Introducing a Diversity-Aware Drupal Extension

Simon Hangl, Ioan Toma, and Andreas Thalhammer

University of Innsbruck, Technikerstr. 21a, A-6020 Innsbruck
{simon.hangl, ioan.toma, andreas.thalhammer}@sti2.at

Abstract. This demonstration paper introduces a diversity-aware extension for the content management system Drupal. It shows how different aspects, such as automatically recognized entities, topics and sentiment scores can be leveraged in a Web user interface. We introduce a coherent approach that enables readers to navigate to further related articles. In particular, we demonstrate new ways to quickly grasp what the articles' sentiments are and which topics they cover before the actual click.

1 Introduction

Nowadays an impressive amount of data is being produced and consumed online each day introducing new challenges for technologies and tools that handle the information management life cycle, from filtering, ranking and selecting, to presenting and aggregating information. Furthermore, existing technologies and tools are based on principles that do not reflect the plurality of opinions and viewpoints captured in the information. Developing methods and software extensions to tools that leverage content analysis at large scale has become a necessity, which the RENDER project[1] is addressing. As a part of this contribution, we introduce a diversity-enabled Drupal module. Drupal is a very popular Content Management System (CMS) with – as of August 2013 – more than 983,000 users and more than 28,000 developers contributing.[2]

The *Diversity Enricher* Drupal extension has been developed as a show case for diversity-enabling technologies. It supports diversity-aware navigation, organization, and presentation of Drupal articles. A demo deployment can be found at `http://render-project.eu/drupal`.

2 Functionality

The *Diversity Enricher* module provides several functionalities that present and process information that can be considered to enrich Drupal articles with more *diverse* information.

[1] RENDER project – `http://render-project.eu`

[2] Numbers taken from the `http://drupal.org` landing page. Retrieved on August 8, 2013

Fig. 1. Article view with diversity aspects

2.1 Diversity Information Extraction

One of the most important things about the Drupal extension is the fact that no user interaction is needed to extract the necessary diversity information. This information is generated by the *Enrycher* service[3] which is publicly available. Enrycher utilizes natural language processing techniques to extract diversity information such as *topics*, *sentiments*, *sentiment scores* or *named entities* captured by the article text. This information is then described by using SIOC [1] in combination with the *Knowledge Diversity Ontology*[4] (KDO) [4].

2.2 Links to Related Articles and Topics

Figure 1 shows an overview on the extension. On the left hand side the original article is presented, whereas on the right hand side the main functionality of the extension is located. There, related articles within the Drupal database are listed, split up according to their extracted overall sentiment. An article is considered to be related, if it has at least one topic in common with the currently shown one and is located in the same cluster of the *Diversity-Aware Ranking Service*[5].

The topics of the related articles can be shown by clicking on the + button in front of the article titles. In addition, tags extracted from the currently shown article are presented in a tag cloud below the related article's tree. The size of the respective tag is determined by its number of occurrences in the triple store. Named entities are recognized within the text and get marked. It is possible to click on all tags, named entities, and topics in order to get articles with the same tag/topic (see Figure 2). As a further diversity feature, each article's sentiment is displayed between the title and the actual text.

[3] Enrycher – `http://enrycher.ijs.si`
[4] KDO – `http://kdo.render-project.eu/`
[5] Diversity-Aware Ranking Service – `http://ranking.render-project.eu`

Fig. 2. Related articles with topic "United States"

2.3 Export Options

The diversity data produced by the Enrycher service can be exported in the following formats: RDF+XML, JSON, and Turtle.

2.4 Import articles from a Sesame triple store

Another important function of the extension is the ability to import additional articles into the Drupal database, if they are stored in an Sesame store and are described with the SIOC [1] and KDO [4] ontologies. This option is only available through the administration interface.

3 Key technologies and implementation

The Drupal extension makes use of several technologies and tools. This section describes how the main parts of the *Diversity Enricher* Drupal extension interact.

Diversity Mining Web Services (Enrycher) The main functionality of the Enrycher service has already been described in Section 2.1. However, the Enrycher service could be replaced by any Web service that is SIOC and KDO compliant. This means, that the service has to support a subset of the SIOC and KDO functionalities - namely the extraction and proper output of *topics*, *sentiments* and *sentiment scores*.

Sesame triple store Sesame is used as data store back-end. All other components operate on the Sesame store by using SPARQL queries. The Enrycher service returns RDF data which can be directly submitted to the Sesame store. The *Diversity-Aware Ranking Service* component and the Drupal tool read from the store using a set of predefined queries.

