http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (23): 7739-7743, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.7739.7743

On Formalization and Reasoning Algorithm in Distributed XBRL System

[1]Ding Pan, [1]Yingmin Zhang and [2]Dong Wang
[1]School of Management, Jinan University, Guangzhou, 510632, China
[2]School of Business Administration, Guangzhou University, Guangzhou, 510006, China

Abstract: Traditional description logic usually used to represent static knowledge. To solve the problem of metadata interoperability in distributed XBRL, the distributed interact model of XBRL was constructed, the distributed characteristics and formal representation requirement of distributed XBRL interact model was analyzed. A distributed temporal description logic $DTDL_{BR}$ and its syntax and semantics was proposed. Finally, the reasoning mapping algorithm was given.

Key words: Distributed XBRL, formalization, reasoning algorithm, distributed temporal description logic,

INTRODUCTION

XBRL (eXtensible Business Reporting Language) is an XML-based markup language designed specifically for financial reporting, disclosure and use. By giving financial accounting data specific category tags, financial reports can be recognized and understood by computer and through the built-in authentication mechanism, reports can be processed and analyzed automatically. XBRL has advantages in cross-platform using, multiple output formats supporting, faster and more accurate searching. It provides a convenience in financial reports' preparation, generation, analyzing, transmission and comparing and enhances the accuracy and reliability of financial data.

Recent work on XBRL is mainly focus on the regulatory of Specification and Taxonomy and the disclosure of XBRL reports, while there is rarely improvement in data intelligent processing and analyzing. The primary cause is XBRL's semantics is achieved by markup language whose readability stuck in the level of natural language and lack of formal representation.

Metadata in XBRL has precise semantics and has apparent distributed, isomerism and temporal characteristics. How to represent the above features is the base of XBRL's intelligent reasoning.

Description Logic (DL) is recommended by W3C for the logical basis of OWL. It is a knowledge representation formal method based on object and is a decidable subset of first-order predicate logic, with strong expressivity and decidability therefore can reason implicit knowledge from explicit knowledge in knowledge base. However, DL is mainly used to represent static knowledge and the corresponding reasoning. XBRL business reports dealing with a large number of tenses data which is cannot be represented by traditional DL.

Scholars have proposed various extended ideas to formalize temporal knowledge. Schmiedel (1990) firstly proposed to combine description logic and interval temporal logic with adding tense operator such as "at", "sometime" and "all-time" to represent temporal knowledge. But this simple combination causes reasoning undecidable. Schild (1993) proposed a temporal description logic ALCT which embedding ALC and point-based tense operators. Lutz (2004) defined the interval-based description logic and gave a general TBox which expand the expressivity and proved reasoning decidable but cannot express the change of tense. Based on interval and Allen's time interval temporal domain, Lutz and Milicic (2007) proposed the algorithm of Tableau under the general concept contain axiom. However, there are few researches on the formalization and especially in the formalization of temporal characteristics of XBRL. Pan *et al.* (2012) proposed a description logic TDL_{BR} which can match the representation of local XBRL tense metadata but not considered the reasoning problem of distributed knowledge. Borgida and Serafini (2003) proposed a distributed description logic (DDL) for the distributed, heterogeneous information to achieve automation of intelligent reasoning.

This study analyzes the distributed and heterogeneous features of XBRL, extends the Pan's TDL_{BR} and Borgida's DDL theory, proposes a distributed temporal description logic ($DTDL_{BR}$) to formalize the terms, attributes and relationships in distributed XBRL reports.

DISTRIBUTED INTERACTIVE MODEL OF XBRL

Introduction of XBRL: XBRL is an open international standardization method for unconstructed information

Corresponding Author: Ding Pan, School of Management, Jinan University, Guangzhou, 560632, China

