Translating Topic Maps to RDF/RDF Schema for The Semantic Web

Shinae Shin
Department of EA & IS Audit, National Information Society Agency, Seoul, 100-775 Korea
Email: sashin@nia.or.kr

Dongwon Jeong
Department of Informatics and Statistics, Kunsan National University, San 68, Miryong-Dong, Gunsan, Jeollabuk-Do, 573-701 Korea
Email: djeong@kunsan.ac.kr

Doo-Kwon Baik
Department of Computer Science and Engineering, Korea University, 5-1, Anam-Dong, Seongbuk-Ku, Seoul, 136-701 Korea
Email: baikdk@korea.ac.kr

This paper focuses on the issue for translating Web metadata, especially Topic Maps and RDF/RDF Schema (RDF(S)). For the Semantic Web, all Web information resources should be accepted and understood by any browsers and web applications. Topic Maps and RDF are both representative international standards for description of Web information resources. However, they have different schemes and syntaxes, and thus much research has been studied to achieve the interoperability between both standards. However, there still remain several issues such as semantic loss, high structural complexity, and so on. This paper proposes an improved method for translating Topic Maps to RDF(S). Our method translates Topic Maps to RDF(S) based on topics, association roles and occurrences (TRO). The proposed method covers explicit semantics and implicit semantics by the object and the semantic mapping between Topic Maps and RDF(S). In addition, our method has strength in terms of the completeness and the naturalness. To illustrate the advantages of our proposal, we conduct experiments with several Topic Maps data sets and perform a comparative evaluation between our proposal and the previous methods.

Keywords: Topic Maps, RDF, RDF schema, Semantic Web, Web Metadata
ACM Classifications: H.0 (Information Systems – General), H.1 (Information Systems – Model and Principles)

1. INTRODUCTION
The vision of the Semantic Web is to extend principles of the Web from documents to data (W3C, 2008). For this, the most important step is to define and describe relations among data (i.e., resources) on the Web (W3C, 2008). Topic Maps and RDF (Resource Description Framework) are recognized as the most representative standards to describe and build Web information and

Copyright© 2009, Australian Computer Society Inc. General permission to republish, but not for profit, all or part of this material is granted, provided that the JRPIT copyright notice is given and that reference is made to the publication, to its date of issue, and to the fact that reprinting privileges were granted by permission of the Australian Computer Society Inc.

Manuscript received: 16 February 2008
Communicating Editor: John Grundy
resources. RDF was initially intended as a foundation for metadata processing of Web resources with the important feature of automation of information and provision of interoperability between applications exchanging machine-understandable information (Kitcharoensakkul et al., 2001). RDF plays the role of a common model, as a kind of glue to integrate the data (W3C, 2008). There have been many efforts to develop tools that can export a web of data into RDF form and the tools include automatic procedures that can produce RDF data on-the-fly as an answer to queries (W3C, 2008).

Many Topic Maps-to-RDF translation methods (Garshol, 2003; Lacher et al., 2001; Moore, 2001; Ogievetsky, 2001 and Ciancarini et al., 2003), have been devised. Although various methods have been proposed, several problems still remain. The Garshol method (Garshol, 2003) uses new vocabularies called RTM (Rdf-Topic map Mapping). The translated results by this method are clear, but this method misses association roles. Also the method deals with associations as one-way relation although these are non-directional. The Stanford approach (Lacher et al., 2001) translates Topic Maps in the context of associations, changing the semantic structure of the original. In addition, this method also misses occurrences. The Moore method (Moore, 2001) generates additional nodes, {Assoc:1, Assacmember:1, Assacmember:2} that do not exist in the original source model. It makes the structure more complex and incomprehensive. This method also does not treat occurrences. Ogievetsky (Ogievetsky, 2001) presented the XTM2RDF converter that uses new vocabularies to translate XML TM to RDF. Therefore, a set of new vocabularies is required for the translation. W3C also surveyed the same issues (Pepper et al., 2006). This survey analyzes the most representative methods, the aforementioned five translation methods and summarizes the shortcomings of the methods with two criteria, ‘Completeness’ and ‘Naturalness’. The analysis result is summarized as follows:

- The Garshol method does not utilize the standard RDF and RDFS predicates, and thus always requires a mapping to be specified.
- The Stanford approach is complete based on PMTM4 (Processing Model for Topic Maps) that is not a complete model for Topic Maps.
- Both the Moore method and The Ogievetsky’s method produce many extra RDF statements reducing ‘Naturalness’.
- The Uniibo proposal (Ciancarini et al., 2003) deals with some features unclearly, e.g., reification of roles and is far less satisfactory, since this adopts additional baggages.

