Novel Topic Maps to RDF/RDF Schema Translation Method

Shinae SHIN^{†a)}, Nonmember, Dongwon JEONG^{††}, Member, and Doo-Kwon BAIK[†], Nonmember

SUMMARY We propose an enhanced method for translating Topic Maps to RDF/RDF Schema, to realize the Semantic Web. A critical issue for the Semantic Web is to efficiently and precisely describe Web information resources, i.e., Web metadata. Two representative standards, Topic Maps and RDF have been used for Web metadata. RDF-based standardization and implementation of the Semantic Web have been actively performed. Since the Semantic Web must accept and understand all Web information resources that are represented with the other methods, Topic Maps-to-RDF translation has become an issue. Even though many Topic Maps to RDF translation methods have been devised, they still have several problems (e.g. semantic loss, complex expression, etc.). Our translation method provides an improved solution to these problems. This method shows lower semantic loss than the previous methods due to extract both explicit semantics and implicit semantics. Compared to the previous methods, our method reduces the encoding complexity of resulting RDF. In addition, in terms of reversibility, the proposed method regenerates all Topic Maps constructs in an original source when is reverse translated.

key words: Topic Maps, RDF, Semantic Web, Web ontology, Web metadata

1. Introduction

1.1 Motivation

The Semantic Web is an intelligent Web implemented with machine-understandable metadata, to define and deliver clear representations of information on the Web. One of the most important issues for the Semantic Web is how to efficiently and precisely build Web information. Currently, Topic Maps [1] and RDF [2], [3] are recognized as representative methods to describe and build Web information. Topic Maps and RDF (Resource Description Framework) have been developed by ISO/IEC JTC 1 and W3C (World Wide Web Consortium), respectively.

Topic Maps defines how to describe and handle information resources to navigate on the Web. Topic Maps captures subjects of information resources, and defines relationships between the subjects. The core elements of Topic Maps are **topics**, **associations**, **association roles** and **occurrences**, and have each type. Roles of these elements are defined in [1] as follows:

- **topic:** a symbol used with a topic map to represent one subject
- **association:** representation of a relationship between one or more subjects
- **association role:** representation of the involvement of a subject in a relationship represented by an association
- **occurrence:** representation of a relationship between a subject and an information resource

RDF is a foundation for processing metadata and provides interoperability between applications that exchange machine-understandable information on the Web [2]. RDF is a data model for objects (resources) and relationships between objects. The basic RDF data model consists of three object types: **resource**, **property**, and **statement**. The RDF statement is composed of three individual parts; **subject**, **predicate**, and **object**. These elements are also defined in [2] as follows:

- **subject:** defines the Web resource for which the information is stored
- **predicate:** defines a characteristic, or property to gather information about on the subject
- **object:** contains the property value for the predicate

RDF Schema is an RDF's vocabulary description language for a semantic extension of RDF. It provides primitives for describing groups of related resources and relationships between these resources [3].

Topic Maps and RDF have many similarities such as an ability to describe and exchange information on the Web, a semantic model base and an adoption of XML encoding. However, they have different schemes and syntaxes.

The semantic web consists of a collection of technologies such as XML, RDF, and ontology, which facilitate interoperability. This is essentially about machine processable web pages [4]. Among these semantic web related technology, RDF plays the role of a common model, as a kind of "glue" which integrates data [5]. There have been many efforts to develop tools that can "export" a web of data into RDF form. The tools include automatic procedures that can produce RDF data on-the-fly in answer to queries [5]. And many translation methods producing RDF from Topic Maps have been also devised. Although the presence of various methods, several problems still remain, such as high complexity, additional dummy nodes and loss of semantics.

We propose an enhanced method of translating Topic

Manuscript received April 22, 2008.

Manuscript revised July 4, 2008.

[†]The authors are with the Department of Computer Science and Engineering, Korea University, Anam-dong, Seongbuk-Gu, Seoul, 136-701 Korea.

^{††}The author is with the Department of Informatics and Statistics, Kunsan National University, Miryong-dong, Gunsan, Jeollabuk-do, 573-701 Korea.

a) E-mail: sashin@nia.or.kr

DOI: 10.1093/ietisy/e91-d.11.2626

Maps into RDF, which essentially covers all of the main Topic Maps constructs, whereas some existing methods do not translated occurrences or association roles. By using *association roles*, our method even can translates implicit semantics in Topic Maps to RDF/RDFS. In addition, we define an evaluation methodology and demonstrate evaluation results from experiments.

1.2 Related Research

Many researchers have been studying on a translating Topic Maps to RDF. The most representative methods are [6]–[10] and [11]. However these methods are somewhat incomplete. The Garshol’s method [7] use new vocabularies called RTM (Rdf-Topic map Mapping). The translated results by this method are clear, but miss *association roles*. Also the Garshol’s method deals with *associations* as one-way relation although *associations* are non-directional. The Stanford’s method [8] translates Topic Maps in the context of *associations*, resulting in a semantic structure which differs from the original. In addition, this method also misses occurrences during translation. The Moore’s method [9] generates additional nodes, {Assoc:1, Assocmember:1, Assocmember:2} that do not exist in the original source model. They make the structure more complex and incomprehensive. This method does not treat occurrences. For XTM2RDF converter by Ogievetsky [10], a new vocabulary is required to translate XML TM to RDF. The Unibo’s method [11] does not cover some features (e.g. reification of association roles).

The comparative study by W3C [6] illustrates interoperability issues between Topic Maps and RDF, and evaluates the existing methods ([7]–[10] and [11]) with the two criteria: ‘*Completeness*’ and ‘*Naturalness*’. The shortcomings analyzed by W3C can be summarized as follows: the Garshol’s method [7] does not utilize “the standard RDF and

RDF Schema predicates,” thus it always requires mapping to be specified; The translation procedure of Moore [9] produces many extra RDF statements, reducing ‘*Naturalness*’; The Stanford’s approach [8] is based on PMTM4 (Processing Model for Topic Maps) which is not a complete model for Topic Maps; The Ogievetsky’s method [10] requires many RDF statements to represent information content that would be more naturally modeled with only a single RDF statement, thus decreasing the ‘*Naturalness*’; The Unibo’s generic translation [11] is far less satisfactory, since this adopts additional *baggages*, resulting in nine RDF statements for a single binary association.

