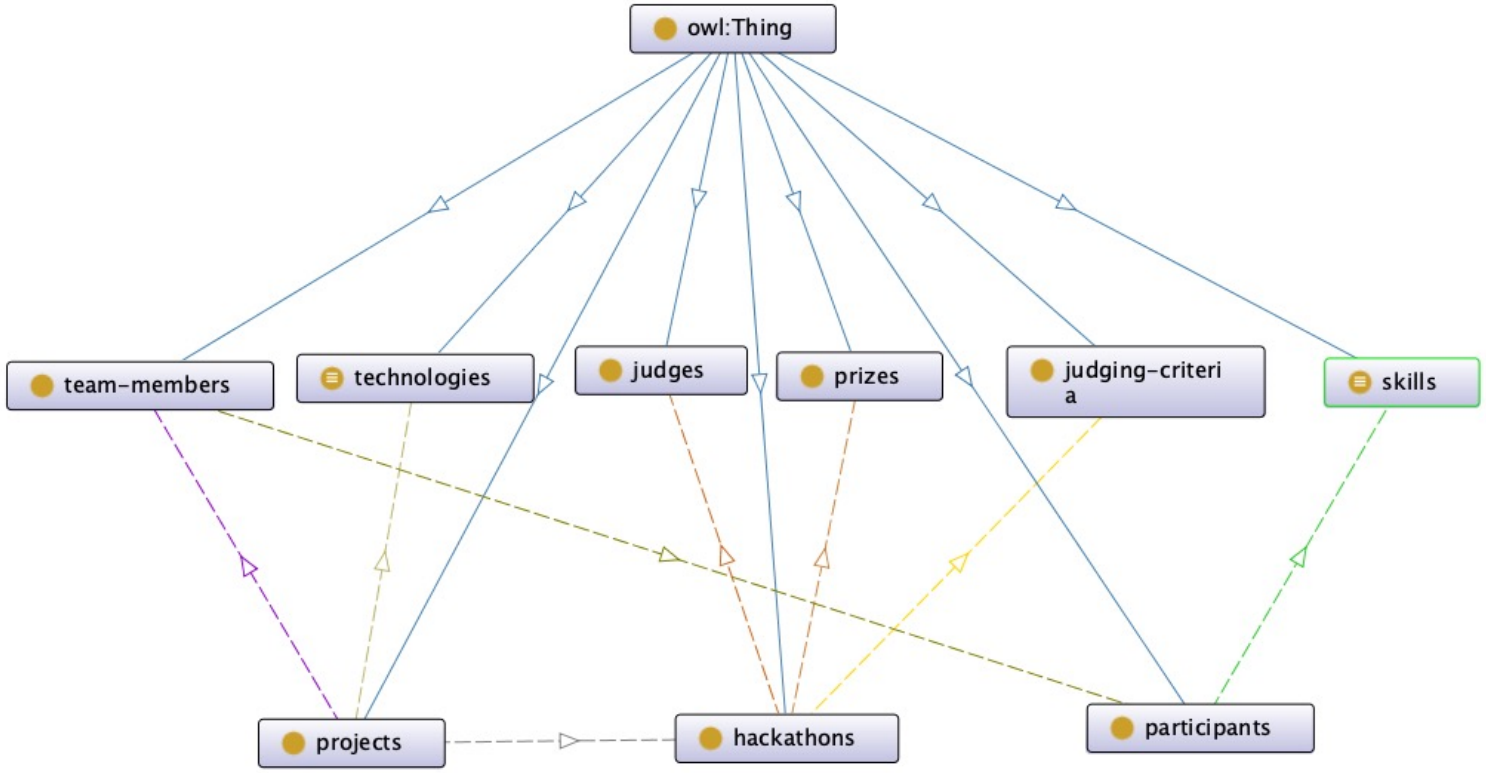
Knowledge Engineering

HacKG

He Zhitao 107*****
Zhang Yuedong 105*****

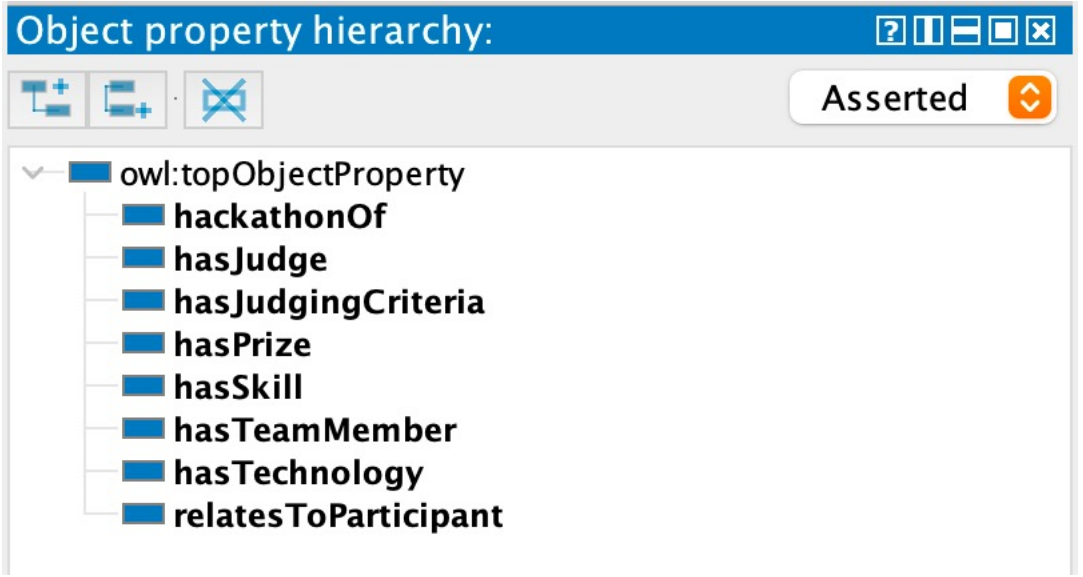
Ontology Design: Hackathons

- Hackathon ontology includes a set of classes and properties about hackathon events.
- The project follows all steps of Knowledge engineering ,Data linking and Querying.

