Abstract:

Data integration and inter-operability among multiple data sources has emerged as an important problem for organizations and business operations. Structural and syntax level heterogeneity have been addressed using data exchange standards. Semantic heterogeneity poses major challenges. Our context of investigations has been geographical data. Semantic Interoperability is a major issue for National Spatial data Infrastructures (NSDIs) and mapping across heterogeneous databases is essential for such interoperability. Mappings based on ontologies provide opportunities for semantic translation of schema elements and hence for database queries across heterogeneous sources. Such semantics based mappings are usually human centered processes. We have developed a framework for semi-automatic mapping using semantic similarity values of descriptions associated with data. Lexical similarity of class and attribute names and class structures constitute knowledge base for mapping between two schemas. We employ semantic mapping based on synonym similarity matches from WorldNet. We use heuristics based propagation of similarities using attribute mapping and superclass-subclass relations. We also apply constraints to filter out inconsistent mappings. The machine based similarity values are seen to be comparable to human generated values of mapping.