Available online at www.sciencedirect.com
SCIENCE DIRECT
Future Generation Computer Systems 21 (2005) 189–198
FGCS
www.elsevier.com/locate/future

Extended resource space model☆

Hai Zhuge[a,b,*], Erlin Yao[b,c], Yunpeng Xing[b,c], Jie Liu[b,c]

[a] *Hunan Knowledge Grid Lab, Hunan University of Science and Technology, Hunan, China*
[b] *China Knowledge Grid Research Group, Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China*
[c] *Graduate School of Chinese Academy of Sciences, Beijing, China*

Abstract

The resource space model (RSM) is a semantic data model based on orthogonal classification semantics for efficiently managing various resources in the future interconnection environment. This paper extends the RSM in theory by formalizing the resource space, investigating its characteristics from the perspective of set theory, defining the resource space schema and developing its normal forms. The topological space properties of the resource space are presented based on the definition of a distance in the space and the construction of a quotient space structure. The proposed theory ensures the RSM to correctly and efficiently specify and manage resources.

Keywords: Normal forms; Resource space model; Resource space schema; Topological space

1. Introduction

Database theories, especially the relational data model and theory have gained a great success in the past 40 years [2–4,11–13]. But, these data models are incapable of effectively managing heterogeneous, distributed and ocean resources in an open and dynamic interconnection environment [14].

The World Wide Web creates a convenient human-oriented information service mode based on the client/server architecture, HTML-based Web page setting, and hyperlink-based information navigation and search. But, hyperlinks and Web pages do not reflect enough semantics to support complex and intelligent applications. The Semantic Web is proposed to create rich and machine-understandable semantics by using markup languages and ontology-relevant mechanisms [1,5,9]. However, the Semantic Web focuses on the semantic representation issue. To achieve the efficiency and effectiveness of resource management, the future interconnection environment still requires semantic rich data models.

The efforts towards the future interconnection environment include Grid (http://www.gridforum.org) [6] and Peer-to-Peer computing [10], which focus on resource sharing in a tightly coupled environment and

☆ Research work was supported by the National Science Foundation of China (Grants 60273020 and 70271007) and the National Grand Fundamental Research 973 Program of China (2003CB316900).
* Corresponding author. Fax: +86 1062562703.
E-mail address: zhuge@ict.ac.cn (H. Zhuge).

0167-739X/$ – see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.future.2004.09.016

a loosely coupled network environment, respectively. Initially, Grid computing does not solve semantic issues. With the emerging of OGSA (Open Grid Service Architecture) [7], the semantic issue becomes one of the kernel issues of Grid services. Currently, researchers have paid attention to the effect of introducing semantics into Peer-to-Peer computing to support high-level intelligent applications and have made attempt to incorporate Grid and Peer-to-Peer to obtain more powerful computing models.

The resource space model (RSM) is a semantic data model for uniformly, normally and effectively specifying and managing resources. Its theory basis is three normal forms based on orthogonal classification semantics and relevant operations [14]. The RSM is based on the following methods and viewpoints:

(1) Uniform resource abstraction;
(2) Orthogonal semantic partition of resources;
(3) Uniform resource operations; and
(4) Universal resource view.

This paper first formally defines the resource space from the perspective of set theory, and then proposes the resource space schema, which represents the inherent characteristics of the resource space and reflects its static and stable properties. The schema helps us on the logic design of the resource space. We extend the normal forms to specify a good design of the resource space schema and guide the logical design of resource spaces. By defining a distance and constructing a quotient space structure on the resource space, we obtain some topological space properties of the RSM.

2. Formalization of resource space

A Resource Space is an n-dimensional space where every point uniquely determines one resource or a set of related resources. A resource space can be represented as RS($X_1, X_2, \ldots, X_n$) or RS in simple, where RS is the name of the resource space and X_i is the name of an axis. We use |RS| to denote the number of dimensions of the RS, i.e., $|\text{RS}| = n$. $X_i = \{C_{i1}, C_{i2}, \ldots, C_{ip}\}$ represents an axis with its coordinates. C denotes the coordinate name in form of a noun or a noun phrase. Any name corresponds to a formal or an informal semantic definition in its domain ontology [15].

Suppose O is a domain terminology set like the data dictionary, and there exists a mapping from O onto domain ontology that explains the semantics. The formal definition of RS is given as follows.

Definition 1. Let $S = 2^O$ be the power set of O. The resource space defined on O can be represented as RS($X_1, X_2, \ldots, X_n$), where RS is the name of the resource space and $X_i = \{C_{i1}, C_{i2}, \ldots, C_{ip}\}$ is an axis, $1 \le i \le n$, C_{ij} is the root of the hierarchical structure of coordinates on X_i, $C_{ij} = \{\langle V_{ij}, E_{ij}\rangle | V_{ij} \in S, E_{ij} = \{\langle v_t, v_s\rangle | v_t, v_s \in V_{ij}, R(v_t) \supseteq R(v_s)\}\}$, $1 \le j \le p$, where $R(v)$ denotes a class of resources represented by v. Every point in RS is an element of the Cartesian product $X_1 \times X_2 \times \cdots \times X_n$, denoted as $p(x_1, x_2, \ldots, x_n)$.

Tuples in the relational data model reflect the attributes of relevant entity, not the semantics [8]. In the RSM, x_i in a point $p(x_1, x_2, \ldots, x_n)$ reflects the partition semantics. Resources represented by a point $(x_1, x_2, \ldots, x_n) \in$ RS can be represented as $R(p(x_1, x_2, \ldots, x_n)) = R(x_1) \cap R(x_2) \cap \cdots \cap R(x_n)$, where $R(x_i)$ denotes a class of resources represented by x_i, $1 \le i \le n$.

3. Resource space schema and normal forms

The resource space schema is the formal description of the resource space. The logic design of a resource space is to specify the resource space schema including the definition of axes and coordinates.

3.1. Definition of the resource space schema

Domain applications determine that resources in a resource space schema must satisfy some given integrity constraints. The resource space schema should describe all these integrity constraints, so the resource space schema is defined as follows.