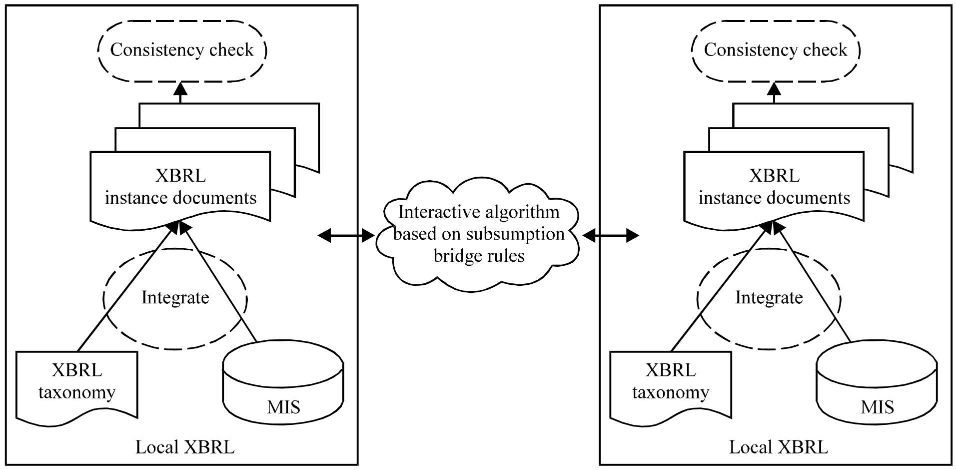

Fig. 1: Distributed XBRL interactive model

process. Since its inception in 1998, XBRL has become one of the most effective forms of internet financial reporting. Major stock exchanges in the world require the listed companies to provide XBRL financial reporting, thus forming a distributed environment of it. XBRL's extensible feature making different countries, regions and industries can extend the basic elements to build suitable taxonomy elements according to their own situation and even in the same company, there are different requirement in metadata representation in different times. This extensible feature is determined by its architecture framework which is consist of Specification, Taxonomy and Instance Document.

Where, Specification is the basic of XBRL architecture framework, mainly used to interpret the structure of XBRL documents, detailing the syntax and semantics of XBRL Taxonomy and Instance Document. Based on the Specification, while the Taxonomy is the region's industrial information "dictionary" that being used to define the attributes and relations of items in the business reports. XBRL Taxonomy is consist of Schema (*.xsd) and Linkbase (*.xml) files. Schema file defines the structure and content schema includes the definition of name, datatype, timetype and related attributes of elements. Linkbase uses XLink to define the relations between elements. Instance Document is the enterprise business reports instances created based on Sepcification and Taxonomy. Taxonomy makes financial reports computer-readable and it is the core and foundation of XBRL financial reports generating and disclosing. Therefore, formalizing the semantics of Taxonomy can realize XBRL reports' identifying, analyzing and reasoning automatically.

Distributed interactive model of XBRL: In order to make distributed and heterogeneous XBRL reports enable to compose and analyze automatically and interact intelligently, we first need to realize the semantic formalization of XBRL Taxonomy and Instance Documents.

This study references the subsumption bridge rule of Borgida's DDL to realize the interact operation of distributed metadata of XBRL, establishes the distributed XBRL semantic metadata interacting model which shown in Fig. 1.

As shown in Fig. 1, based on the specified taxonomy, enterprises extract, integrate data from their own information systems and convert them to local XBRL instance documents and through consistency checks to ensure the correctness of local XBRL data. And then uses the mapping rules to achieve interoperating between metadata in XBRL Taxonomy and instance documents of distributed XBRL.

DESCRIPTION LOGIC

Description Logic is an object-based formal representation method. The basic elements are Concepts and Roles which are used to describe the field of objects and their binary relations. A DL reasoning system contains a Knowledge Box (KBox) and the reasoning mechanism. While KBox includes a terminology set (TBox) and an instance set (ABox), so KBox=<TBox, ABox>. Where TBox is the domain-specific axioms set composed of $C_1 \subseteq C_2$, $R_1 \subseteq R_2$.

Depending on the characteristics of the field of knowledge representation and application requirements, we can extend the description logic to increase the expressivity. TDL_{BR} is the temporal extension of traditional description logic for temporal knowledge representation in XBRL. DDL extends the single description logic to express the distributed knowledge and the related reasoning mechanism.

DISTRIBUTED TEMPORAL DESCRIPTION LOGIC $DTDL_{BR}$

According to the distributed feature of XBRL metadata, this study extends the temporal description logic TDL_{BR}, proposes a concept-identified-supported distributed description logic $DTDL_{BR}$. $DTDL_{BR}$ provides a good representative to the XBRL abstract model and its structuring metadata. It is equipped with decidable reasoning procedure so that can provides rigorous formal representation and reasoning framework to distributed XBRL.

Syntax of $DTDL_{BR}$: $DTDL_{BR}$ reasoning system consists of three components: a plurality of local TDL_{BRi} (where, i = 1, 2, ……, n, is the serial number of a single temporal description logic); a distributed knowledge (KBox) which includes distributed TBox and ABox; a reasoning mechanism of distributed TBox and ABox.

Here we introduce the symbols in TDL_{BR} (Pan *et al*., 2012), in which instance, concept and role express the same as classical description logic. Specific symbols as: C, D is the general concept (and T is the top concept, $\perp$ is the bottom concept, A is the atom concept), R for a role that expresses a binary relation (P is an atom relation), T is for a temporal concept.

