# A DATA MODEL FOR REPRESENTING CHINESE GARDEN

*an E-R Model and Keyword Analysis techniques approach*

SHU-HUI HUANG
*Graduate School of Computational Design, National Yunlin University of Science & Technology*
*Touliu, Taiwan, ROC.*
*g9134709@yuntech.edu.tw*

TENG-WEN CHANG
*Graduate Institute of Architecture, National Chiao Tung University*
*Hsinchu, Taiwan, ROC.*
*tengwen@arch.nctu.edu.tw*

CHUNG-JEN KUO
*Chinese Culture University*
*Taipei, Taiwan, ROC.*
*cjkuo@cjkuo.com*

**Abstract.** There are many discussions about spatial consensus or further to "reuse" the design of Chinese Garden in the modern context. For addressing this question, we need to be able to represent the spatial elements of Chinese Garden digitally. However, only few researches in computational design field start to tackle the problem. Furthermore, most of literatures we found are basically the description of abstract or sometime vague concepts. The scope of this research—spatial composition, an efficient description of spatial elements of Chinese Garden in literatures and their relations will be the goal of this research. In this research, we analyze on spatial elements and characteristics of *Lin Family Compound* to create a suitable database that can offer further researches. For organizing the information we collect, we propose a system called DiGE using the Entity-Relationship Data Model. In addition, we will use natural language processing to deal with the unclear data of words depiction.

## 1. Introduction

Chinese Garden represents a specific type of architecture with strong culture characteristics as well as unique design aura for design studies (an example is shown in Figure 1). Representing Chinese Garden computationally is the key issue for analysis and understanding Chinese Garden. While there are quite a few information available for different aspects of Chinese Gardens, the information itself is not up to the need of either computation or representation. With adapting similar research in design domain—design elements analysis, this research attempts to establish a computational base for representing Chinese Garden with available information.

*Figure 1. Artificial wall and Chi-Ku Library of a Chinese Garden*

While considering Chinese Garden as a design outcome, design and Chinese Garden share a lot of similarity, especially their spatial composition. Therefore, the way of analysis or recognition of the levels of design elements in design researches can be applied onto analyzing the elements of Chinese Garden.

In design domain, there are several approaches towards representing the design elements with information-central approach. Amongst them, Top-Down system and engineering design model (EDM) are two main projects. Top-Down developed by (Mitchell, et al. 1990) describes a building according to the composition of its design elements. With different aspects, EDM developed by (Eastman 1992) provides a powerful language for modeling every design element and their relations. In Top-Down, every element is one kind of symbol and can be easily composed via levels of abstraction, while EDM provides a richer but difficult approach for modeling/representing the design elements.

In the usages of data/information representation, (Eastman 1992) applies data modeling techniques on defining an extensible database schema for the purpose of intelligent support of CAD systems. With similar approach, (Lee 2001) offers a spatial access oriented implementation of a topological data model for 3D urban entities. They are all based on the analysis of design elements in certain design problems and the relations and elements of design are specified in a more structural fashion. Furthermore, two main concepts shared among them are 1) design elements are hierarchical represented and needs some mechanism for classification as well as analysis; 2) data scheme of data elements is the key for representing design. These concepts provide the base of this research: representing Chinese Garden with data modeling techniques.

## 1.1. TYPES OF THE INFORMATION

As the nature of this paper is to represent the Chinese Gardens in terms of computational needs. We start with the analysis of available information. Since Chinese Garden is one of the most important architectural types, there do exist certain amount of information and documents for various aspects of Chinese Garden. There can simply be divided into two catalogues—textural description and spatial information.

