Exciting as the things GraphDB allows you to do (explore heterogenous datasets, build relationships between facts, uncover meaning inside unstructured data, infer new knowledge, to mention just a few), they all start with, to put it mildly, the not so inspiring task of cleaning your data and further transforming it into RDF.

In practice, before the leaps of data-driven insights and actions come the heaps of inconsistent, unfiltered and heterogenous data that need to be cleaned up. For the data worker having to deal with these messy data is not unlike the fifth labor of Hercules where the hero gets the dirty job of cleaning the Augean Stables.

Saving Time and Effort with GraphDB's OntoRefine

With plenty of tools for cleaning and conversion of data, the question of leveraging legacy data is not so much how to get these data transformed into interoperable and easy to query and integrate data pieces (read RDF – the so-called backbone of the Semantic Web) but rather about how to do this with maximum productivity and minimum wasted effort.

And this is where OntoRefine comes into play.

OntoRefine is a new addition to GraphDB that allows you to do many ETL (extract, transform and load) tasks over tabular data through an intuitive user interface. Based on the open source tool for working with messy data – OpenRefine (formerly called Google Refine), and embedded in GraphDB, OntoRefine makes the process of filtering and editing inconsistent data easy and frictionless.

To get back to the Augean Stables parallel, think of OntoRefine as the witty little tool of the brave data hero tasked with the dirty job of data cleanup and transformation.

Before OntoRefine, to turn tabular into interlinked graph data, data had to be loaded in a tool, cleaned manually, further exported and then imported into another tool as to be transformed into RDF. Finally, after yet another import and export, the RDF dataset had to be loaded into GraphDB. With OntoRefine these processes can happen within GraphDB.

Thus cleaning up and transforming a non-RDF dataset is a fast and easy process, leaving

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

more time for the things that really matter: running queries to discover interesting relationships within data, integrating data – in short, enjoying the full power of working with data as a graph.

Key to what OntoRefine does is the heavy lifting of removing inconsistencies, filtering data simultaneously, converting them into RDF and then importing the dataset into the repository. OntoRefine can be used for converting tabular data into RDF and importing it into a GraphDB repository, using simple SPARQL queries and a virtual endpoint. The supported formats include various line-based files, TSV, CSV, *SV, XLS, XLSX, JSON, XML, RDF as XML, and Google sheet.

From the vantage point of understanding the power of working with data as a graph, OntoRefine is a tiny yet important step toward thinking outside the table.

Quick Facts About OntoRefine

- Based on OpenRefine.
- Embedded in GraphDB's.
- Transforms data using SPIN functions.
- Allows cleaning up and transforming data without leaving the GraphDB Workbench.
- Supports the following formats: line-based files, TSV, CSV, *SV, XLS, XLSX, JSON, XML, RDF as XML, Google sheet.

Get, Load, Clean, Import and Enjoy!

To clean up and transform non-RDF data into RDF using OntoRefine, you need to pick a dataset, load it and process it, and then upload it to GraphDB. In the video below you can go through the details of the data cleanup and transformation process. The dataset selected and transformed is from data.amsterdam.nl and contains records of restaurants and cafes in and around Amsterdam, and was available as a CSV file.

Watch the entire video to learn:

- How to create an empty repository and connect to it;

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

- How to import a dataset, preview data and specify various parameters;
- How to create a project and start cleaning data;
- How to edit simultaneously cells containing a particular entry;
- How to apply filters by selecting a subset of possible values and how to edit all entries in a column;
- How to use a SPARQL Construct query to shape our data in a specified way.

To dive even deeper into the technical details behind OntoRefine, check: OntoRefine – overview and features.

More Business Value with Clean and RDF-ized Data

Fast and frictionless experience when cleaning up and RDF-izing within GraphDB means a smoother data processing workflow and above all saving time and effort for focusing on data modeling and analysis. With OntoRefine embedded in the latest version of GraphDB – GraphDB 8, cleaning and transforming tabular data are brought together in one place to let those working with data tap into the full potential of handling data as a graph.

See for yourself how easy and smooth the processes of data cleanup and transformation into RDF with OntoRefine are.