LODQA, PubDictionaries Hands-On

Jin-Dong Kim (DBCLS)

Agenda

- Demo
  - ✔ To understand the LODQA configuration
    - ➔ LODQA (http://www.lodqa.org)
    - ➔ PubDictionaries (http://pubdictionaries.org)
- Exercise
  - ✔ To make a configuration of LODQA for a SPARQL EP

Exercise (Overview)

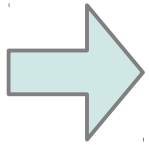
- With Orphanet in Bio2RDF
  1) To get the label, URO mapping
  2) To clean it up
  3) To upload it on PubDictinaries
  4) To write a configuration in JSON while testing with LODQA
  5) To put the JSON file somewhere visible from the net, and specify it

Exercise (Step 1, 2)

- With Orphanet in Bio2RDF (http://cu.orphanet.bio2rdf.org/sparql)

1) To get the label, URO mapping

```
select ?l ?x where {
  ?x rdfs:label ?l
}
```



Get it in TSV

2) To clean it up

```
sed "s/^\"//" sparql.tsv | sed "s/\"$//" | sed "s/\"\t\"/\t/" |
sed "s/\[[^\[]*]\t/\t/" > orphanet-dic.tsv
```

Exercise (Step 3)

3) To upload it on PubDictionaries (http://pubdictionaries.org)
   a) To sign-up.
   b) To create a dictionary ('orphanet') and upload the TSV file.
   c) To wait for 5 minutes for completion of indexing
   d) To check the exception message and confirm it.
   e) To get the URL of the REST service for "Term to Id" mapping

Exercise (Step 3) - continue

- Test the REST service

```
curl -H "content-type:application/json" -d '["dental
ankylosis", "relapsing polychondritis"]'
"http://pubdictionaries.org/mapping/term_to_id?
dictionaries=orphanet&output_format=simple&threshold=
0.6&top_n=0"
```

Exercise (Step 4)

4) To write a configuration in JSON while testing with LODQA

```
{
	"description":"It targets Orphanet.",
	"parser_url":"http://bionlp.dbcls.jp/enju",
	"endpoint_url":"http://cu.orphanet.bio2rdf.org/sparql",
	"ignore_predicates":[],
	"sortal_predicates":[
		"http://www.w3.org/1999/02/22-rdf-syntax-ns#type",
		"http://www.w3.org/2000/01/rdf-schema#subClassOf"
	],
	"dictionary_url":"http://pubdictionaries.org/mapping/...",
	"max_hop":3,
}
```

Exercise (Step 4)

- Configuration example
  - ✔ QALD-Biomed dataset

```
{
    "description":"It targets DrugBank, SIDER, and Diseasome.",
    "parser_url":"http://bionlp.dbcls.jp/enju",
    "endpoint_url":"http://rdf.pubannotation.org/sparql",
    "ignore_predicates":[],
    "sortal_predicates":[
        "http://www.w3.org/1999/02/22-rdf-syntax-ns#type",
        "http://www.w3.org/2000/01/rdf-schema#subClassOf"
    ],
    "dictionary_url":"http://pubdictionaries.org:80/dictionaries/...",
    "max_hop":3,
    "queries":[
        "what genes are associated with alzheimer disease?",
        "what side effects are associated with streptomycin?"
    ]
}
```

Exercise (Step 5)

5) To put the JSON file somewhere visible from the net, and specify it

Exercise (further improvement)

- Refine the dictionary
  - ✔ Use multiple dictionaries.

- Specify
  - ✔ Ignore_predicates
  - ✔ Sortal_predicates

Wrap-up

- LODQA is each to configure
  - ✔ Preparation of the lexical mapping is the biggest trouble,
    - ➔ it affects the performance significantly.
    - ➔ Have your entities typed and connected to each other!
- Feel free to setup LODQA for your EP
  - ✔ On your server
  - ✔ If you want your configuration to appear on the homepage of LODQA, send me the configuration.
    - ➔ Interface for registration will be implemented soon.