sdmay18-27: Data Analytic Tools for Inconsistency Detection in Large Data Sets

Week 9 Report
November 4 - November 17

Team Members

Camden Voigt — *Technical Lead*
Christopher Konopka — *Communication Lead*
Logan Heitz — *Project Lead*
Timothy Rogers — *Quality Lead*

Summary of Progress this Report

We determined that we needed to modify our current implementation of the raw data parser and configuration parser to account for new formatting in the raw data for identifier values. It was also determined that an xml attribute would need to be added to export format to signify that the column should be an index in the table for quick lookup.
Implementation of the exporter class was completed. This class will create a table to store the daily reports that will be used in inconsistency testing.
Implementation of the onthefly matcher was started and has preliminary matching implemented. This will need to be updated to work with the database and to receive one record at a time as they are parsed. By only receiving one record at a time we will not have to store all the records in memory at one time which will vastly reduce on the memory needed for our program.

Pending Issues

Finish implementation of the onthefly matcher
Update raw data parser to improve modularity
Implement testing for the exporter
Implement testing for hashing
Update testing for the raw data parser
Testing of the prototype with large dataset to get speed estimates

Plans for Upcoming Reporting Period

Finish implementation of the onthefly matcher (Christopher Konopka)
Update raw data parser to improve modularity (Logan Heitz)
Implement testing for the exporter (Jackson Voigt)
Implement testing for hashing (Timothy Rogers)

Individual Contributions

Team Member	Contribution	Weekly Hours	Total Hours
Camden Voigt	Created the exporter class and table creation from the configuration file. This will allow us to create tables on the fly to store in coming reports for testing purposes.	11	32
Christopher Konopka	Begun implementation of the onthefly matcher and worked with Timothy on updating the configuration and raw data parser for this part.	12	28.5
Logan Heitz	Implemented the exporter method which will allow for tuples taken from the raw data to be placed into the table. Updated the table implementation and data configuration file to have an option for indexing a table column. Implemented the hasher class that will hash values if needed before placing them in the database.	13	37
Timothy Rogers	Updated the configuration parser and the raw data parser to handle identifier tags in the raw data and to send records one at a time to the inconsistency checker and exporter.	13	36