# THE DESIGN AND IMPLEMENTATION OF A CHINESE GENEALOGY INFORMATION SYSTEM

# 族譜資訊系統中知識管理之研究

*Jian-Hua Yeh**     *Chao-Chen Chen*†

葉　建　華*     陳　昭　珍†

*Post-doctorate    †Professor
*Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan 10617, R.O.C.
†Graduate Institute of Library and Information Studies, National Taiwan Normal University, Taipei, Taiwan 10610, R.O.C.

*博士後研究    †教授
*國立台灣大學資訊工程學系
†國立師範大學圖書資訊學研究所

## 摘　要

對中國人而言，族譜是一個了解家族歷史的最佳參考資料，族譜不但內含家族尋根的重要資訊，同時也包含了提供學術研究的重要參考。台灣族譜資訊網計畫的主要目的，就是想藉由數位化的方式，將族譜資訊進行保存，同時運用先進的數位化以及網路技術，將族譜資訊透過網際網路對外服務。本篇論文將針對台灣族譜資訊網計畫中所設計的資料以及系統架構，做一完整的說明。

**關鍵詞**：　族譜、詮釋資料、世系表、視覺化呈現、台灣族譜資訊網。

## 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 [1] 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 analyzes attributes of Chinese genealogical information, and 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 an 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 [2~4]. The object-oriented model is also applied in genealogy information system design. In this paper, we apply the object structures defined in NTUDLM [5] 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 HTML output 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(a) depicts the attribute and relation structure of a typical object in genealogy information system. Figure 1(b) illustrates the structures among objects.
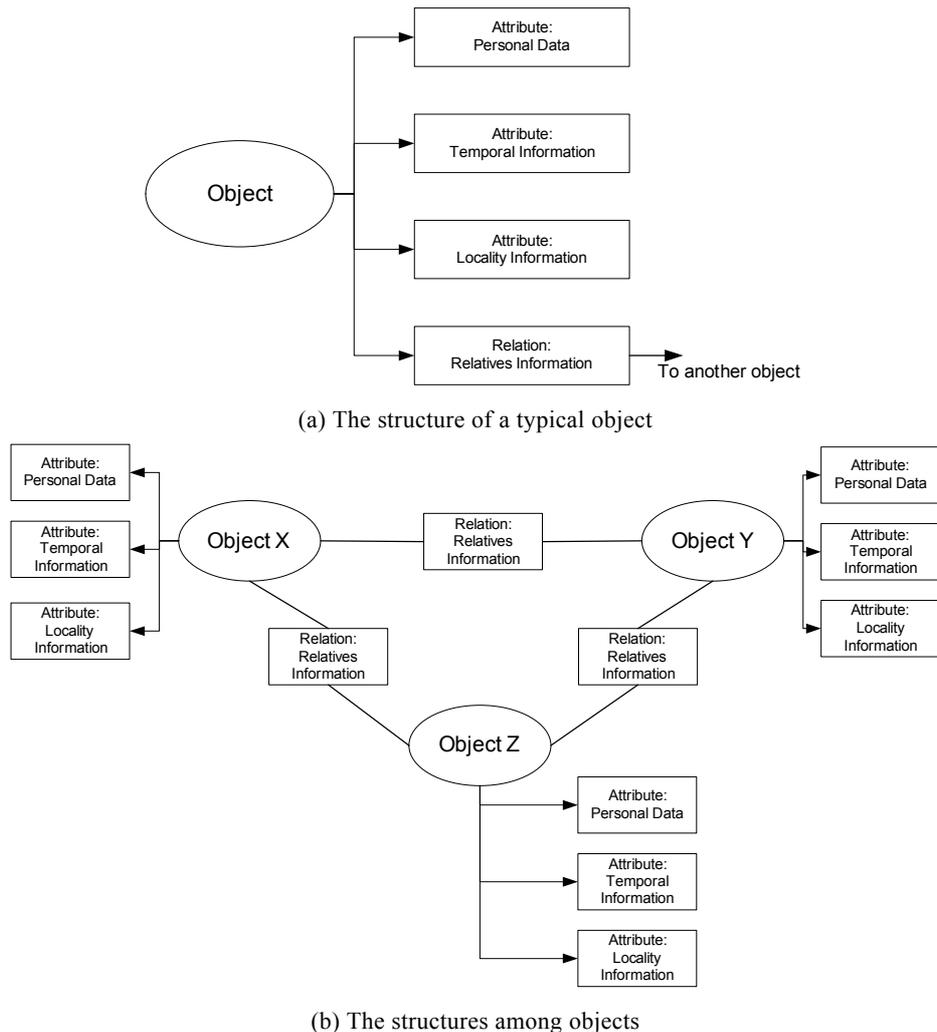


