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.

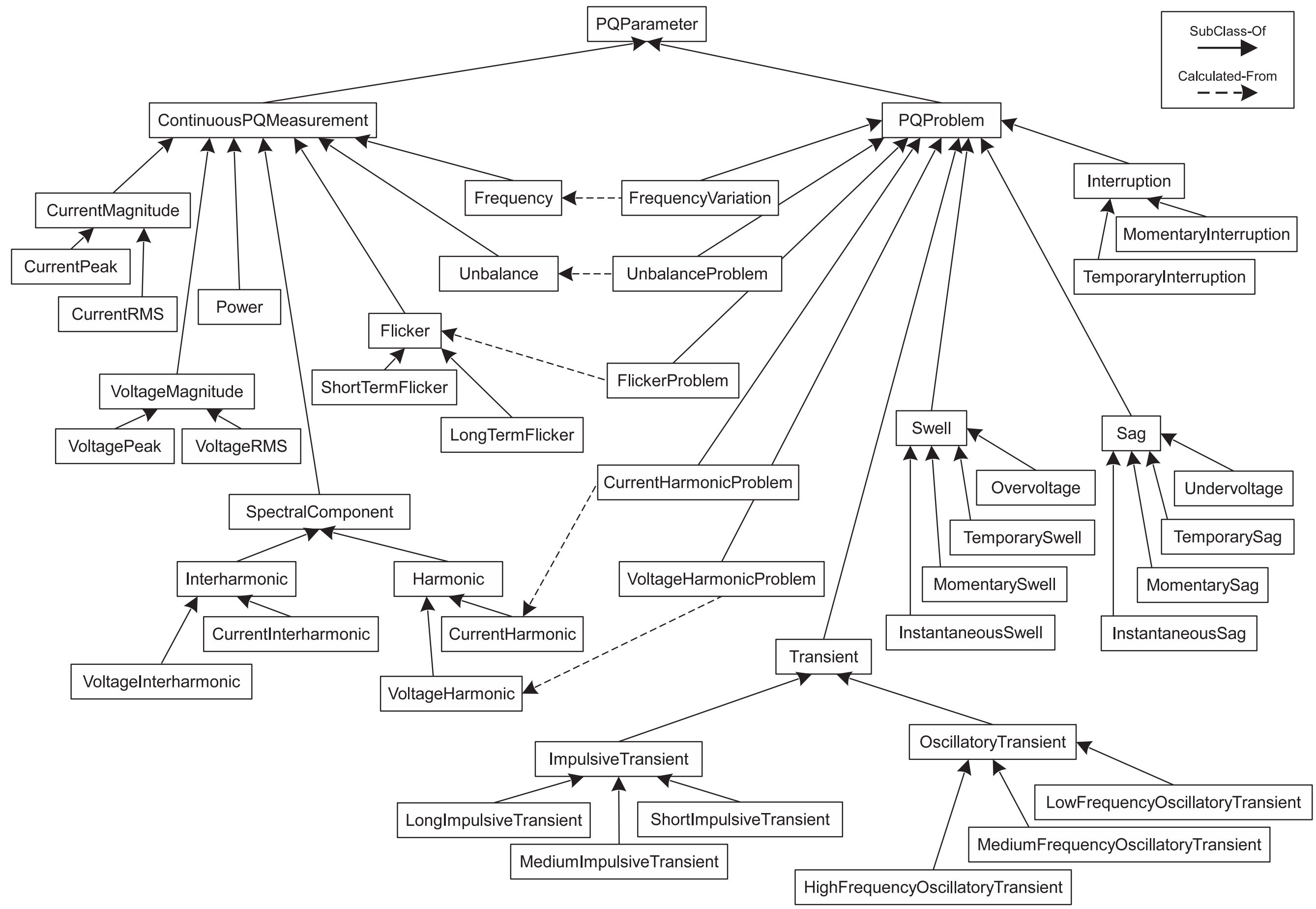

Fig. 1. The taxonomy of the PQParameter concept of PQONT.

among others, have become more ubiquitous and related research problems like determining the optimal locations to install these plants, forecasting the generation by these plants, and their seamless integration to the electrical grid[2] have started to arise. Interested readers are referred to [16,17] for the related literature on research problems regarding the wind-generated power.

- In order to improve the generation, transmission, distribution, and utilization of the electrical energy by making these procedures more efficient and reliable, the concept of *smart grid* has emerged. It is usually defined as an enhanced electrical grid in which conventional components are replaced with smart components, which make use of robust communication and information technology [18,19].
- Various software tools are employed in a typical electrical grid for the monitoring and control of the grid including Supervisory Control And Data Acquisition (SCADA) systems, energy management systems (EMSs), PQ monitoring systems like [20], wide area monitoring systems, and energy forecasting systems, among others. Ontological modeling of the semantics of the underlying electrical data is crucial to help alleviate the effects of possible interoperability problems between these systems. The modeling of electrical devices as related to the domain of the electrical power systems at a certain granularity level will also similarly help enhance the interoperability of other generic tools for the domain, like stock monitoring systems.
- A common vocabulary of the terms within the domain of electrical energy has many application areas related to text analysis, including domain-specific information extraction and text classification systems, and semantic portals, among others. A conveniently created ontology including the domain terms with the interconnections between these terms is an indispensable domain-specific resource for these applications.

