

关联数据查询语言

SPARQL 实战

数字人文项目开发工程师

陈尚松

1

OpenLink Virtuoso

SPARQL语法分析

2

CONTENTS

4

一, JENA操作RDF
二, RESTFUL的调用

SPARQL Endpoint案例

3

SPARQL的概念

SPARQL Protocol And Rdf Query Language

- **SPARQL** for triple store
- **SQL** for relationship database
- 是为RDF开发的一种查询语言和数据获取协议，是为W3C所开发的RDF数据模型所定义，但是可以用于任何可以用RDF来表示的信息资源

SPARQL语法基本分析

使用sparql的两个概念：**三元组**，**match**。

如select * where {?s ?p ?o}这个查询语句，where里面的就是个三元组，表示找到符合这个三元组的所有记录。由于这三个都是未知的变量，所以这条语句的意思就变成了查询RDF里所有的数据。查询到的数据会match到这三个变量上。*号表示三个数据都要取出来。如果像下面这样写：

```
select ?p  
where  
{ ?s ?p ?o}
```

则只会取出match到p上的数据

查询模式

- SELECT - 返回全部或者部分查询结果中的变量；
- ASK - 返回TRUE或FALSE，判断查询结果是否存在；
- DESCRIBE - 返回查询结果中资源的RDF Graph；
- CONSTRUCT - 依据查询结果中的三元组变量，返回新的RDF Graph。

ASK

```
GRAPH:http://mr.library.sh.cn/graph/person
```

```
ASK{
```

```
?s foaf:name ?name. FILTER CONTAINS(?name, '巴金')
```

```
}
```

true

```
ASK{
```

```
?s foaf:name ?name. FILTER CONTAINS(?name, '东野圭吾')
```

```
}
```

false

DESCRIBE

GRAPH:http://mr.library.sh.cn/graph/person

```
DESCRIBE ?s{  
  ?s foaf:name ?name. FILTER CONTAINS(?name, '盛宣怀')  
}
```


CONSTRUCT

CONSTRUCT

```
{?s foaf:name ?o} WHERE {?s foaf:name ?o .FILTER CONTAINS(?o, '盛宣怀')}
}
```

不会返回资源的所有信息，只返回Construct中要求的信息

Prefix Namespace IRI

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

n2 <http://data.library.sh.cn/entity/person/>

rdf <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

xsdh <http://www.w3.org/2001/XMLSchema#>

Subject Item

n2:4lnpsgh342b2n1ru

foaf:name

盛宣怀

SPARQL 查询语法案例

(上图手稿)

PREFIX (Namespace Prefixes)

e.g. PREFIX plant: <http://www.linkeddatatools.com/plants>

SELECT (Result Set)

e.g. SELECT ?name

FROM (Data Set)

