Knowledge Organization and Interlinking of Local Celebrity Documentary Resources from a Digital Humanities Perspective: Postprint

Authors: Wang Ziyi, Chen Chen, Wang Xianghua

Date: 2023-10-08T00:00:00+00:00

Abstract

[Purpose/Significance] Starting from the current status of local celebrity literature resource construction, this study explores knowledge organization and interlinking schemes for multi-source heterogeneous resources of local celebrities from a digital humanities perspective, providing reference for GLAM institutions to develop local figure knowledge management and services. [Method/Process] Referring to relevant celebrity resource development approaches both domestically and internationally, we propose a four-step knowledge organization scheme for local celebrity resources, including heterogeneous literature resource content organization, local celebrity resource ontology construction, entity and entity relationship integration, and local celebrity resource knowledge application. By combining the characteristics of local celebrity literature resources with the Person Resource Description Framework, we constructed a local celebrity literature ontology model CLO. [Results/Conclusion] Taking the Xiangxi poet Tian Mingyu and his manuscript "Kuxuezhai Diary" as a case study, we implemented knowledge organization and relationship revelation of local celebrities and their diary works following the organization steps and using the Protégé tool, thereby verifying the feasibility and operability of this organization scheme. This work not only broadens the perspective of local literature development but also provides reference for the construction of celebrity knowledge bases and the development of characteristic humanities services in ethnic regions.

Full Text

Preamble

Digital Humanities Perspective on Knowledge Organization and Correlation of Local Celebrity Literature Resources

Wang Ziyi[1], Chen Chen[1], Wang Xianghua[2]

[1]School of Tourism and Management Engineering, Jishou University, Zhangjiajie 427000
[2]School of Literature and Journalism, Jishou University, Jishou 416000

Abstract

[**Purpose/Significance**] Starting from the current status of local celebrity literature resource construction, this study explores knowledge organization and correlation solutions for multi-source heterogeneous resources of local celebrities through a digital humanities perspective, providing a reference for GLAM institutions to develop local figure knowledge management and services.

[**Method/Process**] Referencing domestic and international development approaches for celebrity-related resources, we propose a four-step knowledge organization scheme for local celebrity resources: (1) content organization of heterogeneous literature resources, (2) ontology construction for local celebrity resources, (3) entity and relationship fusion, and (4) knowledge application of local celebrity resources. Combining the characteristics of local celebrity literature resources with figure resource description frameworks, we constructed a local celebrity literature ontology model called CLO.

[**Result/Conclusion**] Using the Western Hunan poet Tian Mingyu and his manuscript *Kuxuezhai Diary* as a case study, we implemented knowledge organization and correlation revelation for the local celebrity and his diary works following the organizational steps and using the Protégé tool, thereby verifying the feasibility and operability of this organization scheme. While broadening the development perspective of local literature, this study also provides a reference for building celebrity knowledge bases and developing characteristic humanistic services in ethnic regions.

Keywords: digital humanities, celebrity literature, knowledge organization, ontology construction, Protégé

---

1. Introduction

As a critical nexus between intelligent technology and humanities disciplines, digital humanities encompasses knowledge organization and discovery methods such as natural language processing, spatiotemporal information revelation, and multi-dimensional knowledge co-occurrence that can effectively advance the construction of new liberal arts. Its cutting-edge application examples have consistently attracted attention from resource development institutions and scholars across various fields [1]. Local celebrity literature resources serve as textual carriers of cultural, artistic, economic, and political evolution in specific periods and regions. They constitute important materials for interpreting regional intellectual trends, exploring local historical transformations, and shaping local cultural identity, while also serving as treasure troves of memory revealing the

evolution of local social life, distinctive folk customs, celebrity social networks, economic development trajectories, and political vicissitudes [2].

