Algorithms for processing relaxed functional dependencies

Nguyen Thi Linh

Abstract: This thesis investigates algorithms for processing relaxed functional dependencies (RFDs), which extend classical functional dependencies by accommodating data inconsistencies, noise, and approximations common in real-world datasets. While traditional functional dependencies enforce strict equality, RFDs allow for controlled deviations, making them crucial for data quality assessment, schema discovery, and inconsistency-tolerant querying. The research proposes and evaluates novel algorithms for the efficient discovery, validation, and inference of several RFD types, including approximate, conditional, and metric dependencies. A primary contribution is the development of a scalable, partition-refinement-based algorithm that leverages pruning strategies to reduce the exponential search space inherent in RFD mining. Additionally, the thesis introduces a sound and complete axiomatic system for a unified class of RFDs, enabling logical implication testing. Experimental results on both synthetic and benchmark datasets demonstrate that the proposed algorithms significantly outperform existing methods in runtime and scalability while maintaining high accuracy. The findings provide a robust computational foundation for applying relaxed dependencies in data profiling, cleaning, and integration tasks.

Keywords: relaxed functional dependency, processing algorithm, data dependency, database