данной категории;

- информация о страницах, содержащихся в этих категориях (до 500 страниц из каждой категории, всего 370366 страниц). Информация включает в себя название страницы и идентификатор категории, в которой она находится.

**API Википедии**:

API действий MediaWiki (*MediaWiki action API*) — это веб-служба, предоставляющая удобный доступ к функциям, данным и метаданным вики по HTTP, с помощью URL-адресов вида api.php.

**Использованные API:**

|  |  |
| --- | --- |
| **API** | **Назначение** |
| action=query | Действие «query» используется для получения некоторой информации. Действие «query» является одним из самых важных в API. |
| format=json | Указывает, в каком формате не обходимо получить данные. |
| list=allcategories | Указывает, что необходимо получить список категорий. |
| aclimit=max | Указывает количество категорий, полученных в результате запроса. max – максимально возможное количество для пользователя с конкретными правами (в данном случае 500). |
| acprefix=some prefix | Указывает, с какой подстроки должны начинаться названия категорий. |
| list=categorymembers | Указывает, что необходимо получить список страниц категории. |
| cmlimit=max | Указывает количество страниц категории, полученных в результате запроса. max – максимально возможное количество для пользователя с конкретными правами (в данном случае 500). |
| cmtitle=Category:some category name | Указывает имя категории, для которой будет осуществляться поиск страниц. |
| acprop=size | Указывает, какую информацию о категории необходимо получить. |