Semantic web: the SPARQL language

Why the semantic web?

The main issue of the Internet today is all the content of the web is easily understandable by the human but it is not for a computer. Indeed, the information is not strictly structured and the computer does not understand the true mining of the words.

For example, with the sentence “Winston Churchill died in the city of London” is easy for a human to understand tree things:

- Winston Churchill is a human;
- London is a city;
- London is the “city of death” of Churchill.

This is not that simple for a computer!

The standardization of the semantic web

The World Wild Web Consortium (W3C), the main international standards organization for the World Wide Web (WWW or W3), is creating new specifications to organize the information. The main family of the specification is the Resource Description Framework (RDF).

The graph RDF data model is based upon the idea of triplet. Each information has one subject, one predicate and one object. One resource can be several information.

Go back on our example (the sentence “Winston Churchill died in the city of London”):

- “Winston Churchill” is the subject;
- “city of death” is the predicate;
- “London” is the object.

SPARQL

We have introduce the main idea of the semantic web. The second issue is to get the information in RDF. The SPARQL Protocol and RDF Query Language (SPARQL) is a semantic query language for database. It especially created to manipulate data stored in Resource Description Framework (RDF).

The language was made by the W3C. There are to versions:

- The 1.0 version release on 15 January 2008;
- The 1.1 version, since March 2013.

SPARQL is well adapted for the specific structure of RDF data and use the triplet architecture (subject, predicate and object).

There is two different types of requests we can make:

- SELECT : create a new graph RDF keeping only the resources matching the WHERE parameter (to get information from the graph);
- CONSTRUCT : add new information in the RDF graph (to modify the graph);

The main advantage of SPARQL is that it is analogous to NoSQL database (the relation "document-key-value" is there a "subject-predicate-object" relation). But it also considered as a relation database with three columns (subject, predicate and object). However, each column as a variant type.

Example of request

We want to select all the first episodes of all the series on the series on dbpedia.org:

```
SELECT *
WHERE {
    ?e  <http://dbpedia.org/ontology/releaseDate>    ?date   .
    ?e  <http://dbpedia.org/ontology/episodeNumber>  1       .
    ?e  <http://dbpedia.org/ontology/seasonNumber>   1
}
ORDER BY DESC(?date)
```

The link to try: here or http://dbpedia.org/snorql and copy the request.

Conclusion

SPARQL is a kind of evolution of the SQL. The "web 3.0" will allow computer to manage the information and this language is a way to do that.

This open a lot of possibilities, especially the centralization of the information: the servers will be able to communication together and share information.

Sources

- https://www.w3.org/TR/rdf-sparql-query/
- https://en.wikipedia.org/wiki/World_Wide_Web_Consortium
- https://en.wikipedia.org/wiki/Semantic_Web
- https://en.wikipedia.org/wiki/Resource_Description_Framework
- https://en.wikipedia.org/wiki/Linked_data
- https://en.wikipedia.org/wiki/SPARQL