(a) The structure of a typical object



(b) The structures among objects

**Fig. 1　Object structures**

# 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 a 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.
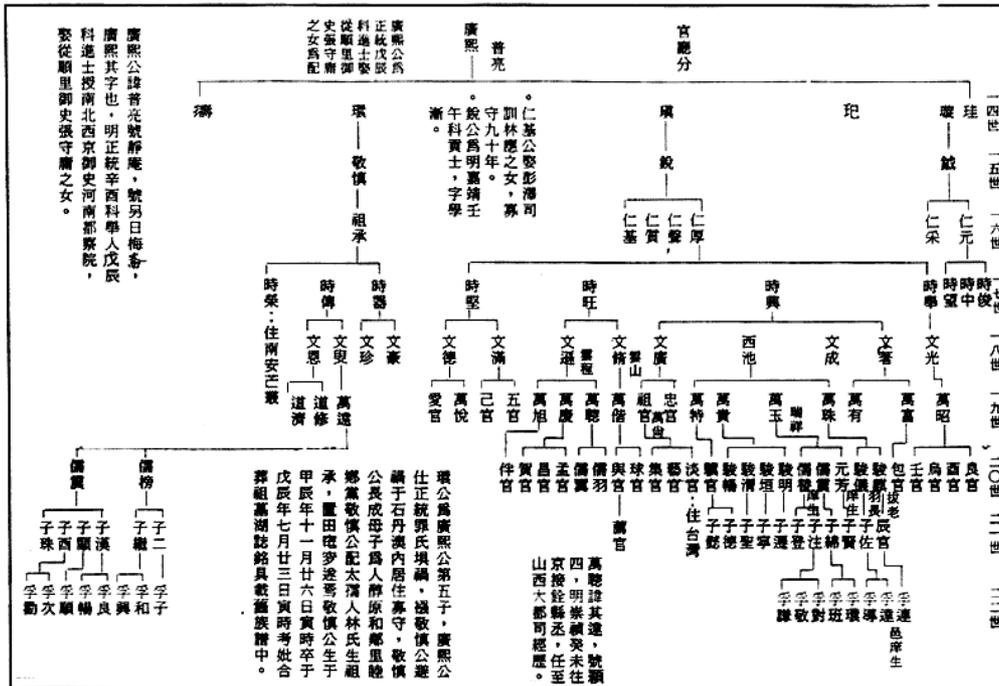


**Fig. 2    A page of traditional genealogy material**

**Table 1    Several chart types of pedigree tree presentation**

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

Because the Chinese genealogy material is based on the paternal system, the presentation of a Chinese pedigree is certainly a tree visualization problem. In the previous visualization researches [1], the visualization concerns of a graph or tree structure is based on several factors: node placement, link management, scalability, interactivity. As a result, the visualization of Chinese pedigree contains the following issues:

- How to present objects in a pedigree chart as many as possible?
- How to prevent user to get confused with a lot of object relationships?
- How to provide better interaction with user?

The above issues lead to the adoption of hyperbolic tree presentation [6~8] in our system, which is able to present a lot of tree nodes in a pedigree chart while keep good interaction with users. A typical hyperbolic tree is shown in Fig. 3.

