GIS-based Family Tree Information Sharing and Service

Di Hu, Guonian Lv*, Yongning Wen, Yiru Feng, Min Chen, Hongtao Zhang
Key Lab of Virtual Geographic Environment, Ministry of Education
Nanjing Normal University
Nanjing, China
*Corresponding author: gnilu@njnu.edu.cn

Abstract—To address the issues of family tree information sharing, mining and spatial-temporal analyzing, this paper introduces database, network and GIS technologies into the research of family tree. It studies the mode of family tree standardization, digitization, spatialization and networking, and discusses the theory foundation, implementing techniques, cooperation platform of family tree information sharing based on GIS. Firstly, according to the elements and characteristics of Chinese family tree by learning bibliographic and content of genealogy, this paper develops a description specification of Chinese family tree information. Then, based on consistently spatial-temporal framework, a family tree spatial-temporal data model cored with person and event is designed, and the system of family tree geographic information service is established. Finally, it proposes to construct a website group of the Family Tree GIS family tree geographic information service is established. Finally, it proposes to construct a website group of the Family Tree GIS integrating surname website, clan website and personal blog, in order to gradually build a community-based, open, multi-level platform for family tree information sharing and serving which serves for both general public and experts.

Keywords- family tree; GIS; consistently spatial-temporal framework; spatial-temporal data model; sharing and service

I. INTRODUCTION

Family tree is a kind of history record that systematically documents a clan of the same ancestor, which contains a large amount of information about society, economy, culture, population, history, geography, nation, custom, religion and so on [1]. Family tree, official history and local gazette are three significant elements in the history of Chinese culture. In recent years, looking up genealogy archives is quite popular and prosperous in Euramerican countries. And meanwhile more and more Chinese people, influenced by overseas Chinese who are interested in finding their ancestors, apply themselves to compile and investigate genealogy. Thus, demand and utilization of genealogy archives are rapidly increased. Currently, the quantity of Chinese genealogy has increased up to 100,000 volumes and more, most of which are collected by national and local libraries or individuals. However, it is a pity that so many genealogy archives are not fully exploited. Many experts point out that developing and sharing genealogy resources provide the benefits on cultural relics, literature, education and rooting to facilitate mining and resolving relevant problems in humanities and social sciences, and further promote the development of society, economy, politics and culture of China [2, 3].

II. REVIEW OF PREVIOUS WORK OF FAMILY TREE

Genealogies from different nations and regions are distinct in characteristics and rich in content, which are concerned by scholars of many fields. Jianfeng Chen et al. worked over the path and reasons of clan’s migration by studying genealogy of celebrity, minority and the masses [4, 5]. Guiling Xu et al. analyzed population distribution, origin, immigration of clan during different historical periods in different nations and regions [6, 7]. Cuirong Liu discussed relationship between clan population and changes in society and economy in Ming and Qing Dynasty [8]. From the perspectives of history, culture, religion, etc., Dallen and Jeanne analyzed the necessity of combining genealogy and geography [9]. Kashuba et al. traced ancestor’s migration sites, and analyzed the space rule on migration of clan via taking map as a tool [10]. Shupeng Chen discussed application of geography in religion, culture and landscape, and proposed to establish information system to manage houses and ancestral temples of Ming and Qing Dynasty [11]. Guonian Lv emphasized that people should pay more attention to study and mine the information about spatial-temporal distribution and migration of clan in genealogy and visually express spatial-temporal spectra of clan, and raised a viewpoint to construct the Family Tree GIS [3].

With the advancement of computer and network techniques, digitalization of family tree has been introduced. Based on established genealogy bibliographic database, genealogy full-text database is in the process of development, which implements functional transfer from family tree bibliographic inquiry to online editing and sharing family tree information. In addition, data exchanging and sharing among different genealogical systems are challenges. GEDCOM (GENealogical Data COMmunication) has been rolled out by the Family History Department of the Church of Jesus Christ of Latter-day Saints to unify the sharing and exchanging format of information of Euramerican countries’ genealogy [12]. The Library of Shanghai, China also has established the Genealogy Metadata Specification which describes bibliographic information of genealogy, in order to share genealogy bibliographic among digital libraries [13].

