

SPARQL Query Examples

2015/07/24

Molecular Profiling Research Center for Drug Discovery (MolProf), AIST

1. fRNAdb

As an example, a SPARQL search for fRNAdb shall be executed by following search conditions.

<Information to obtain from the SPARQL results>

- fRNAdb ID
- its description
- its sequence

<Search Condition>

- Entries whose sequence lengths are from 50 bp to 100 bp.
- Entries whose species are "homo sapiens" (with case ignored)
- Entries whose descriptions contain "miRNA" (with case ignored).
- Max number of SPARQL hits is 25.

A SPARQL query that meet above conditions is as follows:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX aistlsarc: <http://www.molprof.jp/ontologies/aistlsarc.owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
SELECT ?id ?description ?sequence
WHERE {
  ?s rdfs:label ?id .
  ?s aistlsarc:description ?description .
  ?s aistlsarc:sequence ?sequence .
  ?s aistlsarc:length ?len .
  ?s aistlsarc:taxon ?taxon .
  FILTER (?len <= 100 && ?len >= 50) .
  FILTER regex(str(?taxon), "Homo sapiens", "i") .
  FILTER regex(str(?description), "miRNA", "i") .
} LIMIT 25
```

2. SEVENS

As an example, a SPARQL search for SEVENS shall be executed by following search conditions.

<Information to obtain from the SPARQL results>

- SEVENS ID
- its family name
- its sequence

<Search Condition>

- Entries whose sequence lengths are from 300 aa to 600 aa.
- Entries whose species are "homo sapiens" (with case ignored)
- Entries whose family names contain "adrenergic" (with case ignored).
- Max number of SPARQL hits is 20.

A SPARQL query that meet above conditions is as follows:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX aistlsarc: <http://www.molprof.jp/ontologies/aistlsarc.owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
SELECT ?id ?family ?sequence
WHERE {
  ?s rdfs:label ?id .
  ?s aistlsarc:sequence ?sequence .
  ?s aistlsarc:length ?len .
  ?s aistlsarc:family ?family .
  ?s aistlsarc:taxon ?taxon .
  FILTER (?len <= 600 && ?len >= 300) .
  FILTER regex(str(?taxon), "Homo sapiens", "i") .
  FILTER regex(str(?family), "adrenergic", "i") .
} LIMIT 20
```

3. UNIPROT

As an example, a SPARQL search for UNIPROT shall be executed by following search conditions.

<Information to obtain from the SPARQL results>

- UNIPROT ID
- its description
- its sequence

<Search Condition>

- Entries whose descriptions contain “DNA helicase” (with case ignored).
- Max number of SPARQL hits is 30.

A SPARQL query that meet above conditions is as follows:

```
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX aistlsarc:<http://www.molprof.jp/ontologies/aistlsarc.owl#>
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX core:<http://purl.uniprot.org/core/>
SELECT ?id ?description ?sequence
WHERE {
  ?s core:mnemonic ?id .
  ?s core:sequence ?seq .
  FILTER regex(str(?seq), str(substr(str(?s), strlen(str(?s))-6, 6))) .
  ?seq rdf:value ?sequence .
  ?s core:annotation ?anno .
  ?anno a core:Function_Annotation ;
        rdfs:comment ?description .
  FILTER regex(str(?description), "DNA helicase", "i") .
} LIMIT 30
```

4. PDB

As an example, a SPARQL search for PDB shall be executed by following search conditions.

<Information to obtain from the SPARQL results>

- PDB code + chain identifier
- its description
- its sequence

<Search Condition>

- Entries whose sequence lengths are from 100 aa to 400 aa.
- Entries whose species are "Saccharomyces cerevisiae" (with case ignored)
- Entries whose resolutions are ≤ 3.0 .
- Max number of SPARQL hits is 100.

A SPARQL query that meet above conditions is as follows:

```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX PDBx:<http://pdbml.pdb.org/schema/pdbx-v40.xsd>
PREFIX PDBo:<http://rdf.wwpdb.org/schema/pdbx-v40.owl#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT (CONCAT(?id, ?chain) AS ?code) ?desc ?seq
WHERE {
  ?s PDBo:has_atom_sitesCategory ?sitesCa .
  ?sitesCa PDBo:has_atom_sites ?atomSites .
  ?atomSites PDBo:atom_sites.entry_id ?id .
  ?s PDBo:has_entityCategory ?entryCa .
  ?entryCa PDBo:has_entity ?entry .
  ?entry PDBo:entity.pdbx_description ?desc .
  ?entry PDBo:referenced_by_entity_src_gen ?entrySrc .
  ?entrySrc PDBo:entity_src_gen.pdbx_gene_src_scientific_name ?species .
  FILTER regex(?species, "Saccharomyces cerevisiae", "i") .
  ?entry PDBo:referenced_by_entity_poly ?entryPo .
  ?entryPo PDBo:entity_poly.pdbx_seq_one_letter_code_can ?seq .
  FILTER (STRLEN(?seq) < 500 && STRLEN(?seq) > 100) .
  ?entryPo PDBo:entity_poly.pdbx_strand_id ?chain .
  ?s PDBo:has_refine_ls_shellCategory ?resoCa .
  ?resoCa PDBo:has_refine_ls_shell ?reso .
  ?reso PDBo:refine_ls_shell.d_res_high ?highReso .
  FILTER (STRLEN(?highReso) <= 3.0) .
} LIMIT 100
```

5. KEGG Pathway

As an example, a SPARQL search for KEGG Pathway shall be executed by following search conditions.

<Information to obtain from the SPARQL results>

- Pathway name
- its description
- its URL

<Search Condition>

- Entries whose descriptions contain “TCA” (with case ignored).
- Max number of SPARQL hits is 100.

A SPARQL query that meet above conditions is as follows:

```
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX psys:<http://proton.semanticweb.org/protonsys#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX pext:<http://proton.semanticweb.org/protonext#>
SELECT DISTINCT ?pathway ?description ?url
WHERE {
  ?pathway ?p <http://bio2rdf.org/ns/kegg#Pathway> .
  ?pathway rdfs:label ?description .
  FILTER (regex(?description, "tca", "i"))
  ?pathway ?pp ?url .
  FILTER (?pp = <http://bio2rdf.org/ns/bio2rdf#url>)
} LIMIT 100
```