OPTIMA: Tool for Ontology Alignment with Application to Semantic Reconciliation of Sensor Metadata for Publication in SensorMap *

Ravikanth Kolli *and* Prashant Doshi
LSDIS Lab, Dept. of Computer Science
University of Georgia
Athens, GA 30602
{kolli,pdoshi}@cs.uga.edu

Abstract

We present Optima*, a state of the art general purpose tool for ontology alignment that automatically identifies and matches relevant concepts between ontologies. The tool is supported by an intuitive user interface that facilitates the visualization and analysis of ontologies in N3, RDF and OWL and the alignment results. We will demonstrate the usefulness of* Optima *toward the automatic publication of sensor net data feeds in public portals such as Microsoft SensorMap.*

1 Introduction

Instead of a central repository of ontologies, we are witnessing the growth of disparate communities of ontologies that cater to specific applications. Naturally, many of these communities contain ontologies that describe the same or overlapping domains but use different concept names and may exhibit varying structure. Because the development of these ontologies is occurring in a decentralized manner, the problem of matching similar ontologies – often called *alignment* – gains importance.

We have developed Optima – a general purpose tool for automatically aligning concepts and relationships residing in multiple ontologies that address similar domains. Optima predominantly relies on the ontology schemas to efficiently infer matches between the elements and utilizes instances, if available, to refine the matches. It discovers many-one matches between the ontologies thereby allowing multiple concepts of an ontology to be each mapped to a single concept in the other ontology.

A tool related to Optima is OLA[1] that utilizes lexical similarity between concepts to align ontologies (in OWL only). While OLA provides a graphical visualization of the ontologies, it does not allow a visualization of the alignment. In comparison, AlViz [1] highlights clusters of similar concepts, but lacks a sophisticated alignment algorithm and does not show details about which individual concepts are matched. A somewhat related tool is Promptviz that provides a visual representation of the differences between versions of the same ontology. As we outline below, Optima incorporates features that improve on these tools.

*This work is supported by award no. 16048 from Microsoft Research SensorMap 2007 RFP.

[1] http://www.iro.umontreal.ca/~owlola/visualization.html

2 Architecture

Alignment Module. At the core of Optima is a novel graph-theoretic algorithm that formulates the problem of alignment as one of finding the most likely match between two ontologies (optimization), and computes the likelihood using the expectation-maximization (EM) technique. Within the iterative EM approach, we exploit the structural as well as the lexical similarity between the schemas to compute the likelihood of a match. In particular, we utilize the neighborhood of a node in the ontologies to infer a match. As the complete space of candidate alignments tends to be large and to avoid local maximas, we randomly sample a representative set of alignments and augment it with locally improved estimates that exploit general matching heuristics. See [2] for more details on the alignment method and favorable performance evaluations.

RDF and OWL Parser. Optima uses the well-known open source library called *Jena* to parse ontologies expressed in the popular languages of N3, RDF and OWL. *Jena* creates an in-memory representation of the parsed ontology and provides a comprehensive API to access and manipulate the ontology model.

3 Demonstration

Our demonstration of Optima will be interactive: users will be encouraged to obtain a "hands-on" ex-