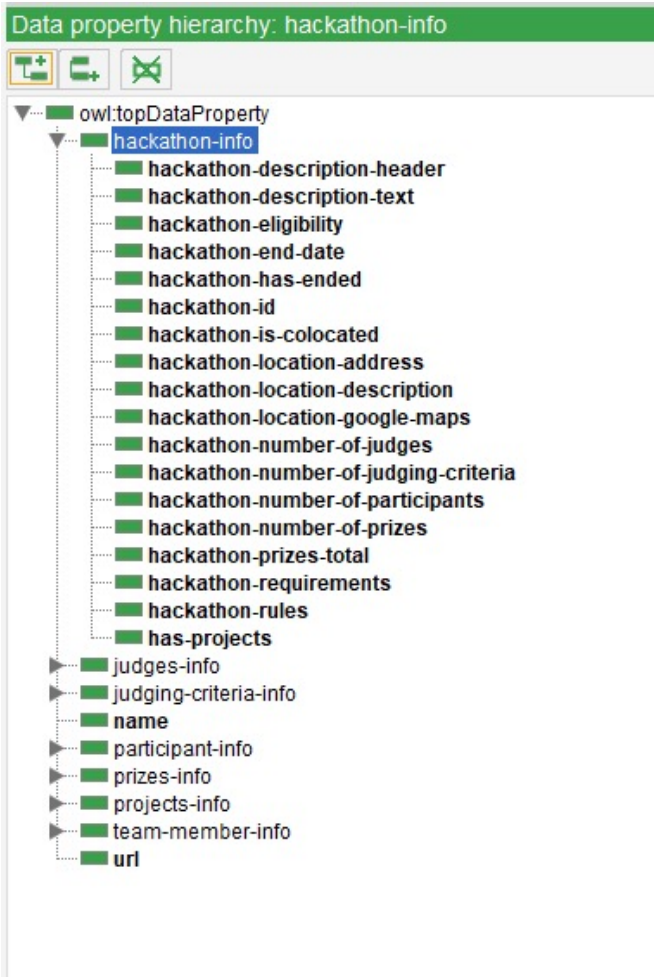


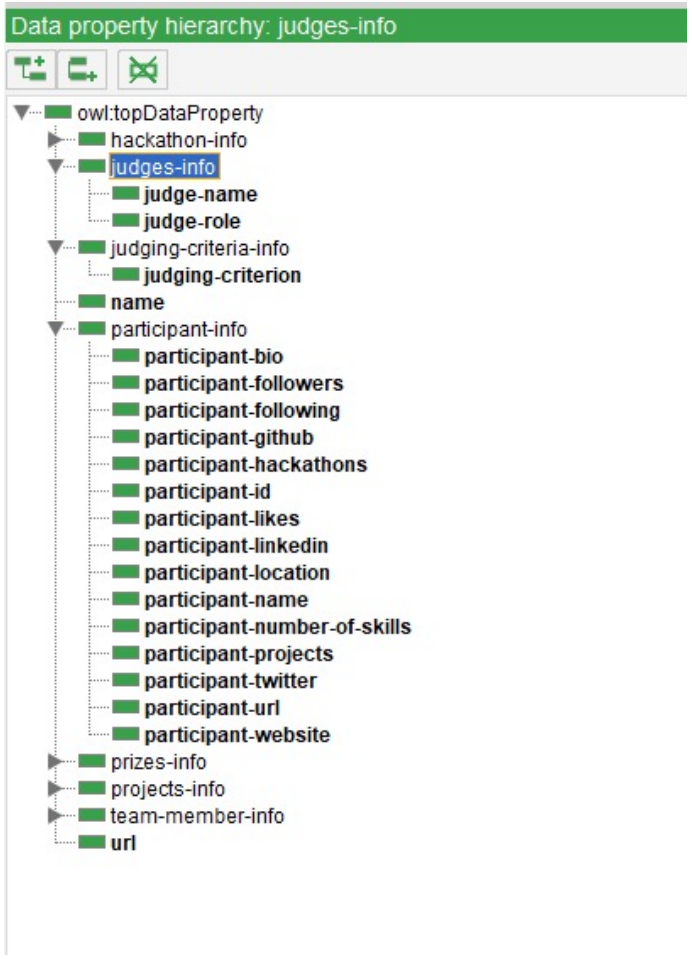
Ontology: Classes and Properties

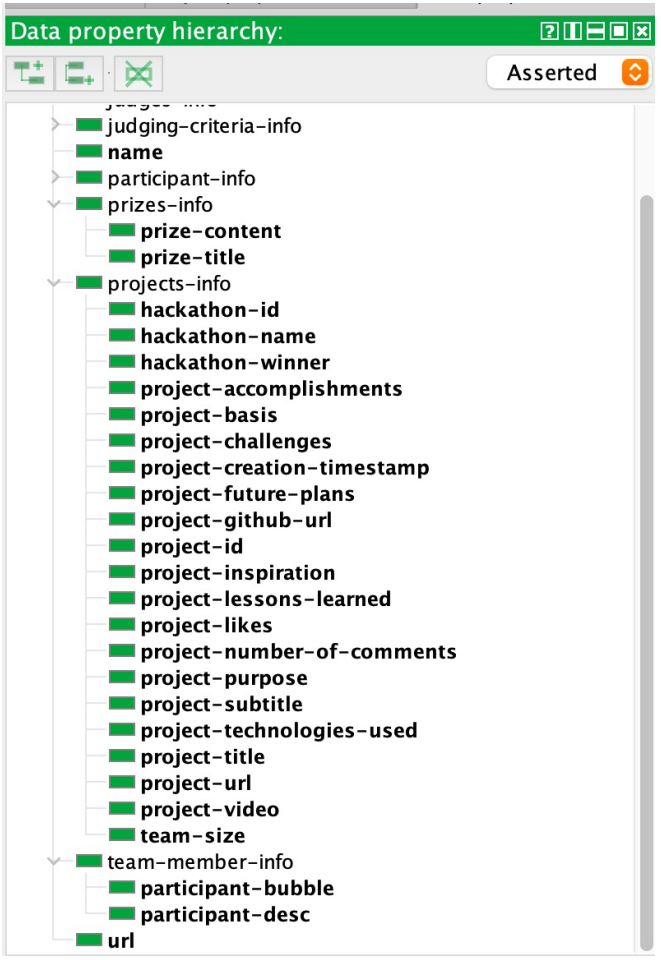
Hackathons: pieces of data describe hackathon events
Prizes: every hackathon event sets up prizes
Judges: every hackathon event has several judges
Judging-criteria: hackathon events have judging criteria
Participants: relate to team members who consist of teams participating in projects
Skills: participants have some skills which can be used as technologies required by projects
Projects: each project relates to a hackathon event and has several team members
Technologies: required to complete projects
Team-members: each project has a team including some members



Ontology: Data Properties







Knowledge engineering: Competency Questions

- 1.What are the projects with at least two participants ?
- 2.Which co-located projects have more than 5 judges ?
- 3.What projects are included in Hackathon events located in New York or West Lafayette ?
- 4.Who are the judges participating in more than one hackathon event ?
- 5.What projects did "Anano Bodokia" participate in ?

Dataset

- The dataset is composed of many separate JSON files and stored in Dropbox, we wrote a Python script to downloads these files, and merges them together.

```python
import wget
from bs4 import BeautifulSoup

#
# # For participants
# output_directory = 'Resources/participants'
# source_path = 'Resources/Dropbox_HTML/participants - Dropbox.html'

# For projects
output_directory = 'Resources/projects'
source_path = 'Resources/Dropbox_HTML/projects - Dropbox.html'

f = open(source_path, "r")
soup = BeautifulSoup(f.read(), 'html.parse
link_form1 = 'https://www.dropbox.com/sh'
link_form2 = '.json'

i = 0
for link in soup.find_all('a'):
    link = f"{link.get('href')[0: -1]}1"
    print(link)

    if link_form1 in link and link_form2 i
        # wget.download(link, out=output_d
        i = i + 1
        if i > 100000:
            break
        print(link)
print(i)
```

```python
import os
import subprocess
import shutil
typeName=""
path=""
newpath=""

def process_files(typeName,path,newpath):
    path="D:/Polimi/M1-S2/KE/project/data/%s"%typeName
    newpath="D:/Polimi/M1-S2/KE/project/data/%s-rdf"%typeName
    integration_path=os.path.join(newpath,typeName+".rdf")
    integration_file=open(integration_path,'a+')
    print(integration_path)

    integration_json_path=os.path.join(newpath,typeName+"_total.json")
    integration_json_file=open(integration_json_path,'a+')