Definition 2. A resource space schema is a 5-tuple: RS$\langle A, C, S, \text{dom}\rangle$, where:

(1) RS is the resource space name, the symbolized semantics of the resource space;
(2) $A = \{X_i | 1 \le i \le n\}$ is the set of axes;
(3) $C = \{C_{ij} | C_{ij} \in X_i, 1 \le i \le n\}$ is the set of coordinates;

(4) S is the power set of the domain ontology O;
(5) dom is the mapping from the axes A and coordinates C to S, dom: $A \times C \rightarrow S$, for any axis $X_i = \{C_{i1}, C_{i2}, \ldots, C_{ip}\}$, $\text{dom}(X_i, C_{ij}) = V_{ij}$, $V_{ij} \in S$, where $1 \leq i \leq n$ and $1 \leq j \leq p$.

In applications, (4) and (5) should be known before the design of the resource space schema, so the resource space schema can be simplified as a 3-tuple: $\text{RS}\langle A, C\rangle$.

The resource space schema is static and stable, but the resource space can be dynamic due to the resource operations on the resource space. The design of a resource space refers to the design of its schema.

In the following we study the normal forms to guarantee a good schema.

3.2. Normal forms of the resource space schema

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.

Definition 3. The first-normal-form of a resource space is a resource space, and there does not exist name duplication between coordinates at any axis.

Definition 4. The second-normal-form of a resource space is a first-normal-form, and for any axis, any two coordinates are independent from each other.

The definition of the third-normal-form is based on the definition of two relations between the axes, i.e., 'fine classification' and 'orthogonal' [15].

Definition 5. Let $X = \{C_1, C_2, \ldots, C_n\}$ be an axis and C'_i be an coordinate at another axis X', we say X fine classifies C'_i (denoted as C'_i/X) if and only if:

(1) $(R(C_j) \cap R(C'_i)) \cap (R(C_k) \cap R(C'_i)) = \Phi$, $1 \leq j < k \leq n$; and
(2) $(R(C_1) \cap R(C'_i)) \cup (R(C_2) \cap R(C'_i)) \cup \cdots \cup (R(C_n) \cap R(C'_i)) = R(C'_i)$ hold.

As the result of fine classification, $R(C'_i)$ is classified into n categories: $R(C'_i/X) = \{R(C_1) \cap R(C'_i), R(C_2) \cap R(C'_i), \ldots, R(C_n) \cap R(C'_i)\}$.

Definition 6. For two axes $X = \{C_1, C_2, \ldots, C_n\}$ and $X' = \{C'_1, C'_2, \ldots, C'_m\}$, we say X fine classifies X' (denoted as X'/X) if and only if X fine classifies C'_1, C'_2, $\ldots$, C'_m.

Definition 7. Two axes X and X' are called orthogonal with each other (denoted as $X \perp X'$) if X fine classifies X' and vice versa, i.e., both X/X' and X'/X hold.

Based on the above definitions, we can give the definition of third-normal-form as follows:

Definition 8. The third-normal-form of a resource space is a second-normal-form, and any two axes of it are orthogonal with each other.

In the following, we give the equivalent definitions from the perspective of set theory.

3.3. Equivalent definitions and corresponding lemmas

For the resource space $\text{RS}(X_1, X_2, \ldots, X_n)$, we use $R(X_i)$ to denote resources represented by axis X_i, where $X_i = \{C_{i1}, C_{i2}, \ldots, C_{ip}\}$, $1 \leq i \leq n$. $R(X_i) = R(C_{i1}) \cup R(C_{i2}) \cup \cdots \cup R(C_{ip})$. In the following discussion, we assume that RS is second-normal-form.

First we give the equivalent definition of 'fine classification'.

Lemma 1. *For two axes* $X_i = \{C_{i1}, C_{i2}, \ldots, C_{ip}\}$ *and* $X_j = \{C_{j1}, C_{j2}, \ldots, C_{jq}\}$ *in the resource space* RS, $X_j/X_i \Leftrightarrow R(X_j) \subseteq R(X_i)$ *holds.*

Proof. The proof includes the following two steps:

(1) If X_j/X_i holds, then according to our definition of 'fine classification', we have:

$$\begin{aligned} R(C_{jk}) &= (R(C_{jk}) \cap R(C_{i1})) \cup (R(C_{jk}) \cap R(C_{i2})) \\ &\quad \cup \cdots \cup (R(C_{jk}) \cap R(C_{ip})) \\ &= R(C_{jk}) \cap (R(C_{i1}) \cup R(C_{i2}) \\ &\quad \cup \cdots \cup R(C_{ip})) \end{aligned}$$

then we have $R(C_{jk}) \subseteq (R(C_{i1}) \cup R(C_{i2}) \cup \cdots \cup R(C_{ip})) = R(X_i)$, $1 \leq k \leq q$. So $(R(C_{j1}) \cup R(C_{j2}) \cup \cdots \cup R(C_{jq})) \subseteq R(X_i)$ holds, i.e., $R(X_j) \subseteq R(X_i)$.

(2) Suppose $R(X_j) \subseteq R(X_i)$. For any C_{jk}, C_{it}, C_{is}, $1 \leq k \leq q$, and $1 \leq t \neq s \leq p$.