3. A high-level domain ontology for electrical energy

Before describing our process to build the domain ontology for electrical energy, in the first subsection below, we present our procedure to extend the two existing ontologies of the energy domain, namely PQONT (Section 2.1) and WONT (Section 2.2), with weighted attributes to arrive at their ultimate versions, called FPQONT and FWONT, respectively[3] as this procedure is also employed during the building process of our electrical energy ontology. The details of the building process of the high-level electrical energy ontology, which is linked to FPQONT and FWONT, are

[2] The electrical grid is defined as an interconnected network through which electrical energy generated in power plants are transmitted to the customers [15]. The main components of the electrical grid are, therefore, the generation subsystem (where energy is produced from various energy sources), the transmission subsystem (which transmits the generated energy from the generation subsystem to the distribution subsystem after making the necessary conversions with transformers), the distribution subsystem (which transmits the received energy to the consumers), and finally the utilization subsystem which comprises the customers utilizing the energy like industrial factories and houses of people.

[3] A preliminary description of this extension procedure and the extended forms of these ontologies are previously described in [21] together with a text categorization system based on the extended ontologies.

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

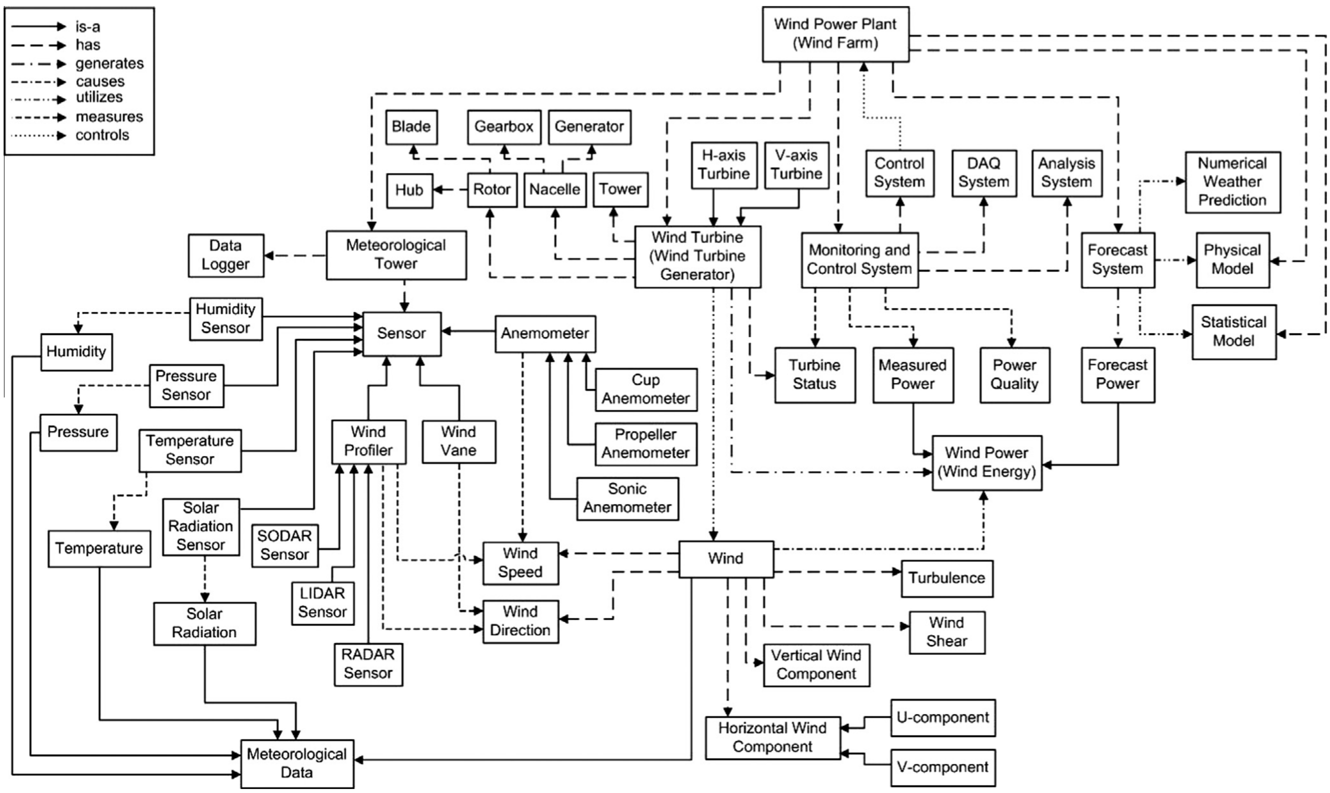

Fig. 2. The taxonomy of the concepts within WONT.

presented in the next subsection and a summary of the main features of the electrical energy ontology is given in the last subsection.

3.1. Extending PQONT and WONT with weighted attributes

The text-related attributes of PQONT and WONT, *label* and *synonymSet*, can seamlessly be utilized to facilitate domain-specific applications like the natural language interface to the electrical PQ database, which relies on PQONT to determine the domain terms within the user queries, as described in [7]. Yet, in relatively more open-domain applications like document/text classification systems, the homonymy of the values of these attributes to other words may cause problems. For instance, although *frequency* is defined as the "*the number of oscillations of the current transmitted in an electrical grid*" within the PQ domain, it has several other definitions such as the "*the number of times an event occurs*". Therefore, the appearance of the word "*frequency*" (the value of the *label* attribute of the *Frequency* concept of PQONT) in a document cannot be considered as strong a clue as "*voltage harmonic*" (the value of the *label* attribute of *VoltageHarmonic* concept) when deciding whether the document is related to the PQ domain, or not, since the latter phrase does not have other uses than the one related to this domain.

