GoodRelations Tools and Applications

Martin Hepp, Andreas Radinger, Andreas Wechselberger, Alex Stolz,
Daniel Bingel, Thomas Irmscher, Mark Mattern, and Tobias Ostheim
E-Business & Web Science Research Group, Universität der Bundeswehr
Werner-Heisenberg-Weg 39, D-85579 Neubiberg, Germany
+49-89-6004-4217

mhepp@computer.org; firstname.lastname@ebusiness-unibw.org

ABSTRACT
The adoption of ontologies for the Web of Data can be increased by tools that help populating respective knowledge bases from legacy content, e.g. existing databases, business applications, or proprietary data formats. In this demo and poster, we show the results from our efforts of developing a suite of open-source tools for creating e-commerce descriptions for the Web of Data based on the GoodRelations ontology. Also, we demonstrate how RDF/XML data can be (1) submitted to Yahoo SearchMonkey via the RDF2DataRSS conversion tool, (2) inspected using the SearchMonkey Meta-Data Inspector, and (3) how common data inconsistencies can be spotted with the GoodRelations Validator.

Categories and Subject Descriptors
I.2.4 **[Artificial Intelligence]:** Knowledge Representation Formalisms and Methods

General Terms
Human Factors, Languages, Economics

Keywords
E-Commerce, GoodRelations, RDF, Google, CMS, Joomla, Virtuemart, osCommerce

1. INTRODUCTION
The GoodRelations ontology [1] provides a conceptual model for a consolidated view on commerce data on the Web, e.g., companies, store locations, offers, product descriptions, pricing, payment, shipment, and warranty information. Since its official release in August 2008, it has gained substantial popularity and is supported by e.g. Yahoo and BestBuy. At the time of writing, there are more than 1,5 million non-toy product model descriptions, 450k individual offers with price information, and 50 k company profiles available. In addition, the popular RDF Book Mashup[1] now exposes Amazon and eBay offers of new and used copies for more than 100k book titles as GoodRelations data.

However, a huge amount of data alone is not sufficient for bringing real value-added by Semantic Web technology, as long as most of the data originates from the inclusion of just a few data sources that are already highly structured and maintained centrally, via T-Box mapping. In other words, it is not sufficient to tap a few large databases and republish them on the Web of Linked Data. Instead, it is equally important to get broad adoption, i.e. many smaller datasets from many contributors, because only this brings the specificity and diversity to the corpus of linked data on the Web which is needed to materialize the full potential. It's not about mashing up a couple of large databases – the true power of linked, structured data on the Web requires millions of content creators to add respective meta-data.

In the realm of e-commerce, that means we must empower any small shop operator in the world to add a growing amount of specific and current meta-data to his or her Web page. Contrary to common belief, such meta-data goes quickly beyond the simple level of linear property-value pairs. While attaching `dc:title` and `foaf:fn` properties to string values may be doable for average site owners and Web developers, this approach becomes unsuited as soon as more sophisticated vocabularies are used. The latter will benefit from more comprehensive tool support, e.g. export and import interfaces for popular application software and form-based online-tools for lay audiences. In the case of FOAF, a large share of available instance data originates from form-based tools like FOAF-a-Matic [2] or built-ins to application software.

In this paper, we give an overview of tools developed in the context of the GoodRelations ontology. Some of the tools presented, namely the RDF2DataRSS converter, can also be useful for other vocabularies.

2. DESCRIPTION OF THE DEMO
In the form of a "tools fair", we will show how the various tools can be used to generate RDF descriptions for a given business from various sources, e.g. (1) form-based data entry for small businesses who need only lightweight descriptions, (2) converting feeds for Google products search from the Web or from popular shop software, or (3) existing XML catalog documents available in the popular BMEcat 2005 format.

Next, we will demonstrate how the resulting data can be published as either RDF/XML, Yahoo DataRSS feeds, or RDFa, and how DataRSS can be submitted to Yahoo using the Yahoo Site Explorer or OPML bulk submissions [3, 4].

Lastly, we will show how the data will be visible for humans and machines in various value-added services. We will use the SearchMonkey Meta-Data Inspector to check the inclusion of a particular feed in the SearchMonkey index. Also, we will show three example Web shops based on osCommerce, Oxid eSales, and Joomla/Virtuemart with enabled GoodRelations export.

3. TOOLS FOR CREATING DATA
In this section, we will briefly describe the tools for creating GoodRelations offer and product data. The tools can be divided into two groups: The first is a form-based tool that helps owners of any business in the world produce a basic yet detailed description of its range of products or services. The second group are stand-alone tools or extensions for application software that help users create RDF descriptions for all items and the business itself from existing structured data, e.g. the product information in shop software.

[1] http://www4.wiwiss.fu-berlin.de/bizer/bookmashup/

3.1 GoodRelations Annotator

Inspired by the impact of the FOAF-a-Matic tool [2] on the availability of FOAF data on the Web, we developed a form-based tool that can be used by any business in the world to create a basic yet detailed description of its range of products, payment and delivery options, store locations, opening hours, and eligible customer types and eligible regions. The tool is available at

http://www.ebusiness-unibw.org/tools/goodrelations-annotator/

Internally, it uses a HTML form with a substantial amount of JavaScript for validating the input, and a Python server-side component that handles the conversion and returns RDF/XML (in the future also RDFa snippets) plus instructions on how to publish the data.