Since RS is a second-normal-form, so $R(C_{it}) \cap R(C_{is}) = \Phi$.
Then we have:

(i) $(R(C_{jk}) \cap R(C_{it})) \cap (R(C_{jk}) \cap R(C_{is})) = R(C_{jk}) \cap (R(C_{it}) \cap R(C_{is})) = R(C_{jk}) \cap \Phi = \Phi$;
(ii) $(R(C_{jk}) \cap R(C_{i1})) \cup (R(C_{jk}) \cap R(C_{i2})) \cup \cdots \cup (R(C_{jk}) \cap R(C_{ip})) = R(C_{jk}) \cap (R(C_{i1}) \cup R(C_{i2}) \cup \cdots \cup R(C_{ip})) = R(C_{jk}) \cap R(X_i)$.
Because $R(C_{jk}) \subseteq R(X_j) \subseteq R(X_i)$, we get $R(C_{jk}) \cap R(X_i) = R(C_{jk})$.
So $(R(C_{jk}) \cap R(C_{i1})) \cup (R(C_{jk}) \cap R(C_{i2})) \cup \cdots \cup (R(C_{jk}) \cap R(C_{ip})) = R(C_{jk})$.
From the definition of the 'fine classification', we get C_{jk}/X_i, $1 \le k \le q$. So X_j/X_i holds.
According to (1) and (2), we reach $X_j/X_i \Leftrightarrow R(X_j) \subseteq R(X_i)$. □

Lemma 1 can be an equivalent definition of the 'fine classification'.

Definition 9. For two axes $X_i = \{C_{i1}, C_{i2}, \ldots, C_{ip}\}$ and $X_j = \{C_{j1}, C_{j2}, \ldots, C_{jq}\}$ in resource space RS, we call X_j/X_i if $R(X_j) \subseteq R(X_i)$ holds.

According to Definition 9, we can get the equivalent definition of the orthogonality.

Lemma 2. *For two axes $X_i = \{C_{i1}, C_{i2}, \ldots, C_{ip}\}$ and $X_j = \{C_{j1}, C_{j2}, \ldots, C_{jq}\}$ in the resource space* RS, *$X_j \perp X_i \Leftrightarrow R(X_j) = R(X_i)$ holds.*

Proof.

(1) If $X_j \perp X_i$ holds, then we have X_j/X_i and X_i/X_j according to the definition of orthogonality. According to Lemma 1 we have $R(X_j) \subseteq R(X_i)$ and $R(X_i) \subseteq R(X_j)$. So $R(X_j) = R(X_i)$ holds.
(2) If $R(X_j) = R(X_i)$ holds, then $R(X_j) \subseteq R(X_i)$ and $R(X_i) \subseteq R(X_j)$ hold. According to Lemma 1, we have X_j/X_i and X_i/X_j. That means $X_j \perp X_i$.

According to (1) and (2), we have $X_j \perp X_i \Leftrightarrow R(X_j) = R(X_i)$. □

Lemma 2 can be an equivalent definition of orthogonality.

Definition 10. For two axes $X_i = \{C_{i1}, C_{i2}, \ldots, C_{ip}\}$ and $X_j = \{C_{j1}, C_{j2}, \ldots, C_{jq}\}$ in resource space RS, we call $X_j \perp X_i$ if $R(X_j) = R(X_i)$ holds.

It is clear that $X_i \perp X_j \Leftrightarrow X_j \perp X_i$, which means the symmetry of the orthogonal relation '⊥' holds.

According to the above theory, we can give a new proof of the transitivity of 'fine classification' and 'orthogonal' relations.

Theorem 1. *'Fine classification' and 'orthogonal' relations are transitive.*

Proof. According to Lemmas 1 and 2, we can get $X_j/X_i \Leftrightarrow R(X_j) \subseteq R(X_i)$ and $X_j \perp X_i \Leftrightarrow R(X_j) = R(X_i)$. Then according to the transitivity of the set relations '⊆' and '=', the transitivity of 'fine classification' and 'orthogonal' relations is clear. □

Now we can give the equivalent definition of 'third-normal-form'.

Theorem 2. *For resource space* RS(X_1, X_2, ..., X_n), RS *is a third-normal-form* $\Leftrightarrow R(X_1) = R(X_2) = \cdots = R(X_n)$, *i.e., every axis X_i can represent all the resources in* RS.

Proof. (1) If RS is a third-normal-form, then $X_1 \perp X_2 \perp \cdots \perp X_n$. According to Lemma 2, $R(X_1) = R(X_2) = \cdots = R(X_n)$ holds. (2) If $R(X_1) = R(X_2) = \cdots = R(X_n)$ holds, we get $X_1 \perp X_2 \perp \cdots \perp X_n$ from Lemma 2, then, according to the transitivity and symmetry of '⊥', we have: for any two axes X_i and X_j in RS, $X_i \perp X_j$ holds. That means RS is third-normal-form. According to (1) and (2), we have: RS is third-normal-form $\Leftrightarrow R(X_1) = R(X_2) = \cdots = R(X_n)$. □

Theorem 2 can be rewritten as an equivalent definition of the third-normal-form.

Definition 11. For the resource space RS(X_1, X_2, ..., X_n), we call RS is a third-normal-form if $R(X_1) = R(X_2) = \cdots = R(X_n)$ holds, i.e., every axis X_i can represent all the resources in RS.

Besides the three normal forms, we can still define other normal forms of the resource space schema for the convenience of resource partition and operations.

3.4. Extension of normal forms—2^+-normal-form and fourth-normal-form

Definition 12. (2^+NF). A resource space RS(X_1, X_2, ..., X_n) is a 2^+-normal-form (2^+NF), if it is a second-normal-form and satisfies: X_2/X_1, X_3/X_2, ..., X_n/X_{n-1}.

Definition 12 means $R(X_1) \supseteq R(X_2) \supseteq \cdots \supseteq R(X_n)$ according to Lemma 1.

If RS is a 2^+-normal-form, then according to the transitivity of fine classification '/', we have: for every two axes X_i and X_j, $1 \le i \ne j \le n$, either X_i/X_j or X_j/X_i holds. So '/' is a full order on the set $\{X_1, X_2, \ldots, X_n\}$. On the other hand, it is obvious that if '/' constitutes a full order on the axes of RS, then RS is a 2^+-normal-form.

In the following, we discuss the properties of the 2^+-normal-form under the operations of resource space as presented in [14,15].