## 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. Figure 4 depicts the system architecture of the system. There are 3 main subsystems, namely, the user interaction subsystem, the repository subsystem, and the metadata maintenance subsystem. The user interaction subsystem is the module that directly interacts with the user. Upon receiving a query from the user, the query processing module sends the query to the repository subsystem. When the repository subsystem returns the corresponding objects based on the query, the user choose the target metadata record, and the genealogy metadata information is presented to the user along with a visualized pedigree chart. Figure 5 shows a sample screenshot of the system.
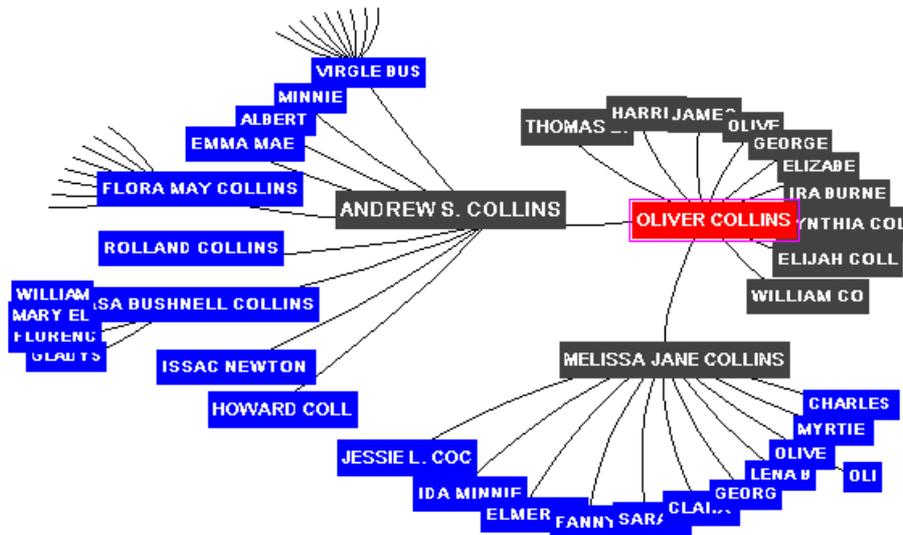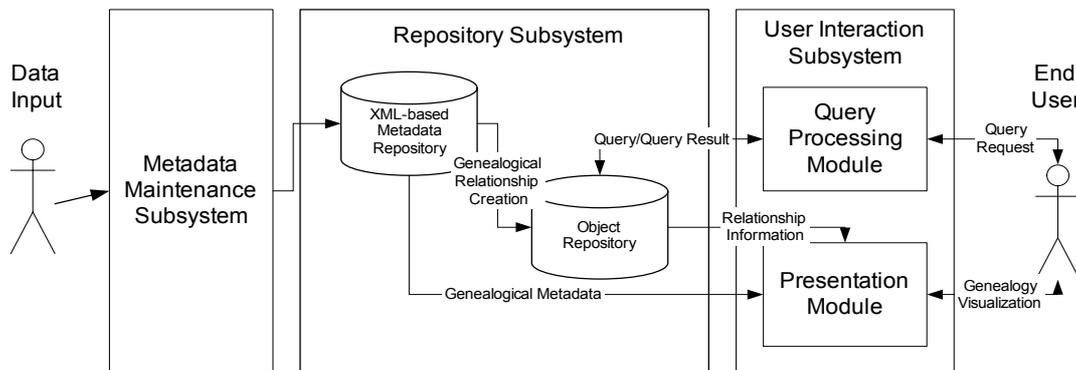


