Probabilistic Resource Space Model for Managing Resources in Cyber-Physical Society

Hai Zhuge, *Senior Member, IEEE* and Yunpeng Xing

Abstract — Classification is the most basic method for organizing resources in the physical space, cyber space, socio space and mental space. To create a unified model that can effectively manage resources in different spaces is a challenge. The Resource Space Model RSM is to manage versatile resources with a multi-dimensional classification space. It supports generalization and specialization on multi-dimensional classifications. This paper introduces the basic concepts of RSM, and proposes the Probabilistic Resource Space Model, P-RSM, to deal with uncertainty in managing various resources in different spaces of the cyber-physical society. P-RSM's normal forms, operations and integrity constraints are developed to support effective management of the resource space. Characteristics of the P-RSM are analyzed through experiments. This model also enables various services to be described, discovered and composed from multiple dimensions and abstraction levels with normal form and integrity guarantees. Some extensions and applications of the P-RSM are introduced.

Index Terms — cyber-physical society, faceted navigation, non-relational data model, resource management, resource space model, semantic link network, cyber-physical-socio services.

1 Introduction

TO create a unified model for effectively organizing and managing various resources with uncertainty in the cyber space, physical space, socio space, and mental space is a fundamental challenge.

1.1 Requirement of Managing Resources in Diverse Spaces

The physical space contains physical resources, which move and transform from one form into another according to physical laws. The resources can be classified from physical structures or features. Many classification tools like bookshelfs and drawers have been invented to effectively organize and manage resources in the physical space.

The socio space includes individuals (humans, behaviors, events, etc), structures, and rules. Individuals are self-organized into classes according to different economic, politic or cultural statuses.

The mental space consists of knowledge in form of concept, experience, commonsense, rule, and theory. Various taxonomies have been created as the model for managing knowledge.

The cyber space contains digital resources and mechanisms for providing various digital services. Many classification mechanisms such as file system and ACM CCS (Computing Classification System) have been designed to effectively manage digital resources in the cyber space.

The physical space, socio space, mental space, and cyber space will cooperate with each other to form the cyber-physical society [39]. How to create a uniform model for organizing versatile resources in diverse spaces?

Classification is the most basic method for organizing various resources in the cyber space, physical space, socio space, and mental space. Usual classification method is one dimensional. There are two major reasons to use multi-dimensional classifications:

1. *Humans need to explore large-scale resource set from multiple dimensions* (*facets*). For example, faceted browsing on the Web enables users to know multi-facet contents of web pages.
2. *Increasing or reducing dimension is an effective way to specialize or generalize knowledge in mind and resources in the cyber space*.

The Resource Space Model (RSM) is a resource management model based on multi-dimensional classifications [35][36]. *A resource space is a multi-dimensional classification space, where each dimension represents a classification method*. The file system can be seen as a one-dimensional classification space in the cyber space. RSM enables users to operate resource spaces in the cyber space according to the classifications in their mental spaces.

1.2 Managing Resources with Uncertainty

Previous resource management approaches only concern the efficiency of managing resources in one space. The future cyber-physical society will be a complex space involved in many uncertain movements, behaviours and events. In many cases, it is hard to clearly classify a set of resources into definite classes. It is necessary to explore the unified model for managing resources in various spaces with uncertainty.

• *Correspondence author: Hai Zhuge is with the Cyber-Physical-Socio Knowledge Grid Research Group, Key Lab of Intelligent Information Processing at the Chinese Academy of Sciences' Institute of Computing Technology, and the Southwest University, China. E-mail: zhuge@ict.ac.cn.*

Manuscript received July. 2009.

Previous approaches to managing resources with uncertainty in the cyber space include two types: incorporating probabilistic methods into information retrieval mechanisms [3][11], and creating an appropriate semantic model.

Most previous data models, like classical relational data model, mainly organize and manage certain data [1][8]. Research has been done to extend traditional data models to manage uncertain data.

Research on uncertain relational data models largely falls into two categories depending on whether the model satisfies the first normal form (1NF) or not. Models satisfying 1NF usually assume that the existence of an entity is uncertain and the probabilities are associated with each tuple to indicate this type of uncertainty [6][12]. Models using non-1NF usually assume that the existence of an entity is certain, but the attribute values of an entity are uncertain [4][14]. They associate probabilities with attributes of a tuple. These two types of probabilistic relational models have their own limitations. The probabilistic relational table satisfying 1NF is limited in ability to represent the probabilities of attribute values, and, using tuple probabilities to specify the probabilities of attributes' values could lead to information loss or combinatorial explosion of tuples. The non-1NF probabilistic relational models often accompany with complicated algebras and querying mechanisms. The ProbView is an attempt to overcome the two types of limitations [21]. It firstly transforms non-1NF data into corresponding annotated 1NF patterns, and then applies all manipulation and query operations to the corresponding 1NF data. But the transformation from non-1NF into 1NF is not an equivalent transformation, so some useful information may be lost during transformation.

Previous probabilistic relational data models mainly concern the existence of a certain entity or the possibility of taking different attribute values of a certain entity [5][29]. Little work has been done on the data model based on uncertain classification.

Relevant research concerns uncertain classification and dataspace [13][15][31]. Dataspace is to realize effective personal information management by integrating resources from various types of data sources that may be uncertain. Research also concerns fuzzy database query language [28], uncertain ontology modeling [30], probabilistic queries on probabilistic database and evaluation [9, 10, 25]. A system integrating research on data management, accuracy and lineage is introduced in [32]. The combination of XML and relational database has been investigated to incorporate both advantages [19].

Much work has been done to manage probabilistic data in XML (eXtensible Markup Language [18][20][24]. A framework is proposed to acquire, maintain and query XML documents with incomplete information, in which the order in documents and DTDs (Document Type Definition) is ignored [2]. A probabilistic XML approach is proposed to resolve conflicts during data integration, where the order in documents and DTDs plays an important role [21]. Complexity for managing probabilistic XML data is analysized in [27].

1.3 Technical Path

This paper firstly introduces the resource space model based on multi-dimensional classifications, and shows its characteristics by comparing with the relational data model in section 2. Then, a Probabilistic Resource Space Model, P-RSM, is proposed in section 3 by mapping the RSM into the probabilistic space. The operations and the integrity constraints of P-RSM are introduced in section 4 and section 5 respectively to complete the model. The characteristics of P-RSM are analysed through experimental comparison in section 6. Section 7 introduces some potential extensions of the model: the satisfactory constraints for effective resource management, transforming 1NF into 2NF, integrating resource space with semantic link network to support advanced applications, and automatically uploading resources into resource space. Section 8 describes the applications of P-RSM in faceted search and managing service resources.

2 The Resource Space Model RSM

2.1 Basic Concepts

A set of resources can be classified by multiple classification methods. If we view a classification method as a dimension, a multi-dimensional classification space can be formed by coordinating the classification methods.

A resource space is an *n*-dimensional classification space represented as $RS(X_1, X_2, \dots, X_n)$, where X_i are *dimensions* (i.e., axes) defined by a set of *coordinates*. A coordinate can be a coordinate tree (i.e., a classification tree), where every node represents a basic concept or a pattern representing a category of resources (e.g., the words sequentially co-occurred in a set of documents can be regarded as a kind of pattern, and communities in socio networks can be regarded as another kind of pattern). A child node is the subclass of its parent node. One point in the space represents the resources of one category.

The hierarchical structure of dimension supports generalization and specialization and it distinguishes the resource space from ordinary distance space.

In the following discussion, $R(C)$ and $R(p)$ denote the resource sets that coordinate C and point p represent respectively.

Axis $X=(C_1, C_2, \dots, C_m)$ forms a *fine classification* on coordinate C_i' at another axis X', (denoted as C_i'/X) if and only if (1) $R(C_k)\cap R(C_p)\cap R(C_i')=\varnothing$ ($k\neq p$, and $k, p\in[1, m]$), and (2) $(R(C_1)\cup R(C_2)\cup \dots \cup R(C_m))\cap R(C_i')=R(C_i')$.

As the result of the fine classification, $R(C_i')$ is classified into m classes: $R(C_i'/X)=\{R(C_1)\cap R(C_i'), R(C_2)\cap R(C_i'), \dots, R(C_m)\cap R(C_i')\}$.

For two different axes X and X', X forms a fine classification on X' (denoted as X'/X) if and only if X forms a fine classification on each coordinate of X'.

X and X' are called orthogonal with each other in classification (denoted as $X \perp X'$) if X'/X and X/X'.

According to above definitions, we have:

$X \perp X'$ if and only if $R(X')\cap R(X)=R(X)$ and $R(X)\cap R(X')=R(X')$, where $R(X)=R(C_1)\cup R(C_2)\cup\dots\cup R(C_m)$.

This indicates the following lemma:

Lemma 1. $X \perp X'$ if and only if $R(X') = R(X)$.

Lemma 1 indicates that *two axes are orthogonal in classification if and only if their expression ability is the same*.

Any *point p* is determined by its projections on all axes, $p[X_i]$ or $p.X_i$ denotes the *projection* of p on X_i. A point can determine a resource set, where each element is called a *resource entry*. Point and resource entry are two basic operation units of RSM. The resources represented by point p is $R(p) = R(p[X_1]) \cap R(p[X_2]) \cap \ldots \cap R(p[X_n])$.

Fig. 1 is an example of a 3-dimensional resource space *Spec-Apart-Gen*(*Specialization*, *Apartment*, *Gender*) that manages information of students in a college. Three axes are *Specialization* = {*math*, *chemistry*, *physics*}, *Apartment* = {1#, 2#, 3#}, and *Gender* = {*male*, *female*}. Each point denotes a class of students. For example, the point (*math*, 1#, *male*) represents all of the male students who are studying mathematics and living in the apartment of type 1# in this college. Each resource entry in this point describes a student of the college.

The coordinate directly at axis is called the top-level coordinate, from which, a classification hierarchy can be defined top-down. Take Fig. 1 for example, the coordinate *chemistry* at axis *specialization* is classified into g_1, g_2 and g_3 in terms of *grade*, and then they can be further classified according to *class*. In this tree, the label of each node is determined by the full path from the root. Thus, the leaf node '*chemistry*.g_1.c_1' can be distinguished from '*chemistry*.g_2.c_1'. The hierarchical resource space can be transformed into the equivalent 'flat' resource space by projecting each leaf node of the coordinate tree onto the axis where the root resides [35][36], but this projection makes a resource space lose abstraction layers.

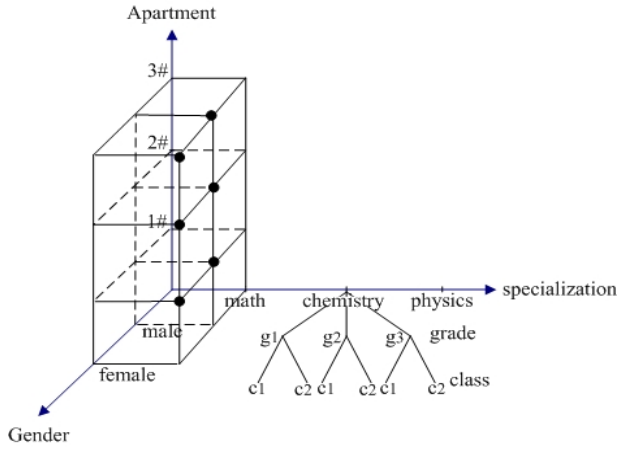


Fig. 1. A 3-dimensional resource space *Spec-Apart-Gen*.

To ensure the correctness of operating resources, RSM defines a set of normal forms. The following are main normal forms:

The first normal form resource space (*1NF*) *is a resource space where there are no duplicated axes and there are no duplicated coordinates at any axis, i.e., there is no duplicated subclasses in each class hierarchy.*

The 1NF is to avoid explicit redundancy.

The second normal form resource space (*2NF*) *is a 1NF resource space where coordinates at any axis are independent of each other, i.e., a coordinate is neither a part of another nor can represent another coordinate at the same axis*.

The 1NF and 2NF, to different extents, enable a resource space to accurately locate a class of resources.

The third normal form resource space (*3NF*) *is a 2NF resource space where different axes are orthogonal with each other.*

The 3NF enables any point to uniquely locate a class of resources. Resources in a 3NF resource space can be accessed from any axis. The 4NF can be further defined by ruling out the empty points in resource space [35].

The resource space shown in Fig.1 satisfies the 1NF and the 2NF, but it may not satisfy 3NF as the gender axis is usually not in orthogonal with the specialty axis (e.g., $R(male) \neq R(male) \cap (R(math) \cup R(chemistry) \cup R(physics))$.

Given the ontology of a domain, coordinates can be accurately specified by a set of concepts in the ontology. Therefore, the normal forms of a given resource space can be automatically verified according to the relations between concepts in the ontology.