2. Translation Method

This section describes the overall algorithm and key rules for the proposed method. The translation algorithm consists of three components as depicted in Fig. 1: input (Topic Maps), translation and output (RDF translated). The translation function is composed of three steps: (1) selecting subjects and objects, (2) generating predicates, and (3) refining to achieve complete RDF statements. In this Section, for better understanding the translation steps, we use Test Data 1 in Fig. 5 as an example to present how the each step really works.

2.1 Selecting subjects and objects

The first step of the translation algorithm is to identify subjects and objects and then to create pairs of a subject and an object. In this step, subjects or objects are selected from topics, association roles and occurrence locators.

The type of a (subject, object) pair falls into one of the followings: a pair of topics, a pair of association roles and a pair of topic-occurrence locator. De-

Fig. 1 The overall structure of the proposed translation algorithm.

Step 1. (subject, object) pairs selection : This rule extracts topics, association, association roles and occurrence locators from a topic map, then translates these into subjects or objects of RDF.

```

1: input Topic Maps;
2: output ( $S, O$ ); /*  $S$ : subjects set,  $O$ : objects set */
3: extract topics ( $t$ ), topic types ( $tt$ ), associations ( $a$ ), association roles ( $r$ ), occurrence locators ( $o$ );
4: create topic pairs ( $t_i, t_j$ ), association role pairs ( $r_i, r_j$ ) and topic-occurrence pairs ( $t_i, ol_i$ );
5: if ( $is(t_i, t_j)$ ) then
6:   translate ( $t_i, t_j$ ) to ( $s_i, o_j$ ) and ( $t_j, t_i$ ) to ( $s_j, o_i$ ); /*  $s$ : subject,  $o$ : object */
7:   translate  $tt_i$  to  $rdf:type$  of  $t_i$  and  $tt_j$  to  $rdf:type$  of  $t_j$ ; /*  $tt_i$ : type of  $t_i$ ,  $tt_j$ : type of  $t_j$  */
8:   translate  $r_i$  to  $rdfs:domain$  of  $t_i$  and  $r_j$  to  $rdfs:domain$  of  $t_j$ ;
9:   translate  $topicName$  of each constructs to  $rdfs:label$  of those;
10: else if ( $is(r_i, r_j)$ ) then
11:   translate ( $r_i, r_j$ ) to ( $s_i, o_j$ ) and ( $r_j, r_i$ ) to ( $s_j, o_i$ );
12:   translate  $rt_i$  to  $rdf:type$  of  $r_i$  and  $rt_j$  to  $rdf:type$  of  $r_j$ ; /*  $rt_i$ : type of  $r_i$ ,  $rt_j$ : type of  $r_j$  */
13:   translate  $topicName$  of each constructs to  $rdfs:label$  of those;
14: else if ( $is(t_i, ol_i)$ ) then
15:   translate ( $t_i, ol_i$ ) to ( $s_i, o_i$ ); /*  $ol$ : occurrence locator */
16:   translate  $tt_i$  to  $rdf:type$  of  $t_i$ ;
17:   translate  $topicName$  of each constructs to  $rdfs:label$  of those;

```

Fig. 2 Algorithm for (subject, object) pairs selection.

pending on the type of a (subject, object) pair (Line 5, 10 and 13 in Fig. 2), different conversion rules should be applied.

- topic based generation (Line 5 to 9)

RDF and Topic Maps are identity-based technologies. A key concept is to use symbols to represent identifiable “things.” In Topic Maps, the term for “a thing” is a subject, and the term for the construct representing a subject is a topic. In RDF the term for “thing” is a resource. The “resource” is treated by RDF as synonymous with “entity,” i.e. as a generic term for anything in the universe of discourse [12]. Therefore, topics of Topic Maps can be translated into RDF resources. Each of the subject, object and predicate in RDF can be a resource [2]. Hence, a topic in Topic Maps can be translated into a subject, an object or a predicate. Topic Maps *per se* does not have directionality between topics, but RDF has. Thus, we assumed a bidirectional topic relationship, which allows topic to become subject or object during translation. This aspect is reflected in Line 6 (in Fig. 2).

When a topic is translated into a subject, a topic type and an association role can be translated into an $rdf:type$ and an $rdfs:domain$, respectively. A topic type is some commonality in a set of subjects, and an association role represents the nature of the participation of a topic in an association. While, an $rdf:type$ is used to state that a resource is an instance of a class. An $rdfs:domain$ means a domain of a property which limits the individuals to which the property can be applied [13]. The other attributes, a $topicName$ and a $scope$ are translated to an $rdfs:label$ and an $rdfs:range$, respectively. A $topicName$ is a name of topic and an $rdfs:label$ is a human-readable version of a resource’s name. A $scope$ is a context within which a statement is valid, an $rdfs:range$ on the other hand is the range of a property which limits the individuals that the property may have as its value.

(Example of topic based generation)

When the topic pair creation algorithm (Line 4) is applied to Test Data 1, we can obtain a pair, (puccini, toscas). Then, as shown below, if the pair is a topic pair (Line 5), the pair turn into RDF subject(s) and object(o) and gain directionality by Line 6. In the example, we omit the $topicNames$ translation results, because of the space limitation.

- (s, o) $\leftarrow$ (puccini, toscas), (toscas, puccini) (by Line 6)
- $rdf:type$ of puccini $\leftarrow$ music:person (by Line 7)
- $rdf:type$ of toscas $\leftarrow$ music:opera (by Line 7)
- $rdfs:domain$ of puccini $\leftarrow$ music:composer (by Line 8)
- $rdfs:domain$ of toscas $\leftarrow$ music:work (by Line 8)

- association role based generation (Line 10 to 13)

A topic represents one subject and a subject can be anything whatsoever, regardless of whether it exists or has any other specific characteristics [1]. Therefore, the proposed method handles an association role as a topic candidate and translates these to subject and object. Akin to the topic based generation method, directionality is achieved and types of association roles are converted into $rdf:type$ of subject or object.