The W3C survey also discusses that mapping methods can be categorized at either the object or the semantic level. The object mapping uses the low-level building blocks of one language to describe the object model of the other. The semantic mapping starts from higher level concepts that carry the semantics of each model. Our proposed method combines the object mapping and the semantic mapping, because our method translates Topic Maps and RDF(S) based on constructs and extracts explicit and implicit semantics depending on semantic relations.

The remainder of the paper is structured as follows. Section 2 presents the concepts, key definitions, and translation rules of our proposal. It also contains a translation example with the proposed method, as a case study. In Section 3 and Section 4, we define an evaluation methodology, conduct experiments, and describe the evaluation result. Finally, Section 5 concludes this paper.

2. TRANSLATION METHOD

2.1 Preliminary
We first formally define Topic Maps and RDF as Definition 1. Basic Topic Maps are organized into two topics (t), which represent subjects of discourse; one associations (a), which represent
relationships between the subjects; two association roles (ar), which represent the involvement of a subject in an association relationship and occurrences (occ), which connect the subjects to pertinent information resources. Each topic, association, association role, and occurrence has its own type. An RDF triple is composed with subject (s), predicate (p) and object (o), respectively. We also define symbols and notations required for description of our definitions (Table 1).

**Definition 1.** The basic Topic Maps model (TM) and the RDF model (RDF Triple) are defined as follows:

\[
\text{TM} = (t_i, t_j, a, ar_i, ar_j, occ) := (I \cup B \cup L) \times (I \cup B \cup L) \times (I \cup B \cup L) \times (I \cup B \cup L) \times (I \cup B \cup L) \times (I \cup L)
\]

\[
\text{RDF Triple} = (s, p, o) := (I \cup B) \times I \times (I \cup B \cup L)
\]

- \(I\): the set of all IRIs (This is the data type of locators using the IRI notation; the IRIs shall be absolute. The identifier of this data type is http://www.w3.org/2001/XMLSchema#anyURI.)
- \(B\): the set of all blank nodes
- \(L\): the set of all RDF Literals

### 2.2 Translation Rules

Figure 1 shows the overall translation framework. In this figure, identification is accomplished to extract addressable resources to be mapped to subjects of RDF. After the identification process, the translation is carried out based on topics, occurrences, and association roles to generate RDF triples. The translation rules consist of four sub-rules to identify resources and to translate Topic Maps according to its components (i.e., topics, association roles and occurrences). The proposed translations rules, MTR is defined as Definition 2.

**Definition 2.** The translation rules, MTR is defined as 4-tuple.

\[
\text{MTR:} = (Id, Mt, Mr, Mocc)
\]

- \(Id\) denotes the identifying part of addressable resources from origin Topic Maps
- \(Mt\) denotes the translation part based on topics
- \(Mr\) denotes the translation part based on association roles
- \(Mocc\) denotes the translation part based on occurrences

<table>
<thead>
<tr>
<th>Notations/\Symbols</th>
<th>Descriptions</th>
<th>Notations/\Symbols</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t)</td>
<td>topic of TM</td>
<td>(nT)</td>
<td>the number of topic</td>
</tr>
<tr>
<td>(tt)</td>
<td>topic type of TM</td>
<td>(nTT)</td>
<td>the number of topic type</td>
</tr>
<tr>
<td>(a)</td>
<td>association of TM</td>
<td>(nA)</td>
<td>the number of association</td>
</tr>
<tr>
<td>(at)</td>
<td>association type of TM</td>
<td>(nAT)</td>
<td>the number of association type</td>
</tr>
<tr>
<td>(ar)</td>
<td>association role of TM</td>
<td>(nR)</td>
<td>the number of association role</td>
</tr>
<tr>
<td>(rt)</td>
<td>association role type of TM</td>
<td>(nRT)</td>
<td>the number of association role type</td>
</tr>
<tr>
<td>(occ)</td>
<td>occurrence of TM</td>
<td>(nOCC)</td>
<td>the number of occurrence</td>
</tr>
<tr>
<td>(occt)</td>
<td>occurrence type of TM</td>
<td>(nOCCT)</td>
<td>the number of occurrence type</td>
</tr>
<tr>
<td>(s)</td>
<td>subject of RDF</td>
<td>(p)</td>
<td>predicate of RDF</td>
</tr>
<tr>
<td>(o)</td>
<td>object of RDF</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 1:** The defined notations and symbols
Translating Topic Maps to RDF/RDF Schema for The Semantic Web