Corollary 1 (Join). *For two resource space* RS_1 *and* RS_2, *let* $RS_1 \cdot RS_2 \Rightarrow RS$, *if both* RS_1 *and* RS_2 *are* 2^+-*normal-form, then* RS *is either* 2^+*-normal-form or not.*

Proof. Suppose $RS_1 = \{X_1, X_2\}$ and $RS_2 = \{Y_1, Y_2\}$, where X_i and Y_i are axes and satisfy X_2/X_1, Y_2/Y_1 and $X_2 = Y_2$. Then, we can join RS_1 and RS_2 together. Let $RS_1 \cdot RS_2 \Rightarrow RS$, so $RS = \{X_1, X_2, Y_1\}$.

(1) If either $R(X_1) \subseteq R(Y_1)$ or $R(Y_1) \subseteq R(X_1)$, correspondingly, we have: either $R(X_2) \subseteq R(X_1) \subseteq R(Y_1)$ or $R(X_2) \subseteq R(Y_1) \subseteq R(X_1)$ holds. According to Definition 12, we have: RS is a 2^+-normal-form.

(2) Else if both $R(X_1) \subseteq R(Y_1)$ and $R(Y_1) \subseteq R(X_1)$ are not satisfied, then both Y_1/X_1 and X_1/Y_1 do not hold. Since '/' is a full order on the axes of RS if RS is a 2^+-normal-form, so RS is not a 2^+-normal-form.

According to (1) and (2), we have RS is either a 2^+-normal-form or not. □

Corollary 1 tells us that 2^+-normal-form does not keep under the join operation. But if we add some conditions, 2^+-normal-form keeps under the join operation.

Corollary 2 (Join). *Let* $RS_1 = \{X_1, X_2, \ldots, X_n\}$ *and* $RS_2 = \{Y_1, Y_2, \ldots, Y_m\}$ *be two* 2^+*-normal-form resource spaces, and* $RS_1 \cdot RS_2 \Rightarrow RS$. *If* $Y_1 = X_n$ *or* $X_1 = Y_m$ *holds, then* RS *is a* 2^+*-normal-form.*

Proof. (1) If $X_1 = Y_m$ holds, according to $RS_1 \cdot RS_2 \Rightarrow RS$, we have $RS = \{Y_1, Y_2, \ldots, Y_{m-1}, X_1, X_2, \ldots, X_n\}$. Since RS_1 and RS_2 are 2^+-normal-form, we have $X_n/X_{n-1}/\cdots/X_2/X_1$ and $Y_m/Y_{m-1}/\cdots/Y_2/Y_1$, then, from the transitivity of '/' we have $X_n/X_{n-1}/\cdots/X_2/X_1 = Y_m/Y_{m-1}/\cdots/Y_2/Y_1$. According to Definition 12, RS is a 2^+-normal-form. (2) If $Y_1 = X_n$ holds, RS is also a 2^+-normal-form according to the same reason as (1). □

Corollary 3 (Disjoin). *If* $RS \Rightarrow RS_1 \cdot RS_2$, *and* RS *is a* 2^+*-normal-form, then* RS_1 *and* RS_2 *are* 2^+*-normal-form.*

Proof. Suppose $RS = \{X_1, X_2, \ldots, X_n\}$, since RS is a 2^+-normal-form, then '/' is a full order on it. Because $RS \Rightarrow RS_1 \cdot RS_2$, then all the axes of RS_1 constitutes a subset of RS, so '/' is also a full order on the axes of RS_1, then we have: RS_1 is a 2^+-normal-form. For the same reason, RS_2 is also a 2^+-normal-form. □

Corollary 3 tells that 2^+-normal-form keeps under the operation Disjoin.

According to the definition of Join and Disjoin operations, we can get $RS_1 \cdot RS_2 \Rightarrow RS$ if and only if $RS \Rightarrow RS_1 \cdot RS_2$. Then, according to Corollary 3 we can give an equivalent corollary as follows:

Corollary 4 (Join). *For two resource spaces* RS_1 *and* RS_2, *let* $RS_1 \cdot RS_2 \Rightarrow RS$, *if* RS_1 *or* RS_2 *is not* 2^+*-normal-form, then* RS *is not a* 2^+*-normal-form.*

From the above corollaries, we can get the following corollary:

Corollary 5 (Disjoin). *If* $RS \Rightarrow RS_1 \cdot RS_2$, *and* RS *is not a* 2^+*-normal-form, then* RS_1 *can be either* 2^+*-normal-form or not, and* RS_2 *can be either* 2^+*-normal-form or not.*

Corollary 6 (Merge). *For two resource spaces* RS_1 *and* RS_2, *let* $RS_1 \cup RS_2 \Rightarrow RS$, *if* RS_1 *and* RS_2 *are* 2^+*-normal-form, then* RS *is a* 2^+*-normal-form.*

Proof. Suppose $RS_1 = \{X_1, X_2, \ldots, X_n\}$ satisfies $X_n/X_{n-1}/\cdots/X_2/X_1$, and $RS_2 = \{Y_1, Y_2, \ldots, Y_n\}$ satisfies $Y_n/Y_{n-1}/\cdots/Y_2/Y_1$. Since $RS_1 \cup RS_2 \Rightarrow RS$, then RS_1 and RS_2 have $n-1$ common axes and one different axis, we can suppose that

$X_i = Y_i$, $1 \le i \ne k \le n$, and $X_k \ne Y_k$. Then RS = $\{X_1, \ldots, X_{k-1}, (X_k \cup Y_k), X_{k+1}, \ldots, X_n\}$. Because $X_{k+1}/X_k/X_{k-1}$ and $X_{k+1} = Y_{k+1}/Y_k/Y_{k-1} = X_{k-1}$, according to Lemma 1, we have $R(X_{k-1}) \supseteq R(X_k) \supseteq R(X_{k+1})$ and $R(X_{k-1}) \supseteq R(Y_k) \supseteq R(X_{k+1})$. So $R(X_{k-1}) \supseteq (R(X_k) \cup R(Y_k)) \supseteq R(X_{k+1})$ holds, which means $R(X_{k-1}) \supseteq R(X_k \cup Y_k) \supseteq R(X_{k+1})$. According to Lemma 1, $X_{k+1}/(X_k \cup Y_k)/X_{k-1}$ holds, then $X_n/\cdots/X_{k+1}/(X_k \cup Y_k)/X_{k-1}/\cdots/X_1$, hence RS is a 2^+-normal-form. □

Corollary 6 tells us that 2^+-normal-form keeps under the Merge operation.