(Example of association role based generation)

A (composer, work) pair is extracted from Test Data 1 by Line 3 as an association role pair. Then, the pair obtains directionality by Line 11 and each of the pair becomes RDF subject or object.

- (s, o) $\leftarrow$ (composer, work), (work, composer) (by Line 11)
- $rdf:type$ of composer $\leftarrow$ music:person (by Line 12)
- $rdf:type$ of work $\leftarrow$ music:opera (by Line 12)

- occurrence based generation (Line 14 to 17)

A topic and an occurrence locator can be translated

Step 2. predicates selection : The association translated into a predicate, *p*, between association roles, *r*, in Topic Maps. The association role, *r*, is also translated into a predicate, *p*, according to the expressive direction.

```

1: input (S, O);
2: output (S, P, O);           /* S: subjects set, P: predicates set, O: objects set */
4: if ((si, oj) == (ti, tj)) then
5:   translate a.rj to p;         /* a.rj: composition of association and association role of tj, p: predicate */
6: else if ((si, oj) == (tj, ti)) then
7:   translate a.ri to p;         /* a.ri: composition of association and association role of ti */
8: else if ((si, oj) == (ri, rj)) then
9:   translate a.tti to p;         /* a.tti: composition of association and type of tj */
10: else if ((si, oj) == (rj, ri)) then
11:   translate a.ttj to p;         /* a.ttj: composition of association and type of ti */
12: else if ((si, oj) == (ti, oli)) then
13:   translate occ to p;          /* occ: occurrence of oli */
14: end

```

Fig. 3 Algorithm for predicates selection.

to an (subject, object) pair. It represents of a relationship between a subject and an information resource. An occurrence is a specialized type of an association in which one participant is an information resource [1]. Thus, our method is devised to translate even a relationship between a subject (topic) and an information resource into an RDF statement. At this time, relationship between a topic and an occurrence locator has an obvious direction. Therefore, a topic and an occurrence locator are translated into a subject and an object, respectively (by Line 14).

(Example of occurrence based generation)

With Test Data 1, Line 4 in Fig.2 produces two topic-occurrence locator pairs, (tosca, 1900-01-14) and (tosca, <http://www.azopera.com/...>), which become (subject, object) pair, respectively, by Line 15.

- (*s*, *o*) ← (tosca, 1900-01-14), (tosca, <http://www.azopera.com/...>)
- **rdf:type** of tosca ← music:opera (by Line 16)

2.2 Generating predicates

The next step is to select predicates. As associations, association roles and occurrence locators in Topic Maps describe the relationships between association roles, topics, and topics and occurrences, respectively, they are translated into predicates in RDF. When the predicates are extracted between topics, a role whose topic corresponds to an object is selectively translated into a predicate, since the relation between the subject and the object is influenced by the association role of the object in the association. Also, when occurrences become predicates, we add an “occ_” as a prefix to distinguish a relationship between a topic and an occurrence from a relationship between topics. Thus the predicates, which have the prefix, “occ_” can be occurrences when these are reverse translated. This step is presented in Fig. 3. (Example of generating predicates)

The (subject, object) pairs, generated in Sect. 2.1 obtain each of those predicates by Fig. 3 as follows;

- For topic pairs:

- *p* of (puccini, tosca) ← composed-by.work (by Line 4, 5)
- *p* of (tosca, puccini) ← composed-by.composer (by Line 6, 7)

- For association role pairs:

- *p* of (composer, work) ← composed-by.opera (by Line 8, 9)
- *p* of (work, composer) ← composed-by.person (by Line 10, 11)

- For topic-occurrence locator pairs:

- *p* of (tosca, 1900-01-14) ← occ_premiere-date (by Line 12, 13)
- *p* of (tosca, <http://www.azopera.com/...>) ← occ_synopsis

2.3 Refining RDF Statements

To achieve a complete translation, there is an important property to be considered: subjects and predicates of RDF should be addressable resources with a URI [2]. However, a topic, an association and an association role of Topic Maps express a “thing” which can be a non-addressable concept expressed by a literal (string) as well as an addressable resource with a URI [14]. Therefore, only an addressable resource among data elements of Topic Maps should be translated into a subject of RDF. To solve this problem, we attached a refinement operation (Step 3, Fig. 4).

3. Experimental Setup

This section describes the criteria of comparison, samples, and models for evaluation.

Step 3. Refining RDF statements: An exception rule prevents a non-addressable concept from being translated into an RDF subject.

```

1: input ( $S, P, O$ );
2: output  $\text{RDF}(S, P, O)$ ;
4: if ( $s_i, L$ ) then          /*  $L$  : a set of literals */
5:   remove ( $s_i, p_i, o_i$ );
6: end

```

Fig. 4 Algorithm for exception processing.

3.1 Comparative Items

We conduct experiments to compare the existing translation methods with our proposed method, by using three criteria of comparison: (1) Semantic loss rate (loss rate of meanings); (2) Reversibility; (3) Encoding complexity of translated result. The first two items correspond to the W3C's evaluation items, *completeness* and the third item is for W3C's *naturalness* [6].

W3C [6] defines *completeness* to evaluate the extent to which a certain method can handles every semantic constructs in the source model and provides means to represent it without loss of semantics in the target model. In order to measure *completeness*, we choose the semantic loss rate and the reversibility as comparative items. The former means the loss rate of original semantics which include both explicit and implicit semantics. The latter is to measure differences between the original and a data which has been translated twice to return to a starting data [11]. W3C [6] defines also another item, *naturalness* as the degree to which the results of a translation naturally reflect the information content in that paradigm to improve readability on the result. For the *naturalness*, W3C compares the number of RDF statements in a translated result. We also use the same way.