The first step of the translation method is to identify addressable resources. Topics Maps and RDFs use symbols to represent identifiable “things”. The term “thing” corresponds to “subject” in Topic Maps, and “resource” in RDFs. **Topic** \((t)\), **association role**, and **occurrence** of Topic Maps expresses a “thing”. Each of the **subject**, **object** and **predicate** in an RDF can be a resource (RDF, 2004). Hence, **topic**, **association role** and **occurrence** in Topic Maps can be translated into **subject**, **object** or **predicate**. The “thing” of Topic Maps can be an addressable resource with a URI and a non-addressable concept expressed by a literal (Pepper, 2002). However, the **subject** of RDF must be an addressable resource with a URI (RDF, 2004). Therefore, only addressable resources of Topic Maps can be translated into **subjects** of RDF.

Topic Maps offers two modes of identification, direct (**subject locators**) and indirect (**subject identifiers**), whereas RDFs offers only one (**URIref**) (Pepper et al., 2006). According to the W3C survey, Stanford ignored the identity issue and Moore did not state explicitly. Both Ogievetsky and Unibo favour **subject locators** and define a separate property for handling **subject identifiers**. Garshol translates both to URIs. The proposed method translates also both **subject identifiers** and **subject locators** into the **URIref** of the RDF node. The identification rule is the Rule 1.

**Rule 1 (Id: the rule for identification of addressable resources)**

\[
\text{INPUT TM;}
\text{INIT QT}_{(t\text{Rear})}, \text{QR}_{(r\text{Rear})}, \text{QOCC}_{(occ\text{Rear})};
\text{WHILE(true)} \{
\text{SELECT } t \text{ FROM TM;}
\text{If hasID}(t) \quad // \text{hasID}(t): \text{a topic with subject locator or subject identifier}
\text{than INSERT } t \rightarrow \text{QT}_{(t\text{Rear})} \text{ and } t\text{Rear}++;
\}
\text{WHILE(true)} \{
\text{SELECT } ar \text{ FROM TM;}
\text{If hasID}(ar) \quad // \text{hasID}(ar): \text{an association role with subject locator or subject identifier}
\text{than INSERT } ar \rightarrow \text{QR}_{(r\text{Rear})} \text{ and } r\text{Rear}++;
\}
\text{WHILE(true)} \{
\text{SELECT } occ \text{ FROM TM;}
\text{If hasID}(occ) \quad // \text{hasID}(occ): \text{an occurrence with subject locator or subject identifier}
\text{than INSERT } occ \rightarrow \text{QOCC}_{(occ\text{Rear})} \text{ and } occ\text{Rear}++;
\}
\]
When \textit{topic} is translated into \textit{subject}, \textit{association role} can be translated into \textit{rdfs:domain}. \textit{rdfs:domain} declares the classes that correspond to \textit{subjects} in the triple model (RDF SCHEMA, 2004). Other attributes, \textit{topicName} and \textit{scope} are respectively translated to \textit{rdfs:label} and \textit{rdfs:range}. For definition of \textit{predicates} from Topic Maps, we need to extract \textit{association roles}, since the relations between \textit{subjects} and \textit{objects} are influenced by \textit{association roles} of the \textit{objects} in the \textit{associations}. The rule is defined by Rule 2.

Rule 2 (Mt : the translation rules based on topics)

\begin{verbatim}
INPUT TM, QT;
SetT = \{QT(t_s) | \forall s = 1 \ldots tRear\};
For 1 \leq s \leq tRear {  
  SELECT t_s FROM SetT;  // This step means only addressable TM(t) is translated into RDF(s).
  SELECT t_o FROM TM, where t_o be related with t_s by an identical association;
  CREATE topic_pairs (t_s, t_o);
  CREATE RDF(p) FROM TM(a_r_o) for topic_pairs (t_s, t_o);
  \text{TRANS (t_s, a_r_o, t_o) \rightarrow RDF(s, p, o);}
  \text{TRANS tt \rightarrow rdf:type, ar \rightarrow rdfs:domain, topicName \rightarrow rdfs:label, scope \rightarrow rdfs:range;}
}
\end{verbatim}

\textit{Association role} is a representation of the involvement of \textit{subject} in a relationship represented by \textit{association}. Thus, \textit{association role} also becomes new \textit{topic} and be translated into \textit{subject} or \textit{object}. The rule for the translation based on \textit{association roles} is defined as the Rule 3.

