

ОСНОВЫ SPARQL

Интеграция данных

SPARQL

SPARQL – SPARQL Protocol And RDF Query Language

Описывается спецификациями комитета W3C.

Язык запросов SPARQL для RDF хранилищ данных

Рекомендация W3C, 15 января 2008

Текущая версия:

<http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/>

Последняя версия:

<http://www.w3.org/TR/rdf-sparql11-query/>



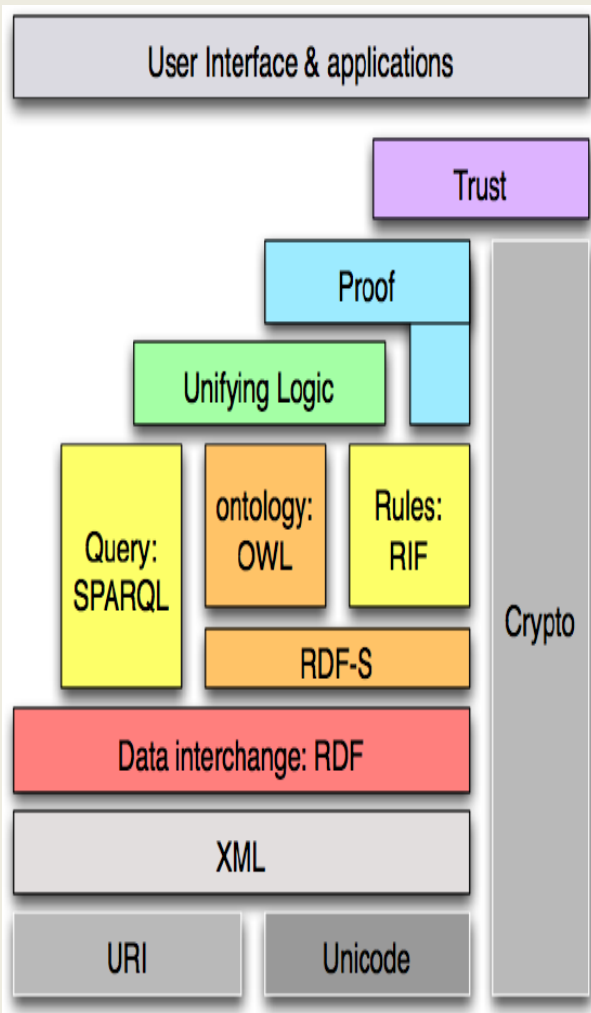
Хранение информации является важнейшей составляющей современных информационных технологий

(65 стр., © Губин А.Н., Филиппов Ф.В.
© ФГОБУВПО СПб ГУТ
им.проф. М.А. Бонч-Бруевича



Хранение информации является важнейшей составляющей современных информационных технологий

SPARQL



Wikidata — это новый проект Фонда Викимедиа. Главная задача — предоставить каждому человеку на планете свободный доступ ко всем возможным знаниям.



SPARQL

Викиданные» ([англ. Wikidata](#)) — [совместно редактируемая база знаний](#), созданная [Фондом Викимедиа](#). Она используется для обеспечения централизованного хранения [данных](#), которые должны содержаться в статьях [Википедии](#) — Содержимое Викиданных распространяется по лицензии [Creative Commons CC0](#).^[4]

Проект официально запущен 30 октября 2012 года, разработка проекта ведётся под руководством [Wikimedia Deutschland](#).




DBpedia занимается тем, что собирает данные из Википедий

В **DBpedia** никакие данные нельзя редактировать напрямую.

Хранение информации является важнейшей составляющей **современных информационных технологий**

Хранилище состоит в основном из элементов, каждый из которых имеет метку, описание и любое количество синонимов.

Элементы имеют уникальный идентификатор с префиксом Q и числовой частью, как, например, Douglas Adams (Q42).



<u>Элемент</u>	<u>Свойство</u>	<u>Значение</u>
Q42	P69	Q691283
Douglas Adams	educated at	St John's College

Утверждения детально описывают каждый элемент. Каждое утверждение состоит из свойства и его значения.
Свойства в Викиданных идентифицируются кодом, имеющим префикс P и числовую часть; например educated at (P69).