The main theory of RSM consists of the basic methodology of resource space, normal forms, operations on resource spaces and their completeness, relations between operations and normal forms, RSM algebra and calculus, expressiveness of query language, search complexity, storage mechanism, and decentralized RSM [35][36]. To establish a powerful semantic model for Web applications, integration and mappings between RSM, OWL, and database were studied [37].

2.2 Comparison between the Resource Space Model and the Relational Database Model

Identity is the basis of the relational database model. RSM regards classification as the basis since it reflects the basic semantics of resources. Human's recognition is based on innate classification ability.

The following example shows the characteristics of the RSM. Multi-layer tables exhibit integrated information of multiple generalization layers. The high layers represent generalization on the lower layers. The lower layer constitutes a fine classification on the higher layer. Fig. 2 is a multi-layer table on university human resources, which naturally constitutes classification trees [36].

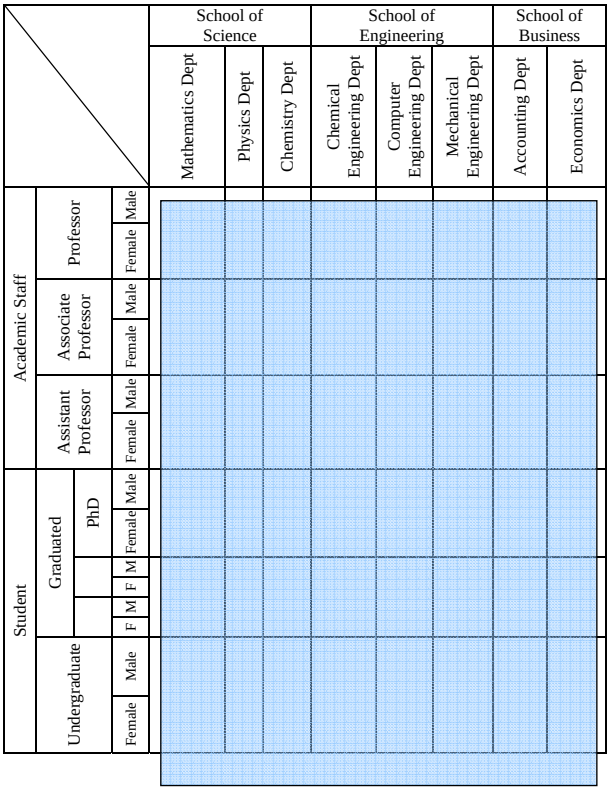
Since the first normal form of the relational data model requires a flat table and atomic values of attributes, it is inappropriate to use a relational table to represent such a multi-layer table. However, it can be naturally converted to the 3-dimensional resource space shown in Fig. 3, which can naturally reserve the generalization levels. The more layers the table has, the more advantages the RSM exhibits.

The essential differences between RSM and relational data model are as follows:

1. *RSM is based on classification*. This enables resources of the same class to be organized closely and retrieve at

the same time. The generalization and specialization on classifications enable users or applications to effectively organize and manage heterogeneous resources according to contents. The traditional relational data model is based on identity, attribute and values as well as the dependence between attributes. It does not support generalization and specialization on attributes.

2. *RSM is a multi-dimensional classification space and its normalization basis is the relations between classifications*. It naturally supports faceted search (navigation or browsing) in a large resource space. The relational data model is based on the flat relational table, and its normalization basis is the functional dependence relations between attributes.
3. *RSM concerns the contents of resources*. The relational data model concerns the attributes of entities, and supports attribute-based operations. For open domain applications, resources become more and more important. The contents of resources cannot be reflected by attributes, for example, the contents of texts and images cannot be reflected by their attributes. If patterns in the resources to be managed are available, some advanced functions of the resource space such as automatic construction, adaptation, and uploading resources are feasible.
4. *RSM supports a universal resource view on resources and generalization and specialization on classifications*. The relational data model can manage multiple tables and support views on them, but it is difficult in maintaining the consistency between large-scale tables, e.g., thousands of tables in some applications. It is also hard for the relational data model to support generalization and specialization.



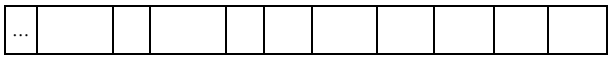


Fig. 2. Multi-layer table for managing university human resources.

The above characteristics indicate that the relational data model is suitable for managing data while the RSM is more suitable for managing classes not only in the cyber space but also in the physical space, the socio space and the mental space.

Although a resource space and a relational database can be transformed from one into the other [37], a good resource space (especially for a multi-dimensional resource space) may not be transformed into a good relational table, vice versa.

The detailed differences between the RSM and other techniques such as the relational database model and the data cube were discussed in chapter 1 of reference [36].

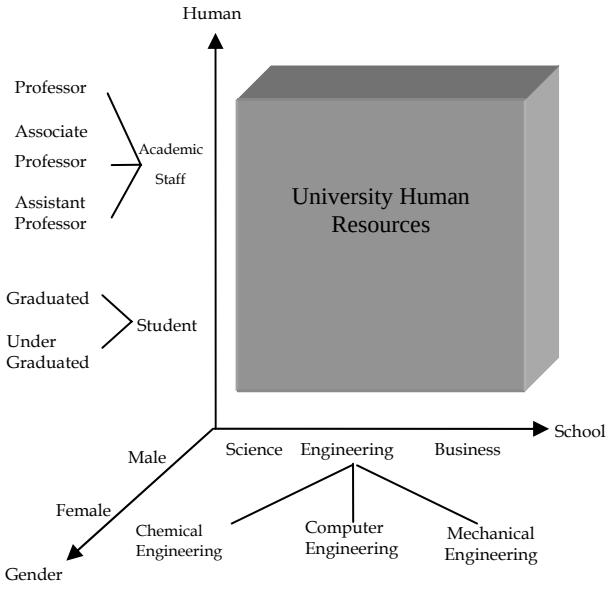


Fig. 3. The resource space for managing university human resources.

2.3 Managing Resources in Cyber-Physical Society

One characteristic of the cyber-physical society is that various resources in the cyber, physical, socio, and mental spaces can be uniformly organized from different scales and abstraction levels in real-time and lifetime. RSM is suitable for supporting this characteristic since classification is the most basic method for organizing resources in these spaces. Various sensors detect the statuses and events in the physical space and socio space, which can be classified and indexed in the resource space.

The following is a three-dimensional resource space that can reflect individuals' behaviors in room through time:

RS(*Behaviors*, *Time*, *Sensortype*), where *Behaviors*, *Time*, *Sensortype* are dimensions, and *Behaviors* =(*MakeFood*(*MakeCoffee*, *CookDish*, *MakeTea*), *Health*(*BloodPresure*, *HeartBeat*, *BrainSignal*,

RespiratoryRate), *Entertainment*(*PlayGame*, *PlayChess*, *PlayCard*, *ListenMusic*)).

Emotions in the mental space and social space will be reflected by the cyber space through the cyber-mental interface and the cyber-socio interface. The cyber space can alarm humans or make recommendations when abnormal phenomena, behaviors and events are detected in the physical space and socio space. The resource space can help the cyber-physical society recognize the phenomena, behaviors, and events from multiple dimensions.

3 The Probabilistic Resource Space Model P-RSM

3.1 Probabilistic Resource Space

A probabilistic event in the Probabilistic Resource Space Model P-RSM is the probability that a resource belongs to a certain class. $Prob(r \in T)$ denotes the membership probability that resource r belongs to a class T. T can represent a class of resources, an axis, a coordinate, a point, or any of their combination by set operations.

The following are two possible strategies for specifying the probabilistic distribution that a given resource belongs to a resource space.

1. Specify the membership probability distribution of every resource on all points in the resource space.
2. Specify the membership probability distribution of every resource on all coordinates at every axis.

The second strategy is more efficient because of the following reasons:

1. The number of points in resource space $RS(X_1, X_2, \dots, X_n)$ is $|X_1|\times|X_2|\times\dots\times|X_n|$, but the number of coordinates is $|X_1|+|X_2|+\dots+|X_n|$, where $|X|$ is the number of coordinates at X. The large number of points makes it difficult to specify and manage the membership probability of every resource to every point.
2. Each axis in a resource space represents a classification method on resources. A point is defined by its projection on all axes. To specify the membership probability distribution that a resource belongs to a point concerns multiple classifications simultaneously. It is easier for users or automatic classification algorithms to specify the membership probability distribution of a resource on the coordinates of every axis.

Definition 1. A probabilistic resource space $<RS(X_1, \dots, X_n), \beta_{ri}: X_i \to [0, 1], i \in [1, n]>$ consists of a resource space $RS(X_1, \dots, X_n)$ and a membership probability function β_{ri}, for any resource r and axis X_i in the resource space, $\beta_{ri}(C)$ denotes the membership probability that r belongs to coordinate C under the condition that r belongs to the parent coordinate of C. If C is a top-level coordinate at X_i, then its parent is X_i.

According to above definition, any resource r in a probabilistic resource space $RS(X_1, \dots, X_n)$ can have n membership probabilistic functions corresponding to the axes. Take the probabilistic resource space $RS(A, B, C)$ in Fig. 4 for example, resource r has the following three membership probabilistic functions: $\beta_{r\text{-}A}: A \to [0, 1]$, $\beta_{r\text{-}B}: B \to [0, 1]$, and $\beta_{r\text{-}C}: C \to [0, 1]$.

Resource r belongs to resource space $RS(X_1, \dots, X_n)$ if and only if there exists at least one axis X_i such that the membership probabilistic function of r at X_i has been explicitly specified.

From the definition of the probabilistic resource space, we can specify the membership probability that a resource belongs to each coordinate. Axis and point are classes of different granularities, both of them concern set operations on coordinates. So, the membership probability that a resource belongs to an axis or a point is a complex probabilistic event. Without knowing the relations between coordinates, and the relations between coordinate and axis, it is difficult to calculate the membership probability that a resource belongs to an axis or a point according to the membership probability that the resource belongs to each coordinate. In the P-RSM, we use a real number interval to specify the possible membership probability that a resource belongs to an axis or a point. According to the membership probability on each coordinate, the membership probability on each axis/point in a probabilistic resource space can be calculated by the following methods.

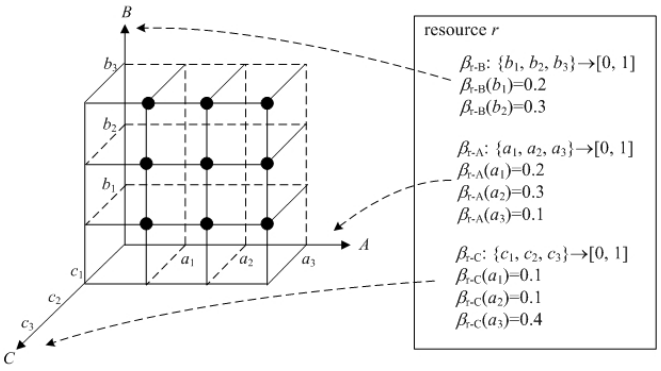


Fig. 4. An example of probabilistic resource space. $\beta_{r\text{-}B}(b_1)=0.2$ means that the probability that resource r belongs to coordinate b_1 at axis B is 0.2.

Methods for calculating membership probability:

1. For axis $X_i=\{C_{i1}, C_{i2}, \dots, C_{im}\}$, the probability that r belongs to X_i falls into the interval $[max\{\beta_{ri}(C_{i1}), \dots, \beta_{ri}(C_{im})\}, min\{1, \beta_{ri}(C_{i1})+\dots+\beta_{ri}(C_{im})\}]$, since $R(X_i)= R(C_{i1}) \cup \dots \cup R(C_{im})$.
2. For point p, the probability that resource r belongs to p is equal to the probability that r simultaneously belongs to $p[X_1]$, $p[X_2]$, …, and $p[X_n]$, that is, $Prob(r \in R(p)) = Prob(r \in (R(p[X_1]) \cap \dots \cap R(p[X_n])))$. The event that events A and B occur simultaneously satisfies $max\{0, Prob(A)+Prob(B) -1\} \le Prob(A \wedge B) \le min\{Prob(A), Prob(B)\}$. Thus, the membership probability that r belongs to p falls into the interval $[max\{0, \dots max\{0, max\{0, Prob(r \in R(p[X_1])) + Prob(r \in R(p[X_2])) - 1\} + Prob(r \in R(p[X_3])) - 1\} \dots + Prob(r \in R(p[X_n])) - 1\}, min\{Prob(r \in R(p[X_1])), \dots, Prob(r \in R(p[X_n]))\}]$.