The above observation, which is also applicable to the concepts of WONT, leads us to extend PQONT and WONT with weighted attributes where the newly added attributes of *membershipValueLabel* and *membershipValuesSynonymSet* are used to hold the membership values (to the domain) for the corresponding values of the *label* and *synonymSet* attributes of the ontology classes. In order to fill in these membership values, Wikipedia disambiguation pages for the applicable concepts are used as follows: the *reciprocal rank*[4] of the use corresponding to the domain definition is used as the value of the corresponding value, i.e., if the nth use in the disambiguation page for a concept label or synonym corresponds the domain definition, then $1/n$ is used as the value of the *membershipValueLabel* attribute or that of the corresponding element within the value of the *membershipValuesSynonymSet* attribute [21]. For instance, at the time of writing this paper, in the disambiguation page for the *Frequency* concept (http://en.wikipedia.org/wiki/Frequency_(disambiguation)), the 3rd use corresponds to the one within the PQ domain, hence, the *Frequency* concept of extended PQONT, namely FPQONT, has 0.33 as the value of its *membershipValueLabel* attribute. The values for the *membershipValuesSynonymSet* attributes are similarly filled in.

The formal definitions of these weighted attributes of PQONT and WONT are provided below:

Definition 1. The weighted attribute *membershipValueLabel* corresponding to the *label* (l) attribute of a concept in an ontology is a membership function $\mu(l)$ defined as follows, where $wiki_d(l)$ denotes the Wikipedia disambiguation page for l and $wiki_a(l)$ denotes the Wikipedia article page for l exactly describing the ontological sense of l:

$$\mu(l) = \begin{cases} 1.0 & \text{if neither } wiki_d(l) \text{ nor } wiki_a(l) \text{ exists, or only } wiki_a(l) \text{ exists} \\ 1/n & \text{if the use matching the ontological sense of } l \text{ is ranked } n\text{th in } wiki_d(l) \\ 0.0 & \text{otherwise} \end{cases}$$

Definition 2. The list-valued weighted attribute *membershipValuesSynonymSet* corresponding to the list-valued *synonymSet* ($ss(l) = \{s_1, s_2, \ldots, s_n\}$) attribute of a concept in an ontology is a membership function $\omega(ss)$ which returns the list of $\mu(s_i)$ (defined above) values for each synonym s_i of the concept: $\omega(ss) = \{\mu(s_1), \mu(s_2), \ldots, \mu(s_n))\}$

The extended versions of PQONT and WONT, namely FPQONT and FWONT, which have been formed through this procedure, are made publicly available at http://www.ceng.metu.edu.tr/

[4] *Reciprocal rank*, or *first answer reciprocal rank*, is a commonly used metric during the evaluation of information retrieval and question answering systems [22,23].

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

~e120329/FPQONT.owl and http://www.ceng.metu.edu.tr/~e120329/FWONT.owl, respectively [21].

3.2. The building process of the domain ontology for electrical energy

In this paper, we do not aim at a wide-coverage ontology which will help related software tools to interoperate (as pointed our in the third item above), because very detailed standards exist for such purposes as the Common Information Model (CIM) standards of International Electrotechnical Commission (IEC) (like [24,25]). Instead, we aim at creating a considerably high-level ontology for the electrical energy domain to satisfy the remaining items above, especially the last one, using the Wikipedia articles related to the domain. This domain ontology also covers the previously proposed ontologies for its subdomains. Since the natural language expressions corresponding to the concepts within the electrical energy domain can be applicable to other unrelated domains, the ultimate ontology for electrical energy is accordingly extended with the weighted attributes to model this phenomenon as described in Section 3.1 for extending PQONT and WONT. Hence, the ultimate Fuzzy Electrical Energy ONTology (henceforth, FEEONT) is created as follows:

1. Following the semi-automated ontology engineering procedure described in [8], used to create WONT, we extract frequent n-grams from the related Wikipedia articles pages for the terms including *electrical energy, electrical power system, power electronics, power station, electrical grid, electricity generation, and renewable energy* and also from the article pages linked to these pages. These n-grams are then used during the determination of the FEEONT concepts.
2. The existing classifications available within the texts of the aforementioned article pages are considered when creating the corresponding class hierarchies of the concepts of FEEONT. Similarly, *part-of* relations between concepts are manually extracted from the article texts, whenever possible.
3. The text-related attributes of *label* and *synonymSet* of PQONT and WONT are included as concept attributes within FEEONT as well. In order to fill in the values for the *label* and *synonymSet* attributes, the names and synonyms usually available at the top of the Wikipedia articles are used. To illustrate, at the time of writing this paper, the initial sentence of the Wikipedia article for the term *power station* (http://en.wikipedia.org/wiki/Power_station) reads as follows: "*A* **power station** *(also referred to as a* **generating station**, **power plant**, **powerhouse**, *or* **generating plant***) is an industrial place for the generation of electric power*". Among the phrases written in boldface, *power plant* is used as the value of the *label* attribute of the corresponding FEEONT concept (*PowerPlant*) and the remaining ones (i.e., {*power station, generating station, powerhouse, generating plant*}) are used to fill in the value of the *synonymSet* attribute of this concept.
4. The attributes introduced to extend PQONT and WONT, i.e, *membershipValueLabel* and *membershipValuesSynonymSet*, are also included within FEEONT. The same procedure based on the utilization of Wikipedia disambiguation pages, as described in Section 3.1, is used to fill in the values for these attributes. Hence, the values of these weighted attributes can readily be used to facilitate related text analysis applications.
5. In order to increase the extent of the applications that FEEONT can contribute to, for each of its concepts we have also included two class (annotation) attributes, namely, *wikiURL* and *dbpediaURL*. As the names of these attributes indicate, they are used to hold the URL of the corresponding Wikipedia [13] article and that of the DBpedia [26] page for the applicable ontology concepts, respectively. Thereby, the ontology concepts are disambiguated and the ontology can further be extended by processing these linked resources.
6. To employ PQONT in multilingual settings, further two language-specific attributes had been added, namely, *translationInTurkish* and *synonymSetInTurkish* which correspond to the values of *label* and *synonymSet*, this time in Turkish [7]. Following that approach, we have similarly added two attributes, *labelTR* and *synonymSetTR* to FEEONT to hold the corresponding values in Turkish. As there is a need to model the relevance of these new values to the domain, we have included two other weighted attributes, namely, *membershipValueLabelTR* and *membershipValuesSynonymSetTR*, to hold the membership degrees of the concept labels and synonyms in Turkish to the domain. Lastly, to hold the URL of the corresponding Wikipedia article in Turkish, we have included an additional attributed called *wikiURLTR*. These five language-specific attributes are included to illustrate how inclusion of a small set of language-specific attributes helps make use of FEEONT in multilingual text analysis applications.
7. Named entity recognition (NER), which is usually described as the extraction of person, location, and organization names, is an important text analysis task and many practical text analysis applications include a NER system as a sub-component. To this end, we have included in FEEONT several classes which can be considered as named entities of organization name type together with their instances (objects). Related organization classes (in gray) and some instances corresponding to the abbreviations of related Turkish organizations (in white) are provided in Fig. 3.[5] Similar to the previous item above, these classes and instances have the *wikiURL* attribute that is used to hold the Wikipedia article page for the concept or the actual instance. For example, the value of the *wikiURL* attribute of TEİAŞ in Fig. 3 is http://tr.wikipedia.org/wiki/Türkiye_Elektrik_İletim_A.Ş.

