

Title: Data Profiling

Author: Radka Hladíková

Department: Department of Software Engineering

Supervisor: Ing. Vladimír Kyjonka

Supervisor's e-mail address: Vladimír.Kyjonka@cze.sas.com

Abstract:

This thesis puts mind on problems with data quality and data profiling. This Work analyses and summarizes problems of data quality, data defects, process of data quality, data quality assessment and data profiling. The main topic is data profiling as a process of researching data available in existing data sources and creating of statistics and information about these data. There is a projected system for evaluating data status in term of its quality. Work is focused on measuring of general characteristic of data, following data defects and its analyses. With the help of data quality SW there is a projected and realized system for evaluation of data quality for real data.

Keywords: data quality, data profiling, data quality metric, DataFlux