### *1.1.1. Textural Description*

Most of information exists for Chinese Garden is some kinds of textural description for the scenery of Chinese Garden. Furthermore, most of available information for Chinese Garden is belonged to this catalogue. The main characteristics shared among them are diverse and trivial, and the amount of information is huge. Approximately, a classification is completed and described as the following:

- Development: It is based on the periods and phrases of the Chinese Garden development to introduce representative works and significant figures, regions or concepts.
- Scenery: There are quite a few varieties of Chinese Garden. And, its classification is on the basis of the varied characters, sizes, and geographical nature environment. This information is to introduce the features and technique of Chinese Garden and architecture scenery.
- Composition: This class is mainly about features, types, functions and deployment techniques and effects of Chinese Garden composition.
- Architecture: This class of information introduces the different models, configuration, composition, materials, types, and effects implemented of Chinese Garden architecture.
- Connection: This class shows the relationships and influences of composition mutually, like pavilion or pond, is discussed in this class.
- Route: Most of Chinese Garden is designed according to some specific touring route. Each route connects some scenic spots and provides a route for tutoring through the Garden. It serves to spread the scenery out and organize the viewing procedures.

- Aesthetics: This class is for discussion of the differences and influences of beauty, garden beauty and Chinese and Western Gardens and to introduce the appreciation of Chinese Garden arts and design quality.
- Case study: It is the introductions, notices and comparisons of Chinese Garden, environment and social background to use words or pictures to depict the significance of Chinese Garden scenery and to introduce important scenic spots.

#### *1.1.2 Spatial Information*

In addition to the textual description of Chinese Garden, the concrete spatial information can be obtained and work as an important collection of information. Moreover, 3D spatial information along with other information is the key for realizing the visual feedback and spatial orientation. It is required in both information representation and user interaction in a virtual-physical co-existence information space shown in (Chang and Lai 2003). Therefore, the spatial information consists of five catalogues as the following:

- Words: Description and explanation of auxiliary data of Chinese Gardens.
- Pictures: Pictures of gardens taken actually.
- 2D/3D drawing: Individual exhibition and explanation of Chinese Garden architectures.
- Panoramic: Display frame of important scenic spots.
- Virtual Reality: Simulation of the Chinese Garden display or guideline of the whole park.

This collection of data is through the measurement over several existing Chinese Garden that is part of a long-term on-going project led by Chung-Jen Kuo and Teng-Wen Chang. Another similar approach (Kuo 2003) thus provides a schematic guideline for information analysis for this research. Based on this resource, another paper (Kuo 2004) is also presented in this conference.

### 1.2. REQUIREMENT

With the information classification above, we tend to elaborate the research problems with several requirements. (1) Information accessibility: The textural description or spatial information of the Chinese Garden is numerous. The information often acquired from the single substantial sources, such as books. Therefore, information browsing is not convenient. (2) Structured information: In addition, in the field of computation, the knowledge configuration of the Chinese Garden needs more structures in order to be any useful. Therefore, this paper will focus on solving problems of that aspect. Thus, two main theories: data model and natural language process are reviewed in the following sections. In addition, due to diverse nature of information, this paper at present develops Data Model on the basis of composition data of Chinese Garden architecture and concrete spatial information. Moreover, since the information itself is mostly descriptions, there are not clear in the mode of data acquisition and the parts of words depiction. For the need of feasibility such for constructing a knowledge base from design cases (Kuo 2004), this study will comply with Natural language processing (cf. Section 2.2 below) to deal with the descriptive nature of information query problems mentioned above.

## 2. Reviews

Following the needs of representation and the nature of information, two main computational techniques are reviewed in this section—data model and natural language process (NLP).

For the sake of digitalizing the Chinese Garden Data, we need to take advantage of the Data Model to analyze Chinese Garden Data and then describe the configuration characteristics of the Chinese Garden Data. Besides, NLP can deal with descriptive problems and make communications between human beings and computers clearer and easier.

### 2.1. DATA MODEL

Data model is an abstract, self-contained, logical definition of the objects and operators. The objects allow us to model the structure of data we are interested and operators are thus for the behaviors among data (Date 2000). Furthermore, a data model is a collection of conceptual tools for describing data, data relationships, data semantics, and data constraints (Korth 1986)—(1) data is the basic unit of the database; (2) data relationship is the relationships between different items in the database; (3) data semantics is the meaning of the data; (4) data constraint is the context within which the data has meaning.

While concerning the modeling technique, an E-R model (Entity Relationship model) is used according to its popularity and usefulness for the domain problem. Briefly speaking, an E-R Model is a detailed, logical representation of the data. In addition, the constructs used in the E-R model can easily be transformed into relational tables. It provides a significant factor in our implementation.

