

OASIS — A New Tool for the Transformation of XML Knowledge Resources into OWL

Pavel Smrž and Martin Povolný and Anna Sinopalnikova

Faculty of Informatics, Masaryk University Brno

Botanická 68a, 602 00 Brno, Czech Republic

E-mail: {smrz,xpovolny,anna}@fi.muni.cz

Abstract

This paper presents OASIS – a new tool that enables (semi)automatic conversion of existing knowledge bases, semantic networks, terminological databases and various other resources to complex ontologies into OWL. The tool is implemented as a client of DEB (Dictionary Editor and Browser) which is able to store, index and efficiently retrieve lexical data. The architecture is based on XML and related W3C standards (XSLT, XML Schema, XPath, DOM).

The main feature which brings the efficiency of the transformation is the extension of a standard XSLT processor with the ability to obtain additional data from the server through the mechanism of nested queries. This technique allows formulation of complex constraints needed in the conversion to OWL.

1 Introduction

The creation of an ontology presents an extremely labor-intensive task when prepared from scratch. The most promising approach in the development of large ontologies is therefore the effort to clean-up, refine and merge the existing resources, such as WordNet (<http://www.cogsci.princeton.edu/wn/>).

This paper presents OASIS — a new system that is developed in the Natural Language Processing Laboratory, Faculty of Informatics, Masaryk University, Brno, Czech Republic, and that is intended to support the transformation of available resources into OWL. Existing knowledge bases, semantic networks, terminological databases, etc. can be (semi)automatically converted to the appropriate representation and the created data can be further edited in OASIS or in specialized ontology editors such as Protege.

The XML-family standards — XML Schema, XSLT, DOM play a crucial role in OASIS. We take XML as a tool for the whole process of ontology creation. The incorporation of the standards can bring the advantage of general applicability of the implemented system for various kinds of data and also easy extensibility of such systems.

Our system implements special features for the efficient management of data organized in a possibly com-

plex network. The primary stimulus of our work has been the aim to provide a universal system for efficient transformation of WordNet-like databases into OWL.

2 System Architecture

The overall architecture of the system is presented in Figure 1. OASIS is implemented as a client of DEB Server.

The DEB server is responsible for the storage and retrieval of data. DEB is based on a specialised module for data storage, on FINLIB text indexing library [Rychlý, 2000] for the conversion of data into a binary format and FININDEX for efficient retrieving. It is in contrast to other systems that do not implement own storage mechanisms. Our system is therefore independent of external tools such as relational databases.

The user can modify the data view by supplying their own XSLT sheet. It gives DEB clients an additional level of flexibility. Last, but not least, most of the described features can be included in a “thin” dictionary client application accessible by standard web browsers.

OASIS uses XSLT for transforming selected parts of existing resources into OWL and also to transform arbitrary data into HTML for visualization purposes. (with the help of a HTML widget). The tool also benefits from the client-side caching of parsed entries in DOM [Hégaret, 2002] and from the use of XPath [Clark and DeRose, 1999] for extraction of important parts of the entries.

3 Transforming Data to OWL

To be able to perform even complex queries, e.g. the transitive closure of a relation, OASIS takes advantage of an extension of the XSLT processor which allows additional queries from the XSLT processor. XSLT sheets can contain request data from the server based on the entry being processed.

The creation of such an extension is possible even without breaking the rules of the XSLT language. We can register a new schema handler within the XSLT processor (by schema we understand the part of URI before the double slash, such as `http`, `ftp`, `file`, etc.). The schema creates a virtual space of XML documents which are results of the queries. From the