Currently, under the 加持 (support) of digital humanities technologies, many hidden precious resources have been reorganized. With the gradual increase of open platforms such as "Genealogy Knowledge Service Platform," "Sheng Xuanhuai Archives Knowledge Base," "Li Zhengdao Digital Resource Center," and "Academic Growth Database of Senior Scientists,"many previously neglected celebrity resources are gaining attention. However, examining the development status of celebrity literature resources in ethnic regions such as the Xiangxi Tujia and Miao Autonomous Prefecture, constrained by human resources, material resources, financial resources, and influence, literature development remains at the stage of partial organization, yielding few achievements and limited forms. There is a need to draw upon digital humanities perspectives and technologies to explore knowledge organization and correlation solutions for local celebrity literature resources.

Through the Web of Science and CNKI databases, we can systematically review achievements in the digital humanities field that focus on figures and their works. These studies can be categorized into four aspects:

1.1 Reorganization of Figure Resources

Xie Man collected literature on female figures across ancient, contemporary, and modern periods in education, culture, art, economy, and technology according to self-built metadata schemes, providing a foundation for personalized knowledge services on female figures [3]. Liu Chaolin et al. explored the precision of named entity recognition for historical event research under language models and conditional random fields, mining core knowledge from over 220 volumes of *Local Gazetteers* and evaluating algorithm accuracy based on extracted official information from various dynasties [4]. Zhao Yifei proposed a knowledge aggregation framework for Chinese name information norms based on commonly used knowledge base entity description standards at home and abroad, providing a foundation for refining figure feature revelation, annotation, and heterogeneous knowledge linking and sharing [5]. Korean scholar J. W. Kim reorganized correspondence from Korean Protestant missionaries between 1880 and 1942, achieving revelation of textual personality and character [6]. Argentine scholar N. Zorrilla et al. used works and manuscripts of early female philosophers as entry points to promote research output and field development through digital humanities paradigms, analyzing factors that excluded early female philosophers from the canon and reflecting on the historical reasons and political backgrounds behind this phenomenon [7].

1.2 Sharing and Linking of Special Collection Resources

Russian scholar Y. M. Lupanova et al. integrated existing "memory places" and "memorial figures" characteristic resource bases for Lomonosov, introducing personal diaries, manuscripts, and archival resources to supplement complete figure images and clarifying that data participation in figure image construction can ef-

fectively enhance students' interest in celebrity lives and experiences, achieving "hero de-idolization" while expanding educational perspectives [8]. American scholar H. Kun et al. reshaped the theme resource base "Roman Historical Figures in Refined Roman Coins" through heterogeneous data sources, character extraction technology, WEB applications, and intelligent operation portals, constructing a relatively complete digital cultural heritage service system [9]. Russian scholar A. Bonch-Osmolovskaya et al. used Tolstoy' s 90-volume important printed literature as a foundation, conducting metadata tagging, letter tagging, and diary tagging for three types of texts (works, letters, diaries) to achieve digital portals and semantic graph indexing, and built a figure knowledge base referencing open linked databases such as DBpedia [10]. Chen Zhiming et al. proposed a Chinese ancient book digital humanities research platform (CABDHRP) supporting Chinese historical research, featuring automatic text annotation (ATAS) and exploration of character social network relationships (CSNRMT). The platform uses the open-source institutional repository DSpace as a digital archive system for resource archiving and image/full-text scanning, links different databases (e.g., CBDB, TGAZ) and ancient text interpretation data sources through JavaScript frameworks, and uses non-relational databases such as Neo4j for figure interaction knowledge retrieval and graph presentation [11].