3. For any coordinate C' and its parent coordinate C at axis X_i, $\beta_{ri}(C')$ is defined as the membership probability that r belongs to C' under the condition that resource r belongs to C, i.e., $\beta_{ri}(C') = Prob(r\in R(C') \mid r\in R(C))$. Since C' is a child of C, $Prob(r\in R(C')) = Prob(r\in R(C') \wedge r\in R(C))$ holds. Since $Prob(r\in R(C')\wedge r\in R(C)) = Prob(r\in R(C)) \times Prob(r\in R(C') \mid r\in R(C))$, we have $Prob(r\in R(C')) = \beta_{ri}(C) \times \beta_{ri}(C')$. So the probability that r belongs to $R(C')$ is $\beta_{ri}(C) \times \beta_{ri}(C')$.

Above formulas can be easily proved. The following are two examples. In Fig. 4, the probability that r belongs to axis A is $Prob(r\in R(A)) \in [max\{\beta_{r\text{-}A}(a_1), \beta_{r\text{-}A}(a_2), \beta_{r\text{-}A}(a_3)\}, min\{1, \beta_{r\text{-}A}(a_1) + \beta_{r\text{-}A}(a_2) + \beta_{r\text{-}A}(a_3)\}] = [0.3, 0.6]$. The probability that r belongs to point $p(a_2, b_2)$ is $Prob(r\in R(p(a_2, b_2))) \in [max\{0, \beta_{r\text{-}A}(a_2) + \beta_{r\text{-}B}(b_2) - 1\}, min\{\beta_{r\text{-}A}(a_2), \beta_{r\text{-}B}(b_2)\}] = [0, 0.3]$. In Fig. 5, the axis *Area* is used to classify scientific publications according to their areas. In the classification hierarchy of coordinate *CS* (*Computer Science*) at axis *Area*, *DB* (*DataBase*) is a subclass of *CS* and *RDB* (Relational DataBase) is a subclass of *DB*. For resource r and its membership probability function β_r, $\beta_r(RDB)$ represents the following conditional probability: $\beta_r(RDB) = Prob(r\in R(RDB) \mid r\in R(DB))$. Similarly, $\beta_r(DB) = Prob(r\in R(DB) \mid r\in R(CS))$. Since *DB* is a subclass of *CS*, the probability that r belongs to *DB* is $Prob(r\in R(DB)) = Prob(r\in R(DB)\wedge r\in R(CS)) = Prob(r\in R(CS)) \times Prob(r\in R(DB) \mid r\in R(CS)) = \beta_r(CS) \times \beta_r(DB)$. In fact, the probability that r belongs to a sub-coordinate is the multiplication of all the conditional probabilities along the path from the top-level coordinate to this sub-coordinate. So the probability that r belongs to *RDB* is $\beta_r(CS) \times \beta_r(DB) \times \beta_r(RDB)$.

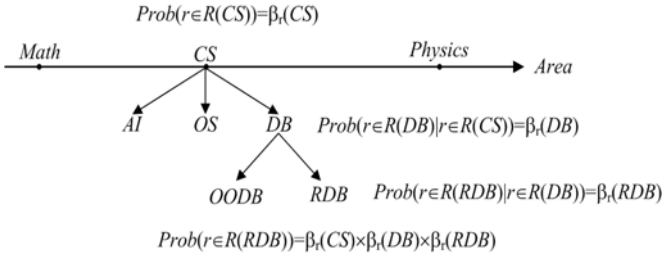


Fig. 5. Conditional probabilities of coordinate hierarchy.

3.2 Normal Forms of the Probabilistic Resource Space Model

Dependence between categories often makes it difficult to correctly classify resources, and it also affects the precision of calculating the membership probabilities that resources belong to points or axes. Normalization of probabilistic resource spaces can help eliminate this dependence.

The 1NF of RSM is used to eliminate the redundancy caused by the duplication of coordinates. It also applies to the P-RSM.

In the resource space satisfying 1NF, dependence relation may exist between coordinates at the same axis. This dependence makes users hard to select an appropriate coordinate when storing and retrieving resources. One solution is to establish a link between dependent coordinates so that the dependent coordinates can be accessed at the same time. Based on this idea, we can define 1.5NF, which can help users to determine the appropriate probability that a resource belongs to a coordinate when operating resource space.

Definition 2. A 1.5 NF resource space is a 1NF resource space, and the co-access links between all interdependent coordinates at the same axis have been established.

The 2NF of RSM is to eliminate the dependency between coordinates so that a resource can be accurately located according to coordinates. The following is the definition of the 2NF probabilistic resource space.

Definition 3. A 2NF probabilistic resource space $RS(X_1, \ldots, X_n)$ is a 1NF resource space and for any pair of coordinates C and C' at X_i ($1\le i\le n$), a resource r satisfies $Prob(r\in R(C) \wedge r\in R(C'))=0$, where $R(C)$ represents the resources that C represents.

For a 2NF probabilistic resource space, there is no such a coordinate that simultaneously belongs to two parent coordinates. But a resource can belong to different coordinates at different probabilities. This is in line with the fact that the resource space designers have clear classification in mind on the resources to be managed while users are sometimes unclear in determining the category of a resource.

Lemma 2. Let X be an axis of a 2NF resource space and $R(X)$ be the resources represented by X, $Prob(r\in R(X)) = \sum_{C\in X} Prob(r \in R(C)) \le 1$ holds.

The approach to transforming 1NF into 2NF will be discussed in section 7.2. The independence between coordinates in 2NF resource space implies the following lemma.

Lemma 3. For a 2NF resource space, there is no common superclass or subclass between coordinates at the same axis.

A 2NF probability resource space can become a 3NF probability resource space by tightly coupling its axes.

Definition 4. Let $X=\{C_1, \ldots, C_m\}$ be an axis, and C' be a coordinate at another axis X', we say that X is a fine classification on C' (denoted as C'/X) if and only if for any resource r:

1. $Prob((r\in R(C')\cap R(C_i)) \wedge (r\in R(C')\cap R(C_j))) = 0$, for $1\le i\ne j\le m$; and,
2. $Prob(r\in R(C')) = \sum_{C\in X} Prob(r \in R(C') \mid r \in R(C)) \times Prob(r \in R(C))$ hold.

According to definition 4 and the total probability theorem, axis X forms a classification on coordinate C' if and only if the probability that resource r belongs to $R(C')$ can be classified into the probabilities that r belongs to $R(C')\cap R(C_1)$, $R(C')\cap R(C_2)$, ..., and $R(C')\cap R(C_m)$ respectively.

Definition 5. A 3NF probabilistic resource space $RS(X_1, \ldots, X_n)$ is a 2NF probabilistic resource space, and for any two different axes X_i and X_j ($1 \le i \ne j \le n$) in RS, $X_i \perp X_j$ holds.

A 3NF probabilistic resource space satisfies the following theorems.

Theorem 1. Let $RS(X_1, \ldots, X_n)$ be a 3NF probabilistic resource space. For any two axes X_i and X_j ($1 \le i, j \le n$) and resource r in RS, $\sum_{C \in Xi} Prob(r \in R(C)) = \sum_{C' \in Xj} Prob(r \in R(C'))$ holds.

Proof. Since RS satisfies 3NF, coordinate C at axis X_i can be classified by axis X_j. So $Prob(r \in R(C)) = \sum_{C' \in Xj} (Prob(r \in R(C) \wedge r \in R(C')))$ holds. Thus, we can get $\sum_{C \in Xi} Prob(r \in R(C)) = \sum_{C \in Xi} \sum_{C' \in Xj} (Prob(r \in R(C) \wedge r \in R(C')))$. On the other hand, coordinate C' at axis X_j can be classified by axis X_i. So, $Prob(r \in R(C')) = \sum_{C \in Xi} (Prob(r \in R(C') \wedge r \in R(C)))$ holds. Thus, we can get $\sum_{C \in Xj} Prob(r \in R(C')) = \sum_{C' \in Xj} \sum_{C \in Xi} (Prob(r \in R(C') \wedge r \in R(C)))$. Therefore $\sum_{C \in Xi} Prob(r \in R(C)) = \sum_{C' \in Xj} Prob(r \in R(C'))$ holds. □

Theorem 1 indicates that for any two axes X_i and X_j of a 3NF probabilistic resource space, the probability that resource r belongs to X_i is equal to the probability that r belongs to X_j.

Theorem 2. Let $RS(X_1, \ldots, X_n)$ be a 2NF probabilistic resource space. For any coordinate C at axis X_i ($1 \le i \le n$), $Prob(r \in R(C)) \ge \sum_{p[Xi]=C} Prob(r \in R(p))$ holds, where p is a point in RS and $p[X_i]$ is the projection of p at axis X_i. If RS satisfies 3NF, we have $Prob(r \in R(C)) = \sum_{p[Xi]=C} Prob(r \in R(p))$.

Proof. Let T be the union of all points whose projections on X_i are C. So $R(T) = R(C) \cap \bigcap_{1 \le j \ne i \le n} \bigcup_{C_j \in X_j} R(C_j)$. Since resource space RS satisfies 2NF, any two points in RS are independent of each other. Thus, we have $Prob(r \in R(T)) = \sum_{p[Xi]=C} Prob(r \in R(p))$. So $\sum_{p[Xi]=C} Prob(r \in R(p)) = Prob(r \in (R(C) \cap \bigcap_{1 \le j \ne i \le n} \bigcup_{C_j \in X_j} R(C_j)))$ holds. $Prob(r \in R(C)) \ge \sum_{p[Xi]=C} Prob(r \in R(p))$ holds. On the other hand, $Prob(r \in R(T)) = Prob(r \in (R(C) \cap \bigcap_{1 \le j \ne i \le n} \bigcup_{C_j \in X_j} R(C_j))) = Prob(r \in (R(C) \cap \bigcap_{1 \le j \ne i \le n} R(X_j)))$ holds. If resource space RS satisfies 3NF, axis X_j ($1 \le j \ne i \le n$) can forms a classification on coordinate C. We can get that $R(C)$ is a subclass of $R(X_j)$. So $Prob(r \in R(T)) = Prob(r \in R(C) \cap \bigcap_{1 \le j \ne i \le n} R(X_j)) = Prob(r \in R(C))$ holds. Therefore $Prob(r \in R(C)) = \sum_{p[Xi]=C} Prob(r \in R(p))$ holds. □

Theorem 2 plays an important role in maintaining the probability values when inserting and updating resources. It will be used in section 3.3 and section 5.2.

3.3 Membership Probability on Points

To specify the membership probabilities that a resource belongs to points is another important issue.

For point p in resource space $RS(X_1, \ldots, X_n)$, $R(p)$ can be represented as $R(p[X_1]) \cap \ldots \cap R(p[X_n])$. So the probability that resource r belongs to p is the probability of the complex event that r belongs to $R(p[X_1])$, $R(p[X_2])$, …, and $R(p[X_n])$ simultaneously. The interval of the membership probability that r belongs to p can be calculated as follows.

For 1NF probabilistic resource space, the probability that r belongs to p falls into the following interval:

$[max\{0, \ldots max\{0, max\{0, Prob(r \in R(p[X_1])) + Prob(r \in R(p[X_2])) - 1\} + Prob(r \in R(p[X_3])) - 1\} \ldots + Prob(r \in R(p[X_n])) - 1\}, min\{Prob(r \in R(p[X_1])), \ldots, Prob(r \in R(p[X_n]))\}]$.

For 2NF probabilistic resource space $RS(X_1, \ldots, X_n)$, according to lemma 2 and theorem 2, the interval for the membership probability that r belongs to p can be obtained by resolving the following linear programming problem:

Object function: $Prob(r \in R(p))$;

Subject to:

1. $\sum_{C \in Xi} Prob(r \in R(C)) \le 1$, $1 \le i \le n$;
2. $\sum_{p[Xi]=C} Prob(r \in R(p)) \le Prob(r \in R(C))$, for any coordinate C at axis X_i, $1 \le i \le n$;
3. $\sum_{p' \in RS} Prob(r \in R(p')) \ge max\{0, \ldots max\{0, max\{0, Prob(r \in R(X_1)) + Prob(r \in R(X_2)) - 1\} + Prob(r \in R(X_3)) - 1\} \ldots + Prob(r \in R(X_n)) - 1\}$; and,
4. $L_i \le Prob(r \in R(p_i)) \le U_i$, for any point p_i in RS, where L_i and U_i are respectively the lower bound and the upper bound of the membership probability that r belongs to p_i set by users. If they do not set explicitly, the default value of L_i is 0, and the default value of U_i is 1.

If RS satisfies 3NF, then the linear programming problem is as follows:

Object function: $Prob(r \in R(p))$;

Subject to:

1. $\sum_{C \in Xi} Prob(r \in R(C)) \le 1$, $1 \le i \le n$;

2. $\sum_{p[Xi]=C} Prob(r \in R(p)) = Prob(r \in R(C))$, for any coordinate C at axis X_i ($1 \le i \le n$);
3. $\sum_{C \in Xi} Prob(r \in R(C)) = \sum_{C' \in Xj} Prob(r \in R(C'))$, for $1 \le i \ne j \le n$
4. $L_i \le Prob(r \in R(p_i)) \le U_i$, for any point p_i in *RS*.

