

Part I

Sparql Endpoint : <https://bio2rdf.org/sparql>

1 Graphs research

1.1 Queries for graphs and results

graph all_graphs_included.sparql : 1 results

graph graphs_with_rdf_type.sparql : 38 results

The graphs_with_rdf_type.sparql returns more results (38)

1.2 List of graphs selected

graph 1 - <http://www.openlinksw.com/schemas/virtrdf#>

graph 2 - b3sifp

graph 3 - <http://www.w3.org/2002/07/owl#>

graph 4 - disgenetv3.0

graph 5 - <http://www.openlinksw.com/schemas/oplweb#>

graph 6 - http://bio2rdf.org/hgnc_resource:bio2rdf.dataset.hgnc.R3

graph 7 - http://bio2rdf.org/pharmgkb_resource:bio2rdf.dataset.pharmgkb.R3

graph 8 - http://bio2rdf.org/sgd_resource:bio2rdf.dataset.sgd.R3

graph 9 - http://bio2rdf.org/hp_resource:bio2rdf.dataset.hp.R3

graph 10 - http://bio2rdf.org/eco_resource:bio2rdf.dataset.eco.R3

graph 11 - http://bio2rdf.org/apo_resource:bio2rdf.dataset.apo.R3

graph 12 - http://bio2rdf.org/go_resource:bio2rdf.dataset.go.R3

graph 13 - http://bio2rdf.org/clinicaltrials_resource:bio2rdf.dataset.clinicaltrials.R3

graph 14 - http://bio2rdf.org/lsr_resource:bio2rdf.dataset.lsr.R3

graph 15 - http://bio2rdf.org/affymetrix_resource:bio2rdf.dataset.affymetrix.R3

graph 16 - http://bio2rdf.org/taxonomy_resource:bio2rdf.dataset.taxonomy.R3

graph 17 - http://bio2rdf.org/ctd_resource:bio2rdf.dataset.ctd.R3

graph 18 - http://bio2rdf.org/drugbank_resource:bio2rdf.dataset.drugbank.R3

graph 19 - http://bio2rdf.org/kegg_resource:bio2rdf.dataset.kegg.R3

graph 20 - http://bio2rdf.org/irefindex_resource:bio2rdf.dataset.irefindex.R3

graph 21 - http://bio2rdf.org/homologene_resource:bio2rdf.dataset.homologene.R3

graph 22 - http://bio2rdf.org/mgi_resource:bio2rdf.dataset.mgi.R3

graph 23 - http://bio2rdf.org/goa_resource:bio2rdf.dataset.goa.R3

graph 24 - http://bio2rdf.org/mesh_resource:bio2rdf.dataset.mesh.R3
graph 25 - http://bio2rdf.org/ncbigene_resource:bio2rdf.dataset.ncbigene.R3
graph 26 - http://bio2rdf.org/ndc_resource:bio2rdf.dataset.ndc.R3
graph 27 - http://bio2rdf.org/sider_resource:bio2rdf.dataset.sider.R3
graph 28 - http://bio2rdf.org/omim_resource:bio2rdf.dataset.omim.R3
graph 29 - http://bio2rdf.org/bioportal_resource:bio2rdf.dataset.bioportal.R3
graph 30 - http://bio2rdf.org/wormbase_resource:bio2rdf.dataset.wormbase.R3
graph 31 - <http://localhost:8890/sparql>
graph 32 - <http://bio2rdf.org/sparql>
graph 33 - <http://bioportal.bio2rdf.org/sparql>
graph 34 - <http://sparql.uniprot.org/>
graph 35 - <http://beta.sparql.uniprot.org/>
graph 36 - <https://query.wikidata.org/sparql>
graph 37 - `virtrdf-label`
graph 38 - http://bio2rdf.org/interpro_resource:bio2rdf.dataset.interpro.R3