e.g. FROM <http://www.linkeddatatools.com/plantsdata/plants.rdf>

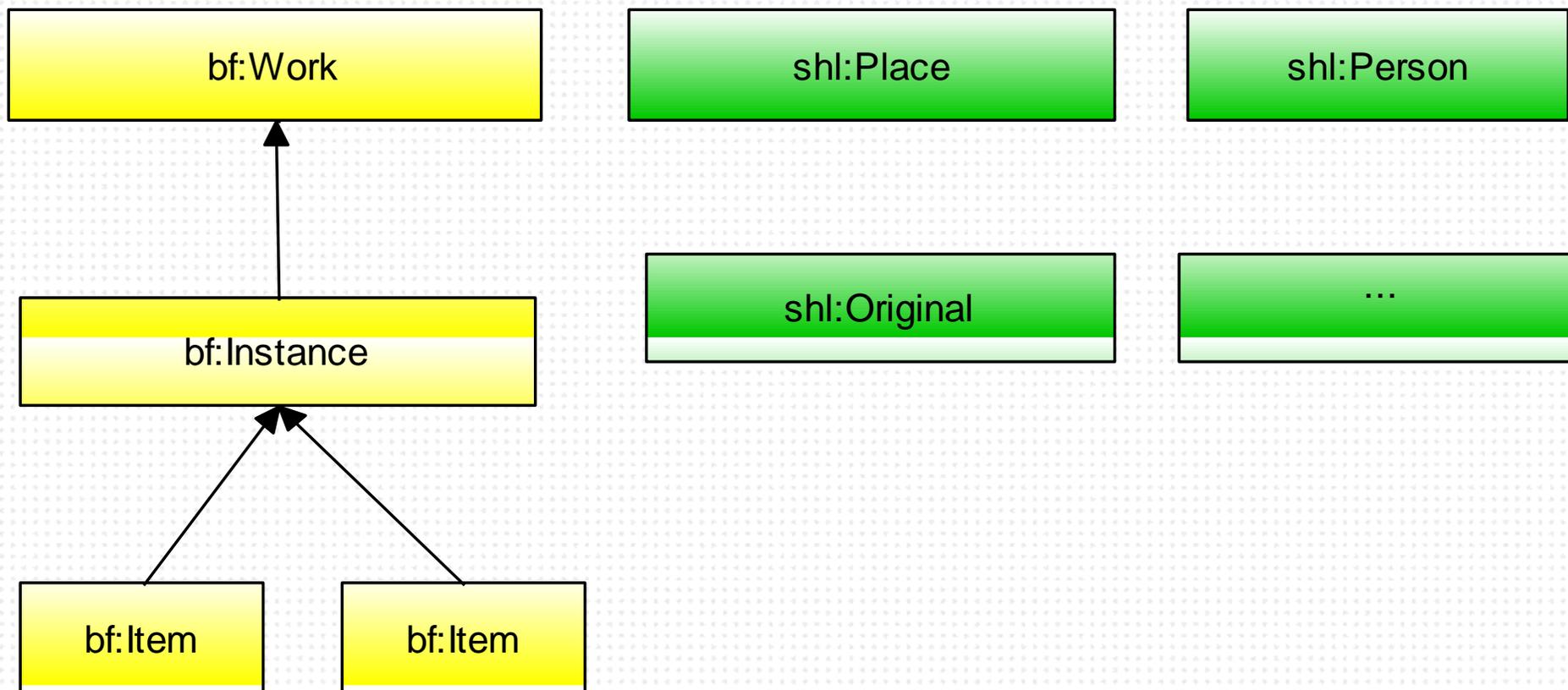
WHERE (Query Triple Pattern)

e.g. WHERE { ?planttype plant:planttype ?name }

ORDER BY, DISTINCT etc (Modifiers)

e.g. ORDER BY ?name

手稿知识库-本体架构



结果修饰

结果修饰符能够影响查询的返回结果，从而使查询在作用范围的广度和作用效果的程度上都有显著地增强。结果修饰符主要包含以下几种：

- **去重**：DISTINCT
- **可选**：OPTIONAL
- **排序**：ORDER BY
- **偏移**：OFFSET 和 **限制**：LIMIT
-

查询某个手稿URI信息

```
WITH GRAPH <http://sg.library.sh.cn/graph/work>
```

```
SELECT * WHERE {<http://data.library.sh.cn/sg/resource/work/kgswt8sfiune3zk> ?s ?p}
```

s	p
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://bibframe.org/vocab/Work
http://purl.org/dc/elements/1.1/creator	"巴金（著）"
http://bibframe.org/vocab/creator	http://data.library.sh.cn/entity/person/6pnc6xdap3cfcup
http://purl.org/dc/elements/1.1/title	"《家》《春》《秋》签名本封面复制件"
http://www.library.sh.cn/ontology/description	"未找到原件"
http://www.library.sh.cn/ontology/description	"签名本复制件"
http://www.library.sh.cn/ontology/description	"附相同内容复制件3页"
http://purl.org/dc/elements/1.1/language	"汉语"
http://purl.org/dc/elements/1.1/subject	"复制件"
http://bibframe.org/vocab/contentCategory	http://data.library.sh.cn/sg/vocab/manuscriptType/zhi-zhi-zi-liao
http://bibframe.org/vocab/originPlace	"不详"

查询所有人的姓名

```
WITH GRAPH <http://mr.library.sh.cn/graph/person>
SELECT ?name
{ ?s a shl:Person; foaf:name ?name
}
```

name
"Alighieri"@cht
"Alighieri"@chs
"Alighieri"@en
"Basic"@cht
"Basic"@chs
"Basic"@en
"CHEN-GEO=KIEN"@cht
"CHEN-GEO=KIEN"@chs
"CHEN-GEO=KIEN"@en
"cher ami"@cht
"cher ami"@chs
"cher ami"@en
"Cher Maitre"@cht
"Cher Maitre"@chs
"Cher Maitre"@en
"Cher Monsieur"@cht
"Cher Monsieur"@chs
"Cher Monsieur"@en

