

# Role of Vocabularies for Semantic Interoperability in Enabling the Linked Open Data Publishing

**Ahsan Morshed**  
*United Nations, Italy*

## EXECUTIVE SUMMARY

*In the spite of explosive growth of the Internet, information relevant to users is often unavailable even when using the latest browsers. At the same time, there is an ever-increasing number of documents that vary widely in content, format, and quality. The documents often change in content and location because they do not belong to any kind of centralized control. On the other hand, there is a huge number of unknown users with extremely diverse needs, skills, education, and cultural and language backgrounds. One of the solutions to these problems might be to use standard terms with meaning; this can be termed as controlled vocabulary (CV). Though there is no specific notion of CV, we can define it as a set of concepts or preferred terms and existing relations among them. These vocabularies play very important roles classifying the information. In this chapter, we focus the role of CV for publishing the web of data on the Web.*

## **INTRODUCTION**

In the spite of explosive growth of the Internet, information relevant to users is often unavailable even when using the latest browsers. At the same time, there is an ever increasing number of documents that vary widely in content, format and quality. The documents often change in content and location because they do not belong to any kind of centralized control. On the other hand, there is a huge number of unknown users with extremely diverse needs, skills, education, and cultural and language backgrounds. One of the solutions to these problems might be to use standard terms with meaning, this can be termed as controlled vocabulary (CV) (Morshed, 2009; Morshed, 2010). Though there is no specific notion of CV, we can define it as a set of concepts or preferred terms and existing relations among them. For example, thesauri, WordNet (Miller, 1998), MeSH (MeSH, 2012), LCSH (LCSH, 2012), all kinds of ontologies, are sorts of CVs. These CVs are used to matching purpose that makes more exible for information extraction. In a semantic or controlled vocabulary (Giunchiglia, 2004), a matching operator takes two-graph like structures, for instance ontologies or classifications and produces matching relationship among them. This semantic matching system is based on two key notions. One of them is the concept of nodes and other is the concept of labels. In semantic matching, labels are written in natural language. These labels are disambiguated using a lexicon (Gale,1992). In this case, they are working as a background knowledge. In this chapter, we will see the contribution of CV for publishing the web of data purposes and review the main applications of controlled vocabularies.

## **CLASSIFICATION OF CONTROLLED VOCABULARY**

In our case, we can classify our controlled vocabularies based on nature, construction perspective and usage. These constructions are based on regions, countries, products, services, vertical markets, clients, customer alliances, structure subsidiaries histories and cultures etc. For instance, two words “Center” and “Centre” both are having same meaning but different spelling in different regions and cultures.

We can classify controlled vocabularies in the following ways.

### **General Controlled Vocabulary**

This class of controlled vocabulary is mainly included in usage and existing relationships among the concepts and entities. For example, the most prominent representation of these vocabularies are Thesaurus, WordNet, Classification, Directories, Lightweight Ontologies (Zhu,2006), etc. (Figure 1).

19 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/role-vocabularies-semantic-interoperability-enabling/77201](http://www.igi-global.com/chapter/role-vocabularies-semantic-interoperability-enabling/77201)

## Related Content

---

### Clustering of Time Series Data

Anne Denton (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 258-263).

[www.irma-international.org/chapter/clustering-time-series-data/10830](http://www.irma-international.org/chapter/clustering-time-series-data/10830)

### Non-Linear Dimensionality Reduction Techniques

Dilip Kumar Pratihar (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1416-1424).

[www.irma-international.org/chapter/non-linear-dimensionality-reduction-techniques/11007](http://www.irma-international.org/chapter/non-linear-dimensionality-reduction-techniques/11007)

### Discovering Knowledge from XML Documents

Richi Nayak (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 663-668).

[www.irma-international.org/chapter/discovering-knowledge-xml-documents/10891](http://www.irma-international.org/chapter/discovering-knowledge-xml-documents/10891)

### Web Mining in Thematic Search Engines

Massimiliano Caramia and Giovanni Felici (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 2080-2084).

[www.irma-international.org/chapter/web-mining-thematic-search-engines/11106](http://www.irma-international.org/chapter/web-mining-thematic-search-engines/11106)

### Secure Computation for Privacy Preserving Data Mining

Yehuda Lindell (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1747-1752).

[www.irma-international.org/chapter/secure-computation-privacy-preserving-data/11054](http://www.irma-international.org/chapter/secure-computation-privacy-preserving-data/11054)