

QALD-1 Open Challenge

DOCUMENT VERSION: March 28, 2011

The QALD-1 open challenge is part of the workshop Question Answering Over Linked Data (QALD-1) at the Extended Semantic Web Conference 2011 in Heraklion, Greece. The challenge is aimed at any kind of question answering system that mediates between a user, expressing his or her information need in natural language, and semantic data. The goal is to evaluate and compare participating systems. To this end, two datasets are provided – DBpedia 3.6 and an RDF export of MusicBrainz – together with a set of 50 test questions each. Beforehand, a training set of 50 natural language questions per dataset is made publicly available, annotated with corresponding SPARQL queries and correct answers. All relevant information for accessing the datasets and questions as well as for participating in the challenge are given in this document.

1 Motivation and Goal	2
2 Relevant information in a nutshell	3
2.1 Open Challenge	3
2.2 Contact	3
3 Datasets: Download and SPARQL end point	4
4 DBpedia 3.6	4
5 MusicBrainz	5
6 Evaluation	7
6.1 Training questions	8
6.2 Submitting results	9
6.3 Evaluation	10
6.4 Test phase	10
6.5 Resources	11
7 Contact and trouble shooting	11

1 Motivation and Goal

In contrast to the Information Retrieval community, where evaluation using standardized techniques, such as those used for the annual TREC competitions, has been common for decades, the Semantic Web community (characterized by the lack of uniformity and the diversity of semantic technologies) has not yet adopted standard evaluation benchmarks for semantic question answering systems that focus on the ability to solve open-ended real life problems over real-world datasets. Aiming to advance in this issue and develop the datasets needed to formally judge the quality of ontology-based question answering approaches at a large scale, two independent sets of tests has been created as part of this challenge . The tests consist of a variety of user questions of different complexity, designed to represent questions that real end users would ask.

As some of the questions are indeed very challenging, feel free to work only on a subset, e.g. leaving out temporal reasoning, Yago hierarchies, etc. We strongly encourage you to report on results even if precision and recall are relatively low, as the goal of the challenge is not to handle all complexities present in the queries, but rather to get a picture of the strengths, capabilities and current shortcomings of question answering systems, as well as to gain insight into how we can develop question answering approaches that deal with the fact that i) the amount of RDF data available on the Web is huge, ii) that this data is distributed and iii) that it is heterogeneous with respect to the vocabularies and schemas used.

And although the competition is tailored towards question answering systems based on natural language, we also encourage other relevant systems and methods, such as dynamic ontology matching, schema integration, word sense disambiguation, fusion and rank technologies that can benefit from QALD-1 evaluation datasets, to report their results (see the call for papers).

2 Relevant information in a nutshell

Coordinators: Christina Unger, Philipp Cimiano, Vanessa Lopez, Enrico Motta

Workshop Website: <http://www.sc.cit-ec.uni-bielefeld.de/qald-1>

2.1 Open Challenge

Datasets:

- DBpedia 3.6
<http://downloads.dbpedia.org/3.6/en>
<http://downloads.dbpedia.org/3.6/links>
- MusicBrainz
<http://greententacle.techfak.uni-bielefeld.de/~cunger/qald1/>
▷ [musicbrainz.tgz](#) (126.4 MB) contains the RDF data, the ontology, and graph files for bulk import

SPARQL end point (for both datasets):

<http://greententacle.techfak.uni-bielefeld.de:5171/sparql>

Training questions:

<http://greententacle.techfak.uni-bielefeld.de/~cunger/qald1/>

- [dbpedia-train.xml](#)
- [musicbrainz-train.xml](#)

Test questions:

<http://greententacle.techfak.uni-bielefeld.de/~cunger/qald1/>

- [dbpedia-test-questions.xml](#)
- [musicbrainz-test-questions.xml](#)

Submission of results and evaluation is done by means of an online form:

- <http://greententacle.techfak.uni-bielefeld.de/~cunger/qald1/evaluation/>

Results for the test questions can be uploaded from April 1 to April 10.

