

Schema.org Mapping Tool for Drupal 8

Sachini A. Herath

Department of Computer Science and Engineering
University of Moratuwa
Sri Lanka
sachini.11@cse.mrt.ac.lk

Abstract—The paper discusses methods of seamlessly integrating Structured Data with Content Management Systems (CMS), to be used by users with no expertise on Linked Data or RDF. First two methods of integrating Linked Data with CMS is presented: mapping Content Structure with Linked Data vocabularies and creating Content Structure based on schemas for structured data. Finally the implementation details of the module developed for Drupal 8, using Schema.org as the mark up vocabulary, is discussed in detail.

Keywords—linked data; CMS; structured data

I. INTRODUCTION

Content Management Systems contributes largely to Hyper Text Markup Language (HTML) Web today. The vast amount of data hosted through them remains unstructured, thus not machine understandable. Meanwhile Semantic Web [1] continues to grow, with many new features and methods of presentations. However connecting these two worlds in a method that can be used by non-experts is still an area largely unexplored.

Drupal is free and open source CMS distributed under GNU General Public License from 2001. The Drupal project's principles encourage modularity, standards, collaboration, ease-of-use, etc. and the architecture has undergone vast changes in its latest version, Drupal 8 (yet to be released) [2]. The content structure of Drupal makes it ideal for this project [3].

Schema.org [4] is an open community driven vocabulary that is sponsored and used by all major search engines such as Google, Yahoo, and Microsoft etc. The vocabularies cover entities, relationships between entities and actions.

A. Content Structure of Drupal

On Drupal, 'Content' defines how data is stored for web content. Each content is an implementation of a Content Type, which is a pre-defined collection of data types (Fields) that relate to each other by an informational context. Content Types are how site editors can input original content on a Drupal site and are the building blocks for structured authoring and content. Content types often work in conjunction with Views, which is one way the content can be served up to the end users - how content types appear and the order they appear can be controlled [5].

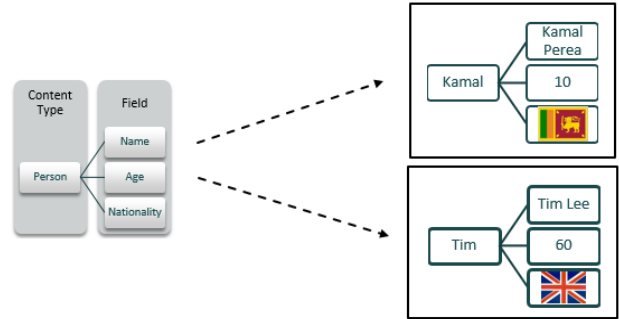


Fig. 1. Drupal Content Structure.

Each Content Type has a set of attributes termed 'fields' as shown in Fig.1. Fields too can be generic or user-defined. Some of the default fields are title and body. Each field has a unique machine name, field type which specifies the type of data to store in the field (e.g. - image, text, integer etc.), and a set of parameters to define how the field is displayed. Fields can be reused across content types.

B. Schema.org vocabulary structure

Schema.org vocabulary introduces hierarchical schema (shown in Fig. 2) that can be presented in Microdata, RDFa, or JSON-LD formats.

C. Linked Data and Web Content

The two content structures described above can be integrated to introduce machine understandable structure for the data, as shown in Fig. 3.

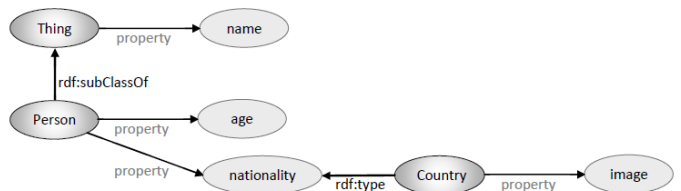


Fig. 2. Hierarchical structure of Schema.org.

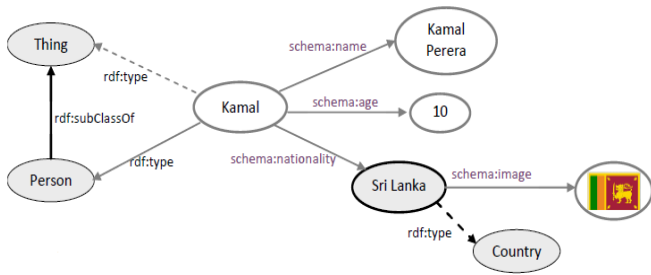


Fig. 3. Content type with RDF data.

```
<html prefix="schema: http://schema.org/">
....
<div typeof="schema:Person">
  My name is
  <span property="schema:name">Kamal Perera</span>.
  I'm a <span property="schema:age">10</span> year old
  from <span rel="schema:nationality">
    <span typeof="schema:Country">
      <span property="schema:name">Sri Lanka</span>
    </span> </span>.
</div>
```

Fig. 4. HTML Text annotated with markup.

