

RSM and SLN: Transformation, Normalization and Cooperation*

Erlin Yao^{1,2}, Yunpeng Xing^{1,2,3}, Jie Liu^{1,2}, and Xiaoping Sun^{1,2}

¹ Institute of Computing Technology, Chinese Academy of Sciences,
Beijing, China

alin.yao@kg.ict.ac.cn

² Graduate School of Chinese Academy of Sciences,
Beijing, China

³ Hunan Knowledge Grid Lab, Hunan University of Science and Technology,
Hunan, China

Abstract. The Resource Space Model RSM and Semantic Link Network SLN are models of future interconnection environment, and the combination of them forms a rich semantic layer for the next-generation web. This paper investigates the combination between RSM and SLN. The methods of the transformation between RSM and SLN are proposed. From the perspective of transformation, the correlations between the normal forms of RSM and SLN are investigated. The investigations show that although RSM and SLN are independently proposed, they are well-cooperated models for the future Semantic Web and Knowledge Web.

1 Introduction

The current web is based on the HTTP and HTML protocols as well as the client/server infrastructure. It enables people to share web information to some extent in form of using browsers to retrieve web pages. It could not meet the needs of the rapid expansion of the Internet applications like e-business. There also exist issues like accuracy and efficiency from the viewpoint of information retrieval. To overcome the shortcomings of the current web, people have made efforts towards the next-generation web. The Semantic Web, the Web Service and the Grid are three new areas, where many leading research institutes and companies have been involved in.

The Resource Space Model RSM is a semantic data model for uniformly, normally and effectively specifying and managing resources. RSM is to organize versatile Web resources according to semantic normal forms [6, 8, 14]. Given a set of coordinates of the space can uniquely determine one or a set of resources. The model includes normal forms based on the orthogonal classification semantics and relevant criteria and methods.

The motivation of Semantic Web is obviously right, however, easy-to-use and easy-to-build are two important criteria that determine the success or failure of new

* This work was supported by the National Basic Research Program of China (973 project no.2003CB317000) and the National Science Foundation of China (Grants 60273020 and 70271007).

techniques. We have proposed the Semantic Link Network SLN as a model to facilitate the Semantic Web [5, 7]. The links among resources (documents, images, and concepts etc) are not simply hyperlinks but semantic links, which support certain semantic reasoning. The semantic link network is the semantic “map” of distributed versatile resources to enable people to publish, manage and browse resources according to the map.

The combination of RSM and SLN forms a rich semantic layer for the next generation Web [9, 11, 12, 15]. Users make and use the orthogonal semantic space and the semantic link network according to their cognitive spaces. The orthogonal semantic space can help users focus their operation destination. The SLN reflects the semantic relationships between versatile resources [13, 16].

In this paper, the integration of RSM and SLN are discussed. First given a semantic link network, the corresponding high-level resource space is constructed, and the correlations between their normal forms are discussed. Then on the contrary, we give a method to construct a semantic link network that is semantically equivalent to a given resource space. In this perspective, the corresponding relationships between the normal forms of RSM and SLN are discussed.

2 The Transformation from SLN to RSM and the Correlations

Each resource defined in the RSM has a mapping image in the SLN. In this section, given a semantic link network, the corresponding high-level resource space is constructed. And the correlations between their normal forms are discussed.

The semantic similar nodes in $SLN(N, SL)$ can be looked as a point in the high-level resource space, so we can get the corresponding high-level resource space from a semantic link network as follows:

For a semantic link network $SLN(N, SL)$, let $SLN = \{C_1, C_2, \dots, C_m\}$, where C_i is a strong connected component of SLN and $C_i \cap C_j = \Phi$, $1 \leq i, j \leq m$. Then $N(SLN) = \cup N(C_i)$, where $N(C_i)$ denotes the nodes of C_i . Suppose we can describe SLN from such aspects in semantics as X_1, X_2, \dots, X_n , then all C_i can be represented as $C_i(x_1, x_2, \dots, x_n)$, where x_k is the value of C_k in the semantic of X_k . If we view every $C_i(x_1, x_2, \dots, x_n)$ as a point $pi(x_1, x_2, \dots, x_n)$ in a resource space, then we get the high-level resource space $RS(X_1, X_2, \dots, X_n)$.

