The role of vocabularies for estimating carbon footprint for food recipies using Linked Open Data

Ahsan Morshed

Intelligent Sensing and Systems Laboratory, CSIRO, Hobart, Australia
{ahsan.morshed, ritaban.dutta}@csiro.au

Fabrizio Celli

Food and Agriculture Organization of United Nation, Rome, Italy
Fabrizio.celli@fao.org

Abstract. The standard terms with known meanings are often called controlled vocabulary or light weight ontologies. These play vital role in the Linked Open Data cloud. These vocabularies capture a central notion of context for a specific domain in the knowledge cloud. The extra information is co-habited with these controlled vocabularies. This short paper shows the role of these vocabularies in calculation of carbon footprint for food recipes using Linked Open Data.

Keywords: Vocabularies; Ontology; LOD

1 Introduciton

Rapid growth of information on the web creates difficulties for accessing the right information by a huge number of unknown users with extremely diverse needs, skill and education, culture and language background [4,6]. They often mislead with wrong information due to right queries. Currently, open data movement also creates the similar kind of problem. To tackle these challenges, we need standard terms with meaning. They are also called controlled vocabularies (CV). There is no notion of CV; we can define it as a set of concepts of preferred terms and existing relation among them. For example, WordNet, MeSH, LCSH, AGROVOC, all kind of light weight vocabularies or classification systems, are sort of CVs [1, 6].

Several organizations are using the vocabularies for tagging their contents [2, 4, 5, 7]. For example, the largest Agriculture vocabulary called "AGROVOC" is used for tagging bibliographic contents at the Food and Agricultural Organization of the United Nations. Furthermore, we believe that the semantic enrichment and heterogeneous collections can be facilitated by using the vocabularies for cross linking between collections, since major classification and thesauri have been made available following linked open data (LOD) principles [4,5,11]. By using the LOD principle [11], we can generate the knowledge at run time. In this paper, we proposed a prototype for generating the carbon footprint for food recipes and further knowledge can be augmented through the different data sources. Therefore, the paper aims to explore the LOD knowledge through the controlled vocabularies.

This paper starts with the classification of CV. In section 3, we propose the prototype for estimating the carbon footprint of food recipes, by calculating the carbon dioxide (CO_2) emission using linked data. Finally, we conclude that the prototype can be extended in the context of soil carbon footprint in Australia.

2 Classification of Controlled Vocabularies

The CVs can be classified based on nature, construction, perspective and usages [6]. For instance, two words "Center" and "Centre" both have the same meaning but different spellings in different regions and cultures. Figure 1 below shows the classification tree for the CVs. Recent efforts by the W3C library Linked Data Incubator Group [12] using the resource description framework (RDF) in the international LOD based libraries made it possible to publish all these vocabularies in machine readable format.

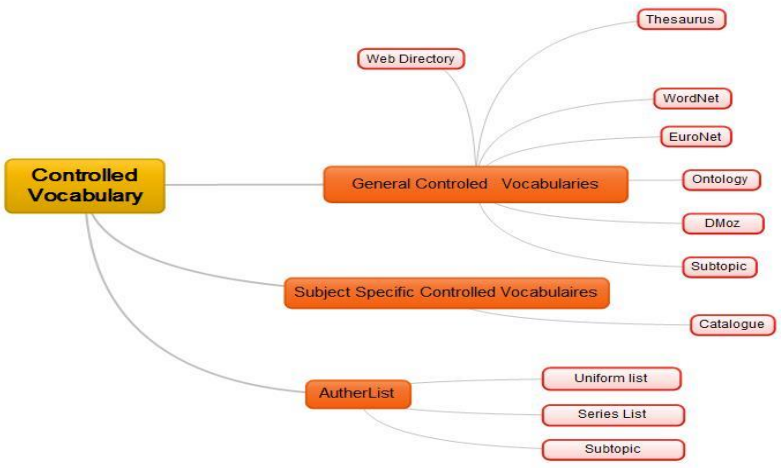

Fig. 1. Classification of Controlled Vocabularies

For generating the machine readable format for these vocabularies, the Simple Knowledge Organization System (SKOS) has gained considerable popularity [7]. The SKOS[1] was developed as a RDF to represent the vocabularies in the semantic web. In RDF of vocabularies, each term is presented as a concept that can be accessible via uniform resource identifier (URI). In addition, by using the URIs, vocabularies can talk to each other by integrating the resources through the HTTP protocols. The term "Milk" from AGROVOC LOD can talk with DBpedia [3, 5] by using the URIs as shown in Figure 2 and RDF data (See Figure 3) can be automatically read by machine.