    w=integration_json_file.write("[")

    old_names=os.listdir(path)
    total_count=0
    for index,old_name in enumerate(old_names):
        if(old_name.endswith('.json')):
            total_count=total_count+1
    for index,old_name in enumerate(old_names):
        if(old_name.endswith('.json')):
            front_name=old_name.split('.')
            print(front_name)
            name=front_name[0]
```

Data Linking

YARRRML : rml.io

- R2RML rules that fit Hackathon ontology model
- Written and parsed using the Yarrrml matey tool.

Multiple rules have been written:

- Rules mapping the hackathons associated with judges, prizes, and Judging Criteria.
- Rules mapping the participants associated with skills.
- Rules mapping the projects associated with hackathons, technologies, and team members.

```
prefixes:
  hac: "http://www.semanticweb.org/hezhang/ontologies/2021/6/Hackathon#"
  xsd: "http://www.w3.org/2001/XMLSchema#"

mappings:
  participants:
    sources:
      - ['participants.json~jsonpath', "$"]
    s: hac:$(participant-id)
    po:
      - [a, hac:participants]
      - [hac:participant-likes, $(participant-likes),xsd:integer]
      - [hac:participant-projects,$(participant-projects),xsd:integer]
      - [hac:participant-url,$(participant-url),xsd:string]
      - [hac:participant-name,$(participant-name),xsd:string]
      - [hac:participant-followers,$(participant-followers),xsd:integer]
      - [hac:participant-location,$(participant-location),xsd:string]
      - [hac:participant-bio,$(participant-bio),xsd:string]
      - [hac:participant-linkedin,$(participant-linkedin),xsd:string]
      - [hac:participant-website,$(participant-website),xsd:string]
      - [hac:participant-github,$(participant-github),xsd:string]
      - [hac:participant-hackathons,$(participant-hachathons),xsd:integer]
      - [hac:participant-following,$(participant-following),xsd:integer]
      - [hac:participant-number-of-skills,$(participant-number-of-skills),xsd:integer]
      - [hac:participant-twitter,$(participant-twitter),xsd:string]

      - p: hac:hasSkill
        o:
          - mapping: participant-skills
            condition:
              function: equal
              parameters:
              - [ str1, 1, s]
              - [ str2, 1, o]


  participant-skills:
    sources:
      - ['participants.json~jsonpath', '$.participant-skills[*]']
    s: hac:$(name)
    po:
      - [a, hac:skills]
      - [hac:url, $(url),xsd:string]
      - [hac:name, $(name),xsd:string]
```