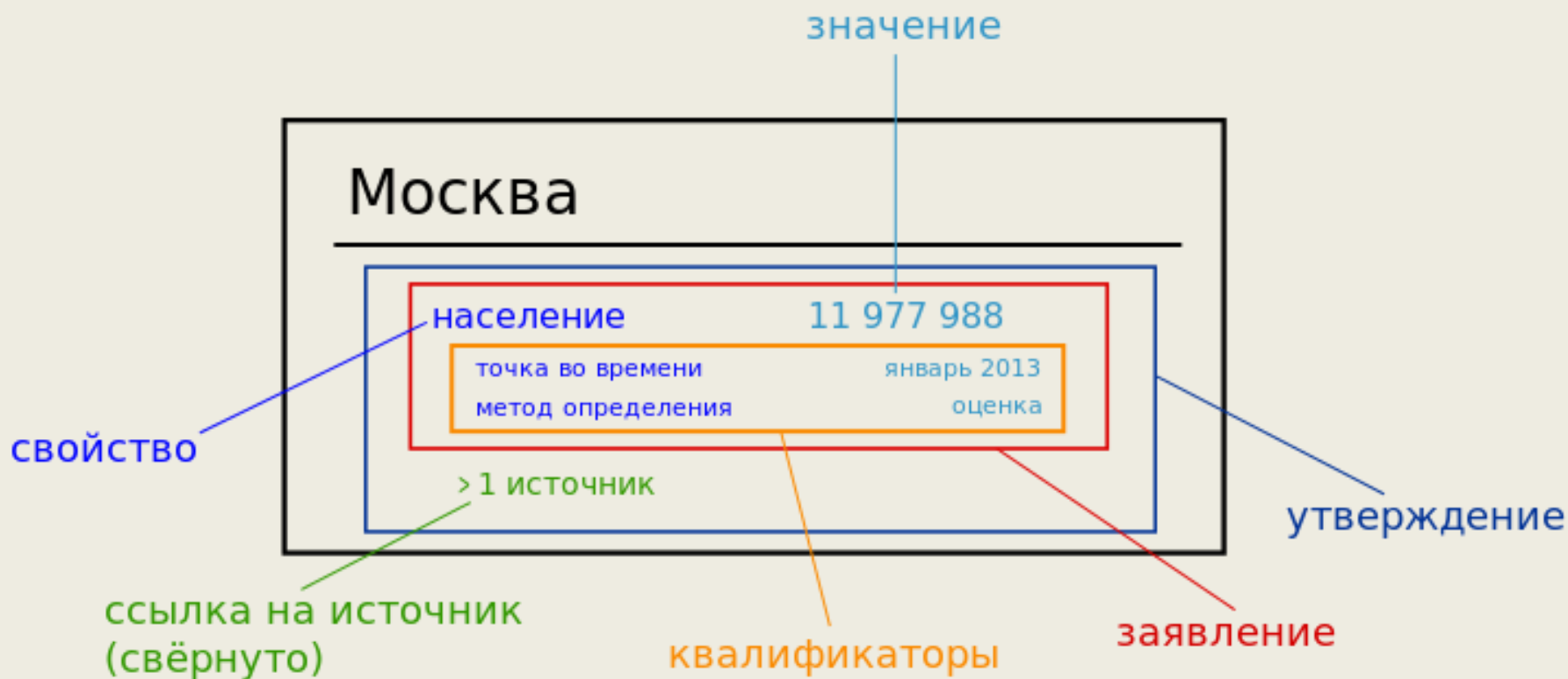
<u>Элемент</u>	<u>Свойство</u>	<u>Значение</u>
Q42	P69	Q691283
Douglas Adams	educated at	St John's College

Свойства могут содержать **ссылки** на внешние базы данных.

Свойство, которое ссылается на элемент во внешней базе данных, вроде баз данных авторитетного контроля, используемых библиотеками и архивами, называется **идентификатор**

<u>Элемент</u>	<u>Свойство</u>	<u>Значение</u>
Q42	P69	Q691283
Douglas Adams	educated at	St John's College

Терминология утверждений в Викиданных



Элемент Викиданных

(самые важные термины Викиданных.)

SPARQL

The image shows a Wikidata entity page for "Дуглас Адамс (Q42)". The page is annotated with labels on the left and right sides, pointing to specific elements. The main content area is a light gray box with a white background. The title "Дуглас Адамс (Q42)" is at the top, followed by a description: "английский писатель, драматург и сценарист". Below this is the name "Адамс, Дуглас | Дуглас Ноэль Адамс" and a link "► На других языках". The "Утверждения" (Statements) section is highlighted with a green border. It contains two statements: "учебное заведение" (school) and "Брентвуд-Скул" (Brentwood School). The "учебное заведение" statement is further detailed with a table of properties: "дата окончания" (1974), "основной учебный предмет" (английская литература), "учёная степень" (бакалавр искусств), and "дата начала" (1971). Below this table is a section for "2 источника" (2 sources), which is expanded to show details for the "Энциклопедия Британика онлайн" source, including the URL, language, date of check, and publisher. The "Брентвуд-Скул" statement has a table with "дата окончания" (1970) and "дата начала" (1959). At the bottom, there is a section for "0 источников" (0 sources).

метка — Дуглас Адамс (Q42) — уникальный идентификатор

описание — английский писатель, драматург и сценарист
Адамс, Дуглас | Дуглас Ноэль Адамс — доп. названия (синонимы)
► На других языках

свойство — учебное заведение — значение

ранг — Колледж Святого Джона — квалификаторы

группа утверждений —

дата окончания	1974
основной учебный предмет	английская литература
учёная степень	бакалавр искусств
дата начала	1971

▼ 2 источника

утверждается в	Энциклопедия Британика онлайн
URL-ссылка на источник	http://www.nndb.com/people/731/000023662/
язык оригинала	английский язык
дата проверки	7 декабря 2013
издатель	NNDB
название	Douglas Adams (английский)

+ добавить источник

Брентвуд-Скул	
дата окончания	1970
дата начала	1959

► 0 источников

+ добавить (утверждение)

