Semantically-Assisted Integration Query Editing in the AquaLogic Data Services Platform

M. Carey[1,2], S. Ghandeharizadeh[2], K. Mehta[1], P. Mork[3], L. Seligman[3], S. Srivastava[1,2], S. Thatte[1]

[1]*BEA Systems, Inc.*   [2]*Computer Science Dept.*   [3]*The MITRE Corporation*
2315 N. First Street   *Univ. of Southern California*   *7515 Colshire Drive*
San Jose, CA 95131   *Los Angeles, CA 90089*   *McLean, VA 22101*

E-mail: mcarey@bea.com

Abstract

This demonstration shows how semantic schema matching technology is being incorporated into the BEA AquaLogic Data Services Platform. Specifically, it demonstrates how the manually-intensive task of authoring a correct and complete integration query is enhanced by semantic guidance. It also highlights the pluggable architecture used to incorporate semantic match knowledge into ALDSP by showing the ALDSP Query Map editor interoperating with the Harmony semantic matching engine from MITRE Corporation.

1. ALDSP Overview

The AquaLogic Data Services Platform (ALDSP) is a middleware software platform for integrating and service-enabling information from disparate enterprise data sources. Drawing data from such diverse source types as relational tables and views, Web services, stored procedures, Java functions, application adaptors, XML files, and delimited files, ALDSP provides a unified, functional, XML-based view of enterprise information that enables a data architect to author coherent data services either graphically or by writing XQuery functions that transform and assemble the target data appropriately [1,3]. ALDSP provides full CRUD (create/read/update/delete) support for the integrated data via the resulting data services.

2. Semantically-Assisted Query Editing

The central tool for an ALDSP data architect is the ALDSP Query Map editor. A data architect uses this editor to transform and compose the various sources' data to provide a desired integrated view. The sources' data and integrated view are characterized using XML Schemas. The graphical integration query authoring problem is thus one of creating an XQuery function to combine and transform the data appropriately.

With today's ALDSP Query Map tool, users are required to know, or to ascertain, exactly how to map the data from the sources' schemas into the target schema. Prerequisite knowledge includes a precise understanding of the schemas involved, e.g., which elements and attributes match which others and what computations must be performed to correctly map their content. For large and/or unfamiliar schemas, this task can be both a difficult (identifying all matches) and tedious (due to problems of scale) process.

BEA is currently extending the ALDSP Query Map editor to consume match set information produced by semantic schema matching tools [2]. These match sets, describing source/target matches at various confidence levels, are used to interactively assist the data architect by drawing their attention to likely matches, displaying additional metadata about the matches, and helping them track their mapping progress. While performing mappings, users give feedback to the semantic match set by accepting/rejecting certain matches as well as capturing complex mapping formulae as hints for future mapping exercises. As hinted at in Figures 1 and 2, the result is a semantically-assisted query authoring process that leaves the data architect in control while streamlining their user experience.

3. References

[1] BEA AquaLogic Data Services Platform 3.0, BEA Systems, Inc., Jan. 2008, http://edocs.bea.com/aldsp/docs30/.

[2] M. Carey et al, Proc. 2007 Workshop on Semantic Data and Service Integration (SDSI), Vienna, Austria, Sept. 2007.

