

SemScape: Visualizing Semantic Web Data Landscapes with Cytoscape 3.0

Andrea Splendiani^{1,2}, Andra Waagmeester³, Carina Haupt⁴, and Helena Deus¹

¹ DERI, Galway, Ireland

² intelliLeaf ltd, Cambridge, UK

³ Maastricht University, NL

⁴ Bonn-Aachen Institute for Information Technology, University of Bonn, DE

Abstract. Core to the success of applying Semantic Web technologies (SWT) towards supporting Life Sciences research is the availability of tools that lower the entry barrier for adoption by biomedical researchers. Researchers need to easily and intuitively exploit and query the wealth of data that is available behind as SPARQL endpoints. Here, we present SemScape, a semantic-web enabled plugin for the popular network biology software Cytoscape. SemScape can be used to query any knowledge bases with a SPARQL endpoint by leveraging familiarity with existing software and intuitiveness of big data exploitation through a mechanism that encapsulates the complexity of data in parametric context dependent queries. We believe SemScape can provide a valuable resource both for data consumers and data publishers.

1 Introduction

The Semantic web and linked data efforts in Life Sciences domains have produced a large amount of structured information ready for querying. However, the entry barrier for researchers willing to use this information on their daily tasks is too great still, posing a particular problem for user interaction and query assembly, especially in the Life Sciences domain. Despite the well-recognized advantages of Linked Data and RDF formalisms in enabling the seamless aggregation of data from multiple distributed sources, these formalisms inherently rely on a format that sets data free from the shackles of a predefined schema.

As a side-effect, data represented in these formats is challenging to query since users cannot rely on organized data maps made available a-priori. While ontologies can provide such a map, there is still a disconnect between ontologies proposed by the semantic web communities, and the data descriptors that are actually used in RDF representations. As such, it is difficult to find ontologies that encompasses and arbitrary mashup of data, which is typical scenario for the use of data published in RDF, in the Life Sciences and beyond. In the life sciences domain, the problem is further compounded by the realization that potential users of the data are typically not experts in semantic technologies, and ease of use as well as domain oriented interfaces are even more relevant.

Hence the need for tools that provide an accessible interface to semantic web

knowledge bases, that can help in discovery the information structure and navigate it, as well as put into context with the typical work environment of researchers.

2 Results

We have developed SemScape, a plugin for the popular network biology software Cytoscape [?], which allows user interaction and queries over remote endpoints. Cytoscape is a network visualization and analysis tool that, while inherently domain independent, provides links and shortcuts to typical biomedical datasets and manipulations. It offers a plugin architecture and has a very active community both of developers and users. SemScape is developed as plugin for the latest version of Cytoscape, 3.0, which is released in beta at the time of this writing and present a completely new architecture from previous versions.

SemScape offers a range of functionalities, which are derived from what was implemented in RDFScape [?] and AGUIA [?]. However this system provides a newer, more efficient and up to date implementation. The functionalities of SemScape can be categorized essentially in discovery and navigation of linked data landscapes.

Discovery functions are based on the automatic extraction of a data map (typically types and relations from a given endpoint).

Navigation functions are based on contextual menus that can be configured by users and data publishers: when right-selecting a node, the system uses attributes of the node (e.g. its type) to look into a query library for queries that can be applied to the node. E.g.: if the node is a protein, the system may find queries for interacting proteins. In general for each resource simple queries such as all statements including the resource as object or subject are provided. Once a selection is performed, results are used to expand the graph, and provenance information is kept for each new element visualized (which in turns allow new contextual queries to the original endpoint). In particular, queries are written as parametric queries, and can be provided by data publisher as a package to which the user can subscribe. For instance a pathway resource provider may provide, together with an endpoint, a set of preconfigured queries to drive user interaction on its knowledge base. The user can keep track of such preconfigured packages and update them if newer versions are proposed by the data publisher. Together with the possibility to execute and visualize generic SPARQL queries, SemScape allow the possibility to extract and define data landscapes, graphs visualizing the semantic links between data sets in a given context (e.g. Cancer).

3 Conclusions

SemScape is a robust plugin developed on the newer version of the widely used network tool Cytoscape and it inherits in its design the experience of several experiments in user interaction of the Semantic Web. We believe it will be a

valuable resource for Life Science researchers willing to make use of these technologies.

4 Availability

SemScape is available <http://code.google.com/p/vsdlc3/> and released under the Apache License 2.0.

5 Acknowledgments

This work was supported by the Google Summer of Code program and all development is by Yigang Zhou. We wish to thank Alex Pico for his support in the GSoC program.

References

1. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*. **13(11)** (2003) 2498–504 .
2. Splendiani A: RDFScape: Semantic Web meets Systems Biology *BMC Bioinformatics* **9(Suppl 4)** (2008) S6
3. Correa MC, Deus HF, Vasconcelos AT, Hayashi Y, Ajani JA, Patnana SV, Almeida JS: AGUIA: autonomous graphical user interface assembly for clinical trials semantic data services. *BMC Medical Informatics and Decision Making*. **10:65** (2010) .