Another factor is normalization. Normalization aims at solving the following five problems: 1) anomalies in operations behavior, 2) existence of inconsistent data, 3) redundancy of data in the database, 4) instability of schemata after changes, 5) different abstraction level in applications

(Thalheim 2000). The goal of conceptual schema design, where the ER approach is most useful, is to capture real world data requirements in a simple and meaningful way that is understandable by both the database designer and the end user (Teorey 1999). In addition, the E-R model is expressed as an entity relationship diagram (E-R diagram), which can provide a way to analyze the usage of entity as well as to arrange attributions of entity in the Chinese Garden.

### 2.2. NATURAL LANGUAGE PROCESSING

Natural language processing (NLP) is a significant way to cope with problems that mainly analyzes the wording of human languages. In addition, NLP carries out all kinds of theories and methods to processed natural languages for the sake of doing valid correspondence between human beings and computers. NLP is usually used with the task of developing programs possessing some capability of understanding a natural language in order to achieve specific goals (Binot 1991).

One simple but powerful of NLP system is called keyword analysis (Covington 1994). The keyword analysis looks for specific words in the sentence and responds to each word in a specific way. A system developed using keyword analysis is called keyword system. Keyword system works well for database querying since each important concept associated with the database has a distinctive name. A keyword system responds only to the words that identify fields, values, comparisons, and the like, and ignores all the other words. Two prominent keyword systems today are AICorp’s Intellect and Symantec’s Q&A, and they both are database query systems.

While the system characteristic of Chinese Garden have proper noun, the query method of ours is very similar to keyword analysis of keyword system. Therefore, we use keyword analysis technique for specifying our query to achieve the features described above.

### 2.3. TOWARD AN E-R MODEL FOR CHINESE GARDEN

In this study, we make use of E-R model to describe the data configuration of Chinese Garden. This is employed to depict the actual database saving configuration of the Chinese Garden. After modeling the Chinese Garden, we take advantage of keyword analysis from NLP to handle the descriptive problems of information query.

## 3. Hypothesis and Research Steps

### 3.1. HYPOTHESIS

Develop a searching engine is to present the data structure of the Chinese Garden: 1) Preliminary data model can't achieve all the requirements of end-users. By re-flash procedure, it can make the data model more flexible. 2) Searching engines provide the functions of searching the database for users. In addition to displaying the data after searching, it can use figures to portray it. 3) Cooperating with keyword analysis, it can meet the users' needs soon; provide more correct and stable information.

### 3.2. RESEARCH STEPS

In this research, we analyze on spatial elements and characteristics of Chinese Garden "Lin Family Compound (林家花園)" (Han and Hung 1973) to create a suitable database that can offer further researches as described before. By analyzing information available from the text as well as pictures of the "Lin Family Compound (林家花園)", we are aiming to find spatial elements as well as some characteristics of information.

Research steps used in this research are—(1) collecting and analyzing two catalogues of information using E-R model techniques; (2) refining the E-R model according to the designers requirement for feasibility in developing the search engine; (3) applying several keyword analysis techniques over the E-R model developed above; (4) implementing the system. The steps of research will explain within analysis and implementation sections.

## 4. Analysis

The main analysis for this paper is to analyze the textual description and spatial information of Chinese Garden and finds the characteristics of spatial elements by text of Chinese Garden. These descriptions of spatial elements might use methods of combining, verb or adjective to show. Following these descriptions we can get many attributes from spatial elements and arrange these possible attributions for database. These attributions are methods of combine, soft or hard, natural or attribute, size and color.

### 4.1. THE E-R MODEL FOR CHINESE GARDEN

The element structure of Chinese Garden is as Figure 2. Each Chinese Garden will be included with more than one Building inside and each

Building will be included with more than one Component inside. Data Model can analyze the information according to the basic structure.

