

# Викиданные: обзор системы



**WIKIDATA**

# Викиданные — это...

**Викиданные (Wikidata)** — это свободная, совместно редактируемая база знаний, созданная Фондом Викимедиа, в которой хранятся структурированные данные о мире.

... и сам проект по созданию такой базы знаний, с собственным сообществом, правилами и процедурами.

<https://www.wikidata.org>

# Принципы Викиданных:

- свободные данные (максимально приближены к общественному достоянию, лицензия CC0)
- открытое редактирование и контроль сообщества
- многоязычность (не как Википедия, нет привязки к языку)
- множественность (нет единственно верного знания, возможны разные точки зрения, противоречия, неопределенности)
- ссылки на источники (храним не только факты)
- структурированные данные (доступные не только для людей, но и для программ)
- легкий доступ (поддержка XML, JSON, RDF)

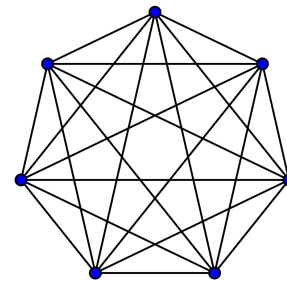
# Задачи проекта и немного истории

Дата открытия проекта: 30 октября (1 ноября) 2012

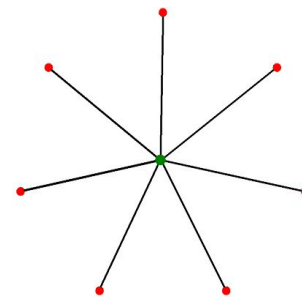
## С какой целью создавался проект?

Представьте 100 разных сайтов и появляется 101. Нам нужно поставить ссылку на 101 у 100 сайтов (или наоборот). Можно написать программу, но это не исключает ошибки, аномалии и т.д. Вот так раньше работали с интервики (межъязыковыми ссылками) в Википедии.

Появилась необходимость централизованно хранить список всех языковых версий статьи.



Было



Стало

Википедия — это ценнейший источник различной статистической и фактологической информации, но есть проблемы:

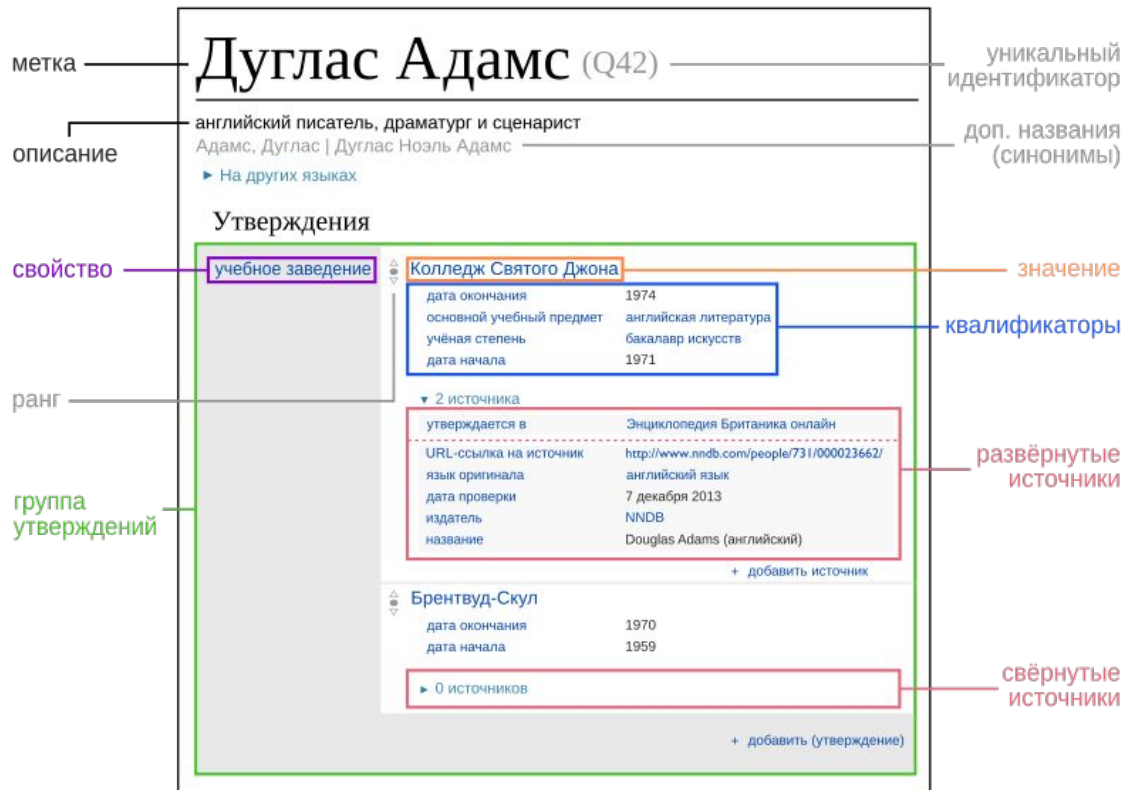
- как достать эти данные? (или их нужно брать откуда-то?)
- какое значение использовать в качестве численности населения Рима, если в английской версии статьи о Риме одно значение, а в итальянской другое, а в статье “города Италии” третье (и еще здесь может различаться в языковых версиях )?

