

A Conversion Method for SCL to CIM Based on Ontology Matching

Shaomin zhang^{1, a}, Lv Shi^{1, b}, Baoyi wang^{1, c}

¹ School of Control and Computer Engineering, North China Electric Power University, Baoding, 071003, Hebei province, China

^aemail: zhangshaomin@126.com, ^bemail:535403645@qq.com, ^cemail:wangbaoyiqj@126.com

Keywords: CIM, SCL, Ontology matching, The mapping relationship, Data exchange.

Abstract. There are structure and content differences between SCL, the substation configuration language defined in IEC61850, and CIM, the common information model contained in IEC61970, which lead to the data exchange barriers between the substation system with the power dispatching center. While the traditional configuration files conversion scheme based on model mapping exists a low efficacy problem in the search for mapping relationships between the two models. For this reason, we construct ontologies for SCL and CIM, establish semantic mappings between SCL and CIM by using ontology matching algorithm in which the rich ontology element features of SCL and CIM are utilized. And according to these mapping relationships, we realize the conversion of the substation instance configuration files SCD/XML to CIM/XML automatically, improving the automation level of the model files conversion.

Introduction

In electric power system, IEC61850 and IEC61970 are two sets of parallel models. IEC61850 is an international standard for the new generation of substation automation system. The standard uses the object-oriented technology to establish a complete semantic information model, and the model of the semantic information is described in the form of the substation configuration language SCL. However, in the dispatch center, the energy management system (EMS) follows the IEC61970 standard, and uses the public information model CIM to describe the power grid model. Since the SCL and CIM has different using scope as well as different designated system, resulting in the data exchange and information sharing in the dispatching center and substation system are not smooth [1]. Therefore, the integration of the two standards is the key to solve the problem of information sharing between the dispatching center and the substation.

Current research on the integration of the two standards is mainly based on the model mutual conversion. Both literatures [2] [3] are transformation scheme based on the model mapping. However there exist a low efficacy in finding the mapping relationships between the two models by pure manual. In this paper, we construct SCL and CIM ontologies by using the modeling tool Protégé, and then adopt the method of ontology matching to find the mapping relationships between CIM and SCL, according to these mapping relationships, we implement the conversion of the SCD/XML document to the CIM/XML document, which improve the level of automation for model files transformation.

The difference analysis between CIM and SCL model

The description for the same object of substation in CIM and SCL provides a realistic possibility for the conversion of the two models [4]. The key differences between CIM and SCL are as follows [5] [6] [7]:

(1) Type differences of conducting equipment classes

Various types of conducting equipments are inherited from the CIM class ConductingEquipment, and instanced by the objects of subclass, Such as Breaker on behalf of circuit breakers; however, in the SCL only exist conducting equipment class tConductingEquipment, and not define specific sub classes to represent the specific conducting equipments, only according to the different enumeration values of tConductingEquipment to indicate the specific conducting equipments. such as type=CBR

said circuit breakers.

(2) The differences of the object identification

Each object in CIM is identified by a unified resource locator URI, while SCL defines the main resource identifier attribute (MRID) for all logical nodes associated with the once equipments, using the path+name uniquely identifies an object.

(3) The differences of the attributes

Different classes defined in the two models have different properties, even if the same class, the properties are also different. For example, the tConductingEquipment in SCL has attributes of name, desc, virtual etc. While the ConductingEquipment in CIM defines more subclasses attributes, such as Breaker, Disconnecter etc. The number of attributes is much more than that of the SCL.

The mapping between CIM and SCL based on ontology technology

Ontology matching

Ontology matching: Ontology matching aims to discover a three tuple $M' = \langle O, O', M \rangle$, which contains two ontologies O and O' to be matched, produces a matching result set $M = \{m_1, m_2, \dots, m_n\}$, in which $m_i (1 \leq i \leq n)$ represents a basic mapping unit and it can be written as a four tuple $m_i = \langle id, e, e', s \rangle$: id is the unique identifier of a matched element for a given matching unit, e and e' are entities respectively belongs to O and O' , and there are $(e, e') = M.s$ represents the degree of the similar between e and e' , there are $s \in [0, 1]$. A complete ontology matching framework is used to support iterative matching, so as to make full use of the existing matching results, so the matching result set M is also used as an input. The ontology matching is based on the consideration of the parameters setting and the external resource situation to find the final matching result set M' . Parameters are the threshold and weight, external resources refer to domain ontology, background knowledge, dictionary, etc. The matching process is shown in Figure 1:

