Creating smarter clients based on top of Linked-Data and Semantic Web

Amet Vikram
Computer Science and Information Systems (CSIS)
BITS Pilani K.K. Birla Goa Campus
Goa, India
amet97vikram@gmail.com

Shravan Bharat Doda
Computer Science and Information Systems (CSIS)
BITS Pilani K.K. Birla Goa Campus
Goa, India
shravanbharatdoda@gmail.com

I. Introduction

This paper is divided into three sections. The first section describes the content in the paper and specifies the objective of the paper. The second section elaborates and elucidates the theory behind the semantic web and other related technology that encapsulates the core function of our software. The final section gives comprehensive information about the tools, and their operation, used in our software.

The objective of this paper is to design and implement a proof of concept and come up with a software for linked data and semantic web.

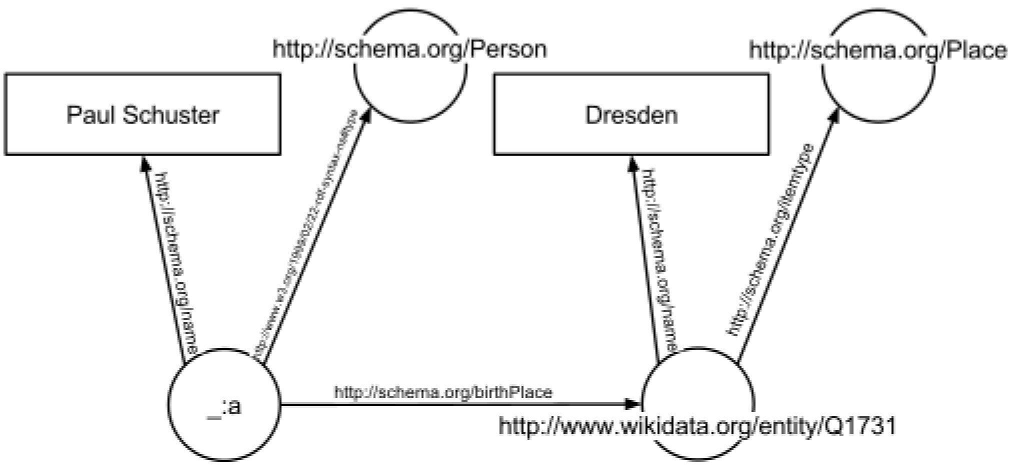
II. Theory

A. Semantic Web

The Semantic Web is an extension of the World Wide Web through standards by the World Wide Web Consortium (W3C). In addition to the classic “Web of documents” W3C is helping to build a technology stack to support a “Web of data,” the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term “Semantic Web” refers to W3C’s vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data.

When applied in the context of the modern internet, it extends the network of hyperlinked human-readable web pages by inserting machine-readable metadata about pages and how they are related to each other. This enables automated agents to access the Web more intelligently and perform more tasks on behalf of users.

```
<div vocab="http://schema.org/" typeof="Person">
  <span property="name">Paul Schuster</span> was born in
  <span property="birthPlace" typeof="Place"
href="http://www.wikidata.org/entity/Q1731">
    <span property="name">Dresden</span>.
  </span>
</div>
```



[The Linked Data machine readable graph resulting from above html code]

Many files on a typical computer can also be loosely divided into human-readable documents and machine-readable data. Documents like mail messages, reports, and brochures are read by humans. Data, such as calendars, address books, playlists, and spreadsheets are presented using an application program that lets them be viewed, searched and combined. Currently, the World Wide Web is based mainly on documents written in Hypertext Markup Language (HTML), a markup convention that is used for coding a body of text interspersed with multimedia objects such as images and interactive forms. Metadata tags provide a method by which computers can categorize the content of web pages. Because of this metadata tagging and categorization, other computer systems that want to access and share this data can easily identify the relevant values.