For the distributed XBRL interoperating model, in order to integrate two or more different XBRL reports we need to resolve the mapping problem of different XBRL. As described, various relations in description logic can be transformed into subsumption relation; therefore, here imports the subsumption rule as the mapping rule for distributed XBRL. Defined as follows:

Definition 1: (Concept Subsumption rules) Suppose C, D are the concept of TDL_{BRi}, TDL_{BRj} respectively, the concept subsumption rules from i to j (i≠j) defined as follow:

- $i:C \xrightarrow{\subseteq} j:D$, former concept subsumption into-bridge rule CIBR which states that concept C of the i-th TDL_{BLi} is the subset of D of the j-th TDL_{BRj}
- $i:C \xrightarrow{\supseteq} j:D$, consequent concept subsumption onto-bridge rule COBR which states that concept D of the j-th TDL_{BLj} is the subset of C of the i-th TDL_{BRi}

Definition 2: (Role Subsumption rules) Suppose R, S are the role of TDL_{BRi}, TDL_{BRj} respectively, the role subsumption rules from i to j defined as follow:

- $i:R \xrightarrow{\subseteq} j:S$, former role subsumption into-bridge rule RIBR which states that role R of the i-th TDL_{BLi} is the subset of S of the j-th TDL_{BRj}
- $i:R \xrightarrow{\supseteq} j:S$, consequent role subsumption onto-bridge rule ROBR which states that rule S of the j-th TDL_{BLj} is the subset of R of the i-th TDL_{BRi}

where subsumption bridge rules has directionality denotes that bridge rules from j to i are not necessarily the inverse of the rules from i to j and in fact there may be no rules in one or both the directions (Jiang et al., 2006; Krieger, 2008).

Definition 3: (Instance Subsumption rules) Suppose a is an instance set of TDL_{BRi}, b, b_1, b_2, ... are instances of TDL_{BRj}, i to j (i≠j) instance subsumption rules include the following two definitions:

- i:a→j:b, where, $b \in \{b_1, b_2, \ldots\}$, is part instance subsumption rules PIR, it denotes that b is one of the instance that related to a
- $i:a \xrightarrow{=} j:\{b_1, b_2, \cdots\}$, is complete instance subsumption rules CIR, that instance a of TDL_{BRi} is complete equal to instance set $\{b_1, b_2, \ldots\}$ of TDL_{BRj}

Definition 4: The knowledge base of distributed temporal description logic DKBox is compose of distributed terminology set DTBox and distributed instance set DABox, that is DKBox = ⟨DTBox, DABox⟩.

Definition 5: DKBox = $\langle \{Tbox_i\}_{i\in I}, BR\rangle$, where, I is the serial set, $\{Tbox_i\}_{i\in I}$ is TBox set of TDL_{BR}, BR = $\{BR_{ij}\}$ is the subsumption bridge rules set. For any k∈I, descriptions in $TBox_k$ must comply the specification of TDL_{BRk} and for any subsumption bridge rule CIBR, COBR, RIBR, ROBR of BR_{ij} and the related concept C, D, related role R, S must comply the specification of TDL_{BRi} and TDL_{BRj}, respectively.

Definition 6: DABox=$\langle \{ABox_i\}_{i\in I}, IR\rangle$, where, I is the serial set, $\{ABox_i\}_{i\in I}$ is ABox set of TDL_{BR}. IR = (IR_{ij}) is instance subsumption bridge rules set from i to j. For any k∈I, descriptions in $ABox_k$ must comply the specification of TDL_{BRk} and for any instance subsumption bridge rule PIR, CIR of IR_{ij} and the related instance a and b,b_1,b_2, ... must comply the specification of TDL_{BRi} and TDL_{BRj} respectively.

Semantics interpretation of $DTDL_{BR}$: $DTDL_{BR}$ is integrated by a number of TDL_{BR} through subsumption bridge rules, so its semantics interpretation can be brought up from TDL_{BR} and the domain association.

I_i denoted as the i-th interpretation of TDL_{BR}, it is compose of the non null interpretation set Δ^{Ii} and interpretation function $\bullet^{Ii}$ ($I_i=(\Delta^{Ii}, \bullet^{Ii})$), mapping with $\bullet^{Ii}$ that concept i:C of TDL_{BRi} is the subset of C^{Ii}, binary relation i:R is R^{Ti} which is the subset of $\Delta^{Ii}\times\Delta^{Ii}$. Detailed semantic interpretation of TDL_{BR} shown in article (Schmiedel, 1990).

