Medical Question Answering Framework to Query Statistical Data

Usman Akhtar, Sungyoung Lee, *Kyung Hee University, South Korea*

***Abstract*— In this paper, we have proposed a framework to handle the Semantic question answering (SDA) over the Linked Open Data (LOD), where data in the form of the RDF data cube. As statistical data published all over the internet there is a need to empowers the non-expert to query in the form of the natural language. Our proposed framework will provide a support to interact in the form of the question answering and produce a SPARQL query to extract the answer from the RDF data cube.**

I. INTRODUCTION

An increasing amount of the statistical data is published on the Linked Open Data (LOD) cloud. Getting insights from the data in more intuitive ways are becoming important. Systems for the Semantic Questions Answering (SQA) plays a vital role to connect with linked open data and provides an intuitive interface by translating natural languages queries into SPARQL syntax. Statistical data need more advance querying methods to empowers non-experts users to draw their own conclusions. Semantic question questioning is extremely important in the Healthcare and Life Sciences (HCLS) and biomedical. We have motivated the need of the semantic question answering, however there are many challenges. First, lack of the processing the RDF by the current question answering system and second is the lack of the query template that extract the information from the Linked Open data.

II. PROBLEM

In summary, from the prior work we can represent the medical statistical data as a RDF data cube. However, how to execute the SPARQL queries over the data cubes is still the challenging issue. We have not seen any work looking at putting all aspect of RDF Cubes together. As argued in the introduction we cannot use existing Question Answering (QA). Since, they do not provide the query template to match the RDF cube.

There are some solution strategies; one is Template Based Question Answering [1] and another is the extracting RDF from the text. However, both these approaches partially support the SPARQL query process on the statistical data. The current solutions are not feasible. But, we have adopted the template based question answering approach and we have extended to support the natural question answering support.

*Research supported by NRF Korea.

Usman Akhtar. Author is a Ph.D. research student in depart. of Computer Science and Engineering in Kyung Hee University, Yongin-si, South Korea (corresponding author: e-mail: usman@ oslab.khu.ac.kr).

Sungyoung Lee, He is a full professor affiliated with Department of Computer Science and Engineering in Kyung Hee University, Yongin-si, South Korea. He is full professor in KHU, (e-mail: sylee@oslab.khu.ac.kr).

III. OUR APPROACH

We propose a framework by combining different approaches to Semantic Question Answering (SQA) over RDF data cute, where natural language questions convert in to a SPARQL query and work on the statistical data. To design we take into the account of the existing SAQ architecture over the linked data.

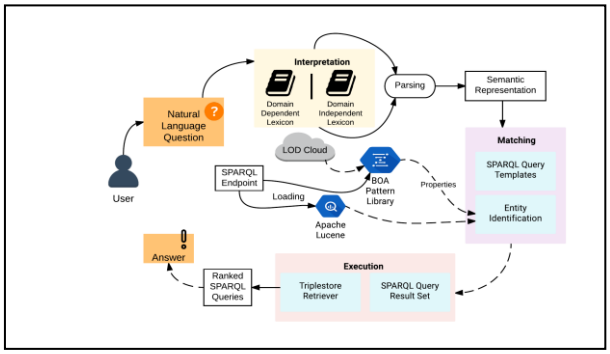

Figure 1. *Our proposed framework*

Figure 1. showing the overview of our proposed approach. The framework consists of the three main states interpretation, Matching and execution. User posted a natural language question over the Linked Open Data that is stored in the triple store, the framework will generate the SPARQL query and gives the answer.

The major research finding is that, more and more medical data is published on the web, the main question is how the user will access the resources is becoming crucial importance.

ACKNOWLEDGMENT

This work was supported by the Industrial Core Technology Development Program (10049079), Develop of mining core technology exploiting personal big data) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea) and This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (2011-0030079).

REFERENCES

[1] C. Unger, L. Bühmann, J. Lehmann, A.-C. Ngonga Ngomo, D. Gerber, and P. Cimiano, "Template-based question answering over RDF data," in *Proceedings of the 21st international conference on World Wide Web*, 2012, pp. 639-648.