1.3 Social Networks and Spatial Distribution of Figures

Huang Junjie et al. proposed a figure research framework composed of symbolic graph models and grouping algorithms, verified its framework's effectiveness and practicality using figure social network architecture formed by the China Biographical Database (CBDB) [12]. Korean scholar H. H. Ji et al. conducted figure research on Joseon scholar Seo Geojeong's kinship and political activities, revealing factors influencing historical figure relationship formation—including state affairs, royal relationships, diplomatic situations, political backgrounds, and regional customs—through network visualization software such as Bubbles [13]. Korean scholar S. Bae applied CBDB, QGIS, Pajek, MARKUS, and DocuSky to generate spatiotemporal graphs and genealogical trees, investigating the imperial examination status and evolution during the Southern Song Dynasty, and 梳理 (sorted out) the social network relationships of core figures in Southern Song historiography based on the experiences and social circles of Lu Jiuyuan, Zhu Xi, and Lü Zuqian [14]. Xu Yongming visualized the travel trajectories of Ming Dynasty dramatist and writer Tang Xianzu using literature materials and geographic information systems such as QGIS, CHGIS, and ARCGIS, and revealed the social relationships among Tang Xianzu, Tu Long, and Wang Daoqun using the CBDB database and Gephi software [15]. Additionally, scholars have used CBDB and CCTS (Chinese Civilization Spatiotemporal Framework) to explore the spatiotemporal evolution of Song Dynasty bureaucratic families [16]; organized Song Dynasty scholars' master-disciple relationships and dynamically revealed figure networks [17]; combed through interpersonal interaction genealogies of Ming-Qing Jinshi families using genealogies [18]; built an official domain ontology based on *Changchun County Gazetteer: Changchun Official*

Annotations to reveal fine-grained knowledge associations among Qing Dynasty officials [19]; and explored the social network and spatiotemporal evolution patterns of exiled poets using *Complete Tang Poems* digital texts [20]. Song Xueyan et al. also used digital humanities tools such as Gephi, QGIS, NLPIR, and LTP to conduct entity extraction, linking, and visualization of names, places, and emotional words, completing research on figure network relationships, spatial presentation of hot events, and emotional polarity discrimination in Wang Shijie' s diaries [21].

1.4 Knowledge Organization and Correlation of Figure Resources

Liu Ningjing et al. proposed an academic celebrity knowledge organization framework referencing FOAF framework, CBDB architecture, Shanghai Library celebrity manuscript model, and CERIF resources, and initially built the Li Zhengdao Digital Resource Center using paper literature, physical objects, and audio-visual materials [22]. Li He et al. designed an ontology model for Republic of China historical events based on trigger word recognition, typical event screening, historical event constituent elements, and existing ontology reusable concepts and attributes, conducting event extraction, reorganization, and microsociety presentation to verify the model' s effectiveness [23]. Yao Tianhong et al. introduced the CIDOC-CRM conceptual model based on the Zhang Xueliang historical resource ontology framework, achieving "event-figure-spatiotemporal" knowledge co-occurrence through semantic annotation and relationship revelation of letter resources, and proposed a knowledge organization scheme based on semantic technology [24]. Wei Jingzhu et al. discussed the application value and implementation approaches of knowledge graphs in digital humanities knowledge organization using the knowledge graph creation of three core figures from the Hundred Schools of Thought (Confucius, Laozi, Mozi) as examples [25]. British scholar A. D. Cheok et al. applied natural language processing technology and machine algorithms to Confucian ritual knowledge and teaching modeling by exploring natural dialogue between humans and computers, developing a knowledge organization system that allows experiencing Confucian teachings through virtual and real interaction. This system enables users to intuitively understand the full picture of various intangible heritages, measure philosophical intentions through interactive dialogue, and generate novel content and answers [26]. Niu Li et al. designed a multi-granularity knowledge organization scheme based on celebrity archive resource memory units, revealing figure information, thoughts, social experiences, and family life behind recorded cases, confirming the value of domain ontology models in global discovery and knowledge mining [27]. Zhang Yunzhong et al. improved the historical and cultural celebrity study tour footprint knowledge organization framework based on CBDB figure database architecture and Shanghai Library' s name data, ancient books, and local gazetteers, referencing poetry websites and figure chronologies. Using relational database systems, Navicat management systems, D2R conversion tools, and LODLIVE visualization software for data storage, transformation, publishing, browsing, querying, and graph construction, they achieved knowledge discovery of historical celebrities' study tour footprints [28].

