Advanced Engineering Informatics xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

A high-level electrical energy ontology with weighted attributes ☆

Dilek Küçük *

Electrical Power Technologies Group, TÜBİTAK Energy Institute, Ankara, Turkey

ARTICLE INFO

Article history:
Received 10 August 2014
Received in revised form 6 April 2015
Accepted 10 April 2015
Available online xxxx

Keywords:
Domain ontology
Electrical energy
Weighted attributes
Ontology learning
Wikipedia
Text analysis applications

ABSTRACT

One of the significant application areas of domain ontologies is known to be text analysis applications like information extraction and text classification systems, and semantic portals. In this paper, we present a high-level ontology for the electrical energy domain. This domain ontology has weighted attributes to cover the inherent fuzziness in the textual representations of its concepts. Additionally, we have included in the ontology the necessary attributes to align the ontology concepts to on-line collaborative knowledge bases like Wikipedia and linked open data sources like DBpedia, other attributes to facilitate its use in multilingual applications, and concepts to hold the named entities in the domain. The ultimate ontology is aligned with the previously proposed ontologies for the energy-related subdomains after extending the latter ones with weighted attributes. We make the ultimate form of the electrical energy ontology, as well as the extended versions of the domain ontologies for the subdomains, available for research purposes. Also included in the paper are sample text analysis applications which mainly exploit the weighted attributes within the ontology.

1. Introduction

Domain ontologies are usually defined as collections of semantic concepts describing a domain together with the interrelations among the concepts and rules governing these concepts [1]. A domain ontology is a valuable semantic resource for the underlying domain since it can act as a shared vocabulary for the domain under consideration, it can help reduce possible interoperability problems between the domain-specific applications and Semantic Web applications, in addition to several other uses, as reviewed in studies like [2].

Ontologies have been proposed for several application domains so far, including those ontologies for bioinformatics [3], petri nets [4], chemical process engineering [5], software product families [6], electrical power quality [7], and wind energy [8], among others. Due to the peculiar needs of the applications in which domain ontologies are to be employed, some ontologies should possess fuzzy characteristics. Hence, fuzzy domain ontologies have also been proposed and exploited in several applications. Examples of fuzzy ontology proposals include the ontology in [9] for medical information retrieval, that in [10] for news summarization, and the one in [11] for the acquisition and share of scientific information on the Web.

An important issue regarding the construction of domain ontologies is that it is a considerably labor-intensive and time-consuming process which is usually carried out by the experts of the domain. As reviewed in studies like [2,12], there are several studies that target at decreasing the overall cost of the ontology building process through the employment of semi-automatic approaches which involve a learning phase, along the way. Yet, even within such procedures, manual intervention by the domain experts is still indispensable for the validation and extension of the automatically built ontology, hence rendering the approaches semi-automatic. During the ontology building process, significant information sources like Wikipedia [13] can be used, since Wikipedia is a community-created knowledge base known to facilitate several applications as reviewed in [14]. To illustrate, the domain ontology for wind energy is built based on the Wikipedia articles through a semi-automatic process, as described in [8].

In this paper, we propose an ontology for the domain of electrical energy, with weighted attributes, where the engineering process of this ontology is based on Wikipedia as the main information source. The existing ontologies proposed for the corresponding subdomains are aligned with the ultimate ontology whenever applicable. The ontology is mainly proposed to facilitate related text analysis applications and sample applications are implemented to demonstrate the utilization of the ontology and linked ones, in such settings. The main contributions of the current study are summarized below:

☆ Handled by C.-H. Chen.

* Tel.: +90 312 2101830; fax: +90 312 2101033.
E-mail address: dilek.kucuk@tubitak.gov.tr

http://dx.doi.org/10.1016/j.aei.2015.04.002
1474-0346/

- A high-level ontology with weighted attributes for the domain of electrical energy is proposed. The electrical energy is a significant engineering domain and has important domain-specific application areas, hence a semantic resource with a considerable coverage can contribute to these applications. Particularly, the weighted attributes of the ontology can facilitate many domain-specific text analysis applications. The ontology is based on the information already available within the community-created Wikipedia articles and hence it can be extended or modified, following the related changes in the Wikipedia articles.
- The previously proposed ontologies for the related subdomains of electrical power quality [7] and wind energy [8] are seamlessly aligned with the ultimate electrical energy ontology.
- The contribution of the proposed ontology to text analysis applications is demonstrated through the implementation of two sample applications, namely a text categorization system for scholarly articles and another one for social media texts, as will be clarified in Section 4 of the paper.

The rest of the paper is organized as follows: In Section 2, the domain ontologies for two subdomains of electrical energy, namely, electrical power quality and wind energy, are reviewed. Section 3 presents the high-level domain ontology for electrical energy, which is also linked with the aforementioned two ontologies. In Section 4, two text analysis applications making use of the ultimate domain ontology are described and finally Section 5 concludes the paper with a summary of the main points of the paper and directions of future work.