Corollary 7 (Split). *Let* RS ⇒ RS$_1$ ∪ RS$_2$, *if* RS *is a* 2^+*-normal-form, then* RS$_1$ *is a* 2^+*-normal-form or not, and* RS$_2$ *is a* 2^+*-normal-form or not.*

Proof. Suppose RS = $\{X_1, X_2, X_3\}$, $X_3/X_2/X_1$ and $X_2 = X_2' \cup X_2''$, then RS$_1$ = $\{X_1, X_2', X_3\}$ and RS$_2$ = $\{X_1, X_2'', X_3\}$.

(1) If either $R(X_3) \subseteq R(X_2')$ or $R(X_2') \subseteq R(X_3)$, correspondingly, we have: either $R(X_3) \subseteq R(X_2') \subseteq R(X_1)$ or $R(X_2') \subseteq R(X_3) \subseteq R(X_1)$. According to Definition 12, we have RS$_1$ is a 2^+-normal-form.

(2) Else if both $R(X_3) \subseteq R(X_2')$ and $R(X_2') \subseteq R(X_3)$ do not hold, then both X_2'/X_3 and X_3/X_2' do not hold. Since '/' is a full order on the axes of RS$_1$ if RS$_1$ is a 2^+-normal-form, so RS$_1$ is not a 2^+-normal-form.

According to (1) and (2), RS$_1$ is a 2^+-normal-form or not. For the same reason, RS$_2$ is a 2^+-normal-form or not. □

Corollary 7 tells us that the 2^+-normal-form does not keep under the Split operation. According to Corollary 7 we have the following corollary:

Corollary 8 (Split). *Let* RS ⇒ RS$_1$ ∪ RS$_2$, RS *is a* 2^+*-normal-form,* RS = $\{X_1, \ldots, X_{k-1}, X_k, X_{k+1}, \ldots, X_n\}$, $X_n/X_{n-1}/\cdots/X_2/X_1$, $X_k = X_k' \cup X_k''$, RS$_1$ = $\{X_1, \ldots, X_{k-1}, X_k', X_{k+1}, \ldots, X_n\}$, *and* RS$_2$ = $\{X_1, \ldots, X_{k-1}, X_k'', X_{k+1}, \ldots, X_n\}$, *if* $X_{k+1}/X_k'/X_{k-1}$, *then* RS$_1$ *is a* 2^+*-normal-form, and if* $X_{k+1}/X_k''/X_{k-1}$, *then* RS$_2$ *is a* 2^+*-normal-form.*

According to the above three corollaries, we have the following corollary:

Corollary 9 (Merge). *For two resource spaces* RS$_1$ *and* RS$_2$, *let* RS$_1$ ∪ RS$_2$ ⇒ RS, *if* RS$_1$ *or* RS$_2$ *is not* 2^+*-normal-form, then* RS *is a* 2^+*-normal-form or not.*

Corollary 10 (Split). *Let* RS ⇒ RS$_1$ ∪ RS$_2$, *if* RS *is not a* 2^+*-normal-form, then* RS$_1$ *is a* 2^+*-normal-form or not, and* RS$_2$ *is a* 2^+*-normal-form or not.*

The 2^+-normal-form is the weakened form of the third-normal-form, we can also get the strengthened form of the third-normal-form as follows:

Definition 13 *(4NF).* A resource space RS(X_1, X_2, …, X_n) is a fourth-normal-form (4NF) if it is a third-normal-form, and for any point $p(x_1, x_2, \ldots, x_n) \in$ RS, $R(p(x_1, x_2, \ldots, x_n)) = R(x_1) \cap R(x_2) \cap \cdots \cap R(x_n) \ne \Phi$ holds.

Because a fourth-normal-form is also a third-normal-form, it has the same properties as the third-normal-form under the resource space operations introduced in [15].

4. Resource space's topological properties

If we define a distance in n-dimensional resource space RS(X_1, X_2, …, X_n), then the distance can induce a topological space. We focus on the second-normal-form resource space, and first define a distance d on axis X_i, $1 \le i \le n$, then construct a distance D on the whole resource space RS according to d.

4.1. Distance on an axis

For a given set G, if there exists a function d: $G \times G \to \Re^+$, where $\Re^+$ represents the set of the non-negative real number, then d is called a distance on G, if it satisfies the following three axioms:

(A1) $d(g_1, g_2) = 0 \Leftrightarrow g_1 = g_2$;
(A2) $d(g_1, g_2) = d(g_2, g_1)$; and
(A3) $d(g_1, g_2) \le d(g_1, g_3) + d(g_3, g_2)$, for any g_1, g_2 and $g_3 \in G$.

For an axis $X = \{C_1, C_2, \ldots, C_n\}$, where $C_i = \langle V_i, E_i\rangle$ is a coordinate. Then, we define the function d on X as follows:

Definition 14. For two given points x_1 and x_2 on axis X,

$$d(x_1, x_2) = \begin{cases} 0 & \text{if } x_1 = x_2, \\ \infty & \text{if } x_1 \in V_i,\ x_2 \in V_j \text{ and } i \neq j, \\ \min\{\text{length}(\Gamma) | \Gamma \\ \quad = (x_1, x'_1, \ldots, x'_m, x_2) \\ \quad \text{where } \langle x_1, x'_1\rangle, \langle x'_j, x'_{j+1}\rangle, \langle x'_m, x_2\rangle \in E_i,\ 1 \le j \le m-1\} & \text{if } x_1 \text{ and } x_2 \in V_i \text{ and } x_1 \neq x_2, \end{cases}$$

where length (Γ) represents the length of the path Γ with weight on each edge. And we make a reasonable assumption: in case of x_1 and $x_2 \in V_i$ and $x_1 \neq x_2$, there always exists a path $\Gamma = (x_1, x'_1, \ldots, x'_k, x_2)$ from x_1 to x_2. So $d(x_1, x_2) <$ length $(\Gamma) < \infty$.