Rule 3 (Mr : the translation rule based on association roles)

\begin{verbatim}
INPUT TM, QR;
SetR = \{QR(ar_s) | \forall s = 1 \ldots rRear\};
For 1 \leq s \leq rRear {  
  SELECT ar_s FROM SetR;
  SELECT ar_o FROM TM, where ar_o is related with ar_s by an identical association;
  CREATE role_pairs (a_s, a_o);
  CREATE RDF(p) FROM TM(a) for role_pairs (a_s, a_o);
  \text{TRANS (r_s, a_s, r_o) \rightarrow RDF(s, p, o);}
  \text{TRANS rt \rightarrow rdf:type;}
}
\end{verbatim}

\textit{Occurrence} can be translated to an RDF statement. \textit{Occurrence} is a specialized type of \textit{association}, where one participant in the \textit{association} shall be an information resource (TM, 2006). So, our method is devised to translate a relationship between \textit{topic} and an information resource into another RDF statement. For this, \textit{topic} and \textit{occurrence locator} are translated into \textit{subject} and \textit{object}, respectively. And \textit{occurrence} and \textit{occurrence type} become \textit{predicate} and \textit{rdf:type} of the \textit{object}, respectively in our method. This rule for the translation based on \textit{occurrences} is defined as the Rule 4.
Translating Topic Maps to RDF/RDF Schema for The Semantic Web

Rule 4. (Mocc denotes the translation phase based on occurrences)

\[
\text{INPUT} \; TM, \; QOCC; \\
\text{SetOcc} = \{QOCC(occ_i) \mid \forall \; s = 1 \ldots occRear\}; \\
\text{For} \; 1 \leq s \leq \text{occRear} \{ \\
\quad \text{SELECT} \; occ_c, \; \text{FROM SetOcc}; \\
\quad \text{SELECT} \; ol_{occ}, \; t_s \; \text{FROM} \; TM, \; \text{where} \; t_s \; \text{is related with an identical occurrence locator with the} \; occ_c; \\
\quad \text{CREATE} \; \text{t\_occ\_pairs}(t_s, \; ol_{occ}) \\
\quad \text{TRANS} \; (t_s, \; occ, \; ol_{occ}) \rightarrow \text{RDF}(s, \; p, \; o); \\
\quad \text{TRANS} \; tt \rightarrow \text{rdfs\text{:type}}, \; r \rightarrow \text{rdfs\text{:domain}}, \; \text{topicName} \rightarrow \text{rdfs\text{:label}}; \\
\} \\
\]

2.3 Case Study

This section shows a translation result of the proposed method with the TM data set that is the most representative basic Topic Maps example used by the W3C survey. This example is small but has all core topic map constructs, TM(\(t\)), TM(\(a\)), TM(\(r\)), and TM(\(o\)). The test data is described in Figure 5 of Section 3.2. Figure 2a illustrates the original Topic Maps data and its translation result by the proposed method in this paper. This figure shows only the major elements of the result in order to help readers’ understanding. Figure 2b illustrates the translated result, RDF statements in N3 (N3, 2006). As shown in Figure 2, the proposed method in this paper handles not only explicit semantics but also implicit semantics. The explicit semantics are the basic meanings which are expressed in Topic Maps. On the other hand, the implicit semantics are newly derived from Topic Maps. The implicit semantics will help knowledge pool formation, improving the Semantic Web quality. The proposed method generates the explicit semantics from \textit{topics}, from relations between \textit{topics} or from relations between \textit{topics} and \textit{occurrences}. The implicit semantics are created from \textit{association roles}.

2.4 Prototyping Implementation

We implemented the translation algorithm with several sample data sets. The system environment for the prototyping implementation is as follows:

- CPU: AMD Athlon(tm) 64 X2 Dual Core Processor 2.11GHz
- Memory: 2.00GB RAM
- OS: MS Windows XP Professional
- Programming Language: Java (JDK 1.6.0_01)
- Integrated Development Tool: Eclipse
- XML Parser: DOM Parser

Figure 3 depicts the architecture of the prototype system implemented for the experiment under the development environment. The prototype system consists of key five components: DOM parser, Extractor, Translator, Storage Manager, and Persistent Storage. The DOM parser reads the Topic Maps data in XML into memory and converts it into an XML DOM object. The XML DOM object can be accessed with Java used for implementation of the prototype system. Java provides APIs to create the XML DOM object. The Extractor component identifies addressable resources in order to extract candidates corresponding to \textit{subjects} in RDF. The component, Translator has a role to generate RDF statements according to the rules described in Section 2.2. The final component,
Translating Topic Maps to RDF/RDF Schema for The Semantic Web

Figure 2: Translation result based on the proposed method.

