Oakland University student builds framework for querying medical imaging data

The Oakland University and School of Engineering and Computer Science communities are invited to attend Sarmad Istephan’s defense of his Ph.D. dissertation. Seating is limited. RSVP with Haroldene Perzyk at perzyk@oakland.edu.

An Extensible Content-Based Support Framework Providing Unlimited Querying of Unstructured Medical Imaging using a Big Data Approach for Data-Driven Medicine

Committee: Mohammad-Reza Siadat, Ph.D. (Chair), Theophilus Ogunyme, Ph.D., Wenjin Zhou, Ph.D., Guangzhi Qu, Ph.D.

With the ever increasing amount of medical imaging data, it is critical to have an extensible content-based support framework that allows for unlimited querying of such data. In this dissertation, this type of framework has been developed to advance data-driven medicine by allowing medical experts to query and test hypotheses on huge volumes of unstructured medical imaging data. The developed framework is efficient, flexible, and accurate. A Big Data approach has been used in the framework’s core to provide ultimate efficiency and flexibility. Specifically, Hadoop Streaming jobs are used to distribute the querying of the unstructured medical imaging data using built-in and user-defined feature extraction modules. At a high level, the framework executes a query in two phases. Phase 1 deals with querying the structured data in the clinical data warehouse that was developed in this dissertation. Phase 2 uses the results of phase 1 to query the unstructured imaging data using modules in Hadoop Streaming. Through testing the framework, several findings have been made. One finding is that using Hadoop’s distributed architecture compared to a traditional single server architecture, the surface to volume and average intensity modules performed up to 40 and 85 times faster, respectively. Therefore, the framework is efficient. The second finding is that the framework is able to integrate and execute new user-defined modules. Therefore, the framework is extensible. The last major finding is through testing a sophisticated and practical medical imaging query, the framework returned all patients who met the query’s structured and unstructured criteria. Therefore, the framework is accurate. These positive results show that the developed framework is an advancement in data-driven medicine providing a powerful mechanism for unleashing the unstructured content of medical imaging data to be queried by medical. Ultimately, this framework provides better patient care.