Resource Space Model: A Survey

Jian Zhou

Key Laboratory of Intelligent Information Processing, Institute of Computing Technology
University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing, China
zhoujian17s@ict.ac.cn

Abstract— **This paper surveys research on the Resource Space Model (RSM). RSM is a multi-dimensional, classification-based, content-based and high-level semantic space model for organizing and managing various resources through multi-dimensional abstraction and specialization. RSM has more powerful resource representation ability than traditional resource management model. It has applications not only in resource management and retrieval, but also in other areas, such as automatic text summarization and question answering system.**

I. Introduction

The Resource Space Model (RSM) is a semantic model for organizing, specifying and managing the contents of various resources from multiple dimensions proposed by Professor Hai Zhuge in 2004 [5]. It supports abstraction and specialization on resources. The RSM was proposed to facilitate reasonable storage management resources and to efficiently and quickly retrieval resources.

The Resource Space (RS) in the RSM is like the table in Relational Database Mode (RDBM). Different from ordinary distance space, a RS is a multi-dimensional classification space where the dimensions (axes) and coordinates are discrete [6]. A dimension represents a classification method and a coordinate represents an abstract class of a set of resource in its dimension with a certain classification method. For example, the file system can be regarded as a one-dimensional resource space, the dimension represents file paths and the coordinate represents a class abstracted from the resource it belongs to. Hence, the RSM is a high-level semantic space model. In order to deal with resources more accurately and rigorously, the normal forms theory and integrity constraints of the RSM was proposed [6].

In 2012, the theory of Probabilistic Resource Space Model (P-RSM) was proposed to deal with uncertainty in managing various resources in different spaces of the cyber-physical society [7]. In P-RSM, the probabilistic resource space can specify the membership probability distribution of every resource on all points in the resource space or all coordinates at every axis of the resource space. Based on this idea, the P-RSM can manage uncertain classifications, support flexible queries, and acquire satisfied query performance.

Some work on automatic dimension discovery of RSM have also been completed. Yu et al proposed a framework of automatically constructing multidimensional resource space, combining Latent Dirichlet Allocation and human background knowledge [11]. He proposed a framework of automatically constructing scientific resources space by establishing macro and micro dimensions through the analysis of scientific documents [12]. A new method of dimension discovery based on word frequency is under study.

II. Basis Of Resource Space Model

A. Structure of the Resource Space Model

Resource, resource space, axis and coordinate are the basis semantic elements of the Resource Space Model. RSM consists of two parts: resource and resource space. Resource space consists of a set of axes, denoted as $RS(X_1, X_2, \ldots, X_n)$, while each axis is partitioned by a set of coordinate, denoted as $X_i = (C_{i,1}, C_{i,2}, \ldots, C_{i,m})$. An axis represents a classification method. A coordinate is an abstract class under the classification method represented by its axis. A coordinate can be a single point, also can be a hierarchical data structure such as a classification tree, where the root node represents the overall category of its child nodes and each child node is a subclass of its parent node. The set of resources represented by the coordinate C is denoted as $R(C)$. Obviously, the set of resources represented by the axis X denoted as $\mathrm{R}(X) = \bigcup_{C \in X} R(C)$. A point p in resource space is determined by its projections on all axes, denoted as $p(p[X_1], p[X_2], \ldots, p[X_n])$ where $p[X_i]$ is the projection of p on the axis X_i. The set of resources represented by the point p is denoted as $R(p) = \bigcap_{i=1}^{n} R(p[X_i])$. **Fig** 1 shows an example of a 3-dimensions scientific papers resource space, the point $p(2018, AI, Conference)$ represents scientific papers resources which is related to artificial intelligence in 2018.

