Reminiscences on Influential Papers

Kenneth A. Ross, editor


Arnon Rosenthal, The MITRE Corporation, arnie@mitre.org


The paper contains two important ideas that have influenced both my research and several MITRE "grand visions" for data sharing. First, it suggested that it is important to capture receivers' interface requirements explicitly, rather than just harvest the queries they issue. The power of this simple idea is discussed in the last paragraph. Second, it provided context rules to describe attribute representations (e.g., simple assertions that the field FuelUnit describes the units for TankerContents, and inferred context: if CUSTOMER.ADDRESS.NATION= "USA" then the CUSTOMER.PRODUCT.PRICE has currency = "US dollar"). Using this sort of knowledge plus libraries of scalar transforms, their group built mediators that generated conversion programs. Over the next decade, I started creating mediators, and ended learning hard lessons about applying such technologies in large enterprises.

We learned that simple context mediation of attributes (when schemas match) was too small a unit of technology insertion. Administrators needed to manually map the context rules to the receiver's schema, an unacceptable burden. Siegel and Madnick's follow-on efforts (Goh et al., ACM TOOLS, 1999) provided a technical solution, but they captured knowledge in an exotic logic known only to a handful of researchers; our customers need a standards-based solution.

Despite these technology transition challenges, the basic idea of context mediation remains compelling. We used it to provide an alternative (simple context rules in OWL) versus current practice of documenting interface semantics in MS Office. We now believe that context mediation is a big win (over hand-coded transforms) if one has structural mappings and if one is mapping many sources to a receiver (or vice versa). We hope that in the near future, high-end integration systems (e.g., IBM's CLIO) will incorporate context mediation.

We also had successes, greatly simplifying several generations of data-sharing visions (large-scale vaporware) for our customers. We applied Siegel and Madnick's first insight (parallel descriptions for sources and receivers) for aspects such as: push versus pull, sending whole tables versus deltas, data quality requirements, data structure encodings, and protocol choice. This work (and also, mediators from MCC and ISI) contributed to the Department of Defense's dropping its vision of a monolithic standard, and managing instead in terms of diverse communities of interest (Rosenthal, et al., SIGMOD Record, December 2004).