In summary, domestic and international research on celebrity literature resources has focused primarily on the above four aspects, with fewer achievements in knowledge organization scheme research for celebrity literature resources. No research has yet emerged on knowledge organization of local celebrities and their works, and there is a lack of corresponding knowledge organization schemes and reusable domain ontology models. Based on this, this paper designs a set of knowledge organization schemes capable of revealing granular knowledge from multi-source heterogeneous resources of local celebrities, aiming to achieve "explicit-implicit" verification of literature knowledge while providing references for local GLAM institutions to develop figure knowledge management and services.

---

2. Design and Analysis of Knowledge Organization and Correlation Scheme for Local Celebrity Literature Resources

Natural language processing and ontology model construction can transform unstructured texts into structured knowledge, thereby achieving knowledge correlation and resource sharing. Since handwritten and printed resources in local celebrity literature require manual collation, the organization scheme for local celebrity literature resources should adopt a human-computer 互助 (mutual assistance) form encompassing the following functions: completing thematic identification and content correction according to editorial conventions (font, arrangement, modern punctuation); using natural language technology for entity recognition and relationship extraction of celebrity literature knowledge (figures, places, events, behaviors, emotions); standardizing domain basic classes and inter-attribute relationships through self-built ontology models (figure ontology, literature ontology, etc.); revealing different granular knowledge networks using humanities graph software (Gephi, QGIS, Cytoscape); storing and querying interconnected knowledge using graph databases (Neo4j); applying knowledge granularity values presented in graphs to distinguish organization result quality and conduct structural adjustments, thereby laying a foundation for local GLAM institutions to develop knowledge services such as figure navigation, semantic retrieval, correlation reasoning, and humanities discovery. The specific design 思路 (approach) is shown in Figure 1 [Figure 1: see original paper], with key steps analyzed below.

2.1 Content Organization of Heterogeneous Literature Resources

Organize celebrity archives and manuscript works according to editorial conventions for similar genre publications. Before text digitization, since local celebrity literature contains many manuscripts with some already damaged, 原生性保护 (primary protection) of damaged manuscripts must be conducted to ensure subsequent text scanning. During digitization, due to personal charac-

teristics in manuscript fonts, illustrations, and tables, an "OCR + human proofreading" model assists in identifying text themes and correcting content, forming standardized, recognizable, and augmentable electronic documents. After digitization, following an "NLP + human proofreading" model, natural language processing platforms such as LTP, NLPIR, and ROST assist researchers in entity recognition and relationship extraction tasks. Extracted elements are verified for accuracy, matching degree, and completeness based on textual content, and texts are corrected according to background historical materials to prepare for subsequent knowledge organization and ontology assignment.

2.2 Ontology Construction for Local Celebrity Resources

After organizing heterogeneous resources, focus on screening and processing semi-structured and unstructured text resources. Based on metadata standard systems such as CNMARC and DC, use natural language processing technologies (entity recognition, relationship extraction) to extract thematic, personal, place, event, emotional, and stylistic knowledge at different granularities from scattered celebrity literature resources. Combining the relevance, universality, and recognizability of extracted knowledge with the China Biographical Database (CBDB) and other knowledge base frameworks, jointly compile figure and literature term vocabularies. Divide basic class hierarchies, definitions, and attributes according to domain core concepts, construct self-built ontology models, and supplement missing classes and expanded attributes through verification and communication, thereby building an ontology model suitable for local celebrity resource organization and development.

