

Master Thesis – IoT Data Catalogues

While platforms and tools such as Hadoop and Apache Spark allow for efficient processing of Big Data sets, it becomes increasingly challenging to organize and structure these data sets. Data sets have various forms ranging from unstructured data in files to structured data in databases. Often the data sets reside in different storage systems ranging from traditional file systems, over Big Data files systems (HDFS) to heterogeneous storage systems (S3, RDBMS, MongoDB, Elastic Search, ...). At AGT International we are dealing primarily with IoT data sets, i.e. data sets that have been collected from sensors and that are processed using Machine Learning-based (ML) analytic pipelines. The number of these data sets is rapidly growing increasing the importance of generating metadata that captures both technical (e.g. storage location, size) and domain metadata and correlates the data sets with each other, e.g. by storing provenance (data set x is a processed version of data set y) and domain relationships.

Tasks

We are looking for a bright, curious and skilled final year student that is interested in developing a prototypical system that semi-automatically creates data catalogues describing our IoT data sets. The successful applicant will have access to a wide range of AGT's IoT data sets from the domains of Smart Cities, Industry 4.0, event management and Social IoT.

We expect the student to develop a crawler that automatically creates structured metadata using semantic web and linked data technologies (RDF, triples stores). Where appropriate machine learning techniques should be used to suggest appropriate metadata. Collected data should be linked to existing metadata that is already available. The developed system should also contain a web-based front-end that allows to review, complete and access both the data catalogues and data sets.

Requirements

- Proficiency in Java and JavaScript programming
- Background or strong interest in linked data, semantic data management, machine learning
- Hands-on mentality with a strong drive to get things done
- Fast learning skills

If you are interested and wish to work in an agile and open-minded international company, learning from experience data scientists, researchers and developers, then don't miss the opportunity to send us your application, including CV and master transcript, via email.

About AGT International

AGT International is privately held multinational company headquartered in Zurich, Switzerland and a pioneer in IoT data management and advanced Big Data analytics solutions. AGT International offers solutions for Connected Cities, Sports & Entertainment, and Manufacturing (Industry 4.0). In Sports and Entertainments our ambition is to change the way life events are managed and experienced. As part of our Joint venture with WME|IMG, a global leader in sports and entertainment events, we have deployed our solutions to events such as the New York Fashion Week or Basket Ball Euroleague Final Four. More information is available at <http://www.agtinternational.com>.

Contact: Dr. Martin Strohbach (mstrohbach@agtinternational.com)