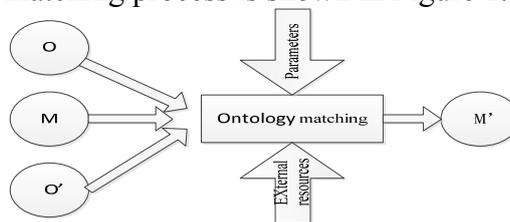


Fig.1 Ontology matching process

The introduction of ontology matching

There are many matching methods, which can be divided into two categories: The ontology matching method based on linguistic features [8-9], which is focused on the elements of the specific entities in the ontology, such as the entities' local name, label and comment. The ontology matching method based on structure features [10-11]. This method analyzes the similarity between the structure of the ontology, and seeks the potential matching pairs according to the relationship between entities. This kind of method mainly includes graph matching, entity hierarchy tree matching, and so on. Graph matching method is suitable for the case that the topological structure of the ontology is similar.

In the mapping between SCL and CIM, the papers [12][14] applied the ontology matching technique, and achieved better mapping results. However, both of them are not flexible enough in building the ontology models, and the semantic information of SCL and CIM model is not fully utilized in the matching process, which leads to a low matching accuracy. For that, it hinders the automatic conversion of the model files, and is not conducive to the information sharing between the dispatching center and the substation, so it is necessary to improve the accuracy of the model matching. For this reason, we propose a hybrid ontology matching algorithm.

The SCL and CIM mapping based on hybrid ontology matching

Hybrid ontology matching is based on the serial combination of linguistic features and structural feature matching method. In building ontology models, corresponding conducting equipment subclasses are added to CIM ontology, and corresponding enumeration values of conductive device are

added to SCL ontology. This measure effectively avoid the loss of information in the matching process. Moreover, We improved the method of similarity computation based on structural feature matching, and make full use of the semantic information of SCL model and CIM model to improve the accuracy of model matching. There are four stages in hybrid ontology matching.

(1) Based on linguistic feature matching stage

In this stage, the initial similarity matrix is obtained by using the element level information of ontology. We extract the names, labels, and comments of entities in SCL and CIM ontologies, and compute the similarity value by string edit distance. Edit distance refers to the minimum number of operations that change two strings into the same string by inserting characters, deleting characters and rewriting characters. We use the formula (1) to convert the edit distance of the strings to the similarity between strings.

$$\text{Sim_ed}(s1,s2)=1-\frac{\text{ed}(s1,s2)}{\max(\text{len}(s1),\text{len}(s2))} \quad (1)$$

In the formula, $\text{Sim_ed}(s1,s2)$ is the similarity between the two strings $s1$ and $s2$, $\text{ed}(s1,s2)$ represents the edit distance between $s1$ and $s2$, and $\text{len}(s1)$ and $\text{len}(s2)$ are respectively the length of $s1$ and $s2$.

The end of this phase we get three different similarity matrices for names, labels, and comments, the three matrices respectively are M_n , M_l and M_c . Then, according to the importance of the matrices, endowing them with different values, we get the final initial similarity matrix M_{Init} by Comprehensive calculating the three matrices. The computational method such as the formula (2).

$$M_{\text{Init}} = \gamma * M_n + \theta * M_l + \omega * M_c \quad (2)$$

The γ , θ and ω are weight parameters. In this paper, we take the $\gamma=0.3$, $\theta=0.1$, $\omega=0.6$, the greater of the weight, the higher importance of the matrix. we set the threshold $\text{thre}_1=0.85$, and mark the matched pair (e_i, e_j) as an anchor if there be $M_{\text{Init}(i,j)} > \text{thre}_1$. The next stage anchors will serve as starting point in the similarity value propagation, and the selection of threshold is the results of several trials.