2 Querying <https://bio2rdf.org/sparql>

2.1 Results by query

Query class.sparql - distinct results : 793

Query property.sparql - distinct results : 1487

Query label_class.sparql - distinct results : 717

Query label_property.sparql - distinct results : 1133

Query subclass.sparql - distinct results : 0

Query subproperty.sparql - distinct results : 8

Query property_domain.sparql - distinct results : 279

Query property_range.sparql - distinct results : 167

Query indi_range_property.sparql - distinct results : 1434

Query indi_domain_property.sparql - distinct results : 8763

Query entity_type.sparql - distinct results : 4603693

Total of unique results found 4618474 on <https://bio2rdf.org/sparql>

2.2 Error Log

- Error on query/graphs_with_a_triple_min.sparql : HTTP Error 504: Gateway Time-out
- Error on query/label_property.sparql graph 7 http://bio2rdf.org/pharmgkb_resource:bio2r : HTTP Error 504: Gateway Time-out
- Error on query/label_property.sparql graph 17 http://bio2rdf.org/ctd_resource:bio2rdf.dat : HTTP Error 504: Gateway Time-out
- Error on graph <http://www.openlinksw.com/schemas/virtrdf#> query/subclass.sparql : EndpointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)'
- Error on graph b3sifp query/subclass.sparql : EndpointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)'
- Error on graph <http://www.w3.org/2002/07/owl#> query/subclass.sparql : EndpointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph `disgenetv3.0 query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph `http://www.openlinksw.com/schemas/oplweb# query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph `http://bio2rdf.org/hgnc_resource:bio2rdf.dataset.hgnc.R3 query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph `http://bio2rdf.org/pharmgkb_resource:bio2rdf.dataset.pharmgkb.R3 query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph `http://bio2rdf.org/sgd_resource:bio2rdf.dataset.sgd.R3 query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph `http://bio2rdf.org/hp_resource:bio2rdf.dataset.hp.R3 query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph `http://bio2rdf.org/eco_resource:bio2rdf.dataset.eco.R3 query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph http://bio2rdf.org/apo_resource:bio2rdf.dataset.apo.R3 query/subclass.sparql : EndPointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph http://bio2rdf.org/go_resource:bio2rdf.dataset.go.R3 query/subclass.sparql : EndPointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph http://bio2rdf.org/clinicaltrials_resource:bio2rdf.dataset.clinicaltrials.R3 query/subclass.sparql : EndPointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph http://bio2rdf.org/lsr_resource:bio2rdf.dataset.lsr.R3 query/subclass.sparql : EndPointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph http://bio2rdf.org/affymetrix_resource:bio2rdf.dataset.affymetrix.R3 query/subclass.sparql : EndPointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph http://bio2rdf.org/taxonomy_resource:bio2rdf.dataset.taxonomy.R3 query/subclass.sparql : EndPointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph http://bio2rdf.org/ctd_resource:bio2rdf.dataset.ctd.R3 query/subclass.sparql : EndPointInternalError: The endpoint returned

the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph http://bio2rdf.org/drugbank_resource:bio2rdf.dataset.drugbank.R3query/subclass.sparql : EndpointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph http://bio2rdf.org/kegg_resource:bio2rdf.dataset.kegg.R3query/subclass.sparql : EndpointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph http://bio2rdf.org/irefindex_resource:bio2rdf.dataset.irefindex.R3query/subclass.sparql : EndpointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph http://bio2rdf.org/homologene_resource:bio2rdf.dataset.homologene.R3query/subclass.sparql : EndpointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph http://bio2rdf.org/mgi_resource:bio2rdf.dataset.mgi.R3query/subclass.sparql : EndpointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph http://bio2rdf.org/goa_resource:bio2rdf.dataset.goa.R3query/subclass.sparql : EndpointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph `http://bio2rdf.org/mesh_resource:bio2rdf.dataset.mesh.R3 query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: `b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)`
- Error on graph `http://bio2rdf.org/ncbigene_resource:bio2rdf.dataset.ncbigene.R3 query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: `b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)`
- Error on graph `http://bio2rdf.org/ndc_resource:bio2rdf.dataset.ndc.R3 query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: `b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)`
- Error on graph `http://bio2rdf.org/sider_resource:bio2rdf.dataset.sider.R3 query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: `b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)`
- Error on graph `http://bio2rdf.org/omim_resource:bio2rdf.dataset.omim.R3 query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: `b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)`
- Error on graph `http://bio2rdf.org/bioportal_resource:bio2rdf.dataset.bioportal.R3 query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: `b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)`
- Error on graph `http://bio2rdf.org/wormbase_resource:bio2rdf.dataset.wormbase.R3 query/subclass.sparql` : `EndPointInternalError`: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph <http://localhost:8890/sparql> query/subclass.sparql : EndPointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph <http://bio2rdf.org/sparql> query/subclass.sparql : EndPointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph <http://bioportal.bio2rdf.org/sparql> query/subclass.sparql : EndPointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph <http://sparql.uniprot.org/> query/subclass.sparql : EndPointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph <http://beta.sparql.uniprot.org/> query/subclass.sparql : EndPointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph <https://query.wikidata.org/sparql> query/subclass.sparql : EndPointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph [virtrdf-label](#) query/subclass.sparql : EndPointInternalError: The endpoint returned the HTTP status code 500