1 Supported by National Natural Science Foundation of China (40901186 and 40730527) and National High Technology Research and Development Program of China (2009AA12Z228).
Through the analysis of the previous work of family tree, four critical problems in the research of genealogy are summarized as follows: (1) Genealogy has great value for academic research, but it is difficult to obtain genealogy archives for researchers. And consequently, many studies just take individual genealogy as an example. (2) Map and GIS are important methods and technologies for studying spatial-temporal distribution and migration of clan, but not well used. So, spatial-temporal information implied in genealogy has not been well mined yet. (3) Since current situation is located in the construction step from genealogy bibliographic database to full-text database, data mining and spatial-temporal analyzing for genealogy is not completely supported by applying computer technology. (4) Genealogy information sharing is in a low level, and now China only has finished the sharing goal of bibliographic metadata.

III. DESCRIPTION SPECIFICATION OF CHINESE FAMILY TREE INFORMATION

A. Analysis of Bibliographic and Content of Chinese Family Tree

In the process of family tree’s development, the purpose and content of genealogy are increasing and changing, and the style is gradually improved and perfected. It has laid a solid foundation for regularly expressing and sharing genealogy information that many experts made some beneficial exploration in bibliographic and content of genealogy. Bibliographic information of genealogy is a description associated with bibliography itself which provides necessary information for people to understand and obtain genealogy. It consists of genealogy place, genealogy name, compiler and compiling mode, version, carrier form, binding form, annotation, abstract, collection unit and so on. Among them, abstract is significant content, generally including the simple description on the earliest ancestor, migrated ancestor, migrated origin and volumes.

Genealogy is rich in content which basically contains genealogical name, preface, genealogical comment, legend, catalogue, compiler detail, origin, lineage chart, Zibe, honor record, biography, clan rule and domestic discipline, ancestral temple, tomb, clan property, contract, writing, serial number and so on [1, 2]. Modern genealogy is usually attached with demographic chart, compared table of time, ancient and modern placename references etc. There is no standard for genealogical content. Some are more, some are less, some are brief and some are detailed. The genealogy also includes a large number of pictures and maps other than text, and modern genealogy may even contains a few audio and video materials. Person, relationship, event, time and place are key content of genealogy. Relationship implied in lineage chart is very complex which includes father-child, brotherhood, husband-wife, etc. In father-child relationship, heir frequently occurs in China Ancient, while adoption is very usual in China Today. In husband-wife relationship, polygamy is a universal social phenomenon in China Ancient, while divorce and remarriage also exist in China Today. In addition, some men become the visit son-in-law. Time information in genealogy are primarily composed of compiling time, birth time, death time and event time and expressed in various ways, most of which are in dynasty-reign title year or reign title-trunk lateral year or both, along with the Christian Era year which usually appear in modern genealogy. Place information in genealogy include spatial location or extent of genealogy place, birthplace, death place, tomb and ancestral temple, event place, migration place and so on, which are mainly represented by placename and simple map, and ancient placename is used more often.

B. Developing the Description Specification of Chinese Family Tree Information

A description specification of family tree information is the foundation to implement genealogy digitization and sharing genealogy information. The Library of Shanghai, China has established The Genealogy Metadata Specification which defines and normatively describes genealogy bibliographic resources. Thus, it is possible to share and interoperate genealogy resources among digital libraries. The specification only extracts the earliest ancestor, migrated ancestor and important person as abstract which is just a metadata of bibliographic, not a comprehensive description about genealogy content. GEDCOM has been rolled out by The Family History Department of The Church of Jesus Christ of Latter-day Saints. Based on it, modified specifications occur, such as GenXML, GedXML, GENTECH Genealogical Data Model and so on. So, expressing and sharing genealogy information of Euramerican countries have been implemented. However, there are still a few problems in describing relationship, event and spatial-temporal information. In fact, these specifications are so difficult to express unique content and element of Chinese family tree that they cannot satisfy the requirement of exchanging and sharing information of Chinese genealogy. There are a large number of genealogies in China, the informatization of genealogy also has being introduced, and demand for genealogy information consulting and sharing increases gradually. Therefore, it is urgent to produce a specification or standard that suits for Chinese genealogies to share and exchange information of family tree. According to the advantages of the Genealogy Metadata Specification in bibliographic information, GEDCOM in Euramerican countries’ genealogy information and KML, GML from GIS in spatial-temporal information, this paper puts forward new concepts of double time tag and period placename, then completely and explicitly express spatial-temporal information, and finally produce the Description Specification of Chinese Family Tree Information cored with person and event.