Theorem 3. *d is a distance on axis X.*

Proof. Axioms (A1) and (A2) are obvious, the following proof will focus on axiom (A3). The proof considers the following two cases, for any x_1, x_2 and x_3 on axis X:

(1) If $d(x_1, x_2) = \infty$, then $d(x_1, x_3) = \infty$ or $d(x_3, x_2) = \infty$ holds. So $d(x_1, x_2) \le d(x_1, x_3) + d(x_3, x_2)$ holds (if $d_1 = \infty$ and $d_2 = \infty$, we always assume that $d_1 = d_2$ herein).
(2) If $d(x_1, x_2) < \infty$, then there exists V_i, such that $x_1 \in V_i$ and $x_2 \in V_i$. In this case we have:
  (i) if $x_3 \notin V_i$, then $d(x_1, x_3) = d(x_3, x_2) = \infty$. So $d(x_1, x_2) \le d(x_1, x_3) + d(x_3, x_2)$ holds;
  (ii) if $x_3 \in V_i$, then $d(x_1, x_3) < \infty$ and $d(x_3, x_2) < \infty$. Since the set of all the paths that satisfy conditions (between x_1 and x_3 or between x_2 and x_3) is finite, we can find paths Γ_1 and Γ_2, such that $d(x_1, x_3) = \text{length}(\Gamma_1)$ and $d(x_3, x_2) = \text{length}(\Gamma_2)$, where $\Gamma_1 = (x_1, x'_1, \ldots, x'_s, x_3)$ and $\Gamma_2 = (x_3, x''_1, \ldots, x''_t, x_2)$. Then, $\Gamma_1 \cup \Gamma_2$ is a path from x_1 to x_2, so $d(x_1, x_2) \le \text{length}(\Gamma_1 \cup \Gamma_2)$ holds. So $d(x_1, x_2) \le \text{length}(\Gamma_1 \cup \Gamma_2) = \text{length}(\Gamma_1) + \text{length}(\Gamma_2) = d(x_1, x_3) + d(x_3, x_2)$.

According to (1) and (2), d constitutes a distance on axis X. □

4.2. Distance in resource space

We first give the definition of function D on RS, and then prove that it is a distance on RS.

Definition 15. For any two points $p_1(x_1, x_2, \ldots, x_n)$ and $p_2(y_1, y_2, \ldots, y_n)$ in the resource space RS($X_1, X_2, \ldots, X_n$), we define $D(p_1, p_2) = \left(\sum_{i=1}^{n} d^2(x_i, y_i)\right)^{1/2}$, where d is the distance on axis X_i, $1 \le i \le n$.

Theorem 4. *D is a distance in* RS.

Proof. According to the proof that d satisfies all the axioms of distance, we can prove that $D(p_1, p_2)$ satisfies axioms (A1) and (A2). So we focus on the proof that D satisfies axiom (A3): $D(p_1, p_2) \le D(p_1, p_3) + D(p_3, p_2)$, for any p_1, p_2 and $p_3 \in$ RS.

(1) Just as in Theorem 3, in the cases that $D(p_1, p_2) = \infty$ or $D(p_1, p_3) = \infty$ or $D(p_3, p_2) = \infty$, the proof is trivial. So here we will only discuss the case that $D(p_1, p_2) < \infty$ and $D(p_1, p_3) < \infty$ and $D(p_3, p_2) < \infty$.
(2) Because $D(p_1, p_2) < \infty$ and $D(p_1, p_3) < \infty$ and $D(p_3, p_2) < \infty$, we suppose that there exists $p_3(z_1, z_2, \ldots, z_n) \in$ RS, such that

$$D(p_1, p_2) = \left(\sum_{i=1}^{n} d^2(x_i, y_i)\right)^{1/2},$$

$$D(p_1, p_3) = \left(\sum_{i=1}^{n} d^2(x_i, z_i)\right)^{1/2} \quad \text{and}$$

$$D(p_3, p_2) = \left(\sum_{i=1}^{n} d^2(z_i, y_i)\right)^{1/2}.$$

According to the Cauchy inequality $\sum_{i=1}^{n} u_i v_i \le \left(\sum_{i=1}^{n} u_i^2\right)^{1/2} \times \left(\sum_{i=1}^{n} v_i^2\right)^{1/2}$, $u_i \ge 0$, $v_i \ge 0$, $1 \le i \le n$.

We get $2\sum_{i=1}^{n} d(x_i, z_i) \times d(z_i, y_i) \le 2\left(\sum_{i=1}^{n} d^2(x_i, z_i)\right)^{1/2} \times \left(\sum_{i=1}^{n} d^2(z_i, y_i)\right)^{1/2}$.

That means

$$2\sum_{i=1}^{n} d(x_i, z_i) \times d(z_i, y_i) \le 2D(p_1, p_3) \times D(p_3, p_2). \tag{i}$$

And we have

$$\sum_{i=1}^{n} d^2(x_i, z_i) = D^2(p_1, p_3) \tag{ii}$$

and

$$\sum_{i=1}^{n} d^2(z_i, y_i) = D^2(p_3, p_2) \tag{iii}$$

hold.

Sum both sides of Eqs. (i)–(iii), we get

$$\begin{aligned} &2\sum_{i=1}^{n} d(x_i, z_i) \times d(z_i, y_i) \\ &\quad + \sum_{i=1}^{n} d^2(x_i, z_i) + \sum_{i=1}^{n} d^2(z_i, y_i) \\ &\le 2D(p_1, p_3) \times D(p_3, p_2) + D^2(p_1, p_3) \\ &\quad + D^2(p_3, p_2). \end{aligned}$$

That means $\sum_{i=1}^{n}(d(x_i, z_i) + d(z_i, y_i))^2 \le (D(p_1, p_3) + D(p_3, p_2))^2$.

