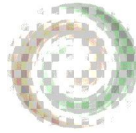


Simplifying e-Business Collaboration by providing a Semantic Mapping Platform



Abels, Sven¹; Sheikhhasan Hamzeh¹; Cranner, Paul²

¹ TIE Nederland BV, 1119 PS Amsterdam, Netherlands

² University of Sunderland, Sunderland, UK

{ sven.abels | hamzeh.sheikhhasan }@tie.nl; paul.cranter@sunderland.ac.uk

Abstract: Within this paper we introduce the STASIS approach for creating a comprehensive application suite that allows enterprises to simplify the mapping process between data schemas based on semantics as opposed to syntax. This paper initially introduces the current schema mapping problem and outlines the limitations of existing solutions. The STASIS approach is then presented and contrasted with other semantic projects such as “foam”. In the main part of this paper we explain our vision of creating an easy to use user interface that allows users to create semantic schema mappings while keeping the semantic base formats.

1 Introduction and Problem Description

In the last decade, a significant number of different e-Business related systems have been developed such as catalogue management tools and online shopping systems. Whenever two or more companies need to collaborate, they invariably need to exchange information electronically between such systems. In an ideal scenario, this information exchange is performed in an automatic way allowing business partners to exchange information seamlessly. However, because of the large number of diverse information systems the data format (syntax) of messages usually differs from company to company, or sometimes even within the same company if more than one software product is used. The situation is similar to the Tower of Babel involving many different people that want to work together on a specific task without understanding each other.

This makes it a very difficult challenge to exchange information in an interoperable way. Interoperability in this context means “the ability of two or more systems or components to exchange information and to use the information that has been exchanged” [1]. In those cases, users have to either agree on a common standard or they have to individually map the data format of one business partner to the data format of the other one. The STASIS project (www.stasis-project.net), funded by the European Commission’s FP6 programme [2] is addressing this problem. It provides a set of semantic methods for easing the mapping process between XML, Flatfile, EDI and Database schemas.

It is addressing a simple key question which is important for almost all European companies: “If I have information in my format, and it is integrated into my systems, I want to put minimal effort into mapping this into any format (standardised or not) to do electronic business with another party” [2].

2 Existing Solutions

Mapping processes between schemas today focus on a rather syntactical approach. Tools such as Altova XMLSpy¹ or TIE Integrator² are used to successively map different attributes of two different schemas to each other. Those attributes are in most cases either mapped in a 1:1 relationship or using various connectors such as Methlets (see [3] for explanations and details) for creating complex mappings. This approach has been pretty successful in the past and it allows a fast mapping creation if the schema is not too complicated. However, restricting the mapping creation to a syntactical mapping process has some disadvantages. These include:

- An in-depth knowledge of both the source and destination schemas is required in advance. New schemas need to be studied in detail before mapping. This is a very expensive step and cannot be avoided
- Mappings are often hard to understand and error-prone because humans normally think in terms of the actual semantics instead of the syntax – e.g. companies usually want to map two addresses instead of mapping Element [ELT4711] to [ELT0815]
- Mappings typically cannot be reused, and must be created from scratch when mapping to a new schema
- Concepts such as inheritance and logical constraints are not supported
- The mappings must be performed by a technical person who has schema knowledge whereas what is wanted is the mapping of business information by business people

¹ <http://www.altova.com>

² <http://www.tieglobal.com>

3 STASIS Idea and Approach

The STASIS project aims to address many of the problems outlined above and also both introduce the concept of market-driven semantics – “Competitive Semantics” – and to promote a neutral standard mapping format which can be exploited by existing transformation tools and technologies. Instead of focussing on syntactical mappings, STASIS concentrates on identifying semantic entities and mapping those semantic pieces. For example, two elements [ELT1] and [ELT2] may be grouped into one logical entity called “Address” which is mapped to a well defined concept of an Address.

This process is based on ontologies used to define and link semantic entities within a schema. The link to the original syntax is still made but this is transparent to the user.

Unlike approaches such as foam [4] and GLUE [5] STASIS does not intend to create an automatic Ontology Mapping, nor does it rely on semantic mapping approaches as analysed in [6]. Instead, it is our goal to provide an easy to use GUI, allowing users to identify semantic elements in an easy way. Once this identification has been performed STASIS lets users map their semantic entities to those of their business partners where possible assisted by STASIS. This allows users to create mappings in a more natural way by considering the meaning of elements rather than their syntactical structure.

All mappings that have been created by STASIS, as well as all semantic entities, are managed in a distributed registry and repository network called the SRRN (see [7]). This gives STASIS another significant advantage over traditional mapping creation tools as STASIS may reuse all mappings. This allows STASIS to make some intelligent mapping suggestions by reusing mapping information from earlier semantic links. For example, imagine two companies A and B that would like to map their business schemas in order to exchange information. Let’s assume that both have conducted business with company C already in the past. In this situation, STASIS can map the semantic entities of A and B automatically because it knows about their mapping to a common schema from company C. This will become increasingly beneficial as more companies begin to use STASIS.