3.2 Google Product Feed Converter

A huge number of online merchants are already creating data feeds of their items and offers for Google, because such is needed for getting the items listed in Google's products search. Many popular Web shop applications also contain a respective export functionality. Google accepts multiple formats, i.e. RSS 1.0en, RSS 2.0de, Atom 0.3de, Atom 1.0en, and Tab-delimited (en). The feed formats are not as detailed as GoodRelations and do not contain all the information needed for a useful self-description, because they are meant for manual submission, not anonymous publication on the Web. Accordingly, the tool asks the user to provide a few additional details, which are then used for the conversion. The tool is available at

http://www.ebusiness-unibw.org/tools/google-product-feed-converter/

3.3 Joomla/Virtuemart Extension

Joomla/Virtuemart is a popular CMS plus a shopping cart extension. GoodRelations4Joomla is an extension that activates RDFa and RDF/XML export of the offer and item details. The tool is available at

http://code.google.com/p/goodrelations-for-joomla/.

A demo shop is at http://www.stalsoft.com/dip/joomla/.

3.4 osCommerce Shop and Extension

osCommerce is a free Web shop application with an estimated 15,000 installations worldwide. GoodRelations4osCommerce is an extension that activates RDFa and RDF/XML export of the offer and item details. The tool is available at

http://code.google.com/p/goodrelations-for-oscommerce/

A demo shop is at

http://www.stalsoft.com/dip/oscommerce/catalog/index.php.

3.5 BMEcat2GoodRelations

BMEcat 2005 [5] is one of the major XML-based exchange formats for catalog data interchange, mainly in B2B business relationships. Also, major shop and product information management (PIM) applications can create BMEcat data. BMEcat2GoodRelations is an online tool for converting a given catalog document into RDF/XML. The tool is available at:

http://www.ebusiness-unibw.org/tools/bmecat2goodrelations/

4. RDF2DATARSS CONVERTER

Yahoo SearchMonkey considers RDF data only when provided either as RDFa embedded directly into the crawled Web page, or via dedicated data feeds in the proprietary Yahoo feed format "DataRSS" [3]. Since RDFa is often not an option for site owners (e.g. because modifications on the live system can be subject to heavy restrictions), DataRSS is frequently the only way to feed data into the SearchMonkey index. In order to empower companies to submit their RDF data to Yahoo, we created a tool that converts any RDF/XML or N3 file into that feed format. The tool is available at

http://www.ebusiness-unibw.org/tools/rdf2datarss/.

The tool is also available as a Web service in a REST-style manner. Large amounts of feed files can be submitted to Yahoo by means of OPML (Outline Processor Markup Language) documents, see [3] and [4].

5. GOODRELATIONS VALIDATOR

In order to help site owners validate the correctness of their descriptions, we are in the process of completing a validation tool which (1) checks syntactical correctness, (2) compliance with the formal account of the GoodRelations ontology, and (3) searches for typical mistakes or contradictions in commerce data. Internally, it processes a list of SPARQL queries that search for problematic patterns and lists their URI plus a description of the problem. For example, it searches for expired offers as follows:

```
PREFIX gr:<http://purl.org/goodrelations/v1#>
SELECT ?o, ?date WHERE
{?o rdf:type gr:Offering. ?o gr:validThrough ?date.
FILTER (?date < "2009-08-07T00:00:00Z"^^xsd:dateTime)
}
```

The tool is currently under development at

http://www.ebusiness-unibw.org/tools/goodrelations-validator/

6. SEARCHMONKEY INSPECTOR

As a quick tool for checking whether user-provided RDF data is properly included in the Yahoo SearchMonkey index, we developed a simple tool that imitates a search in Yahoo and shows RDF data, if available, as N3.

http://goodrelations-search.appspot.com/

Hint: When trying the tool, enter "wwwurl:http://www.heppnetz.de/searchmonkey/product.html" as the query string.

Acknowledgments: We would like to thank Peter Mika and Evan Goer from Yahoo for valuable feedback on the RDF2DataRSS tool.

REFERENCES

[1] Hepp, M.: GoodRelations: An Ontology for Describing Products and Services Offers on the Web. Proceedings EKAW2008, Springer LNCS Vol. 5268, (2008) pp. 332-347.

[2] FOAF-a-Matic, http://www.ldodds.com/foaf/foaf-a-matic.

[3] Creating the Feed (Yahoo! SearchMonkey Guide), http://developer.yahoo.com/searchmonkey/smguide/creating_feeds.html.

[4] OPML: Outline Processor Markup Language, available at http://www.opml.org/.

[5] Schmitz, V., Leukel, J., Kelkar, O.: Specification BMEcat 2005. Stuttgart/Essen (2005).