

# D3SPARQL: JavaScript library for visualization of SPARQL results

Toshiaki Katayama<sup>1</sup>

<sup>1</sup>Database Center for Life Science, Research Organization of Information and Systems,  
178-4-4 Wakashiba, Kashiwa-shi, Chiba 277-0871, Japan  
ktym@dbcls.jp

**Abstract.** Semantic Web technologies are being widely applied in life sciences. Major bioinformatics data centers started to provide heterogeneous biomedical datasets in RDF and expose them at SPARQL endpoints. SPARQL query is used to search those endpoints and the results are obtained as a SPARQL Query Results XML Format or a SPARQL Query Results JSON Format, both are essentially tabular structured data. To effectively represent the SPARQL results, appropriate visualization methods are highly demanded. To create and control dynamic graphical representation of data on the Web, the D3.js JavaScript library is getting popularity as a generic framework based on the widely accepted Web standards such as SVG, JavaScript, HTML5 and CSS. A variety of visualization examples implemented with the D3.js library is already available, however, each of them depends on assumed JSON data structure that differs from the JSON structure returned from SPARQL endpoints. Therefore, it is expected to largely reduce development costs of Semantic Web visualization if a JavaScript library is available which transforms SPARQL Query Results JSON Format into JSON data structures consumed by the D3.js. D3SPARQL is developed as a generic JavaScript library to fill this gap. D3SPARQL can be used to query SPARQL endpoints as an AJAX call and provides various callback functions to visualize the obtained results. Biological applications will be shown in this software demo along with our integrated semantic genome database, the TogoGenome application. The D3SPARQL library is freely available at <https://github.com/ktym/d3sparql>.

**Keywords:** Semantic Web, SPARQL, JavaScript, visualization

## Introduction

RDF data in life sciences including UniProt [1], Bio2RDF [2], wwPDB [3], EBI RDF (BioModels, BioSamples, ChEMBL, Expression Atlas, Reactome) [4], NCBI (PubChem) [5], GlycoRDF [6] and INSDC/DDBJ [7] have been published since 2008. Therefore, it is demanded to develop applications to effectively utilize those semantic datasets so that end-users can benefit from integration of heterogeneous biomedical data. RDF data is queried by the SPARQL query language and the results will be returned as a SPARQL Query Results XML Format [8] or a SPARQL Query

## 2 Toshiaki Katayama<sup>1</sup>

Results JSON Format [9]. Both of those result formats contain a set of key-value pairs and essentially are tabular structured data. Usually, the results are simply rendered as a HTML table, however, it is more suitable to be visualized as a graph, tree, chart, geographic map or other graphical representation depending on the nature of obtained data in many cases.

Because SPARQL uses HTTP as its application protocol and the result can be obtained as a JSON object, the JavaScript language naturally fits for handling the data on the Web. There are many JavaScript libraries have been developed to visualize JSON data, among them, D3.js [10] is one of the most powerful library to manipulate, visualize and control data and provides dynamic and interactive graphics on a Web page based on the standard technologies including SVG, JavaScript, HTML5 and CSS.

However, the data structure of the SPARQL Query Results JSON Format is different from the JSON format which existing D3.js implementations expect. The D3.js library comes with several built-in algorithms (`d3.layout`) to calculate graphical properties for typical visualizations. Thus, the purpose of this work, D3SPARQL, is to transform a JSON data obtained from SPARQL endpoint into a data structure accepted by those layout algorithms implemented in the D3.js.

## Results

D3SPARQL is an open source library which can be embedded in any Web page, performs a SPARQL query via AJAX call, transforms the result and visualizes data with the help of D3.js library. For example, the following SPARQL query retrieves all taxonomic subtrees of a given organism “Hypsibiidae” from the UniProt database.