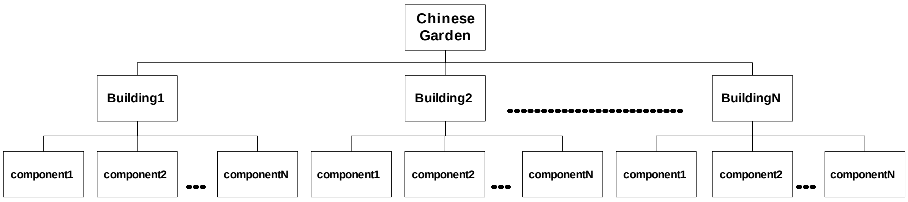


*Figure 2. The structure of Chinese Garden document.*

*4.1.1 Entity*

In a basic database, Entities are the principal data objects about which information is to be collected. Entity also can be a physical existing object, like man, car or building, also can be an abstract object. In correspondence to the concept, the elements of Chinese Garden, like inn and pavilion are catalogued into physical object. And, like functions, traits, all are regarded as abstract objects.

*4.1.2 Attribute*

Attributes refer to some basic traits, which describe the objects or relationship. So, attributes are characteristics of the entity that can provide detailed description of entity. Like a pavilion can be described by the name, functions, location, or traits, and so on. Each entity will have the attribute, which describe the attribute values.

*4.1.3 Relationship*

A relationship is an association between two or more entities. Therefore, it stands for some kind of specific meaning. For instance, in Figure 3 discussing about the relationship type between each entity and building is “include”.

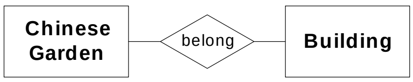


*Figure 3. The relationship for Garden and Building.*

*4.1.4 Constraints*

An E-R scheme may define certain constraints to which the contents of a database must conform.

- Mapping constraints: Mapping constraints represent the quantity of relationship from one entity to the other. For instance, each garden gets to have more than one building. Therefore, we can get the conclusion that the relationship between Chinese Garden and building is one again maple. This kind of condition has to be narrated in the data model scheme. Mapping constraints consist of 4 types as 1) One to one. 2) One to many. 3) Many to one. 4) Many to many.

- Participation constraints: The existence of Entity will be determined by the relationship between each entity confirmation or not. The idea can be called as Partial Participation and Total Participation.

Below E-R diagram consists of entities, attributes, relationships, and constraints.

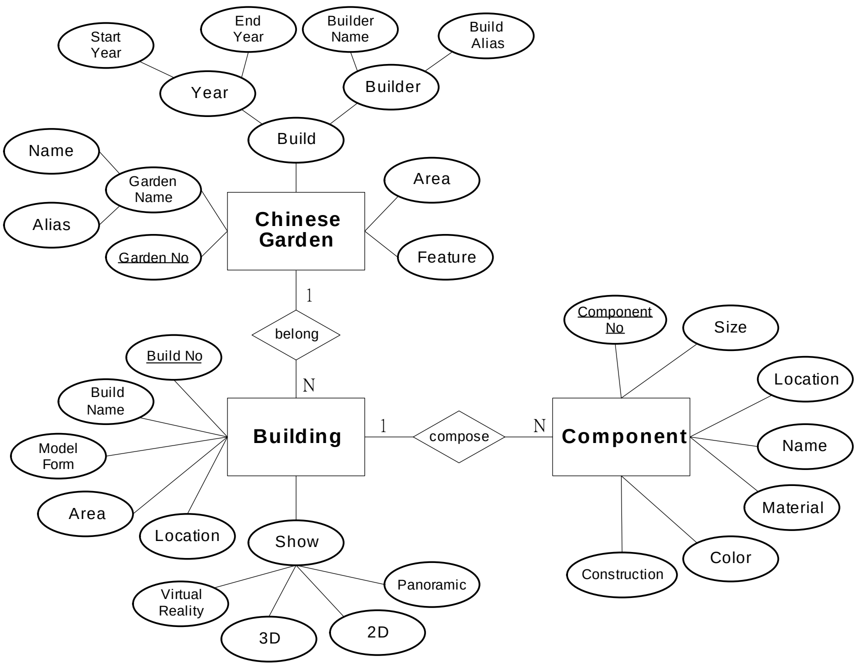


*Figure 4. E-R diagram notation for Chinese Garden.*

### *4.1.5 Normalization*

After fulfilling the design of data guidelines, it is able to use normalization to check whether the design is suitable or not. The purpose of normalization is to reduce the repetition and avoid the abnormal update cases. When a common database achieves its Third Normal Form, it almost meets data correction abnormal cases required by our model. Consequently, this study will take advantage of a common normalization procedure—First Normal Form (1NF), Second Normal Form (2NF) and Third Normal Form (3NF) to examine schema in order to achieve the connection of 3NF.