Для решения всех этих насущных проблем и создали Викиданные.

Система Викиданные — это...

**Wikidata = MediaWiki + Wikibase**

# Модель данных



Полезные ссылки: [Введение](#), [Экскурсии по Викиданным](#), [Словарь](#)

# Доступ к данным

Мы можем скачать еженедельные копии базы данных (дампы) или осуществлять поэлементный доступ следующими способами:

- интерфейс связанных данных (с пом. URI:  
<http://www.wikidata.org/entity/Q42>,  
<http://www.wikidata.org/wiki/Special:EntityData/Q42.json>)\*
- MediaWiki API
- Боты (доступ к API может осуществляться ботом)
- Точка доступа SPARQL



# Сервис запросов Викиданных — это...

это точка доступа SPARQL — служба, позволяющая делать запросы к БЗ.

Есть два способа использования службы: через пользовательский интерфейс или непосредственно через HTTP-запросы GET и POST к <https://query.wikidata.org/sparql>.

Например (2 способ):

<https://query.wikidata.org/sparql?query=SELECT%20?item%20WHERE%20{?item%20wdt:P31%20wd:Q146.}>

# немного RDF (Resource Description Framework)

RDF представляет собой формальный язык для описания данных.

Схема утверждения: «субъект-предикат-объект»

Подставим вместо этих слов URI и получим утверждение на языке RDF.

В терминах Викиданных схема выглядит так:

«элемент-свойство-значение»

## еще немного RDF

Пример: утверждение “Стаббс занимал должность мэра” будет на RDF выглядеть так:

<https://www.wikidata.org/wiki/Q7627362>

<https://www.wikidata.org/wiki/Property:P39>

<https://www.wikidata.org/wiki/Q30185>.

# SPARQL

— это язык запросов к данным, представленным по модели RDF, а также протокол для передачи этих запросов и ответов на них.

Общая схема SPARQL-запросов имеет вид:

```
PREFIX ... #сокращения для URI
```

```
SELECT ... #переменные, значения которые надо вернуть
```

```
WHERE {
```

```
    ... #ограничения
```

```
}
```

# Создание запроса на языке SPARQL

Шаблон тройки; это тройка, в которой на месте субъекта, предиката или объекта могут стоять переменные, состоящие из «?» и имени переменной. Если шаблоны троек объединяются в фигурные скобки, то во все переменные с одним именем подставляется одно значение.

Короткая запись для запросов в Викиданных :

[wd:Q7627362](#) [wdt:P39](#) [wd:Q30185](#).

# Примеры запросов

Запрос 1: про кота

Запрос 2: правдивое утверждение

Запрос 3: полное утверждение

Запрос 4: полное утверждение и квалификаторы

Запрос 5: правильный запрос 4

Запрос 6: использование службы Label Service

# Запросы с различными синтаксическими конструкциями

Запрос 1: “картины Клода Моне с изображениями”

Запрос 2:

?элемент представитель Картина.

?элемент создатель Клод Моне.

?элемент располагается в ?локация.

?локация страна Россия.

Запрос 3: найти все произведения искусства

# Обзор основных операторов и функций

Запрос 1: карта картин Ван Гога с подсчетом количества находящихся картин в данной локации

Запрос 2: использование оператора FILTER и приведение к типу Дата-Время

Запрос 3: использование BIND



# Возможные области применения:

- машиночитаемая база знаний о мире
- обучение ИИ
- возможность создания поисковых средств с высоким качеством ответа без использования ИИ, а лишь на основании данных из этой БЗ.
- аналитика: полезные энциклопедические данные, которые вы не найдете нигде, кроме Linked Data (Например, города, страны, исторические данные, институты, организации, госструктуры, фильмы, композиторы, музыканты, художники, политики, спортсмены, исторические фигуры);
- улучшение ассоциативного понимания человеческой речи, извлечение фактов, анализ текстов

# Примеры проектов, работ, исследований

- курс на Викиверситете “[Программирование Викиданных](#)” — исследование различных предметных областей и исследование полноты темы, много учебных проектов, задач, примеров запросов.
- мобильное приложение на Андроид по изучению языков [Artolela](#).
- поиск имплицативных знаний в Викиданных с помощью анализа формальных понятий, [статья](#).
- исследование культурного контекста в Википедии, классификация всех статей Википедии по культурно-контекстным признакам, [статья](#).