Data Linking

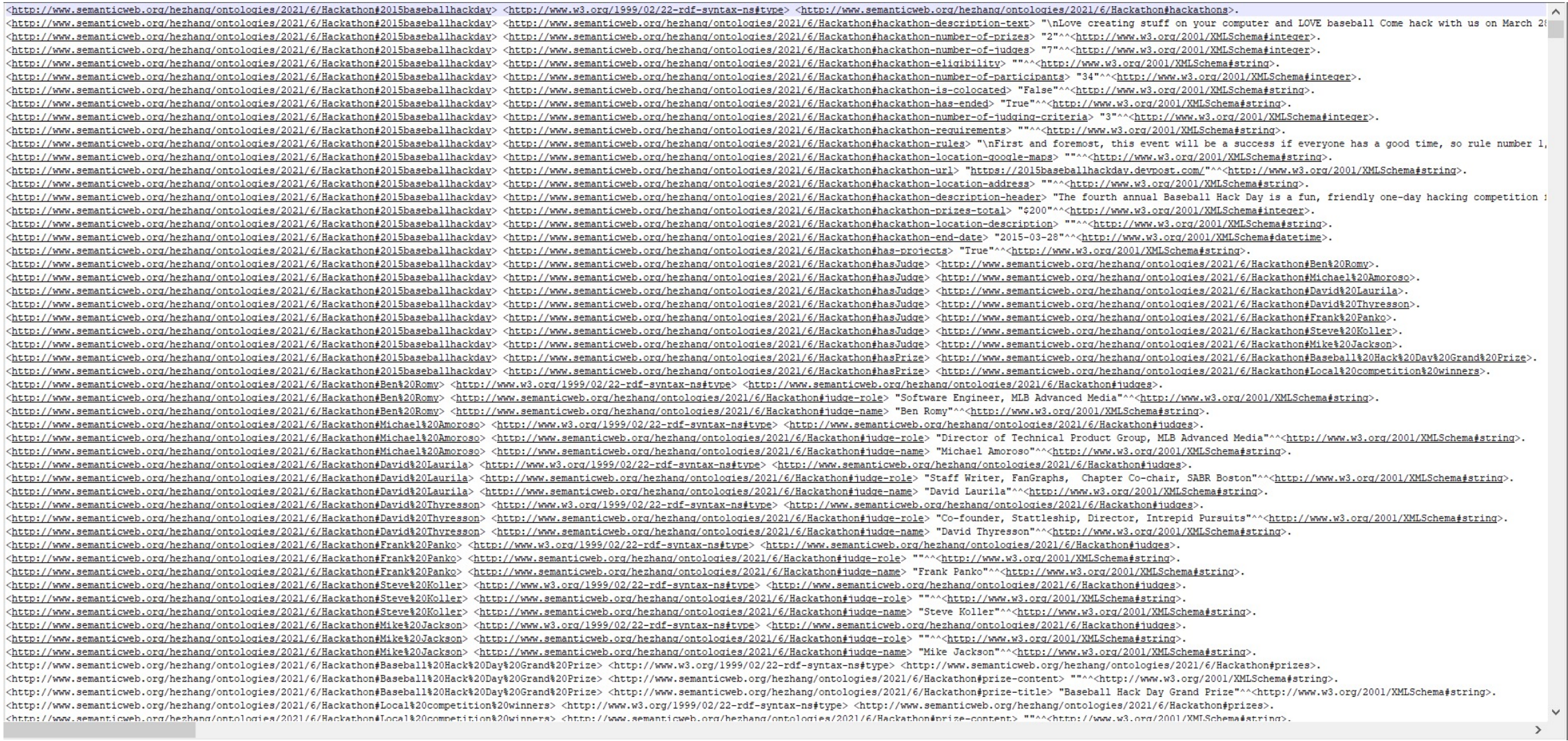
```
prefixes:
  hac: "http://www.semanticweb.org/hezhang/ontologies/2021/6/Hackathon#"
  xsd: "http://www.w3.org/2001/XMLSchema#"

mappings:
  hackathons:
    sources:
      - ['hackathons.json~jsonpath', "$"]
    s: hac:$(hackathon-id)
    po:
      - [a, hac:hackathons]
      - [hac:hackathon-description-text, $(hackathon-description-text),xsd:string]
      - [hac:hackathon-number-of-prizes,$(hackathon-number-of-prizes),xsd:integer]
      - [hac:hackathon-number-of-judges,$(hackathon-number-of-judges),xsd:integer]
      - [hac:hackathon-eligibility,$(hackathon-eligibility),xsd:string]
      - [hac:hackathon-number-of-participants,$(hackathon-number-of-participants),xsd:integer]
      - [hac:hackathon-is-colocated,$(hackathon-is-colocated),xsd:string]
      - [hac:hackathon-has-ended,$(hackathon-has-ended),xsd:string]
      - [hac:hackathon-number-of-judging-criteria,$(hackathon-number-of-judging-criteria),xsd:integer]
      - [hac:hackathon-requirements,$(hackathon-requirements),xsd:string]
      - [hac:hackathon-rules,$(hackathon-rules),xsd:string]
      - [hac:hackathon-location-google-maps,$(hackathon-location-google-maps),xsd:string]
      - [hac:hackathon-url,$(hackathon-url),xsd:string]
      - [hac:hackathon-location-address,$(hackathon-location-address),xsd:string]
      - [hac:hackathon-description-header,$(hackathon-description-header),xsd:string]
      - [hac:hackathon-prizes-total,$(hackathon-prizes-total),xsd:integer]
      - [hac:hackathon-location-description,$(hackathon-location-description),xsd:string]
      - [hac:hackathon-end-date,$(hackathon-end-date),xsd:datetime]
      - [hac:has-projects,$(has-projects),xsd:string]
      - p: hac:hasJudge
        o:
          - mapping: hackathon-judges
            condition:
              function: equal
              parameters:
              - [ str1, 1, s]
              - [ str2, 1, o]

      - p: hac:hasPrize
        o:
          - mapping: hackathon-prizes
            condition:
              function: equal