Well normalized data makes programming (relatively) easy, and works very well in multi-platform, enterprise wide environments [1, 2]. The normal forms are to guarantee the correctness and effectiveness of operations [3, 4].

Then the correlations between the normal forms of the semantic link network and the resource space will be discussed.

Corollary 1 (INF): The first-normal-form of a semantic link network SLN is also the first-normal-form of the corresponding resource space RS .

Proof: If SLN is INF, then there does not exist semantic-equivalent nodes in it. From the definition of strong connected component, there can't exist the same strong component in SLN . Suppose RS is the high-level resource space, then there can not be the same points in it. So we can get that there does not exist name duplication between coordinates at any axis in RS . Then RS is also the first-normal-form.

Corollary 2 (2NF): The second-normal-form of a semantic link network SLN is also the second-normal-form of the corresponding resource space RS .

Proof: If SLN is 2NF, then there does not exist inconsistent semantic links and duplicate semantic links between the same pair of nodes, this guarantees that the strong components of SLN are got correctly in semantics. Suppose $\{C_1, C_2, \dots, C_m\}$ is the set of strong components of SLN and RS is the high-level resource space, then $RS = \{p_1, p_2, \dots, p_m\}$, where p_i is corresponding to C_i , $1 \leq i \leq m$. If RS is not the second-normal-form, then there exist some points p_i and p_j are not independent from each other, i.e., $R(C_i) \cap R(C_j) \neq \Phi$, which means $C_i \cap C_j \neq \Phi$. This is not consistent with that C_i and C_j are different strong components in SLN , so the RS is also the second-normal-form.

Corollary 3 (3NF): The third-normal-form of a semantic link network SLN is also the third-normal-form of the corresponding resource space RS .

Proof: If SLN is 3NF, then there does not exist isolated nodes (accessible from each other), therefore all the strong components of SLN are accessible from each other. Suppose RS is the high-level resource space, we can get that any of its point are reachable from the others, then every axis X_i can represent all the resources in RS , which is equivalent to that any two axes of RS are orthogonal with each other [10], so RS is also the third-normal-form.

The above three corollaries tell that although we have defined the normal forms of the RSM and SLN respectively, they have the common properties in solving the redundancy and inconsistency, are the same in some sense.

3 The Transformation from RSM to SLN and the Correlations

In the above section, given a semantic link network, the high-level resource space is constructed and the correlations between them are discussed. But much semantics is lost in the procedure of from SLN to RSM. Then on the contrary, we will compare the RSM with the SLN in the perspective of constructing a semantic link network that is semantically equivalent to a given resource space.

Given a resource space $RS (P, E)$, we can construct a $SLN (N, SL)$ that is semantically equivalent to it as follows:

First let $N (SLN) = P (RS)$, then for any two nodes n_1 and n_2 in the SLN , we can refer them as $p_1(x_1, x_2, \dots, x_n)$ and $p_2(y_1, y_2, \dots, y_n)$ in RS respectively. Then we define $SL (n_1, n_2) = \{ SL(x_i, y_i) \mid 1 \leq i \leq n \}$, where

$$SL(x_i, y_i) = \begin{cases} \text{Im } p_i, & \text{if } \langle x_i, y_i \rangle \in E(RS) \text{ or } x_i = y_i; \\ \Phi, & \text{otherwise.} \end{cases}$$

In the resource space $RS (P, E)$, if $\langle x_i, y_i \rangle \in E(RS)$ or $x_i = y_i$, then $R(x_i) \supseteq R(y_i)$, i.e., x_i implies y_i . So here we can look every Imp_i as an instance of the ‘Implication link’ [5], $1 \leq i \leq n$. Then we can get $SL (SLN)$ from $SL (SLN) = \cup_{n_1, n_2 \in N} SL(n_1, n_2)$.