- 1NF: Remove the repeated data. Every table will have one primary key. And at the same time, every column of every row in the every table will only save the atomic data (indivisible). For example, “Garret-Location” is not atomic. We can separate the data into two values.
- 2NF: Remove the Functional Dependency, which only takes place in the main key consisted by many columns. It means that some columns only have dependency with partial columns of the main key and not with others. For instance: “Garden No + Building No→ Garden Area”,” Garden No + Building No→ Building Types”. If deleting Building No only influences Building Type column, not to have effects on Garden Area. The kind of relation is called Functional Dependency.
- 3NF: Remove Transitive Dependency. Transitive Dependency is said that in the one table, one column can decide the value of other column, however, these columns also another column to determine the values of the rest. For example, Garden No+ Building No→Component No, Building No + Component No→Construction. This type of relation is named Transitive Dependency.

#### *4.1.6 Denormalization*

Sometimes after normalization, the effectiveness will be influenced by the normalization so that it will cause the problems of the speeds to deal with data. Therefore, under the consideration of speeds, it needs to perform denormalizaion. In other words, when performing denormalization of database, it also examines the system effectiveness. When the effectiveness is not stable, it must do appropriate denomalization. But, it has to carefully control the probable repetition brought about denormalization in order to achieve the perfect database design.

### 4.2. KEYWORD ANALYSIS

The goal of keyword analysis is to analyze, understand and generate languages that humans use naturally for computer. Here, we will parse the text and communicate between human and computer.

Keyword analysis work well for database querying because each important information and associated is using keyword to store in database. Via the keyword analysis, it can delete unnecessary words, find the synonym, and source the keyword out as figure 5.

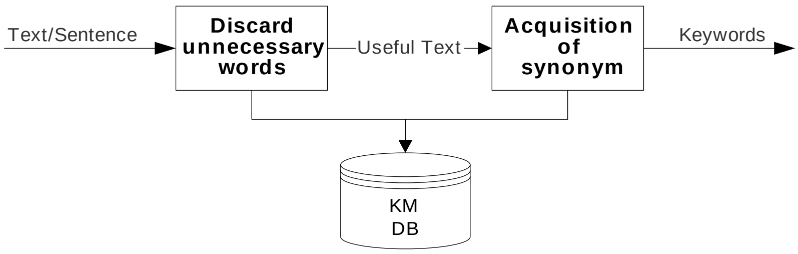


*Figure 5. The process of keyword analysis.*

- Discard unnecessary words: Discard unnecessary words from Text/Sentence, for example: "Show Garden", "Show" is unnecessary word and it might get dropped. System can check with Text/Sentence by this mechanism.
- Acquisition of synonym: Create a synonym, which provide us enough information to find out the words. After we finish the first two processes, we will get one or more keywords.

## 5. Implementation

For organizing the information we collect, we propose a system called Digital Chinese Garden Elements (DiGE) using the E-R Model. E-R Model completed will integrate with normalization and denormalization to do processing so that the data configuration of the Chinese Garden will be well-organized. The result of E-R Model will be implemented using a MySQL database on a Linux platform. In addition, we will create a search engine over this database. We will use Tomcat to be server at Linux platform and use Java/Jsp as the programming language.

The information we collect will not be limited by the textual information but also the 2D or 3D spatial information of some particular examples. In the final paper, we will use the example "Lin Family Compound (林家花園)" to describe how the E-R model is applied to this problem as well as the usages of DiGE.

DiGE consists of the three sections: Refining Data Model, Translator and Search Process, (cf. figure 6). DiGE uses the Refine Data Model to allow users adjusting original Data Model Structure. It takes advantage of Translator to convert the Natural Language keyed in by the users into Query Language and then provide Search Process to show the information and the related structure by the Search Process.