              parameters:
              - [str1, 1, s]
              - [str2, 1, o]

      - p: hac:hasJudgingCriteria
```

```
prefixes:
  hac: "http://www.semanticweb.org/hezhang/ontologies/2021/6/Hackathon#"
  xsd: "http://www.w3.org/2001/XMLSchema#"

mappings:
  projects:
    sources:
      - ['projects.json~jsonpath', "$"]
    s: hac:$(project-id)
    po:
      - [a, hac:projects]
      - [hac:hackathon-winner,$(hackathon-winner),xsd:string]
      - [hac:hackathon-name,$(hackathon-name),xsd:string]
      - [hac:project-technologies-used, $(project-technologies-used),xsd:integer]
      - [hac:project-video,$(project-video),xsd:string]
      - [hac:project-basis,$(project-basis),xsd:string]
      - [hac:project-challenges,$(project-challenges),xsd:string]
      - [hac:project-future-plans,$(project-future-plans),xsd:string]
      - [hac:project-purpose,$(project-purpose),xsd:string]
      - [hac:project-title,$(project-title),xsd:string]
      - [hac:project-inspiration,$(project-inspiration),xsd:string]
      - [hac:project-lessons-learned,$(project-lessons-learned),xsd:string]
      - [hac:project-accomplishments,$(project-accomplishments),xsd:string]
      - [hac:project-likes,$(project-likes),xsd:integer]
      - [hac:project-subtitle,$(project-subtitle),xsd:integer]
      - [hac:project-github-url,$(project-github-url),xsd:string]
      - [hac:project-number-of-comments,$(project-number-of-comments),xsd:integer]
      - [hac:team-size,$(team-size),xsd:integer]
      - [hac:project-creation-timestamp,$(project-creation-timestamp),xsd:datetime]
      - [hac:project-url,$(project-url),xsd:string]

      - p: hac:hackathonOf
        o:
          - mapping: hackathons
            condition:
              function: equal
              parameters:
              - [ str1, $(hackathon-id), s]
              - [ str2, $(hackathon-id), o]

      - p: hac:hasTechnology
        o:
          - mapping: technologies
            condition:
              function: equal
              parameters:
              - [ str1, 1, s]
              - [ str2, 1, o]
```

