

GR2RSS: Publishing Linked Open Commerce Data as RSS and Atom Feeds

Technical Report

TR-2014-1

Version: October 19, 2014

Alex Stolz, alex.stolz@unibw.de

Martin Hepp, martin.hepp@unibw.de

E-Business and Web Science Research Group,
Universität der Bundeswehr München

Werner-Heisenberg-Weg 39, D-85579 Neubiberg, Germany

GR2RSS: Publishing Linked Open Commerce Data as RSS and Atom Feeds

Alex Stolz and Martin Hepp

E-Business and Web Science Research Group, Universität der Bundeswehr München
Werner-Heisenberg-Weg 39, D-85579 Neubiberg, Germany
{alex.stolz,martin.hepp}@unibw.de

Abstract. The integration of Linked Open Data (LOD) content in Web pages is a challenging and sometimes tedious task for Web developers. At the same moment, most software packages for blogs, content management systems (CMS), and shop applications support the consumption of feed formats, namely RSS and Atom. In this technical report, we demonstrate an on-line tool that fetches e-commerce data from a SPARQL endpoint and syndicates obtained results as RSS or Atom feeds. Our approach combines (1) the popularity and broad tooling support of existing feed formats, (2) the precision of queries against structured data built upon common Web vocabularies like schema.org, GoodRelations, FOAF, VCard, and WGS 84, and (3) the ease of integrating content from a large number of Web sites and other data sources in RDF in general.

1 Introduction

Despite the growing amount of structured data on the Web, the useful integration into Web pages still lags behind opportunities. In the field of e-commerce, major retail sites like *sears.com*, *bestbuy.com*, *wayfair.com*, *rakuten.de* and numerous smaller shops have added RDFa and Microdata markup to their page templates, exposing more than 30 million offers that are updated on a daily basis. A couple of SPARQL endpoints already collate this structured e-commerce data. However, fetching useful information typically involves contacting the right endpoints, crafting proper SPARQL queries, and eventually converting results into data formats understood by target applications. Average Web developers and site owners quickly get overwhelmed by the technical challenges imposed by these tasks. At the same time, there exist popular data formats for publishing dynamic content on the Web, namely RSS [4] and the Atom Syndication Format [3,5]. Both feed formats have excellent tool support; i.e., most software packages for blogs, content management systems (CMS), and shopping carts provide integration capabilities for external sources using RSS or Atom.

In this technical report, we show an approach that combines the broad tooling support of established data feed standards with the precision of queries

against structured data collected from multiple Web sites and RDF data sources, built upon common Web vocabularies such as schema.org¹, GoodRelations [1], FOAF², VCard³, and WGS 84⁴.

2 GR2RSS Tool

The on-line tool⁵ that we have developed fetches GoodRelations e-commerce data from a SPARQL endpoint and syndicates obtained results as RSS or Atom feeds. Fig. 1 outlines the general system architecture of our tool. The generated data feeds serve as carriers to facilitate the consumption and integration of linked open commerce data in Web pages. This way site owners have a means to add relevant, dynamic content to their Web pages, thereby attracting visitors and improving search engine rankings. Conversely, product vendors gain additional visibility of their products for free, since items republished by virtue of site owners and bloggers link back to shops where the products are actually offered.

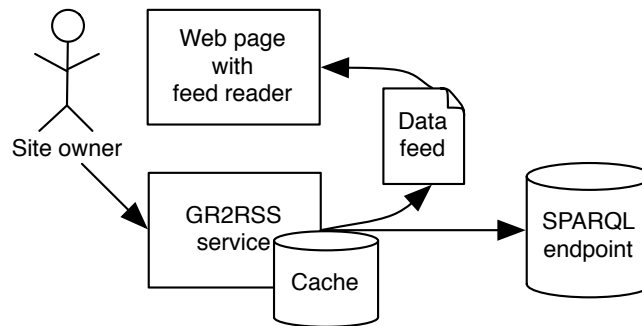


Fig. 1. Conceptual architecture of the GR2RSS on-line service

In the following, we summarize the core technical contributions of our on-line service:

1. *Query builder*: The tool supports three different levels of expertise. Depending on the selected search mode, a more or less sophisticated query builder is presented to the user. The following three search modes are available:
 - *Basic*: Single input field for keyword searches.
 - *Extended*: Query builder with support for filtering and sorting results, price currency conversion for products, and location-aware store searches.

¹ <http://schema.org/>

² <http://xmlns.com/foaf/spec/>

³ <http://www.w3.org/TR/vcard-rdf/>

⁴ <http://www.w3.org/2003/01/geo/>

⁵ <http://www.ebusiness-unibw.org/tools/gr2rss/>

- *Expert*: Input field for entering raw SPARQL queries and a set of possible variables which bindings can be processed by the feed templates.
- 2. *Prefetching and caching*: In order to limit load at the SPARQL endpoints, we implemented server-side caching of the generated feeds. Our caching mechanism stores records of successfully executed queries in a local database management system for serving future requests and creates corresponding cache files. After a given cache period (e.g. a day), future requests trigger cache invalidations whereby the cached files are replaced by freshly generated feed content.
- 3. *Geo information*: RSS and Atom feed formats provide extension mechanisms to support custom vocabularies. We used the GeoRSS⁶ vocabulary to include positional data in data feeds. These annotations allow, for instance, to extract and display location data on a map, e.g. on *Google Maps*.
- 4. *Viral use of RDFa*: Based on the idea of embedding RSS and Atom feeds in blog systems and CMS, we decided to piggyback RDFa statements as entity-encoded HTML in feed entries [7]. Thereby we can obtain a viral publication effect, because RDFa preserves the URIs of the original entities and thus prevents the proliferation of identifiers. In other words, any Web page integration of feed content contributes to the promotion of the product offers at the origin. Moreover, we employ *foaf:page* links to provide a means for tracking the document URI at which the particular content reappears.
- 5. *Currency conversion*: The currency conversion at the endpoint is realized using a materialization of exchange rates based on the Exchange Rate Ontology⁷ (XRO). These currency exchange rates need to be available before any currency conversion task in SPARQL can take place. A service based on XRO that provides regularly updated exchange rates in RDF is available at <http://www.currency2currency.org/> [6].