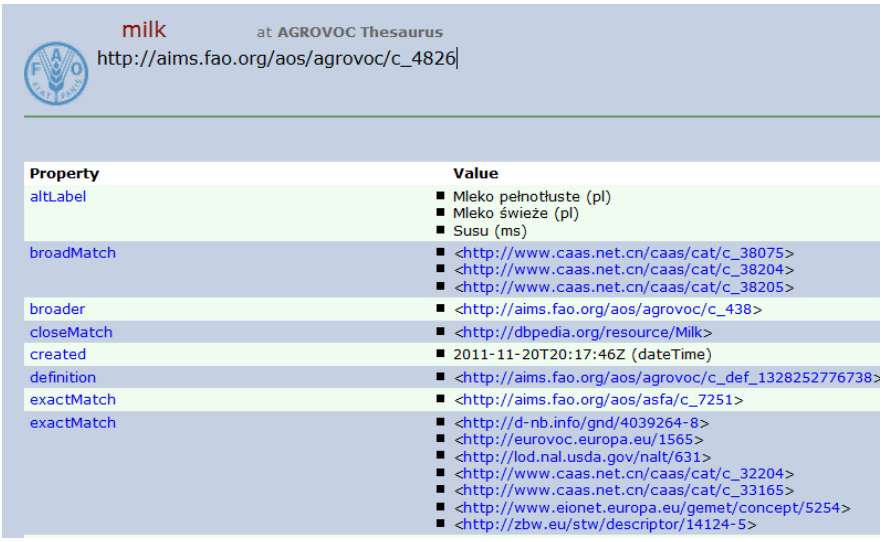

Fig. 2. AGROVOC LOD

Furthermore, it is also possible to read the match URIs from these RDF data.

```
− <rdf:RDF>
  − <j.0:Concept rdf:about="http://aims.fao.org/aos/agrovoc/c_4826">
      <j.0:prefLabel xml:lang="sk">mlieko</j.0:prefLabel>
      <j.6:prefLabel rdf:resource="http://aims.fao.org/aos/agrovoc/xl_pt_1299486417939"/>
      <j.6:prefLabel rdf:resource="http://aims.fao.org/aos/agrovoc/xl_es_1299486417617"/>
      <j.6:prefLabel rdf:resource="http://aims.fao.org/aos/agrovoc/xl_en_1299486417597"/>
      <j.3:isComponentOf rdf:resource="http://aims.fao.org/aos/agrovoc/c_15739"/>
      <j.0:closeMatch rdf:resource="http://dbpedia.org/resource/Milk"/>
      <j.0:definition rdf:resource="http://aims.fao.org/aos/agrovoc/c_def_1328252776738"/>
      <j.0:narrower rdf:resource="http://aims.fao.org/aos/agrovoc/c_16077"/>
      <j.0:narrower rdf:resource="http://aims.fao.org/aos/agrovoc/c_16083"/>
```

Fig. 3. A snapshort of RDF

1. http://www.w3.org/2004/02/skos/

The roles of these CVs are to help for gathering and searching sustainable information from the web. This information creates an opportunity for famers to make educated them through the web technologies. However, we section 3, we presents

3 Consumption of Linked Open Data: Use Case of Carbon footprint

Current global crisis of food creates the necessity to make climate friendly food choice by avoiding waste. Wasting food costs extra money and increases the carbon footprint. This sustainability issue affects the social life in the country. To tackle the challenge, users need to know in advance how much CO_2 emission is caused by a given food recipe [8, 9, 10]. To calculate the carbon footprint, we considered the following processes involved in a food recipe based on [9]

- The production of the ingredients
- The transport of the ingredients
- The cooling of the ingredients
- The preparation of the ingredients

By taking the above processes into account, the information about the emission of CO_2 have been extracted by the application programming interfaces (API) provide by AMEE[2]. Figure 4 shows the total carbon emission when "Rice", "Milk" and "Bacon" are selected as ingredients for calculating the CO_2.

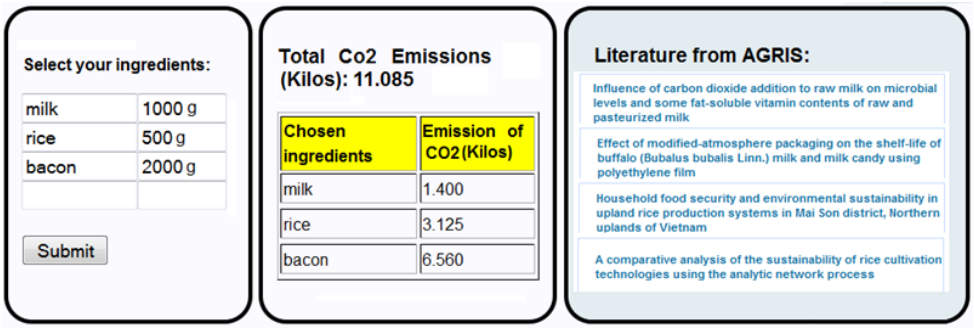

Fig. 4. Information consumption through LOD

Furthermore, we added more sustainable information about these ingredients from the international system for agricultural science and technology (AGRIS) literatures by using the LOD principles. The AGRIS database is a collection of more than 5 million bibliographic records in the agricultural domain, enhanced by the AGROVOC thesaurus which is extensively used by cataloguers world-wide to enrich data indexing in