3.2 Evaluation Models

This section shows evaluation models as the criteria for comparison.

3.2.1 Evaluation Model for the Semantic Loss Comparison

A semantics loss rate of each method (m) is used to evaluate the completeness of translation. When the number of original semantics is $originS$ and the extractable semantic number from the translated result is $translatedS(m)$, the number of lost semantics $n(Loss)$ is $(\sum originS - \sum translatedS(m))$. At this time, $originS$ is a semantic numbers of source topic map, $n(TM)$ and $translatedS(m)$ is a semantic numbers extracted from a translated RDF, $n(RDF)$. The semantic loss rate $LossR$ is defined as the calculation formula (1).

$$\begin{aligned}
 LossR(m) &= \frac{\sum originS - \sum translatedS(m)}{\sum originS} \\
 &= \frac{n(TM) - n(RDF)}{n(TM)} = \frac{n(Loss)}{n(TM)} \quad (1)
 \end{aligned}$$

For example, with Test Data 1, $\sum originS$ in Fig. 5 by method in Sect. 4.1.1 is 6. While $\sum translatedS$, the number of semantics in translated RDF by the Stanford's method, is 2 such as Fig. 6. Thus, for the Stanford's method, $LossR(x)$ is 0.67 $((6 - 2)/6)$. It means that the Stanford's method lost 67% of original semantics.

3.2.2 Evaluation Model for the Reversibility Comparison

The evaluation item, reversibility is suggested by [11]. But an explicit formula was not shown to calculate it [11]. Thus we make a formula (2). This calculates the number of same topic map constructs in the round tripped result with the original source topic map constructs. Since some existing methods miss subject topics, associations, occurrences, or association roles when reverse translated, the reversibility is calculated by counting these constructs. An $OriginTMconst$ is a sum of the numbers of topics, associations, occurrences, and association roles in the original source and a $RevTransTMconst$ is a sum of the numbers of these elements in the reverse translated, which are identical to those in the original. A value of the reversibility is more nearer to 1 (100%), the reversibility get better.

$$\begin{aligned}
 Rev(m) &= \frac{\sum RevTransTMconst(m)}{\sum OriginTMconst} \\
 &= \frac{\sum RevTransTM(\# ofTs, \# ofAs, \# ofOs, \# ofRs)}{\sum OriginTM(\# ofTs, \# ofAs, \# ofOs, \# ofRs)} \quad (2)
 \end{aligned}$$

Ts : topics, As : associations,

Os : occurrences, Rs : roles

For example, Test Data 1 (Fig. 5) has two topics, one association, two association roles and two occurrences, in all 7 ($\sum OriginTMconst = 7$). If once RDF has been produced from Test Data 1 by the Stanford's method and then the RDF is reverse-translated into Topic Maps, two occurrence will be lost ($\sum RevTransTMconst = 5$) (Table 3) and $Rev(m)$ is 0.71 ($5/7$).

3.2.3 Evaluation Model for the Encoding Complexity Comparison

An encoding complexity can be evaluated with the number of resulting RDF statements. The formula for the encoding complexity, $Complex(pm, em)$, is defined as (3). The pm and em indicate the proposed method and the existing methods,

```
[puccini : person = "Giacomo Puccini"]
[tosca : opera = "Tosca"]

(tosca, premiere-date, [[1900-01-14]])
(tosca, synopsis, "http://www.azopera.com/learn/synopsis/tosca.shtml")
composed-by( tosca : work, puccini : composer )

a. Test Data 1: The basic topic map for Tosca

[Garcia_Lorca : person = "Garcia Lorca"]
[Granada : city = "Granada"]
[Spain : country = "Spain"]
[La_Alhambra : palace = "Alhambra"]
[First_Deep_Song_Festival : music_event = "First deep song festival"]
[Manuel_de_Falla : person = "Manuel de Falla"]
[La_Vida_breve : music = "La vida breve"]

born-in(Garcia_Lorca : poet, Granada : city)
located-in(Granada : city, Spain : country)
located-in(La_Alhambra : place, Granada : city)
held-in(First_Deep_Song_Festival : music_event, La_Alhambra : place)
attend-in(Garcia_Lorca : poet, First_Deep_Song_Festival : music_event)
collaborated-with(Garcia_Lorca : poet, Manuel_de_Falla, composer)
attend-in(Manuel_de_Falla: composer, First_Deep_Song_Festival : music_event)
wrote(Manuel_de_Falla: composer, La_Vida_breve : music)

b. Test Data 2: Topic map for Garcia Lorca
```

Fig. 5 Test Data 1 and Test Data 2 for the experiment.**Table 1** Test Data sets for the experiment.

Topic Maps constructs	Test Data 1	Test Data 2	Test Data 3
# of topic types	2	6	3
# of topics	2	7	685
# of association types	1	6	3
# of associations	1	8	2378
# of association role types	2	9	5
# of association roles	2	16	5
# of occurrence types	2	-	3
# of occurrences	2	-	750

• **Test Data 1** : The basic topic map for Tosca, **Test Data 2** : The Garcia Lorca topic map, **Test Data 3** : The CIA World Factbook topic map

respectively. The $n(pm)$ and $n(em)$ present the numbers of RDF statements that have been translated by these methods. If the $Complex(pm, em)$ for a particular translation method (em) is 2.0, it means the complexity of the translated RDF by em is twice as great as that of the proposed method.

$$Complex(pm, em) = \frac{Complex(em)}{Complex(pm)} = \frac{n(em)}{n(pm)} \quad (3)$$

For instance, with the proposed method (pm), 13 RDF statements are generated on Test Data 1 ($Complex(pm) = 13$). Meanwhile, with the Stanford's method (em), 36 RDF statements are obtained (Sect. 4.2.3) ($Complex(em) = 36$). Thus, $Complex(pm, em)$ is 2.77 (36/13), which means the complexity of the Stanford's method is x2.77 higher than that of the proposed method.

