

An RDF Platform for Generating Web API for Open Government Data

Pattama Krataithong^{1,2}, Marut Buranarach¹, Nuttanont Hongwarittorn²
and Thepchai Supnithi¹

¹ Language and Semantic Technology Laboratory
National Electronics and Computer Technology Center (NECTEC), Pathumthani, Thailand
{marut.bur,pattama.kra}@nectec.or.th

² Department of Computer Science, Faculty of Science and Technology
Thammasat University, Pathumthani, Thailand
nth@cs.tu.ac.th

Abstract. Most of datasets in open data portals are mainly in tabular format in spreadsheet, e.g. CSV and XLS. To increase the value and reusability of these datasets, the datasets should be made available in RDF format that can support better data querying and data integration. However, publishing and querying RDF requires different knowledge and skills. In this poster, we present a platform for publishing and querying the dataset in RDF that does not require the user's knowledge of RDF and SPARQL. This framework supports semi-automatic construction of RDF data and RESTful APIs from the datasets in tabular format. The framework provides automatic schema detection, i.e. data type detection, and ontology and RDF data mapping generation. RESTful API is provided on top of the SPARQL data querying service for each published RDF dataset. A platform prototype was developed and demonstrated using some datasets from the Data.go.th website. Some current research directions include automatic dataset API generation based on Web crawler and validator and development of intelligent search engine over the dataset APIs.

Keywords: dataset management, open data platform, RDF data publishing

1 Introduction

The number of datasets on the Thailand open government data portal, i.e. Data.go.th, is continually increasing. Majority of datasets on these portals are in tubular formats such as Excel and CSV. Based on the 5-star open data model¹, Resource Description Framework (RDF) is a standard data format that can support linked open data. There are two important standards for integrating data. First, RDF is a standard format for integrating data based on URI and XML syntax. Second, the Web Ontology Language (OWL) is important for linked data based on classes and properties.

¹ <http://5stardata.info/en/>

Consuming RDF data is usually achieved by querying via an SPARQL endpoint. A developer who wants to use the SPARQL endpoint must have the knowledge about SPARQL and RDF. Our work proposes that Web API is an easier way for retrieving RDF-based open data. There are several advantages of providing Web API over the RDF datasets including:

- Data as a service – developers who do not have background in RDF and SPARQL can query a dataset via a RESTful API service.
- Standard data format– developers do not need to study a new data format, the query results will be returned in the standard JSON format.

In this poster, we present a platform that provides a data management support for RDF data publishing and consuming. The platform was developed using the Ontology Application Management (OAM) framework [1]. The platform prototype was available at the Demo-api.data.go.th website, which exemplifies deployment of the platform using some datasets from the Data.go.th website.

2 RDF Dataset Management Process

Fig. 1a shows the RDF dataset publishing process. The RDF dataset generations consist of four processes [2]: 1) User Management and Authentication 2) Dataset Preparing and Import 3) Schema Detection and Verification and 4) OWL and RDF data generation. The requirements of input data are as follows: 1) the dataset must consist of only one table (one spreadsheet), 2) the table must have one header row, 3) header of the table may be written in English or Thai language.

Fig. 1b shows the layers of Dataset Service API at Demo-api.data.go.th. The website provides the data as a service through RESTful APIs for each dataset which was converted in RDF format and published on this portal. A form-based search interface for each dataset is formed based on the ontology of the dataset. For each dataset, the data querying service is automatically provided as RESTful API by means of the OAM framework. Application developers can query each dataset via APIs and the returned search results are provided in JSON format.

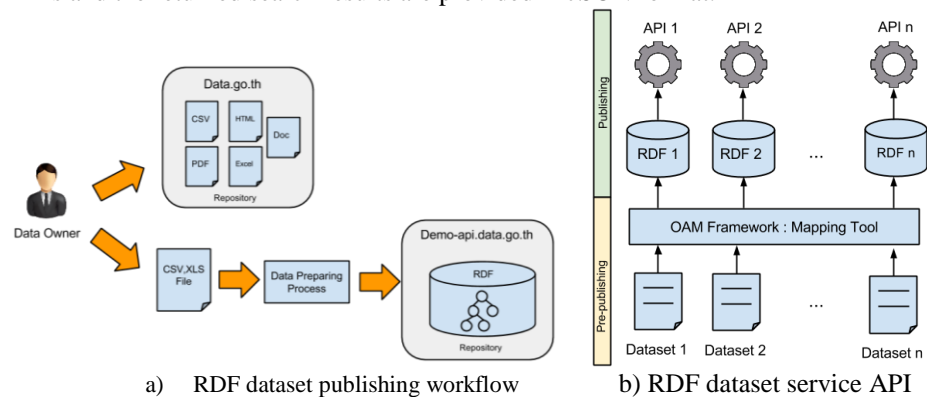


Fig. 1 RDF Dataset Management Process for Demo-api.data.go.th

3 Usage Scenarios

Fig. 2 shows an interface of the dataset publishing functions for each user. The user can choose to create new, update or delete an RDF dataset. Fig. 3 shows an interface of the schema detection and verification step. In accessing each created dataset the user can choose to search or view the RDF dataset. Fig. 4 shows access to the dataset API via the ontology-based search interface.

The APIs are provided for three main functions: getting all dataset names, getting description of a dataset schema and querying a dataset [3]. An example of API for querying a dataset by search conditions is shown in Fig. 5.