As *RS* satisfies 3NF, item 3 of above constraint will be satisfied if both item 1 and item 2 are satisfied.

Theorem 3. For a probabilistic resource space *RS* that satisfies 1NF, 2NF or 3NF, the membership probability interval that resource *r* belongs to point *p* can be obtained in polynomial time of the number of points in *RS*.

Proof. For resource space $RS(X_1, \ldots, X_n)$, if *RS* satisfies 1NF, then the membership probability interval that resource *r* belongs to point *p* is [*max*{0, ... *max*{0, *max*{0, $Prob(r \in R(p[X_1]))$ + $Prob(r \in R(p[X_2]))$ − 1} + $Prob(r \in R(p[X_3]))$ − 1} … + $Prob(r \in R(p[X_n]))$ − 1}, *min*{$Prob(r \in R(p[X_1]))$, …, $Prob(r \in R(p[X_n]))$}]. It is obvious that both the lower bound and the upper bound can be computed by $n-1$ steps.

If *RS* satisfies 2NF, identifying the membership probability intervals that resource *r* belongs to points can be converted to the following linear program problem *LP*:

Object function:

$Prob(r \in R(p_j))$, for any point p_j in *RS* ;

Subject to:

1. $\sum_{C \in Xi} Prob(r \in R(C)) \le 1$, $1 \le i \le n$;
2. $\sum_{p[Xi]=C} Prob(r \in R(p)) \le Prob(r \in R(C))$, for any coordinate C at axis X_i, $1 \le i \le n$;
3. $\sum_{p' \in RS} Prob(r \in R(p')) \ge max\{0, \ldots max\{0, max\{0, Prob(r \in R(X_1)) + Prob(r \in R(X_2)) - 1\} + Prob(r \in R(X_3)) - 1\} \ldots + Prob(r \in R(X_n)) - 1\}$; and,
4. $L_i \le Prob(r \in R(p_i)) \le U_i$, for any point p_i in *RS*.

In *LP*, both $Prob(r \in R(C))$ and $Prob(r \in R(X_i))$ ($1 \le i \le n$) are constants and $Prob(r \in R(p_j))$ are variables. It is obvious that both the number of variables and the number of inequalities in *LP* are in polynomial with the number of points in *RS*. Since a linear programming problem is tractable in polynomial time, the membership probability interval that any resource *r* belongs to point *p* can be obtained in polynomial time of the number of points in *RS*.

Similarly, we can prove that when *RS* is in 3NF, the membership probability interval that any resource *r* belongs to point *p* can also be calculated in polynomial time of the number of points in *RS*. □

4 Operations of Probabilistic Resource Space Model

4.1. Point Query

The first query approach of the P-RSM is point query. The result of a point query is a set of points, each of which contains a set of resources with membership probability.

For a resource space *RS*, the point query operation is for selecting the desirable points according to the given restrictions. This type of query can be denoted as $\sigma_p(RS)=\{p \mid p \in RS \wedge F_p(p)\}$, where F_p is a logical expression. The basic form of F_p is: $p_m[X_i]\ \theta\ Y$, where Y may be $p_n[X_j]$ or just a noun and noun phrase defined in domain ontology, p_m and p_n are points and θ represents $=, \ne, <, \le, \ge$ or $>$. F_p is usually a logical combination of the basic forms by using $\wedge$, $\vee$ and $\neg$.

The P-RSM uses the following statement to support point queries. The *conditional expression* in this statement is the logical combination of restrictions on the projections of points on axes.

SELECT POINT *p* FROM $RS(X_1, \ldots, X_n)$
[WHERE <*conditional expression*>]

The point query supports query of multiple points. Take Fig. 6 for example, to query all resources in points $p_1(a_2, b_1, c_1)$ and $p_2(a_2, b_2, c_1)$, user should use the following logical expression $\sigma_p(RS)=\{p \mid p \in RS \wedge p[A]=a_2 \wedge p[C]=c_1 \wedge (p[B]=b_1 \vee p[B]=b_2)\}$ and the following point query statement:

SELECT POINT *p* FROM *RS*(*A*, *B*, *C*)
WHERE $p[A]=a_2$ AND $p[C]=c_1$ AND ($p[B]=b_1$ OR $p[B]=b_2$)

As the consequence of the query, resources with membership probabilities belonging to points $p_1(a_2, b_1, c_1)$ and $p_2(a_2, b_2, c_1)$ will be returned.

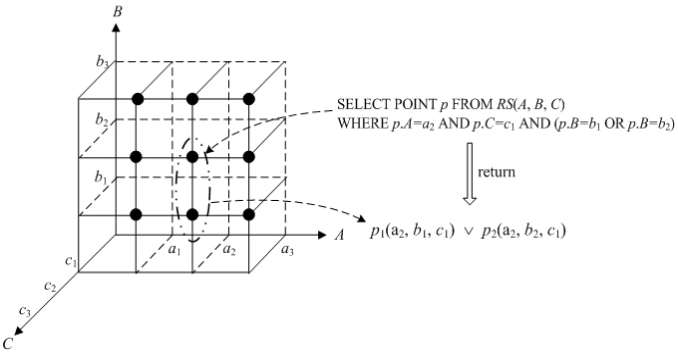


Fig. 6. An example of point query.

4.2 Resource Modification

In the original RSM, before a resource *r* can be inserted into a resource space *RS*, we have to identify the coordinates that *r* belongs to each axis in *RS*.

Take Fig. 7 for example, the resource space *RS*(*Classes*, *Courses*, *Gender*) is used to manage student information according to their *classes*, *courses* and *gender*. Resource *r* can be inserted into point (*Database*, C_2, *Male*) if it belongs

to *Database* at axis *Courses*, C_2 at axis *Classes* and *Male* at axis *Gender*.

From the perspective of probability, *r*(*Courses*=*Database*, *Classes*=C_2, *Gender*=*Male*) implies that the membership probability functions of resource *r* at axes *Courses*, *Classes* and *Gender* are $\beta_{\text{r-Courses}}$, $\beta_{\text{r-Classes}}$ and $\beta_{\text{r-Gender}}$ respectively such that:

1. $\beta_{\text{r-Courses}}(Math)=0$, $\beta_{\text{r-Courses}}(Operating\ System)=0$, and $\beta_{\text{r-Courses}}(Database)=1$.
2. $\beta_{\text{r-Classes}}(C_1)=0$, $\beta_{\text{r-Classes}}(C_2)=1$, and $\beta_{\text{r-Classes}}(C_3)=0$.
3. $\beta_{\text{r-Gender}}(Male)=1$ and $\beta_{\text{r-Gender}}(Female)=0$.

The process of inserting a resource into a probabilistic resource space is the same as the original resource space except that the membership probability functions in the P-RSM can take value within [0, 1].

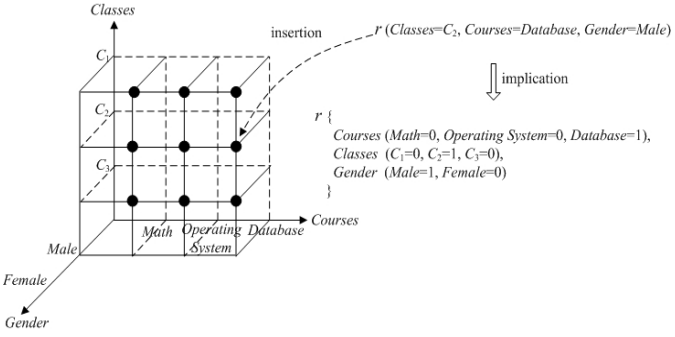


Fig. 7. Insert a resource into a resource space.

The following is the general insertion statement for inserting a resource *r* into a resource space *RS*. β_k is the membership probability function of *r* belonging to coordinate C_k at axis X_k (k=1, 2, …, n).

INSERT *r* ((X_1=<C_1, β_1>, …, X_n=<C_n, β_n>)
INTO *RS*(X_1, …, X_n)

P-RSM also supports the following delete operation and update operation, which is seldom used in the cyber-physical society:

DELETE *r* FROM *RS*
[WHERE <*conditional expression*>]

UPDATE *r*(<β_i, …, β_j>) INTO *RS*(X_i, …, X_j)
[WHERE <*conditional expression*>].

4.3 Operations on Probabilistic Resource Spaces

Join, Disjoin, Merge and Split are four major operations of the original RSM. P-RSM has the following corresponding operations.

1. **Join**. If two resource spaces $RS_1(X_1, …, X_m, Y_1, …, Y_n)$ and $RS_2(Y_1, …, Y_n, Z_1, …, Z_k)$ specify the same type of resources and have n common axes, then they can be put together as one resource space $RS(X_1, …, X_m, Y_1, …, Y_n, Z_1, …, Z_k)$ such that RS_1 and RS_2 share these n common axes and $|RS|=|RS_1| + |RS_2| - n$. For any resource *r* in *RS*, the membership probability functions of *r* at axes X_i ($1\le i\le m$), Y_j ($1\le j\le n$) and Z_h ($1\le h\le k$) are the same as those functions in RS_1 and RS_2.

Let $p(x_1, …, x_m, y_1, …, y_n, z_1, …, z_k)$, $p_1(x_1, …, x_m, y_1, …, y_n)$, and $p_2(y_1, …, y_n, z_1, …, z_k)$ be the points in *RS*, RS_1 and RS_2 respectively. The event that resource *r* belongs to point *p* corresponds to the following two events occur simultaneously: *r* belongs to p_1, and *r* belongs to p_2. If the membership probability interval that *r* belongs to p_1 is [L_1, U_1] and the membership probability interval that *r* belongs to p_2 is [L_2, U_2], then we can obtain the following restriction: $max\{0, L_1+L_2-1\}\le Prob(r\in R(p))\le min\{U_1, U_2\}$. The membership probability interval that *r* belongs to *p* can be calculated as introduced in section 3.

2. **Disjoin**. A resource space $RS(X_1, …, X_m, Y_1, …, Y_n, Z_1, …, Z_k)$ can be separated into two resource spaces $RS_1(X_1, …, X_m, Y_1, …, Y_n)$ and $RS_2(Y_1, …, Y_n, Z_1, …, Z_k)$ that store the same type of resources as that of *RS* such that they have n common axes and $k + m$ different axes, and $|RS|=|RS_1| + |RS_2| - n$. For any resource *r* in RS_1, the membership probability functions of *r* at axes X_i ($1\le i\le m$) and Y_j ($1\le j\le n$) are the same as those functions in *RS*.

For point $p(x_1, …, x_m, y_1, …, y_n)$ in RS_1, let p_i ($1\le i\le k$) be the point in *RS* such that p_i has the same projection at axes X_1, …, X_m, Y_1, …, Y_n as point *p*. Suppose that the membership probability interval that resource *r* belongs to p_i is [L_i, U_i], where $1\le i\le k$. Then, we can obtain the following restrictions:

1. If *RS* only satisfies 1NF, then $max\{L_1, …, L_k\} \le Prob(r\in R(p)) \le min\{1, U_1+ … +U_k\}$ holds;
2. If *RS* only satisfies 2NF, then $L_1+ … + L_k \le Prob(r\in R(p))\le 1$ holds; and,
3. If *RS* satisfies 3NF, then $L_1+ … +L_k \le Prob(r\in R(p)) \le U_1+ … +U_k$ holds.

3. **Merge**. If two resource spaces $RS_1(X_1, …, X_{n-1}, X')$ and $RS_2(X_1, …, X_{n-1}, X'')$ store the same type of resources and satisfy: a) $|RS_1|=|RS_2|=n$; and, b) they have $n-1$ common axes, and there exist two different axes X' and X'' satisfying the merge condition, then they can be merged into one *RS* by retaining the $n-1$ common axes and adding a new axis $X^*=X'\cup X''$. *RS* is called the merge of RS_1 and RS_2, denoted as $RS_1\cup RS_2\Rightarrow RS$, and $|RS|= n$. For any resource *r* in *RS*, the membership probability functions of *r* at axes X_i ($1\le i\le n-1$) are the same as those for RS_1. Let β' and β'' be the membership probability functions of *r* at axes X' and X'' respectively. Then, the membership probability function β of *r* at axis X^* is defined as follows: for any coordinate *C* at axis X^*, $\beta(C)=\beta'(C)$ if *C* is at axis X', otherwise $\beta(C)=\beta''(C)$.

4. **Split**. A resource space $RS(X_1, …, X_{n-1}, X)$ can be *split* into two resource spaces RS_1 and RS_2 that store the resources in *RS* and have $|RS|-1$ common axes by splitting axis *X* into two: X' and X'', such that $X=X' \cup X''$. For any resource *r* in RS_1, the membership probability functions of *r* at axes X_i ($1\le i\le n-1$) are the same as those in *RS*. Let β be the membership probability function of *r* at axis *X*. Then, the membership probability function β' of *r* at axis X' is defined as follows: for any coordinate *C* at axis X', $\beta'(C)=\beta(C)$ holds.