Combining with demand of genealogy research and application, and based on the analysis content of Chinese genealogy aforementioned in part 2.1, Chinese family tree information has been divided into three parts in this paper: family tree bibliographic, entry content and lineage chart. The first part describes metadata of family tree bibliography including genealogical name, genealogical place, clan’s name, compiler, version, abstract, collection unit and so on; the second describes the full content of Chinese family tree except for lineage chart; and the third one describes details of each person and family in lineage chart, relation and event. There is a great deal of information about time and space in genealogy, which are important to show clan’s lineage relationship,
multiplication and migration of clan. In terms of description of
time information, this paper designs a double time tag that
adopts two optional tags which simultaneously taking Chinese
traditional time and AD time into consideration. In fact, the
place in genealogy is a placename in specific time. As for the
frequent changes of placename, this paper introduces a concept
of period place that outlines temporal characteristic of a place,
and describes spatial location of a place with longitude and
latitude. The name of period place is described by a series of
placename ordered by administrative level. Lineage chart is the
most important information, among which the relationship
between persons is the core. This paper takes person as center,
and designs main family and affiliated family. Each family
consists of the husband and the wife, and the whole families in
a genealogy constitute a clan. The relationship between persons
is showed separately by clan object relation tag, which related
to a person, family or clan with identifier. Event in genealogy
involving with different individuals, families and even the
whole clan carries detailed information about time and place. It
is described with event tag, and relates to participant with
identifier. Based on these, the Description Specification of
Chinese Family Tree Information is constructed in this paper
by employing simple and composite element as basic elements.
The specification is implemented by XML and applies in the
Family Tree GIS for data exchanging and sharing. It is valid
and feasible.

IV. A FAMILY TREE SPATIAL-TEMPORAL DATA MODEL

A. Consistently Spatial-temporal Framework

According to the Description Specification of Chinese
Family Tree Information, time and space information in
genealogy have been expressed completely and regularly. Due
to be lack of a consistently spatial-temporal datum and
transformation rules, it is impossible to carry out temporal-
spatial position for family tree information by using computer.
Therefore, it could not construct spatial-temporal spectra of
clan. This paper proposes a consistently spatial-temporal
framework by using the Christian era year, Julian date and time
as temporal datum, the Chinese history administrative division
as spatial datum. It converts Chinese traditional time to the
Christian era year, and date of Julian and time according to
transformation rules of time. Thus, time information can be
positioned. It maps ancient placename to specific space
location or extent through ancient and modern placename
encoding. So, the space information also can be positioned.
Then, it could realize expressing time and space information in
a unique datum.

The basis of consistently temporal framework is to
construct a multi-precision, general time model. Multi-
precision is to meet the needs of multidisciplinary, multi-fields
research. This paper designs a two-level, multi-precision time
precision with historical-period-dynasty-year-month-day and
hour-minute-second-millisecond precision exists at the same time.
General indicates the model contemporarily supports Chinese
traditional year and the Christian era year, and date of
Gregorian calendar and Lunar calendar. It designs entities for
historical period, dynasty, dynasty stage, reign title etc. using
object-oriented method. Each entity records its start time and
end time with precision is year. And affiliations among them
are built one by one. In terms of date, it designs lunar month
entity with attributes of month type and the first date of a
month expressed in Gregorian calendar with precision is day.
Hour-second-millisecond precision uses the time model of
ISO19108. Based on this model, time database and time
transfer engine are developed, computer automatic parsing of
traditional time is achieved, and mutual conversion and
topological comparison between Chinese traditional time and
Christian era time with computer are implemented in this paper.