查询名字中包含李的人

```
WITH GRAPH <http://mr.library.sh.cn/graph/person>
```

```
SELECT ?name
```

```
{?s a shl:Person; foaf:name ?name
```

```
. FILTER contains(?name, '李')
```

```
}
```

```
}
```

name
"李一哲"@cht
"李一哲"@chs
"李一安"@cht
"李一安"@chs
"李專波"@cht
"李专波"@chs
"李世佐"@cht
"李世佐"@chs
"李中全"@cht
"李中全"@chs
"李中法"@cht
"李中法"@chs
"李中流"@cht
"李中流"@chs
"李丹"@cht
"李丹"@chs

查询名字为李鸿章的人

查询名字是李鸿章，且语言标记是简体的人

```
WITH GRAPH <http://gen.library.sh.cn/graph/person>
SELECT ?name
{?s a shl:Person; bf:label ?name
. FILTER contains(lang(?name), 'chs')
. FILTER contains(?name, '李鸿章')}
}
```

name
"李鸿章"@chs
"李鸿章"@chs
"李鸿章"@chs

查询所有人及出生时间

```
WITH GRAPH<http://mr.library.sh.cn/graph/person>
```

```
SELECT ?name ?bir
```

```
{?s a sh:Person; foaf:name?name
```

```
  OPTIONAL {?s sh:birthday ?bir .}
```

```
. FILTER contains(?name, '张')
```

```
}
```

```
ORDER BY desc (?bir)
```

name	bir
"张青山"@chs	"1967.12"
"张炜"@chs	"1956"
"张卫"@chs	"1955"
"张抗抗"@chs	"1950"
"张之先"@chs	"1946"
"张广志"@chs	"1946"
"张雷平"@chs	"1945"
"张新泉"@chs	"1941"
"张桂铭"@chs	"1939.9"
"张香华"@chs	"1939"
"张胜泽"@chs	"1937"
"张洁"@chs	"1937"
"张贤亮"@chs	"1936.12"

结果修饰-DISTINCT

```
WITH GRAPH<http://mr.library.sh.cn/graph/person>
```

```
SELECT DISTINCT ?name
```

```
{?s a shl:Person; foaf:name ?name. FILTER contains(?name, '李鸿章')  
}
```

name
"李鸿章"@chs

结果修饰-OFFSET, LIMIT

```
WITH GRAPH<http://mr.library.sh.cn/graph/person>
```

```
SELECT DISTINCT ?name
```

```
{?s a sh:Person; foaf:name ?name. FILTER contains(?name, '李')
```

```
. FILTER contains(lang(?name), 'chs')
```

```
}
```

```
OFFSET 0 LIMIT 5
```

name

"李一哲"@chs

"李一安"@chs

"李专波"@chs

"李世佐"@chs

"李中全"@chs

统计查询

```
WITH GRAPH<http://mr.library.sh.cn/graph/person>
```

```
SELECT count(?s) as ?pCount  
{?s a shl:Person; foaf:name ?name. FILTER contains(?name, '李')  
. FILTER contains(lang(?name), 'chs')  
}
```

pCount

2229

复杂查询

```
SELECT DISTINCT ?uri ?title ?typeValue WHERE {  
  ?uri a bf:Work ;  
  dc:title ?title ;bf:contentCategory ?typeUri .  
  {SELECT ?typeUri ?typeValue FROM <http://gen.library.sh.cn/graph/baseinfo> WHERE  
  {  
    ?typeUri a bf:Category ;  
  
    bf:label ?typeValue. FILTER (lang(?typeValue) = 'chs') }} .FILTER (CONTAINS(?title, '章程'  
')) }ORDER BY ?title OFFSET  
  
0 LIMIT 100
```