Diversity-Aware Ranking Service This service is used to retrieve related articles with differing sentiments. It operates on a Sesame triple store. The core of the ranking service is a clustering algorithm that operates using a distance metric based on topics and sentiment scores. Articles that have at least one topic in common with the current article are preselected and then clustered by topic. All articles that are in the same cluster as the currently browsed one are then marked as related.

Drupal Integration The tool is connected to Drupal with so-called *hooks*. An implemented hook is called each time a certain event occurs. The hooks of the *Diversity Enricher* module are

- New Article Created: As soon as an article is created, the raw text data is submitted to Enrycher, which extracts the diversity information. This information is then stored to the local Sesame store.
- Article Viewed: If the article is viewed the first time and it has been in the database before the Drupal extension has been activated, this hook acts the same as in the case for *New Article Created*. Additionally, the information needed to present is generated (by using ranking, SPARQL queries) and then presented beside the raw article text.
- Article Changed: If an article is changed, Enrycher is again asked for diversity information and the store is updated with the new enrichment.
- Article Deleted: If the article is removed, all links to the diversity information is deleted from the Drupal database.

4 Related work

The integration of semantic technologies into CMSs brings clear benefits especially for improving search, integration and intelligent management of the content. During the last years several approaches have been published on how semantics can be used within CMSs in general and Drupal in particular.

Since version 7, Drupal natively supports RDF representation of posts, making use of vocabularies like SIOC, FOAF, Dublin Core, and SKOS. Although the new RDF module in Drupal easily enables publishing LOD, it does not provide means for the automatic creation of links to relevant LOD resources.

The approaches described in [2] and [3] enable the production and consumption of Linked Data in CMSs. In [2], two Drupal modules are introduced, one for creating RDFa annotations and another one for generating a SPARQL endpoint for any Drupal site out of the box. The RDFa export module also enables content providers to use their own vocabulary with RDF mappings management. [3] presents RDFaCE, a WYSIWYM (What You See Is What You Mean) editor that extends traditional WYSIWYG editors by RDF statement and RDFa output capabilities. This also enables the reuse of Linked Data sources such as DBpedia. Both approaches focus on the manual or semi-automatic annotation of articles with named entities and topics.

VIE.js[6] is a JavaScript-based semantic interaction framework. It facilitates annotation and interaction with textual and RDFa-annotated content on Web pages. It is used in combination with Apache Stanbol[7] that supports the extension of CMSs with semantic services. Another annotation framework is given by the OpenCalais[8] Drupal extension that uses the OpenCalais API of Thomson Reuters to annotate posts with named entities, facts, and events.

While the above approaches focus on the named entity or topic aspects, we introduce a new dimension given by the active utilization of automatic sentiment extraction. Eventually, this is expected to support the content creation and perception process (given a more fine-grained sentiment and opinion extraction). Also, in contrast to the above approaches, our approach focuses on providing a complete and fully automatic cycle to support the management of diversity; from text analysis and annotation to different visualization methods within Drupal.

5 Current Work

We developed a diversity-aware Drupal extension coined *Diversity Enricher*. The module is currently available at `http://drupal.org/sandbox/sti-innsbruck/1991696`. As of the time of writing (i.e., August 12, 2013) the extension is within a review process to achieve "full project status" within the `http://drupal.org` Web portal. Amongst our next steps will be the qualitative evaluation of the *Diversity Enricher* Drupal module.

Acknowledgment This research was partly funded by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 257790 (RENDER project).

References

1. John G. Breslin, Andreas Harth, Uldis Bojars, and Stefan Decker. Towards semantically-interlinked online communities. In *The Semantic Web: Research and Applications*, volume 3532 of *Lecture Notes in Computer Science*, pages 500–514. Springer Berlin Heidelberg, 2005.
2. Stephane Corlosquet, Renaud Delbru, Tim Clark, Axel Polleres, and Stefan Decker. Produce and consume linked data with drupal! In *Proc. of the 8th Intl. Semantic Web Conf. (ISWC2009)*, Lecture Notes in Computer Science, pages 763–778. Springer, 2009.
3. Ali Khalili, Sören Auer, and Daniel Hladky. The rdfa content editor - from wysiwyg to wysiwym. In *Proc. of COMPSAC 2012*, pages 531–540. IEEE Computer Society, 2012.
4. Andreas Thalhammer, Ioan Toma, Rakebul Hasan, Elena Simperl, and Denny Vrandečić. How to represent knowledge diversity. Poster at the 10th intl. Semantic Web Conf. (ISWC2011), 10 2011.

[6] VIE.js Semantic Interaction Framework – `http://viejs.org/`
[7] Apache Stanbol – `http://stanbol.apache.org/`
[8] OpenCalais Drupal module – `http://drupal.org/project/opencalais`