5 PROBABILISTIC INTEGRITY CONSTRAINTS

Just as the integrity constraint in the relational data model [1], the integrity constraint in RSM is to ensure the consistency during operations. It concerns entity integrity, membership integrity, referential integrity, and user-defined integrity [35]. In P-RSM, the meaning of some constraint rules changes and some new rules should be obeyed.

5.1 The Key

As a coordinate system, the RSM supports accurate resource positioning by giving coordinates. However, it is sometimes unnecessary and even arduous to give the coordinates at all axes to identify a point, especially for high-dimensional resource spaces. The *key* is for efficiently locating resources according to some axes. In RSM, a point that does not represent any resource is called a null point. Otherwise, it is called a non-null point. The following is the notion of the *key* in RSM.

Let CK be a subset of axis set $\{X_1, \ldots, X_n\}$, p_1 and p_2 be any two non-null points of resource space $RS(X_1, \ldots, X_n)$. CK is called a *candidate key* of resource space RS if we can derive $p_1[X_i]=p_2[X_i]$, $X_i \in \{X_1,\ldots, X_n\}$ from $p_1[X_j]=p_2[X_j]$, $X_j \in CK$.

In P-RSM, point p is a null point if and only if for any resource r, $Prob(r \in R(p))=0$ holds.

The *key* in the probabilistic resource space is defined as follows.

Definition 6. Let CK be a subset of the axis set $\{X_1, \ldots, X_n\}$, p_1 and p_2 be any two points in resource space $RS(X_1,\ldots, X_n)$ such that $p_1[X_i]=p_2[X_i]$, $X_i \in CK$. CK is called a *candidate key* of resource space RS if $Prob(r_1 \in R(p_1) \wedge r_2 \in R(p_2))=0$ holds for any two resources r_1 and r_2, and there exists an axis X_j such that $X_j \in \{X_1, \ldots, X_n\} - CK$ and $p_1[X_j] \neq p_2[X_j]$.

Above definition implies a kind of resource dependency: If the event that r_1 belongs to p_1 occurs, the probability that r_2 belongs to p_2 is 0, i.e. $Prob(r_2 \in R(p_2) \mid r_1 \in R(p_1)) = 0$, vice versa.

Most previous probabilistic relational data models manage entities one by one and seldom concern the relationship between entities. They usually assume that the uncertainty of one entity is independent of another entity. P-RSM considers some dependency between resources. Semantic links can be established between points to reflect this kind of dependence [38].

The following theorem applies to the situation where the probabilistic events about two resources should not be supposed to be independent of each other.

Theorem 4. Let CK be a candidate key of 3NF resource space $RS(X_1,\ldots, X_n)$ and CK' be a subset of $\{X_1,\ldots, X_n\}$ such that $CK \subset CK'$. Let p_1 and p_2 be two points in RS such that $p_1[X_i]=p_2[X_i]$ ($X_i \in CK$) and $p_1[X_j] \neq p_2[X_j]$ ($X_j \in CK'-CK$). For any two resources r_1 and r_2, the events $r_1 \in \mathop{\mathrm{I}}_{X \in CK' \wedge p_1[X]=C} R(C)$ and $r_2 \in \mathop{\mathrm{I}}_{X \in CK' \wedge p_2[X]=C} R(C)$ are not independent of each other, and $Prob(r_1 \in \mathop{\mathrm{I}}_{X \in CK' \wedge p_1[X]=C} R(C) \wedge r_2 \in \mathop{\mathrm{I}}_{X \in CK' \wedge p_2[X]=C} R(C)) = 0$.

Proof. Suppose that both $Prob(r_1 \in \mathop{\mathrm{I}}_{X \in CK' \wedge p_1[X]=C} R(C)) \neq 0$ and $Prob(r_2 \in \mathop{\mathrm{I}}_{X \in CK' \wedge p_2[X]=C} R(C)) \neq 0$ hold. Since RS satisfies 3NF, both $\mathop{\mathrm{I}}_{X \in CK' \wedge p_1[X]=C} R(C) = \mathop{\mathrm{U}}_{X \in CK' \wedge p_1[X]=p[X]} R(p)$ and $\mathop{\mathrm{I}}_{X \in CK' \wedge p_2[X]=C} R(C) = \mathop{\mathrm{U}}_{X \in CK' \wedge p_2[X]=p'[X]} R(p')$ hold. If $Prob(r_1 \in \mathop{\mathrm{I}}_{X \in CK' \wedge p_1[X]=C} R(C) \wedge r_2 \in \mathop{\mathrm{I}}_{X \in CK' \wedge p_2[X]=C} R(C)) \neq 0$, then there must exist at least two points p_3 and p_4 such that $p_1[X_i]=p_3[X_i]$, $p_2[X_i]=p_4[X_i]$ ($X_i \in CK'$) and $Prob(r_1 \in R(p_3) \wedge r_2 \in R(p_4)) \neq 0$ hold. This contradicts to the fact that CK is a candidate key of RS. So $Prob(r_1 \in \mathop{\mathrm{I}}_{X \in CK' \wedge p_1[X]=C} R(C) \wedge r_2 \in \mathop{\mathrm{I}}_{X \in CK' \wedge p_2[X]=C} R(C)) = 0$ holds. □

5.2 Integrity Constraints in Probabilistic Resource Space Model

Modification of resources may result in inconsistency in resource spaces. P-RSM needs some special integrity constraint rules to deal with the inconsistency.

Since $\beta_{ri}(C)$ represents the probability that resource r belongs to coordinate C, it is reasonable to require $0 \leq \beta_{ri}(C) \leq 1$. For axis X_i, $R(X_i) = \mathop{\mathrm{U}}_{C' \in X_i} R(C')$ holds. If any two coordinates at X_i are independent of each other, $Prob(r \in R(X_i)) = \sum_{C' \in Xi} \beta_{ri}(C')$. So $\sum_{C' \in Xi} \beta_{ri}(C') \leq 1$ holds.

Rule 1. For resource space $RS(X_1, \ldots, X_n)$, let β_{ri} be the membership probabilistic function of resource r at axis X_i, $1 \leq i \leq n$. For any coordinate C at X_i, $0 \leq \beta_{ri}(C) \leq 1$ must hold. If any two coordinates at X_i are independent of each other, then $\sum_{C' \in Xi} \beta_{ri}(C') \leq 1$ holds.

The insertion, modification of resources and merge operations between resource spaces may violate Rule 1.

Rule 2. For resource space $RS(X_1, \ldots, X_n)$ and resource r, let β_{ri} and β_{rj} be the membership probabilistic functions of r at X_i and X_j ($1 \leq i, j \leq n$) respectively. If X_j can form fine-classification on X_i and any two coordinates at X_i are independent of each other, then $\sum_{C \in Xi} \beta_{ri}(C) \leq \sum_{C' \in Xj} \beta_{rj}(C')$ holds. If X_i is orthogonal with X_j, i.e., $X_i \perp X_j$ holds, then $\sum_{C \in Xi} \beta_{ri}(C) = \sum_{C' \in Xj} \beta_{rj}(C')$ holds.

The following is the reason of Rule2. If X_i / X_j holds, we can reach that $R(X_i) \subseteq R(X_j)$ holds. So $Prob(r \in R(X_i)) \leq Prob(r \in R(X_j))$ holds. Since $R(X_i) = \mathop{\mathrm{U}}_{C' \in Xi} R(C')$ and $R(X_j)$

$= \bigcup_{C' \in Xj} R(C')$, both $Prob(r \in R(X_i)) = \sum_{C \in Xi} \beta_{ri}(C)$ and $Prob(r \in R(X_j)) = \sum_{C' \in Xj} \beta_{rj}(C')$ hold. Thus, $\sum_{C \in Xi} \beta_{ri}(C) \le \sum_{C' \in Xj} \beta_{rj}(C')$ holds.

If X_i is in orthogonal with X_j, both X_i/X_j and X_j/X_i hold, then both $\sum_{C \in Xi} \beta_{ri}(C) \le \sum_{C' \in Xj} \beta_{rj}(C')$ and $\sum_{C \in Xi} \beta_{ri}(C) \ge \sum_{C' \in Xj} \beta_{rj}(C')$ hold. So $\sum_{C \in Xi} \beta_{ri}(C) = \sum_{C' \in Xj} \beta_{rj}(C')$ holds.

Rule 3. For any 3NF resource space $RS(X_1, \ldots, X_n)$ and resource r, let β_{ri} be the membership probabilistic function of r at X_i ($1 \le i \le n$). For any coordinate C at X_i and point p in RS, $\sum_{p[Xi]=C} Prob(r \in R(p)) = \beta_{ri}(C)$ holds.

According to theorem 2, in any 3NF resource space, the probability that r belongs to coordinate C can be classified into the points having projection of C on axis X_i, i.e., $Prob(r \in R(C)) = \sum_{p[Xi]=C} Prob(r \in R(p))$ holds. Rule 3 should be checked to make sure the satisfaction of theorem 2 when inserting or updating resources.

So far we have presented the basis of the P-RSM.

6 Analysis

P-RSM shows distinguished characteristics compared with RSM in managing resources.

6.1 Experimental Data and Schemas of resource space

Without loosing generality, our experimental data are the papers collected from the World Wide Web conference from 2001 to 2007. These papers fall into 13 topics such as *Browser & Interfaces*, *Data Mining*, *e-Applications*, *Search*, and *Semantic Web*. To manage these papers, the following two resource space schemas can be designed according to the RSM.

1. 1NF resource space RS_1(*Topics*, *Years*, *Locations*), where *Topics*={*Browser&Interfaces*, *DataMining*, *e-Applications*, *Practice&Experience*, *Performance&Scalability*, *Ubiquitous*, *Search*, *Sercurity&Reliability*, *SemanticWeb*, *WebEngineering*, *XML&WebData*, *WebServices*, *Ontologies*, *E-Learning*, *WebMining*, *Multimedia*}, *Years*={2001, 2002, 2003, 2004, 2005, 2006, 2007} and *Locations*={*Hong Kong*, *Hawaii*, *Budapest*, *New York*, *Chiba*, *Edinburgh*, *Banff*}. Since the *Topics* axis has several coordinates that are not independent of each other, such as *Semantic Web* and *Ontologies*, the resource space schema RS_1 does not satisfies 2NF. Two resource space instances ORS_1 and PRS_1 having the same schema as RS_1 are created for the original RSM and the P-RSM respectively.
2. 2NF resource space RS_2(*Topics*, *Years*, *Locations*), where *Topics*={*Browser & Interfaces*, *DataMining*, *e-Applications*, *Practice & Experience*, *Performance & Scalability*, *Ubiquitous*, *Search*, *Sercurity & Reliability*, *SemanticWeb*, *WebEngineering*, *XML & WebData*, *WebServices*, *Multimedia*}, *Years*={2001, 2002, 2003, 2004, 2005, 2006, 2007} and *Locations*={*Hong Kong*, *Hawaii*, *Budapest*, *New York*, *Chiba*, *Edinburgh*, *Banff* }. Since all coordinates at each axis in the resource space schema RS_2 are independent of each other, RS_2 satisfies 2NF. Two resource space instances ORS_2 and PRS_2 having the same schema as RS_2 are created for RSM and P-RSM respectively.

The membership probability of each paper belongs to each topic is calculated by using the Naïve Bayes model. A keyword vector $\boldsymbol{x}$ based on Boolean model is used to represent paper. For k topics T_1, …, and T_k, the probability $p(T_i|\boldsymbol{x})$ is used to represent the possibility that a given paper belongs to topic T_i. $p(T_i|\boldsymbol{x})$ is evaluated by $p(\boldsymbol{x}|T_i) \times \boldsymbol{P}(T_i)/p(\boldsymbol{x})$, where $\boldsymbol{P}(T_i)$ is the prior probability. The required training samples are the papers published in WWW2002 and WWW2005.

6.2 Effect of Query

Using RSM to manage uncertain resources, users need to judge which coordinates a given resource belongs to with the membership probabilities. Misjudgment will lead to misplacing the resource in the resource space.

To manage uncertain resources in resource space, the following strategies are adopted:

1. For resource r and any axis X of a 2NF RS, select the coordinate C at X such that the membership probability that r belongs to C is the maximum among all the coordinates at X. Then, insert resource r into coordinate C.
2. If RS satisfies only 1NF, select the coordinate C at X such that the membership probability that r belongs to C is the maximum among all the coordinates at X, and then insert r into coordinate C and the coordinate C' as long as C' is not independent of C and the membership probability that r belongs to C' is greater than 0.