[3] M. Carey et al, "Data Delivery in a Service-Oriented World: The BEA AquaLogic Data Services Platform", Proc. ACM SIGMOD 2006, Chicago, IL, June 2006.
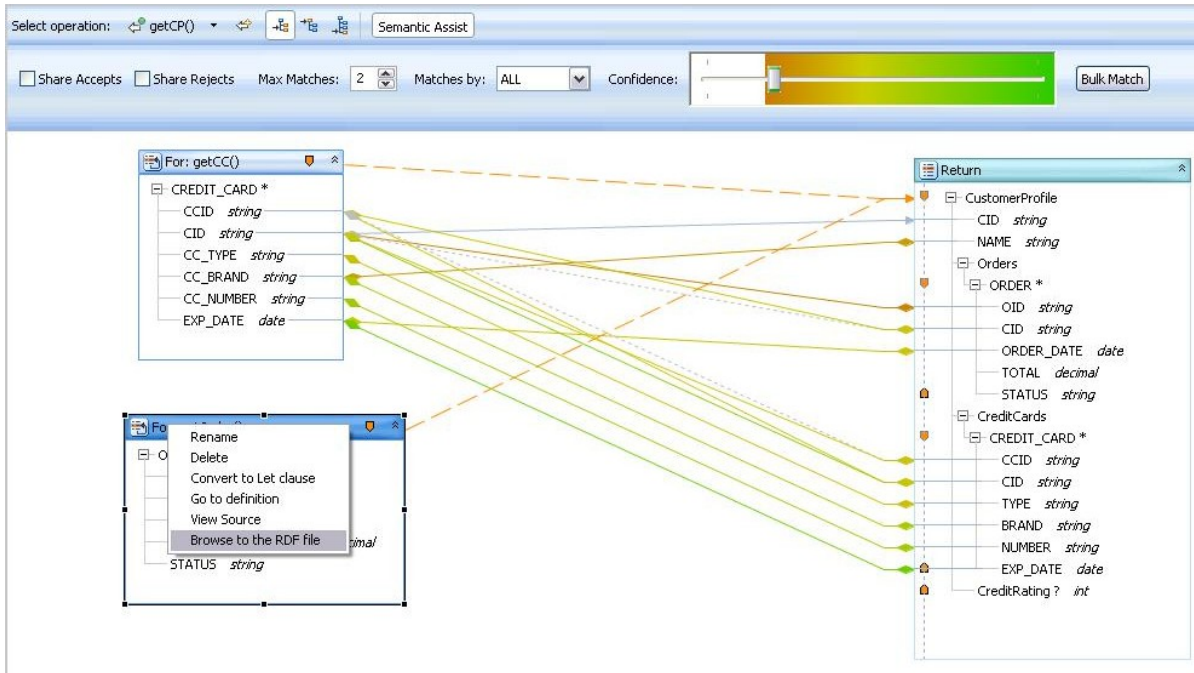
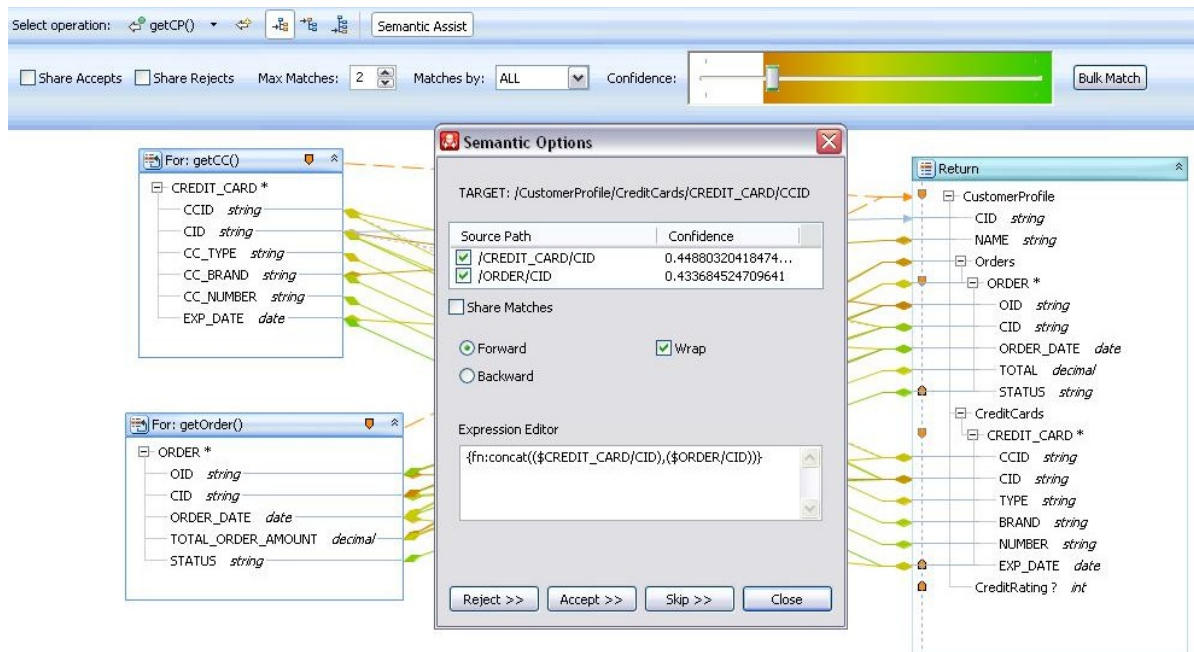**Figure 1. Importing an RDF file containing a semantic match set**

Figure 2. Authoring an integration query with semantic match set guidance