Response: b'Virtuoso 42000 Error The estimated execution time 1201097 (sec) exceeds the limit of 9000 (sec)

- Error on graph `http://bio2rdf.org/interpro_resource:bio2rdf.dataset.interpro.R3`
query/subclass.sparql : `EndPointInternalError`: The endpoint returned
the HTTP status code 500

Response: `b'Virtuoso 42000 Error The estimated execution time 1201097`
(sec) exceeds the limit of 9000 (sec)
- Error on query/subproperty.sparql graph 7 `http://bio2rdf.org/pharmgkb_resource:bio2rdf.c`
: HTTP Error 504: Gateway Time-out
- Error on query/subproperty.sparql graph 17 `http://bio2rdf.org/ctd_resource:bio2rdf.datase`
: HTTP Error 504: Gateway Time-out
- Error on query/indi_range_property.sparql graph 17 `http://bio2rdf.org/ctd_resource:bio2r`
: HTTP Error 504: Gateway Time-out
- Error on graph `http://bio2rdf.org/hgnc_resource:bio2rdf.dataset.hgnc.R3`
query/indi_domain_property.sparql : `EndPointNotFound`: It was not
possible to connect to the given endpoint: check it is correct

Response: `b'<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML//EN">`
`<html> <head> <title>Error HTTP/1`
- Error on query/indi_domain_property.sparql graph 7 `http://bio2rdf.org/pharmgkb_resour`
: HTTP Error 504: Gateway Time-out
- Error on query/indi_domain_property.sparql graph 17 `http://bio2rdf.org/ctd_resource:bio`
: HTTP Error 504: Gateway Time-out
- Error on query/indi_domain_property.sparql graph 25 `http://bio2rdf.org/ncbigene_resour`
: HTTP Error 504: Gateway Time-out
- Number of errors : 48

3 Process before indexation

3.1 JSON results conversion after distinct operation

All string results were converted into valid json results

4 Partition in separated queries

Results volume is superior to 2 millions and has been split into partitions

- Number of parts : 3
- Size of partition : 2000000
- Partition 1 - query/entity_type_1.sparql : 2000000 results
- Partition 2 - query/entity_type_2.sparql : 2000000 results
- Partition 3 - query/entity_type_3.sparql : 603693 results

These partitions are bulked separately and merged into same index

5 Elastic Search Indexation

Index 'class' already exists and will be completed
Storing 793 results in index 'class' took 1.312 secs

Index 'property' already exists and will be completed
Storing 1487 results in index 'property' took 1.354 secs

Index 'label_class' already exists and will be completed
Storing 717 results in index 'label_class' took 0.081 secs

Index 'label_property' already exists and will be completed
Storing 1133 results in index 'label_property' took 0.15 secs

Index 'subproperty' already exists and will be completed
Storing 8 results in index 'subproperty' took 0.937 secs

Index 'property_domain' already exists and will be completed
Storing 279 results in index 'property_domain' took 0.414 secs

Index 'property_range' already exists and will be completed
Storing 167 results in index 'property_range' took 0.07 secs

Index 'indi_range_property' already exists and will be completed
Storing 1434 results in index 'indi_range_property' took 0.814 secs

Index 'indi_domain_property' already exists and will be completed
Storing 8763 results in index 'indi_domain_property' took 1.362 secs

Query subpartition 1 for entity_type
Total of results to store : 2000000
Partition size : 200000
Number of subpartitions : 10
Timesleep : 5 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 44.467 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 37.025 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 21.498 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 50.566 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 93.899 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 45.273 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 51.201 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 31.071 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 32.37 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 58.719 secs

Query subpartition 2 for entity_type
Total of results to store : 2000000
Partition size : 200000
Number of subpartitions : 10
Timesleep : 5 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 43.356 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 63.292 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 35.282 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 43.259 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 58.248 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 39.16 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 43.397 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 24.49 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 52.984 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 40.292 secs

Query subpartition 3 for entity_type
Total of results to store : 603693
Partition size : 200000
Number of subpartitions : 4
Timesleep : 2 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 30.867 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 33.478 secs

Index 'entity_type' already exists and will be completed
Storing 200000 results in index 'entity_type' took 57.413 secs

Index 'entity_type' already exists and will be completed
Storing 3693 results in index 'entity_type' took 0.366 secs

6 Getting results and indexing times

Getting results for <https://bio2rdf.org/sparql> took 49 min (2961.698 secs)

Indexation for <https://bio2rdf.org/sparql> took 20 min (1253.919 secs)