2.2 Contact

Updates on the open challenge will be published on the QALD-1 mailing list:

<https://lists.techfak.uni-bielefeld.de/cit-ec/mailman/listinfo/qald-1>

In case of question, problems, and comments, please contact Christina Unger (cunger@cit-ec.uni-bielefeld.de).

3 Datasets: Download and SPARQL end point

The open challenge uses two RDF datasets: DBpedia 3.6 (<http://dbpedia.org>) and MusicBrainz (<http://musicbrainz.org>). In order to work with the datasets, you can either download them or use the provided SPARQL end point.

The DBpedia 3.6 dataset comprises all files contained in the following two directories:

- <http://downloads.dbpedia.org/3.6/en>
- <http://downloads.dbpedia.org/3.6/links>

The MusicBrainz dataset is an RDF export of all classes and the most important properties of the MusicBrainz database. It builds upon the MusicBrainz ontology and can be downloaded here:

- <http://greententacle.techfak.uni-bielefeld.de/~cunger/qald1/musicbrainz.tgz>

This package contains the RDF data¹, the ontology (`musicbrainz.owl`), and graph files for a bulk import. For more information on bulk importing the files in Virtuoso, please see:

<http://ods.openlinksw.com/wiki/main/Main/VirtBulkRDFLoader>

The SPARQL end point for both datasets is the following:

<http://greententacle.techfak.uni-bielefeld.de:5171/sparql>

Evaluation will take place with respect to this SPARQL end point (and not the official DBpedia end point), in order to ensure invariable and therefore comparable results.

In the following, the datasets are described in more detail.

4 DBpedia 3.6

DBpedia is a community effort to extract structured information from Wikipedia and to make this information available. The RDF dataset provided for the challenge is the official DBpedia 3.6 dataset for English, including links, most importantly to YAGO categories.

For more information on the DBpedia dataset, please refer to:

<http://wiki.dbpedia.org>

¹But only a subset of all track information, due to performance problems.

For detailed information on the YAGO class hierarchy, please see:

<http://www.mpi-inf.mpg.de/yago-naga/yago/>

Namespaces relevant for querying DBpedia 3.6 are the following:

```
rdfs: <http://www.w3.org/2000/01/rdf-schema#>
rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
onto: <http://dbpedia.org/ontology/>
res: <http://dbpedia.org/resource/>
prop: <http://dbpedia.org/property/>
yago: <http://dbpedia.org/class/yago/>
xsd: <http://www.w3.org/2001/XMLSchema#>
foaf: <http://xmlns.com/foaf/0.1/>
```

5 MusicBrainz

MusicBrainz is a collaborative effort to create an open content music database. The RDF dataset provided for the challenge contains all of MusicBrainz' artists, albums, and tracks, as well as a subset of important relations between them. In the following, the classes and properties will be briefly explained. The ontology (`musicbrainz.owl`) is also contained in the download package.

The following namespaces are relevant for the training and test questions:

```
mm: <http://musicbrainz.org/mm/mm-2.1#>
ar: <http://musicbrainz.org/ar/ar-1.0#>
rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
rdfs: <http://www.w3.org/2000/01/rdf-schema#>
owl: <http://www.w3.org/2002/07/owl#>
dc: <http://purl.org/dc/elements/1.1/>
ct: <http://www.sc.cit-ec.uni-bielefeld.de/ontologies/musicbrainz#>
xsd: <http://www.w3.org/2001/XMLSchema#>
```

MusicBrainz comprises three classes:

- *Artist* (`mm:Artist`): a person or a group
- *Album* (`mm:Album`): albums, CD singles, interviews, audio books, etc.
- *Track* (`mm:Track`): a track on a CD, single, compilation, etc.

The property `dc:title` maps an artist, album, or track to a string representing its title, and the property `dc:creator` maps an album or track to an artist.

The property `mm:artistType` specifies an artist to either be of type `mm:TypePerson` or of type `mm:TypeGroup`.

The relations `mm:beginDate` and `mm:endDate` relate an artist with a date. In the case of a person, these are the day of birth or death, respectively. In the case of groups, the dates represent the date when a group was founded or broke up, respectively. Dates are of form `"1970-04-10"^^xsd:date`, i.e. typed literals, unless the day and/or month are unknown – in that case, they are untyped literals of form `"1970-00-00"`.