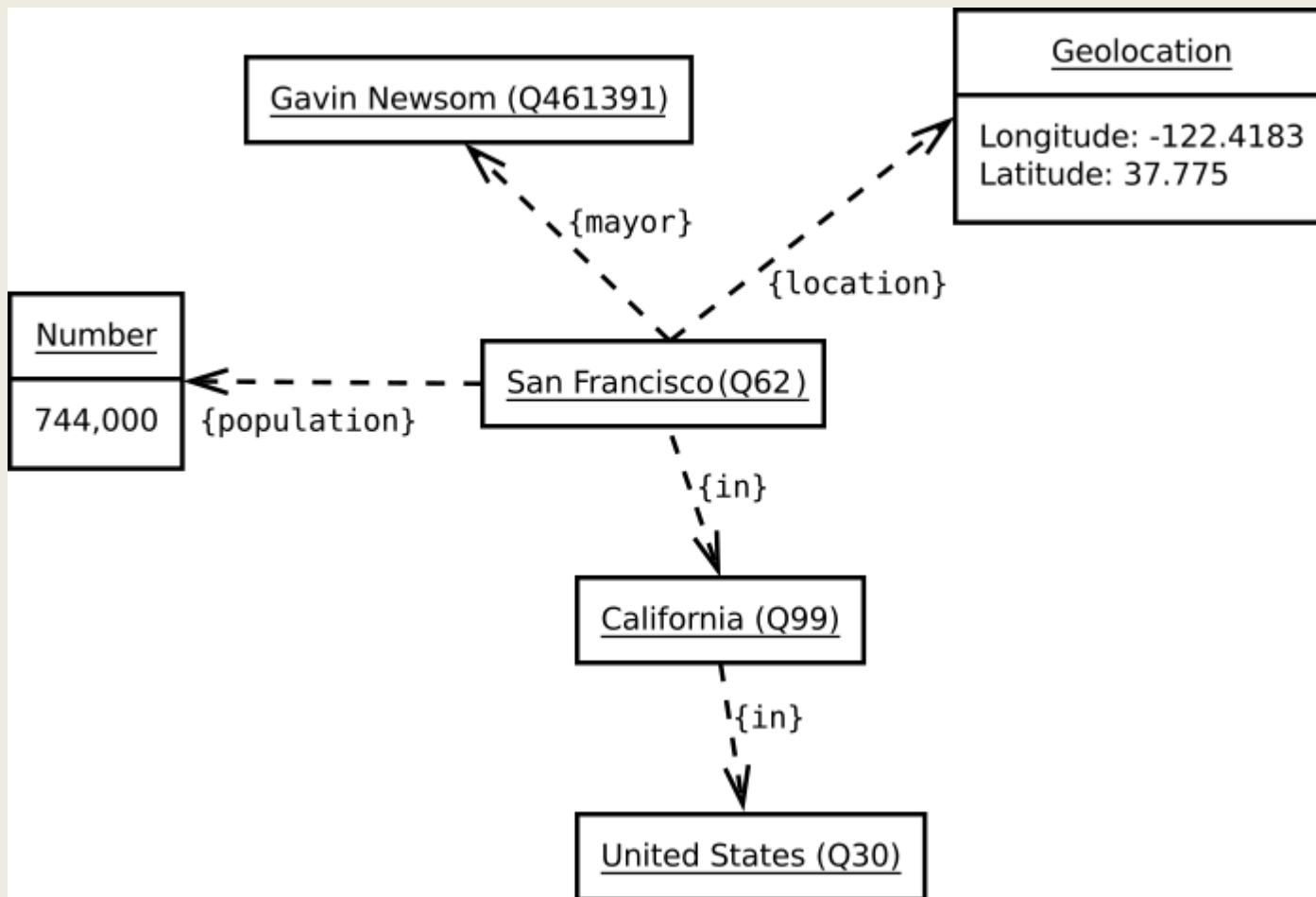
развёрнутые источники

свёрнутые источники

Пример.

Связанные между собой
элементы и их данные.

SPARQL



Элементом Викиданных (*item*) называется сложная структура данных, описывающая некое понятие.

Идентифицируется элемент строкой вида **Q<число>**, который называют **идентификатором**.

Идентификатор присваивается элементу в момент его создания и с тех пор никогда не меняется.



Санкт-Петербург



Город



[Q656](#)

Санкт-Петербург



Гостиница



[Q4407648](#).

Элементы Викиданных



S	Desc
wd:Q114375	мемуары
wd:Q124761	Эпитома
wd:Q131084	либретто
wd:Q179461	священный текст
wd:Q181001	эротика
wd:Q182659	баллада
wd:Q189867	Булла

Свойство описывает значение данных заявления.

Свойства, в сочетании со значениями, формируют [оператор](#) в Википедии.

Свойства также используются в [квалификаторах](#) .

Свойства имеют свои собственные страницы в Викидате и связаны с элементами, в результате чего создается связанная структура данных.



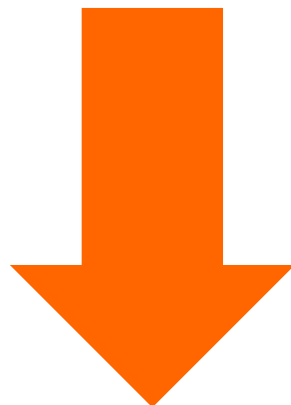
Свойства похожи на элементы.

Подобно элементам, каждое свойство имеет метку, описание и даже псевдоним или псевдонимы, которые могут быть добавлены на нескольких языках.

Подобно элементам, **свойства** также имеют уникальный идентификатор.

Элементы начинаются с префикса Q и находятся в основном пространстве имен Wikidata.

Свойства имеют префикс P и сохраняются в пространстве имен свойств.



Title	ID	Data type	Description	Examples	Inverse
publisher	P123	Item	организации или лицо, ответственное за публикацию книг, периодических изданий, игр или программного обеспечения	The Oxford English Dictionary <publisher> Oxford University Press	-
programming language	P277	Item	язык (языки) программирования, на которых разрабатывается программное обеспечение	Stockfish <programming language> C++	-
operating system	P306	Item	операционная система (ОС), на которой работает программное обеспечение или ОС, установленная на оборудовании	Mozilla Firefox <operating system> Microsoft Windows	-
software version identifier	P348	String	числовые или символьные идентификаторы версий программного обеспечения или формата файлов, текущих и прошлых	Bugzilla <software version identifier> 4.5.1	-

В Викиданных есть пара универсальных свойств, обязательных к использованию во всех элементах без исключения.