复杂查询

uri	title	typeValue
http://data.library.sh.cn/sg/resource/work/hh64ey6w4c0pdcx2	"上海侨友文化经济协会章程"	"纸质资料"@chs
http://data.library.sh.cn/sg/resource/work/qkt4u105nofbqtve	"中国共产党章程"	"纸质资料"@chs
http://data.library.sh.cn/sg/resource/work/33y1soplxovploq4	"中国南社与柳亚子研究会章程(修改草案)"	"纸质资料"@chs
http://data.library.sh.cn/sg/resource/work/jeuwvfgbverzdh78	"中国通俗文学学会章程及公告"	"纸质资料"@chs
http://data.library.sh.cn/sg/resource/work/78osyg8343jg2g1q	"报纸发行条例及代派章程"	"纸质资料"@chs
http://data.library.sh.cn/sg/resource/work/6u3vylqjhchm1idl	"新声音乐艺术中心章程"	"纸质资料"@chs
http://data.library.sh.cn/sg/resource/work/b5vljq8ow6os17pf	"民治新闻学院章程"	"纸质资料"@chs
http://data.library.sh.cn/sg/resource/work/sukakzkk3ofomfea	"生活教育社南京分社章程"	"纸质资料"@chs
http://data.library.sh.cn/sg/resource/work/6pq6c1l7kv0bx4ra	"生活教育社章程"	"纸质资料"@chs
http://data.library.sh.cn/sg/resource/work/uwnfhq4v78tngb0q	"租赁著作权章程"	"纸质资料"@chs
http://data.library.sh.cn/sg/resource/work/u07hyv37ki3mgqba	"租赁著作权章程"	"证书"@chs

数据结构查询

//查询本体中所有属性

```
SELECT DISTINCT ?s  
WHERE { ?s a rdf:Property . } ORDER BY ASC(?s)
```

s

<http://bibframe.org/vocab/agent>

<http://bibframe.org/vocab/category>

<http://bibframe.org/vocab/categoryType>

<http://bibframe.org/vocab/categoryValue>

<http://bibframe.org/vocab/contentCategory>

<http://bibframe.org/vocab/contribution>

<http://bibframe.org/vocab/contributor>

<http://bibframe.org/vocab/creator>

<http://bibframe.org/vocab/dimensions>

<http://bibframe.org/vocab/edition>

<http://bibframe.org/vocab/extent>

<http://bibframe.org/vocab/heldBy>

<http://bibframe.org/vocab/identifiedBy>

<http://bibframe.org/vocab/instanceOf>

<http://bibframe.org/vocab/itemOf>

数据结构查询

//查某个属性的相关信息

```
SELECT ?p ?o WHERE { <http://bibframe.org/vocab/label> ?p ?o . }
```

p	o
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/1999/02/22-rdf-syntax-ns#Property
http://www.w3.org/2000/01/rdf-schema#label	“规范名”
http://www.w3.org/2000/01/rdf-schema#domain	http://bibframe.org/vocab/Resource
http://www.w3.org/2000/01/rdf-schema#range	http://www.w3.org/2000/01/rdf-schema#Literal
http://www.w3.org/2000/01/rdf-schema#comment	“表示某一属性的值的字符串。”

数据结构查询

////查询某个URI下的详细信息，并且取值为URI的属性，返回标记:1

```
SELECT ?v (isURI(?v) AS ?t) WHERE {  
<http://data.library.sh.cn/sg/resource/work/0x7xcnazr1llzjqy> ?p ?v .}
```

v	t
http://bibframe.org/vocab/Work	1
"巴金（著）"	0
http://data.library.sh.cn/entity/person/6pnc6xdap3cfkcup	1
"《巴金全集》(共二十六册)"	0
"巴金"	0
"汉语"	0
"签名本"	0
"综合作品集"	0
http://data.library.sh.cn/sg/vocab/manuscriptType/qian-ming-ben	1
"北京"	0

数据结构查询

//查询某个URI所属的类

```
SELECT ?c WHERE {<http://data.library.sh.cn/sg/resource/work/0x7xcnazr1llzjqy> a ?c .}
```

c

<http://bibframe.org/vocab/Work>

SPARQL -INSERT

INSERT DATA

```
{ GRAPH <http://gen.library.sh.cn/graph/work>
  { <http://data.library.sh.cn/sg/resource/work/0x7xcnazr1llzjqy>
    bf:label 'XXXXXX '@cht.
  }
}
```