Unlike RSM, P-RSM will maintain all the membership probabilities of each paper on each topic. A probabilistic query can be associated with a confidence threshold. For example, "get all papers of which the topic is *search*, and the membership probability is greater than or equal to 0.2".

The following experiments evaluate the recall and precision for querying the resource spaces of the two models. The recall is the ratio of the number of the returned relevant papers to the total number of the relevant papers and the precision is the ratio of the number of the returned relevant papers to the total number of the returned papers.

We refer to the maximum among the membership probabilities of a given paper on each topic as its probability upper bound. According to the probability upper bound, the papers are classified into eight categories: the papers of which probability upper bound is less than or equal to 0.3, 0.4, …, or 1.

The following is the general form of a point query used in the experiment, where α, β and γ are the membership probabilities or the upper bounds that C, C' and C'' belongs to axes *Topics*, *Years* and *Locations* respectively.

SELECT POINT p FROM $RS_i(Topics, Years, Locations)$ WHERE $p[Topics]=(C, \alpha)$ & $p[Years]=(C', \beta)$ & $p[Locations]=(C'', \gamma)$ WITH CONFIDENCE μ.

The first experiment compares the recall and the precision of the two models. Fig. 8 and Fig. 9 plot the average recall and precision of the resource spaces ORS_1, ORS_2 and PRS_1. When querying the probabilistic resource space PRS_1 and setting the confidence threshold as 0.2. The following results can be drawn from the experiment:

1. The probability upper bound can indicate whether the membership probability that a paper belongs to several topics is approximately equal or not. Both the recall and the precision are quite low when the membership probability distribution that a paper belongs to the topics is even. This is due to the fact that it is easier to misjudge when the probabilities that a paper belongs to several topics are almost equal. Both the recall and the precision are gradually improved with the increase of the probability upper bound.
2. Using RSM to manage uncertain information, both the recall and precision of 1NF resource space is better than 2NF resource space. It is mainly because a paper can belong to several topics in the 1NF resource space whereas a paper can belong to only one topic in the 2NF resource space.
3. The probabilistic resource space PRS_1 has better recall and precision than the original resource spaces ORS_1 and ORS_2. It is mainly because the probabilistic resource space can store all the probabilistic information that a paper belongs to several topics regardless of their independence.

The second experiment is to evaluate the impact of confidence threshold on the recall and the precision when querying the probabilistic resource spaces. Fig. 10 indicates the trend in the recall and the precision of the probabilistic resource space PRS_1 with the increase of confidence threshold.

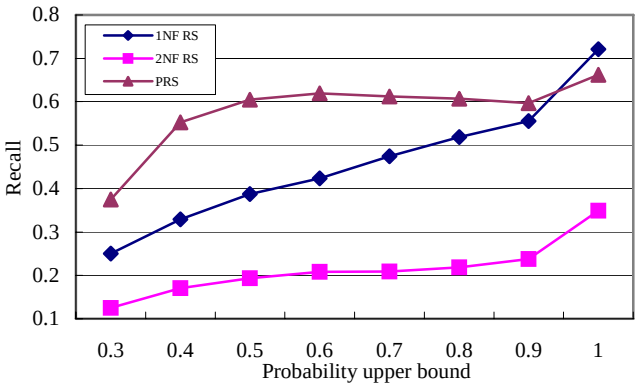


Fig. 8. Recall comparison.

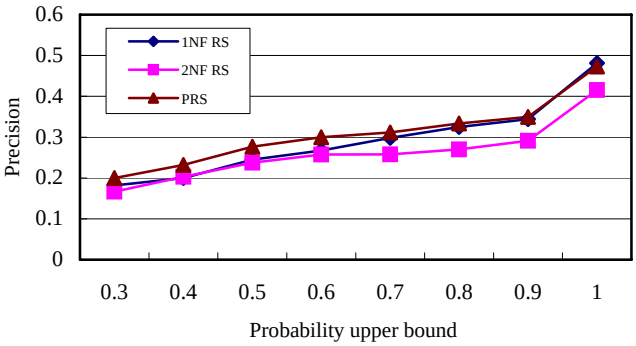


Fig. 9. Precision comparison.

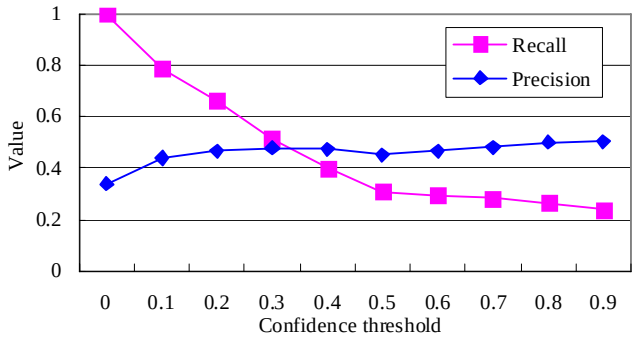


Fig. 10. Recall and precision of a probabilistic resource space.

The following results can be drawn from the experiment:

1. The recall of the probabilistic resource space goes down and the precision of probabilistic resource space goes up with the increase of the confidence threshold. When querying the probabilistic resource space, only the papers for which the membership probability is equal to or larger than the confidence threshold can be returned. Thus, the number of returned relevant papers goes down and the total number of the returned papers goes down more.
2. Theoretically, the recall of the probabilistic resource space will be 100% when the confidence threshold is 0. This is because if the confidence threshold is 0, all the papers probably belonging to a topic will be returned. On the other hand, the 100% recall is due to the trad-off of the low precision.

6.3 Resource Distribution

This experiment is to know how resources are distributed in RSM and P-RSM. The formula $\sqrt{\sum_{1\le i\le n}(|p_i|-m/n)^2}$ is used to evaluate the distribution of resources, where m is the total number of resources to be managed, n is the total number of points in a resource space and $|p_i|$ is the number of resources in point p_i.

Fig. 11 compares the paper distributions in the resource spaces ORS_1, ORS_2, PRS_1 and PRS_2. The following results can be drawn from this experiment:

1. Resources are distributed more evenly in the probabilistic resource space than in the resource space.
2. Normal forms have more impact on resource distribution in the resource spaces than the distribution

in the probabilistic resource spaces. This is because a resource in the probabilistic resource space can be inserted into a point if its membership probability belonging to this point is larger than 0. But in the RSM, a resource cannot be simultaneously inserted into two points that are independent of each other.

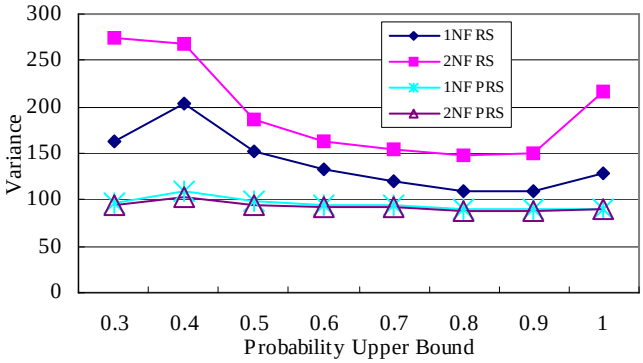


Fig. 11. Resource distribution comparison.

7. EXTENSIONS

7.1 Satisfactory Constraints

A resource space that does not satisfy 2NF is useful in applications since sometimes dependent coordinates cannot be avoided in applications due to the understanding gap between the resource space creater, the users who input resources, the users who retrieve resources, and the authors of the resources. To ensure effective resource operations, constraints can be set according to applications.

Constraint 1. For any resource r in the resource space, there exists an axis $X=(C_1, \ldots, C_n)$ such that $\eta \le \sum_{C_k \in X} \beta_{r\text{-}X}(C_k) \le 1$, where $0<\eta \le 1$ is the *minimum satisfactory degree* in an application.

The minimum satisfactory degree reflects users' confidence in determining the category of resources when inputting resources and the users' demand on the accuracy of retrieving resources. η can be adapted by users according to the change of requirements. Coordinate C_k is selected by users or applications. *Constraint 1 ensures that a resource in a probabilistic resource space can be retrieved at certain satisfactory degree from at least one axis.*

Constraint 2. For any resource r in the resource space, constraint 1 holds for every axis.

Constraint 2 ensures that a resource in a probabilistic resource space can be retrieved from every axis at the minimum satisfactory degree.

P-RSM allows one resource to have multiple indexes. As the consequence, operations on the resource space need to search all coordinates. This is a trade-off between flexibility and complexity.

7.2 Transforming 1NF into 2NF

A 1NF resource space can be transformed into a 2NF or a 3NF resource space by merging the inter-dependent coordinates at one axis into a complex coordinate. The new coordinate should represent the original coordinates.

Let C_1 and C_2 be coordinates at the same axis, and C be the new coordinate representing C_1 and C_2, the coordinate merge operation concerns the following three cases:

1. $R(C) \supseteq R(C_1) \cup R(C_2)$, i.e., C expands the semantic range covered by C_1 and C_2 to represent more resources. In this case, $Prob(r \in R(C)) \ge Prob(r \in R(C_1)) + Prob(r \in R(C_2))$ for resource r.
2. $R(C) \subseteq R(C_1) \cup R(C_2)$, i.e., C shrinks the semantic range covered by C_1 and C_2 to represent less resources. In this case, $Prob(r \in R(C)) \le Prob(r \in R(C_1)) + Prob(r \in R(C_2))$.
3. $R(C)=R(C_1) \cup R(C_2)$, i.e., C preserves the semantic range covered by C_1 and C_2. This case happens if the merge operation neither deletes nor increases any subclass. In this case, $Prob(r \in R(C))=Prob(r \in R(C_1)) + Prob(r \in R(C_2))$.

A 1NF resource space can become a 2NF resource space after merging all interdependent coordinates at every axis. Two independent coordinates can also be merged into one complex coordinate for such requirements as reducing null points and keeping balance between the capacities of coordinates in representing resources.

Lemma 4. A resource space will keep its normal form after merging two coordinates C_1 and C_2 at one axis into one complex coordinate C such that $R(C)=R(C_1) \cup R(C_2)$.

The above lemma is true for 1NF, 1.5NF, and 2NF probabilistic resource spaces. For 3NF probabilistic resource space, the merge operation preserves classification and it does not change the orthogonality of the resource space. On the other hand, the total resources represented by the axis keeps unchanging after carrying out the classification-preserve coordinate merge, so the merge operation keeps the 3NF according to the criterion (lemma 1) introduced in [36].

The merge solution is feasible when the scale of resources represented by the new coordinate is appropriate for effective resource retrieval. A basic criterion is that the resources in a resource space are evenly distributed in general. When the solution is not applicable, 1.5NF can be adopted to regulate the resource space.

7.3 Incorporating Semantic Link into RSM

Humans wave various semantic link networks during lifetime. The above discussion has mentioned two types of semantic links: *co-access* link between coordinates, and *probability dependence* link between points. More types of semantic links like *citation* link between papers can be established between resources [40]. Generally, a semantic link $A \text{—} \gamma \rightarrow B$ represents the relation γ between semantic nodes A and B. Relevant semantic links construct a semantic link network denoted as $<N, L, Rules>$, where N, L and *Rules* are sets of semantic nodes, semantic links, and rules for reasoning on semantic links.

Different from static graph, a semantic link network is dynamic in nature due to its reasoning ability. Semantic communities will emerge and changed during the evolution

of the network [38].

A semantic node can be anything: a class or an instance. A semantic link can link resources, coordinates, and points in resource spaces. A resource space organizes resources by multi-dimensional classification according to the contents of resources, and it supports generalization and specialization on classifications.

The *complex semantic space model* integrating the resource space model and the semantic link network model can be represented as follows:

<*RSS*, *L*, *Rules*, *Ontology*, *Operations*>, where

1. *RSS* is a set of resource spaces. In every resource space, each coordinate has a weight—the function of the number of resources it specifies and the times of being accessed.
2. *L* is a set of semantic links between resource spaces, between points, between axes, between coordinates, or between resources.
3. *Rules* consists of three parts: *reasoning rules* for deriving semantic links; *influence rules* for reflecting the influence of operating semantic link network on the resource spaces and the influence of operating resource space on the semantic link network; and, *operation rules* for regulating operations.
4. *Ontology* is a class hierarchy that explains the coordinate hierarchies in the resource spaces and the semantic link network.
5. *Operations* includes the operations on the resource space and the operations on the semantic links.

In addition to the faceted browsing or navigation, the integrated model provides an SQL-like query language for functioning services. The following are three examples of the query:

SELECT POINT *p'* FROM *RS*
WHERE *p'* links *p* [WITH RELATION *r*].

SELECT POINT * FROM *RS*
WHERE * links *p* [WITH RELATION *r*].

SELECT POINT * FROM *RS*
WHERE * links *p* [WITH RELATION *r*]
[AFTER/BEFORE REASONING].

The complex semantic space model can reflect not only the classification on resources but also the linkage, reasoning, and influence between resources [41].

