

Grant Number: IIS-0238386

Institution of PI: Duke University

PI: Jun Yang

Title: CAREER: Techniques and Applications of Derived Data Maintenance

Research Objectives:

To develop techniques and applications of derived data maintenance. Derived data is obtained by applying structural or computational transformation to base data, and is widely used to facilitate access to base data. Examples include caches, indexes, views, etc. Derived data must be *maintained* when base data is updated. We seek new solutions to traditional problems (e.g., view maintenance and data warehousing) as well as new applications (e.g., XML indexing and continuous query processing).

Significant Results:

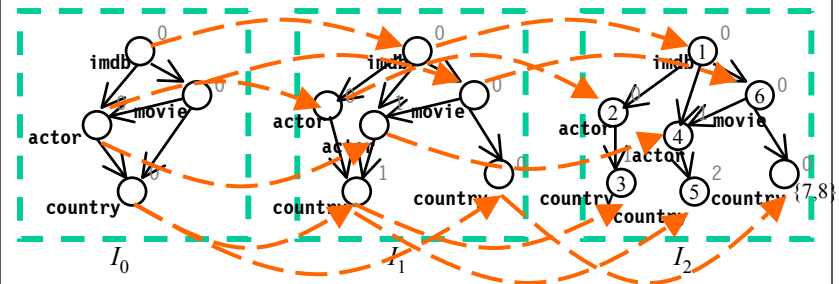
During the first year, we have made the following contributions:

1. Incorporation of caching in view maintenance (ICDE 2003, Master's thesis);
2. Incorporation of caching in stream data processing (submitted);
3. Incorporation of caching in XML indexing (ICDE 2004);
4. Incremental maintenance of XML structural indexes (SIGMOD 2004).

Approach:

Traditionally, derived data maintenance has been tackled separately in different contexts (e.g., index updates and view maintenance), although these problems share the same underlying theme. New and more complex data management tasks call for creative combinations of traditionally disjoint ideas. A good example is semantic caching, which incorporates the idea of views into a cache. We seek to discover more techniques that *combine multiple flavors of derived data to provide better solutions to problems*.

Graphic:



Broader Impact:

Our research will benefit a number of important new applications, including XML data management and continuous query processing. While our focus is on database problems, we hope that our techniques, developed by combining synergetic ideas from different research fields, will in turn contribute back to these fields and make impact beyond databases. Our research has also be incorporated into the new database curriculum at Duke University.

An example $M^*(k)$ index, a structural index for XML that incorporates ideas of caching and auxiliary views. It refines itself on a demand according to query workload, and is also the first structural index based on local similarity that is incrementally maintainable when base XML document is updated.

References: He et al., ICDE 2004; Yi et al., SIGMOD 2004.