Content Mining and Visualization of Traditional Genealogies of China
—Deployed on the Genealogy of Wu's in Gaoqian, Zhejiang

Zhenyuan Zhu [0000-0003-3964-253X]

School of Information Resource Management,
Renmin University of China, Beijing 100086, China
`zhuzhenyuan@ruc.edu.cn`

Abstract. **[Background]** Facing the elite figures in the traditional genealogy of China, this paper advocates that the traditional genealogy archives should be transformed into multi-dimensional resources by means of combining historical and humanistic analysis with digital technologies in order to analyze and display the family elite culture. **[Method]** This paper takes the genealogy of Wu's from Zhejiang Province of China as the research material, and constructs the genealogy ontology, in order to reveal the trend of digitalization of traditional genealogy research. **[Results]** This paper constructs the concept system and relationships of the genealogy ontology, and takes Wu Shilai, the ancestor of the 23rd generation of Wu's, as an example to realize the visualization of Chinese traditional genealogy. The significance of the visualization of traditional genealogy archives in "the construction of new rural talents" and "the construction of family heritage" is put forward.

Keywords: Chinese Traditional Genealogy, Content Mining, Visualization, Ontology Construction.

1 Introduction

As a valuable material in the fields of history, demography, sociology, economics, and folklore, traditional genealogy of China is now becoming an emerging research object in the field of digital humanities. Papery genealogy also faces the practical problem of preservation and protection. The aging, incompleteness, disaster and other conditions make the digital protection of genealogy urgent. How to transform these papery and plane genealogies into digital, multi-dimensional, comprehensive, time-sensitive resources is a problem worth studying.

From the literature review, with the requirements of the digitalization of ancient books and the development of digital humanities, the digitalization of genealogy archives has become a leading concern in genealogy research. In recent years, digital humanities, as an emerging research field, has developed rapidly, prompting many modern information technologies to be applied to the exploration of traditional humanistic materials, such as ancient books, scriptures, local Chronicles and archives. As one

of the most important archives of a family, genealogy is getting more attention from scholars. Scholars have discussed the documentation and digitalization construction of traditional genealogy, and combed the established genealogy catalogues and genealogy website [1], genealogy databases. At present, organizations represented by archives, libraries and family associations in China and churches abroad are committed to the collection of genealogies, sorting out genealogies from the perspective of collection construction, and some of the genealogy websites and databases have realized the function of browsing full-text pictures of genealogy. In addition, some genealogy websites have also realized the functions of people search and family root seeking [2, 3, 4].

In the study of genealogy resource construction, metadata construction of genealogy is the mainstream of research, the Xia C.J team of the Shanghai Library used bibliographic metadata to organize the contents of the genealogy [5], and realized the data mining of genealogy through the linked data and ontology technology [6]. In the existing research on genealogy contents, the visualization of genealogy lineage is a major aspect. Xiao Y et al. [7] constructed genealogy lineage model by using XML from the perspective of genealogy data description. Ontology has become an effective tool for genealogy visualization [8, 9], Jiang Y [10], based on ontology technology, described the relationships between figures in the genealogy by constructing knowledge map of genealogy, and realized the visualization of lineage of genealogy by constructing the system. Traditional genealogy contains a lot of literature value, historical value and sociological value, which need to be explored urgently.

In the history of the Wu's family in Gaoqian Village, Zhejiang Province, there were elite figures such as Wu Fu from the Northern Song Dynasty, Wu Jian from the southern Song Dynasty and Wu Shilai from the Ming Dynasty. They were all outstanding officials in Chinese history, not only have a role model for the descendants of Wu's, but also produce a significant social impact then and now. The Wu's family has completely preserved the "Wu's Genealogy", which has been passed down for thousands of years, with a long time span and well-preserved content.

Through the method of combining historical and humanistic analysis with content mining and visualization techniques, the content of Wu's Genealogy is excavated, the influence of different social relations and cultural customs on family changes is explored, and the social value and cultural value reflected by elite figures are summarized.

Through the method of ontology construction, the visualization of traditional genealogy is realized, and the relationship between people, geography, time, events and other elements in the family tree is displayed, which more intuitively shows the change of traditional family history and the inheritance of Chinese culture.

