

**San Francisco State University  
Department of Computer Science**

***Technical Report SFSU-CS-TR-06.09***

**Title:      Web Services Joins Using ER Design Tool  
             XML Databus**

**Author(s):   Yan Liu**

**Date:         05/04/06**

**Abstract:**

With increasing demands for dealing with large scale distributed and heterogeneous data sources, XML and Web Services are being extensively and frequently used for data storage, data exchange and data integration. The CSC675 ER Design Tool can help users design a database by mapping Entity Relationship (ER) schemas into equivalent implementation schemas for different data models (relational, XML, ODMG, and spreadsheet). Its data transfer Web Service can exchange and integrate data among these data models based on an XML Databus. The goal of this project is to extend the data transfer Web Service to build a new join Web Service, which can execute distributed joins among distributed and heterogeneous data store Web Services.

Remote data stores are wrapped into the data transfer Web Service with existing dump and load operations. The data transfer proceeds by dumping data in a data store from its native data format to XML Databus format, and then loading data to another data store (usually with a different data model) from XML Databus format to the native data format. In this new project, the join Web Service uses the data transfer Web Service to invoke two data stores to dump data and ship them to a common relational data store. The data is joined in the relational database by using the semantically equivalent attributes manually identified in both schemas on the join service client side.

This project is developed using the Java Web Services Developer Pack (JWSDP). This includes the JAX-RPC API, used to implement Web Services, and the JAXP API, used to deal with XML documents. In order to demonstrate that the CSC 675 ER Design tool and its extension can be used for biological data integration, TSeq and GBSeq data sets (downloaded from the NCBI-NIH web site) are used to do experiments. These data sets are transformed from native XML format to XML Databus format using Extensible Stylesheet Language Transformation (XSLT). Performance analysis experiments show the join Web Service is effective, in that the join operation takes less than 2 seconds based on filtering the equivalent of over  $10^6$  (denormalized) records.

**Keywords:**   Data Integration, Web Service, XML, JWSDP, JAXP, JAXP, XSLT

**Copyright:**   Yan Liu