

Задания 3.1 и 3.2. Запрос на SPARQL



Тушканова Ольга Николаевна

Санкт-Петербург
2020

SPARQL

Язык обращения к данным, представленным в формате RDF, а также протокол для передачи таких запросов и ответов на них.

Является рекомендацией консорциума W3C и одной из технологий семантической паутины.

Предоставление SPARQL-точек доступа (англ. SPARQL-endpoint) является рекомендованной практикой при публикации данных во всемирной паутине.

Формы запросов

SELECT запрос

- Извлекает необработанные значения из точки доступа SPARQL и возвращает результаты в формате таблицы.

CONSTRUCT запрос

- Извлекает информацию из точки доступа SPARQL в формате RDF и преобразовывает результаты к определенной форме.

ASK запрос

- Формирует запрос типа Истина/Ложь.

DESCRIBE запрос

- Получает описание RDF-ресурса. Реализация поведения DESCRIBE-запросов определяется разработчиком точки доступа SPARQL.

Общая схема SELECT запроса

PREFIX

префиксные объявления - служат для указания сокращений универсальных идентификаторов ресурса (URI), используемых в запросе

FROM ...

источники запроса - определяют какие RDF-графы запрашиваются

SELECT ...

состав результата - определяет возвращаемые элементы данных

WHERE {...}

шаблон запроса - определяет, что запрашивать из набора данных

ORDER BY ...

модификаторы запроса - ограничивают, упорядочивают, преобразуют результаты запроса

Некоторые ключевые слова

PREFIX — служит для сокращения URI

OPTIONAL — обозначает необязательный шаблон

GRAPH — применяет шаблон к именованным графам

DISTINCT — обеспечивает уникальность решений в ответе на запрос

LIMIT — задает максимальное количество выводимых результатов

OFFSET — опускает в результате первые n решений

ORDER BY — позволяет отсортировать результат по возрастанию (ASC) или по убыванию (DESC)

MINUS — позволяет задать множество, которое необходимо «отнять»

Пример

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT ?name

 ?email

WHERE

{

 ?person **a foaf:Person** .

 ?person **foaf:name** ?name .

 ?person **foaf:mbox** ?email .

}

Точка доступа SPARQL

Служба, поддерживающая протокол запросов SPARQL.

Позволяет пользователю делать запросы к базе знаний.

Сервер обрабатывает запрос и возвращает ответ в машиночитаемом формате.

API к базам знаний

Представление результатов должно быть реализовано программным обеспечением вызывающей стороны.

Виды точек доступа:

- общего назначения - запросы по любым указанным RDF-документам, находящимся в Сети

- локальные - данные только от одного ресурса

Sparql Endpoints

Project	SPARQL endpoint	comment
Wikidata	endpoint	See also SPARQL federation input
BBC Programmes and Music	endpoint	Powered by OpenLink Virtuoso; also supports Faceted Browsing and Exploration
Bio2RDF	List of 40 SPARQL endpoints	uses OpenLink Virtuoso
BioGateway	endpoint	BioGateway provides many parameterizable SPARQL queries, both biological as ontological, on RDF graphs that were optimized for querying. The graphs have relational closures. Empowered by OpenLink Virtuoso.
BBC Backstage (HP Labs)	endpoint	uses joseki 3

Задание 3.1. Запрос на SPARQL

Номер	Запрос
1	список фильмов, которые получили Оскар в том же году
2	список актрис, которые играют в рекомендованном фильме и не имеют Оскар
3	места рождения актрис фильма на карте
4	список фильмов в том же жанре, вышедших на той же киностудии
5	самого младшего участника съемочного процесса
6	список актеров, которые играют в рекомендованном фильме, старше 40 на момент выхода фильма с фотографиями
7	список фильмов режиссера рекомендованного фильма
8	список героев рекомендованного фильма с указанием актеров и актрис, сыгравших эти роли
9	таймлайн с датами рождения и фотографиями участников съемочного процесса
10	список фильмов, которые получили награду на Каннском фестивале в том же году

Задание 3.1. Запрос на SPARQL

Номер	Запрос
11	список фильмов сценариста рекомендованного фильма
12	список фильмов, которые стоили более чем в 2 раза дороже, чем рекомендованный фильм
13	образовательные учреждения, в которых учился режиссер, на карте
14	список фильмов, вышедших на той же киностудии
15	список фильмов в том же жанре, вышедших в том же году
16	список фильмов не в жанре этого фильма, вышедших на той же киностудии
17	список наград, которые имеет сценарист рекомендованного фильма
18	список фильмов, которые получили Оскар или награду на Каннском фестивале в том же году
19	список фильмов, снятых в той же стране, что и рекомендованный фильм
20	список актеров фильма, которые живы к сегодняшнему дню

Задание 3.1. Запрос на SPARQL

Номер	Запрос
21	места рождения актеров фильма на карте
22	список имен детей всех актеров фильма с указанием имени актера
23	даты рождения жен (подруг) режиссера фильма
24	список стран рождения актеров фильма, отсортированный по популяции, с указанием имени актера
25	список актрис, для которых фильм стал первым в карьере
26	список актрис фильма, которые живы к сегодняшнему дню
27	список актеров, для которых фильм стал первым в карьере
28	самого старшего участника съемочного процесса
29	список фильмов, которые вышли в том же году, что и рекомендованный фильм
30	список наград, которые имеет режиссер рекомендованного фильма
31	образовательные учреждения, в которых учился сценарист, на карте

Как сдать задание 3.1?

1. Выбрать запрос по своему НОВОМУ (колонка В) порядковому номеру из списка.
2. Написать запрос в <https://query.wikidata.org/> для фильма, полученного в результате выполнения задания 2.1 (рекомендация с контекстом).
3. Загрузить ссылку на запрос на курс в «Сдать задание №3.1. Запрос на SPARQL».
4. Получить мой комментарий.
5. Подойти ко мне на практике и обсудить код.
6. После 24.11 баллы снижаются на 30%.

Задание 3.2. Запросы на SPARQL в коде

1. Реализовать запрос на SPARQL с помощью библиотеки SPARQLWrapper.
2. Использовать <https://www.wikidata.org/w/api.php>.
3. Интегрировать код, реализующий запрос на SPARQL, в код к заданию 2.1 так, чтобы запрос выполнялся для фильмов, полученных в результате работы рекомендующего алгоритма.
4. Код выгрузить на GitHub в свой репозиторий.
5. Ссылку на код выгрузить на курс в «Сдать задание №3.2. Запрос на SPARQL в коде».
6. После 24.11 баллы снижаются на 30%.

Баллы

- Задание 3.1: 0-10 баллов.
- Задание 3.2: 0-10 баллов.

Полезные ссылки

- Wikidata SPARQL endpoint
<https://query.wikidata.org/>
- Wikidata API
<https://www.wikidata.org/w/api.php>
- SPQRQL overview
<https://www.w3.org/TR/sparql11-overview/>
- SPARQL cheat sheet
<https://www.slideshare.net/LeeFeigenbaum/sparql-cheat-sheet>
- SPARQL endpoints
<https://www.w3.org/wiki/SparqlEndpoints>
- RDF Schema
<https://www.w3.org/TR/rdf-schema/>
- SPARQL for beginners
<https://www.youtube.com/watch?v=kJph4q0Im98>

Спасибо за внимание!