a) The translation result – graph form

```
# translation result of explicit semantic from Topic Maps
[rdf:subject :puccini;
 rdf:type music:person;
 rdfs:domain music:composer;
 rdfs:label "Giacomo Puccini";
 :compose.opera [  
    rdf:object :tosca;
    rdf:type music:opera;
    rdfs:label "Tosca";]] .
[rdf:subject :tosca;
 rdf:type music:opera;
 rdfs:domain music:work;
 rdfs:label "tosca";
 :compose.composer [  
    rdf:object :puccini;
    rdf:type music:person;
    rdfs:label "Giacomo Puccini";]
 ] .
[rdf:subject :tosca;
 rdf:type music:opera;
 rdfs:domain music:work;
 rdfs:label "tosca";
 :premiere-date [  
    rdf:object :1900-01-14;
    rdf:type music:premiere-date;
    rdfs:label "Premiere-date";]] .
```

b) Part of the translation result – N3 description
Storage Manager stores the translated data, i.e., the RDF data to persistent storage. In this paper, the storage model in Jena is used as persistent storage. The Jena storage is based on the RDB model, which is widely used for storing data.

With the prototype, we verify the translated results by our algorithm. Figure 4 shows a snapshot of the implementation. In this figure, we first select a Topic Maps data sample, and then the data is displayed in the area, “Selected TM data”. When the button in the middle of the window is clicked, the translated result is displayed in the area, “Translated Result”. The button labeled “Save” can be used to permanently save the translated result. The detailed descriptions of the TM data and the translated results are described in other sections, and thus we omit the descriptions of them here.

3. EVALUATION
This section describes the comparative items and evaluation models, and the test data sets.

3.1 Comparative Items and Evaluation Models
We conduct experiments to compare our proposed method with the existing translation methods by using the two comparative items: (1) Reversibility, (2) Structural complexity of translated result.
These items correspond to the W3C evaluation items (Pepper et al, 2006), \textit{Completeness} and \textit{Naturalness} respectively. The W3C survey report used these criteria for the evaluation of the existing research. However, W3C did not show any concrete formula for the evaluation. Thus we define the formulas for the evaluation on the comparative items.

\textbf{A. Completeness – The reversibility}

The W3C’s survey defines \textit{completeness} to evaluate the extent to which a certain method can handle every semantic construct in the source model and provides a means to represent it without loss of semantics in the target model. In order to measure the \textit{completeness} of our proposal, we evaluate on reversibility. Reversibility measures differences between the original and the result which has been translated twice to return to the starting model. Reversibility is measured by comparison between the numbers of the topic map constructs in the round trip result and the original source topic map constructs. Therefore, reversibility is an evaluation item for the more accurate experiment on \textit{completeness}. Reversibility is calculated by counting core constructs such as \textit{subject topics}, \textit{associations}, \textit{occurrences}, or \textit{association roles} in the Topic Maps generated by the reverse translation. \textit{OrigTMconst} is the number of \textit{topics}, \textit{associations}, \textit{occurrences}, and \textit{association roles} in the original source, and \textit{RveTMconst} is the number of constructs by the reverse translation.

\begin{equation}
    Rve(m) = \frac{\sum RveTMconst(m)}{\sum OrigTMconst} = \frac{n(RveT) + n(RveA) + n(RveO) + n(RveR)}{n(OrigT) + n(OrigA) + n(OrigO) + n(OrigR)}
\end{equation}

\begin{itemize}
    \item \textit{n(OriginT)}: the number of \textit{topics} in the original source topic maps
    \item \textit{n(OriginA)}: the number of \textit{associations} in the original source topic maps
    \item \textit{n(OriginO)}: the number of \textit{occurrences} in the original source topic maps
    \item \textit{n(OriginR)}: the number of \textit{association roles} in the original source topic maps
    \item \textit{n(RveT)}: the number of \textit{topics} in the reverse translated topic maps
    \item \textit{n(RveA)}: the number of \textit{associations} in the reverse translated topic maps
    \item \textit{n(RveO)}: the number of \textit{occurrences} in the reverse translated topic maps
    \item \textit{n(RveR)}: the number of \textit{association roles} in the reverse translated topic maps
\end{itemize}

\textbf{B. Naturalness – The structural complexity}

The W3C survey defines \textit{naturalness} as the degree of readability of the translation result and counts the number of RDF statements of the translated result. We also use the same method, and this paper defines \textit{naturalness} as structural complexity. Structural complexity is evaluated by analyzing the number of resulting RDF statements. The formula for structural complexity, \textit{Cx(m)} is defined below (2). \textit{pm} and \textit{em} indicate the proposed method and one of the existing methods respectively. \textit{n(RDFstate(pm))} and \textit{n(RDFstate(em))} mean the numbers of RDF statements that have been translated by the proposed method and the existing methods.