**Fig. 3   A sample hyperbolic tree**
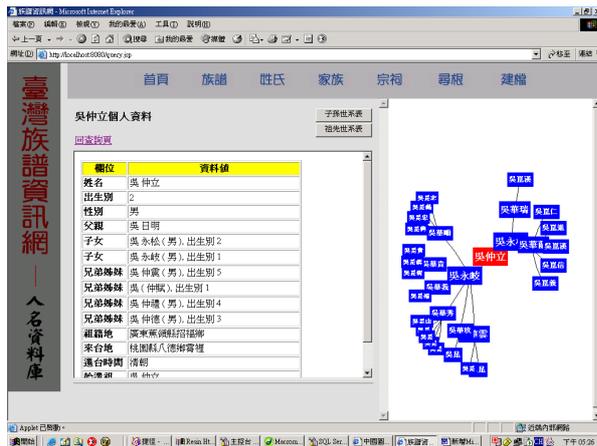


**Fig. 4   System architecture**

**Fig. 5    A sample screenshot of pedigree presentation**



**Fig. 6    A sample screenshot of the metadata maintenance subsystem**

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 processing module, the object repository is an object-oriented data store and its objects structures have been elaborated in Section 3.   Upon receiving a query from the query processing module, the object repository looks into the object structure and extracts the objects of interest.   Then it returns the extracted objects to the query processing module for packaging.

The XML-based metadata repository is the module that stores the contents of the genealogy metadata in a physical form.   On the other hand, the object repository stores the contents of the genealogy metadata in an abstract form.   The main reason why the same contents are stored in both an abstract form and a physical form (XML-based) is to facilitate both query processing and maintenance of the database contents.   The contents stored in an abstract form in the object repository is

structured to facilitate query processing and pedigree visualization, while the contents stored in a physical form in the XML-based metadata repository is structured to facilitate maintenance of metadata contents. 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.

This paper elaborates the object-oriented structures employed in the system to store the relationship information and the process of visualizing relationship data.   Though the discussion in this paper focuses on management and visualization of relations among the objects in a digital library that contains mostly genealogy information, similar idea can be applied to process other types of knowledge based on the contents of various types of digital libraries.

# REFERENCES

[1]  G. D. Battista, P. Eades, R. Tamassia and I. G. Tollis, *Annotated Bibliography on Graph Drawing*, *Computational Geometry*: *Theory and Applications*, Vol. 4, No. 5, 1994, pp. 235–282.

[2]  Y. W. Arms, C. Blanchi, and E. A. Overly, "An architecture for information in digital libraries," *D-Lib Magazine*, February 1997.

[3]  R. B. Schatz and H. C. Chen, "Building large-scale digital libraries," *Computer Theme Issue on the US Digital Library Initiative*, May 1996.

[4]  R. Kahn and R. Wilensky, "A framework for distributed digital object services," Technical report, CNRI, May 1995.

[5]  J.-H. Yeh, J.-Y. Chang and Y.-J. Oyang, "Content and knowledge management in a digital library and

museum," *Journal of American Society for Information Science (JASIS), Special Topic Issue on Digital Libraries*, Vol. 51, No. 4, March 2000, pp. 371–379.

[6] T. Munzner, "H3: Laying out large directed graphs in 3D hyperbolic space," *Proceeding of Information Visualization '97*, Phoenix, Arizona, 1997.

[7] T. Munzner and P. Burchard, "Visualizing the structure of the world wide web in 3-D hyperbolic space," *Proceedings of VRML '95*, San Diego, CA, 1995, pp. 33–38.

[8] T. Teraoka and M. Maruyama, "Adaptive information visualization based on the user's multiple viewpoints—interactive 3D visualization of the www," *Proceeding of Information Visualization '97*, Phoenix, Arizona, 1997.
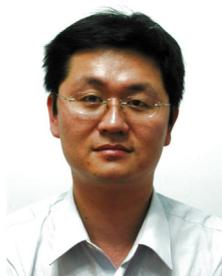
**Jian-Hua Yeh** (葉建華) received the B.S., M.S., and Ph.D. degrees in computer science and information engineering from National Taiwan University, Taipei, Taiwan, in 1992, 1995, and 2000, respectively. He is currently a postdoctoral fellow of Institute of Computer Science and Information Engineering, National Taiwan University. His current research interests include data clustering, knowledge discovery and management, digital library architecture, and concept and ontology processing.

**Chao-Chen Chen (**陳昭珍**)** received the B.S. degree in library science from Catholic Fu-Jen University, Taipei, Taiwan, in 1980, and the M.S. and Ph.D. degrees in library and information science from National Taiwan University, Taipei, Taiwan, in 1984 and 1994, respectively. Her current research interests include digital libraries and digital archives, library automation, information organization and metadata, electronic publishing, and electronic learning.