Fig. 4 shows the relationships shown as graphic in Fig. 3 depicted as RDFa markup in the HTML body.

II. RELATED WORK

Drupal has introduced RDF module as one of its core modules starting from Drupal 7, which injects the defined Resource Description Framework (RDF) mappings to theme functions and templates [6]. Moreover Drupal 7 has a considerable collection of contributing modules which focuses on Linked Data. RDF Extensions [7] is one such module and it presents the functionality to map Content types and fields with vocabularies, much similar to the first component described in this paper.

Other popular CMS such as Joomla and WordPress have modules for introducing Linked Data, although due to the differences in architecture they are not tightly coupled with the underlying content structure.

III. DESIGN AND IMPLEMENTATION

The implementation was done for Drupal 8 as a contributing module [8]. The main module, RDF UI, contains the tools for mapping content structure with Schema.org vocabularies while the submodule, RDF Builder, focus on creating content structure based on schemas for structured data.

A. RDF UI

Schema.org core schema is fetched through a Hyper Text Transfer Protocol (HTTP) request to the module. The fetched

data is processed as a graph using EasyRDF library for PHP [9] and cached for faster access.

The existing create and edit forms for Content types are altered using hooks to introduce Schema.org mapping. The altered process view is depicted in Fig. 5 and UI in Fig. 6.

Once a Content type is created the mapping for fields and properties are to be selected. The candidate properties for auto-completion will be properties of the selected Type or its parent Types, and are filtered removing deprecated properties. The process view for specifying field mapping is shown in Fig. 6 and the UI in Fig. 7.

jQuery autocomplete widget is used for the dropdown menu for selecting the Types and Properties. However since it presents the limitation of categorizing properties according to the Type it belongs to, it is advised to use Select2 autocomplete combo-box for the select widget [10].

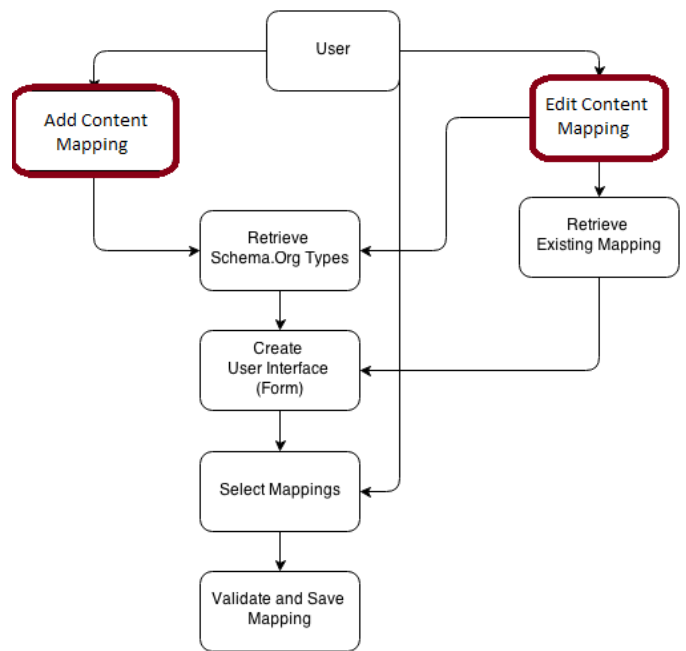


Fig. 5. Process view for mapping Content type and Schema.org type.

Fig. 6. Content type mapping form.

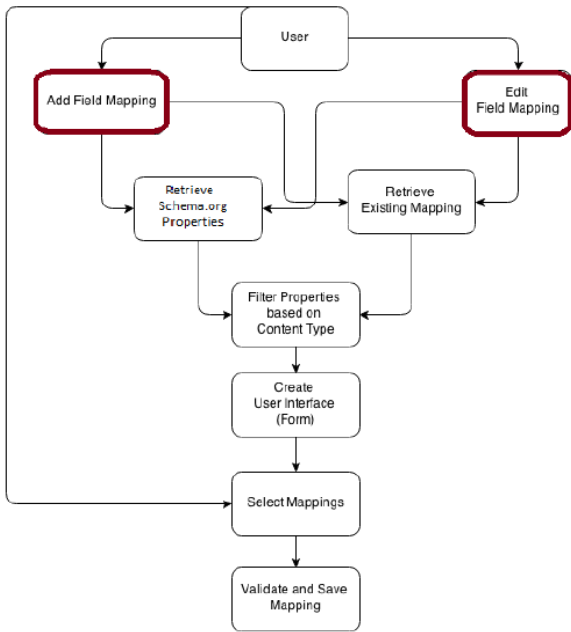


Fig. 7. Process View for mapping fields and Schema.org properties.