The taxonomy of the concepts within the ultimate form of FEEONT, which is created following the above steps, is presented in Fig. 4. We should note that FEEONT is a domain ontology that is semi-automatically constructed using the related Wikipedia articles and it is not claimed to be a complete ontology. There is always room for extending it and correcting possible errors that can be found within. Furthermore, it needs manual/automatic revisions as related Wikipedia articles are edited and new article pages related to the domain are created.

For the purposes of our study, FEEONT is designed as a considerably high-level ontology, in which several concepts (for instance, the concept of *PowerPlant*) can further individually be modeled with sub-ontologies as there are several subconcepts that are not covered in details. Furthermore, only the basic relations of *is-a*, *is-part-of*, *measures*, *is-generated-in*, *is-utilized-in*, and *happens-in* are modeled within FEEONT and not all of the instances of these relations are demonstrated in Fig. 4 due to illustrative reasons. Because, several sub-concepts of the *ElectricalPowerSystemComponent*, like *PowerConverter*, are actually physically located in its other sub-concepts, like *ElectricalSubstation*, yet, it such relations are not shown in the taxonomy in order not to compromise the presentation of the ontology.

[5] The open forms are these organization name abbreviations are as follows:
EPDK: Enerji Piyasası Denetleme Kurulu (*Energy Market Regulatory Authority*)
EÜAŞ: Elektrik Üretim A.Ş. (*Electricity Generation Company*)
TEİAŞ: Türkiye Elektrik İletim A.Ş. (*Turkish Electricity Transmission Company*)
TÜREB: Türkiye Rüzgar Enerjisi Birliği (*Turkish Wind Energy Association*)
YEGM: Yenilenebilir Enerji Genel Müdürlüğü (*General Directorate of Renewable Energy*).

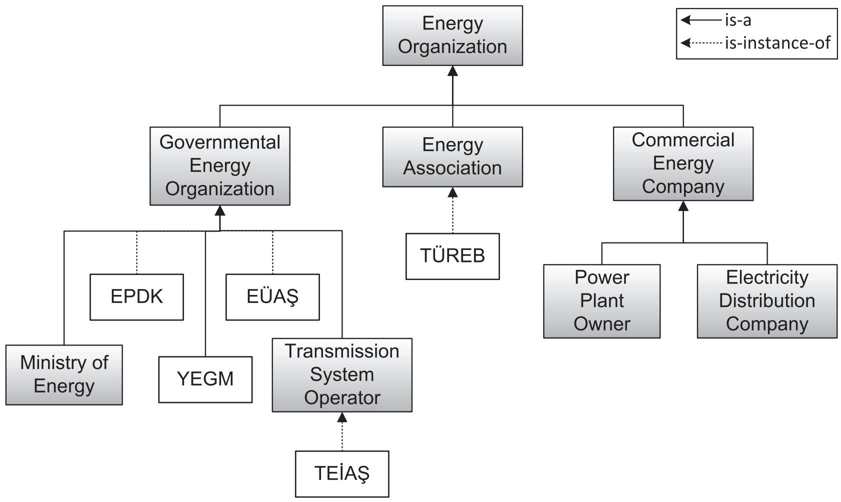

Fig. 3. A simple taxonomy of energy organization classes in FEEONT with instances.

The values for the attributes of two FEEONT classes, namely *PowerPlant* and *ProgrammableLogicController* are provided in Table 1 and in Table 2, respectively, for illustrative purposes.

As we have pointed out above, some of the FEEONT concepts can well be further modeled by other sub-ontologies. Below, we discuss the implemented/prospective alignment of related sub-ontologies to FEEONT:

- The concept of *ElectricalPowerQuality*, shown in gray to distinguish it in Fig. 4, is related to the domain of electrical PQ which has already been modeled by PQONT and its successor version, FPQONT, as presented in the previous sections. Since PQONT and FPQONT are modeling the measured PQ parameters, they can be integrated into FEEONT through the *ElectricalPowerQuality* concept. Basically, we align FPQONT and FEEONT by linking the *PQParameter* class of FPQONT with the *ElectricalPowerQuality* class of FEEONT.
- Similarly, the concepts of *Wind* and *WindPlant* (shown in gray in Fig. 4) of FEEONT correspond to the *Wind* and *WindPowerPlant* concepts of WONT and its successor FWONT. These matching concepts within FEEONT are also linked with the related FWONT concepts, thereby, FWONT and FEEONT are also aligned.[6]
- Another possible way of conveniently extending the coverage of FEEONT is to integrate an ontology for the domain of the smart grid technologies and include the concepts of this ontology through the *SmartGrid* concept (shown in gray in Fig. 4) of FEEONT. This extension is currently left as a direction of future work.