Because the semantic links are directed, $SL(x_i, y_i) \neq SL(y_i, x_i)$ in most cases. If $x_i = y_i$, then $SL(x_i, y_i) = SL(y_i, x_i) = Imp_i$. And on the contrary, if $SL(x_i, y_i) = Imp_i$ and

$SL(y_i, x_i) = Imp_i$, then $x_i = y_i$. It is obvious that the link Imp_i is transitive. Note that for any node $n(x_1, x_2, \dots, x_n)$ in the resulting SLN (N, SL) , $R(x_i) = \cup\{R(y_i) | SL(x_i, y_i) = Imp_i\}$. Then from the steps of constructing, it is clear that the resulting $SLN(N, SL)$ is semantically equivalent to the resource space $RS(P, E)$.

In a resource space RS , an axis with hierarchical coordinates can be transformed into an axis with flat coordinates if only the leaf nodes of each hierarchy are considered. So we focus on the flat cases in this section.

The corresponding semantic link network SLN of RS becomes more regular when RS is in the flat case. In this case, for any two points $p_1(x_1, x_2, \dots, x_n)$ and $p_2(y_1, y_2, \dots, y_n)$ in RS , we have $x_i = y_i$ or $R(x_i) \cap R(y_i) = \Phi$, $1 \leq i \leq n$, i.e., both x_i implies y_i and y_i implies x_i hold or neither hold. Correspondingly, for any two nodes $n_1(x_1, x_2, \dots, x_n)$ and $n_2(y_1, y_2, \dots, y_n)$ in the SLN , $SL(x_i, y_i) = SL(y_i, x_i) = Imp_i$ or $SL(x_i, y_i) = SL(y_i, x_i) = \Phi$, i.e., there are no semantic links between x_i and y_i .

Given a resource space RS , there is a SLN that is semantically equivalent to it, so we can compare the normal forms of RS and the corresponding SLN to study the correlations between the normal forms of RSM and SLN.

Corollary 4 (1NF): The first-normal-form of a resource space RS is also the first-normal-form of the corresponding SLN .

Proof: If RS is 1NF, then there does not exist name duplication between coordinates at any axis, so there can't be the same points in RS . Suppose SLN is the corresponding semantic link network of RS , from $N(SLN) = P(RS)$, we can get that there does not exist semantic-equivalent nodes in SLN , so SLN is also the first normal form.

Corollary 5 (2NF): The second-normal-form of a resource space RS is also the second-normal-form of the corresponding SLN .

Proof: If RS is 2NF, then for any axis, any two coordinates are independent from each other. Suppose SLN is the corresponding semantic link network of RS , then there does not exist semantic-equivalent nodes in SLN . And for any pair of nodes $n_1(x_1, x_2, \dots, x_n)$ and $n_2(y_1, y_2, \dots, y_n)$ in the SLN , the semantic links between them fall into the two cases: $SL(x_i, y_i) = SL(y_i, x_i) = Imp_i$ or $SL(x_i, y_i) = SL(y_i, x_i) = \Phi$, so there does not exist inconsistent semantic links and duplicate semantic links between n_1 and n_2 , then we say that the SLN is the second normal form.

Corollary 6 (3NF): The third-normal-form of a resource space RS is also the third-normal-form of the corresponding SLN .

Proof: If RS is 3NF, then any two axes of it are orthogonal with each other, which is equivalent to $R(X_1) = R(X_2) = \dots = R(X_n)$, i.e., every axis X_i can represent all the resources in RS [10], this guarantees that any points in the form of $\{p(x_1, x_2, \dots, x_n) | x_i \in X_i\}$ are meaningful. Suppose SLN is the corresponding semantic link network of RS , from $N(SLN) = P(RS)$, we can get that any nodes in SLN in the form of $\{n(x_1, x_2, \dots, x_n) | x_i \in X_i\}$ are also meaningful. Then for any two nodes $n_1(x_1, x_2, \dots, x_n)$ and $n_2(y_1, y_2, \dots, y_n)$ in the SLN , if there exists some i , $1 \leq i \leq n$, such that $x_i = y_i$, then $SL(x_i, y_i) = SL(y_i, x_i) = Imp_i$, then n_1 and n_2 can be accessed from each other. Else if $x_i \neq y_i$, for $1 \leq i \leq n$ hold, let $n_3 = n_3(x_1, x_2, \dots, x_{n-1}, y_n)$, from $x_1 \neq y_1$ and $x_n \neq y_n$, we can get $n_3 \neq n_1$ and $n_3 \neq n_2$. Then there exists semantic links $n_1 - Imp_1 \rightarrow n_3$, and $n_3 - Imp_n \rightarrow n_2$, so n_1 and n_2 can also be