They date, when an artist joined or left a band (or when an artist married another artist and the like) are also modelled by means of `mm:beginDate` and `mm:endDate`. To this end, the artist is not directly related to a group but to a blank node that is then related to a group, to a begin date and an end date. This blank node represents an instance of this group with the artist as its single member. All of the following properties between artists are modelled this way:

- `ar:memberOfBand`
- `ar:collaboratedWith`
- `ar:marriedTo`
- `ar:involvedWith`
- `ar:isParent`
- `ar:isSibling`

For example, a SPARQL query asking whether John Lennon was a member of the Beatles until the 10th of April 1970 would look as follows (leaving out prefixes):

```
ASK WHERE {
  ?artist      dc:title      "John Lennon" .
  ?artist      ar:memberOfBand ?bandinstance .
  ?bandinstance ar:toArtist   ?band .
  ?band        dc:title      "The Beatles" .
  ?bandinstance mm:endDate    "1970-04-10"^^xsd:date .
}
```

The property `ct:tag` relates an album to a user assigned tag, i.e. a string, for example genre specifications like Jazz and Rock.

The following properties provide information on albums:

- `mm:releaseStatus` maps albums to one of the following:
 - `mm:StatusOfficial`
 - `mm:StatusBootleg`
 - `mm:StatusPromotion`
- `mm:releaseType` maps albums to one of the following:
 - `mm:TypeAlbum`
 - `mm:TypeSingle`

- `mm:TypeEP`
- `mm:TypeCompilation`
- `mm:TypeSoundtrack`
- `mm:TypeSpokenword`
- `mm:TypeInterview`
- `mm:TypeAudiobook`
- `mm:TypeLive`
- `mm:TypeRemix`
- `mm:TypeOther`

The following properties relate albums and artists:

- `ar:performer`
- `ar:instrument`
- `ar:vocal`
- `ar:performingOrchestra`
- `ar:conductor`
- `ar:composer`
- `ar:lyricist`
- `ar:producer`

The property `mm:trackList` relates an album to a numbered list of tracks (e.g., `<rdf:Seq rdf:li mm:track/fdae8b76-f5b8-4458-bf32-81885000980b>` and so on).

The property `mm:duration` relates a track to its duration in seconds (with three decimals precision).

6 Evaluation

The task of the challenge is to extract correct answers for natural language questions from a given RDF repository. Participating systems will be evaluated with respect to precision and recall. Moreover, participants are encouraged to report performance, e.g. the average time their system takes to answer a query, in the paper.

6.1 Training questions

In order to get acquainted with the datasets and possible questions, a set of 50 training questions for each dataset can be downloaded at the following locations: <http://greententacle.techfak.uni-bielefeld.de/~cunger/qald1/>

- DBpedia
 - ▷ [dbpedia-train.xml](#) (HTML encoding)
 - ▷ [dbpedia-train-CDATA.xml](#) (CDATA encoding)
- MusicBrainz
 - ▷ [musicbrainz-train.xml](#) (HTML encoding)
 - ▷ [musicbrainz-train-CDATA.xml](#) (CDATA encoding)

These training questions are annotated with corresponding SPARQL queries as well as answers retrieved from the provided SPARQL end point. Annotations are provided in an XML format. The overall document is enclosed by a tag that specifies an ID for the dataset (i.e. DBpedia or MusicBrainz).

```
<dataset id="dbpedia-train">
  <question id="1">...</question>
  ...
  <question id="50">...</question>
</dataset>
```