Following the above points related to its design, FEEONT is implemented using the *Protégé* ontology editor [28,29] and its final form is made available at http://www.ceng.metu.edu.tr/e120329/FEEONT.owl for research purposes. To the best of our knowledge, FEEONT stands as the first domain ontology proposed for the domain of electrical energy and a convenient model for the semantics of this domain is a significant contribution due to the impact of the domain and its wide range of related application areas.

3.3. A summary of the main features of the proposed ontology

The ultimate FEEONT ontology for the electrical energy domain has the following significant features which make it an important knowledge source for the domain.

1. *High-Levelness*: FEEONT is a high-level ontology with significant coverage for the electrical energy domain. Its coverage and contribution to practical applications are illustrated in Section with two applications. There is always room for extending FEEONT and aligning it with the related energy ontologies.
2. *Weighted Attributes*: FEEONT has weighted attributes to better model the inherent ambiguities in the textual representations of its concepts. As described in details in the previous subsection, these attributes model the relevance of the values of the labels and synonyms of the ontology concepts to the domain.
3. *Multilinguality*: In order to illustrate its employment in multilingual text analysis applications, FEEONT is equipped with sample attributes for Turkish to hold the Turkish labels and synonyms for the ontology concepts. Similar attributes can be included in the ontology for other languages in a similar fashion, possibly making use of machine translation systems together with a subsequent expert revision.
4. *Links to Related Resources*: FEEONT is linked to related FPQNT and FWONT ontologies, as well as to the collaborative resources of Wikipedia and its Turkish version, and lastly to DBpedia which is available as a linked open data source. As a future work based on the current study, we plan to link FEEONT with other related domain resources, which provide linked open energy data, like *reegle* [30], a portal for clean energy and *OpenEI* [31] which is an initiative to publish energy-related information. In Fig. 5, FEEONT's current links to related resources are shown with solid arrows and its prospective links considered as future work are shown with dotted arrows. As a statistical summary, out of all 110 classes in FEEONT, 69 of them are linked to their Wikipedia and DBpedia pages and 36 of these 69 classes are also linked to their Turkish Wikipedia pages. The links to DBpedia and Turkish Wikipedia pages are formed automatically based on the Wikipedia links.

4. Text analysis applications exploiting the energy ontologies with weighted attributes

In this section, we present two applications that make use of the extended ontologies described so far. The first one is a text

[6] In this work, we align the ontologies instead of merging them, and hence introduce links from the corresponding concepts of FEEONT to that of FPQONT and FWONT, following the definitions of ontology alignment and merging in [27].

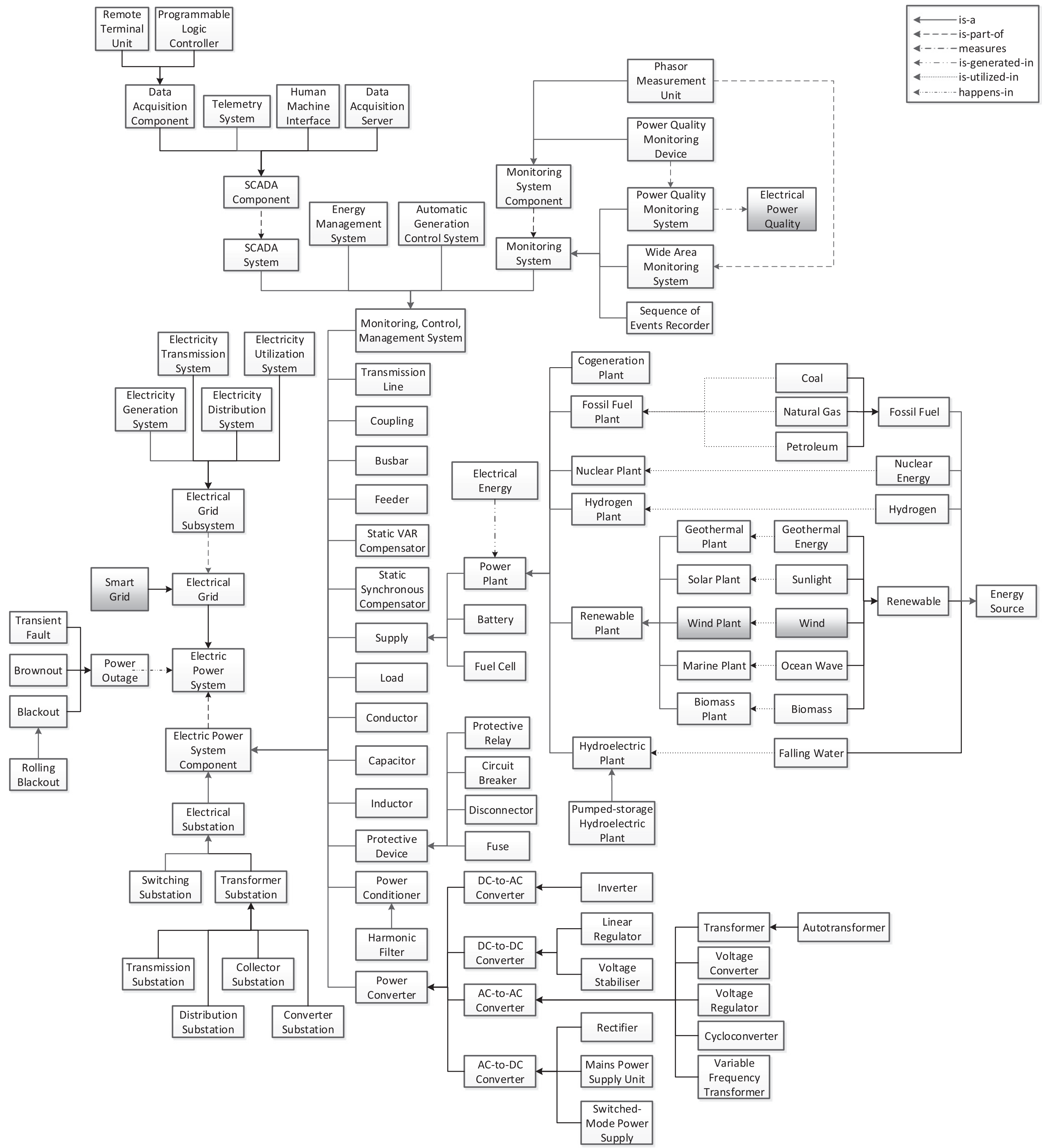

Fig. 4. The taxonomy of the concepts within FEEONT.

