SCHEMA COERCION: USING DATABASE META-INFORMATION TO FACILITATE DATA TRANSFER

by

Terence Critchlow

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Computer Science

The University of Utah

June 1997

ABSTRACT

As more information becomes available, the ability to quickly incorporate new and diverse data sources into existing database systems becomes critical. Schema coercion addresses this need by defining the mapping between databases as a collection of mappings between corresponding constructs. This work defines a comprehensive schema coercion tool: it transforms schemata into corresponding ER representations, identifies correspondences between them, and uses these correspondences to generate a program that automatically transfers data between the databases. In addition to creating a useful tool, this work addresses the significant theoretical problems associated with resolving representational and semantic conflicts between heterogeneous data sources. The approach advocated by this dissertation associates confidences with correspondences, and meta-information with schemata. This approach has successfully reduced the amount of interaction required to define several coercions, including a complex coercion between diverse genetics databases.