

DataGuides: Enabling Query Formulation and Optimization in Semistructured Databases

Reviewer: Youngdae Kim

October 30, 2007

Summary

This paper presents *DataGuide*, a structural summary of semistructured database (especially graph-structured database). Since in semistructured database there exists no fixed schema in advance, user could have difficulties in formulating queries and there were also difficulties for query planner to find an optimized query plan. DataGuide stores all the possible labels without duplication which comprise the structural information of the database and some sample values for the labels, hence users can formulate queries on their needs and query planner can optimize a given query by referencing the DataGuide. In addition to the query formulation and optimization support, DataGuide can dynamically update the summaries as the underlying semistructured database schema changes with as small computation as possible.

Comments

Since unlike the traditional database adding or deleting an object causes schema changes in semistructured database, it is very important for summary data structure to have dynamic update property. DataGuide supports dynamic update with as small computation as possible, so it is very useful as a summary data structure for semistructured database.

Not only summarizing the underlying structure of semistructured database, a *strong DataGuide* can store annotations (a property of objects reachable by the label) in DataGuide. Examples of annotation can be sample values or statistical information of objects and users can have benefits from this annotations since annotations help to formulate queries more appropriately.

Although DataGuide has great features, DataGuide can be redundant information when the underlying semistructured database has no redundant label paths. This means that there is a one-to-one correspondence between label path and data path instance. In this case, DataGuide cannot compact the original database and is just a duplication of the original database. But I think that this rarely happens.