categorization system for scholarly articles based on FPQONT and the second one is a similar social media analysis application based on FEEONT. The details of these applications are provided in the corresponding subsections below.

4.1. A text categorization system for scholarly articles

Based on the concepts within FPQONT, we have implemented a text categorization system for scholarly articles which categorizes a given article as relevant to the PQ domain or not.[7] We have used the below relevance score in our categorization system, making use of the weighted attributes of FPQONT:

[7] This application has been previously described in [21] together with the extensions to PQONT and WONT, which are summarized in Section 3.1 of the current paper.

Table 1
The details of the *PowerPlant* class.

Class attribute	Value
dbpediaURL	http://dbpedia.org/page/Power_plant
label	power plant
labelTR	elektrik santrali
membershipValueLabel	1.0
membershipValueLabelTR	1.0
membershipValuesSynonymSet	{1.0, 1.0, 1.0, 1.0, 1.0, 1.0}
membershipValuesSynonymSetTR	{1.0, 1.0, 1.0}
synonymSet	{power station, generating station, powerhouse, generating plant, energy plant, energy station}
synonymSetTR	{enerji santrali, güç santrali, santral}
wikiURL	http://en.wikipedia.org/wiki/Power_plant
wikiURLTR	http://tr.wikipedia.org/wiki/Elektrik_santrali

Table 2
The details of the *ProgrammableLogicController* class.

Class attribute	Value
dbpediaURL	http://dbpedia.org/page/Programmable_logic_controller
label	programmable logic controller
labelTR	programlanabilir mantıksal denetleyici
membershipValueLabel	1.0
membershipValueLabelTR	1.0
membershipValuesSynonymSet	{1.0, 0.033}
synonymSet	{programmable controller, PLC}
wikiURL	http://en.wikipedia.org/wiki/Programmable_logic_controller
wikiURLTR	http://tr.wikipedia.org/wiki/PLC

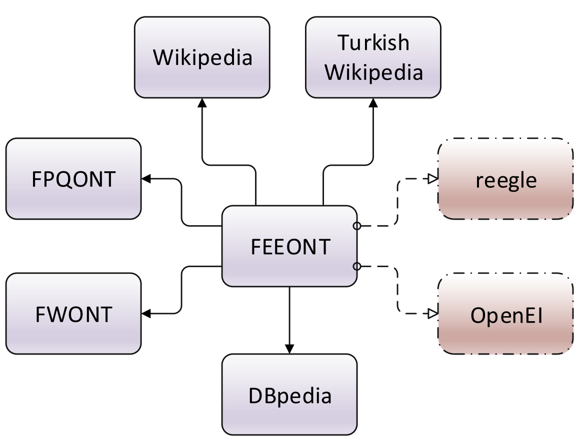

Fig. 5. The existing and prospective links between FEEONT and related resources.

$$relevance\ score = \sum_{i=0}^{c}\left[freq(l_i) * \mu(l_i) + \sum_{j=0}^{ss_i} freq(s_{i,j}) * \mu(s_{i,j})\right]$$

To clarify the symbols in the above formula: c is the number of concepts in FPQONT, l_i denotes the *label* value of ith concept, $\mu(l_i)$ is the value of the *membershipValueLabel* of the ith concept, ss_i is the number of synonyms for the ith concept (e.g., the size of the value of its *synonymSet* attribute), $s_{i,j}$ is the jth element within the value of *synonymSet* of the ith concept, $\mu(s_{i,j})$ is the jth element within the value of *membershipValuesSynonymSet* of the ith concept (e.g., the membership value of the jth synonym to the corresponding concept), and $freq(x)$ denotes the frequency of the phrase (a label or synonym), x, within an article [21].

Table 3
Performance evaluation results of the baseline and the FPQONT-based categorizers [21].

	True positives	False negatives	True negatives	False positives	Accuracy (%)
Baseline categorizer	10	0	86	14	87.27
FPQONT-based categorizer	10	0	99	1	99.09

The implemented categorization system classifies a given article as relevant to the PQ domain, if the *relevance score* for the article is greater than a threshold. This threshold value is taken as 1.0 in our experiments. In order to test the categorization system, we have compiled a text corpus of 110 article abstracts (with a total of about 17,700 words) comprising 11 subsets of 10 articles, corresponding to 11 distinct research domains where one of these domains is the electrical PQ domain [21].

In order to compare the performance of the proposed text categorizer, we have also implemented a baseline categorizer which instead takes all membership values (for labels and synonyms) as 1.0 during the calculation of the *relevance score*. This baseline categorizer also takes the threshold value as 1.0 during categorization. The performance evaluation results of these categorizers are given in Table 3, with breakdowns of their outputs (in columns 2–5) and their accuracies (in the last column).