Each of the questions specifies an ID for the question (don't worry about them not being ordered), the natural language string of the question, a corresponding SPARQL query, as well as the answers this query returns. For example:

```
<question id="123">
  <string>Which caves have more than 100 entrances?</string>
  <query>
    PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
    PREFIX onto: <http://dbpedia.org/ontology/>
    PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
    SELECT ?uri ?string
    WHERE {
      ?uri rdf:type onto:Cave .
      ?uri onto:numberOfEntrances ?entrance .
      FILTER (?entrance > 100) .
      OPTIONAL { ?uri rdfs:label ?string . }
      FILTER (lang(?string) = "en") }
  </query>
  <answers>
    <answer>
      <uri>http://dbpedia.org/resource/Kanheri_Caves</uri>
      <string>Kanheri Caves</string>
    </answer>
    <answer>
      <uri>http://dbpedia.org/resource/Ox_Bel_Ha_Cave_System</uri>
      <string>Ox Bel Ha Cave System</string>
```



```
</answer>
</answers>
</question>
```

The answers will be either a literal (a boolean, date, number, or string), or a list of resources, for which both the URI as well as an English label or name (if it exists) is specified.

Please note that dates are enclosed with the tag `<date>` and are provided in the format YYYY-MM-DD, for example:

```
<question id="41">
  <string>When was Capcom founded?</string>
  <query>
    PREFIX res: <http://dbpedia.org/resource/>
    PREFIX prop: <http://dbpedia.org/property/>
    PREFIX onto: <http://dbpedia.org/ontology/>
    SELECT ?date
    WHERE {
      res:Capcom prop:foundation ?date .
    }
  </query>
  <answers>
    <answer>
      <date>1983-06-11</date>
    </answer>
  </answers>
</question>
```

This format is required when you submit results containing a date as answer.

6.2 Submitting results

All submissions are required to comply to the XML format specified above. For all questions, the dataset ID and question IDs are obligatory. Beyond that, you are free to specify either a SPARQL query or the answers (or both), depending on which of them your system returns. You are also allowed to change the query (insert quotes, reformulate, extract and use only keywords, use some structured input format, and the like). If you do so, please document these changes, i.e. replace the provided question string by the input you used.

Submissions will be accepted and evaluated by means of the following online form:

<http://greententacle.techfak.uni-bielefeld.de/~cunger/qald1/evaluation/>

6.3 Evaluation

For each of the questions, your specified answers or the answers your specified SPARQL query retrieves will be compared to the answers provided by the gold standard XML document. For each resource that answers a query, you are free to either provide their URI, their English label or name (if it exists), or both. All options count as correct answers, as long as the answer list contains all and only the correct resources.

The evaluation tool computes precision, recall and F-measure for every question.

$$\begin{aligned}\text{Recall} &= \frac{\text{number of correct system answers}}{\text{number of gold standard answers}} \\ \text{Precision} &= \frac{\text{number of correct system answers}}{\text{number of system answers}} \\ \text{F-measure} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\end{aligned}$$

The tool then also computes the overall precision, recall and F-measure along the standard definitions:

$$\begin{aligned}\text{Recall} &= \frac{\text{number of correctly answered questions}}{\text{number of questions}} \\ \text{Precision} &= \frac{\text{number of correctly answered questions}}{\text{number of answered questions}} \\ \text{F-measure} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\end{aligned}$$

All these results are printed in a simple HTML output; additionally you get a list of all question that your tool failed to capture correctly.

During training phase, you are allowed to submit results as often as you wish.

6.4 Test phase

During test phase, 50 different questions for each dataset without annotations are provided at the following locations:

<http://greententacle.techfak.uni-bielefeld.de/~cunger/qald1/>

- [dbpedia-test-questions.xml](#)
- [musicbrainz-test-questions.xml](#)

Results can be submitted from April 1 to April 10, via the same online form as used during training phase (note the drop down box that will allow you to specify *test* instead of *training*):

<http://greententacle.techfak.uni-bielefeld.de/~cunger/qald1/evaluation/>

The only difference is that evaluation results are not displayed. You can upload results as often as you like (e.g. trying different configurations of your system); in this case the result file with best precision and recall will count.

6.5 Resources

You are free to use all resources (e.g. WordNet, GeoNames, dictionary tools, and so on).

7 Contact and trouble shooting

If you have any questions or comments, including worries about the training and test questions, trouble with the datasets, the SPARQL end point, or the online submission and evaluation form, please contact Christina Unger: cunger@cit-ec.uni-bielefeld.de.