BHAVI 2019 Winter Quarter Report*

Shiladitya Dutta

created March 19, 2019, revised March 19, 2019

Primary Research Project

Description: The main objective at the beginning of Winter Quarter was to make CoVaSEA more usable. I believe that this goal has been accomplished in a variety of ways. Currently, the main progress that has been made during Winter Quarter is the development of a GUI for CoVaSEA. This GUI consists of 4 main options. The first option is for a complete search which includes an external web crawler search and an internal SPARQL query search. The second option is for only an external web crawler search to convert records and the third option is an internal SPARQL query search to analyze existing records. The last option is the ability to change the environment variables of CoVaSEA to allow for easier set-up. In addition, the SPARQL query builder form has been finalized. It consists of three main sections. The first section is choosing the conditions for the SPARQL query search. The second section is choosing the outputs for the search and the last section is choosing any post-conditions.

In addition to the addition of new features, I have attempted to make this project more easy-to-use and reproducible. Firstly, I have added an external environment variables file so that the environment variables can be changed without having to change anything in the code. I have also written a user installation manual such that it is easier to learn how to install the system. There has also been other minor changes such as moving the semantic metadata records file to the CoVaSEA program file so that the environment variable doesn't need to be changed. I have also tested these features with another student. Other than these, I have also continued work on the lexical-to-semantic translator. Most notably, I have spent time to improve the parsing capabilities and runtime.

Citations:

- [1] - A paper describing a semantic reasoning engine using direct logic to make inferences about semantic graphs in a database.
- [2] - This paper describes Apache Jena, a semantic triplestore database. Jena could be a possible triplestore system for CoVaSEA.
- [3] - SHACL is supposed to act as a potential replacement for SPARQL. SHACL differs from SPARQL in SPARQL is condition based while SHACL is shape based.
- [4] - This paper describe a linear logic theorem prover. This is another possible approach to semantic reasoning in the CoVaSEA internal search engine.
- [5] - A paper on PELLET a semantic reasoning engine. PELLET is a relatively well-known OWL 2 based semantic reasoner.
- [6] - This article details Google's approach to lexical-to-semantic markup translation. Currently, most translations systems, including CoVaSEA's, utilize a multi-stage discourse representation theory parser. However, Google has taken a new approach by simply having a neural network model perform the parsing. This approach is very interesting and elegant, though it does have its own fair share of problems.

*Document typeset March 19, 2019.

- [7] - An article describing an open source software library which implements a neural machine parse: Opennmt.

Conferences/Competitions:

- GGSF - March 23rd (Already included in Fall Quarter report, however competition technically occurring in Spring Quarter) Also it should be noted that California State Science Fair (CSSF) and/or Intel International Science and Engineering Fair (ISEF) may occur depending on my performance at GGSF. Both of these competitions should occur during Spring Quarter.

Statement: I wish to continue my current primary research project.

Secondary Research Project

Description: Nothing to report. I didn't have a secondary research project since it was originally planned to be with Arnav Bansal. However, Arnav Bansal left the program thus I didn't have a secondary research project.

Statement: For my secondary research project, I wish to contribute to Adam Craig's research on FAIR metrics. This is due to its potential relevance to my current work and my involvement in the project thus far.

Software Education Project

- I am not currently enrolled in any CSSE coursework at my school.
- I have not completed or am currently enrolled in any CSSE coursework online. However, I am currently doing USA Computing Olympiad (USACO). USACO is a competitive coding competition in the same vein as TopCoder, ACM ICPC, or HackerRank. USACO administers a series of 4-hour online tests from December to March. Each test consists of 3-4 algorithmic coding problems to complete. Though not necessarily coursework, it is worth noting this activity.
- I have continued reading the Algorithm Design Manual. Most notably I have spent more time reading about graph theory and tree traversal algorithms. This is because these could possibly be useful for the tree parsing section of CoVaSEA. I have also read more about data structures in an attempt to learn possible ways to optimize the runtime of CoVaSEA. [8]

Discussion

I would suggest the compilation of a complete list of BHAVI sayings and phrases. This would include phrases such as "Real software that really works", "RTFM", "Deliver the deliverables", and "Elite program for elite students". This would provide a simple and somewhat humorous introduction to the cardinal values of BHAVI for new students. I have no other suggestions or requests in regards to the BHAVI program.

References

[1] N. Bassiliades, G. Antoniou, and I. Vlahavas, "A defeasible logic reasoner for the semantic web," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 2, no. 1, pp. 1–41, 2006.

[2] J. J. Carroll, I. Dickinson, C. Dollin, *et al.*, "Jena: Implementing the semantic web recommendations," in *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, ACM, 2004, pp. 74–83.

[3] J. Corman, J. L. Reutter, and O. Savković, "Semantics and validation of recursive shacl," in *International Semantic Web Conference*, Springer, 2018, pp. 318–336.

[4] J. Rao, P. Küngas, and M. Matskin, "Composition of semantic web services using linear logic theorem proving," *Information Systems*, vol. 31, no. 4-5, pp. 340–360, 2006.

[5] E. Sirin, B. Parsia, B. C. Grau, *et al.*, “Pellet: A practical owl-dl reasoner,” *Web Semantics: science, services and agents on the World Wide Web*, vol. 5, no. 2, pp. 51–53, 2007.

[6] Y. Wu, M. Schuster, Z. Chen, *et al.*, “Google’s neural machine translation system: Bridging the gap between human and machine translation,” *arXiv preprint arXiv:1609.08144*, 2016.

[7] G. Klein, Y. Kim, Y. Deng, *et al.*, “Opennmt: Open-source toolkit for neural machine translation,” *arXiv preprint arXiv:1701.02810*, 2017.

[8] S. S. Skiena, *The algorithm design manual: Text*. Springer Science & Business Media, 1998, vol. 1.