3.3 Test Data Sets for Experiment

Three Topic Maps examples are used as Test Data sets. The first (Fig. 5 (a)) is the most representative basic topic map example used by W3C's survey [6]. We also deal with more

complicated topic map examples (Test Data 2 and Test Data 3) to increase the reliability of the experimental results. Test Data 2 (Fig. 5 (b)) is "The Garcia Lorca," an extended version of Steve Pepper's example [14]. Test Data 3 is the topic map of the "CIA World Factbook," which can be found on the Ontopia web site [15]. The brief descriptions of the three topic map are shown in Table 1. Test Data 1 and Test Data 2 are described with LTM [16] in Fig. 5. Some constructs (e.g. topicName) of these Test Data sets and a complete description of Test Data 3 are not available here, because of the space limitation. By using the three distinct types of Test Data, we aimed to show how various features of topic maps are converted by the methods.

4. Experimental Results

This section presents experimental results and evaluations, with three Test Data sets. For this, the Test Data sets from Sect. 3.3 were translated using our proposed method and the other methods, respectively. Then, the results from each method were analyzed by the evaluation criteria designated

in Sect. 3.2.

4.1 Translation Results

Firstly, we describe the translation results from each method. We depict each of the translation results as a graph, to aid in a clear understanding of the translation processes and results. The graph represents only basic RDF triple elements (subjects, objects, and predicates) because representing all elements (including attributes) of RDF will not clarify the translation results. In here, our illustration of the translation results is limited to Test Data 1, which is simple, but furnished with all of the main elements, topics, association roles, associations and occurrences. Illustrations of the results from Test Data 2 and 3 were not included, since the translation process is virtually identical and full descriptions for all the three data sets will not be allowed due to the space limitation. However, in Sect. 4.2, all translated results from the three Test Data sets are used, to show precise comparison evaluations.

4.1.1 Semantics Extraction from the Source Test Data

It is very difficult to find complete semantics of Topic Maps automatically. Therefore, all semantics (explicit semantics and implicit semantics) were manually extracted from Topic Maps in order to achieve more precise comparison. In order to this, we conducted an hybrid ontology engineering method in which the forward engineering and the reverse engineering methods are combined; each is a distinct type of ontology engineering method. For the reverse engineering, we first defined a pattern of semantic structure of Topic Maps, on the basis of Topic Maps data model [1]. And then, semantics were extracted by applying that pattern to a basic Topic Maps sample which has been used in the previous studies. Here, we used the sample model of W3C survey. In the mean time, we asked experts in field of Topic Maps that they extract semantics in the basic Topic Maps sample as forward method. The semantics extracted by the each method were compared and analyzed. And again, the semantic pattern on the basis of Topic Maps data model was completed to conduct to a manual extraction. The pattern is applied to Test Data sets as a manual semantic extraction method.

Explicit semantics are basic meanings which are expressed in topic map. The explicit semantics are extracted based on topics, relations between topics and relations between topics and occurrences. Implicit semantics are meanings that have been newly derived from topic map, although are not intended to express originally. Implicit semantics provide a larger pool of knowledge, improving the Semantic Web quality. The implicit semantics are extracted from association roles. The extracted RDF semantics from Test Data 1 are shown below. As shown from this extraction result, Test Data 1 has six types of semantics, which includes both explicit and implicit semantics.

explicit semantics

$M_{a1} = (\text{puccini}, \text{composed-by}, \text{tosca})$

$M_{a2} = (\text{tosca}, \text{composed}, \text{puccini})$

$M_{a3} = (\text{tosca}, \text{premiere-date}, 1900-01-14)$

$M_{a4} = (\text{tosca}, \text{synopsis}, \text{'http://www.azopera.com/learn/synopsis/tosca.shtml'})$

implicit semantics

$M_{a5} = (\text{composer}, \text{compose}, \text{opera})$

$M_{a6} = (\text{opera}, \text{composed-by}, \text{composer})$

4.1.2 Translation Based on the Previous Methods

The Stanford's method basically translates t-nodes (topic nodes) and a-nodes (association nodes) of Topic Maps into subjects and predicates of RDF, respectively. Figure 6 presents translation results for Test Data 1, based on the Stanford's method. This method translates Topic Maps in the context of associations, and produces the semantic structure which differs from the original.

The Moore's method defines a relationship between a topic and an association role as an RDF statement, and again generates a relationship between the defined RDF statements. In a process of translation, this method generates additional nodes, (Assoc:1, Assocmember:1, Assocmember:2) that do not exist in the original source model. The translation result from Test Data 1 for this method is presented in Fig. 7.

The Ogievetsky's method creates new vocabulary called RTM (Rdf-Topic maps Mapping) for transforming Topic Maps to RDF. This method translates a topic to an RDF statement of type `rtm:topic`. The topic's subject locator becomes the URI of the subject of the statement; otherwise a blank node is created [6]. The translation result from Test Data 1 for this method is presented in Fig. 8.

Fig. 6 Translation result for Test Data 1, based on the Stanford's method. The circles symbolize subjects (s) or objects (o). Arrows mean predicates (p).

Fig. 7 Translation result based on the Moore's method.

Table 2 The number of RDF semantics and the semantic loss rates from each method's translated results.

	Test Data 1		Test Data 2		Test Data 3	
	$\sum \text{translated}S(m)$	% of LossR(m)	$\sum \text{translated}S(m)$	% of LossR(m)	$\sum \text{translated}S(m)$	% of LossR(m)
Stanford's method	2	67%	16	47%	4,756	14%
Moore's method	2	67%	16	47%	4,756	14%
Ogievetsky's method	3	50%	8	73%	3,128	43%
Garshol's method	3	50%	8	73%	3,128	43%
Unibo's method	4	33%	16	47%	5,506	0%
Proposed method	6	0%	30	0%	5,512	0%

- $\sum \text{translated}S(m)$: the number of RDF semantics that are generated by each methods, which is the variable of the formula (1)
- $\text{LossR}(m)$: the semantic loss rates that are calculated by the formula (1)

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@Prefix ntmr: <http://software.korea.ac.kr/
tmrdrfmapping/new-tm-rdf#> .

