

ABSTRACT

The rapid movement of data flow today makes many private companies and state-owned companies competing to process data owned by each organization, the things that are required to process the data, one of those is having the good quality of the data owned, in fact many companies and state-owned companies do not have good data quality, the biggest contributor of the problem is because of the many applications that have stand-alone databases for each application it makes a lot of redundancy. To overcome these problems there are a profiling process and that can be used to perform cleansing, the method can be achieved with a variety of applications from paid to open, but the high price of paid applications make many organizations use open applications. In this research, the researcher performs the data profiling using value similarity and completeness method, the result of the algorithm that has been made will be implemented using Pentaho Data Integration and will be compared with other application to see the difference of result from the algorithm that has been made.

Keywords : *data profiling, value similarity, data completeness, open source, data, cleansing*