[2] https://www.amee.com/pages/about

agricultural information systems. AGRIS has a public sparql endpoint and is exposed as Linked Open Data (more than 130 million triples) to better meet its customers' expectations by providing entity-based access to bibliographic data and by mashing up this information with related data sources by taking advantage of the many formal alignments between AGROVOC and other knowledge organization systems [2,3]. Figure 5 gives an example of an AGRIS RDF record. AGRIS LOD can be browsed through OpenAGRIS[3], an RDF mashup application that allows users to query the AGRIS content and - using AGROVOC URIs as backbone to call webservices or sparql endpoints – to interlink to external sources of information (World Bank, Nature OpenSearch, Europeana, GBIF, Biodiversity International, etc.) discovering as much information as possible (text, media, maps, statistics) about a specific topic or research area.

```
<bibo:Article rdf:about=" http://agris.fao.org/aos/records/LV2009000697">
        <dct:identifier> LV2009000697</dct:identifier>
        <dct:title xml:lang="eng">
                Problems of dairy farming and their solutions in Latvia
        </dct:title>
        <dct:creator><foaf:Person>
                <foaf:name><![CDATA[Vitola, I.]]></foaf:name>
        </foaf:Person></dct:creator>
        <dct:issued>2010</dct:issued>
        <dct:subject rdf:resource="http://aims.fao.org/aos/agrovoc/c_4829"/>
        <dct:subject rdf:resource="http://aims.fao.org/aos/agrovoc/c_33560"/>
        <bibo:abstract xml:lang="eng">
            Dairy farming is one of the basic industries of agriculture in Lat-
            via. Milk production costs sharply increase, …
        </bibo:abstract>
        <bibo:language>eng</bibo:language>
        <dct:isPartOf rdf:resource="http://aims.fao.org/serials/ c_37701e1f"/>
</bibo:Article>
```

Fig. 5. The RDF/XML serialization of an AGRIS record

4 Conclusion

In our paper, we have shown the role of LOD for estimating the carbon footprint for a given food recipe by using the CVs. This is a research prototype which can be extended and reused for other applications. In future, we hope to use the prototype for estimating the soil carbon emission for a given location in context of LOD.

[3] http://agris.fao.org/openagris/

References

1. Morshed, A., Johannsen, G., Keizer, J., & Zeng, M. L. (2010). Bridging end users' terms and AGROVOC concept server vocabularies.
2. Morshed, A., Caracciolo, C., Johannsen, G., & Keizer, J. (2011, September). Thesaurus alignment for linked data publishing. In International Conference on Dublin Core and Metadata Applications (pp. 37-46)
3. Caracciolo, C., Stellato, A., Morshed, A., Johannsen, G., Rajbhandari, S., Jaques, Y., & Keizer, J. (2012). The AGROVOC Linked Dataset. Semantic Web
4. Bizer, C. (2009). The emerging web of linked data. Intelligent Systems, IEEE, 24(5), 87-92.
5. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). Dbpedia: A nucleus for a web of open data. In The semantic web (pp. 722-735). Springer Berlin Heidelberg.
6. Morshed, A. (2012). Role of Vocabularies for semantic interoperability in enabling the Linked Open Data publishing", International Journal of Database Management Systems,Volume 4, Issue 11,2012, DOI: 10.5121/ijdms.2012.4502. 21
7. Hayman, S. (2007). Folksonomies and tagging: New developments in social bookmarking. In *Ark Group Conference: Developing and Improving Classification Schemes*.
8. Hertwich, E. G., & Peters, G. P. (2009). Carbon footprint of nations: A global, trade-linked analysis. *Environmental science & technology*, *43*(16), 6414-6420.
9. Stichnothe, H., Hospido, A., & Azapagic, A. (2008). Carbon footprint estimation of food production systems: The importance of considering methane and nitrous oxide. *Aspects of Applied Biology*, *87*, 65-75.
10. Friedrich, D., & Peetz, S. Methodology for Calculating Carbon-Footprints of Recipes (Third Draft). See at http://www.greenlicious.org/downloads/methodology.pdf
11. Tim Berners-Lee. See at http://www.w3.org/DesignIssues/LinkedData.html/.
12. Library Linked Data Incubator. See at http://www.w3.org/2005/Incubator/lld/charter