

Data Quality – Not Your Typical Database Problem

Mourad Ouzzani

Qatar Computing Research Institute
mouzzani@qf.org.qa

Abstract.

Textbook database examples are often wrong and simplistic. Unfortunately Data is never born clean or pure. Errors, missing values, repeated entries, inconsistent instances and unsatisfied business rules are the norm rather than the exception. Data cleaning (also known as data cleansing, record linkage and many other terminologies) is growing as a major application requirement and an interdisciplinary research area.

In this talk, we will start by discussing some of the major issues and challenges facing creating effective and efficient data cleaning solutions. Then we will discuss some challenges and criticize current conservative approaches to this very critical problem. Finally we will discuss some of our work at QCRI in this area.