

ABSTRAK

Cepatnya pergerakan data saat ini membuat banyak perusahaan swasta serta BUMN berlomba untuk mengolah data yang dimiliki oleh masing – masing organisasi, hal – hal yang diperlukan untuk mengolah data tersebut salah satunya adalah memiliki kualitas yang dari data yang dimiliki, pada kenyataannya banyak perusahaan dan BUMN yang belum memiliki kualitas data yang baik penyumbang terbesar dari masalah tersebut adalah karena banyaknya aplikasi yang memiliki database yang berdiri sendiri untuk masing – masing aplikasi hal tersebut membuat banyaknya data yang terulang. Untuk menanggulangi masalah tersebut terdapat proses *profiling* yang nantinya dapat digunakan untuk melakukan *cleansing*, metode tersebut dapat diraih dengan berbagai macam aplikasi dari yang berbayar hingga yang bersifat terbuka, namun tingginya harga dari aplikasi berbayar membuat banyak organisasi memanfaatkan aplikasi terbuka. Pada penelitian ini peneliti melakukan data *profiling* menggunakan metode *value similarity* dan *data completeness*, hasil dari algoritma yang sudah dibuat tersebut akan diimplementasikan menggunakan Pentaho Data Integration dan dilakukan komparasi dengan aplikasi lain untuk melihat perbedaan hasil dari algoritma yang sudah dibuat.

Kata kunci : data *profiling*, *value similarity*, *data completeness*, *open source*, data, *cleansing*