The evaluation results in Table 3 show that both categorizers are good at classifying truly relevant articles as *relevant* reflected with their correct categorization of all 10 articles on electrical PQ as *relevant*. For these articles, the baseline categorizer calculates far higher *relevance score* values (ranging between 7.0 and 17.0) compared to the ones calculated by the FPQONT-based categorizer (ranging between 1.5 and 8.084) which is an expected result considering the fact that the latter one uses weights between [0–1] for the encountered domain terms while the former one uses the crisp weight of 1.0 for all such terms. Regarding the not-relevant 100 articles in our data set, the baseline categorizer incorrectly classifies 14 of them as *relevant* while the FPQONT-based categorizer only misclassifies one of them. Although our corpus of scholarly articles is limited in size, the FPQONT-based categorizer achieves an improvement of about 13.5% in accuracy over the baseline categorizer. Thereby, the proposed categorizer demonstrates the practical contribution of FPQONT to real text analysis applications.

4.2. A social media analysis application for electrical energy

Different social media platforms like Twitter, Facebook, and LinkedIn enable many people around the world to instantly publish information. Since textual information produced through these applications is huge in size[8] and contains very important information for trend analysis and monitoring applications, among others, significant research on social media analysis has started to emerge [32]. These works include those ones on named entity recognition on tweets [33], event extraction from tweets [34], tweet summarization [35], and cross-media linking like interlinking news articles and social media texts [32].

In this subsection, we present the evaluation results of a tweet categorization application that we have implemented based on our categorization scheme presented in the previous subsection. To this end, we have used FEEONT's (and linked ontologies') labels and synonyms in Turkish to relevant extract tweets from a compiled set of tweets in Turkish. This tweet set comprises about

[8] For instance, as it is reported in Twitter blog (https://blog.twitter.com/2013/new-tweets-persecond-record-and-how), more than 500 million tweets are produced per day.

200,000 tweets in Turkish which are compiled randomly for one hour each day during the consecutive 28-day period between the dates of 16th of November 2013 and 13th of December, 2013.

Using this tweet categorization scheme, the application has classified 110 tweets as *relevant* to the electrical energy domain. In order to evaluate the performance of the system, we have used the precision metric, which is the ratio of the number of true positives over all items classified as *relevant* (hence, the denominator being 110 in our case). The number of true positives is found as 73 and hence the precision of the tweet categorization scheme implemented is 66.36%.

Although the number of tweets classified as *relevant* in the whole tweet set (about 0.055%) is quite low which may be due to some missed relevant tweets, such missed tweets can be covered after FEEONT and the linked domain ontologies are further extended. The precision rate is not very high but is still promising as the performance result of the first such application, to the best of our knowledge. Most of the false positives are due to the use of some domain terms in idiomatic expressions and the homonymy of some terms in Turkish to some foreign words which are used in song lyrics, etc. in other languages within the tweets. Below, we provide three sample tweets that are relevant to the domain which are classified so by our application:

> *Elektrik enerjisi talebi yüzde 5, ithalatı yüzde 27 arttı (Özel)* http://t.co/KxhnHLbH5r
> ('*Electrical energy demand increases by 5 percent and its import by 27 percent (Private)* http://t.co/KxhnHLbH5r')
>
> *Borçka-Ahıska Elektrik İletim Hattı Hizmete Açıldı* http://t.co/kY1KPCWX7l
> ('*Borçka-Ahıska Electricity Transmission Line Is Inaugurated* http://t.co/kY1KPCWX7l')
>
> *Karbona bağlılığımız devam ediyor:PwC Düşük Karbon Ekonomi Endeksi fosil yakıt kullanım artışını inceliyor #carbon* http://t.co/FIGIcaRFsP
> ('*Our dependence on carbon continues:PwC Low Carbon Economy Index is examining the increase of fossil fuel use #carbon* http://t.co/FIGIcaRFsP')

We believe that performing such an experiment on a tweet data set in English may produce far better performance results, as both FEEONT and its linked ontologies have more coverage for the domain terms in English due to the larger size of Wikipedia compared to Turkish Wikipedia in terms of both the number and extent of the articles included. Still, this tweet categorization application for the electrical energy domain can be used in an online fashion within a semantic energy portal so that the users can be informed about the recent relevant tweets published worldwide.

5. Conclusion

In this paper, we propose a high-level domain ontology with weighted attributes for the electrical energy domain which is a significant engineering field with several plausible application areas. We describe the procedures to determine the ontology concepts and to extend the ontology with weighted attributes, mainly based on the Wikipedia articles and disambiguation pages. The existing domain ontologies which have been previously proposed for the subdomains of electrical energy, including a power quality ontology and a wind energy ontology, are also interlinked (aligned) with the ultimate ontology. Finally, two text analysis applications are implemented in which the proposed electrical energy ontology and linked ontologies are employed as the sources of semantic information. The proposed ontology is a high-level one with significant coverage and has weighted attributes to handle fuzziness, multilinguality, and links to other domain ontologies as well as to other semantic information sources like Wikipedia and linked open data sources like DBpedia.

Directions of future work based on the current study include automatic extension of this ontology using Wikipedia articles and other textual Web-based information sources, in addition to linking it with other existing ontologies for significant subdomains such as the smart grid, and also with related portals like *reegle* and *OpenEI*. Further applications can be implemented based on the ontologies and evaluated on larger data sets. Another prospective future study is to automatically turn the electrical energy ontology into a multilingual resource with more coverage, supporting several other languages, by automatic processing of the Wikipedia article pages for different languages.

