A Semantic-Based Platform for Efficient Online Communication

Zaenal Akbar, José María García, Ioan Toma, Dieter Fensel

Semantic Technology Institute, University of Innsbruck, Austria
firstname.lastname@sti2.at

Abstract. To achieve an effective and efficient way of disseminating information to ever growing communication channels, we propose an approach that separates the information and communication channels and interlinks them with an intermediary component. The separation enables various dimensions to reuse the information and communication channels in transactional communication. In this paper we introduce our online communication platform, which is comprised of several components. The important roles of semantic web technologies to the platform are explained in detail, including a use case to show the contributions of semantic web in supporting the effectiveness and efficiency of information dissemination.

Keywords: semantic web, online communication, platform, information dissemination

1 Introduction

In today's internet era, the number and kinds of information dissemination channels are growing exponentially and changing constantly. Websites, e-mails, blogs and social media have become the mainstream means of communication. Nevertheless, information dissemination is not only about finding suitable channels, but also fitting the content to the available channels. These are the main challenges for effective and efficient information dissemination, and for online communication in general.

Our solution to overcoming these challenges is to decouple information from channels, defining separate models for each of them, and then interlinking them with an intermediary component [1]. Semantic technologies play important roles in our solution: analysis and understanding of the natural language statements, information modeling and sharing with common vocabularies, matchmaking information and channels using a rules-based approach [2].

In this paper, we focus on the information modeling (including annotations) part such that the matchmaking of information to appropriate channels can be performed efficiently. First, we present the overall architecture, then we discuss how semantics contribute to the solution and finally we show a use case, followed by the conclusion and future works.

2 The Online Communication Platform

Shown in Fig. 1, the online communication platform consists of several components which are grouped based on their conceptual functions:

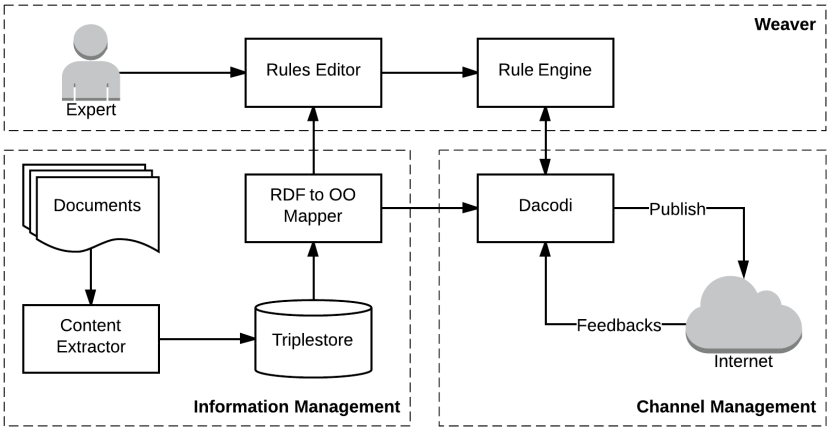

Fig. 1. The Online Communication platform architecture

Information Management is responsible for gathering the content from data sources (annotated and un-annotated) and representing them into the common vocabularies. First, the contents are extracted by a *Content Extractor* (implemented using Any23 [1]), then stored onto a *Triplestore* such as OWLIM [2]. Further, an *RDF to OO Mapper* (implemented using RDFBeans [3]) maps the stored triples onto object-oriented models to be used by the other components. For annotated sources where the sources have been annotated with the selected vocabularies, the content can be extracted automatically. For un-annotated sources, a manual mapping is required to inter-relate the database items (i.e. table fields) to relevant terms in the desired vocabularies.

Weaver is responsible for matching the information to appropriate channels through a rule based system. A *Rule Editor* enables experts to create and maintain rules through an integrated user interface and access-controlled rules repository. The rules are then matched to the facts in the working memory of the rule-based system by a *Rule Engine*. In our implementation we use Drools [4].

Channel Management is responsible for distributing the information to the selected channels according to the defined rules. *Dacodi* [5] offers various functionalities for distributing the content to the selected communication channels, as well as for collecting and analyzing feedback from those channels [3].

[1] `http://any23.apache.org`
[2] `http://www.ontotext.com/owlim`
[3] `http://rdfbeans.sourceforge.net`
[4] `http://drools.jboss.org`
[5] `http://dacodi.sti2.at`

3 Applying Semantic Technologies to Online Communication

Semantic technologies contribute mainly to content modeling, namely in how to obtain content from distributed and heterogeneous sources (i.e. through annotation) and represent them in a common representation to make an efficient match between information and desired channels possible. The matching is not between content sources but between the common representation to channels.

