Proposal for Mid-Sem Innovative Project
Database Management Systems
MC-302

Submitted By:

Aiman Siddiqua   Apoorva
2k18/MC/008   2K18/MC/019

PROJECT:

Analysing and Reproducing Results of the Paper "Similarity Metrics for SQL Query Clustering"

PAPER DETAILS:

Title: Similarity Metrics for SQL Query Clustering [1]

Authors: G. Kul, D. T. A. Luong, T. Xie, V. Chandola, O. Kennedy and S. Upadhyaya

Journal: *IEEE Transactions on Knowledge and Data Engineering*, Vol. 30, No. 12, December 2018

OBJECTIVE:

" Database access logs are the starting point for many forms of database administration, from database performance tuning to security auditing, to benchmark design, and many more. Unfortunately, query logs are also large and unwieldy, and it can be difficult for an analyst to extract broad patterns from the set of queries found therein. Clustering is a natural first step towards understanding the massive query logs."

<div align="right">– From the abstract of "Similarity Metrics for SQL Query Clustering"</div>

The paper aimed to evaluate three different heuristics for computing similarity of SQL queries and propose a feature engineering method to improve their accuracy.

Our goal in this project will be to understand and analyse the paper, conduct a literature review of similar papers, and reproduce the results using the code and data published on Github by the authors.
The deliverable will be a report on the above.