The complex semantic space model can also be the mental model for recognition, understanding and interaction [39]. Fig.12 depicts the scenario of incorporating semantic link into resource space. The dotted arrows represent the inter-coordinate semantic links and the inter-point semantic links. Users can select either the link style or the classification style as the main operation interface of the integrated model.

Semantic node can be extended to represent physical object, human, event, energy and thought in the cyber-physical society [40][42]. Probabilistic semantic links A—(γ, pr)→B can be established to reflect uncertain relation or influence γ between points, between coordinates, and between semantic nodes at probability pr.

Various uncertain interactions and the probabilistic semantic link network's self-reasoning mechanisms evolve the network semantics. The effect of semantic networking and the abstraction ability of resource space cooperate with and influence each other to evolve the structure of the cyber-physical society.

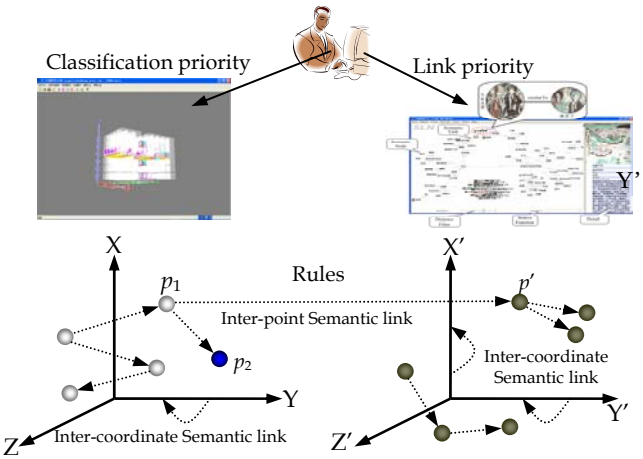


Fig. 12. The complex semantic space model integrating the resource space model and the semantic link network model.

7.4 Automatically Detecting and Uploading Resources into Resource Space

Just like appending data to database, the original RSM requires users to upload resources into the resource space through interface. It is significant to explore the approach to automatically detecting and uploading resources into the resource space. It is feasible if the patterns of resources can be found. A pattern reflects the classification of resources from a certain facet.

Discovering patterns in resources is to find an appropriate classification method. Different types of resources may need different classification methods. Existing frequent pattern discovery, schema matching, text clustering, and community discovery approaches can be helpful references [16, 23, 26, 34, 38]. A fundamental issue is to define appropriate distances on various resources, e.g., the Euclidean distance between physical objects, the distance between concepts in ontology, and the distance between vectors of texts.

Fig. 13 depicts the approach to automatically uploading resources into a probabilistic resource space. According to the current classification in the mental space, the user needs to outline the Classification Tree (denoted as *CT*) for each axis in the cyber resource space, where the higher layer consists of more general classifications, and the lower layer consists of more specific classifications. A classification tree generally takes the following form: $CT = C|(CT, \ldots, CT)$, where C represents a class.

The resource space can automatically detect and upload resources by matching the pattern in resources and the classification trees of the resource space. The following is the general approach.

1. Search the resources in the cyber space according to the classification trees of the resource space. Form the

candidate resource set by merging the search result with the temporal set (initial state is empty).
2. Discover the resource pattern in the candidate resource set according to the classification trees of the resource space and the given matching degree. Put the resources without clear pattern into the temporal set.
3. For each cluster in the pattern, calculate its membership distribution on each axis.
4. The membership distribution of the cluster on the points of the resource space can be obtained by unifying the membership distributions of the clusters on the *CT*s of every axis according to section 3.3.
5. Upload the resources in each cluster into the corresponding point in the resource space, and remove these resources from the candidate resource set.
6. Repeat from step 1 until the candidate resource set does not change.

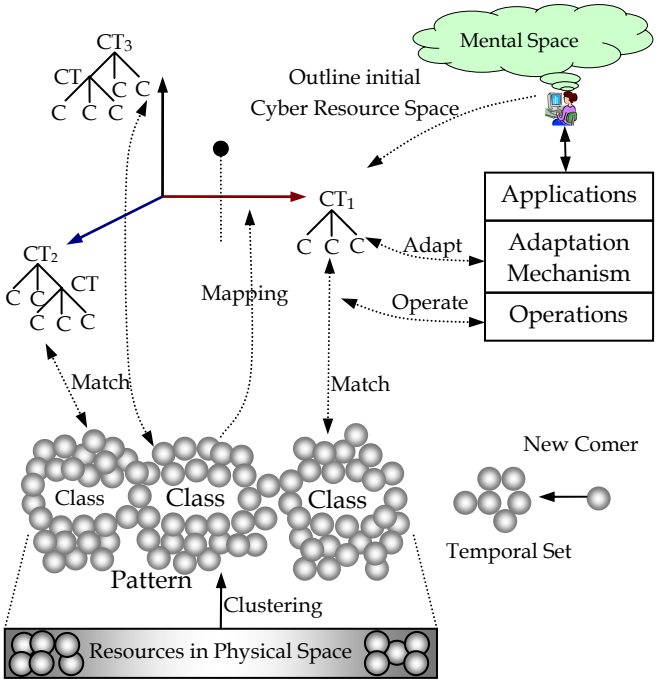


Fig. 13. Automatically detecting and uploading resources into a probabilistic resource space.

During long-term use of the resource space, the original classifications of the resource space may not be able to reflect the up-to-date classifications due to the following causes [7]:

1. Classifications in human mind evolve with continuous co-experience in multiple spaces. New classes will be generated, and the existing classes will be merged or separated from time to time.
2. Patterns in resources may change with continuous addition of new resources. Therefore, the classification trees need to be adjusted to reflect the changes.

Adaptation needs to consider the subjective aspect (user's opinion and social opinion), the objective aspect (the patterns in resources), and the principles of resource space. The basic micro adaptation includes the following operations:

1. Remove a coordinate. This will lead to the removal of all its sub-coordinates if it has, and the removal of all the corresponding points and all the resources in the point.
2. Add a coordinate. This will lead to the addition of the corresponding points in the space.

The other operations such as coordinate split and merge operation can be realized by using the above two operations. To implement a smart resource space that can adapt itself, the resource space needs to trace the active points in the user's mental space and previous queries so that adaption is in line with the evolution of user's mental space.

8. Application Examples

8.1 Faceted Search in Cyber-Physical Society

Resources have multi-facet characteristics. Faceted search (browsing or navigation) is to refine search result from multiple facets during search [17][33]. For example, a set of papers can be refined from the facets of topic, region, publisher, or published time. A faceted search consists of a series selection on these facets. The target is gradually approached from different facets with the progress of search. So far, faceted search lacks theory support although some systems have been developed.

P-RSM can be the theory that supports faceted search not only on the web but also in the cyber-physical society. If the resource spaces in the cyber space share some dimensions (axes) with the user mental space, users are easy to find the required resources. The more dimensions the two spaces share, the easier the users find the targets as they are more familiar with the classification of the resources.

Fig. 14 depicts the idea of the faceted search in the cyber-physical society. Users are individuals of the socio space. They can interact with and influence each other in lifetime through socio networks in the socio space while operating the cyber space.

P-RSM supports faceted search with the following advantages:

1. It supports generalization and specialization on classifications by co-naviation in the cyber space and the mental space.
2. It enables the mental space and the socio space to interact with the physical space through the cyber spaces, in addition to direct interaction.
3. It enables users to access relevant class hierarchies of resource space at different probabilities when their mental spaces lack accurate classification about the target.
4. It enables users to explore a large-scale resource set with flexible classification in mind and to obtain multiple interested classes. The process of navigation is also a process of learning from the patterns in resources and from the community opinion on classification.
5. The *key* in P-RSM enables users to operate a resource space according to a part of its axes.

6. An adaptive resource space can adapt itself according to the change in the cyber, physical, mental, and social spaces as well as the interaction between these spaces.

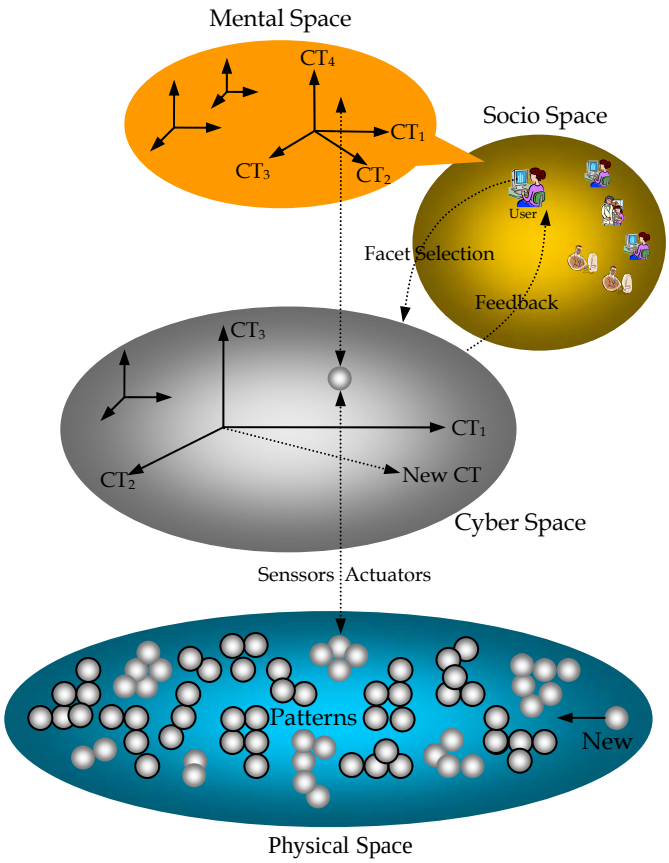


Fig. 14. P-RSM based faceted search in the cyber-physical society.

Faceted search or navigation has the following effects in the mental space:

1. Search motivation is initialized as the effect of the emerging active points in user's mental space. During the search process, classifications in the mental resource space are enriched, completed, refined, or adjusted.
2. The facet selections during faceted search reflect users' opinion on classifications in the mental space. For example, selecting *gender* facet after *degree* facet reflects the following commonsense: the gender dimension is orthogonal with the degree dimension in classifying students.
3. The target class of resources reflects the active point in the user's mental space.
4. The times of repeatedly searching a point in the resource space reflect the active extent of the corresponding point in the mental space.
5. Enriching the mental space by adding coordinates to existing axes.
6. Specializing or generalizing the mental space by adding axis to or removing axis from the space.
7. Tracing all the faceted search processes during long-term use of the resource space reflects the structure of the user's mental space and the active regions. This is very important in exploring minds and in realizing on-demand information services, e.g., recommending the appropriate resources to the user when he/she is thinking about it.

P-RSM based faceted searching plays an important role in service, learning, and science.

8.2 Managing Cyber-Physical-Socio Services

The cyber-physical society links the cyber services (e.g., information services), physical services (e.g., natural resource services), and socio services (e.g., human services) to form the *cyber-physical-socio services*. P-RSM can be used to organize, reflect, and locate the services from multi-dimensions. Fig.15 describes an example of organizing the cyber-physical-socio services.

Hotel services can be reflected by the following 3-dimensional resource space in users' mental spaces according to experience: *RS*(*Quality*, *Location*, *Facility*). The *quality* dimension consists of the following coordinates: *1-star*, *2-star*, *3-star*, *4-star*, and *5-star*. The *location* dimension consists of multiple levels: the first-level coordinates are *countries*, the second-level coordinates are *provinces* or *cities*, the third-level coordinates could be *districts*. The number of levels depends on experience. The facility dimension consists of the following coordinates: *entertainment*, *room*, *food*, and *meeting*. The entertainment coordinate consists of such sub-coordinates as *swimming*, *fitness*, *spa*, and *sauna*. The *room* coordinate consists of such sub-coordinates as *standard*, *moderate*, *superior*, *deluxe*, and *suite*. The food coordinate consists of such sub-coordinates as *Chinese*, *Indian*, and *Western*. If the user concerns price, a 4-dimensional resource space: *RS'*(*Quality*, *Location*, *Facility*, *Price*) can be used.

A resource space with the same structure can be created in the cyber space according to the mental resource space. Information about the physical space such as the features of hotels and the regions can be collected and organized in the space through various sensors. Services of all hotels can be viewed. The cyber resource space enables users to query from the following dimensions: *quality*, *location*, *and facility* according to the active points in the mental space. User's query incites the emerged hotels with relevant quality, facility and location through semantic links.

In the cyber-physical society, users are not isolated. Users can influence each other through socio networks in the socio space [42]. Various resource spaces will emerge in minds and evolve with socio interactions. This is different from previous information systems that need rigid design, and that designers and users are separated to play different roles.