2 Research Method

2.1 Data Collection and Preprocessing

Gaoqian Village Committee provided the original "Wu's Genealogy", with a total of 59 volumes, including the biographies of Wu's elite figures and the lineage map of Wu's, as shown in Figure 1. Firstly, we use OCR software to perform text recognition and transcription on the scanned genealogy image. Through proofreading, punctuation

and interpretation, the papery genealogy ultimately forms a genealogical text for analysis.
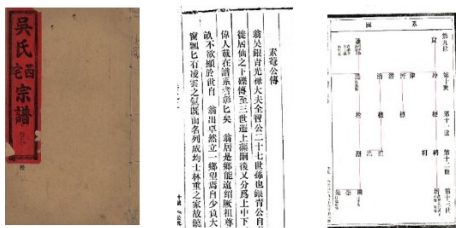

Fig. 1. Wu's Genealogy

2.2 Research Framework

The main research contents of this paper include the following aspects, as shown in Figure 2:

1. **Collection and preprocessing of genealogical data:** This paper adopts multi-source research materials, mainly Wu's Genealogy, supplemented historical materials, local chronicles, field research results, pictures and other materials.
2. **Analysis of the content and structure of traditional genealogy:** Chines Traditional genealogy contains family history, family lineage, biographies and other contents, it is necessary to organize and classify these contents, including the content composition of the genealogy, the stylistic rules and layout, the language features and so on. This part of work is to grasp the whole picture of traditional genealogy from a macro perspective and pave the way for the content analysis of family figures.
3. **Analysis of the content of family figures:** Content analysis of the family figures is the focus of this paper, including the classification of status of Wu's elite figures (officials, fame, sage, madam, etc.), the features analysis of elite figures (times, age, gender, location, etc.) and analysis of the virtues of Wu's elite figures (loyalty, filial valiant, etc.) and other aspects of the research content.
4. **Construction of knowledge organization system of genealogy:** By summarizing and integrating the above research contents, we build the knowledge organization system of genealogy, which mainly includes the construction of genealogy glossary, the construction of the concept system of genealogy, and relationship building of concepts of genealogy. The construction of the knowledge organization system of genealogy is the prerequisite for the visualization of elite figures.
5. **Visualization of traditional genealogy of China:** This part is the application practice of content mining results. Based on the constructed knowledge system of genealogy, it realizes the visual representation of elite figures with the help of ontology technology, generates the semantic relationship of elite figures based on Protégé, and realizes the visualization and organization of elite figures' contents.
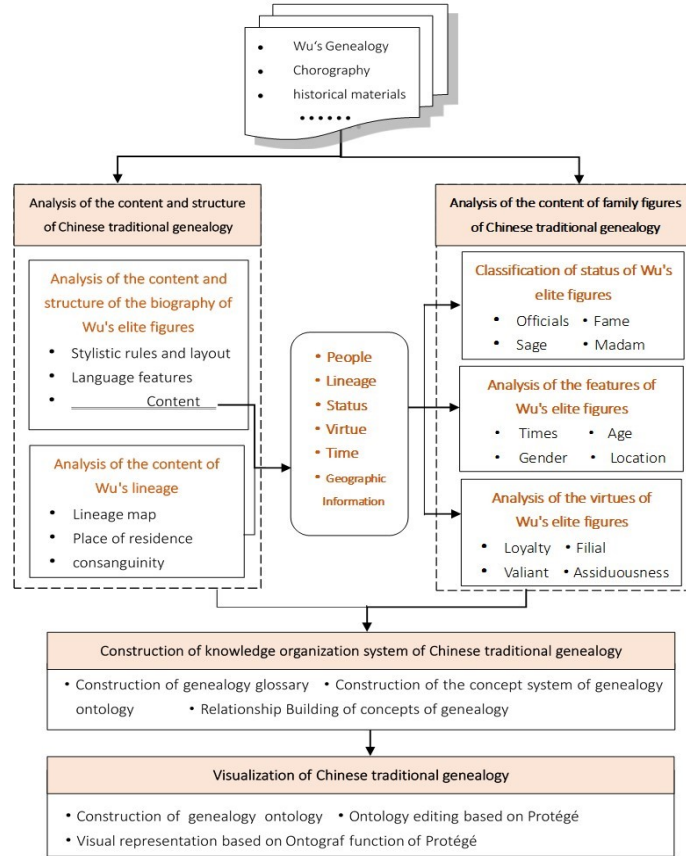
Fig. 2. Research Framework

