The semantics of data refers to the method machines use to identify the different meanings of data. The labeling of this data is a process that allows computers to more easily handle the data in the future. The technology is used all the time, in applications all around us. For instance, Google uses semantic data labeling to sort and organize specialized search results. When you search for “Flight to LAX,” Google uses the different data types to sort them into a table, offering a simple and intuitive representation of search results.

Dr. Knoblock has been working on using a mix of machine learning and semantic data modeling to autonomously assign semantic labels to data sources. What this means is that the types of data (name, date, age, etc.) no longer need to be manually labeled; the entire process can be done by the computer. This technology will be used to quickly and efficiently scan the web for certain criteria. Specifically, it is being designed for a program that will be able to detect a specified type of content on the web for the US Army, such as illegal weapons sales and human trafficking ads.

Because of the instructive and developmental nature of my STEM coursework, my classes constantly push me into the areas of computer science with which I am unfamiliar. This includes new languages, unfamiliar programs, new ideas and techniques, and a fair amount of collaboration. The process of studying and working with advanced systems, along with the quick pace of the lab environment has improved my proficiency in all these capacities. Throughout the experience, I have had to quickly learn new programs, familiarize myself with new computing languages, and learn about new processes, all while relying on others in the lab, and collaborating with more experienced peers.

Additionally, my coursework is normally based around learning the finer points of computer science. Though some examples used in class relate to real life applications of computer science, I am not normally exposed to the uses of what I’ve learned. In the lab, however, I got to see, and take part in, the kind of work that is actually done with what I learned in class.

In the upcoming year, I plan on taking two additional computer science classes. The first will be in android coding, the second, in advanced Java. I will also apply what I have learned in my upcoming classes, and on ongoing projects, as well as the projects still to come. For instance, with my new knowledge of Python, I will finally be able to finish a number of programs on which I had already been working.

During my time in the lab, I felt as though I wasn’t qualified to be there. I felt as if the work being done was so far out of my intellectual range that I shouldn’t have even been in the building. So, I asked the other SHINE students about my feelings. Throughout my inquiries between the students, there was a general consensus: You never really begin to feel like you understand what’s going on. It’s important to remember that, though you are not a PhD level student, you are being asked to do PhD level work. With that in mind, always keep trying to do what is asked of you, and make sure to try and understand why. The real benefit of the program is not always finishing the tasks given to you, but learning how to perform them in the first place.

Special thanks to Dr. Craig Knoblock, Minh Pham, Dr. Pedro Szekely, Dr. Maja Mataric, Dr. Katie Mills, Lancelot Wathieu, Ian Andrade

Skills Learned
- Efficient on-the-spot research
- Python
- Terminal Commands
- Github
- MatLab
- Eclipse
- PyCharm
- Tomcat Servers

Examples of Terminal Commands
http://action-andi.com/

Program to convert XML files to HTML pages written in Python
http://jetbrains.com/