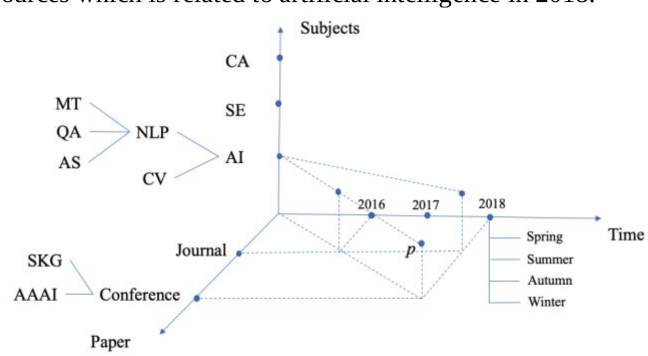

Fig 1: An example of a 3-dimensions web resource space

A resource space schema [6] is a 5-tuple: $\{RS, A, C, S, dom\}$, where:

- RS: the name of the resource space;
- A: the set of axes, denoted as $\mathrm{A} = \{\mathrm{X_i} | 1 \leq \mathrm{i} \leq \mathrm{n}\}$;
- C : the set of coordinates, denoted as $\mathrm{C} = \{\mathrm{C_{ij}} | \mathrm{C_{ij}} \in \mathrm{X_i}, 1 \leq \mathrm{i} \leq \mathrm{n}\}$;
- S: the power set of the domain ontology;

- dom: the mapping from the axis A and coordinates C to S, dom: $A \times C \rightarrow S$, for any axis $X_i = \{C_{i1}, C_{i2}, \ldots, C_{ip}\}$, $dom(X_i, C_{ij}) = V_{ij}$, $V_{ij} \in S$, where $1 \le i \le n$ and $1 \le j \le p$.

B. *The Normal Form of the Resource Space*

Like the RDBM has its normal form, such as 1NF, 2NF, 3NF, BCNF and so on, RSM also has its normal form.

To deal with resources accurately and rigorously, Professor Hai Zhuge defined the following normal forms for the space as criteria for designing proper resource spaces [6].

1) *The first-normal-form resource space (1NF):* there does not exist same name (semantic repetition) between coordinates at any axis in the resource space.

2) *The second-normal-form resource space (2NF):* if there is no intersection between $R(C)$ and $R(C')$, denoted as $R(C) \cap R(C') = \emptyset$, the coordinate C is called independent from another coordinate C'. The 2NF resource space means that on the basis of the 1NF resource space, any two coordinate points on the same axis are independent of each other for all axes in the resource space.

3) *The third-normal-form resource space (3NF):* the 3NF resource space means that on the basis of the 2NF resource space, any two axes of the resource space are orthogonal with each other. From another point of view, the 3NF resource space ensures that each axis in the resource space has the same ability to express resources under the condition of satisfying the 2NF resource space, that is $\forall i \forall j\, R(X_i) = R(X_j), i \neq j \text{ and } i, j \in [1, n]$.

4) *The forth-normal-form resource space (4NF):* the 4NF resource space means that on the basis of the 2NF resource space, any two axes of the resource space are regularly orthogonal with each other. From another point of view, the 4NF resource space means that there are no empty points on the basis of the 3NF resource space, that is $\forall p\, R(p) \neq \emptyset, p \in RS$.

C. *The Operations of the Resource Space*

Professor Hai Zhuge defined four basic operations of resource space [5, 6].

1) *Join Operation:* for two resource spaces RS_1 and RS_2, if the RS_1 and RS_2 can be joined as a new resource space RS, then two conditions need to be met:

- RS_1 and RS_2 specify the same type of resources;
- RS_1 and RS_2 have n $(n \ge 1)$ common axes;

denoted as $RS_1 \cdot RS_2 \Rightarrow RS$. RS, RS_1 and RS_2 share these n common axes and $|RS| = |RS_1| + |RS_2| - n$, where $|RS|$ represents the number of dimensions of the RS.

2) *Disjoin Operation:* the disjoin operation is the reverse of the join operation. A resource space RS can be disjoined into two resource space RS_1 and RS_2 if the following conditions are satisfied:

- RS_1 and RS_2 specify the same type of resources as that of RS;
- RS_1 and RS_2 have n $(1 \le n \le min(|RS_1|, |RS_2|)$ common axes and $|RS| - n$ different axes;
- $|RS| = |RS_1| + |RS_2| - n$;

denoted as $RS \Rightarrow RS_1 \cdot RS_2$.

3) *Merge Operation:* for two resource spaces RS_1 and RS_2, if the RS_1 and RS_2 can be merged as a new resource space RS (denoted as $RS_1 \cup RS_2 \Rightarrow RS$), then following conditions need to be met:

- RS_1 and RS_2 specify the same type of resources;
- RS_1 and RS_2 have the same number of dimensions, $|RS_1| = |RS_2| = n$;
- RS_1 and RS_2 have $n - 1$ common axes;
- RS_1 and RS_2 exist two different axes X_1 and X_2 satisfying the merge condition.