Data Linking: RDF triples

RMLMAPPER: rmlmapper.jar

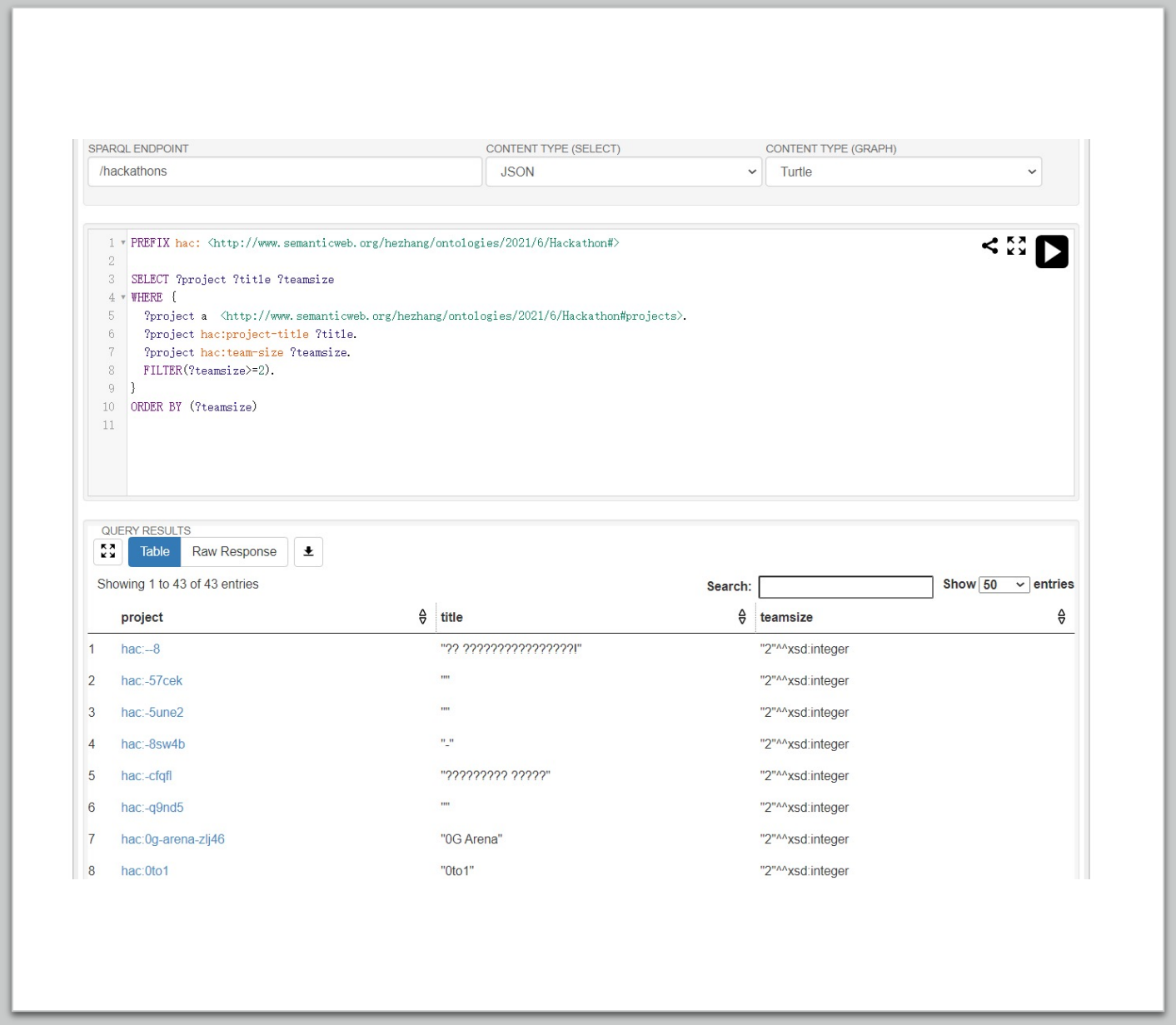
- Generated RDF triples from the rml rules



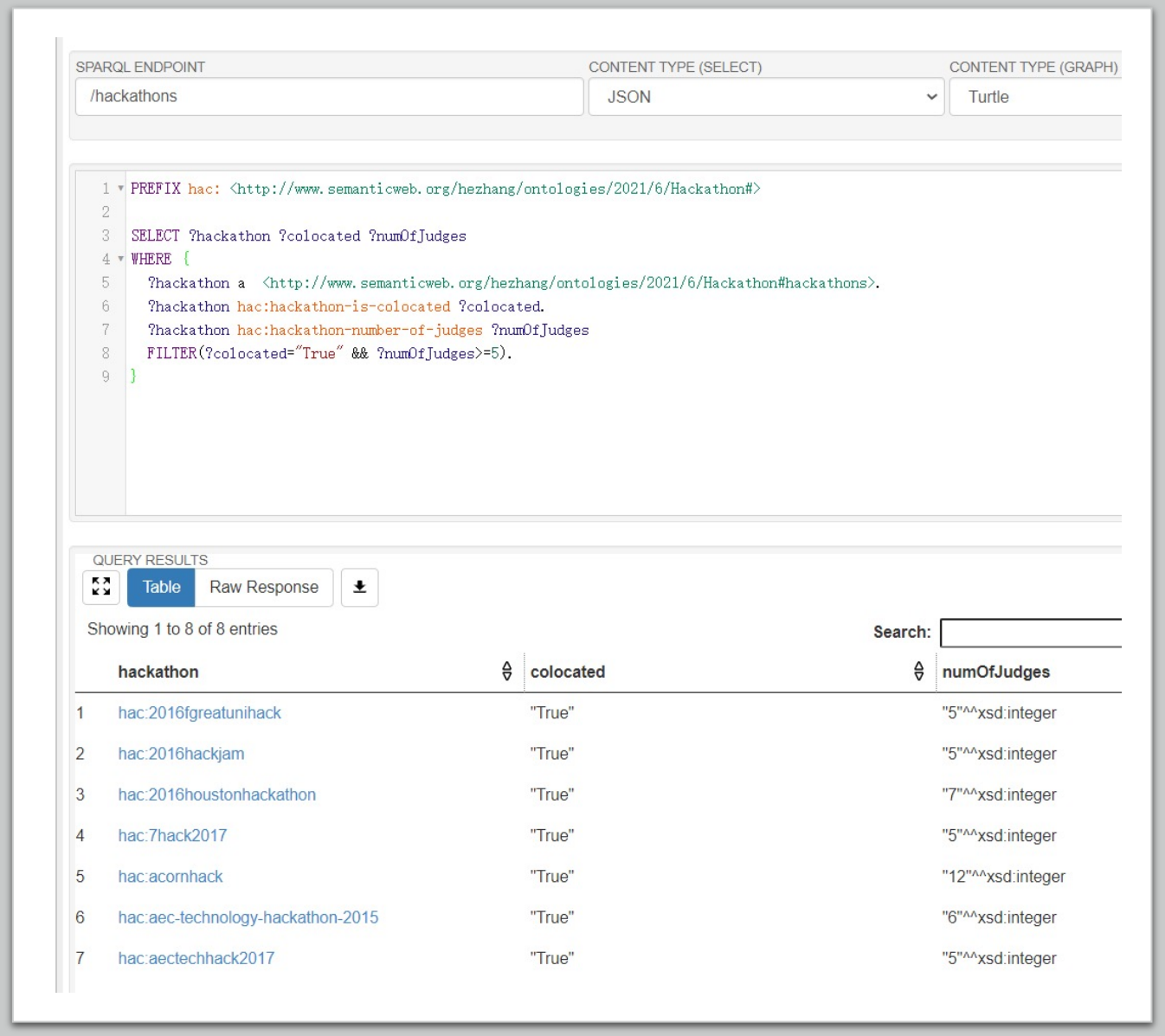
Publishing & Querying

1. SPARQL endpoint was created using Apache- Jena-Fuseki.
2. Converted RDF format files to NT format and upload them to Fuseki as dataset.
3. Translated competency questions to sparql.