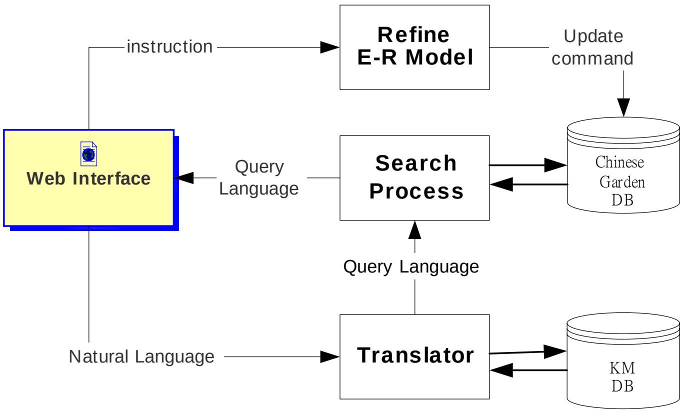


*Figure 6. The process of System.*

5.1. REFINE E-R MODEL

DiGE will refine the Data Model automatically according to the modifying information. For instance, system can add an extra column "Garden" into Garden Table whose data structure will be revised as Figure 7. The part highlighted by the red line is the part of the changeable data configuration.

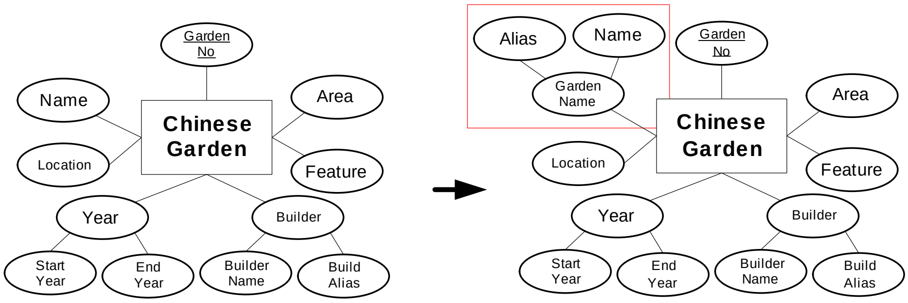


*Figure 7. Refining the Data Model*

5.2. KEYWORD ANALYSIS COMPONENT

Keyword analysis component of DiGE will provide the functionality of keyword analysis. First, DiGE will delete the unnecessary words, get the possible keywords, and then finally find the synonym. DiGE will check the database according to the keyword. The information process is shown as next:

- Step1: User input a query string such as "*About Chinese Gardens, In front of Ding-Jing Hall, a pair of moon-door face each other from across the space enclosed by the courtyard walls in Lin Family Compound*".
- Step2: keyword analysis component will delete the unnecessary words. After process, we can get a more useful string "Chinese Gardens, Ding-Jing Hall, moon-door, courtyard walls, Lin Family Mansion and Garden".
- Step3: Find the synonym. With algorithm, we can then find the synonym as shown in Table 1.

TABLE 1. Synonym

| **Synonym** | **Source** |
|---|---|
| Garden | Chinese gardens |
| Wall | courtyard walls |

- Step4: Obtain keywords. Finally, the keywords we get are Garden, Ding-Jing Hall, moon-door, Wall, Lin Family Mansion and Garden.

### 5.3. A SEARCH ENGINE

Search Engine allows users to check the Chinese Garden data and relationship. After we finish the translation, we will get one or more keywords. Hence DiGE will use these keywords to retrieve data from database. Furthermore, search engine will follow the sequence of column and search the information.

- If the keyword is identical with the column name, it means the column name is what DiGE is looking for.
- 100 % identical with the keyword.
- Similar with the contents of column.

Finally, the result will be show as 2 parts like data and relationship by search sequence as we mentioned in previous pages. Result as Figure 8.