RS_1 and RS_2 can be merged into one resource space RS by retaining the $n - 1$ common axes in the new space and including a new axis $X = X_1 \cup X_2$, so $|RS| = n$.

The last two conditions can be extended as follows:

- RS_1 and RS_2 have $n - \mathrm{k}$ $(1 \le k < n)$ common axes;
- there exists one-one mapping between the rest k axes of the two spaces such that the merge condition can be satisfied.

Then RS_1 and RS_2 can be merged into one resource space RS by retaining the rest $n - k$ common axes in the new space and including k new axes, each of which is the union of the corresponding axes.

4) *Split Operation:* the split operation is the reverse of the merge operation. A resource space RS can be split into two resource space RS_1 and RS_2 if the following conditions are satisfied:

- RS_1 and RS_2 specify the same type of resources as that of RS;
- RS_1 and RS_2 have $|RS| - 1$ common axes by splitting an axis X into two: X' and X''.

denoted as $RS \Rightarrow RS_1 \cup RS_2$.

Also split operation can be extended. The second condition can be extended as follows:

- RS_1 and RS_2 have $n - k$ $(1 \le k < n)$ common axes by splitting each of the k axes into two.

The 1NF, 2NF and 3NF resource space combined by the above operations are still 1NF, 2NF and 3NF resource space respectively.Fig 2 shows examples of four basic operations of resource space.

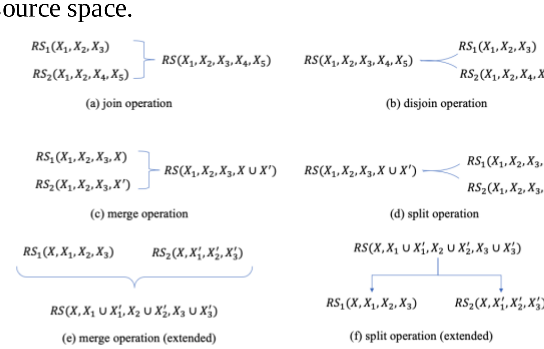

Fig 2: Examples of four basic operations

III. About The Dimension

The dimension is the most special aspect and the core of RSM. The dimensions of RSM are like the attributes of relation tables in RDBM in some aspect. But attributes only reflect only certain properties of a resource itself. A dimension represents a perspective on a set of resources. A RSM with multi-dimensions views a set of resources from different perspectives. In other ways, dimensions like the different topics on resources. And the coordinates are like a set of abstract words or classes with clearly hierarchical structure which can describe the such topic which is like its dimension. Based on this, people can use the topic discovery algorithm such as LDA to automatically construct the dimensions and coordinates in the resource space [11]. Hence, the dimensions and coordinates in RSM are a set of resources based while the attributes and properties in RDBM are individual-based.

A resource has external feature and internal feature. The internal feature reflects the content of resources while the external feature helps the distinguish one resource from the other resources [6]. Based on this, dimensions can be divided into two parts: the external dimensions and the internal dimensions correspond to the macro dimensions and micro dimensions [12]. For example, there is a scientific resource space, the external dimensions can be directly get from the set of scientific resources such as the Year axis, Discipline categories axis, Author axis, Publisher axis and so on. The hierarchical coordinates on external dimensions can be dig out by hierarchical clustering algorithms according to resources. Then it can be adjusted in some ways to obtain the optimal hierarchical structure. And internal dimensions are content-based, three elements for the contents of scientific documents are Task, Process and Material [12]. The hierarchical coordinates in internal dimensions can be dig out by the Natural Language Processing (NLP) technology such as Named Entity Recognition (NER) and Relation Extraction (RE) by Machine Learning (ML).

IV. The Relationship Between RSM And Other Resource Management Model

A. RSM and RDBM

RSM is classification-based while RDBM is focus on attributes of individuals. In RSM, people can observe the resources as a whole and then classifies them top-down by common sense for high-level classification and domain specific knowledge for low-level classification to establish a uniform hierarchical coordinates system. The data model is a relation table in RDBM which does not support generalization and specialization on attributes [6, 7].