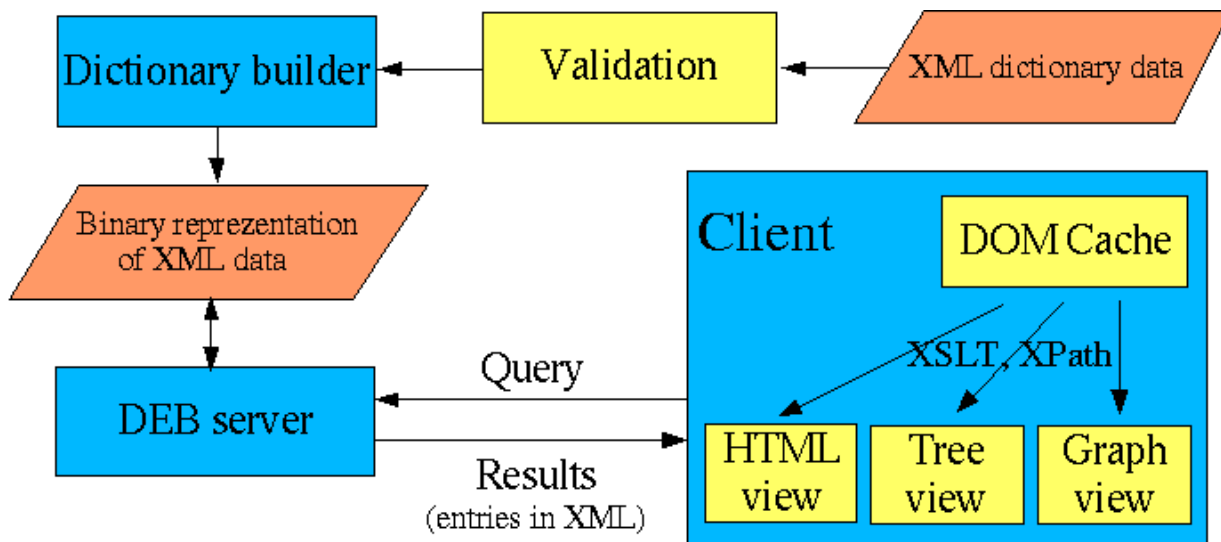


Figure 1: Schema of OASIS data interchange



Figure 2: XSLT processor running queries nested in the XSLT script

XSLT processor point of view, accessing the server data is the same as accessing any other external resources.

The dataflow schema which has been described in detail in [Smrz and Povolny, 2003] is shown in Figure 2.

The same mechanism is used for transformations of relations selected from available resources into their OWL representation. The process can be divided into the following steps:

1. User selects the relevant data in the OASIS client;
2. OASIS runs one of the parametrized predefined XSLT scripts based on the retrieved data;
3. The resulting part of the ontology in OWL can be visualised.

4 Conclusions

OASIS is able to transform selected data from different resources into OWL. It takes advantage of an extension of a XSLT processor with the ability to get additional data by querying dictionary server DEB.

In our future work on the server side we would like to implement XPath evaluation inside the dictionary server to make its interface closer to W3C standards. Another task is to make the server work metadata. We will also consider the implementation of some data versioning features in the server.

We would also like to separate the core functionality from OASIS and make it a separate layer in order to get a three-level architecture which would further ease

the development of specialized applications for thin clients such as WWW browsers.

Current work on standardization of language resources [Lenci and Ide, 2002] focuses on creation of unified format for representation of multilingual lexical data. We would like to offer OASIS as a standard tool in this context.

References

- [Clark and DeRose, 1999] James Clark and Steve DeRose. XML Path Language (XPath) Version 1.0, 1999. <http://www.w3.org/TR/xpath>.
- [Hégaret, 2002] Philippe Le Hégaret. Document Object Model (DOM) Bindings, 2002. <http://www.w3.org/DOM/Bindings>.
- [Lenci and Ide, 2002] Alessandro Lenci and Nancy Ide. The MILE Lexical Model, linguistic and formal architecture, 2002. ISLE Workshop.
- [Rychlý, 2000] Pavel Rychlý. *Corpus Managers and Their Effective Implementation*. PhD thesis, Faculty of Informatics, Masaryk University, Brno, 2000.
- [Smrz and Povolny, 2003] Pavel Smrz and Martin Povolny. DEB — Dictionary Editing and Browsing. In Nancy Ide, Laurent Romary, and Graham Wilcock, editors, *Proceedings of the 3rd Workshop NLPXML 2003, EACL 2003 Workshop*, pages 49–55, Budapest, Hungary, 2003.