STI Community – An approach for a semantically enhanced company platform

Carmen Brenner, Andreas Thalhammer, and Dieter Fensel

STI Innsbruck Technical Report

September 11, 2012

This technical report will introduce the background and the goal of the STI International Community platform, an open online platform with the goal to collect information on company and business entities that are active in the field of semantic technologies. Based on the ideas of Crowdsourcing the data is created and enriched by the semantic community itself therefore introducing the community's knowledge into the dataset. The platform is accessible at http://community.sti2.at.

Introduction

The interest in and the number of applications enriched by semantic technology have grown fast in the last years. No longer are terms like Semantic Web, Big Data or Linked Open Data only associated with research, but semantic technology is on the cusp of mainstream adoption and represents a huge potential for advances in all areas of ICT. But it has also become more and more difficult for interested parties to keep track of all these companies that try to bring semantic technology into our daily lives with a variety of products. Information on companies that are active in the field of semantic technologies is usually diverse and widely distributed, and hard to get by. Additionally, in many cases, the data on a company or a product is not provided in a machine-readable format like RDF, but rather in an only human-readable way.

Out of this dilemma the idea was born to create a list of companies that are active in the field of semantic technology in order to have a central point where interested parties can easily browse for companies and their products. We initiated the project in December 2010 collecting information on companies working in the semantic technology field and the first draft was publicly announced to the semantic technology community on December 20, 2010 over the W3C Semantic Web mailing list[1]. The feedback received was huge and mostly very positive, and a great number of companies were proposed to be added to the list[2].

---

[1] http://lists.w3.org/Archives/Public/semantic-web/

[2] The whole conversation on the W3C Semantic Web mailing lists is available at: http://lists.w3.org/Archives/Public/semantic-web/2010Dec/0190.html

Such change and update requests to the list were numerous, which showed that a simple list that only offers a short description of the companies with a link to their Web sites as it started out at the beginning was simply not enough. There was demand to allow the semantics community to easily add new companies or edit existing profiles, and therefore maintain their public data themselves.

The idea to build an open online platform was therefore initiated in co-operation with STI Research and Consultancy[3]. The focus and vision of STI Research and Consultancy is to ensure that semantics will become part of everyday life, and such an open online platform was a huge step in supporting the community.

The platform structure

This section will shortly describe the structure of the platform including the technologies used and the vocabulary.

The STI Community platform was designed to be an online platform with the goals

1. To collect data on companies that are active in the field of semantic technologies,

2. To support developers and vendors of semantic technology in their wish to position their businesses and products in the Internet, and

3. To give users who are searching for semantic applications and semantic technology vendors a place to start their search.

[3] http://research.sti2.org

The term "open" in this context refers to the concept of Crowdsourcing[4] where the profiles of companies are added and maintained by the community.

Technologies

The STI International Community platform makes by itself heavy use of semantic technology. All information on companies is stored as data triples in an underlying triple store. For this purpose we use OWLIM[5], a well-known semantic repository that supports the load of tens of billions of RDF statements, uses non-trivial inference and delivers outstanding multi-user query performance. It is entirely developed in Java and uses Forest[6], a lightweight web interface that brings together tools for execution of SPARQL queries, visual exploration of RDF triples, and full text. Forest allows easy connectivity to the underlying OWLIM triple store via Sesame a de-facto standard framework for processing RDF data that includes parsing, storing, inferencing and querying of/over such data.

Vocabulary of the platform

The vocabulary used to describe companies is based on several existing and well-distributed vocabularies including RDF Schema[7], Friend of a Friend (FOAF)[8], Dublin Core[9] and GoodRelations[10], but has also a small vocabulary of its own where the category and the attributes of a company are specified. The

---

[4] http://en.wikipedia.org/wiki/Crowdsourcing

[5] http://www.ontotext.com/owlim

[6] http://www.ontotext.com/

[7] http://www.w3.org/TR/rdf-schema/

[8] http://www.foaf-project.org/

[9] http://dublincore.org/documents/dcmi-terms/

[10] http://www.heppnetz.de/ontologies/goodrelations/v1.html

mini-vocabulary of the STI International Community platform specifies the different categories and the properties a company can have.

Company categories (e.g. Vendor, User or Marketplace) are in this vocabulary defined as subclasses of RDF Schema's Organization. We have pre-defined the following company attributes: *companyProperty*, *description*, *logo*, *name* and *homepage. companyProperty* will used by the platform to describe all user generated company attributes.