Каждый элемент должен иметь одно из свойств:

-либо [это частный случай понятия \(P31\)](#),

-либо [подкласс от \(P279\)](#)^[11].

Оба свойства имеют тип данных «элемент».

Конкретные (не являющиеся группой или какой-то разновидностью объектов) понятия должны иметь значением свойства [это частный случай понятия \(P31\)](#) некоторое общее понятие, например:



Элемент	Значение свойства это частный случай понятия (P31)
Земля (Q2)	внутренняя планета (Q3504248)
Роберт Шуман (Q7351)	человек (Q5)
Джомолунгма (Q513)	гора (Q8502)
Ich weiß nicht, was soll es bedeuten... (Q14540816)	стихотворение (Q5185279)

терминологии и модели Wikidata
([Викиданные:Глоссарий](#))




Заявление (*claim*) — это отдельный фрагмент данных об [элементе](#), на [странице](#) которого содержится это заявление.

Заявление состоит из [свойства](#) (например, «Местоположение») и [значения](#) (скажем, «Германия»), а также может быть— «нет значения» и «неизвестное значение».

Заявление может содержать [квалификаторы](#), например, временные, говорящие, что заявление действительно в течение некоторого промежутка времени.

Заявление образует часть [утверждения](#) на странице [элемента](#), где оно может быть дополнено [ссылками на источники](#) и [рангами](#). Заявления могут также появляться на страницах данных, не являющихся элементами.



SPARQL

The image shows a Wikidata profile for Douglas Adams (Q42) with several annotations explaining SPARQL concepts:

- метка** (label): Points to the name "Дуглас Адамс (Q42)".
- описание** (description): Points to the text "английский писатель, драматург и сценарист".
- свойство** (property): Points to the "учебное заведение" (school) property.
- ранг** (rank): Points to the "Утверждения" (statements) section.
- группа утверждений** (statement group): Points to the "Колледж Святого Джона" statement.
- Заявление** (statement): Points to the "Колледж Святого Джона" statement.
- уникальный идентификатор** (unique identifier): Points to "(Q42)".
- доп. названия (синонимы)** (alternative names/synonyms): Points to "Адамс, Дуглас | Дуглас Ноэль Адамс".
- значение** (value): Points to "Колледж Святого Джона".
- квалификаторы** (qualifiers): Points to the table of qualifiers for the school statement.
- развёрнутые источники** (expanded sources): Points to the source information for the school statement.
- свёрнутые источники** (collapsed sources): Points to the source information for the "Брентвуд-Скул" statement.

Утверждения

учебное заведение	Колледж Святого Джона
дата окончания	1974
основной учебный предмет	английская литература
учёная степень	бакалавр искусства
дата начала	1971

2 источника

утверждается в	Энциклопедия Британника онлайн
URL-ссылка на источник	http://www.oxfordjournals.org/abstract/doi/10.1093/acprof:osao/0000000000000000/
язык оригинала	английский язык
дата проверки	7 декабря 2013
издатель	MNDB
название	Douglas Adams (английский)

+ добавить источник

Брентвуд-Скул	
дата окончания	1970
дата начала	1959

0 источников

+ добавить (утверждение)

SPARQL

Для многих заявлений требуются дополнительные **квалификаторы** для адекватного и точного формирования данных в Wikidata.



Квалификаторы используются для указания консенсусного значения



Для нескольких значений

Для одиночных значений

Для спорных предметов и согласия сообщества

Альмаматер

Время действия данных



Другие точки зрения могут быть добавлены в качестве дополнительных значений с использованием квалификаторов

SPARQL

Для многих заявлений требуются дополнительные **квалификаторы** для адекватного и точного формирования данных в Wikidata.



Квалификатор (*qualifier*) — это часть заявления, которая сообщает нечто об этом конкретном заявлении, часто в описательной форме. Квалификатор может быть термином в рамках определённого словаря, но также может быть и произвольной описательной фразой. Решение о том, представляют ли собой эти термины или фразы свободный текст или являются частью некоторого словаря, вероятно, будет приниматься сообществом Викиданных.

SPARQL

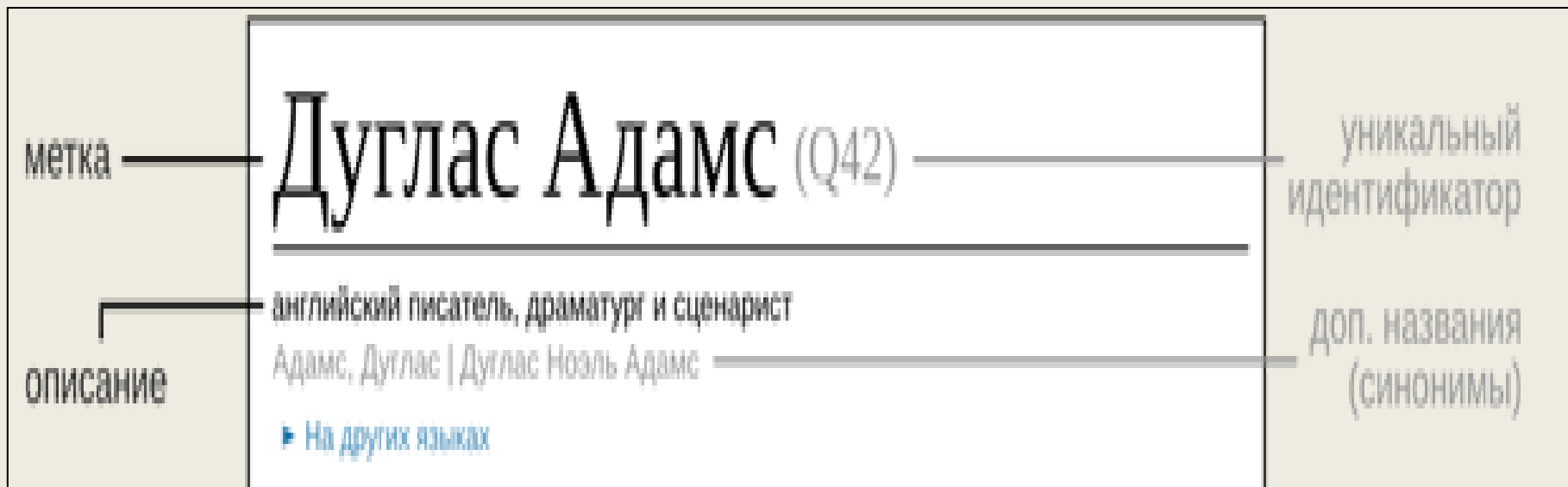
Синонимы или **псевдонимы** (*aliases*; в пользовательском интерфейсе они обозначаются фразой «*Also known as*») — это альтернативные названия элемента или свойства.