Query 1. What are the projects with at least two participants?

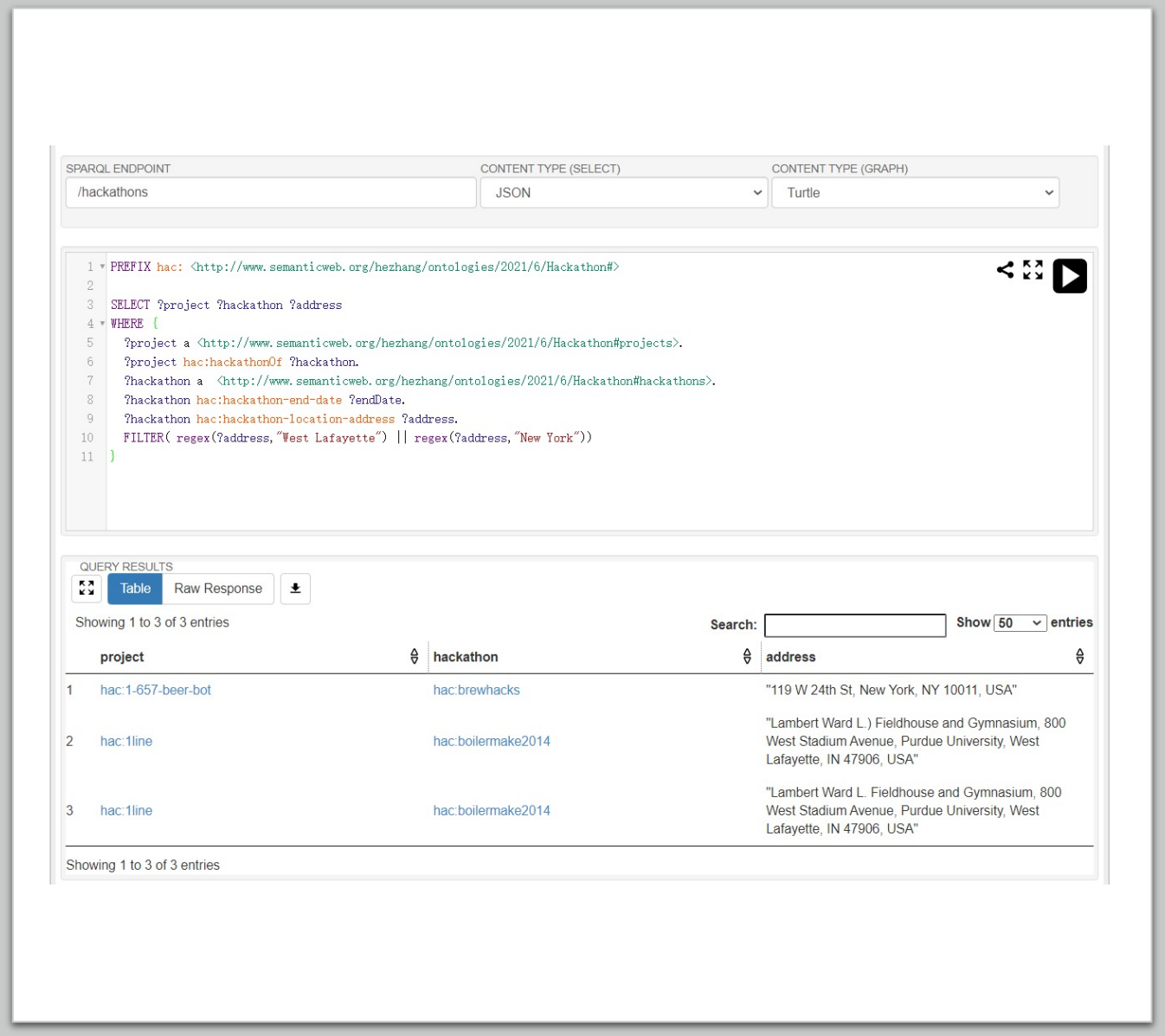


Query 2. Which "Co-located" hackathons have more than 5 judges?



	hackathon	colocated	numOfJudges
1	hac:2016fgreatunihack	"True"	"5"^^xsd:integer
2	hac:2016hackjam	"True"	"5"^^xsd:integer
3	hac:2016houstonhackathon	"True"	"7"^^xsd:integer
4	hac:7hack2017	"True"	"5"^^xsd:integer
5	hac:acornhack	"True"	"12"^^xsd:integer
6	hac:aec-technology-hackathon-2015	"True"	"6"^^xsd:integer
7	hac:aectechhack2017	"True"	"5"^^xsd:integer

Query 3. What projects belongs to Hackathon which location in New York or West Lafayette?