When creating content using the Content type template the specified mappings will be injected in the HTML body by the core RDF module. The mappings in HTML can be extracted and viewed using the Google Structured Data Testing Tool [11].

B. RDF Builder

RDF Builder module introduces the novel concept of using the Schema.org schema structure to specify and create Content Types. This not only makes the life of Content developers easier but also creates content structure expected by the search engines.

RDF builder uses a multi-step form to create new Content Types as shown in Fig. 8. Once a Type is selected, the properties of the selected Type and its parent Types are listed as candidates for fields, as shown in Fig. 9. Based on the information available in the vocabulary many of the field properties can be filled automatically. One example is the field type widgets to be used for fields, which are suggested based on the data type of each property.

Upon submission the new Content type and fields are created and stored along with RDF mappings. The user then can use the usual tools to customize the created Content Type and fields.

This tool presents a faster and efficient alternative to the typical Drupal workflow.

C. Testing and Maintainance

All the functionality presented in the two modules are tested using SimpleTest [12] automated Testing tool.

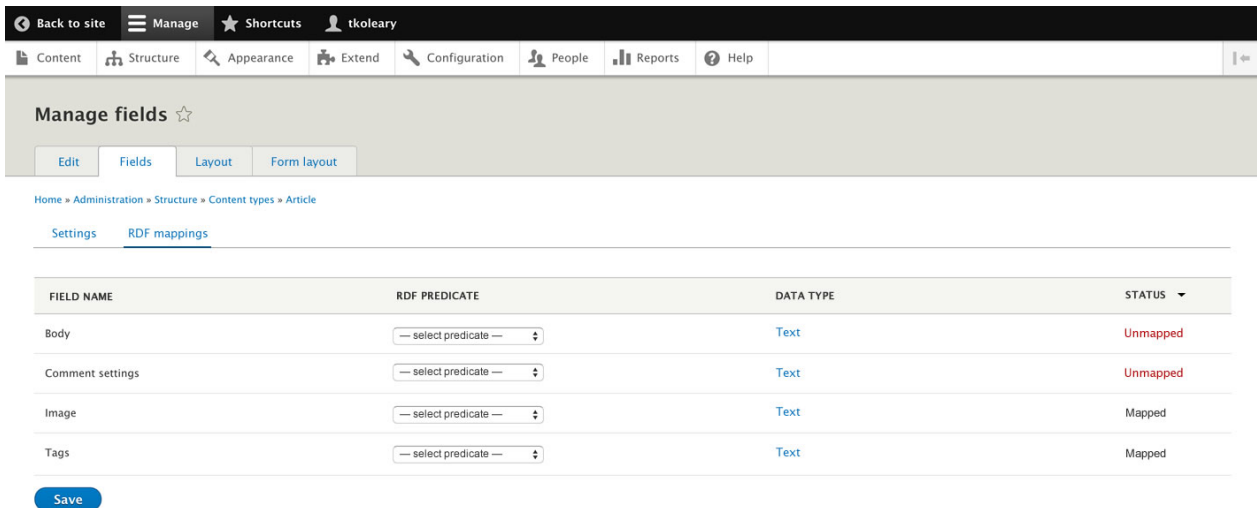


Fig. 8. Field mapping form.

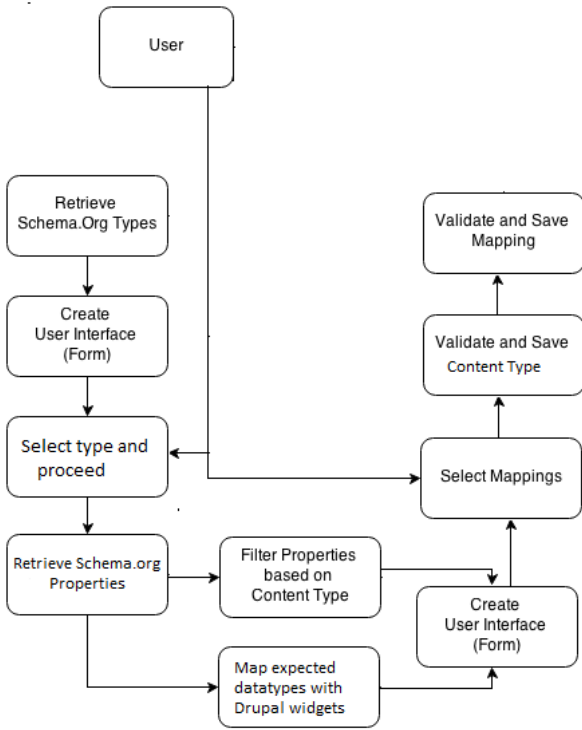


Fig. 9. Process View for RDF Builder.