Semantic Web involves publishing in languages specifically designed for data: Resource Description Framework (RDF), Web Ontology Language (OWL), and Extensible Markup Language (XML). HTML describes documents and the links between them. RDF, OWL, and XML, by contrast, can describe arbitrary things such as people, meetings, or airplane parts. These technologies are combined in order to provide descriptions that supplement or replace the content of Web documents. Thus, content may manifest itself as descriptive data stored in Web-accessible databases, or as markup within documents (particularly, in Extensible HTML (XHTML) interspersed with XML, or, more often, purely in XML, with layout or rendering cues stored separately). The machine-readable descriptions enable content managers to add meaning to the content, i.e., to describe the structure of the knowledge we have about that content. In this way, a machine can process knowledge itself, instead of text, using processes similar to human deductive reasoning and inference, thereby obtaining more meaningful results and helping computers to perform automated information gathering and research.

An example of a tag that would be used in a non-semantic web page:

```
<item>blog</item>
```

Encoding similar information in a semantic web page might look like this:

```
<item rdf:about="http://example.org/semantic-web/">Semantic Web</item>
```

[Example of Non-Semantic and Semantic web page]

B. Linked Data

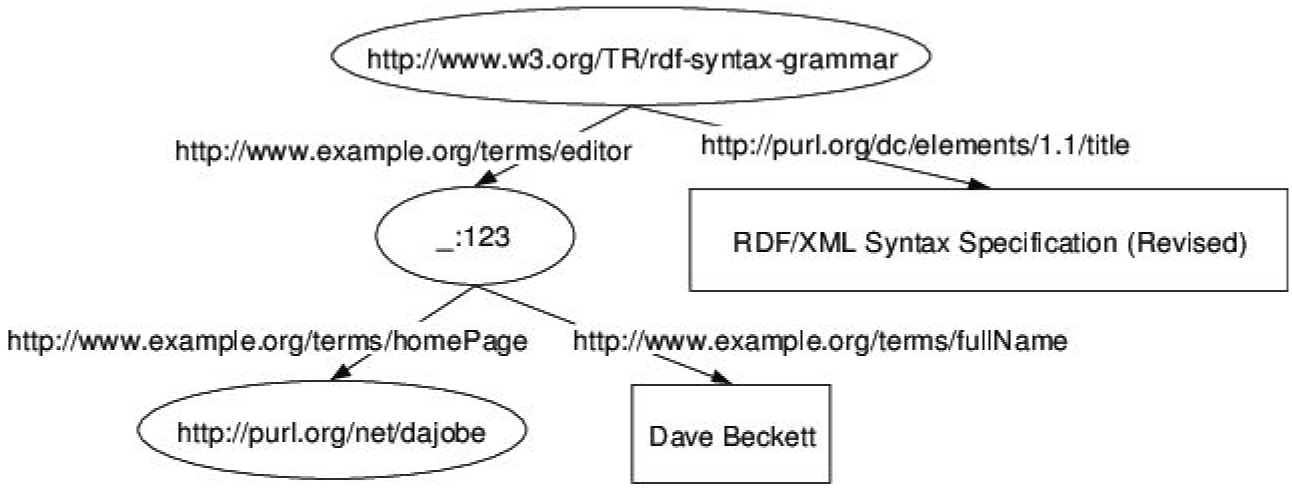
linked data is a structured data which is interlinked with other data so that it becomes more useful through semantic queries. It builds upon standard Web technologies such as HTTP, RDF and URIs, but rather than using them to serve web pages only for human readers, it extends them to share information in a way that can be read automatically by computers. Part of the vision of linked data is for the internet to become a global database.

To achieve and create Linked Data, technologies should be available for a common format (RDF), to make either conversion or on-the-fly access to existing databases (relational, XML, HTML, etc). It is also important to be able to setup query endpoints to access that data more conveniently. W3C provides a palette of technologies (RDF, GRDDL, POWDER, RDFa, the upcoming R2RML, RIF, SPARQL) to get access to the data.

C. Resource Description Framework (RDF)

RDF is a standard model for data interchange on the Web. RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed. RDF extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a “triple”). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications. This linking structure forms a directed, labeled graph, where the edges represent the named link between two resources, represented by the graph nodes. This graph view is the easiest possible mental model for RDF and is often used in easy-to-understand visual explanations.