And from $d(x_i, y_i) \le d(x_i, z_i) + d(z_i, y_i)$ (d is a distance), we get $\sum_{i=1}^{n} d^2(x_i, y_i) \le \sum_{i=1}^{n}(d(x_i, z_i) + d(z_i, y_i))^2 \le (D(p_1, p_3) + D(p_3, p_2))^2$. That means $D^2(p_1, p_2) \le (D(p_1, p_3) + D(p_3, p_2))^2$.

So $D(p_1, p_2) \le D(p_1, p_3) + D(p_3, p_2)$. According to (1) and (2), D is a distance on RS. □

Now we have proved that D is a distance on RS, so the resource space RS(X_1, X_2, ..., X_n) constitutes a metric space (RS, D) with the distance D in it. The distance D in RS naturally induces a discrete topological space (RS, ρ). The following section discusses the properties of the topological space (RS, ρ).

4.3. Topological properties of resource space

According to the definition of distance d, we have $d(x_1, x_2) < \infty \Leftrightarrow x_1$ and x_2 belong to the same coordinate hierarchy.

Definition 16. For two points $p_1(x_1, x_2, \ldots, x_n)$ and $p_2(y_1, y_2, \ldots, y_n)$ in the resource space RS, p_1 is called connective to p_2 if $D(p_1, p_2) < \infty$. For a set of points P in RS, P is called a connective branch, if for any two points p_i and p_j in P, p_i is connective to p_j.

According to above definition, we can get the following corollary.

Corollary 11. *In the Resource Space* RS(X_1, X_2, ..., X_n), *if a set of points P constitutes a connective branch, then for any two points* $p_1(x_1, x_2, \ldots, x_n)$ *and* $p_2(y_1, y_2, \ldots, y_n)$ *in P,* x_i *and* y_i $(1 \le i \le n)$ *belong to the same coordinate hierarchy*.

Proof. If P constitutes a connective branch, then for any two points $p_1(x_1, x_2, \ldots, x_n)$ and $p_2(y_1, y_2, \ldots, y_n)$ in P, $D(p_1, p_2) < \infty$.

Since $D(p_1, p_2) = \left(\sum_{i=1}^{n} d^2(x_i, y_i)\right)^{1/2}$, we can get $d(x_i, y_i) < \infty$, $1 \le i \le n$. Hence, x_i and y_i belong to the same coordinate hierarchy. □

Corollary 11 tells us that if two points in the resource space are connective to each other, then their corresponding coordinates belong to the same coordinate hierarchy.

It is clear that the connective relation (denoted as $\sim$) constitutes an equivalent relation on the topological space RS. So RS/$\sim$ constitutes a quotient space of the space RS. The following corollary describes the structure of the quotient space RS/$\sim$.

Corollary 12. *The quotient space* $\mathrm{RS}/\!\sim\, = \{p'(C_{i1}^1, C_{i2}^2, \ldots, C_{in}^n) | C_{ik}^k$ *is a root coordinate on axis* X_k *in* RS(X_1, X_2, ..., X_n), $1 \le k \le n\}$, *where* $p'(x_1, x_2, \ldots, x_n)$ *in* RS/$\sim$ *represents the connective branch including point* $p(x_1, x_2, \ldots, x_n)$ *in* RS.

Proof.

(1) It is clear that any point $p'(C_{i1}^1, C_{i2}^2, \ldots, C_{in}^n)$ is in RS/$\sim$. So $\mathrm{RS}/\!\sim\, \supseteq \{p'(C_{i1}^1, C_{i2}^2, \ldots, C_{in}^n) | C_{ik}^k$ is a root coordinate on axis X_k in RS$\}$.

(2) For any point $p(x_1, x_2, \ldots, x_n)$ in RS(X_1, X_2, ..., X_n), according to Corollary 11, we get that there exists a root coordinate C_{i1}^1 on axis X_1, ..., and C_{in}^n on axis X_n, such that x_1 in $C_{i1}^1, \ldots, x_n$ in C_{in}^n. So $p(x_1, x_2, \ldots, x_n)$ is in the connective branch of $p'(C_{i1}^1, C_{i2}^2, \ldots, C_{in}^n)$, which means

$p'(x_1, x_2, \ldots, x_n) = p'(C_{i1}^1, C_{i2}^2, \ldots, C_{in}^n)$. Then we have $RS/\sim\ \subseteq \{p'(C_{i1}^1, C_{i2}^2, \ldots, C_{in}^n)|C_{ik}^k$ is a root coordinate on axis X_k in RS$\}$.

According to (1) and (2), $RS/\sim\ = \{p'(C_{i1}^1, C_{i2}^2, \ldots, C_{in}^n)|C_{ik}^k$ is a root coordinate on axis X_k in $RS(X_1, X_2, \ldots, X_n)$, $1 \le k \le n\}$ holds. □

In the quotient space $RS/\sim$, we can define a distance $D_\sim$ on $RS/\sim$ as the induced distance of the distance D on RS. $D\sim(p'_1, p'_2) = \min\{D(p_1, p_2)|p_1 \in p'_1$ and $p_2 \in p'_2\}$, where p'_1 and p'_2 represent the connective branch including p_1 and p_2, respectively. Then for any $p'_1, p'_2 \in RS/\sim$, $p'_1 \neq p'_2$, $D\sim(p'_1, p'_2) = \infty$, $D\sim(p'_1, p'_2) = 0$, which means that $RS/\sim$ constitutes a discrete topological space with the distance $D_\sim$ on it.

The Resource Space RS enables us to locate resources by coordinates. The quotient space $RS/\sim$ enables us to search in a more abstract space.

Theorem 5. *A point exists in* RS *if and only if it belongs to a point of* $RS/\sim$.

Proof.