Наиболее распространённый или самый важный из вариантов — это **метка**. Синонимы помогают людям найти элемент, даже если они ищут не то значение, которое сохранено как метка.

метка	Дуглас Адамс (Q42)	уникальный идентификатор
описание	английский писатель, драматург и сценарист Адамс, Дуглас Дуглас Нозль Адамс ▶ На других языках	доп. названия (синонимы)

SPARQL

Метка или **обозначение** (от англ. *label*; также может обозначаться как **название** — *name*) — это самое важное из названий, под которыми известна описываемая сущность ([элементов](#) или [свойство](#)). Каждая сущность может иметь только одну метку на каждом из языков. В рамках проекта метки не обязаны быть уникальными. Для того, чтобы отличать сущности с одинаковыми метками, используются [описания](#) и [синонимы](#) (псевдонимы). При этом у [страниц](#) Викиданных не метки, а [заголовки](#) (*title*), не следует это путать.



SPARQL

Используемые префиксы.



Prefix	Full URL	Usage	Example
wikibase:	http://wikiba.se/ontology-beta#[1]	Wikibase ontology	wd:Q2 wikibase:Item
	Nodes		
wdata:	/Special:EntityData/	Набор данных, описывающий определенную сущность	wdata:Q2 schema:about wd:Q2 .
wd:	/entity/	Wikibase entity - item or property.	wd:Q2 p:P9 wds:Q2-82a6e009-4f93-28dc-3555-38bbfc3afe6a
wds:	/entity/statement/	Statement node, describes claim about entity.	wds:Q2-a4078553-4ec1-a64a-79e7-c5b5e17b2782 wikibase:Statement
wdv:	/value/	Value node	wdv:87d0dc1c7847f19ac0f19be978015dfb202cf59a wikibase:Value

SPARQL

Используемые префиксы.



Prefix	Full URL	Usage	Example
wdref:	/reference/	Reference node	<p>wds:Q3-24bf3704-4c5d-083a-9b59-1881f82b6b37</p> <p>prov:wasDerivedFrom</p> <p>wdref:87d0dc1c7847f19ac0f19be978015dfb202cf59a</p> <p>.wdref:87d0dc1c7847f19ac0f19be978015dfb202cf59a a</p> <p>wikibase:Reference .</p>
	Predicates		
wdt:	/prop/direct/	Truthy assertions about the data, links entity to value directly.	<p>wd:Q2 wdt:P9</p> <p><http://acme.com/></p>

SPARQL

Используемые префиксы.



Prefix	Full URL	Usage	Example
wdtn:	/prop/direct-normalized/	Truthy assertions about the data, links entity to normalized value directly.	<code>wd:Q2 wdtn:P9 <http://acme.com/ABCDE></code>
p:	/prop/	Links entity to statement	<code>wd:Q2 p:P9 wds:Q2-82a6e009-4f93-28dc-3555-38bbfc3afe6awd</code>
wдно:	/prop/novalue/	Class to use when the entity has <i>novalue</i> for this property.	<code>wd:Q2 a wдно:P9 .</code>
ps:	/prop/statement/	Links value to statement	<code>wds:Q3-24bf3704-4c5d-083a-9b59-1881f82b6b37 ps:P8 "13000000000-01-01T00:00:00Z"^^xsd:dateTime</code>

SPARQL

Используемые префиксы.



Prefix	Full URL	Usage	Example
psv:	/prop/statement/value/	Links deep value to statement	wds:Q3-24bf3704-4c5d-083a-9b59-1881f82b6b37 psv:P8 wdv:87d0dc1c7847f19ac0f19be978015dfb202cf59a
psn:	/prop/statement/value-normalized/	Links normalized value to statement node	wds:Q3-24bf3704-4c5d-083a-9b59-1881f82b6b37 psn:P8 wdv:87d0dc1c7847f19ac0f19be978015dfb202cf59a
pq:	/prop/qualifier/	Links qualifier to statement node	wds:Q3-24bf3704-4c5d-083a-9b59-1881f82b6b37 pq:P8 "-13000000000-01-01T00:00:00Z"^^xsd:dateTime

SPARQL

Используемые префиксы.



Prefix	Full URL	Usage	Example
pqv:	/prop/qualifier/value/	Links qualifier deep value to statement node	wds:Q3-24bf3704-4c5d-083a-9b59-1881f82b6b37 pqv:P8 wdv:87d0dc1c7847f19ac0f19be978015dfb202cf59a
pqn:	/prop/qualifier/value-normalized/	Links normalized qualifier value to statement node	wds:Q3-24bf3704-4c5d-083a-9b59-1881f82b6b37 pqn:P8 wdv:87d0dc1c7847f19ac0f19be978015dfb202cf59a
pr:	/prop/reference/	Links reference to value	wdref:87d0dc1c7847f19ac0f19be978015dfb202cf59a pr:P8 "- 13000000000-01- 01T00:00:00"^^xsd:dateTime

SPARQL

Используемые префиксы.