SPARQL -UPDATE

//属性更新

GRAPH: <http://gen.library.sh.cn/graph/person>

delete {?s bf:role <<http://data.library.sh.cn/sg/vocab/role/1>>}

insert {?s bf:role <<http://data.library.sh.cn/sg/vocab/role/2>>} where {?s bf:role
<<http://data.library.sh.cn/sg/vocab/role/1>>}

SPARQL -DELETE

//删除bf:title属性

GRAPH:http://gen.library.sh.cn/graph/work

```
delete {?s bf:title ?title}
where {?s a bf:Work ; bf:title ?title .}
```

//删除某整条数据

GRAPH::http://gen.library.sh.cn/graph/place

```
delete {?s ?p ?o}
where {?s ?p ?o. filter strstarts(str(?s),'http://data.library.sh.cn/entity/place/m42oi32rxq8q85vf')}
```

利用JENA操作RDF

Jena的理解：

Jena是一个Java开发工具包, 用于Semantic Web(语义网)中的应用程序开发；Jena是开源的，在下载文档中有Jena的完整代码。Jena框架主要包括:

- a) 以RDF/XML、三元组形式读写RDF
- b) RDFS，OWL，DAML+OIL等本体的操作
- c) 利用数据库保存数据
- d) 查询模型
- e) 基于规则的推理

下载地址：<http://jena.sourceforge.net/downloads.html>

利用JENA操作RDF

RDF的理解：

每个RDF模型中的arc(既一组关系，因为RDF语言本身就是一组Triple) 可以看做是一个statement。它对资源进行维护，包含以下三个部分：

Subject：就是arc定义的地方，可以理解为Resource

Predicate：表示arc的属性

Object：表示arc指向的对象（或者常量）

利用JENA操作RDF

编写一个RDF

编写一个RDF主要有三个步骤：

- (1) 使用ModelFactory的createDefaultModel()方法构造一个空的model。
- (2) 使用model的createResource方法，通过URI构建资源。
- (3) 向Resource中添加属性，使用addProperty方法。

读写RDF

Jena支持从文件中读写RDF模型

读取：用read方法读入输入流

或者用FileManager.get().readModel(model, 文件路径)

写入：用write方法

控制prefix

在rdf中，定义的url会自动给予一个变量名，如果要改变这个名字，使用自定义的名字，就要用setNsPrefix (设置的变量名，要设置的uri) 方法

利用JENA操作RDF

参考资料

- [1] \\Jena-2.4\doc\ontology Jena 2 Ontology API文档
- [2] \\Jena-2.4\doc\javadoc Jena 2 Java文档
- [3] \\Jena-2.4\src-examples\jena\examples\ontology\describeClass
- [4] <http://www.chengtao.name/modules/wordpress> 维基博客
- [5] iMarine <http://iMarine.blog.163.com>

SPARQL Endpoint

- SPARQL Endpoint是遵循SPARQL协议的查询终端，
- 用户可以通过SPARQL查询语言对知识库进行查询操作。SPARQL终端提供Web交互界面，支持用户直接输入查询语句进行查询，同时也支持机器用户通过HTTP URI+SPARQL语句的Restful方式调用。
- SPARQL Endpoint支持GET/POST请求方式；支持各种查询类型，如CONSTRUCT（构建）、DESCRIBE（描述）；查询结果以XML/JSON/N3等多种方式输出

手稿知识库

- SPARQL Endpoint

<http://data.library.sh.cn:8892/sparql>

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text

```
SELECT DISTINCT ?uri ?title ?typeValue WHERE {  
    ?uri a bf:Work ; dc:title ?title ;bf:contentCategory ?typeUri .  
  
    {SELECT ?typeUri ?typeValue FROM <http://gen.library.sh.cn/graph/baseinfo> WHERE {  
        ?typeUri a bf:Category ;  
        bf:label ?typeValue. FILTER (lang(?typeValue) = 'chs' ) }  
    } .FILTER (CONTAINS(?title, '章程')) }ORDER BY ?title OFFSET  
0 LIMIT 100|
```

Sponging:

Use only local data (including data retrieved before), but do not retrieve more ▼

Results Format:

HTML ▼

Execution timeout:

0 milliseconds (values less than 1000 are ignored)

Options:

 Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#))