RSM is a multi-dimensional classification space and it concerns the contents of resources. A coordinate can be defined by an abstract class which represents a set of resources from a specific semantic perspective. Hence, the resources managed by RSM can be any type of resources while RDBM only manages the atomic data. RSM has more powerful resource representation ability than RSDM.

RSM and RDBM can be transformed into each other.

B. RSM and Semantic Link Network

Semantic Link Network (SLN) is a self-organized semantic model for representing and operating the semantic structure of complex system. SLN consists of semantic nodes and semantic links between the nodes. Its semantic nodes represent categories of things and the semantic links between nodes represent the semantic relations between nodes [1, 2]. People can mine deeper relationships between semantic nodes by applying semantic reasoning rules on the semantic link network.

RSM focuses on the classification dimension of resources while SLN focuses on the semantic relationship between semantic nodes. RSM and SLN can be transformed into each other without changing their normal forms. A RSM can be semantically equivalent transformed to an SLN by preserving implication links and equality links on the hierarchical coordinates structure. Fig 3 is an example of the transformation from RSM to SLN.

RSM does not focus on semantic relationships between resources while SLN does not focus on classification of resources. Then, RSM and SLN can be integrated together to support richer semantic modelling and applications as combining the RSM's classification semantics and the SLN's link semantics.

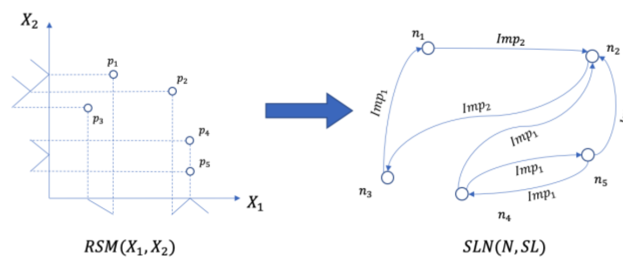

Fig 3: An example of semantically equivalent transformation from RSM to SLN

C. Comparison Between Different Models

RSM focuses on classification semantics, SLN focuses on the links between semantic nodes, RSM+SLN focuses on the both classification semantics and link semantics and RDBM is mainly based on the identity, attribute, and values as well as the dependence between attributes.

RSM does not focus on the relationship between resources, hence, it has no reasoning ability. SLN and RSM+SLN model have strong reasoning ability while RDBM has middle reasoning ability.

RSM, SLN, RSM+SLN and RDBM have normal forms theory and integrity constraints to ensure that dealing with resources more accurately and rigorously.

RSM+SLN model contains the strongest semantic information. The most important thing for human beings to know the world is to rely on the classification of resources and the relationship between resources. And the combined model contains both.

D. Other

The Web ontology language (OWL) is an ontology language designed to provide a common way to handle the content of web information. And it can be transformed into RSM by mapping OWL description onto the RSM's elements, for example mapping the class labels and property labels onto resource space, axes and hierarchical coordinates, mapping the individuals onto resources, mapping restriction and characteristic onto constraint of RSM. The generated RSM could lose some of the semantic information described in OWL. But when RSM is transformed into OWL description, the generated OWL description can be semantically equal to the RSM, as the classification semantics can be fully defined in OWL.

V. DISCUSSION

The RSM differs from other resource management models lie in dimensions and hierarchical coordinates. A fine RSM can automatically analyse resources to get dimension and hierarchical coordinates, and can automatically maintain resource space. As more and more resources are managed, the axes and hierarchical coordinates will change according to the resources to obtain the optimal space structure. Hence, how to automatically discover the appropriate dimensions precisely and construct the appropriate hierarchical coordinates correctly are the core issues to be considered next in RSM. A good designed hierarchical coordinates system should reflect the precisely classification semantics close to resources and spend less time searching and managing resources.

People can deal with resources from different perspectives according to the specific set of resources. For the photo RSM, the time dimension, the place dimension and the colour dimension maybe very important. But for the e-mail RSM, there does not exist the colour dimension. So, how to build a general resource space model according to the various resources is also a problem worth studying.

Dimensions and coordinates mining should be based on resources. For text RSM, dimensions and coordinates are obtained by analysing and processing text. When natural language processing technology develops to application level, text RSM will make great breakthroughs in dimension and coordinate automatic discovery.