\begin{equation}
    Cx(m) = \frac{Cx(em)}{Cx(pm)} = \frac{n(RDFstate(em))}{n(RDFstate(pm))}
\end{equation}

\textbf{3.2 Test Data Sets for the Experiment}

In this paper, three Topic Maps examples are used to increase the reliability of the experimental results. By using the three distinct types of test data, we tried to show how various features of Topic Maps are converted by the methods.

\begin{itemize}
    \item Test data 1 is the representative data used by much of the existing research including the W3C survey as referred to in the section 2.3.
\end{itemize}
Translating Topic Maps to RDF/RDF Schema for The Semantic Web

Test data 2 is the Topic Map for “The Garcia Lorca”, an extended version of Steve Pepper’s example (Pepper, 2002). The test data 2 has relatively a large number of association types and association role types than the test data 1, but has no occurrence.

Test data 3 is the Topic Map of “The Italian opera” which can be found on the Ontopia web site (Ontopia, 2005). It is significantly large in size.

Table 2: Test data sets for the experiment

<table>
<thead>
<tr>
<th>TM Constructors</th>
<th>nTT</th>
<th>nT</th>
<th>nAT</th>
<th>nA</th>
<th>nRT</th>
<th>nR</th>
<th>nOCCT</th>
<th>nOCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Data 1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Test Data 2</td>
<td>6</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>16</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Test Data 3</td>
<td>49</td>
<td>1918</td>
<td>28</td>
<td>3479</td>
<td>28</td>
<td>28</td>
<td>23</td>
<td>1488</td>
</tr>
</tbody>
</table>

Test data 1 is described with LTM (LTM, 2006) in Figure 5, but test data 2 and test data 3 are not available here because of space limitations. A brief description of the three Topic Maps compositions is shown in Table 2.

4. EXPERIMENT RESULTS

This section shows the experiment and evaluation results on the existing methods and the proposed method with the three Topic Maps test data. For the comparative evaluation on the proposed method, the Stanford method, the Moore method, and the Unibo method are selected, but the Garshol method and the Ogievetsky method are excluded because both methods modify the original Topic Maps source with new vocabularies. First, we show the translated results according to the existing three methods respectively. Then, the evaluation results are analyzed by the comparative items in Section 3.1.

4.1 Analysis on the Translation Results based on the previous three methods

First, we analyze the RDF statements generated by each method in order to help in understanding of the translation results. To increase objectivity, the statements by the existing methods are taken from the W3C survey. The Stanford method basically translates t-nodes (topic nodes) and a-nodes
Translating Topic Maps to RDF/RDF Schema for The Semantic Web

### composed-by association


_:puccini-tosca-assoc
rdfs:type tms:Association;
tms:associationTemplate _:composed-by.
```

### reified statement representing composer role


_:puccini-composer-role
rdfs:type rdfs:Statement;
rdfs:subject _:puccini-tosca-assoc;
rdfs:predicate tms:roleLabel;
rdfs:object _:puccini;
tms:roleLabel _:composer.
```

### reified statement representing work role


_:puccini-work-role
rdfs:type rdfs:Statement;
rdfs:subject _:puccini-tosca-assoc;
rdfs:predicate tms:roleLabel;
rdfs:object _:tosca;
tms:roleLabel _:work.
```

Figure 6: Translation result on the test data 1 based on the Stanford's method.

(association nodes) of Topic Maps into subjects and predicates of RDF. This method translates Topic Maps in the context of associations, resulting in changes of semantic structure from the original. Among the translated results, one corresponds to “composed-by” relationship is shown in Figure 6. The Moore method defines the relationship between topic and association role as a RDF statement and again generates the relationship between the defined RDF statements. In the process of translation, this method generates additional nodes, (Assoc:1, Assocmember:1, Assocmember:2) that do not exist in the original source model. Among the results, one corresponds to “Assocmember:1” relationship is shown in Figure 7. The Unibo method defines the relationships between topics as RDF statements. The Unibo method does not handle association roles and


# topic 1: puccini
_:puccini
<http://www.empolis.com/rdf/tmviewer#tm-topicname>
#:topic1.

#:topic1
<http://www.empolis.com/rdf/tmviewer#tm-topicnamevalue>
"Giacomo Puccini".

#:puccini
<http://www.empolis.com/rdf/tmviewer#tm-instanceof>
"http://psi.ontopia.net/music/#person".

# topic 3: composer
<http://www.empolis.com/rdf/tmviewer#tm-topicname>
#:topic3.

#:topic3
<http://www.empolis.com/rdf/tmviewer#tm-topicnamevalue>
"Composer".
<http://www.empolis.com/rdf/tmviewer#tm-subjindicator>
"http://psi.ontopia.net/music/#composer".

# association

`:assocmember-1
<http://www.empolis.com/rdf/tmviewer#tm-roledefiningtopic>
<http://psi.ontopia.net/music/#composer>.

`:assocmember-1
<http://www.empolis.com/rdf/tmviewer#tm-rollplayingtopic>
_:puccini.
```

