ABSTRACT

SINGLE COLUMN CARDINALITIES, DATA PATTERN AND *VALUE DISTRIBUTION* ANALYSIS FOR *DATA PROFILING* USING AN *OPEN SOURCE PLATFORM*

By

SANDY RAMA AMETHYST

1202142181

The importance of data quality might have a major impact on the company's existing business processes. But there are still many companies that yet to understand the importance of data quality. Many cases that often occurs to the quality of data in many companies in Indonesia is that the inputted data are not filtered, so there are issues about the absence of a data, data patern is not standardized and data that are not distributed evenly. These cases can be handled with several processes, one of which is data profiling. Data profiling is a proses of collecting an information of a data so it can be measured its quality, accuracy and consistency. Data profiling can be done by using various tools either from paid tools to an open source tools where each has its own advantages in accordance with the needs. In this research, the main focus on data analysis is by conducting data profiling using cardinalities, data pattern, and value distribution methods using open source tools. The results of the profiling will be implemented in logical form in open source application and will be compared with another open source tools.

Keyword : data *profiling, open source, cardinalities, data pattern, value distribution*