# translation result of explicit semantic from Topic Maps
[ rdf:subject :puccini;
  rdf:type music:person;
  rdfs:domain music:composer;
  :composed-by.work[
    rdf:object :tosca;
    rdf:type music:opera;]] .

[ rdf:subject :tosca;
  rdf:type music:opera;
  rdfs:domain music:work;
  :composed-by.composer [
    rdf:object :puccini;
    rdf:type music:person;]] .

[ rdf:subject :tosca;
  rdf:type music:opera;
  rdfs:domain music:work;
  :occ_premiere-date [
    rdf:object :1900-01-14;]] .

[ rdf:subject :tosca;
  rdf:type music:opera;
  rdfs:domain music:work;
  :occ_synopsis [
    rdf:object :http://www.azopera.com/learn/synopsis/tosca.shtml;]] .

# translation result of implicit semantics from Topic Maps
[ rdf:subject :work;
  rdf:type music:opera;
  :composed-by.person [
    rdf:object :composer;
    rdf:type music:person;]] .
[ rdf:subject :composer;
  rdf:type music:person;
  :composed-by.opera [
    rdf:object :work;
    rdf:type music:opera;]] .

# labeling topics
:tosca          rdfs:label "Tosca" .
:Puccini        rdfs:label "Giacomo Puccini" .
:person         rdfs:label "Person" .
:composer       rdfs:label "Composer" .
:opera          rdfs:label "Opera" .
:work           rdfs:label "Work" .
:composed-by    rdfs:label "Composed by" .
:premiere-date  rdfs:label "Première date" .
:synopsis        rdfs:label "Synopsis" .

```

Fig. 12 N3 description of Fig. 11.

```

[puccini : person = "Giacomo Puccini"]
[tosca : opera = "Tosca"]
[composer : person = "Composer"]
[work : opera = "Work"]

{tosca, premiere-date, 1900-01-14}
{tosca, synopsis, "http://www.azopera.com/learn/
synopsis/tosca.shtml"}

