

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-interopability-enabling/77201

Related Content

Soft Subspace Clustering for High-Dimensional Data

Liping Jing, Michael K. Ngand Joshua Zhexue Huang (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1810-1814).

www.irma-international.org/chapter/soft-subspace-clustering-high-dimensional/11064

Cost-Sensitive Learning

Victor S. Shengand Charles X. Ling (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 339-345).

www.irma-international.org/chapter/cost-sensitive-learning/10842

Integration of Data Mining and Operations Research

Stephan Meisel (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1046-1052).

www.irma-international.org/chapter/integration-data-mining-operations-research/10950

Secure Building Blocks for Data Privacy

Shuguo Han (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1741-1746).

www.irma-international.org/chapter/secure-building-blocks-data-privacy/11053

An Introduction to Kernel Methods

Gustavo Camps-Valls, Manel Martínez-Ramónand José Luis Rojo-Álvarez (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition* (pp. 1097-1101).

www.irma-international.org/chapter/introduction-kernel-methods/10958