This mechanism for describing resources is a major component in the W3C's Semantic Web activity: an evolutionary stage of the World Wide Web in which automated software can store, exchange, and use machine-readable information distributed throughout the Web, in turn enabling users to deal with the information with greater efficiency and certainty. RDF's simple data model and ability to model disparate, abstract concepts has also led to its increasing use in knowledge management applications unrelated to Semantic Web activity. A collection of RDF statements intrinsically represents a labeled, directed multi-graph. This in theory makes an RDF data model better suited to certain kinds of knowledge representation than are other relational or ontological models. However, in practice, RDF data is often stored in relational database or native representations (also called Triplestores—or Quad stores, if context such as the named graph is also stored for each RDF triple). As RDFS and OWL demonstrate, one can build additional ontology languages upon RDF.



[An example of RDF graph]

III. Software Overview

A. Tools Used

- Language : Python
- Version Control System : Git

B. About Tobias

tobias is the basic module from which API designers/engineers/developers can build up their own automated infrastructure to leverage smart clients. It is a framework that enables REST APIs to be described semantically using RDF. It is based on JSON-LD. It is a Python implementation and can navigate a websites autonomously by reading their API documentation. It is a generic client because it can query and retrieve data from any, Websites API provided their vocabulary.

C. Software Demonstration

This project was created in order to demonstrate how the above concepts come into play in real world.
It uses spotify for dynamically fetching metadata about tracks, albums and artists. This metadata is arranged as described by a custom created json-ld vocabulary in a rdf graph format.

```
>>> get songs
+-----------------------------------------------------+--------------+-------------------------------------------------------+
|                     track.name                      | track.length |                       track.uri                       |
+-----------------------------------------------------+--------------+-------------------------------------------------------+
|              Don't Look Back in Anger               |     4:53     | https://open.spotify.com/track/6qZoQsBTQW2tix6KViDC7F |
|       Wish You Were Here - 2011 Remastered Version  |     5:34     | https://open.spotify.com/track/7aE5WXu5sFeNRh3Z05wwu4 |
| Rocket Man (I Think It's Going To Be A Long Long Time) |  4:41     | https://open.spotify.com/track/2zvot9pY2FNl1E94kc4K8M |
|            Here Comes The Sun - Remastered          |     3:5      | https://open.spotify.com/track/45yEy5WJywhJ3sDI28ajTm |
|                     Wildflowers                     |     3:10     | https://open.spotify.com/track/2Pr1nZpt8A8WP7QYpyq6L3 |
|                   Romeo And Juliet                  |     6:2      | https://open.spotify.com/track/2fF7mQOxo1SZuIMHTQgQD8 |
|      Comfortably Numb - 2011 Remastered Version     |     6:22     | https://open.spotify.com/track/082cLCIXNPg2ruTrENz4Vt |
|                     Tiny Dancer                     |     6:17     | https://open.spotify.com/track/4BGJSbB5rAcg4pNzD4gfxU |
|                   Father And Son                    |     3:41     | https://open.spotify.com/track/2tbdjsdJ9xtyunKnZBkGxF |
|                Let It Be - Remastered               |     4:3      | https://open.spotify.com/track/5V1AHQugSTASVez5ffJtFo |
|                Sailing To Philadelphia              |     5:53     | https://open.spotify.com/track/5ti5SeraA9Ci7OtbLhEkz2 |
|         Every Breath You Take - Remastered 2003      |     4:13     | https://open.spotify.com/track/5C0LFQARavkPpn7JgA4sLk |
|                   Wonderful Tonight                 |     3:45     | https://open.spotify.com/track/4qmHeQBYieqj3VMAPascsE |
+-----------------------------------------------------+--------------+-------------------------------------------------------+

b'Query internal execution time: 0.334922 milliseconds'
>>> get albums
+------------------------------------------------------------------+-----------------------------------------------------------------------+
|                            album.name                            |                            album.art_url                              |
+------------------------------------------------------------------+-----------------------------------------------------------------------+
|                          Stop The Clocks                         | https://open.spotify.com/image/4ea8f898e556dd416c58e1dd6894e55667a4a39b |
|         Wish You Were Here [Remastered] (Remastered Version)     | https://open.spotify.com/image/b608097cc4e5074c9793615cd67810dacecdb7eb |
|                    Honky Chateau (Remastered)                    | https://open.spotify.com/image/7a27ef49fb0bbd6291de42cd98c82a9377835db8 |
|                     Abbey Road (Remastered)                      | https://open.spotify.com/image/a650b9dadd2b2d66ab9d7788abdcbfab45b2997d |
|                           Wildflowers                            | https://open.spotify.com/image/0efb93effa150fcbfd061ca2f197a45611dded98 |
|                    Making Movies (Remastered)                    | https://open.spotify.com/image/edfc889a740fea5ecc1170bb9f8ccabe27e91c31 |
|        The Wall [2011 - Remaster] (2011 Remastered Version)      | https://open.spotify.com/image/a414a431992c75fac336467d50e54f574c1929f2 |
|                       Madman Across The Water                    | https://open.spotify.com/image/f86510ed4dd2af194a06d1c1dffa5ea1df7bf32e |
|                  Tea For The Tillerman (Remastered)              | https://open.spotify.com/image/fcdfb079d3d2976cfc40bff7576da442a12e418e |
|                      Let It Be (Remastered)                      | https://open.spotify.com/image/a8b7d6b28fc36248278e128c33e1a0dfb8cffb88 |
| The Best Of Dire Straits & Mark Knopfler - Private Investigations | https://open.spotify.com/image/3255e4ed7c3533ed9c17330c698270cb86f0afee |
|                     Synchronicity (Remastered)                   | https://open.spotify.com/image/cf93b4c9b557054cbf3f6a1b00c08731c6a9b662 |
|              Slowhand 35th Anniversary (Super Deluxe)            | https://open.spotify.com/image/0e7c06beb95545231b2efd5d2832ed593ca16239 |
+------------------------------------------------------------------+-----------------------------------------------------------------------+
```