compose-by(tosca : work, Puccini : person)
composed-by(composer : person, work : opera)

```

Fig. 13 LTM description about the reverse translated result on Fig. 12.

spectively. For the Unibo's method, it is 27%. By contrast, the proposed method produces no semantic loss.

4.2.2 Reversibility

The reversibility is calculated with formula (2). For Test Data 1, the test results of the reversibility of the previous translation methods are borrowed from the W3C's survey [6] for reliable evaluation. According to the study, the Moore's method does not handle the occurrences, therefore the occurrences are disappeared when the translated data is translated back to Topic Maps. The Ogievetsky's method deals with all Topic Maps constructs. The Unibo's method missed the subject, Puccini, in the reverse translated result [6]. Therefore, this method does not regenerate associations and reverse translated only two semantics about relationships between the toscas and each occurrence. For the Garshol's method, the association roles are omitted from the round-tripped statements although all of the resulting statements match with the original. However, all constructs produced by the proposed method match with the original. Table 3 presents the reverse translated constructs ($\text{RevTransTMconst}(m)$) by each method on each Test Data set.

Table 4 and Fig. 14 illustrate evaluation results about the reversibility which are calculated by the formula (2).

4.2.3 Encoding Complexity

Table 7 and Fig. 15 present the encoding complexity of RDF translated by the proposed method and the existing methods. W3C's survey [6] already included an experiment with Test Data 1 for the existing methods. We de-

ment, all occurrences, type-instance relationships, and names of typing topics are omitted [6]. The Garshol's method deals with association as one-way predicate. $\sum \text{translated}S(m)$, the number of semantics in the translated result by each method calculated (Table 2). As shown in Fig. 14, on average, the Stanford's method and the Moore's method have the same semantic loss rates of 43%. The Ogievetsky's method and the Garshol's method are 55%, re-

a. Comparison of average semantic loss rates by each method

b. Comparison of average reversibility by each method

Fig. 14 Synthesized evaluation results on the semantic loss rates and the reversibility.**Table 3** The original Topic Maps constructs and the reverse translated Topic Maps constructs.

<i>OriginTMconst and RevTransTMconst(m)</i>	# of topics			# of associations			# of association roles			# of occurrences		
	TD1	TD2	TD3	TD1	TD2	TD3	TD1	TD2	TD3	TD1	TD2	TD3
Original Source	2	7	685	1	8	2,378	2	16	3	2	-	750
Stanford's method	2	7	685	1	8	2,378	2	16	3	0	-	0
Moore's method	2	7	685	1	8	2,378	2	16	3	0	-	0
Ogievetsky's method	2	7	685	1	8	2,378	2	16	3	2	-	750
Garshol's method	2	7	685	1	8	2,378	0	0	0	2	-	750
Unibo's method	1	0	250	0	0	0	0	0	0	2	-	750
Proposed Method	2	7	685	1	8	2,378	2	16	3	2	-	750

Table 4 Experimental results for the reversibility.

Translation method	Test Data 1	Test Data 2	Test Data 3
Stanford's method	$Rev=5/7=0.71$ (71%)	$Rev=31/31=1.00$ (100%)	$Rev=3,066/3,816=0.80$ (80%)
Moore's method	$Rev=5/7=0.71$ (71%)	$Rev=31/31=1.00$ (100%)	$Rev=3,079/3,816=0.81$ (81%)
Ogievetsky's method	$Rev=7/7=1.00$ (100%)	$Rev=31/31=1.00$ (100%)	$Rev=3,816/3,816=1.00$ (100%)
Gashol's method	$Rev=5/7=0.71$ (71%)	$Rev=15/31=0.48$ (48%)	$Rev=3,813/3,816=0.99$ (99%)
Unibo's method	$Rev=3/7=0.43$ (43%)	$Rev=0/31=0.00$ (0%)	$Rev=1,003/3,816=0.26$ (26%)
Proposed method	$Rev=7/7=1.00$ (100%)	$Rev=31/31=1.00$ (100%)	$Rev=3,816/3,816=1.00$ (100%)

defined estimation rules for measuring the number of translated RDF statements based on the results of W3C's survey. A precise evaluation of encoding complexity is obtained by comparing the numbers of RDF statements required to express the same semantics. As explained in Sect. 4.2.1 and Table 2, some translation methods did not deal with occurrences or association roles. Therefore, among resulting RDF statements created by the proposed method and the existing methods, we only count the RDF statements which are also found to have the same semantics, excluding translated results on the occurrences and association roles. According to W3C's survey [6] for Test Data 1 (excluding occurrences), the Stanford's method, the Moore's method, the Ogievetsky's method, the Garshol's method and the Unibo's method produced 36, 20, 50, 7 and 20 statements, respectively. If we assume that the Garshol's method translates associations in both directions, this method generates 11 statements. On the other hand, the proposed method produced 13 statements. The defined estimation rule for measuring the number of translated RDF statements is shown in Table 5.

Fig. 15 Comparison of average encoding complexity.

The number of RDF statements produced applying the estimation rule (Table 5) to Test Data sets (Table 1). The results are shown in Table 6. As shown in Table 7 and Fig. 15, the encoding complexity of the Stanford's method is on average x2.39 higher than that of the proposed method, the Moore's method is x1.36 higher, the Ogievetsky's method is x1.68 higher, the Garshol's method is x0.91 higher, and the Unibo's method is x1.24 higher. In conclusion, encoding complexity of the RDFs translated by the proposed method

Table 5 Estimation rules for measuring translated RDF statements.

Translation methods	Estimation Rules for measuring translated RDF statements
Stanford's method	<ul style="list-style-type: none"> - 10 RDF state. for each association, 12 RDF state. for each topic-basename relationship - 1 RDF state. to define node type for each topic, topic types and association - 4 RDF state. to define other node types (associationTemplate, role-label, role-topic and role-basename)
Moore's method	<ul style="list-style-type: none"> - 3 RDF state. for each topic, association, and topic type - 5 RDF state. for association
Ogievetsky's method	<ul style="list-style-type: none"> - 5 RDF state. to define basenames for each topic, topic type and association - 5 RDF state. for each topic-instance and association - 3 RDF state. for each type of topic types, 2 RDF state. for each type of associations, 1 RDF state. for each type of topics
Gasohol's method	<ul style="list-style-type: none"> - 8 RDF state. for each association (assume that this method has translated associations in both directions) - 1 RDF state. to define label for each topic type and association type
Unibo's Method	<ul style="list-style-type: none"> - 2 RDF state. for each topic, 7 RDF state. for each association - 3 RDF state. to define each topic type and associations type
Proposed method	<ul style="list-style-type: none"> - 8 RDF state. for each association - 1 RDF state. for each labeling of topics, topic types and association types

Table 6 Estimated RDF statements (*Complex(pm)* and *Complex(em)*).

Translation method	Test Data	Estimated RDF statements (approximate)
Stanford's Method	TD 2	- 189 state. = $8(a) \times 10 + 7(t) \times 12 + ((6(tt) + 9(rt) + 6(at)) \times 1 + 4)$
	TD 3	- 32,015 state. = $2378(a) \times 10 + 685(t) \times 12 + ((3(tt) + 5(rt) + 3(at)) \times 1 + 4)$
Moore's Method	TD 2	- 103 state. = $(7(t) + 8(a) + 6(tt)) \times 3 + (8(a) \times 5)$
	TD 3	- 25,844 state. = $(685(t) + 2378(a) + 3(tt)) \times 3 + (2378(a) \times 7)$
