Coding Journal

Curran Kelleher

Abstract

A journal of programming ideas from December 2010.

1 ASH as a universal platform unifying data and interactive content

12/26/2010 ASH session states are RDF graphs. Therefore they can embody anything representable in RDF.

The annotation of a CSV file into UDC could be a sessioned process. At the end, there would be a URL for the end state of the import process, which would contain all the information necessary to interpret the original CSV file using UDC tools.

Each ontology class could be subclassed by an ASH resource. Then for example, "CSVFileReference" could be subclassed with "CSVFileReference-GASHObject", which would be an interactive visual representation of the "CSVFile" class. In this case, the RDF representing the actual CSV reference would be implied by the existence of the ASH visual representation. This would mean that any RDF knowledge base created visually using ASH resources can be queried using SPARQL like a database.

2 A Serverless Universal Data Cube

12/28/10

Queryable UDC servers can be replaced with documentation of how existing (posted on the internet) files can be interpreted withing the UDC framework:

- URL: www.example.com/exampledata.csv

- Type: data table
- format: CSV
- dc:Title: Example Data
- dc ..... etc.
- numColumns: 5
- numRows: 50
- columnMeanings:
- column 2: US State Name (dimension using lookup table www.example.com/stateNamesToMe
- column 1: Population (measure)

Such an RDF file can tell a UDC client exactly all the information it needs to be able to download, parse and interpret the original data, without the need for a queryable server as a middleman. Such a server could still exist, and probably in many cases would speed things up, but is fundamentally unnecessary.

The same is true for geometry collections.

A dedicated "admin" program can be written which is a wizard which takes as input of a CSV (and fixed width and other formats) file URL, goes through a bunch of steps, and in the end produces a data descriptor file which a UDC client can use to get at the original data.

A few reasons why Java technology was chosen over HTML5:

- Cross site requests are allowed in Java and disallowed in Javascript
- Java is a rich platform for the Semantic Web (many libraries are available), Javascript is not
- Use of OpenGL from Java is a mature technology, WebGL is not.
- Java can easily create and use an in-memory or on disk database, such Javascript technology is not mature
- I'm familiar with Java, not Javascript
- Java has multithreading

3 Table vs Cube metadata

12/29/10

The table parsing metadata can be separate from the data cube annotation:

Table parsing metadata:

- URL: www.example.com/exampledata.csv
- Type: data table
- format: CSV
- dc:Title: Example Data
- dc ..... etc.
- numColumns: 5
- numRows: 50

Data cube metadata:

- columnMeanings:
- column 2: US State Name (dimension using lookup table www.example.com/stateNamesToMe
- column 1: Population (measure)

One example data source is here: ftp://ftp.bls.gov/pub/special.requests/cew/DOCUMENT/lay