CSCI8380 (Spring 2012): Paper Review Form

Reviewer Name: Mehdi Allahyari

Paper Name: gStore: Answering SPARQL Queries via Subgraph Matching

Section I. Overview

A. Reader Interest

1. Which category describes this manuscript?
   ___Practice/Application/Case Study/Experience Report
   _✔__Research/Technology
   ___Survey/Tutorial/How-To

B. Content

1. Please explain how this manuscript advances this field of research and/or contributes something new to the literature.

   Answer: *This manuscript introduces and explains a new method based on graph approach to store and do queries on RDF data. It describes how SPARQL queries can be answered precisely by using this technique, which converts the RDF data and SPARQL query into graph and sub graph matching query respectively, indexing them and applying search algorithms. It also proposes an efficient algorithm to perform online updates over RDF datasets with the lowest overhead. With respect to dramatic increase in using RDF data, querying and processing RDF data are two fundamental issues, and this manuscript is a useful help in terms of facing these challenges.*

C. Presentation

1. Does the introduction state the objectives of the manuscript in terms that encourage the reader to read on?
   _✔__Yes
   ___Could be improved
   ___No

2. How would you rate the organization of the manuscript? Is it focused? Is the length appropriate for the topic?
   _✔__Satisfactory
   ___Could be improved
   ___Poor

3. Please rate and comment on the readability of this manuscript.
   ___Easy to read
   _✔__Readable - but requires some effort to understand

___Difficult to read and understand
___Unreadable

Section II. Evaluation

Please rate the manuscript. Explain your choice.

___Award Quality
___Excellent
__✔_Good
___Fair
___Poor

Section III. Detailed Comments (provide your thoughts/criticism about the ideas in the paper; not only summarize the paper but have a critical look here)

Answer: *The main focus in this paper is to address the issues of existing solutions for processing SPARQL queries in an efficient way, and the authors propose a new graph based method to handle this problem. Their idea is to convert the RDF and SPARQL query into directed graphs with vertices and edges where every vertex and edged has a label as well. The RDF graph has index to be effectively searched, and the goal is finding the matches of SPARQL query graph in RDF graph.*

The authors have tried to express their new solution in a coherent and fluent way, but critically speaking, I believe the paper is very technical, not easily digestible and needs quite a bit of knowledge in Math and graph theory. Knowing SPARQL is also required to understand it fully.

Additional Comments:

1. Provide one aspect that you liked the most in this paper.

   The idea of using encoded graph is quite interesting to me. Besides, a new indexing technique is used to speed up the access to nodes, which sounds great. I think combining these method together helps to reduce the time and space complexity of the algorithms used, from the graph perspective point of view.

2. Provide one aspect that you disliked the most in this paper.

   Even though the SPARQL queries and the RDF data were pretty simple, generating and encoding graph seems a bit tedious, complicated and kind of confusing. Therefore I am not sure how this technique is feasible for very large RDF data and complex SPARQL queries.

Section IV. Discussion Points (provide at least 3 discussion topics/questions related to ideas/techniques described in the paper; these will be used for discussions in the class)

1. *One of the most important aspects of any graph algorithm is the correctness of that. In this paper various algorithms are used for encoding and for query matching, but the authors haven't discussed about the* ***Correctness*** *of the algorithms.*

2. *How the encoding is scalable for complex SPARQL queries?*

3. *Why adjacency list is used for storing the RDF data? Is there any other data structure, which might be better than adjacency list in terms of access time and space complexity?*