(2) Based on structure feature matching stage

In this stage, we take graph matching method to achieve similarity matrix. The edges of the graph represent the relationship between entities, and different relationships are not treated differently. So we can use one type of edge to represent the relationship subClassOf, kind-of and part-of which are considered in this paper. Given of the direction of the similarity transmission, we choose directed graph as the matching structure. We need to map the OWL ontology to a directed graph before matching. Part of directed graph that are constructed from SCL and CIM ontology classes are shown in Figure 2 and Figure 3.

If the matched pair $(t_{\text{Bay}}, \text{Bay})$ is selected as the anchor in the linguistic feature matching stage, its similarity will propagate to the neighbor nodes, such as $(t_{\text{Terminal}}, \text{Terminal})$, in the matching process. We construct the similarity propagation graph shown in Figure 4. In the graph, W_i represents the similarity propagation factor, namely the coefficient of the anchor spread its similarity value to its neighbor nodes, there are $i=1,2,3,\dots$. (The serial number does not represent the order, but the value is different).

For any matched pair (m,n) of the similarity propagation graph, we calculate its similarity value $S(m,n)$ by formula (3).

$$S^{k+1}(m,n) = S^k(m,n) + \sum_{i \in M} W_i * S_i^k \quad (3)$$

In the formula, $S^{k+1}(m,n)$ represents the similarity values of the matched pair (m,n) in the $(k+1)$ th iteration. Similarly, $S^k(m,n)$ represents the similarity values of the matched pair (m,n) in the k th iteration. M represents the matched pairs that has a neighbor relationship with (m,n) . S_i^k represents the similarity value of the matched pair i which is the neighbor of the (m,n) in the k th iteration. When $k=0$, the similarity values are the vales of similarity matrix M_{Init} , and W_i is the similarity propagation factor.

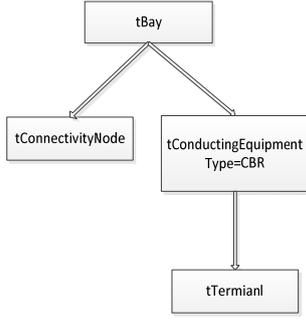


Fig.2 SCL model fragment directed graph

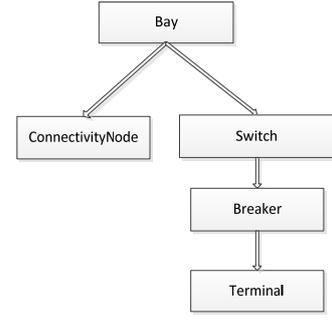


Fig.3 CIM model fragment directed graph

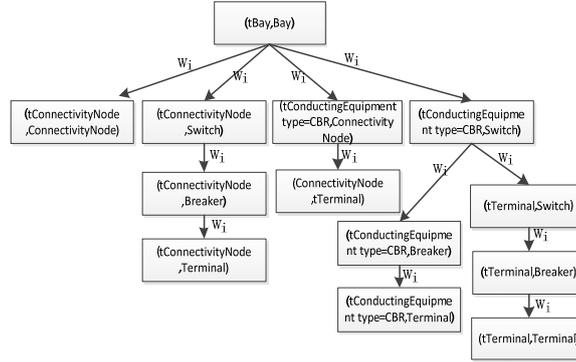


Fig.4 Similarity propagation graph

The similarity propagation factor has different computation methods in different fields, and it is also frequently adjusted in the process of iterative calculation. In this paper, we propose a new method for calculating the propagation factor based on the characteristics of ontology elements. The method makes full use of the semantic information and topological structure of SCL and CIM model. In this paper, the calculation method of W_i is shown in the formula (4).

$$W_i = \frac{\sum_{i \in M} Arc_i}{N * (N + 1)} M_{Init} \quad (4)$$

In the formula, N represents the maximum distance from matched pair (m,n) to anchor, we measure it by the number of edges that when passing through. M represents a set of pairs to be matched that are neighbors of (m,n) . Arc_i represents the number of edges that entities from M distance with m and n respectively. M_{Init} is the initial similarity matrix.

(3) Iterative process

The end of each iteration, select matched pair which its similarity value exceeds the threshold $thre_2$ as a new anchor, adjust the similarity propagation factor, and then normalize the matrix computed in this iteration. Then conduct the next round of match, until the similarity matrix convergence, or can't select a new anchor, we end the matching algorithm. In this paper, we can get better results by taking $thre_2=0.8$. The end of the iteration, we get the final similarity matrix M_{struct} , according which we extract the matched pairs.