The project uses redisgraph to create the graph database and cypher to query the graph database. The spotify metadata is fetched with the help of dbus module native to all linux systems. It also supports some other eye-candy features such as opening up the album-art using feh or visualizing the graph with the help of graphviz.

The vocabulary mainly consists of three classes and 6 properties. The classes include track, album and artist. Apart from these, the standard music ontology can also be used for processing the metadata.

References

- **[RFC5988]** Web Linking. M. Nottingham. IETF. October 2010. Proposed Standard.
- **[RFC6570]** URI Template. J. Gregorio; R. Fielding; M. Hadley; M. Nottingham; D. Orchard. IETF. March 2012. Proposed Standard.
- https://www.w3.org/standards/semanticweb/
- https://www.w3.org/RDF
- **[RFC2119]** Key words for use in RFCs to Indicate Requirement Levels. S. Bradner. IETF. March 1997. Best Current Practice.
- [**BCP47**] Tags for Identifying Languages. A. Phillips; M. Davis. IETF. September 2009. IETF Best Current Practice.
- **[Turtle**] RDF 1.1 Turtle. Eric Prud'hommeaux; Gavin Carothers. W3C. 25 February 2014. W3C Recommendation.
- Kuriakose, John (September 2009). "Understanding and Adopting Semantic Web Technology". Cutter IT Journal. CUTTER INFORMATION CORP.
- Buffa, Michel; Dehors, Sylvain; Faron-Zucker, Catherine; Sander, Peter (2005). "Towards a Corporate Semantic Web Approach in Designing Learning Systems: Review of the Trial Solutions Project". International Workshop on Applications of Semantic Web Technologies for E-Learning. Amsterdam, Holland.
- Grigoris Antoniou, Frank van Harmelen (March 31, 2008). A Semantic Web Primer, 2nd Edition. [The MIT Press].