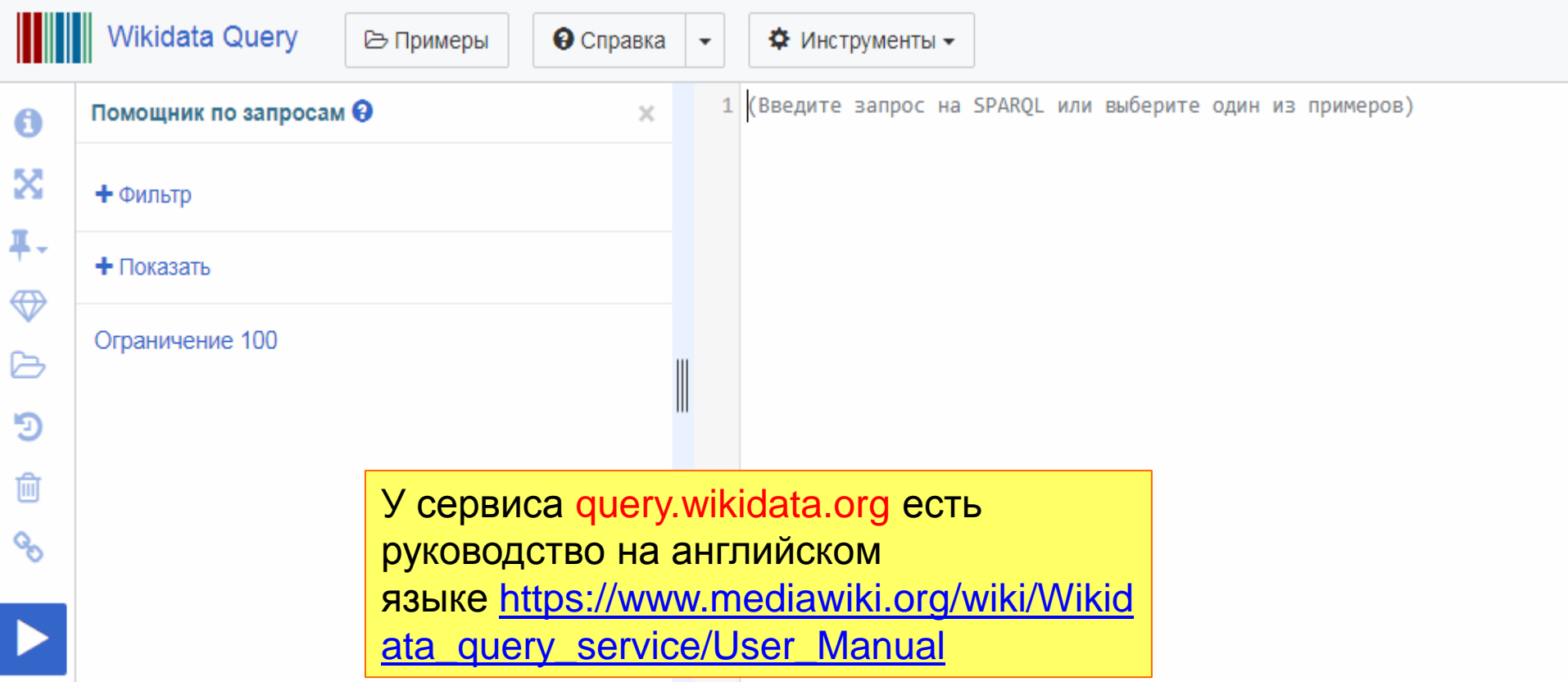


Prefix	Full URL	Usage	Example
prv:	/prop/reference/value/ /	Links reference to deep value	wdref:87d0dc1c7847f1 9ac0f19be978015dfb2 02cf59a prv:P8 wdv:87d0dc1c7847f19 ac0f19be978015dfb20 2cf59a
prn:	/prop/reference/value -normalized/	Links reference to normalized value	wdref:87d0dc1c7847f1 9ac0f19be978015dfb2 02cf59a prn:P8 wdv:87d0dc1c7847f19 ac0f19be978015dfb20 2cf59a

<https://query.wikidata.org>

Один из способов обращаться к Викиданным - с помощью программного интерфейса.

[Wikidata Query](https://query.wikidata.org)



The screenshot shows the Wikidata Query interface. At the top, there is a navigation bar with the Wikidata logo, the text "Wikidata Query", and three buttons: "Примеры" (Examples), "Справка" (Help), and "Инструменты" (Tools). Below the navigation bar is a sidebar on the left with various icons and a "Помощник по запросам" (Query Assistant) section. The main area is a text input field for a SPARQL query, with a placeholder text "(Введите запрос на SPARQL или выберите один из примеров)".