The basis of consistently spatial framework is to construct a
multi-scale, spatial-temporal data model of placename. Multi-
scale is to organize placename according to grade of Chinese
history administrative division. Here, spatial-temporal data
model highlights life period of placename, changing time of
placename and change of space location and extent. Systems,
such as CCTS and CHGIS, use map layer to organize Chinese
history administrative division by a dynasty or a year. They do
not take time as an attribute or a dimension, not take all
attributes of placename such as name, space location or extent,
subordinating relation, administrative grade into consideration,
and couldn’t express whole evolution process of placename in
its life cycle. Combining administrative division and ancient
and modern placename, and embracing object-oriented idea,
this paper expresses space extent information of placename by
using map layers organized by dynasty or year, as well as other
information of placename by using basic placename entity and
its affiliated entities. The basic placename entity emerges when
setting a new placename or changing name of an old
placename, and ends when changing name of an old name or
withdrawing a placename. It includes general name, proper
name and life cycle properties etc. The affiliated entities of
placename include name entity, space location entity,
subordinating relation entity depending on basic entity of
placename which are separately used to describe change
information and time of placename. The time involved in
placename information is expressed by using multi-precision,
general time model with precision is day. This paper
establishes Chinese history administrative division database,
ancient and modern placename database and seamlessly
connects the two databases, and develops ancient and modern
placename encoding service.

B. Family Tree Spatial-temporal Data Model

Family tree spatial-temporal data model bases on time
model and spatial-temporal data model of placename, and
includes family tree bibliographic model, item content model
and lineage record model. The first model describes
bibliographic information in sets of genealogies and relations
among a set of genealogies and a volume of genealogy, and
relation between bibliographic information and clan. The
second expresses item content excluding lineage record, and
relations between item content and clan. The third one
expresses information of individual, family and clan, events
related to them, and relations among them. The former two
models organize information in hierarchical structure
according to the analysis of the bibliographic information and
content of Chinese family tree described in part 2.1. They are
simpler than the third model. Due to space limitations, just a
brief outlook of them is given in this paper. Family tree lineage record model is the core of family tree spatial-temporal data model, as shown in Figure 1.

Individual entity includes given name, alias, nickname, sex, nationality information etc. Family entity is made up of husband and wife entities and it can be divided into main family entity and affiliated family entity. Affiliated family entity is related to main family entity, and records second wife and husband. Clan entity includes clan name, totem information etc. Clan object relation entity indicates relations among individual, family and clan. Main object attribute indicates the subject of clan relation, affiliated object attribute indicates the object with respect to main object. Both of them can be individual, family or clan. Relation type attribute marks type of main object and affiliated object, including individual-individual, individual-family, individual-clan, family-family, family-clan, clan-clan and other relations. Relation attribute illustrates relation between main object and affiliated object. Event entity includes attributes of name, subject, time, place, type, description. Subject attribute associated with individual, family and clan type attribute includes individual birth, death, family construction, migration and clan sacrifice. Family tree spatial-temporal data model cored with person and event information etc. Clan object relation entity indicates relations among individual, family and clan. Clan entity includes clan name, totem information etc. And it is also possible to get family tree or surname news according to surname, time or place. So, latest trends and information of family tree can be shared immediately. GIS data service includes Calendar data, Chinese history administrative division data, ancient and modern placename data and image data service. It could get time data in different ways according to specified filters, and get name, space location and extent, subordinating relation, administrative grade and change information of placename. And it is also possible to get tile maps of Chinese history administrative division since the Qin Dynasty and nowadays image. So, time expressed in different ways and basic data of historical geography can be shared fully.

Family tree geographic information function service includes family tree relevant function and GIS function service. Among these, family tree relevant function service includes family tree bibliographic and full-text retrieving, data converting, correlation and statistical analyzing service. It could retrieve family tree bibliographic and full-text, convert family tree data among different formats, import and export data based on the Description Specification of Chinese Family Tree Information, analyze similarity and correlation among genealogies according to main characteristics of family tree, and analyze count and features of family tree and population. GIS function service includes time encoding, ancient and modern placename encoding, GIS spatial-temporal analyzing, spatial-temporal spectra analyzing service. It could convert and analyze spatial-temporal distribution and evolvement of clan and population using time series analysis and spatial analysis methods. The spatial-temporal spectra of clan could be vividly presented in static map and dynamic map with timeline.

V. FAMILY TREE GEOGRAPHIC INFORMATION SERVICE AND SHARING PLATFORM

A. Family Tree Geographic Information Service

GIS Web Service is web service technology applies in the field of GIS, and meanwhile is a group of software function entities which are related to geographic information in internet.

It interacts with outside through interface of GIS web service defined by service description. Based on GIS, sharing of family tree information is in-depth combination of GIS and family tree, not only sharing family tree information, but also sharing data and function including spatial-temporal information and GIS function. Family tree geographic information service consists of two parts: data service and function service.