Query 4. Who are the judges who participated in more than one hackathon event?

```
PREFIX hac: <http://www.semanticweb.org/hezhang/ontologies/2021/6/Hackathon#>
SELECT ?judge ?role ?total_hack
WHERE{
  ?judge hac:judge-role ?role
   {
SELECT ?judge (count(?hackathon) as ?total_hack)
WHERE {
  ?judge a <http://www.semanticweb.org/hezhang/ontologies/2021/6/Hackathon#judges>.
  ?hackathon a <http://www.semanticweb.org/hezhang/ontologies/2021/6/Hackathon#hackathons>.
  ?hackathon hac:hasJudge ?judge.
  ?judge hac:judge-name ?name.
}group by (?judge) order by desc(?total_hack)
  }FILTER(?total_hack>1)
}
```

QUERY RESULTS

Table | Raw Response

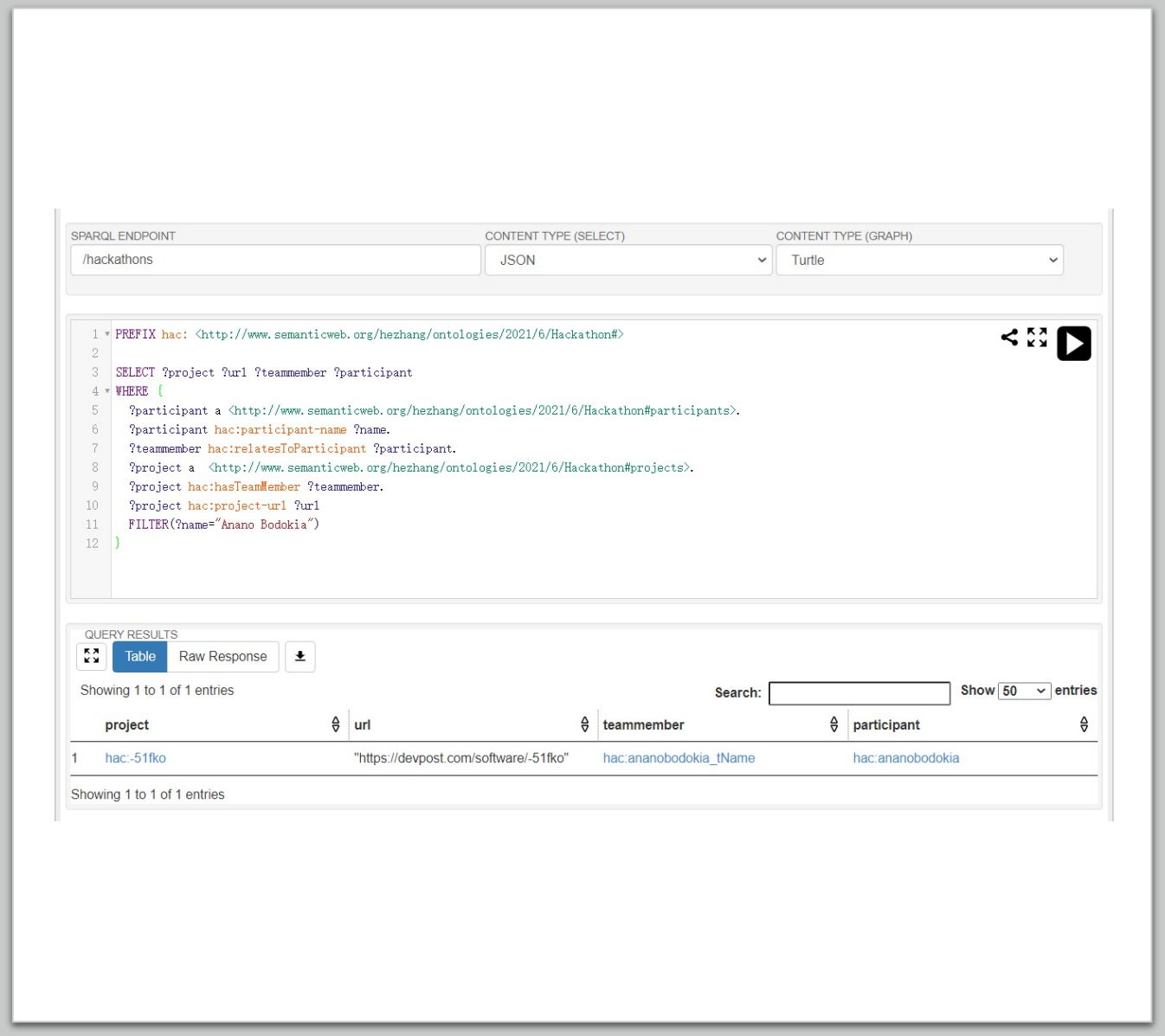
Showing 1 to 7 of 7 entries

Search: Show 50 entries

	judge	role	total_hack
1	hac:Chandran%20Elumalai	"Software Manager GE"	"2"^^xsd:integer
2	hac:Chris%20Baldwin	"Software Manager GE"	"2"^^xsd:integer
3	hac:Habib%20Khan	"Software Engineer GE"	"2"^^xsd:integer
4	hac:Patrick%20Crawford	"National Director"	"2"^^xsd:integer
5	hac:Rob%20Otani	"Principal at Thornton Tomasetti"	"2"^^xsd:integer
6	hac:Ryan%20Luo	""	"2"^^xsd:integer
7	hac:Ryan%20Luo	"Google"	"2"^^xsd:integer

Showing 1 to 7 of 7 entries

Query 5. What projects did "Anano Bodokia" participate in ?



Thanks for your attention