У сервиса query.wikidata.org есть руководство на английском языке https://www.mediawiki.org/wiki/Wikidata_query_service/User_Manual

Виды литературных произведений

```
SELECT ?S ?Desc
WHERE
{
  ?S wdt:P279 wd:Q7725634 .
  OPTIONAL
  {
    ?S rdfs:label ?Desc
    filter(lang(?Desc)="ru") .
  }
}
Limit 10
```

подкласс от (P279)
каждый частный случай данного элемента также частный случай следующего(-их) элемента(-ов)

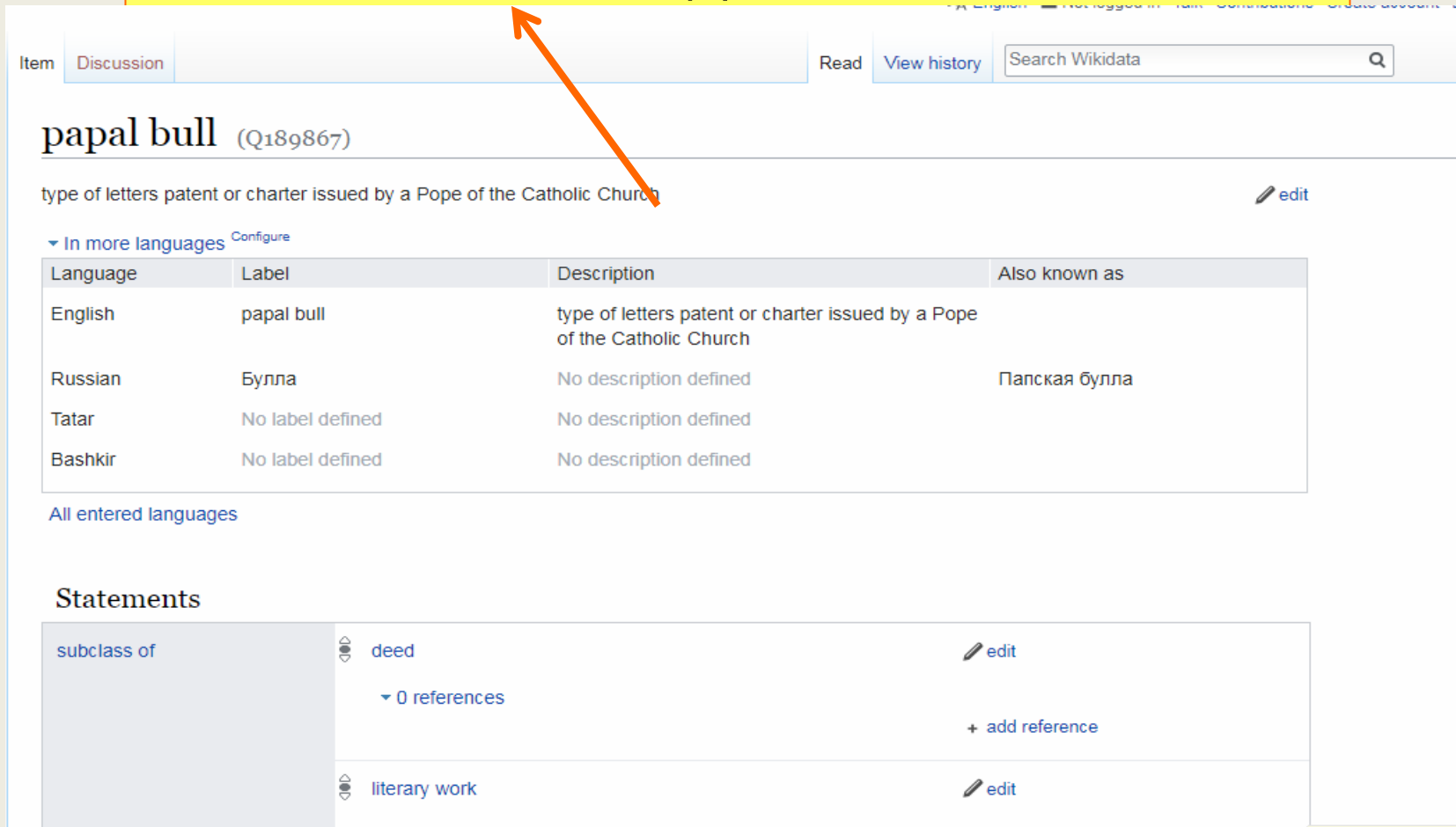
литературное произведение (Q7725634)
работа, состоящая в основном из текста

Виды литературных произведений

S	Desc
wd:Q114375	мемуары
wd:Q124761	Эпитама
wd:Q131084	либретто
wd:Q179461	священный текст
wd:Q181001	эротика
wd:Q182659	баллада
wd:Q189867	Булла
wd:Q193195	фельетон
wd:Q268416	беллетристика
wd:Q472808	короткая повесть



Тип письма патент или устав, выданный Папой Римским Католической Церкви



Item Discussion Read View history Search Wikidata

papal bull (Q189867)

type of letters patent or charter issued by a Pope of the Catholic Church [edit](#)

[In more languages](#) Configure

Language	Label	Description	Also known as
English	papal bull	type of letters patent or charter issued by a Pope of the Catholic Church	
Russian	Булла	No description defined	Папская булла
Tatar	No label defined	No description defined	
Bashkir	No label defined	No description defined	

[All entered languages](#)

Statements

subclass of	deed edit
	0 references + add reference
	literary work edit

SPARQL запросы в WIKIDATA

Пример, показывающий список президентов США и их супругов:

```

SELECT ?p ?pLabel ?w
?wLabel
WHERE
{
  wd:Q30 p:P6/ps:P6 ?p .
  ?p wdt:P26 ?w .
  SERVICE wikibase:label {
    bd:serviceParam
    wikibase:language "ru" .
  }
}
    
```

США (Q30)
страна в Северной Америке

глава правительства (P6)
высшее административное лицо, глава исполнительного органа города, муниципалитета, страны или другой правительственной структуры