Definition 7: In order to illustrate the interpretation relation between TDL_{BRi} and TDL_{BRj} ($i\neq j$), we import the binary relation $rij\subseteq\Delta^{Ii}\times\Delta^{Ij}$,$i\neq j\in I$, if $r_{ij}(C^{Ii})$ to designate the relation between interpretation of concept i:C and the concept of TDL_{BRj}, then the distributed interpretation of $DTDL_{BR}$ defined as $DI=\langle\{I_i\}_{i\in I}, \{r_{ij}\}_{I,j\in I}\rangle$.

Definition 8: the satisfiability of a distributed interpretation $DI=\langle\{I_i\}_{i\in I}, \{Ir_{ij}\}_{I,j\in I}\rangle$ denoted as $DI\vDash_d$, it means that we can find an interpretation in the Knowledge Box, whatever in the distributed TBox or ABox and has the following forms:

- **Satisfiability of concept:** $DI\vDash_d i:C \xrightarrow{\Sigma} j:D$, if the interpretation $r_{i,j}(C^{Ii})\Sigma D_{Ij}$ (where, Σ = {$\subseteq$, $\supseteq$}, C , $D^{Ij}\neq\varnothing$) holds, then DI satisfies the concept of DTBox that concept i:C to j:D is satisfiable
- Satisfiablity of subsumption relation: if concept C is included in D, then for DI satisfies $C^{Ii}\subseteq D^{Ii}$, denoted as $DI\vDash_d i:C\sqsubseteq D$
- **Satisfiabilty of terminology:** if a distributed interpretation $DI\vDash_d$ satisfies all the assertions of each TBox (denoted as $DI\vDash_d TBox_i$) and meets all bridge-rules interpretation (denoted as $DI\vDash_d BR_{ij}, BR_{ij}\in BR$), called that the terminology set satisfiable
- **Satisfiability of instance's relations:** if the interpretation of relation $b^{Ij}\in r_{i,j}(a^{Ii})$ satisfied then call instance i:a partly contain j:b satisfiable, denoted as $DI\vDash_d i:a \longrightarrow j:b$; if :

$$r_{i,j}(a^{I_i})=\{b_1^{I_j},b_2^{I_j},\cdots\}$$

established, then instance i:a completely contain j:b = (b_1, b_2, ...) is satisfiable, denoted as $DI\vDash_d i:a \xrightarrow{=} j:\{b_1,b_2,\cdots\}$
- **Satisfiability of DABox:** if distributed interpretation $DI\vDash_d$ all assertions of each ABox, that $DI\vDash_d i:C(a)$, $DI\vDash_d i:P(a,b)$, we call that DI is satisfiable to $ABox_i$, denoted as $DI\vDash_d ABox_i$ if DI meets interpretation of all instance in, denoted as $DI\vDash_d in, in\in IN$, that DI is satisfiable to distributed instance set DABox, denoted as $DI\vDash_d DABox$
- **Satisfiability of DKBox:** if $DI\vDash_d$ satisfy all assertions and instances of DKBox, call that $DI\vDash_d$ is a model of DKBox, denoted as $DI\vDash_d DKBox$

REASONING ALGORITHM OF $DTDL_{BR}$

Whether a concept C_0 in ALCN is satisfiable is to find the interpretation $I = (\Delta^I, \bullet^I)$ that make $C_0{}^I$ satisfied. If can find it that under the interpretation I, then concept C_0 is satisfiable, otherwise it is unsatisfiable. In fact, the interpretation is uncountable that it is impossible to traverse all interpretations to judge whether it is satisfiable. In this context, Tableau algorithm being put forward, its basic idea is constructing a model to prove C_0 is satisfiable.

Reasoning in $DTDL_{BR}$ includes local and global parts. Representation and reasoning of version migrating in taxonomy belongs to local XBRL consistency which goal is to find a local interpretation I_i meets $I_i\vDash C\sqsubseteq D$, because in the ALC language, reasoning problems can be transformed into satisfiability judgment.

This study concerns the interoperability after local consistency checking, namely the global inference. In the distributed XBRL interactive model, local XBRLs interoperate based around the subsumption rules, therefore, the global reasoning in $DTDL_{BR}$ is through subsumption rules to find a distributed interpretation DI satisfies $DI\vDash_d i:C\sqsubseteq D$,denoted as $(C\sqcap\neg D)^{I_i}=\varnothing$.