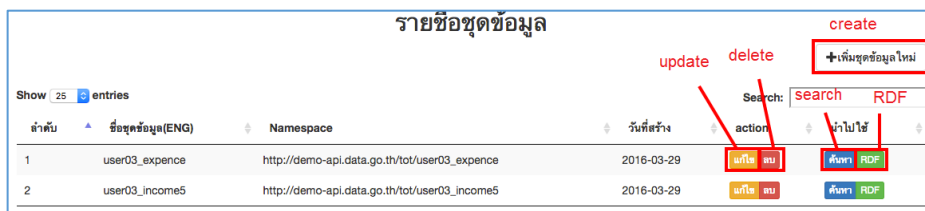


Fig. 2 User interface for listing all datasets and functions for each user

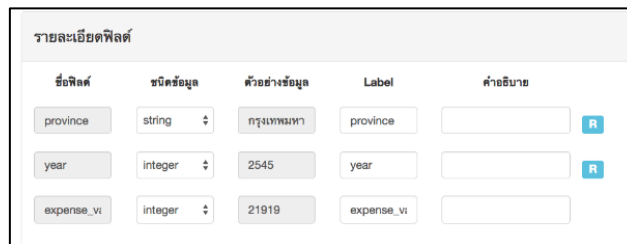


Fig. 3 User interface for the schema verification step

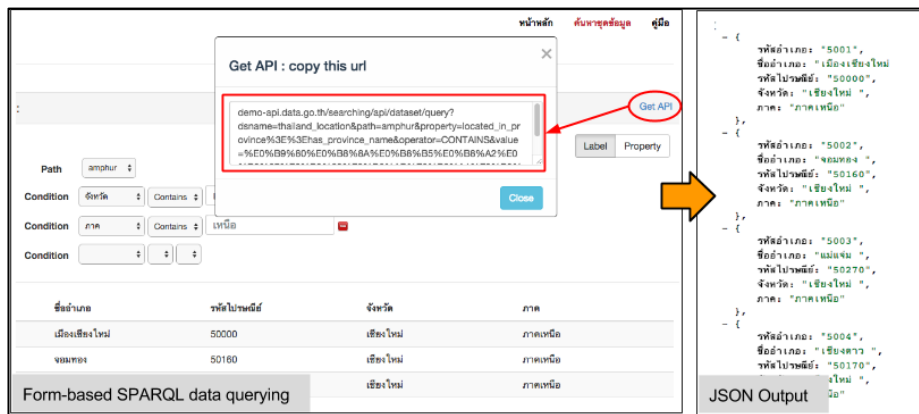


Fig. 4 Access to the API via the ontology-based search interface

```

http://demo-api.data.go.th/searching/api/dataset/query?dsname=thailand_location&
path=amphur&property=located_in_province>>located_in_region>>has_region_name&o
perator=CONTAINS&value=ภาคเหนือ&property=located_in_province>>has_province_
name&operator=CONTAINS&value=เชียงใหม่&limit=100&offset=0

```

Fig. 5 Example API for querying a dataset by search conditions

4 Discussion and Research Directions

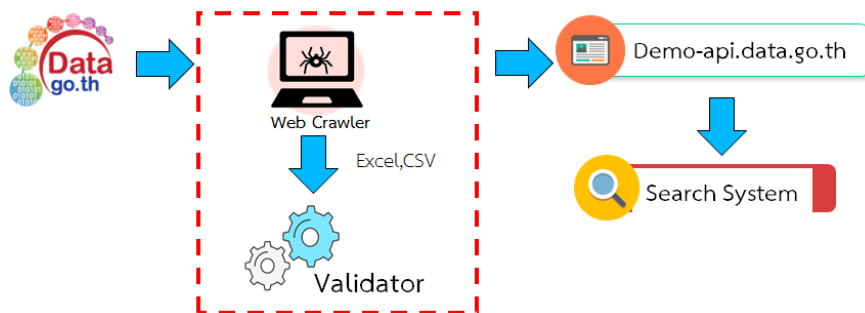


Fig. 6 An automatic approach for generating APIs for the datasets of Data.go.th

This poster describes a semi-automatic framework for generating RDF dataset from open tabular data. This framework allows the users to publish their datasets in RDF format and query the data via Web API with no required knowledge about RDF and SPARQL. One of the difficulties is that some datasets are not in the valid tabular format [4]. In addition, human intervention is still required, which limited scalability of the framework. One of our research directions is to develop a Web crawler and validator to automatically retrieve and create the APIs from all valid datasets of the Data.go.th website. We are also developing an intelligent search system, which allows the user to search the data in the datasets via the APIs using a semi-natural-language UI. The automatic approach for generating APIs for the datasets is shown in Fig 6.

References

1. Buranarach, M. et al.: OAM: An Ontology Application Management Framework for Simplifying Ontology-Based Semantic Web Application Development. *Int. J. Softw. Eng. Knowl. Eng.* 26, 01, 115–145 (2016).
2. Krataithong, P., Buranarach, M., Supnithi, T., and Hongwarittorn, N.: Semi-Automatic Framework for Generating RDF Dataset from Open Data. In: *Proc. of the 11th International Symposium on Natural Language Processing (SNLP2016)* (2016).
3. Krataithong, P., Buranarach, M., and Supnithi, T.: RDF Dataset Management Framework for Data.go.th. In: *Proc. of the 10th International Conference on Knowledge, Information and Creativity Support Systems (KICSS2015)* (2015).
4. Ermilov I, Auer S, Stadler C (2013) User-driven semantic mapping of tabular data. *Proc 9th Int Conf Semant Syst - I-SEMANTICS '13* 105. doi: 10.1145/2506182.2506196