супруг(а) (P26)
супруг или супруга персоны

Label service

SPARQL запросы в WIKIDATA

Пример, показывающий список президентов США и их супругов:

p	pLabel	w	wLabel
wd:Q9640	Линдон Джонсон	wd:Q233648	Джонсон, Леди Берд
wd:Q9588	Ричард Никсон	wd:Q158707	Никсон, Пэт
wd:Q9582	Джеральд Рудольф Форд	wd:Q213122	Форд, Бетти
wd:Q8612	Эндрю Джонсон	wd:Q234295	Маккардл-Джонсон, Элиза
wd:Q8007	Франклин Делано Рузвельт	wd:Q83396	Элеонора Рузвельт
wd:Q1124	Билл Клинтон	wd:Q6294	Хиллари Клинтон
wd:Q207	Джордж Уокер Буш	wd:Q152019	Лора Буш
wd:Q91	Авраам Линкольн	wd:Q215062	Мэри Тодд Линкольн
wd:Q76	Барак Обама	wd:Q13133	Мишель Обама

SPARQL запросы в WIKIDATA

```
SELECT ?p ?pLabel ?w ?wLabel WHERE
{
  wd:Q30 p:P6/ps:P6 ?p .
  ?p wdt:P26 ?w .
  ?p rdfs:label ?pLabel.
  ?w rdfs:label ?wLabel.
}
Limit 10
```



SPARQL запросы в WIKIDATA



wd:Q23	George Washington	wd:Q191789	Martha Washington
wd:Q23	George Washington	wd:Q191789	Martha Washington
wd:Q23	ጆርጅ ዋሽንግተን	wd:Q191789	Martha Washington
wd:Q23	George Washington	wd:Q191789	Martha Washington
wd:Q23	George Washington	wd:Q191789	Martha Washington
wd:Q23	واشنطن جورج	wd:Q191789	Martha Washington
wd:Q23	ጆርጅ ዋሽንግተን	wd:Q191789	Martha Washington
wd:Q23	واشینطون جورج	wd:Q191789	Martha Washington
wd:Q23	George Washington	wd:Q191789	Martha Washington
wd:Q23	George Washington	wd:Q191789	Martha Washington

SPARQL запросы в WIKIDATA

Пути свойств

Пути свойств - это способ очень коротко записать путь свойств между двумя элементами. Самый простой путь - это просто одно свойство, которое образует обычную тройку:

```
?Пункт WDT:P31 ?класс .
```

Можно добавлять элементы пути с помощью косой черты (/).

```
?item wdt:P31/wdt:P279 ?класс .
```

SPARQL запросы в WIKIDATA

Что эквивалентно



```
? item wdt : P31 ? temp1 .  
? temp1 wdt : P279 ? temp2 .  
? temp2 WDT : P279 ? класс .
```

Города в России и их телефонные коды

```
SELECT ?item ?relationid ?itemLabel
?countrydialingcode ?localdialingcode
```

WHERE

```
{ ?item wdt:P17 wd:Q159.
?item wdt:P31 wd:Q515.
OPTIONAL {
  ?item wdt:P402 ?relationid. }
?item
  wdt:P473 ?localdialingcode.
wd:Q159
wdt:P474 ?countrydialingcode}
```

SELECT выбирает данные для вывода

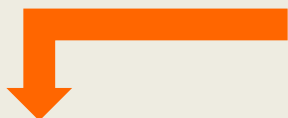
выбираем субъекты (?item), у которых явно указано, что они в стране - *Россия* (Q159)

выбираем субъекты (?item), у которых явно указано, что они суть - *города* (Q515)

значения предиката "ID отношения из OpenStreetMap" P402 могут быть просто не заполнены у отдельных объектов, такие объекты будут исключены из результатов.

выбираем все возможные субъекты и объекты у предиката P473 (код города).

SPARQL



Запрос на все книги сэра Артура Конан Дойла:

```
SELECT ?book ?bookLabel
WHERE
{
  ?book wdt:P50 wd:Q35610.
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "ru". }
}
LIMIT 100
```

автор (P50)
создатель произведения

Артур Конан Дойл (Q35610)
английский писатель





book	bookLabel
wd:Q45192	Собака Баскервилей
wd:Q223131	Этюд в багровых тонах
wd:Q228119	Последнее дело Холмса
wd:Q265546	Скандал в Богемии
wd:Q392147	Приключения Шерлока Холмса
wd:Q461026	Союз рыжих
wd:Q466166	Шерлок Холмс при смерти
wd:Q499059	Картонная коробка
wd:Q500688	Знак четырёх
wd:Q602654	Москательщик на покое

Обработка запрос, который определит заголовок (P1476) , иллюстратор (P110) , издатель (P123) и дату публикации (P577) .



```
SELECT ?book ?title ?illustratorLabel ?publisherLabel ?published
WHERE
{
  ?book wdt:P50 wd:Q35610.
  OPTIONAL { ?book wdt:P1476 ?title. }
  OPTIONAL { ?book wdt:P110 ?illustrator. }
  OPTIONAL { ?book wdt:P123 ?publisher. }
  OPTIONAL { ?book wdt:P577 ?published. }
  SERVICE wikibase:label { bd:serviceParam wikibase:language "ru". }
}
LIMIT 100
```





book	title	illustratorLabel	publisherLabel	published
Q7748730	The Lost Special			1 января 1898 г.
Q7753508	The New Annotated Sherlock Holmes		W. W. Norton & Company	17 ноября 2004 г.
Q7768531	The Terror of Blue John Gap			
Q7772936				
Q12857071				
Q15891529				1 января 1896 г.
Q16385389				
Q16387251	The Narrative of John Smith			1 января 2011 г.
Q17352362	L'Aristocratique célibataire			1 января 1924 г.
Q17352366				1 января 1892 г.