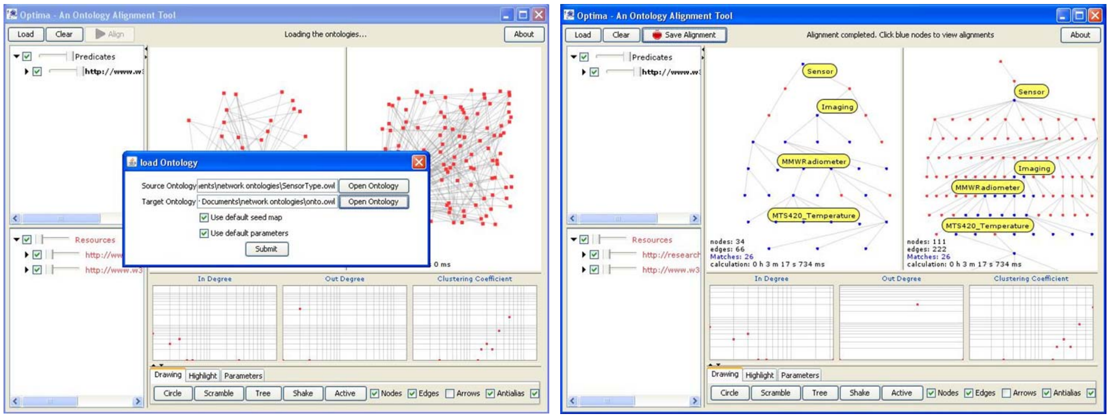

Figure 1. Screenshots of Optima's user interface showing two sensor ontologies loaded [left], and results of running the alignment with some of the aligned nodes (in blue) highlighted [right].

perience with the tool. Optima targets the specific application of reconciling metadata on sensor data feeds with the ontologies of sensor data publishers such as Microsoft SensorMap [3], to facilitate automatic publication. Thus, we will use sensor net ontologies to demonstrate its functionality and applicability.

3.1 Tool Description

Optima is implemented as a Java (v5) based application capable of executing on both Windows and Linux platforms, and directly off the Web using Java Web Start. A screencast of Optima is available at *http://lsdis.cs.uga.edu/projects/sensormap/*. Optima's user interface builds on an open source ontology browser called Welkin which is a part of MIT's Simile project. As we show in Fig. 1, Optima provides an intuitive interface that allows the user to load and view the ontologies. To facilitate browsing the ontologies, several different layouts of the ontology graphs such as *tree*, *circle*, *scramble* (randomly generated locations) and *active* are available. In addition, nodes may be clustered and those belonging to different namespaces may be filtered out from the display to reduce clutter. On performing the alignment, nodes of both the ontologies that are matched are highlighted and the user may select any of these nodes to identify its corresponding match in the other ontology. The progress of the alignment task is shown to the user and may range from a few seconds to hours depending on the size and complexity of the ontologies. An option to *save* the discovered alignment in XML is provided as well. To facilitate or steer the alignment, users may initially enter *seed matches* by selecting pairs of nodes from the ontologies.

3.2 Application

Identification and categorization of metadata on incoming sensor data feeds is a key step for automatic publication and querying of streaming data on Web based portals such as the SensorMap [3]. This is because the type of the streaming data often determines how the data is rendered and which are the other data feeds with which it interacts. This motivates the data publishers to develop registration mechanisms for obtaining the type and other metadata on the streaming data feeds (e.g., SensorMap utilizes Web services) and to develop taxonomies of the possible sensor types and other metadata. Data providers are compelled to use the, possibly sophisticated and tedious, mechanisms for providing the metadata for their sensor feeds.

Our investigation reveals that several data providers develop and curate detailed metadata on their feeds, which often assumes the form of ontologies in RDF or OWL. We aim to link the provider-ontologies with those of the publishers, thereby automating the step of metadata registration and identification, thus reducing the burden on both the data providers and publishers.

3.3 Examples

We have acquired sensor net ontologies from researchers involved with deployed sensor networks and those publishing their sensor data on SensorMap. We will utilize these real-world ontologies in conjunction with SensorMap's SensorType ontology as examples to demonstrate automatic metadata reconciliation using Optima.

References

[1] M. Lanzenberger and J. Sampson. AlViz - A Tool for Visual Ontology Alignment In *IV*, pp. 430-440, 2006.

[2] P. Doshi and C. Thomas. Inexact matching of ontology graphs using expectation-maximization. In *AAAI*, pp. 1277-1282, 2006.

[3] A. Santanche, S. Nath, J. Liu, B. Priyantha, and F. Zhao. Senseweb: Browsing the physical world in real time. In *Demo Abstracts, IPSN*, 2006.