Family tree geographic information data service includes family tree data service and GIS data service. Among these, family tree data service includes family tree bibliographic and content metadata, full-text service. It could get and share metadata or full-text according to specified filters. And it is also possible to get family tree or surname news according to surname, time or place. So, latest trends and information of family tree can be shared immediately. GIS data service includes Calendar data, Chinese history administrative division data, ancient and modern placename data and image data service. It could get time data in different ways according to specified filters, and get name, space location and extent, subordinating relation, administrative grade and change information of placename. And it is also possible to get tile maps of Chinese history administrative division since the Qin Dynasty and nowadays image. So, time expressed in different ways and basic data of historical geography can be shared fully.

Family tree geographic information function service includes family tree relevant function and GIS function service. Among these, family tree relevant function service includes family tree bibliographic and full-text retrieving, data converting, correlation and statistical analyzing service. It could retrieve family tree bibliographic and full-text, convert family tree data among different formats, import and export data based on the Description Specification of Chinese Family Tree Information, analyze similarity and correlation among genealogies according to main characteristics of family tree, and analyze count and features of family tree and population. GIS function service includes time encoding, ancient and modern placename encoding, GIS spatial-temporal analyzing, spatial-temporal spectra analyzing service. It could convert and analyze spatial-temporal distribution and evolvement of clan and population using time series analysis and spatial analysis methods. The spatial-temporal spectra of clan could be vividly presented in static map and dynamic map with timeline.

B. Platform of Family Tree Geographic Information Sharing

Absorbing the VGI idea, the collaborative construction and sharing of family tree information are encouraged. This paper proposes to construct a website group of the Family Tree GIS integrating surname website, clan website and personal blog, so as to adequately make use of enthusiasm provided by public for sharing family tree information, and intend to explore a new mode of quickly collecting and organizing massive family tree information. The architecture of the Family Tree GIS website group is shown in figure 2.
The platform of the Family Tree GIS website group is the basic three-tier architecture: surname website, clan website and personal blog. Users can create family tree, edit and share information through a consistently interface. Any clan or individual can quickly build surname website or personal blog by using the service provided by the platform. Outside the platform, it uses a Web Crawlers of family tree to search and parse news in internet to provide news service of family tree for the platform. So, the platform and internet are combined deeply to promote socialization of sharing family tree information together. According to hierarchical design principles, the platform consists of three layers in the implementation: data layer, service layer and application layer. The architecture of the platform of family tree information sharing and serving is shown in Figure 3.

![Figure 2. Architecture of the family tree GIS website group](image)

Figure 2. Architecture of the family tree GIS website group

Data layer provides data supporting for the platform, including time database, Chinese history administrative division database and ancient and modern placename database. Service layer builds on data layer and is the core of platform including data service and function service of family tree geographic information. Application layer is the interface of platform. Users can visit interested website through website group and construct individual application by using service provided by the platform.

VI. CONCLUSION

Based on the analysis of bibliographic and content of Chinese genealogy, this paper develops a description specification of Chinese family tree information. According to various ways of time expression combined with fuzzy described spatial position, a multi-precision, general time model and a multi-scale, spatial-temporal data model of placename are brought to solve these problems. So, all genealogical information can be shared in a consistently spatial-temporal datum. Based on these, a family tree spatial-temporal data model cored with person and event is designed. This model has advantages such as presenting sufficient information, well-organized spatial-temporal information and flexible expansion. Absorbing the VGI idea, the collaborative construction and sharing of family tree information are encouraged. It designs a service system cored with family tree searching, time coding, ancient and modern placename encoding and family tree full-text retrieving service. Finally, the assumption to construct a website group of the Family Tree GIS integrating surname website, clan website and personal blog is proposed, in order to gradually build a community-based, open and multi-level platform of family tree information sharing and serving which serves for the public and experts, and supply a series of new ideas, methods, techniques and tools for better protecting and utilizing family tree.

ACKNOWLEDGEMENT

The authors would like to thank Feng Yu for her support and Mingguang Wu, Ziping Wang for their initial review, and meanwhile thank the group of the Family Tree GIS for their collaborative work.

REFERENCES


[12] Family History Department of The Church of Jesus Christ of Latter-day Saints, The GEDCOM STANDARD Release 5.5. The Church of Jesus Christ of Latter-day Saints, 1996.