accessed from each other by the chain $n_1\text{-Imp}_1 \rightarrow n_3\text{-Imp}_n \rightarrow n_2$. So any two nodes in the *SLN* are accessible from each other, i.e., the *SLN* is the third-normal-form.

The above three corollaries indicate that although we have defined the normal forms of the RSM and *SLN* independently, they have the common properties in solving the redundancy and inconsistency, and are the same in some sense.

4 Conclusion

The investigation of the relationship between RSM and *SLN* indicates that although RSM and *SLN* are proposed respectively, they are well-cooperated models for the future semantic and knowledge networking environment [15, 16]. Based on the theory of the RSM [6, 8, 10, 14] and *SLN* [5, 13], this paper proposes a method of constructing the corresponding high-level resource space from a semantic link network, investigates the correlations between the normal forms of the *SLN* and RSM, and provides a method of constructing a *SLN* that is semantically equivalent to a given resource space.

References

1. Codd, E. F.: A Relational Model of Data for Large Shared Data Banks, *Communications of the ACM*, 13 (6) (1970) 377-387
2. Codd, E. F.: Normalized Database Structure: A Brief Tutorial, *ACM SIG-FIDET Workshop on Data Description, Access, and Control*, San Diego, California, Nov., 1971
3. Mok, W. Y.: A Comparative Study of Various Nested Normal Forms, *IEEE Transactions on Knowledge and Data Engineering*, 14 (2) (2002) 369-385
4. William, K.: A Simple Guide to Five Normal Forms in Relational Database Theory, *Communications of the ACM*, 26 (2) (1983) 120-125
5. Zhuge, H.: Active e-Document Framework ADF: Model and Platform, *Information and Management*, 41 (1) (2003) 87-97
6. Zhuge, H.: Resource Space Model, Its Design Method and Applications, *Journal of Systems and Software*, 72 (1) (2004) 71-81
7. Zhuge, H.: *The Knowledge Grid*, World Scientific, 2004
8. Zhuge, H.: Resource Space Grid: Model, Method and Platform, *Concurrency and Computation: Practice and Experience*, 16 (14) (2004) 1385-1413
9. Zhuge, H.: China's E-Science Knowledge Grid Environment, *IEEE Intelligent Systems*, 19 (1) (2004) 13-17
10. Zhuge, H., Yao, E., Xing, Y. and Liu, J.: Extended Normal Form Theory of Resource Space Model, *Future Generation Computer Systems*, 21 (2005) 189-198
11. Zhuge, H.: Semantic Grid: Scientific Issues, Infrastructure, and Methodology, *Communications of the ACM*, 48 (4) (2005) 117-119
12. Zhuge, H.: The Future Interconnection Environment, *IEEE Computer*, 38 (4) (2005) 27-33
13. Zhuge, H., Liu, J., Feng, L., Sun, X. and He, C.: Query Routing in a Peer-to-Peer Semantic Link Network, *Computational Intelligence*, 21 (2) (2005) 197-216
14. Zhuge, H. and Xing, Y.: Integrity Theory for Resource Space Model and Its Application, *Keynote at WAIM2005, LNCS 3739*, pp.8-24, 2005
15. Zhuge, H. and Shi, X.: Toward the Eco-Grid: A Harmoniously Evolved Interconnection Environment, *Communications of the ACM*, 47 (9) (2004) 79-83
16. Zhuge, H.: Exploring an Epidemic in an E-Science Environment, *Communications of the ACM*, 48 (9) (2005) 109-114