(1) If a $p(x_1, x_2, \ldots, x_n)$ exists in RS, then according to Corollary 12, there exists $p'(C_{i1}^1, C_{i2}^2, \ldots, C_{in}^n)$ in $RS/\sim$ such that $p(x_1, x_2, \ldots, x_n)$ is in the connective branch of $p'(C_{i1}^1, C_{i2}^2, \ldots, C_{in}^n)$. So $p(x_1, x_2, \ldots, x_n)$ belongs to a point of $RS/\sim$.
(2) If a $p'(x_1, x_2, \ldots, x_n)$ exists in $RS/\sim$, then according to Corollary 12, all the points in the connective branch $p'(x_1, x_2, \ldots, x_n)$ are in RS, so it also belongs to RS.

According to (1) and (2), we can reach that a point exists in RS if and only if it also belongs to a point in $RS/\sim$. □

The above theorem provides with a top-down refinement search strategy in a large-scale resource space: from the quotient space down to the resource space, and it also guarantees that all the resources in resource space RS can be found through $RS/\sim$.

5. Conclusions and future work

This paper extends the theory of the Resource Space Model in the following aspects: (1) formally define the resource space from the perspective of set theory; (2) propose the concept of the resource space schema; (3) propose the 2^+NF and the 4NF of the resource space schema and relevant theory; and (4) unveil the topological properties of the resource space.

Future work includes: defining a sufficient and self-contained operation sets for the RSM, studying the properties of the normal forms under more operations of the RSM, developing new mathematical properties of the RSM, and investigating the integrity constraints of the resource space.

Acknowledgements

The authors thank all members of China Knowledge Grid Research Group (http://kg.ict.ac.cn, http://www.knowledgegrid.net) for their diligent work and contribution to this paper. This work was supported by National Fundamental Research 973 Program (2003CB317000) and National Science Foundation of China (60273020 and 70271007).

References

[1] T. Berners-Lee, J. Hendler, O. Lassila, Semantic Web. Sci. Am. 284 (5) (2001) 34–43.
[2] E.F. Codd, A relational model of data for large shared data banks, Commun. ACM 13 (6) (1970) 377–387.
[3] E.F. Codd, Normalized database structure: a brief tutorial, in: ACM SIG-FIDET Workshop on Data Description, Access, and Control, San Diego, California, November, 1971.
[4] A. Eisenberg, J. Melton, Sql:1999, Formerly Known as Sql3, SIGMOD Rec. 28 (1) (1999) 131–138.
[5] D. Fensel, et al., OIL: an ontology infrastructure for the semantic web, IEEE Intell. Syst. 16 (2) (2001) 38–45.
[6] I. Foster, Internet computing and the emerging grid, 2000 http://www-fp.mcs.anl.gov/∼foster/.
[7] I. Foster, et al., Grid services for distributed system integration, Computer 35 (6) (2002) 37–46.
[8] J. Han, M. Kambr, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2000.
[9] A. Maedche, S. Staab, Ontology learning for the semantic web, IEEE Intell. Syst. 16 (2) (2001) 72–79.
[10] D. Schoder, K. Fischbach, Peer-to-peer prospects, Commun. ACM 46 (2) (2003) 27–29.
[11] M. Stonebraker, P. Brown, D. Moore, Object-Relational DBMSs: Tracking the Next Great Wave, Morgan Kaufmann Publishers, San Francisco, 1999.
[12] Z. Tari, J. Stokes, S. Spaccapietra, Object normal forms and dependency constraints for object-oriented schemata, ACM Trans. Database Syst. 22 (4) (1997) 513–569.

[13] K. William, A simple guide to five normal forms in relational database theory, Commun. ACM 26 (2) (1983) 120–125.
[14] H. Zhuge, Resource space model, its design method and applications, J. Syst. Softw. 72 (1) (2004) 71–81.
[15] H. Zhuge, Resource space grid: model, method and platform, Concurrency and computation: practice and experience 16 (14) (2004) 1385–1413.

Hai Zhuge is the professor and director of the Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences. He leads the China Knowledge Grid Center (http://www.knowledgegrid.net), which includes China Knowledge Grid Research Group (http://kg.ict.ac.cn), Dunhuang Knowledge Grid Lab, and Hunan Knowledge Grid Lab. The center has over 30 young researchers. He is the chief scientist of the China Semantic Grid project—a 5-year national grand fundamental research plan project. He was the keynote speaker at the 2nd International Conference on Grid and Cooperative Computing (GCC03) and the 5th International Conference on Web-Age Information Management (WAIM04), and the Chair of the 2nd International Workshop on Knowledge Grid and Grid Intelligence (KGGI04). He has organized a journal special issue on Semantic and Knowledge Grid. He is serving as the Area Editor of the *Journal of Systems and Software*, the Associate Editor of *Future Generation Computer Systems*, and on the editorial boards of the *Information and Management*, *Journal of Computer Science and Technology*, and *Electronic Commerce Research and Applications*. His current research interests are the models, theories, methods and applications of the future interconnection environment. He is the author of two books and over 60 papers appeared mainly in leading international conferences and journals such as *Communications of the ACM*; *IEEE Computer*; *IEEE Intelligent Systems*; *IEEE Computing in Science and Engineering*; *IEEE Transactions on Systems, Man, and Cybernetics*; *Information and Management*; *Decision Support Systems*; *Journal of Systems and Software*; *Concurrency and Computation: Experience and Practice*; *Knowledge-based Systems*; *Information and Software Technology*; *Expert Systems with Applications*, and, *Future Generation Computer Systems*. He is an IEEE Senior Member.

Erlin Yao is a Ph.D. student at the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. His current research interest is the theory and formalization of the Knowledge Grid, especially in the resource space model and semantic link network.

Yunpeng Xing is a Ph.D. student at the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. His research interest is integrity theory of resource space model.

Jie Liu is a Ph.D. student at the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. Her research interests mainly include: P2P Computing, Semantic Web service integration, and semantic heterogeneous data integration. She has published six papers in international journals and received Microsoft Fellow Award in 2003. Now she is a member of the IEEE Computer Society.