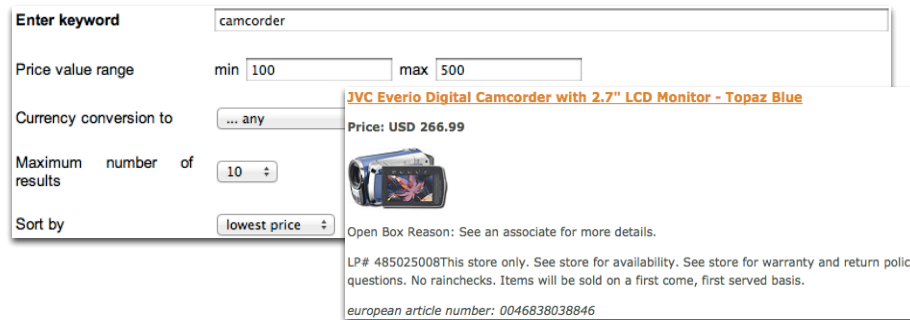
Our on-line tool was written in PHP and uses Javascript for user interaction. It runs over Linked Open Commerce data stores, and is compatible with Virtuoso SPARQL endpoints that support the *bif:contains* feature, a built-in function that executes over a full-text search index. For the future, we are planning to rewrite parts of the code in order to make the tool SPARQL-1.1-compliant.

3 Demonstration

In the following, we demonstrate an example of integrating a generated feed into a Web page. Suppose that we are looking for products within a price range of 100 and 500 dollars that contain the keyword “camcorder” and include product pictures. The populated form fields of the query builder are depicted in Fig. 2. The figure also shows a single camcorder item after integrating the feed into a Web page. Similar examples can be found in the tool documentation⁸ and

⁶ <http://georss.org/>

⁷ <http://purl.org/xro/>



Enter keyword: camcorder

Price value range: min 100 max 500

Currency conversion to: ... any

Maximum number of results: 10

Sort by: lowest price

JVC Everio Digital Camcorder with 2.7" LCD Monitor - Topaz Blue
 Price: USD 266.99

Open Box Reason: See an associate for more details.

LP# 485025008This store only. See store for availability. See store for warranty and return policy questions. No rainchecks. Items will be sold on a first come, first served basis.

europaean article number: 0046838038846

Fig. 2. Query builder for *camcorder* search and Web page integration of the feed example page⁹.

4 Related Work

We compared our work with existing approaches, namely (1) single feed definition dialogs, as offered by major sites (e.g. eBay and Amazon), and (2) feed aggregation services, i.e. Yahoo Pipes¹⁰ and DERI Pipes [2]. The former approaches typically fail at integrating different data sources (e.g. *list five cheapest offers among Amazon and eBay feeds*), whereas aggregation services are limited to filter results by brittle regex-based expressions (e.g. *show only shops in New York*) and lack simple unit conversion (e.g. *display all prices in euros*) (cf. [7]).

5 Conclusions

The presented on-line service aims to address the issue of Linked Open Data (LOD) content integration in Web pages. For this purpose, it generates RSS and Atom feeds from structured e-commerce data fetched from a SPARQL endpoint, thereby exploiting the excellent tool support for content syndication formats. The service allows for different levels of query building assistance, implements caching, incorporates geo-location data and RDFa annotations, and employs currency conversion at the endpoint. We consider our approach to be of similar use for other fields of Linked Open Data outside the narrow scope of e-commerce.

References

1. Hepp, M.: GoodRelations: An Ontology for Describing Products and Services Offers on the Web. In: Proceedings of the 16th International Conference on Knowledge Engineering (EKAW 2008). pp. 329–346. Acritezza, Italy (2008)

⁸ <http://www.ebusiness-unibw.org/tools/gr2rss/docs>

⁹ <http://www.ebusiness-unibw.org/tools/gr2rss/examples>

¹⁰ <http://pipes.yahoo.com/>

2. Le-Phuoc, D., Polleres, A., Hauswirth, M., Tummarello, G., Morbidoni, C.: Rapid Prototyping of Semantic Mash-ups through Semantic Web Pipes. In: Proceedings of the 18th International World Wide Web Conference (WWW 2009). pp. 581–590. Madrid, Spain (2009)
3. Nottingham, M., Sayre, R.: The Atom Syndication Format (2005), <http://tools.ietf.org/html/rfc4287>
4. RSS Advisory Board: RSS 2.0 Specification, <http://www.rssboard.org/rss-specification>
5. Sayre, R.: Atom: The Standard in Syndication. IEEE Internet Computing 9(4), 71–78 (2005)
6. Stolz, A., Hepp, M.: Currency Conversion the Linked Data Way. In: Proceedings of the First Workshop on Services and Applications over Linked APIs and Data (SALAD2013). pp. 44–55. CEUR Workshop Proceedings, Montpellier, France (2013)
7. Stolz, A., Hepp, M.: From RDF to RSS and Atom: Content Syndication with Linked Data. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media (Hypertext 2013). pp. 236–241. ACM, Paris, France (2013)