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX up: <http://purl.uniprot.org/core/>
SELECT ?root_name ?parent_name ?child_name
WHERE
{
  VALUES ?root_name {"Hypsibiidae"}
  ?root up:scientificName ?root_name .
  ?child rdfs:subClassOf+ ?root .
  ?child rdfs:subClassOf ?parent .
  ?child up:scientificName ?child_name .
  ?parent up:scientificName ?parent_name .
}
```

The D3SPARQL makes an AJAX call against a given SPARQL endpoint via a `d3sparql.query` function to retrieve the result in a SPARQL Query Results JSON Format. Then, user can choose a visualization pattern such as a `d3sparql.sunburst` as a callback function which internally calls a `d3sparql.tree` function to transform the resulted JSON data into a tree structured JSON format (Fig. 1). Finally, an interactive SVG image is generated and shown on a Web page (Fig. 2) in which any part of the subtree is clickable to zoom in.

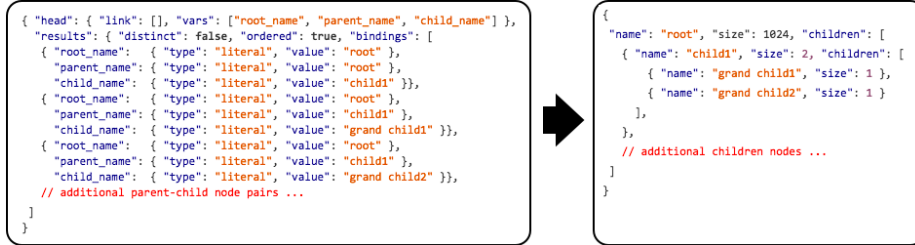


Fig. 1. Example of a *d3sparql.tree* JSON transformation for a tree structured data

Visualization patterns currently implemented in the D3SPARQL library are given in Table 1. Our integrated semantic genome database, TogoGenome [10], partly uses the library as a biological application which will also be shown in the software demo session.

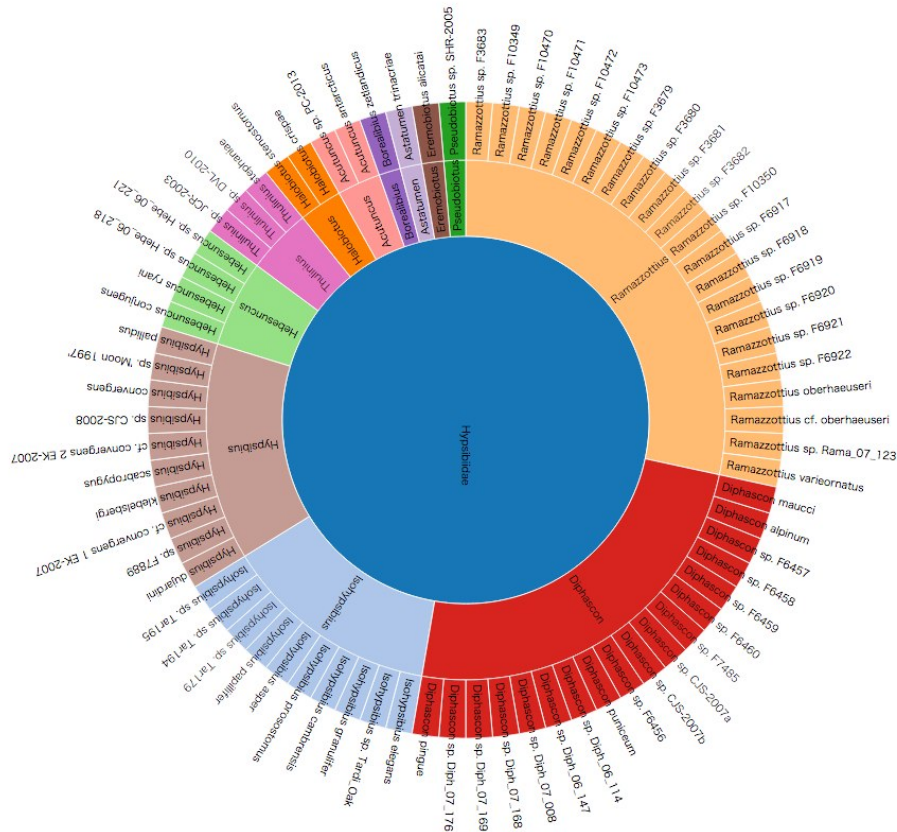


Fig. 2. Example of a *d3sparql.sunburst* visualization of the SPARQL result showing a species tree under the selected organism “Hypsibiidae” from the UniProt database

**Table 1.** Visualization categories and types implemented in the D3SPARQL library

Category	Type
Charts	Bar chart, Pie chart, Scatter plot
Graphs	Force graph, Sankey graph
Trees	Round tree, Dendrogram, Treemap, Sunburst, Circlepack
Maps	Geographic map
Tables	HTML table

## Discussions

Even though the D3SPARQL itself is a generic JavaScript library for any SPARQL endpoints, there are many cases where the library can accelerate development of life science applications. It is planned to implement more biological visualizations to utilize biomedical semantic Web resources in the future.

## References

1. Uniprot Consortium: Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 42, D191–8 (2014).
2. Callahan, A., Cruz-Toledo, J., Ansell, P., Dumontier, M.: *The Semantic Web: Semantics and Big Data*. Springer Berlin Heidelberg, Berlin, Heidelberg (2013).
3. Kinjo, A.R., Suzuki, H., Yamashita, R., Ikegawa, Y., Kudou, T., Igarashi, R., Kengaku, Y., Cho, H., Standley, D.M., Nakagawa, A., Nakamura, H.: Protein Data Bank Japan (PDBj): maintaining a structural data archive and resource description framework format. *Nucleic Acids Res.* 40, D453–60 (2012).
4. EBI RDF, <http://www.ebi.ac.uk/rdf/>.
5. PubChemRDF, <https://pubchem.ncbi.nlm.nih.gov/rdf/>.
6. GlycoRDF, <http://glycoinfo.org/>.
7. Kodama, Y., Mashima, J., Kosuge, T., Katayama, Toshiaki Fujisawa, T., Kaminuma, E., Ogasawara, O., Okubo, K., Takagi, T., Nakamura, Y.: The DDBJ Japanese Genotype-phenotype Archive for genetic and phenotypic human data. *Nucleic Acids Res.* (submitted).
8. SPARQL Query Results XML Format, <http://www.w3.org/TR/rdf-sparql-XMLres/>.
9. SPARQL Query Results JSON Format, <http://www.w3.org/TR/sparql11-results-json/>.
10. D3.js, <http://d3js.org/>.
11. TogoGenome, <http://togogenome.org/>.