Figure 7: Translation result based on the Moore’s method
generates additional nodes, such as *baggage*s. Among the translated RDF statements, the "composed-by" *association* is shown in Figure 8.

### 4.2 Evaluation and Discussion

**A. Completeness – The evaluation result on reversibility**

Figure 9 illustrates the experiment result on the reversibility of the proposed method and the existing methods. For test data 1, the results of the reversibility of the previous translation methods are borrowed from the W3C survey for reliable evaluation. According to the study, the Stanford and the Moore method do not handle occurrences, and the Unibo method missed subject Puccini, in the reverse translated results. However, the proposed method produced all constructs. Table 3 presents the reverse translated constructs (RveTMconst(m)) by each method on Test Data sets and Table 4 shows the experimental results on the reversibility.
B. Naturalness – The structural complexity evaluation result

Figure 10 presents the structural complexity of the translation results by the proposed method and the existing methods. As explained in Section 2.3, the proposed method creates the implicit semantics as well as the explicit semantics. None of the existing methods can handle the implicit semantics. In addition to, according to the W3C survey, the Moore method does not cover occurrences and the Stanford method omits occurrences, and names of typing topics. Therefore, in the RDF statements created by the translation methods, we only count the numbers of RDF statements required to express the same semantics. The W3C survey shows the evaluation results of the existing methods with test data 1. Therefore, we defined the estimating rules for measuring the number of translated RDF statements based on the results of the W3C survey. The defined estimating rules are defined in Definition 3. The compositions of the test data sets are described in Table 1. The numbers of the produced RDF statements by the translation methods and the calculated results according to the Definition 3 are shown in Table 5.

Table 3: The original source Topic Maps constructs and the reverse translated Topic Maps constructs

<table>
<thead>
<tr>
<th>TM constructs</th>
<th>Test Data 1</th>
<th>Test Data 2</th>
<th>Test Data 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nT nA nR nOCC</td>
<td>nT nA nR nOCC</td>
<td>nT nA nR nOCC</td>
</tr>
<tr>
<td><em>OrigTMconst</em></td>
<td>2 1 2 2</td>
<td>7 8 16 -</td>
<td>1,918 3,479 28 1,488</td>
</tr>
<tr>
<td><em>RveTMconst</em>(Stanford)</td>
<td>2 1 2 0</td>
<td>7 8 16 -</td>
<td>1,918 3,479 28 0</td>
</tr>
<tr>
<td><em>RveTMconst</em>(Moore)</td>
<td>2 1 2 0</td>
<td>7 8 16 -</td>
<td>1,918 3,479 28 0</td>
</tr>
<tr>
<td><em>RveTMconst</em>(Unibo)</td>
<td>1 0 2 2</td>
<td>0 0 16 -</td>
<td>1,488 0 28 1,488</td>
</tr>
<tr>
<td><em>RveTMconst</em>(Proposed)</td>
<td>2 1 2 2</td>
<td>7 8 16 -</td>
<td>1,918 3,479 28 1,488</td>
</tr>
</tbody>
</table>

Table 4: Experimental results for the Reversibility

<table>
<thead>
<tr>
<th>Reversibility</th>
<th>Test Data 1</th>
<th>Test Data 2</th>
<th>Test Data 3</th>
<th>E(Rve)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Rve</em>(Stanford)</td>
<td>0.71</td>
<td>1.00</td>
<td>0.78</td>
<td>0.83 (83%)</td>
</tr>
<tr>
<td><em>Rve</em>(Moore)</td>
<td>0.71</td>
<td>1.00</td>
<td>0.78</td>
<td>0.83 (83%)</td>
</tr>
<tr>
<td><em>Rve</em>(Unibo)</td>
<td>0.43</td>
<td>0.00</td>
<td>0.43</td>
<td>0.29 (29%)</td>
</tr>
<tr>
<td><em>Rve</em>(Proposed)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00 (100%)</td>
</tr>
</tbody>
</table>

• E(Rve) denotes a mean of the reversibility on the three test data sets.