Ogievetsky's method	TD 2	- 217 state. = $(7(t) + 6(tt) + 8(a)) \times 5 + (7(t) + 8(a)) \times 5 + 6(tt) \times 3 + 6(at) \times 2 + 7(t) \times 1$
	TD 3	- 31,345 state. = $(685(t) + 3(tt) + 2378(a)) \times 5 + (685(t) + 2378(a)) \times 5 + 3(tt) \times 3 + 3(at) \times 2 + 685(t) \times 1$
Gasohol's method	TD 2	- 76 state. = $8(a) \times 4 \times 2 + 6(tt) + 6(at)$
	TD 3	- 19,030 state. = $2378(a) \times 4 \times 2 + 3(tt) + 3(at)$
Unibo's Method	TD 2	- 106 state. = $7(t) \times 2 + 8(a) \times 7 + ((6(tt) + 6(at)) \times 3)$
	TD 3	- 18,034 state. = $685(t) \times 2 + 2378(a) \times 7 + ((3(tt) + 3(at)) \times 3)$
Proposed Method	TD 2	- 83 state. = $8(a) \times 8 + 7(t) + 6(tt) + 6(at)$
	TD 3	- 19,715 state. = $2378(a) \times 8 + 685(t) + 3(tt) + 3(at)$

* t : topic, a : association, r : association role, tt : topic type, at : association type, rt : association role type, ot : occurrence type

Table 7 Experimental results for the encoding complexity.

Translation method	Test Data 1	Test Data 2	Test Data 3
Stanford's method	$Complex=36/13=2.77$	$Complex=189/83=2.78$	$Complex=32,015/19,715=1.62$
Moore's method	$Complex=20/13=1.54$	$Complex=103/83=1.24$	$Complex=25,844/19,715=1.31$
Ogievetsky's method	$Complex=50/13=3.85$	$Complex=217/83=2.61$	$Complex=31,345/19,715=1.59$
Gashol's method	$Complex=11/13=0.85$	$Complex=76/83=0.92$	$Complex=19,030/19,715=0.97$
Unibo's method	$Complex=20/13=1.54$	$Complex=106/83=1.28$	$Complex=18,034/19,715=0.91$

is much lower than those of the others.

5. Conclusion and Future Work

The Semantic Web is one of the most promised candidates as the Web tomorrow, whose basis is on RDF and RDF Schema. The purpose of the Semantic Web is to make data on the Web available not only for human beings but also for automated processing, which would be specialized for the integration and reuse of data across various applications [18]. There have been many efforts to develop tools that can "export" a web of data into RDF form. Although the presence of various methods, several problems still remain, such as high encoding complexity, loss of semantics and constructs.

This paper proposes a novel TM-to-RDF translation

method in order to resolve the limitations of the existing methods. As demonstrated by the evaluation results in Sect. 4, our proposal is superior to the previous translation methods. In terms of the semantic loss rate, the proposed method translated even "implicit" semantic relationships in Topic Maps from association roles and occurrence. The proposed translation method provided a more improved semantics extraction rate (about 27% ~ 55%) compared to the existing methods (see Fig. 14(a)). On the reversibility, the proposed method generated all Topic Maps constructs in an original source when is reverse translated. With regard to the encoding complexity, our method showed a better result with reduction in the number of RDF statements. The encoding complexity of our method decreases more than 24% when compared with the existing methods except the Garshol's method (see Fig. 15).

In further study, we will deal with the non-binary relationship translation issue. Topic Maps permits non-binary relationships: The existing methods for translating associations with more than two role-players are unsatisfactory [6]. Also, research on relationships with other elements of the Semantic Web will be conducted. Through studies about TMCL (Topic Map Constraint Language) and OWL (Web Ontology Language), more information about converting Topic Maps to RDF can be obtained. Eventually, this will enable us to realize the practical Semantic Web.

References

- [1] ISO/IEC 13250-2 Information technology — Topic Maps — Part 2: Data model. ISO/IEC JTC1, 2006.
- [2] RDF: Concepts and Abstract Syntax, W3C, 2004.
- [3] RDF Vocabulary Description Language 1.0: RDF Schema, W3C, 2004.
- [4] D. Harrisa, L. Khana, R. Paulb, and B. Thuraingham, "Standards for secure data sharing across organizations," *Computer Standards & Interfaces*, vol.29, no.1, pp.86–96, 2007.
- [5] W3C Semantic Web Frequently Asked Questions, <http://www.w3.org/RDF/FAQ#Manual>, W3C, 2008.
- [6] S. Pepper, F. Vitali, L.M. Garshol, N. Gessa, and V. Presutti, A Survey of RDF/Topic Maps Interoperability Proposals, W3C, 2006.
- [7] L.M. Garshol, "Living with Topic Maps and RDF," *Proc. XML Europe 2003*, London, England, 2003.
- [8] M.S. Lacher and S. Decker, "On the integration of Topic Maps and RDF data," *Proc. Semantic Web Workshop*, California, USA, 2001.
- [9] G. Moore, "RDF and Topic Maps—An exercise in convergence," *Proc. XML Europe 2001*, Berlin, Germany, 2001.
- [10] N. Ogievetsky, "XML Topic Maps through RDF glasses," *J. Markup Languages: Theory and Practice*, vol.3, no.3, pp.333–364, 2001.
- [11] P. Ciancarini, R. Gentilucci, M. Pirruccio, V. Presutti, and F. Vitali, "On the integration of RDF and Topic Maps," *Proc. Extreme Markup Languages 2003*, Montreal, Canada, Aug. 2003.
- [12] RDF Schema, W3C, 2004.
- [13] OWL Web Ontology Language Overview, W3C, 2004.
- [14] S. Pepper, Ten Theses on Topic Maps and RDF, <http://www.ontopia.net/topicmap/materials/rdf.html>, 2002.
- [15] Ontopia, A CIA World Factbook Topic Map, <http://www.ontopia.net/omnigator/plugins/export/export.jsp?tm=factbook.hytm&filename=factbook.hytm.ltm&type=xml&format=ltm>
- [16] Linear Topic Maps Notation, <http://www.ontopia.net/download/ltm.html>, 2006.
- [17] Notation 3, <http://www.w3.org/DesignIssues/Notation3>, 2006.
- [18] T. Morita, N. Izumi, N. Fukuta, and T. Yamaguchi, "A graphical RDF-based meta-model management tool," *IEICE Trans. Inf. & Syst.*, vol.E89-D, no.4, pp.1368–1377, April 2006.

Shinae Shin received the B.S. degree from the Kyonggi University in 1993, and the M.S. degrees from the Korea University in 1999, Seoul, Korea, all in computer science. She is a Ph.D. candidate in the Department of Computer Science & Engineering, in the Korea University. In 1993 she joined NIA (National Information Society Agency) in Korea, where she is a senior research fellow. Her primary research interests include Metadata, Semantic Web, Ontology and EA (Enterprise Architecture).

Dongwon Jeong received his B.S degree in Computer Science from Kunsan National University, Korea, in 1997. He received his M.S. and Ph.D. degree in Computer Science from Choongbuk National University, Korea, in 1999 and from Korea University, Korea, in 2004 respectively. He was a Visiting Research Scholar, School of Information Sciences & Technology, Pennsylvania State University, US, 2005. Now, he is a Professor in the Dept. of Informatics and Statistics, Kunsan National University, Korea from 2005. His research interests include Data Integration, Semantic Web, Semantic Sensor Network, Semantic GIS and Security.

Doo-Kwon Baik received his B.S. degree in Mathematics from Korea University, Seoul, Korea, in 1974. He received his M.S. and Ph.D. degrees in Computer Science from Wayne State University at U.S.A., in 1983 and 1986, respectively. Currently, he is a full professor in the Department of Computer Science and Engineering, Korea University, Seoul, Korea. Also, he is a Korea committee chair of ISO/IEC JTC1/SC32-Data Management and Interchange, 1992-Present. His research interests include Data Engineering, Software Engineering, and Modeling and Simulation.