To achieve a reusable and interoperable information model, we selected vocabularies (whole or partial) from the Linked Open Vocabularies [6]:

1. Dublin Core [7], all metadata terms to support resource description
2. Friend of a Friend [8], a vocabulary to describe people, the links between them, the things they create and do
3. Good Relations [9], a vocabulary to describe e-commerce products and services
4. Schema.org [10], a collection of tags to markup a page in ways recognized by major search engines

These vocabularies are widely used, especially Schema.org which has been adopted by webmasters to increase their webpages' visibility in search engines.

We show these contributions in detail within the Tourismusverband (TVb) Innsbruck [11] use case. As one of the big tourism boards in Austria, its goal is to achieve the highest visibility possible in search engines as well as to be present in various social channels [4]. It has a lot of content types (i.e. Place, Event, Trip) to be disseminated to numerous channels (i.e. Facebook, YouTube).

a) The TVb Innsbruck content sources (i.e. Blog, services from touristic providers) were annotated with the selected terms of Schema.org.

```
<div itemscope itemtype="http://schema.org/Event">
  <span itemprop="name">Farmer's Market</span>
  <div itemprop="startDate" datetime="2014-07-14T07:00">
    14.07.2014 - 01.01.2015</div>
  <time itemprop="endDate" datetime="2015-01-01T12:00"/>
  <span itemprop="location" itemscope
    itemtype="http://schema.org/PostalAddress">Location:
    <span itemprop="streetAddress" content="Markthalle">
      Markthalle (Herzog-Siegmund-Ufer 1-3, AT-6020, Innsbruck)</span>
    <meta itemprop="streetAddress" content="Herzog-Siegmund-Ufer 1-3"/>
    <meta itemprop="addressRegion" content="Innsbruck"/>
    <meta itemprop="postalCode" content="6020"/>
    <meta itemprop="addressCountry" content="AT"/></span>
</div>
```

In this example, information about *Event* is annotated with the term `Event` from Schema.org by using microdata format [12].

[6] `http://lov.okfn.org`
[7] `http://dublincore.org`
[8] `http://www.foaf-project.org`
[9] `http://purl.org/goodrelations/`
[10] `http://schema.org`
[11] `http://www.innsbruck.info`
[12] `http://www.w3.org/TR/microdata/`

b) The publication rules were defined to guide the publication of extracted contents to selected channels.

```
rule "Event Publication Rule"
when item : Event()
then insert(new ItemToBePublishedIn(item, facebookWall))
     insert(new ItemToBePublishedIn(item, youtube))
end
```

In this rule, each time a new *Event* was found in the extracted contents, it was then prepared to be published to the `facebookWall, youtube` (instances of TVb's Facebook and YouTube accounts respectively).

4 Evaluation, Conclusions and Future Work

In order to evaluate our work, we compared the number of visitors to the TVb's website before and after annotating the content. Compared to the same period in 2013, the number of visitors increased by 8.63% between Jan-Feb 2014, which may be caused by the annotation. Also, the platform is currently being tested by 6 people at TVb as a substitution to their social media dissemination tool.

The platform was comprised of several components and used semantic web technologies to integrate various information sources, extracting and representing the content into common vocabularies to enable efficient matchmaking to appropriate channels using a rules-based approach. There are four vocabularies currently supported and in the future, we would like to add more vocabularies (i.e. Schema.org Action, SIOC [13]) to enhance the channel management, in order to improve the feedback collection, for example.

Acknowledgements We would like to thank all members of the OC working group [14] for their valuable feedback. This work was partly funded by the EU FP7 under grants no. 600663 (Prelida), 257641 (PlanetData) and 284860 (MSEE).

References

1. Fensel, A., Toma, I., García, J.M., Stavrakantonakis, I., Fensel, D.: Enabling customers engagement and collaboration for small and medium-sized enterprises in ubiquitous multi-channel ecosystems. Computers in Industry **65**(5) (2014) 891–904
2. Akbar, Z., García, J.M., Toma, I., Fensel, D.: On using semantically-aware rules for efficient online communication. In Bikakis, A., Fodor, P., Roman, D., eds.: Rules on the web. From theory to applications. Volume 8620 of LNCS. Springer (2014) 37–51
3. Toma, I., Fensel, D., Oberhauser, A., Fuchs, C., Stanciu, C., Larizgoitia, I.: Sesa: A scalable multi-channel communication and booking solution for e-commerce in the tourism domain. In: The 10th International Conference on e-Business Engineering (ICEBE). (Sept 2013) 288–293
4. Akbar, Z., Fensel, A., Fensel, D., Fuchs, C., Garcia, J., Juen, B., Lasierra, N., Stanciu, C.V., Toma, I., Tymaniuk, S.: Tvb innsbruck semantic pilot analysis. White paper, Semantic Technology Institute, University of Innsbruck (May 2014) `http://oc.sti2.at/TR/TVBInnsbruck`.

[13] `http://sioc-project.org`

[14] `http://oc.sti2.at`