4 A Roundtrip: From OWL to EMF and Back

In order to realize STASIS, the consortium has selected the Eclipse Framework as a base for creating the Graphical User Interface of the mapping editor. Eclipse provides much functionality that may be reused and will therefore reduce the development time in most scenarios.

A definition of what a semantic entity is, and on how a schema is represented in STASIS, is expressed in the STASIS Logical Data Model (LDM), which itself is expressed in OWL. This logical data model is the base for the graphical editor of STASIS. In order to create this editor, STASIS is reusing two other core concepts of Eclipse: EMF and GMF. - EMF (Eclipse Modelling Framework, see [8]) is used to create a definition of the LDM model in eCore on MOF level (see [8]). GMF [8] is a framework that allows creating a graphical editor that is based on EMF. The overall process is illustrated in Figure 1:

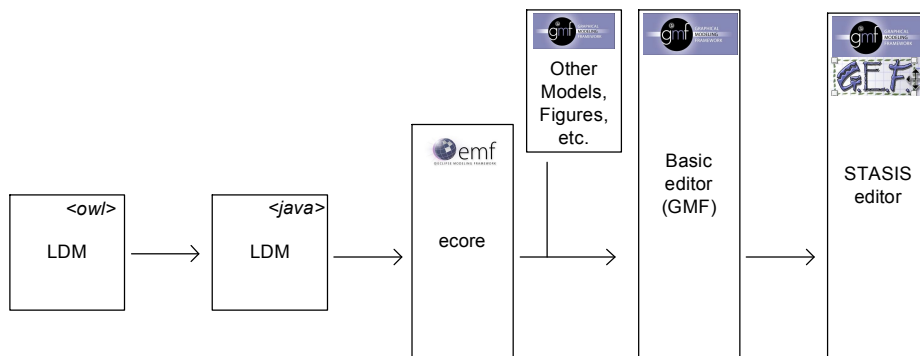


Figure 1: Editor Creation

As seen in the figure, the overall LDM definition is exported into annotated java classes using the Protégé³ EMF export methods. In the next step, the Eclipse functionality is used to derive an eCore file as well as an EMF editor and finally to generate a GMF editor, which is the base for the STASIS schema editor component.

Using OWL parser libraries such as Kazuki⁴ allows us to continue using OWL based ontologies with our STASIS editor as displayed in Figure 2.

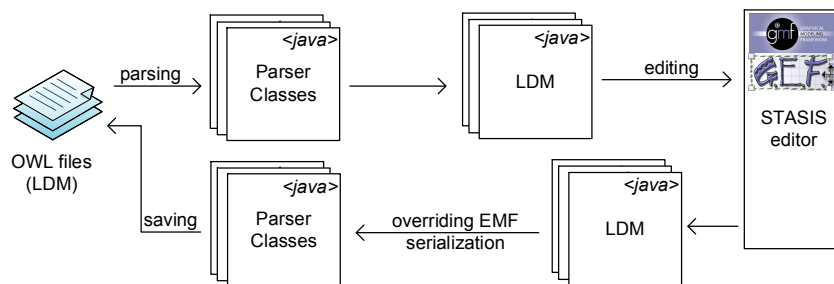


Figure 2: Roundtrip

³ <http://protege.stanford.edu>

⁴ <http://projects.semwebcentral.org/projects/kazuki>

5 Further Activities and Conclusion

The STASIS project and its implementation is ‘work in progress’ and is due to be completed in 2009. In the current state, a basic GMF editor has been defined which may already be used to create new schemas graphically and to map schema elements to semantic entities. However, there is still a huge potential for extending STASIS.

Features that will be integrated in the next phase include a query environment allowing users to use an ontology query language (SparQL) to query for mappings stored in the distributed network of STASIS (SRRN). In addition to this, STASIS will be able to export mappings into XSLT and it will be able to import existing schemas that have been defined in XSD.

References

1. Institute of Electrical and Electronics Engineers: IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries (1990)
2. STASIS project, Software for Ambient Semantic Interoperable Services, <http://www.stasis-project.net> (2008)
3. TIE Integrator, An easy-to-use system that can transform and apply business rules to data in any format, passing through enterprise systems around the clock, Product Brochure, <http://www.tieglobal.com> (2004)
4. Ehrig, M., Sure, Y. Ontology Mapping by Axioms. In: Proceedings of the 3rd Conference Professional Knowledge Management. Springer (Heidelberg) (2005)
5. Doan, A. et al: Ontology matching: A machine learning approach. In Handbook on Ontologies in Information Systems, Springer-Verlag (2004)
6. Abels, S.; Haak, L.; Hahn, A.: Identification of Common Methods Used for Ontology Integration Tasks. Proceedings of the first international ACM workshop on Interoperability of Heterogeneous Information Systems, ACM (2005)
7. SEAMLESS project, Small Enterprises Accessing the Electronic Market of the Enlarged Europe by a Smart Service Infrastructure, <http://www.seamless-eu.org> (2008)
8. The Eclipse Project, <http://www.eclipse.org> (2008)