VI. SUMMARY

RSM is a multi-dimensional, classification-based, content-based and high-level semantic space model. The biggest difference from the other resource management model is the classification-based.

In the future, the resource can be managed by a complex semantic model. Integrating the RSM with SLN in application level, database in storage level. Converting OWL, XML and other resource description languages into RSM through transformation technology while converting the resources into SLN. Based on this, relationships hidden behind resources can be easily excavated and it can help people better understand the world and change the world.

ACKNOWLEDGMENT

This research was supported by the National Science Foundation of China (No. 61876048 and No. 61640212) and by the International Research Network on Cyber-Physical-Social Intelligence consisting of Institute of Computing Technology at Chinese Academy of Sciences, University of Chinese Academy of Sciences, Aston University, and Guangzhou University.

REFERENCES

[1] H. Zhuge, The Knowledge Grid, World Scientific, Singapore, 2004.

[2] H. Zhuge, The Knowledge Grid: Toward Cyber-Physical Society, World Scientific, Singapore, 2012.

[3] H. Zhuge, Resource Space Grid: Model, Method and Platform, *Concurrency and Computation: Practice and Experience*, vol.16, no.14, 2004, pp.1385 - 1413.

[4] H. Zhuge, Fuzzy resource space model and platform. *Journal of Systems and Software*, vol.73, no.3, 2004, pp.389 - 396.

[5] H. Zhuge, Resource Space Model, Its Design Method and Applications, *Journal of Systems and Software*, vol.72, no.1, 2004, pp.71 - 81.

[6] H. Zhuge, The Web Resource Space Model, Springer, 2008.

[7] H. Zhuge and Y. Xing, Probabilistic Resource Space Model for Managing Resources in Cyber-Physical Society, *IEEE Transactions on Service Computing*, vol.5, no.3, 2012, pp.404 - 421.

[8] H. Zhuge, Multi-Dimensional Summarisation in Cyber-Physical Society, Morgan Kaufmann, 2016.

[9] H. Zhuge and L. He, Automatic maintenance of category hierarchy. *Future Generation Computer Systems*, vol.67, 2017, pp.1 - 12.

[10] H. Zhuge, Y. Xing, P. Shi, Resource space model, OWL and database: Mapping and integration, *Acm Transactions on Internet Technology*, vol.8, no.4, 2008, pp.1 - 31.

[11] X. Yu, L. peng and Z. Huang, A framework for automated construction of resource space based on background knowledge, *Future Generation Computer Systems*, vol.32, 2014, pp.222 - 231.

[12] L. He, Automatic Construction of Scientific Resource Space and its Application, Beijing: Institute of Computing Technology Chinese Academy of Sciences, 2018.

[13] H. Zhuge, Semantic linking through spaces for cyber-physical-socio intelligence: A methodology, *Artificial Intelligence*, vol.175, 2011, pp.988 - 1019.

[14] H. Zhuge and B. Xu, Basic operations, completeness and dynamicity of cyber physical socio semantic link network CPSocio-SLN, *Concurrency and Computation: Practice and Experience*, vol.23, no.9, 2011, pp.924 - 939.

[15] H. Zhuge and J. Zhang, Automatically constructing semantic link network on documents, *Concurrency and Computation: Practice and Experience*, vol.23, no.9, 2011, pp.956 - 971.

[16] H. Zhuge and Y. Sun, The schema theory for semantic link network, *Future Generation Computer Systems*, vol.26, no.3, 2010, pp.408 - 420.

[17] H. Zhuge, Interactive Semantics, *Artificial Intelligence*, vol.174, 2010, pp.190 - 204.

[18] 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.

[19] M. Cao, X. Sun and H. Zhuge, The contribution of cause-effect link to representing the core of scientific paper — The role of Semantic Link Network, *Plos One*, vol.13, no.6, 2018.

[20] X. Sun and H. Zhuge, *Summarization of scientific paper through reinforcement ranking on Semantic Link Network*, IEEE Access, vol. 6, 2018, pp. 40611-40625.

[21] H. Zhuge, Cyber-Physical-Social Intelligence on Human-Machine-Nature Symbiosis, Springer, 2019