

# Knowledge Management in a Chinese Genealogy Information System

Jian-Hua Yeh<sup>1</sup> and Chao-Chen Chen<sup>2</sup>

<sup>1</sup>Department of Computer Science and Information Engineering, National Taiwan University.  
System designer, Taiwan Genealogy Online Project, Taipei, Taiwan, R.O.C.  
jhyeh@mars.csie.ntu.edu.tw

<sup>2</sup>Graduate Institute of Library and Information Studies, National Taiwan Normal University.  
Project Leader, Taiwan Genealogy Online Project, Taipei, Taiwan, R.O.C.  
cc4073@cc.ntnu.edu.tw

**Abstract.** The Chinese genealogies are the best resources for the family histories. They contain excellent information for people's root tracing and for academic researches. Our Taiwan Genealogy Online Project aims to digitalize genealogy material and to provide online services. This paper explains the design theory and implementation ideas of our system for Chinese genealogy.

**Keywords:** genealogy, metadata, pedigree, visualization, Taiwan Genealogy Online Project.

## 1 Introduction

It is estimated that there are about sixty thousand titles of Chinese genealogies around the world. These huge amounts of genealogies are excellent resources for people's root tracing, besides, for researchers, there are many valuable undiscovered phenomena in those genealogies to be revealed. And it is a good point to put those valuable resources to be widely used. But unfortunately they are traditionally book form or microfiche form. Those forms of materials are not convenient for full text information retrieval. For the sake of efficiencies, it needs a new transformation of those genealogies. At the present time, "digitalizing" is the good choice. But there is little related research on the digitalization of Chinese genealogies, so we present one comprehensive research project about that. In this paper, we will introduce how to using the XML technology to create the metadata of genealogy, maintaining relations among the individuals, and how to developing management and visualization utilities to representing pedigree information.

## 2 Genealogical Metadata Design Issues

In this project, we analyze attributes of Chinese genealogical information, then design three sets of metadata: (1) format for ancestors; (2) format for families; (3) format for surnames.

The part (1) records the relations of an individual and his families. This is the basic materials for building pedigrees. The part (2) records the origin and the history of a family, including its organization rules and residence places. The part (3) gives general descriptions for each surname. The three formats altogether give a complete development history of patriarchal clan.

In addition, this project also implements a management system for Chinese genealogy. Our system can automatically link individuals to their ancestors or offspring and establish pedigree charts. The system is interfaced with Web. Users can create metadata of genealogy through Internet.

### 3 Object Structures for Representation of Genealogical Information

The researches of digital libraries have been inspired in recent years. Among these researches, a lot of them are based on an object-oriented model [6,7,8]. The object-oriented model is also applied in genealogy information system design. In this paper, we apply the object structures defined in NTUDLM [9] to manage genealogical information along with XML-based metadata. The rationale of this design is to facilitate both management and visualization of object relationships: the XML-based metadata is able to render Web output more directly, while the object relationships can be visualized more efficiently. The object structure in the genealogy information system is used to keep both metadata attributes and relations to other objects. While an attribute describe an object from one aspect, a relation associates one object with another object. As a result, the relationships among objects will form an object network. Figure 1 illustrates the structures among objects.

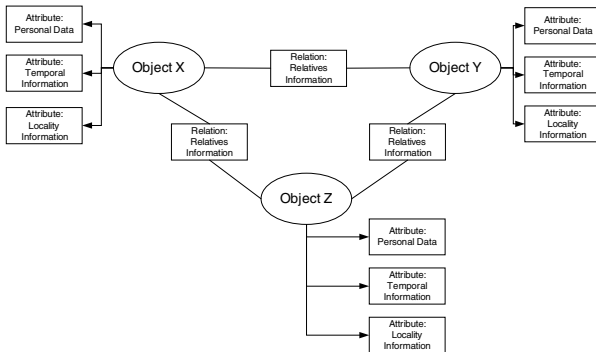


Fig. 1. The structures among objects

### 4 Genealogical Information Visualization

The visualization issues in genealogy information system are based on the object structures and object network discussed in the previous section. Because the objects in the system form object networks, the visualization of the network will become a

graph presentation problem. However, the pedigree information contained in the Chinese genealogy is characterized as a paternal system, as far as the traditional genealogy material concerns, the pedigree is presented as a tree structure, as shown in Fig. 2. There are many related pedigree-drawing approaches found in historical material use tree-like structure as presentation format, as listed in Table. 1.