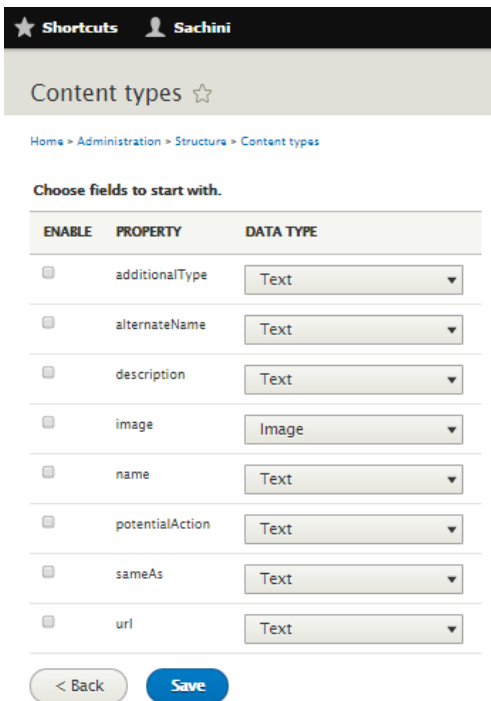


Fig. 10. RDF Builder.

IV. CONCLUSION AND FUTURE WORK

The tools developed for this project enables non-expert users to easily integrate Linked Data to their websites and define better structure for the data to optimize search. Functionality such as suggestions for fields, given a Content Type and matching relevant properties and types based on the display name of Content types and fields can increase the user experience of this module.

The next step for improving the Semantic Web Technology in Drupal 8 would be matching actual content with relevant entities through URIs.

ACKNOWLEDGMENTS

The author thank Stephane Corlosquet, Senior Software Engineer, Acquia Inc. for mentoring the project and Google Summer of Code initiative for sponsoring the project in 2014.

REFERENCES

- [1] W3.org, "Semantic Web - W3C", 2015. [Online]. Available: <http://www.w3.org/standards/semanticweb/>. [Accessed: 05- Jul- 2015].
- [2] Buytaert.net, "Why the big architectural changes in Drupal 8 | Dries Buytaert", 2015. [Online]. Available: <http://buytaert.net/why-the-big-architectural-changes-in-drupal-8>. [Accessed: 04- Jul- 2015].
- [3] S. Corlosquet, "Bootstrapping the Web of Data with Drupal", Master Degree in Computer Science, National University of Ireland, Galway, 2009.
- [4] Schema.org, 2015. [Online]. Available: <http://schema.org/>. [Accessed: 05- Jul- 2015].
- [5] Drupal.org, "Working with content types and fields (Drupal 7 and later) | Drupal.org", 2011. [Online]. Available: <https://www.drupal.org/documentation/modules/field-ui>. [Accessed: 04- Jul- 2015].
- [6] Api.drupal.org, "RDF Mapping API | rdf.module | Drupal 8 | Drupal API", 2015. [Online]. Available: <https://api.drupal.org/api/drupal/core!modules!rdf!rdf.module/group/rdf/8>. [Accessed: 05- Jul- 2015].
- [7] Drupal.org, "RDF Extensions | Drupal.org", 2015. [Online]. Available: <https://www.drupal.org/project/rdfx>. [Accessed: 05- Jul- 2015].
- [8] Drupal.org, "RDF UI | Drupal.org", 2015. [Online]. Available: <https://www.drupal.org/project/rdfui>. [Accessed: 05- Jul- 2015].
- [9] N. Humfrey, "EasyRdf - RDF Library for PHP", *easyrdf.org*, 2015. [Online]. Available: <http://www.easyrdf.org/>. [Accessed: 05- Jul- 2015].
- [10] Select2.github.io, "Select2 - The jQuery replacement for select boxes", 2015. [Online]. Available: <https://select2.github.io/>. [Accessed: 05- Jul- 2015].
- [11] Google Developers, "Structured Data Testing Tool | Google Developers", 2015. [Online]. Available: <https://developers.google.com/structured-data/testing-tool/>. [Accessed: 05- Jul- 2015].
- [12] Drupal.org, "SimpleTest | Drupal.org", 2015. [Online]. Available: <https://www.drupal.org/project/simpletest>. [Accessed: 05- Jul- 2015].