

QUANT User Manual

Description: Question Answering Curator (QUANT) is a tool that helps to define a SPARQL of a question so that it can be run against a given SPARQL endpoint and produce a semantically correct result set. Furthermore, QUANT will check whether each question's attribute has a correct value. If the value is incorrect, QUANT will provide a suggestion for the correct value and let the user do the correction.

Data: Dataset that contains questions and their attributes. The attributes are SPARQL, Answer Type, Out of Scope, Aggregation, Onlydbo, Hybrid, Multilingual Keyword List, Multilingual Question List.

Output: a New dataset that contains a correction result. The new dataset will be based on DBpedia version 2016-10 and written in JSON format.

How to use QUANT?

1. Access QUANT at this following address: <http://131.234.29.28:8080/webqald>
2. Enter username and password
3. Go to **Dataset** menu to find a list of questions that will be curated. Go to the details page of a question. There will be three options available i.e. doing correction, give a confirmation that the question does not need any changes, and remove the question. The following are explanations for all options:
 - a) Doing correction: change the value of any question's attribute so that the value becomes correct regarding particular condition. This task can be done by clicking **Start Correction** button
 - b) Give a confirmation that the question does not need any changes: state that all attributes of the question already have correct value so it does not need to do a correction or remove the question. This task can be done by clicking **No Changes Needed** button.
 - c) Remove the question: delete the question from the dataset as its SPARQL is not able to return results from current SPARQL endpoint and there is no way to modify the SPARQL to let it able to return results. This task can be done by clicking **Remove Question** button
4. In doing the correction, check the value of these following attributes: **SPARQL, Answer Type, Out of Scope, Aggregation, Onlydbo, Hybrid, Multilingual Keyword, Multilingual Question** and provide the correct value for them if their existing values are incorrect. QUANT will provide a suggestion for an attribute that either does not have value or the

existing value is incorrect. Multilingual Keyword could have additional treatment since some of the questions might not have keywords value so QUANT will provide the suggestion of generated keywords first before giving the translations suggestion. SPARQL value is used to determine the value of other attributes. Therefore, start to do checking on SPARQL value base on the value of Answer from Current Endpoint. After finished the correction, click **Done** button so that the correction will be recorded. During the correction process, annotator can decide whether the question should be removed (**Remove Question** button) or cancel the correction (**Cancel** button).

5. If there is no value at **Answer from Current Endpoint** field, check whether QUANT provides **View SPARQL suggestion** button (below **SPARQL** text field). If it is the case, check the suggestion and take it if it is considered as a correct value. Otherwise, update the SPARQL value by yourself by doing correction directly in **SPARQL** text field. A valid SPARQL must contain one returning/projection variable. If there is a changing of SPARQL value, QUANT will update value at **Answer from File** and **Answer from Current Endpoint** text field and will adjust suggestion value for other attributes.

How to Get Correct Values?

1. These are rules to get the correct value for each attribute if SPARQL is already correct in the sense it returns results from a current endpoint.

a) Answer Type

Answer Type is used to describe a data type of Answer from Current Endpoint's value. There are 5 possibilities value of it i.e. Boolean, Date, Number, Resource, and String. The following are samples value of each data type:

- Boolean : true, false
- Date : 1957-03-25
- Number : 17, 1.0363E11
- Resource : http://dbpedia.org/resource/New_Jersey
- String : England

If Answer from Current Endpoint attribute does not have value, see point 5.

b) Aggregation

Aggregation describes whether SPARQL attribute's value contains one or more aggregation function such as COUNT, SUM, AVG, MIN, MAX, SAMPLE, GROUP_CONCAT, GROUPCONCAT, VECTOR_AGG, and COUNT DISTINCT. If it is the case, aggregation value will be true. Otherwise, it will be false.

c) Hybrid

Hybrid describes whether it is required to do searching both Linked Data and textual data for an answer. A hybrid question will have a SPARQL that contains phrase text:query in it. If it is the case, its value will be true. Otherwise, it will be false.

d) Onlydbo

Onlydbo describes whether SPARQL attribute's value does not contain some URI outside of DBpedia, for instance, <http://yago-knowledge.org/resource/yagoLegalActor> or <http://xmlns.com/foaf/0.1/nick>. If it is the case, Onlydbo's value will be set as true. Otherwise, it will be false.

e) Out of Scope

Out of Scope describes whether SPARQL query is not able to retrieve results from current SPARQL endpoint or the returned results are semantically incorrect. If there is a result (some result) from the current endpoint and the result is semantically correct, Out of Scope value must be false. It will be true if there is no result exists or the result is not correct semantically (for example question "is Pluto a planet?").

f) Multilingual Keyword

This attribute should show keywords of the question in 11 different languages. English is the main language. If there are no keywords exist, annotator should provide the keywords. After that, annotator should provide the translations of the keywords. QUANT will provide generated keywords suggestion for a question that does not have keywords. Annotator should make sure whether the suggested keywords are accurate in term of fulfilling these following rules:

- Separated by comma
- The verbs are an infinitive form or in their tense
- Singular/plural of nouns
- Question words

QUANT will also provide suggested keywords translations in 10 languages. For those questions that already have keywords, QUANT will provide keywords translations suggestion only.

g) Multilingual Question

Like the keywords, a question also should have translations in 10 languages. Some questions might have no translations at all and others exist with some translations (still not complete in 10 languages). For both of cases, annotator needs to complete the translations by considering QUANT's question translations suggestion.

2. Curation result and the log of activities can be seen on **User Activities** menu. This page also completed with a menu to print the result of the curation in JSON format.