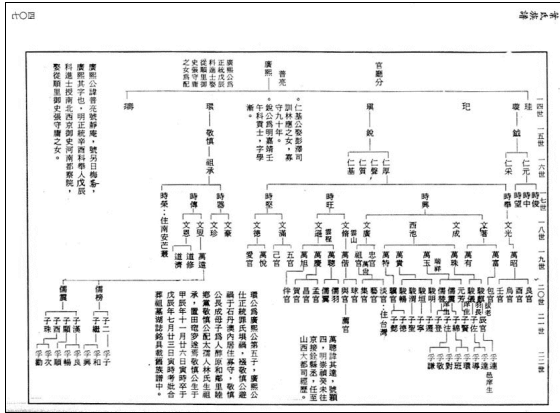









Fig. 2. A page of traditional genealogy material

Table 1. Several chart types of pedigree tree presentation

Chart Type	Sample	Chart Type	Sample	Chart Type	Sample
Tree chart	 Tree chart	Ancestor chart	 Ancestor chart	Descendant list	 Descendant
Box chart	 Box chart	Collapsed chart	 Collapsed	Group list	 Group
Fan chart					

Since the Chinese genealogy material is based on the paternal system, the presentation of a Chinese pedigree is therefore a tree visualization problem. In the previous visualization researches [1], the visualization which concerns of a graph or tree structure is based on several factors: node placement, link management, scalability, and interactivity. These issues lead to the adoption of hyperbolic tree presentation [3,4,5] in our system, which is able to present a lot of tree nodes in a pedigree chart while keep good interaction with users. The presentation of a hyperbolic tree is shown in the sample figure in the next section.

## 5 Genealogy Information System Architecture

The design of object structures and the adoption of visualization techniques are two important bases of the genealogy information system. Fig. 3 depicts the system architecture of the system. Fig. 4 shows a sample screenshot of the system.

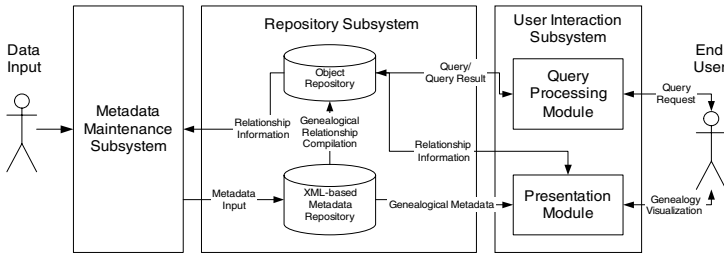


Fig. 3. System architecture

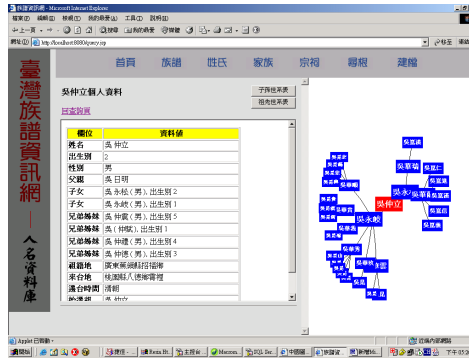


Fig. 4. A sample screenshot of pedigree presentation

The object repository is the module that stores and manages the metadata records in the system and is responsible for resolving the user’s query. To the query process module, the object repository is an object-oriented data store and its objects structures have been elaborated in section 3. The XML-based metadata repository is the module that stores the contents of the genealogy metadata in a physical form.[2] On the other hand, the object repository stores the contents of the genealogy metadata in an abstract form.. When an update to the metadata contents is to be carried out, the update is first made to the structure of the XML-based metadata repository. Then the structure of the contents repository is compiled to create a new object structure for the object repository.

## 6 Conclusion

This paper discusses the design of a Chinese genealogy information system aimed at providing management and visualization utilities to extend the applications of digital libraries to a new dimension. The discussion particularly focuses on maintaining relations among the objects in the digital library from the genealogy metadata information stored in the digital library. The motivation is to develop management and visualization utilities that represent pedigree information. With such utilities, human beings are able to maintain and visualize genealogical information in a more efficient way.

## References

1. Battista, G.D., Eades, P., Tamassia, R., and Tollis, I.G., Annotated Bibliography on Graph Drawing. *Computational Geometry: Theory and Applications*, 4(5), 235–282, 1994.
2. Chen, Chao-chen, Chen, Hsueh-hua. and Chen, Kuang-hua. "The Design of Metadata Interchange for Chinese Information and Implementation of Metadata Management System." *Bulletin of the American Society for Information Science and Technology*, 27:5(June/July 2001): 21–27.
3. Munzner, T., H3: Laying Out Large Directed Graphs in 3D Hyperbolic Space. *Proceeding of Information Visualization '97*, Phoenix Arizona, 1997.
4. Munzner, T. and Burchard, P., Visualizing the Structure of the World Wide Web in 3-D Hyperbolic Space. *Proceedings of VRML'95*, San Diego, CA, 33–38, 1995.
5. Teraoka, T. and Maruyama, M., Adaptive Information Visualization based on the User's Multiple Viewpoints – Interactive 3D Visualization of the WWW. *Proceeding of Information Visualization '97*, Phoenix Arizona, 1997.
6. William Y. Arms, Christophe Blanchi, and Edward A. Overly. An architecture for information in digital libraries. *D-Lib Magazine*, February 1997.
7. Schatz, Bruce R. and Chen, Hsinchun. Building large-scale digital libraries. *Computer theme issue on the US Digital Library Initiative*, May 1996.
8. Kahn, Robert and Wilensky, Robert. A framework for distributed digital object services. *Technical report*, CNRI, May 1995.
9. Yeh, Jian-Hua, Chang, Jia-Yang, and Oyang, Yen-Jen, "Content and Knowledge Management in a Digital Library and Museum", *Journal of American Society for Information Science(JASIS)*, Special Topic Issue on Digital Libraries, 51(4), pp. 371–379, March 2000.