Users can query a set of points according to the emerging active points in the mental space by giving one coordinate or several coordinates at every axis as follows:

SELECT POINT *p*

FROM *RS*(*Quality*, *Location*, *Facility*, *Price*)
WHERE
p[*Quality*]=((*4-star*, *0.7*), (*5-star*, 0.3)) &
p[*Location*]= (*China.Beijing*, *0.7*) &
p[*Facility.room*]=(*standard*, *0.5*) &
p[*Price*]=(200USD, 0.6)).

An item in a point includes: *service content* and *evaluation*. The service content may contain a set of photos of the hotel, descriptive texts, videos, and even the real-time situations (e.g., available room and real-time traffic) sensed through various sensors deployed in the cyber-physical society. The evaluation will lead to *unsatisfied*, *satisfied*, or *excellent* results summarized from the comments of previous guests.

A multi-dimensional resource space can organize and manage cyber-physical-socio services and enables users to retrieve the services from different facets and abstraction levels. The future cyber-physical space can link the hotel services to relevant services such as shopping and sightseeing to provide comprehensive services, and will enable users to virtually present to feel the services.

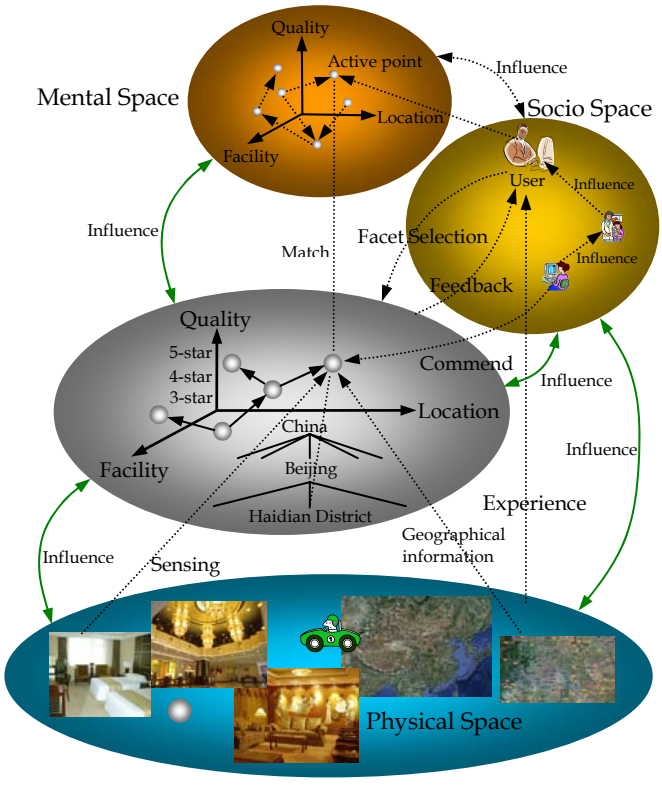


Fig.15. Managing cyber-physical-socio services. Changing the pattern in one space may influence the pattern in the other spaces.

9. Conclusion

The Resource Space Model RSM is a general model for organizing and managing various resources with multi-dimensional classifications. By mapping the RSM into the probabilistic space, this paper establishes the Probabilistic Resource Space Model P-RSM, which can organize and manage various resources with uncertainty by multi-dimensional classifications. The normal forms, operations, and integrity constraints are extended under probabilistic condition. The P-RSM can manage uncertain classifications, support flexible queries, and acquire satisfied query performance. It also enables different resource spaces to be merged or separated with certain normal form and integrity constraint guarantees.

The formation and evolution of natural species can also be viewed from the angle of continuous multi-dimensional classification. Classification is also the most basic method for humans to understand, organize and manage various resources in the cyber space, physical space, socio space, and mental space. P-RSM can be the basic model for organizing various resources in the cyber-physical society.

Humans consciously and subconsciously wave semantic link networks in the cyber, physical, socio and mental spaces, and carry out reasoning while co-experiencing in these spaces in lifetime. A complex semantic space model integrating RSM and SLN reflects the nature of the cyber-physical society. It is suitable for organizing and managing various resources in the cyber-physical society. Since the complex semantic space model reflects the fundamental intelligent mechanisms — *classification*, *link* and *reasoning*, it also supports cyber-physical-socio intelligence [42].

Acknowledgment

Research was supported by the National Science Foundation of China (61075074) and the National Basic Research Program of China (973 Project No. 2003CB317000). We thank the reviewers for their helpful comments on the earlier version of this paper.

References

[1] S. Abiteboul, R. Hull and V. Vianu. Foundations of Databases. Addison-Wesley, Reading, MA, 1995.
[2] S. Abiteboul et al, "Representing and Querying XML with Incomplete Information". *ACM Transactions on Database Systems*, 31(1) (2006) 208-254.
[3] R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval, Addison Wesley, 1999.
[4] D. Barbara, et al. "The Management of Probabilistic Data", *IEEE Transactions on Knowledge and Data Engineering*, 4(5)(1992)437-502.
[5] O. Benjelloun, A.D. Sarma, A.Halevy and J. Widom, "ULDBs: Databases with Uncertainty and Lineage," VLDB2006, pp.953 – 964.
[6] R. Cavallo and M. Pittarelli. "The Theory of Probabilistic Databases," VLDB 1987, pp. 71-81.
[7] C.K.Chang, H.Jiang, H.Ming and K.Oyama, Situ: A Situation-Theoretic Approach to Context-Aware Service Evolution, *IEEE Transactions on Service Computing*, vol.2, no.3, 2009, pp.261-275.
[8] E. F. Codd. "A Relational Model of Data for Large Shared Data Banks," *Communications of the ACM*, 13 (6) (1970) 377-387.
[9] N. Dalvi and D. Suciu. "Efficient Query Evaluation on Probabilistic Databases," VLDB 2004, pp. 864-875.
[10] N. Dalvi and D. Suciu. "Answering Queries from Statistics and Probabilistic Views," VLDB 2005, pp. 805-816.
[11] N. Dalvi and D. Suciu. "Management of probabilistic data: foundations and challenges," SIGMOD 2007, pp. 1-12.
[12] D. Dey and S. Sarkar. "A Probabilistic Relational Model and Algebra," *ACM Transactions on Database Systems*, 21(3) (1996) 339-369.
[13] X.Dong, and A.Halevy, "Index Dataspace," SIGMOD 2007, pp.43-54.

[14] N. Fuhr and T. Rolleke. "A Probabilistic Relational Algebra for the Integration of Information Retrieval and Database Systems," *ACM Transactions on Information Systems*, 15(1) (1997) 32-66.
[15] A.Halevy, M.Franklin and D.Maier, "Principles of Dataspace Systems," In *Proceedings of the 25th ACM Symposium on Principles of Database System*, 2006, 1-9.
[16] J.Han and M.Kambert, *Data Mining: Concepts and Techniques*. Morgan Kaufmann. 2005.
[17] M. A. Hearst, "Clustering Versus Faceted Categories for Information Exploration," *Communications of the ACM*, 49 (4) (2006) 59-61.
[18] E. Hung et al. "Probabilistic interval XML," *ACM Transactions on Computational Logic*, 8(4), (2007) , no. 24..
[19] H. Jiang, H. Lu, W. Wang and J.X. Yu, "XParent: An Efficient RDBMS-Based XML Database System," ICDE, 2002, 335-336.
[20] B. Kimelfeld and Y. Sagiv. "Matching twigs in probabilistic XML," VLDB 2007, pp. 27-38.
[21] L.V.S. Lakshmanan, et al. "ProbView: A Flexible Probabilistic Database System," *ACM Transactions on Database Systems*, 22(3) (1997) 419-469.
[22] M. Keulen et al. "A Probabilistic XML Approach to Data Integration," ICDE 2005, pp 459-470.
[23] J. Madhavan, P.A. Bernstein, A. Doan, and A. Halevy, "Corpus-based schema matching," ICDE 2005, pp. 57- 68.
[24] A. Nierman and H. V. Jagadish. "ProTDB: Probabilistic data in XML," VLDB 2002, pp. 646-657.
[25] C. Ré and D. Suciu. "Materialized views in probabilistic databases: for information exchange and query optimization," VLDB 2007, pp. 51-62.
[26] G.Salton, A.Wong, and C.S.Yang, "A Vector Space Model for Automatic Indexing," *Communications of the ACM*, vol.18, no.11, 1975, pp.613-620.
[27] P. Senellart and S. Abiteboul. "On the Complexity of Managing Probabilistic XML Data," PODS 2007, pp.283-292.
[28] Y. Takahashi, "Fuzzy Database Query Languages and Their Relational Completeness Theorem," *IEEE Transactions on Knowledge and Data Engineering*, 5(1) (1993)122-125.
[29] Y. Tao, et al., "Indexing Multi-dimensional Uncertain Data with Arbitrary Probability Density Functions," VLDB2005, pp.922 – 933.
[30] Q. T. Tho, S. C. Hui, A.C.M. Fong, and T. H. Cao, "Automatic Fuzzy Ontology Generation for Semantic Web," *IEEE Transactions on Knowledge and Data Engineering*, 18(6) (2006) 842-856.
[31] F. Wang, "Fuzzy Supervised Classification of Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, 28(2) (1990) 194-201.
[32] J. Widom. "Trio: A System for Integrated Management of Data, Accuracy, and Lineage," The 2nd *Biennial Conference on Innovative Data Systems Research*, CIDR 2005, pp. 262-276.
[33] K.P. Yee, et al., "Faceted Metadata for Image Search and Browsing," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (2003)401 – 408. Ft. Lauderdale, Florida, USA.
[34] G. Zheng and A. Bouguettaya, Service Mining on the Web, *IEEE Transactions on Service Computing*, vol. 2, no.1. 2009, pp.65-78.
[35] H. Zhuge. The Knowledge Grid, World Scientific, 2004.
[36] H. Zhuge. The Web Resource Space Model, Springer, 2007.
[37] H. Zhuge, Y.Xing and P.Shi, "Resource Space Model, OWL and Database: Mapping and Integration," *ACM Transactions on Internet Technology*, 8/4, 2008.
[38] H. Zhuge, "Communities and Emerging Semantics in Semantic Link Network: Discovery and Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol.21, no.6, 2009, pp. 785-799.
[39] H. Zhuge, "Interactive Semantics, "*Artificial Intelligence*, 174(2010)190-204.
[40] H. Zhuge, Socio-Natural Thought Semantic Link Network: A Method of Semantic Networking in the Cyber Physical Society, Keynote at IEEE AINA 2010, Perth, Australia, April, 20-23, 2010.
[41] H. Zhuge and J. Zhang, "Topological Centrality and Its e-Science Applications," *Journal of the American Society for Information Science and Technology*, 61(9)(2010)1824-1841.
[42] H. Zhuge, "Semantic Linking through Spaces for Cyber-Physical-Socio Intelligence: A Methodology," *Artificial Intelligence*, 2011.

Hai Zhuge is a professor and the chief scientist of the Key Lab of Intelligent Information Processing at Chinese Academy of Sciences' Institute of Computing Technology. He is a joint professor of the Southwest University, China. He was the chief scientist of the Semantic Knowledge Grid Project of the National Basic Research Program of China. He is the pioneer of the cyber-physical-socio Knowledge Grid research. His research concerns the classification-based Resource Space Model (RSM), the self-organized Semantic Link Network model (SLN), the Knowledge Flow, the scalable knowledge grid platform, and the Cyber-Physical Society. Based on RSM and SLN, he established a complex semantic space model as the unified model for managing resources in different spaces and support cyber-physical-socio intelligence. He presented over ten keynotes at international conferences on his innovations. He initiates the International Conference on Semantics, Knowledge and Grids (SKG, www.knowledgegrid.net). He is an associate editor of the Knowledge and Information Systems and the IEEE Intelligent Systems. He serves as the reviewer of several national foundations such as NSF of Austria, NSF of China, SFI of Ireland, and NSF of USA. He is the author of two monographs The Knowledge Grid and The Web Resource Space Model. He was the top scholar in relevant area according to a Journal of Systems and Software assessment report. His publications appeared in AIJ, ACM TOIT, CACM, Computer, IEEE TKDE, IEEE TPDS, IEEE TSC, and JASIST. His work was cited by journals and conferences such as ACM TOSEM, ACM TAAS, IEEE TKDE, IEEE TSE, JPDC, WWW, ICSE, CIKM and ISWC. He received 2007's Wang Xuan Award of China Computer Federation for his fundamental theory of the Knowledge Grid. He is a Distinguished Scientist of the ACM and Senior Member of IEEE. Email: zhuge@ict.ac.cn. Webpage: www.knowledgegrid.net/~h.zhuge.

Yunpeng Xing is a research fellow of the Cyber-Physical-Socio Knowledge Grid Research Group at Institute of Computing Technology in Chinese Academy of Sciences. He was the PhD student of the Research Group. His research interest is R&D of the Resource Space Model. Email: ypxing@kg.ict.ac.cn.