Motivation

- PhoneLab is one of the largest crowd-sensing platforms
  based on the Android system.
- The data gathered is flat-structured, making derivation of meaningful information extremely cumbersome and time consuming.
- The goals are to (i) make the measurement data more accessible, (ii) increase interoperability and reusability of data gathered from different sources, (iii) develop extensible data representation to support future development of the PhoneLab platform.

Oracle Data Mining

- Oracle supports Oracle Data Miner with easy to use GUI and easy integration with R enterprise.
- Oracle 12c supports multitenant architecture

Oracle Data Mining Algorithms

Part 1: Converting JSON TO RDF (Ontology) format

- Apache Jena allows creating the RDF model which is a statement that consists of subject(resource), predicate(property) and the object(resource or literal)

Part 2: Loading RDF into Oracle 12c

- Oracle supports unstructured(RDF) data storage and provides advanced analytics.
- RDF stored using Jena Library. Oracle provides a Jena Adapter that provides a java interface to connect to Oracle to perform batch loading.
- Billions of Triples, Terabytes of Storage

Part 3: Applying Oracle Data Mining Techniques

- Different data mining tools available like R, Oracle Data Miner
- Find user pattern and behavior using classification/clustering

Context Extraction

- SPARQL queries to extract information from triples

Implementation

Figure 1: Data Mining Techniques

Figure 2: SPARQL query example

Figure 3: Raw JSON data

Figure 4: Jena Triple Example

Figure 5: Jena Adapter Bulk Loader

Figure 6: Apache Jena Triple Example

Figure 7: Oracle Data Miner Workflow

Figure 8: Oracle Data Miner with easy to use GUI and easy integration with R enterprise.

Figure 9: Apache Jena Triple Example

Conclusion

- RDF allows easy context and correlation extraction
- Apache Jena framework and Oracle seamless integration
- ODM supports mining RDF and apply various models

References