(4) Generate mapping table

Comparing the similarity value in M_{struct} with the $thre_2$, we extract the set of matched pairs when $M_{struct(i,j)} > thre_2$. There are wrong matched pairs, we need to select them according to the power system knowledge base [12], and keep the final right pairs. Select the entity SCL as a key, the entity of CIM as a value, constructing HashTable.

Experiment and analysis

Experimental environment and matching cases

Experiments are carried out in the Eclipse development platform using Java language. We use the open source modeling software Protégé to construct SCL and CIM ontologies, and save as a OWL file. And using the application interface Jena that developed by HP laboratory to resolve and

operate ontologies.

In the experiment, we built SCL and CIM ontology model about substation equipment level, which includes the substation, voltage, voltage level, connection point, end point and related conductive equipment. And, there are 75 kinds of CIM entities, 70 SCL entities and attributes are involved in the matching. And there exist 65 correct pairs to be matched.

Experimental result evaluation

After the execution of the hybrid ontology matching algorithm, we find 60 correct mapping pairs. In order to enhance the comparability of the results, we select ontology matching tools AgreementMaker[13] and OM system[14] which are also based on linguistic features and structural features matching as contrast. While we use the accuracy rate of P, recall value R and F-Measure to evaluate matching results.

$$P = \frac{\text{The number of correct mapping pairs that founded}}{\text{The number of all mapping pairs that founded}} \quad R = \frac{\text{The number of correct mapping pairs that founded}}{\text{The number of correct mapping pairs that existed}} \quad F\text{-Measure} = \frac{2 \times P \times R}{P + R}$$

Firstly, we use the data set of 248 ~ 266 provided by the OAEI website for testing, this test case is used as a general comprehensive test case to test the comprehensive performance of a matching tool. P, R, F-Measure values of the three are shown in Table 1. Then, take the model of power domain ontology that constructed in this experiment as test cases. P, R, F-Measure values of the three are shown in Table 2.

Tab.1 Comprehensive test evaluation contrast				Tab.2 Field test evaluation contrast			
	P	R	F-Measure		P	R	F-Measure
AgreementMaker	0.75	0.82	0.78	AgreementMaker	0.67	0.58	0.62
OM system	0.79	0.87	0.83	OM system	0.80	0.74	0.77
The hybrid ontology matching algorithm	0.55	0.60	0.57	The hybrid ontology matching algorithm	0.93	0.86	0.89

As can be seen from table 1, the hybrid ontology matching algorithm, P, R, and F-Measure values are lower than the other two matching tools. While in table 2, the P, R and F-Measure of the hybrid ontology matching algorithm are better than the other two. From which we conclude that the matching algorithm proposed in this paper can greatly improve the accuracy of semantic ontology mapping in power system. Improving the accuracy of the mapping is an important basic work for the automation of SCL and CIM model files.

The process of Model file conversion

After the above matching process, we find most of the mapping relationship between SCL and CIM model, and generate a mapping table of the two models, according which we can achieve the conversion of the substation instance file SCD/XML to the CIM/XML file in the document converter. In the conversion, we use Java-Dom4j to parse out the class, attribute values of the instance files. The last, we verify whether the generated CIM file is in compliance with the IEC61970 standard. The conversion process in the document converter is shown in Figure 5.

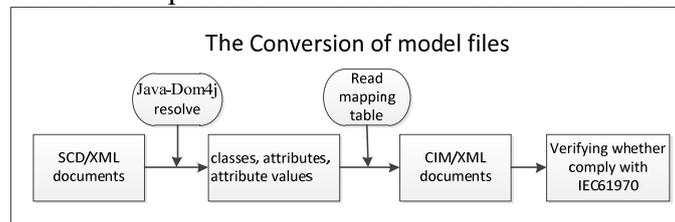


Fig.5 The conversion process of the model files

The verification of Model file conversion

To actually verify the conversion of the model files, we configured a typical substation's wiring structure using the configuration tool SCLManager Kalkitech, it consists of 1 transformers T1, 1 circuit breakers QA1, 2 voltage levels D1 and E1, and 2 bays Q1 and Q2, and some electrical connection points. After the configuration is complete, the SCD/XML file of the substation structure is generated, and its code fragment is as follows:

```
<Terminal connectivityNode="baden220_132/E1/Q2/L3" substationName="baden220_132"
voltageLevelName="E1" bayName="Q2" cNodeName="L3"/></TransformerWinding>
</PowerTransformer><VoltageLevel name="D1"> <Voltage multiplier="k"
unit="V">220</Voltage> <Bay name="Q1">
```

