SEDEX: Scalable Entity Preserving Data Exchange

Data exchange is the process of generating an instance of a target schema from an instance of a source schema such that source data is reflected in the target. Generally, data exchange is performed using schema mapping, representing high level relations between source and target schemas. In this paper, we argue that data exchange solely based on schema level information limits the ability to express semantics in data exchange. We show such schema level mappings not only may result in entity fragmentation, they are unable to resolve some ambiguous data exchange scenarios. To address this problem, we propose Scalable Entity Preserving Data Exchange (SEDEX), a hybrid method based on data and schema mapping that employs similarities between relation trees of source and target relations to find the best relations that can host source instances. Our experiments show SEDEX outperforms other methods in terms of quality and scalability of data exchange.