In the following your will see the XML for the vocabulary as described above.

```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xml:base="http://community.sti2.at/company/"
  xmlns:dc="http://purl.org/dc/terms/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:stc="http://community.sti2.at/spec.rdf#"
  xmlns:xhv="http://www.w3.org/1999/xhtml/vocab#"
  xmlns:xml="http://www.w3.org/XML/1998/namespace"
  xmlns:owl="http://www.w3.org/2002/07/owl">


<owl:Ontology
      rdf:about="http://community.sti2.at/spec.rdf">
      <dc:title>Semantic Technology Companies vocabulary</dc:title>
      <dc:description>Semantic Technology Companies RDF vocabulary, described
using W3C RDF Schema and the Web Ontology Language.</dc:description>
</owl:Ontology>


<rdfs:Class rdf:about="http://community.sti2.at/spec.rdf#Vendor">
  <rdfs:isDefinedBy rdf:resource="http://community.sti2.at/spec.rdf"/>
  <rdfs:label>Vendor</rdfs:label>
  <rdfs:comment>A vendor of Semantic Technologies.</rdfs:comment>
```

```
    <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
</rdfs:Class>


<rdfs:Class rdf:about="http://community.sti2.at/spec.rdf#Marketplace">
  <rdfs:isDefinedBy rdf:resource="http://community.sti2.at/spec.rdf"/>
    <rdfs:label>Marketplace</rdfs:label>
  <rdfs:comment>A Semantic Data marketplace.</rdfs:comment>
  <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
</rdfs:Class>


<rdfs:Class rdf:about="http://community.sti2.at/spec.rdf#User">
  <rdfs:isDefinedBy rdf:resource="http://community.sti2.at/spec.rdf"/>
  <rdfs:label>User</rdfs:label>
  <rdfs:comment>A user of Semantic Technologies.</rdfs:comment>
  <rdfs:subClassOf rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
</rdfs:Class>


<rdf:Property rdf:about="http://community.sti2.at/spec.rdf#companyProperty">
  <rdfs:label>Super property of Semantic Technology Companies</rdfs:label>
</rdf:Property>


<rdf:Property rdf:about="http://purl.org/dc/terms/description">
  <rdfs:subPropertyOf
rdf:resource="http://community.sti2.at/spec.rdf#companyProperty"/>
</rdf:Property>


<rdf:Property rdf:about="http://xmlns.com/foaf/0.1/logo">
  <rdfs:subPropertyOf
rdf:resource="http://community.sti2.at/spec.rdf#companyProperty"/>
</rdf:Property>


<rdf:Property rdf:about="http://xmlns.com/foaf/0.1/name">
  <rdfs:subPropertyOf
rdf:resource="http://community.sti2.at/spec.rdf#companyProperty"/>
</rdf:Property>


<rdf:Property rdf:about="http://xmlns.com/foaf/0.1/homepage">
```

```
   <rdfs:subPropertyOf
rdf:resource="http://community.sti2.at/spec.rdf#companyProperty"/>

</rdf:Property>
```

Crowdsourcing the data

As stated before, the platform encourages the users to extend this vocabulary and therefore offers the possibility of adding new categories and new company attributes. New attributes and categories are modeled as RDF properties having certain domains, but as the determination of the range has proved to be a non-trivial task in other applications like DBpedia[11], we decided to use simple strings for the initial system. An extension on a later point is possible and planned.

A basic profile of a company represented in RDF/XML can look like this:

```
<stc:Vendor rdf:about="DA0ED2E7-F1E8-4464-AE72-88F2755F5BFE">

      <dc:description>Ontotext provides core semantic technology distinctive
for its performance, scale, and compliance with standards.

      Its semantic database OWLIM is proven to be the fastest and most
scalable RDF(S)/OWL engine. Ontotext develops KIM - the most popular semantic
annotation and search platform. WSMO Studio and wsmo4j are at the heart of the
recent developments in Semantic WS and BPM.

      Ontotext's technology delivers real-world applications in Life Sciences,
Web Search, BI, Telecommunications, Media Monitoring, Online Recruitment, and
other areas.</dc:description>

      <foaf:name>Ontotext</foaf:name>

      <foaf:homepage rdf:resource="http://www.ontotext.com/"/>

</stc:Vendor>
```

---

[11] http://www.dbpedia.org

The example describes the company *Ontotext*. It is categorized into the category *stc:Vendor* with the unique ID *DA0ED2E7-F1E8-4464-AE72-88F2755F5BFE*. It has a company description (*dc:description*), a company legal name (*foaf:name*) and a link to its Web site (*foaf:homepage*).

Accessing the data

The complete data set is accessible through a Web interface but also through a SPARQL endpoint. The Web interface of the community platform allows the user to browse through the list of company profiles, add new profiles or modify existing ones. An additional search interface with filtering options for different company categories simplifies the search for a specific company.

Users with experience in the query language SPARQL can use the provided SPARQL endpoint for more sophisticated queries, or explore the complete data set as RDF dumb.

The status

The current dataset of the platform includes the company profiles of around 500 companies from all over the globe that are active in the field of semantic technology. The data has been collected with the help of the community (e.g. over the W3C Semantic Web mailing list) and by investigating the data of contributors of research projects in the area of semantic technology.

Future Work

For the future, it might be interesting to allow users of the platform to introduce existing vocabularies into the data set. So instead of having *companyProperty* as an abstract description, users would be able to select specific attributes for company properties. E.g. the address of a company might be added as a Location in GoodRelations *gr:location*[12]. This seems to be no trivial task and thorough implementation.

Addressing recent problems with the technology behind the platform (specifically Forest), it seems to be reasonable to change the technologies used. It might be worth looking into the matter of using Drupal 7.x as a basis for a re-implementation of the STI Community platform.

Drupal 7.x is an open source content management system (CMS) with a big community that is very active in the semantic field. It offers out-of-the-box user management, logging, versioning and RDFa support, and can be easily extended e.g. with a SPARQL endpoint and other annotations. A big advantage would be that it is more easily maintained than the current platform implementation.

---

[12] http://www.heppnetz.de/ontologies/goodrelations/v1.html#Location