Taking the SCD/XML file as input, after conversion we get CIM/XML file, which is verified in accordance with the IEC61970 standard by verification tool CIM Validate. The generated CIM file code fragment is as follows:

```
<cim:Substation rdf:ID="201301"> <cim:Naming.name>Baden220_132</cim:Naming.name>
<cim:Naming.description>Baden220_132</cim:Naming.description> </cim:Substation>
<cim:VoltageLevel rdf:ID="201302"> <cim:Naming.name>220kV</cim:Naming.name>
<cim:VoltageLevel.MemberOf_Substation rdf:resource="#201302" /> </cim:VoltageLevel>
```

Conclusions

Between the substation and the master station, there are two different description models, which lead to data sharing barriers between them. In this paper, we established the mapping relationships between SCL and CIM models by using the method of ontology matching, and then implemented the model files conversion of a typical substation structure. However, the established mapping relationships does not involve all the corresponding information of SCL and CIM, so it can not realize the conversion of all substation structure model files. There are a lot of work to be done before we put the scheme into practical application.

Acknowledgement

In this paper, the research was sponsored by the Scientific Research Project of Hebei Province (Project No. Z2012077).

References

- [1] Li Xin, He Zhicheng, Xu Zhong. Research on the mapping of IEC61850 and IEC61970 semantic model based on Ontology [J]. low voltage apparatus, 2010,21:57-61.
- [2] Zhu Botong, Cheng Zhihai, Tang Zhiqiang, Jiang Wei. Research on the interoperability of Smart Substation and dispatching center based on CIM model [J]. Power system protection and control, 2013,10:93-97.
- [3] Zhang Haidong, Zhang Hong, Song Xin, Geng Mingzhi, Chen Ailin. The method of transforming SCD model to CIM/E [J]. power system automation, 2012,15:91-95.
- [4] Tan Wenshu. The introduction for Substation communication network and system protocol IEC61850 [J]. power grid technology, 2001,09:8-11+15.
- [5] Zhang Huichen. Research on the method of information integration for the dispatching center and substation based on Ontology [D]. North China Electric Power University, 2013
- [6] Luo Jian, Zhu Botong, Cai Ming, Zhao Guofu. Study on the interoperability of CIM and SCL model based on XML CIM [J]. and SCL model 2011,17:134-138.
- [7] Huang song. Research on the information integration scheme of intelligent substation based on IEC61850 and CIM [D]. Shanghai Jiao Tong University, 2012.
- [8] Zou Lijun, Ling Xing Wang, Yao Wangshu. A matching process combination feature information based ontology [J]. Computer Engineering, 2013,09:271-276.
- [9] Ngo D H, Bellahsene Z, Coletta R. A generic approach for combining linguistic and context profile metrics in ontology matching [M]. On the Move to Meaningful Internet Systems: OTM 2011. Springer Berlin Heidelberg, 2011: 800-807.

- [10] Lu Ying. Research and implementation of ontology matching algorithm based on structure feature [D]. Soochow University, 2012
- [11] Zhang Yue. Research on the efficient adaptive ontology matching algorithm [D]. Soochow University, 2012
- [12] Santodomingo R, Rodriguez-Mondejar J, Sanz-Bobi M A, et al. Towards the automatic alignment of CIM and SCL ontologies[C]. IEEE International Conference on Smart Grid Communications,2011: 422-427.
- [13] Cruz I F, Stroe C, Caimi F, et al. Using AgreementMaker to align ontologies for OAIE 2011[C]. Proceedings of the Sixth International Workshop on Ontology Matching, 2011: 114-121.
- [14] Santodomingo R, Rohjans S, Uslar M, et al. Ontology matching system for future energy smart grids[J]. Engineering Applications of Artificial Intelligence, 2014, 32: 242-257.