SIGMOD 2010 RWE Review on Paper "Sampling Dirty Data for Matching Attributes"

1   Experiment Overview

In this report, we will provide some analysis on the repeatability and workability on the source codes provided by the authors of SIGMOD 2010 paper "Sampling Dirty Data for Matching Attributes". According to the comments from the primary reviewer, the second reviewer only runs a simple re-evaluation on the experiments listed in the paper.

Based on the source codes and running scripts provided by the authors, we have set up some experimental environment on a Windows XP (SP2) operating system equipped on a Dell desktop with Intel Duo-Core CPU E6550 at 2.33 GHz, 3.25 GB main memory and 250 GB hard disk. The programs are run in environment with Java Development Kit 1.6.0 and MySQL Server 5.1.

2   Summary from Primary Reviewer

In this section, we give a brief summary on the RWE results by the primary reviewer on this paper.

The primary reviewer has successfully repeated all the experiments in the package. He only points out some problems with the installation notes. After testing the program by myself, I also noticed that it is important to set JDK and MySQL appropriately to continue the experiments.

3   Repeatability and Workability

I have re-run all the scripts in the program package and validate the correctness of the results. The only real problem I met is modifying the scripts correctly when I use another table name from the data downloaded from the web site.

Since I'm not very familiar with relational data, I didn't find any appropriate data set for workability tests. But from my experience on the repeatability tests, I am sure the program will generally work, if the user fully understand the scripts for corresponding modifications.

4 Conclusion

The program works well on the given data sets, even when the version of the database system is only different from the specific one.