

NAMING CONVENTION FOR PROFILER

OVERVIEW

The output files corresponding to overview go in “overview” folder

1. Overview.nodes.tsv

This file contains overview of the graph like

- a. Number of nodes in the graph.
- b. Number of properties in the graph.
- c. Number of edges in the graph.
- d. Number of classes in the graph.

2. Overview.degree.tsv

This file contains information about the in/out degree of the graph:

- a. Average in-degree of the graph
- b. Minimum in-degree of the graph
- c. Maximum in-degree of the graph
- d. Average out-degree of the graph
- e. Minimum out-degree of the graph
- f. Maximum out-degree of the graph

3. Overview.class.tsv

This file contains information about top K classes of the graph based on number of instances of the class.

4. Overview.property.[datatype].tsv

There are total 13 files for each of the data type and contains information about top K properties of each datatype based on number of statements.

CLASS OVERVIEW-

The output files corresponding to class overview go in “class_overview” folder

Following set of files is present for each of the top K classes found earlier

1. Class_overview.[class_name].overview.tsv

Contains basic information about the class like name, description, aliases, subclasses, superclasses, number of subclasses, number of superclasses etc.

2. Class_overview.[class_name].examples.tsv

Contains top three instances of the class based on PageRank

3. Class_overview.[class_name].outgoing_properties.tsv

Contains details about top K outgoing properties from the instances of the class.

4. Class_overview.[class_name].incoming_properties.tsv

Contains details about top K incoming properties to the instances of the class.

PROPERTY OVERVIEW-

The output files corresponding to class property overview go in “property_overview” folder

1. Property_overview.top.tsv

Contains top K properties based on number of statements and not divided based on datatype

This file is present for each of the top K properties found above.

2. Property_overview.[property_name].overview.tsv

Contains details about the property like label, alias etc.

This is done for each of the top K properties of Datatype: Quantity

3. Property_overview.quantity.[property_name].units_distribution.tsv

It finds the list of units and number of statements for each of the units. Its sorted in descending order based on number of statements.

4. Property_overview.quantity.[property_name].[unit].value_distribution.tsv

For a property it finds the distribution of values for most prevalent unit. Binned at an interval of ten units.

This is done for each of the top K properties of Datatype: time

5. Property_overview.time.[property_name].year_distribution.tsv

For a property it finds the distribution of dates binned per year.

This is done for each of the top K properties of Datatype: geo_coordinates

6. Property_overview.geo_coord.[property_name].distribution.tsv

For a property it finds the distribution of coordinates. Which can then be plotted on a global map.

7. Property_overview.geo_shape.random.tsv:

This file contains the random sample of the nodes with data-type: geo_shape.