Figure 10: The evaluation result for the structural complexity
Definition 3. The estimation rules for measuring translated RDF statements are defined as follows:

\[
\begin{align*}
C_x(Stanford) &= nA \times 10 + nT \times 12 + (nTT + nRT + nAT) \times 4 \\
C_x(Moore) &= (nT + nA + nTT) \times 3 + nA \times 5 \\
C_x(Unibo) &= nT \times 2 + nA \times 7 + (nTT + nAT) \times 3 \\
C_x(Proposed) &= nA \times 8 + nT + nTT + nAT
\end{align*}
\]

<table>
<thead>
<tr>
<th>Complexity</th>
<th>Test Data 1</th>
<th>Test Data 2</th>
<th>Test Data 3</th>
<th>E(Cx)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>nRDFstate</td>
<td>Cx(m)</td>
<td>nRDFstate</td>
<td>Cx(m)</td>
</tr>
<tr>
<td>Cx(Stanford)</td>
<td>36</td>
<td>2.77</td>
<td>189</td>
<td>2.78</td>
</tr>
<tr>
<td>Cx(Moore)</td>
<td>20</td>
<td>1.54</td>
<td>103</td>
<td>1.24</td>
</tr>
<tr>
<td>Cx(Unibo)</td>
<td>20</td>
<td>1.54</td>
<td>106</td>
<td>1.28</td>
</tr>
<tr>
<td>Cx(Proposed)</td>
<td>13</td>
<td>-</td>
<td>83</td>
<td>-</td>
</tr>
</tbody>
</table>

- E(Cx) denotes a mean of the structural complexities on the three test data sets.

Table 5: Estimated RDF statements and Experiment results on the structural complexity

5. CONCLUSION AND FUTURE WORK

We proposed an enhanced Topic Maps-to-RDF(S) translation method based on TRO (topics, association roles, and occurrences). As mentioned in Section 1, the proposed method also supports the object and semantic mapping level translation. Our method differs from other existing methods in some aspects.

First, the previous methods handle only the explicit semantics, but our method extracts the implicit semantics based on association roles as well as the explicit semantics. Second, according to the evaluation results in Section 4.2, our method regenerates the original source constructs without loss, but the existing methods lose some of original constructs. The Standford, the Moore, and the Unibo method showed loss rates of, 16%, 16%, and 77% respectively. Third, with regard to the structural complexity, the proposed method reduced the structural complexity. The structural complexities of the Standford, the Moore, and the Unibo method are on 2.39-fold, 1.3-fold, and 1.26-fold higher respectively when compared with our proposal.

The W3C survey discusses the many semantic mapping issues of ‘Identity’, ‘Names’, ‘Scope’, ‘Occurrences’, ‘Binary and Non-binary relations’, and so on. On the ‘Identity’, the proposed method translated both subject identifiers and subject locators of Topic Maps to URIs of RDFs (Rule 1). When considering ‘Name’ and ‘Scope’, our method treated a base name in Topic Maps as rdfs:label and scope as rdfs:range (Rule 2, Rule3, and Rule 4). The base name of Topic Maps and rdfs:label are both human-readable resource’s names. scope and rdfs:range are a context within which a statement is valid and a range of a property which limits the individuals that the property may have as its value respectively. Our method treated occurrences as a specialized type of associations whether these are internal or external occurrences and translated the relationships between subjects (topics) and occurrences into other RDF statements (Rule 4). Regarding the binary and the non-binary issues, our method handled only the binary relation. Topic Maps does not have directionality between topics, but RDF has. As shown in Figure 1, we assumed a bidirectional topic relationship, which allows topic to be subject or object during translation. In further study, we will deal with the translation issue associated with a non-binary relationship.
REFERENCES


BIOGRAPHICAL NOTES

Shinae Shin received the BS degree from the Kyonggi University in 1993, and the MS degrees from the Korea University, Seoul, Korea, in 1999, all in computer science. She is a PhD 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 BS degree in computer science from Kunsan National University, Korea, in 1997. He received his MS and PhD 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, USA in 2005. He has been a Professor in the Department of Informatics and Statistics, Kunsan National University, Korea since 2005. His research interests include Data Integration, Semantic Web, Semantic Sensor Network, Semantic GIS and Security.
Translating Topic Maps to RDF/RDF Schema for The Semantic Web

Doo-Kwon Baik received his BS degree in mathematics from Korea University, Seoul, Korea, in 1974. He received his MS and PhD degrees in Computer Science from Wayne State University in USA, 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 Modelling and Simulation.