3 Preliminary Results

3.1 Construction of the Concept System of Genealogy Ontology

By investigating the relevant historical figures ontology and combining the characteristics of Wu's Genealogy, this paper divides the concepts in the genealogy ontology into 6 basic categories: people, lineages, status, virtues, time, and geographic information:

- **People:** People is the most important content recorded in the genealogy, and are the main body of the family's development, mainly including the elite figures who owned the biographies and related figures such as their daughters, the authors of their biographies, and so on.
- **Lineage:** Lineage is an important part of genealogy, showing the continuing development of a family, from the first generation to the next generation.

- **Status:** Status is an important part of describing the role of people in society and the social relations of people in genealogy, as well as their achievements and honors. Status includes the fame or office of a person, and the honor he has received.
- **Virtues:** Virtue is a generalization of a person's moral characteristics, reflecting the people's merits.
- **Time:** Time refers to the dynasty in which the people are located.
- **Geographic information:** Geographic information refers to the description of the geographical location of the people in genealogy, including the place of residence, place of official and other geographic information.

3.2 Relationship Building of Genealogy Ontology

In ontology construction, an important component are properties of ontology. There are two kinds of properties in ontology. The first one is datatype properties, which describes the relationship between an instance and a basic datatype. The second one is object properties, which describe the relationship between instances and serve as the "link" between instances. The property building of ontology can be used to map the relationships between concepts in the ontology.

The genealogy ontology in this paper mainly contains 25 kinds of relationships, as shown in Figure 3, including 7 datatype properties (such as "name", "countesyName") and 18 kinds of object properties (such as "hasStatus", "exileIn").
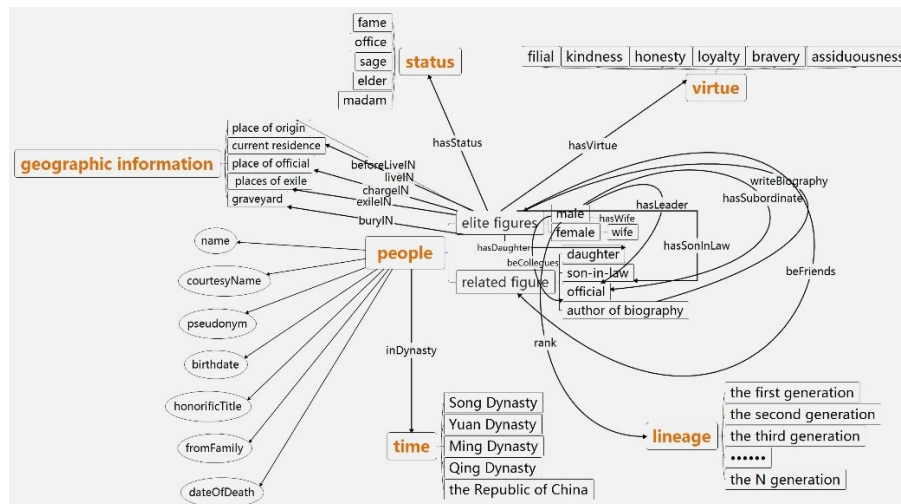

Fig. 3. Relationships of Genealogy Ontology

Through 25 kinds of relationship mapping, the people, time, geographic information, status and other elements involved in the genealogy can be linked to form a relationship network.