Run Query

Reset

uri	title	typeValue
http://data.library.sh.cn/sg/resource/work/hh64ey6w4c0pdcx2	“上海侨友文化经济协会章程”	“纸质资料”@chs
http://data.library.sh.cn/sg/resource/work/gkt4ul05nofbqtve	“中国共产党章程”	“纸质资料”@chs
http://data.library.sh.cn/sg/resource/work/33y1soplxovploq4	“中国南社与柳亚子研究会章程(修改草案)”	“纸质资料”@chs
http://data.library.sh.cn/sg/resource/work/jeuwvfgbverzdh78	“中国通俗文学学会章程及公告”	“纸质资料”@chs
http://data.library.sh.cn/sg/resource/work/78osyg8343jg2g1g	“报纸发行条例及代派章程”	“纸质资料”@chs
http://data.library.sh.cn/sg/resource/work/6u3vy1qjhchmlidl	“新声音乐艺术中心章程”	“纸质资料”@chs
http://data.library.sh.cn/sg/resource/work/b5vljq8ow6os17pf	“民治新闻学院章程”	“纸质资料”@chs
http://data.library.sh.cn/sg/resource/work/sukakzkk3ofomfea	“生活教育社南京分社章程”	“纸质资料”@chs
http://data.library.sh.cn/sg/resource/work/6pq6c117kv0bx4ra	“生活教育社章程”	“纸质资料”@chs
http://data.library.sh.cn/sg/resource/work/uwnfhq4v78tngb0g	“租赁著作权章程”	“纸质资料”@chs
http://data.library.sh.cn/sg/resource/work/u07hyv37ki3mgqba	“租赁著作权章程”	“证书”@chs

RESTful概念

RESTful是一种设计模式，或者说是一种设计规范，实际上它有的只是几个原则，当一个应用满足这些原则的时候，可以认为它是RESTful的。

符合下面设计规范的应用，这些规范包括：

模型表示 (Representations)

消息 (Messages)

URIs一致接口 (Uniform interface)

无状态 (Stateless)

资源之间的链接 (Links between resources)

缓存 (Caching)

RESTful用处

RESTful可以认为是一种建立在HTTP协议之上的设计模式，充分的利用了HTTP协议的特定，

使用URL来表示资源，用各个不同的HTTP动词来表示对资源的各种行为。

这样做的好处就是资源和操作分离，让资源的管理更加规范。

使用对象：1，自己开发的应用，2，第三方。其他有很多开放的API服务他们都是使用RESTful规范，如各种API。

RESTful调用

- <http://data.library.sh.cn:8890/sparql>

Client需要提交的参数：

default-graph-uri=<http://mr.library.sh.cn/graph/person>

query=`select * where {?s foaf:name ?name. filter contains(str(?name),'陈')}`

format=json

```
{ "head": { "link": [], "vars": ["s", "name"] },
  "results": { "distinct": false, "ordered": true, "bindings": [
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/7vo2mhzimpml7l1i" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u4E1C\u9648" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/p554rqcjm1z3wodf" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u4E25\u9648" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/iiqoa5rlxwk929eq" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E00\u98DE" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/z5d3f2in7nokie8p" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E00\u9E23" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/tuevi3n6bhx3jkt5" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E15\u663E" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/9x2qc8je5luuk4gq" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E16\u4E94" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/yc675zecqpmefr6" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E16\u4FCA" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/ivycvyeltrh7ivaq" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E16\u7389" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/nu7ki5x5pfdstpqu" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E19\u6EE2" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/5l6eron9rqihgazr" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E19\u83B9" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/dw2jy4ijqucfoemv" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E1A\u6052" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/fupolcalt8tr7ec3" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E4B\u592B" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/6k12xkrffow5lcsb" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E4B\u9E64" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/6kc6bota3yb12jbx" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E66\u6C49" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/jvs1sjoly2prpib3" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E66\u7389" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/btzrdansgbdh3yvf" }, "name": { "type": "literal", "xml:lang": "chs", "value": "\u9648\u4E91" } },
    { "s": { "type": "uri", "value": "http://data.library.sh.cn/entity/person/i3do83vvy5o4q12e" }, "name": { "type": "literal", "xml:lang": "chs", "value":
```

RESTful调用

- 更多具体调用实例，请参照：[手稿开放数据接口](#)。



THANKS.