Reasoning algorithm of $DTDL_{BR}$: According to distributed reasoning feature this study extends traditional Tableau algorithm, establishes the interactive Tableau algorithm for distributed XBRL. Main idea is: firstly, constructing a local Tableau tree, secondly, checking whether can end the local Tableau tree through subsumption mapping rules. If cannot end it, then the concepts is satisfiable, otherwise it is unsatisfiable (Serafini and Tamilin, 2005). Specific steps designed as follows:

Start input: $L_0 = \{C_0(x_0)\}$;

Root note: x_0;

Output: whether $TBox_i\vDash i:C\sqsubseteq i:D$ can be established

Distributed Tableau tree building process:

- Build a local complete tree through general Tableau algorithm, if there were confliction then concept in local KBox is unsatisfiable and end algorithm, otherwise go to step (2)
- For any $j\in I$, $j\neq i$ in TDL_{BRj} and TDL_{BRi}, judge whether the following statements can be met: If exists the COBR rules from j to i, the concept C in TDL_{BRi} being contained to G in TDL_{BRj}; IF exists the CIBR rules from j to i, the concept H in TDL_{BRj} being contained to D in TDL_{BRi}. Go to step (3)

- Judge whether TDL_{BRj} meet j:G⊆j:H (that is $TBox_j \models j:G \subseteq j:H$). If true, then go to the end and the concepts are satisfiable, otherwise go to step (4)
- Judge whether there exist k∈I, k≠j≠i that meets the statements in step (2), if exists then go to step (2), else the concepts are unsatisfiable and end the algorithm

Nature of $DTDL_{BR}$'s reasoning algorithm: $TDLR_{BR}$ is an subset of ALCN, the Tableau algorithm has been proved that reasoning of ALCN-concepts is reliability, terminability and completeness. $DTDL_{BR}$ is a collection of multiple TDL_{BRi}, so it has reliability, terminability and completeness either.

De Giacomo (1995) proved the inference algorithm of ALCNis exponential time complete. $DTDL_{BR}$ did not change the complexity of algorithm, so its complexity is exponential time complete either.

CONCLUSIONS

XBRL architecture framework determines that metadata in XBRL has typical distributed, heterogeneous and temporal features. This study extends the TDL_{BR}, proposes a distributed and temporal-extended description logic $DTDL_{BR}$, uses it to represent XBRL's distributed metadata, constructs the corresponding reasoning algorithm and proves its decidability.

ACKNOWLEDGMENT

This study was supported by the National Nature Science Foundation of China (No. 71171097, 70872020, 70771044).

REFERENCES

Borgida, A. and L. Serafini, 2003. Distributed Description Logics: Assimilating Information from Peer Sources. In: Journal on Data Semantics, Spaccapietra, S., S. March and K. Aberer (Eds.). Springer, Berlin, Heidelberg, pp: 153-184.

De Giacomo, G., 1995. Decidablity of class-based knowledge representation formalisms. Ph.D. Thesis, Dipartimento di Informatica e Sistemistica, Universita di Roma, La Sapienza.

Jiang, Y., Z. Shi, Y. Tang and J. Wang, 2006. A distributed dynamic description logic. J. Comput. Res. Dev., 43: 1603-1608.

Krieger, H.U., 2008. Where Temporal Description Logics Fail: Representing Temporally-Changing Relationships. In: Advances in Artificial Intelligence, Dengel, A.R., K. Berns, T.M. Breuel, F. Bomarius and T.R. Roth-Berghofer (Eds.). Springer, Berlin, Heidelberg, pp: 249-257.

Lutz, C. and M. Milicic, 2007. A tableau algorithm for DLs with concrete domains and GCIs. J. Autom. Reasoning, 38: 227-259.

Lutz, C., 2004. Combining interval-based temporal reasoning with general tboxes. Artif. Intell., 152: 235-274.

Pan, D., Y.M. Zhang and D. Wang, 2012. On formalization of metadata in XBRL based on a temporal description logic TDL_{BR}. J. Inform. Comput. Sci., 9: 4477-4484.

Schild, K., 1993. Combining Terminological Logics with Tense Logic. In: Progress in Artificial Intelligence, Filgueiras, M. and L. Damas (Eds.). Springer, Berlin, Heidelberg, pp: 105-120.

Schmiedel, A., 1990. Temporal terminological logic. Proceedings of the 8th National Conference on Artificial Intelligence, July 29-August 3, 1990, Boston, MA., USA., pp: 640-645.

Serafini, L. and A. Tamilin, 2005. DRAGO: Distributed Reasoning Architecture for the Semantic Web. In: The Semantic Web: Research and Applications, Gomez-Perez, A. and J. Euzenat (Eds.). Springer, Berlin, Heidelberg, pp: 361-376.