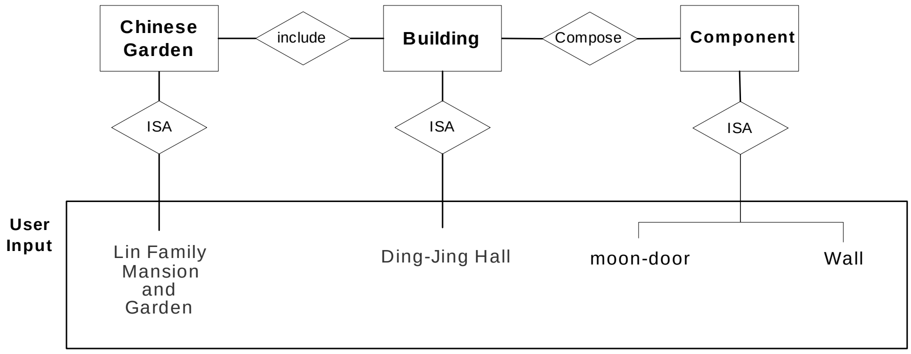


*Figure 8. Data Relation.*

Structure relationship will chance by following the re-vision of data model. With this kind of procedure, users can access the structure of data model and modify the structure according to the demands of end-user, and finally, check the refine by search engine.

## 6. Conclusion and future works

In this paper, it makes use of E-R model analysis to classify composition and features of Chinese Garden based on different types of informaton. Reorganizing the garden information not only provides descriptive information by words, 2D, 3D and virtual reality but also presents the relation of data configuration. Besides, DiGE will utilize keyword analysis to deal with diverse descriptive queries for providing a feasible query and search engine as well as the patterns of information search.

The benefits of DiGE are numerous. For example, with the DiGE, we are able to extend the information model to other Chinese Gardens. In addition, the patterns among information structures of different gardens can be analyzed and studied while we have more garden information available.

Another important direction is applying the machine learning onto DiGE to help the communication and coordination for the data model built-in. Consequently, we expect that it can integrate with machine learning technology in the near future and develop available learning mathematical calculations. During the process of using DiGE, agents can automatically acquire knowledge to set up suitable rule, then will learn the user need and voluntarily update the configuration of data model. Using machine learning will bring more flexibility to face the problems that needs to handle, thus to further understand this specific type of architecture.

## References

Binot, J-L: 1991, "Natural language processing and logic." *in* T A., (ed.), *From Natural Language Processing to Logic for Expert System*, John Wiley & Sons, New York, pp: 49-116.

Chang, T-W and Lai, I-C: 2003, Reflective Navigation in Information Landscape: A 3D Museum for Civic Buildings, *in* A Choutgrajank (ed.), *the 8th Computer Aided Architectural Design Research in Asia (CAADRIA) conference*, Bangkok, Thailand, pp: 641-650.

Covington, MA: 1994, *Natural Language Processing for Prolog Programmers*, Prentice Hall.

Date, CJ: 2000, *An Introduction to database systems*, 7th, Addison Wesley.

Eastman, CM: 1992, A Data Model Analysis of Modularity and Extensibility in Building Databases, *Building and Environment*, **27**: 135-148.

Han, PT and Hung, WH: 1973, *Pan-Chiao Lin Family Compound*, Arch. Dept. Tunghai Univ.

Korth, HF: 1986, *Database system concepts*, McGraw-Hill.

Kuo, C-J: 2003, Spatial Analysis of Chinese Garden Designs with Machine Learning, *in*, *CAADRIA2003*, Bangkok Thailand.

Kuo, C-J: 2004, Machine Learning for Diagrammatic Spatial information, *in*, *CAADRIA2004*, Seoul Korea.

Lee, J: 2001, A Spatial Access Oriented Implementation of a Topological Data Model for 3D Urban Entities, *in*, *UCGIS Summer Assembly*, Buffalo, NY.

Mitchell, WJ, et al.: 1990, "Top-Down Knowledge-Based Design." *in* M McCullough, et al., (eds.), *The Electronic design studio: architectural knowledge and media in the computer era*, MIT Press, Cambridge (Massachusetts / USA), pp: 137-148.

Teorey, TJ: 1999, *Database modeling & Design*, Morgan Kaufmann.

Thalheim, B: 2000, *Entity-Relationship Modeling: Foundations of database Technology*, Springer.