2. Existing ontologies related to the electrical energy domain

In this section, we overview two ontologies related to the electrical energy domain: the first one is an ontology for electrical power quality, called PQONT [7], and the second one is an ontology for wind energy, which is henceforth referred to as WONT [8]. After these overviews given in the first two subsections below, the motivation for building the domain ontology for electrical energy is presented in the last subsection.

2.1. Domain ontology for electrical power quality (PQONT)

PQONT is an ontology for the domain of electrical power quality (PQ) which is manually engineered by the domain experts, making use of related textbooks and international standards. It basically models the PQ parameters which are used to assess the quality of the electrical energy transmitted and distributed within the electricity grid. Hence, the main PQONT concepts[1] correspond to the continuously measured PQ parameters such as frequency, power, harmonics, flicker, in addition to PQ problems such as sags, swells, and interruptions, among others. The taxonomy of the *PQParameter* concept of PQONT, which is the core class of the ontology, is illustrated in Fig. 1, as excerpted from [7].

Each of the concepts within PQONT shown in Fig. 1 has two particular attributes of *label* and *synonymSet*, in addition to their domain-specific attributes. The *label* attribute holds the natural language expression corresponding to the concept and the *synonymSet* attribute is used to hold the list of synonyms that can be used to refer to this concept within natural language texts. Hence, these two attributes are mainly included within the ontology to facilitate related domain-specific text analysis applications. To illustrate, the concept of *Sag* (a subclass of *PQProblem* concept which in turn is a subclass of *PQParameter* concept at the top level) of PQONT which is defined in the ontology as "*temporary reduction of the voltage below a predetermined threshold*" has *sag* as the value of its *label* attribute and has {*voltage sag, dip, voltage dip*} as the value of its *synonymSet* attribute.

PQONT has been used by a natural language interface to query data from a PQ database where this interface basically exploits the values of the aforementioned text-related attributes of PQONT concepts to detect the domain-specific terms within a query expression posed through the interface [7]. For instance, if a user submits a query of the form "*get me the list of voltage dips at X transformer substation between the dates of D_1 and D_2*", then PQONT is used to detect the "*voltage dips*" phrase within the query and map it to the *Sag* concept (as *voltage dip* is included within the value of the *synonymSet* value of the concept). Then, the database table corresponding to the *Sag* concept is queried to retrieve and to supply the user with the list of PQ problems, of *sag* type, recorded for the queried *X transformer substation* within between the dates of D_1 and D_2 [7]. In order for PQONT and the interface to support languages other than English, similar text-related attributes can be added to PQONT. For instance, to make PQONT and the interface applicable to Turkish, the attributes of *translationInTurkish* and *synonymSetInTurkish* are added to PQONT to hold the natural language expression and the list of synonyms of the concept in Turkish, respectively. PQONT has been made publicly available for research purposes at http://www.ceng.metu.edu.tr/~e120329/PQONT.owl as a Web Ontology Language (OWL) file [7].

2.2. Domain ontology for wind energy (WONT)

WONT is a domain ontology for the wind energy domain, similar to PQONT, but has been built through a semi-automatic procedure using the related Wikipedia article contents [8] as opposed to PQONT which has been manually engineered. Basically, starting with the Wikipedia article on "*wind power*" (http://en.wikipedia.org/wiki/Wind_power), the procedure extracts the frequent n-grams from this article and the articles linked to it where stopwords and named entities (i.e., person, location, and organization names) are not considered during the n-gram extraction phase. Then these n-grams are used to form the concepts, their attributes, and the values of the *label* and *synonymSet* attributes, like the corresponding attributes of PQONT, whenever applicable. Hence, the ultimate form of WONT includes the necessary concepts to model the basic components of the typical *wind power plant* (also commonly referred to as a *wind farm*) as well as the concepts to model the meteorological data like *wind speed*, *wind direction*, *temperature*, and *pressure* [8].

The taxonomy of WONT concepts is provided in Fig. 2 as excerpted from [8] and WONT is made publicly available at http://www.ceng.metu.edu.tr/~e120329/WONT.owl as an OWL file.

2.3. Motivation for building the electrical energy ontology

PQONT and WONT are two significant ontologies modeling the subdomains of the larger domain of electrical energy. Yet, to the best of our knowledge, the domain of electrical energy lacks an ontology which can be utilized in several application settings, and PQONT and WONT only cover, though important but, the semantics of two subdomains of electrical energy. To illustrate, the following considerably recent developments within the electrical energy domain call for the related semantics to be modeled in a convenient ontological structure:

- The share of the use of *renewable energy sources* within the overall electrical energy generation keeps increasing. Hence, power plants using these sources like wind power plants, solar plants,

[1] The terms *concept* and *class* both refer to *ontology concepts* and are used interchangeably throughout the paper.

Please cite this article in press as: D. Küçük, A high-level electrical energy ontology with weighted attributes, Adv. Eng. Informat. (2015), http://dx.doi.org/10.1016/j.aei.2015.04.002