3.3 Visualization of Chinese Traditional Genealogy

We use Wu Shilai, the ancestor of the 23rd generation of Wu's, as an example. Wu Shilai (A.D 1527—1590), served in the Jiajing reign of the Ming Dynasty, with Zhang Juzheng, Xu Jie, Yan Song as colleagues. Wu Shilai was upright and brave, inheriting the good spirits of the Wu's family, which makes him to be a representative of the Wu's. The biography of Wu Shilai is recorded in the Wu's Genealogy, describing the role of Wu Shilai's patriotism. "Loyalty" has become a summary of Wu Shilai's moral quality, which also becomes an important spiritual culture of Wu's descendants.
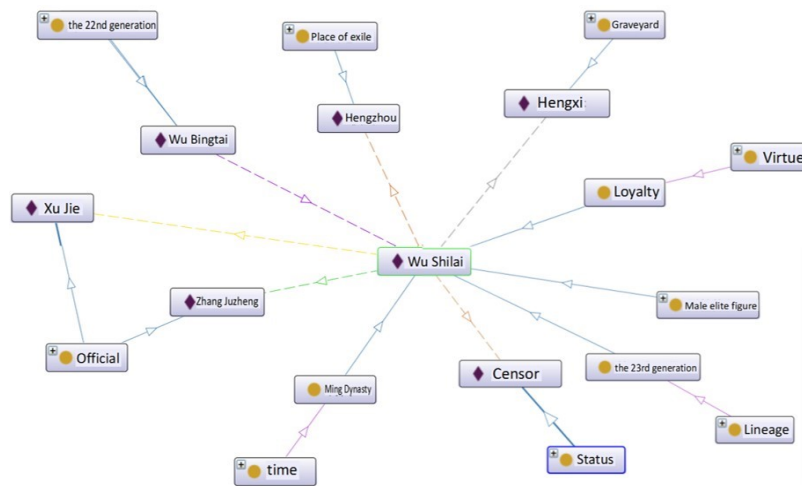

Fig. 4. Semantic relationship of "Wu Shilai" based on ontograf of Protégé

Through the construction of semantic relationship based on ontology, as shown in Figure 4, Wu Shilai's important aspects of life are displayed, especially his social relations can be more clearly reflected, forming a multi-dimensional visual display of this elite figure.

4 Discussion and Implication

China's traditional genealogy archives record the history of a family's prosperity, the deeds of sages of the family, and the social relations of the family. It contains rich social and cultural values and is a valuable archive of China.

Through the combination of historical humanistic analysis and digital technologies, we can dig deep into the rich contents of the genealogy and realize the visualization of the genealogy, and thus transform the traditional papery genealogy into multi-dimensional digital resources.

- The visualization of traditional genealogy reflects the requirements of digitalization of research, providing methodological inspiration and reference for the study of Chinese traditional genealogy archives.
- The study of elite figures in Chinese traditional genealogies provides a historical trace for the construction of "new rural talents". In the process of "new rural construction", new rural talents are important parts, and the research of the elites of the traditional genealogy archives will provide a solid impetus for the rural culture construction.
- The research of traditional genealogies provides a powerful way for the construction of "family heritage". Family generation of spiritual culture is an important part of the genealogy, which is hidden in the deeds of the family's sages. By digging into the deeds of the elites and displaying them through visual means, the inheritance of the family spirit and culture will be better promoted.

The research of traditional genealogies has rich academic values and important practical significance. As a frontier direction of research on digitalization, content mining and visualization of genealogy is worthy of further exploration.

References

1. Zhao W.: Research and Collation of Genealogies. China Science and Technology Information 25 (5), 62-63+66 (2013).
2. Ancestry | Genealogy, Family Trees & Family History Records, https://www.ancestry.com/, last accessed 2019/9/21.
3. Werelate-Main Page-Genealogy, https://www.werelate.org/wiki/Main_Page, last accessed 2019/9/25.
4. Free Family History and Genealogy Record-FamilySearch.org, https://www.familysearch.org/records/archives/web/, last accessed 2019/9/25.
5. Cuijuan X.,Wei L.,Lei Z., et al.: A Genealogical Ontology in the Form of BIBFRAME Model. Library Tribune(11), 5-19 (2014).
6. Cuijuan X., Wei L. et al. :A Genealogy Data Service Platform Implemented with Linked Data Technology.Journal of Library Science in China 42(3):27-38 (2016).
7. Yu X.:Study on genealogy lineage data model, National Library of China Publishing House, Beijing (2016).
8. Jidong Z.: Study on knowledge management of genealogy based on ontology molecule. Library work and study (2),78-83 (2017).
9. Ying J., Hui D.:Ontology Based Knowledge Modeling of Chinese Genealogical Record.In :WSCS '08 Proceedings of the IEEE International Workshop on Semantic Computing & Systems. IEEE Computer Society, pp.33-34 (2008).
10. Ying J.，Jing Z.，Lingxuan Z.: Ontology based knowledge graph model of genealogical record and retrieval system. Electronic design engineering 25(12),161-165 (2017).