References

[1] A. Gómez-Pérez, M. Fernández-López, O. Corcho, Ontological Engineering, third ed., Springer-Verlag, 2004.
[2] S. Grimm, A. Abecker, J. Völker, R. Studer, Ontologies and the semantic web, in: J. Domingue, D. Fenseland, J.A. Hendler (Eds.), Handbook of Semantic Web Technologies, Springer, 2011, pp. 507–580.
[3] R. Stevens, C. Goble, I. Horrocks, S. Bechhofer, Building a bioinformatics ontology using OIL, IEEE Trans. Inf. Technol. Biomed. 6 (2) (2002) 135–141.
[4] D. Gašević, V. Devedžić, Petri net ontology, Knowl.-Based Syst. 19 (4) (2006) 220–234.
[5] J. Morbach, A. Yang, W. Marquardt, Ontocape–a large-scale ontology for chemical process engineering, Eng. Appl. Artif. Intell. 20 (2) (2007) 147–161.
[6] T. Asikainen, T. Männistö, T. Soininen, Kumbang: a domain ontology for modelling the variability in software product families, Adv. Eng. Inform. 21 (2007) 23–40.
[7] D. Küçük, O. Salor, T. İnan, I. Çadırcı, M. Ermiş, PQONT: a domain ontology for electrical power quality, Adv. Eng. Inform. 24 (1) (2010) 84–95.
[8] D. Küçük, Y. Arslan, Semi-automatic construction of a domain ontology for wind energy using Wikipedia articles, Renew. Energy 62 (2014) 484–489.
[9] D. Parry, A fuzzy ontology for medical document retrieval, in: Proceedings of the Australasian Workshop on Data Mining and Web Intelligence, 2004, pp. 121–126.
[10] C.-S. Lee, Z.-W. Jian, L.-K. Huang, A fuzzy ontology and its application to news summarization, IEEE Trans. Syst. Man Cybern.–Part B: Cybern. 35 (5) (2005) 859–880.
[11] N. Xue, S. Jia, J. Hao, Q. Wang, Scientific ontology construction based on interval valued fuzzy theory under Web 2.0, J. Softw. 8 (8) (2013) 1835–1842.
[12] L. Drumond, R. Girardi, A survey of ontology learning procedures, in: Proceedings of the 3rd Workshop on Ontologies and Their Applications, 2008, pp. 1–12.
[13] Wiki – Wikipedia. <http://en.wikipedia.org/wiki/Wiki> (accessed 02.04.15).
[14] O. Medelyan, D. Milne, C. Legg, I.H. Witten, Mining meaning from Wikipedia, Int. J. Hum Comput Stud. 67 (9) (2009) 716–754.
[15] Electrical grid – Wikipedia. <http://en.wikipedia.org/wiki/Electrical_grid> (accessed 02.04.15).
[16] T. Burton, D. Sharpe, N. Jenkins, E. Bossanyi, Wind Energy Handbook, first ed., John Wiley & Sons, 2001.
[17] P. Jain, Wind Energy Engineering, first ed., McGraw-Hill, 2011.
[18] F. Li, W. Qiao, H. Sun, H. Wan, J. Wang, Y. Xia, Z. Xu, P. Zhang, Smart transmission grid: vision and framework, IEEE Trans. Smart Grid 1 (2) (2010) 168–177.
[19] Smart Grid – Wikipedia. <http://en.wikipedia.org/wiki/Smart_grid> (accessed 02.04.15).
[20] T. Demirci, A. Kalaycıoğlu, D. Küçük, O. Salor, M. Güder, S. Pakhuylu, T. Atalık, T. İnan, I. Çadırcı, Y. Akkaya, S. Bilgen, M. Ermiş, Nationwide real-time monitoring system for electrical quantities and power quality of the electricity transmission system, Gener. Transm. Distrib. IET 5 (5) (2011) 540–550.
[21] D. Küçük, D. Küçük, A. Yazıcı, On fuzzy extensions to energy ontologies for text processing applications, in: Proceedings of the International Symposium on Computer and Information Sciences, 2014, pp. 87–95.
[22] E.M. Voorhees, D.M. Tice, The TREC-8 question answering track evaluation, in: Proceedings of the Eighth Text REtrieval Conference (TREC-8), 1999, pp. 83–105.
[23] D.R. Radev, H. Qi, H. Wu, W. Fan, Evaluating Web-based question answering systems, in: Proceedings of the Language Resources and Evaluation Conference, 2002, pp. 1–4.
[24] International Electrotechnical Commission (IEC) Standard, IEC 61968-11: Application Integration at Electric Utilities – System Interfaces for Distribution Management – Part 11: Common Information Model (CIM) Extensions for Distribution.
[25] International Electrotechnical Commission (IEC) Standard, IEC 61970-301: Energy Management System Application Program Interface (EMS-API) – Part 301: Common Information Model (CIM) Base.
[26] DBpedia. <http://wiki.dbpedia.org/> (accessed 02.04.15).

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

[27] N.F. Noy, M.A. Musen, An algorithm for merging and aligning ontologies: automation and tool support, in: Proceedings of the Workshop on Ontology Management at the Sixteenth National Conference on Artificial Intelligence (AAAI-99), 1999, pp. 1–11.

[28] The Protégé Ontology Editor and Knowledge Acquisition System. <http://protege.stanford.edu> (accessed 02.04.15).

[29] N.F. Noy, M. Sintek, S. Decler, M. Crubezy, R.W. Fergerson, M.A. Musen, Creating Semantic Web contents with Protégé-2000, IEEE Intell. Syst. 16 (2) (2001) 60–71.

[30] Reegle–Clean Energy Info Portal. <http://www.reegle.info/> (accessed 02.04.15).

[31] OpenEI–Open Energy Information. <http://en.openei.org/wiki/Main_Page> (accessed 02.04.15).

[32] K. Bontcheva, D. Rout, Making sense of social media streams through semantics: a survey, Semantic Web 5 (5) (2014) 373–403.

[33] A. Ritter, S. Clark, Mausam, O. Etzioni, Named entity recognition in tweets: an experimental study, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2011, pp. 1524–1534.

[34] A. Ritter, Mausam, O. Etzioni, S. Clark, Open domain event extraction from twitter, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012, pp. 1104–1112.

[35] D. Chakrabarti, K. Punera, Event summarization using tweets, in: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011, pp. 66–73.

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