2.3 Entity and Relationship Fusion

After information extraction, we cannot ignore the large amount of non-standard expressions (nicknames, abbreviations, dialects, etc.) in non-formally published literature (diaries, speeches, interview transcripts). Such expressions generate redundant or erroneous information that affects accurate definition of figure relationships and biased behaviors, reducing instance ontology clarity and humanities graph construction quality. Therefore, fusion methods are needed to resolve and disambiguate multi-source knowledge. For entities and relationships with synonymous but different names, use similarity functions or reasoning models to resolve co-reference conflicts. For entities and relationships with identical names but different meanings, use specific lists and linking systems to eliminate referential ambiguity, merging multi-source heterogeneous knowledge to solve knowledge matching and correlation dilemmas, providing a basis for dynamic local celebrity resource management, "explicit-implicit" knowledge discovery, and shared interaction platform design.

2.4 Knowledge Application of Local Celebrity Resources

Merged local celebrity literature resources are stored in knowledge bases in layers as knowledge units. Users can conduct figure navigation, semantic retrieval,

and humanities knowledge discovery according to needs or purposes. For example, link local figure and related literature knowledge bases through Resource Description Framework (RDF) for figure navigation; use query languages such as SPARQL and Cypher (Neo4j) for target knowledge network retrieval; discover associations among local celebrity resource instances through logical operations and reasoning mechanisms; and achieve characteristic resource knowledge discovery from perspectives such as textual content, paragraph emotions, spatiotemporal networks, and interactive behaviors using humanities graph software.

---

3. Characteristics of Local Celebrity Literature Resources and Description Frameworks

3.1 Characteristics of Local Celebrity Literature Resources

Local celebrity literature resource development is constrained by human and material resources, mostly existing as unorganized manuscripts and unstructured texts to be processed. As resource development evolves, basic classes and attributes require continuous supplementation and adjustment. Local celebrity literature resources constitute a relatively special category among characteristic resources preserved by GLAM institutions, comprising directly related literature (manuscripts, diaries, poetry, lectures, monographs) and indirectly related literature (biographies, novels, genealogies, local gazetteers). To organize different genres, carriers, and thematic literature and figure materials using ontology models requires analyzing celebrity literature resource characteristics and referencing figure knowledge frameworks. Characteristics can be summarized as follows:

1. **Rich resource types.** Local celebrity resources are important collections recording regional evolution, historical changes, folk culture, and religious beliefs. Celebrity diaries, poetry, and lectures are also important bases for revealing ethnic changes, language characteristics, local customs, and mythology. Celebrity photos, videos, and artifacts (handicrafts, stone carvings, woodcuts) are important materials for local cultural memory projects. Therefore, during resource organization and classification, besides processing paper-based textual literature, attention must be paid to other forms such as audio-visual recordings (photos, tapes, videos) and physical objects (handicrafts, stone carvings, printing blocks).
2. **Complex text structures.** Since most celebrity literature resources are preserved as manuscripts, different works vary significantly in cultural background, publication norms, language preferences, paragraph conception, and font forms. Limited by machine collation quality, using algorithms and machine learning for core knowledge recognition and extraction from celebrity literature resources yields unsatisfactory accuracy and 容易

出现 (easily produces) vocabulary redundancy and semantic mismatches. Therefore, following editorial conventions, manual processing of characters, syntax, and word order in texts to be processed is necessary to ensure that classical Chinese, vernacular (spoken and written) from different periods can be translated into uniform textual forms to improve language processing quality.

3. **Broad resource perspectives.** Directly related literature (e.g., self-authored works and personal records) provides instance materials for literature resource organization, while indirectly related literature (e.g., celebrity life experiences and career changes) provides corroborating materials for figure knowledge 梳理 (organization). Additionally, reference materials and citation literature must be introduced to improve the applicability of figure resource organization schemes—expanding text volume through figures and works mentioned in resources, and enhancing text breadth through poems and events in reference materials.

4. **Concealed thematic relationships.** Due to complex content and diverse forms of celebrity resources, relying solely on natural language processing technology makes it difficult to achieve high-accuracy relationship extraction and discovery of heterogeneous texts on the same topic. Therefore, for relatively concealed related themes, use "figures" as the 基点 (anchor point), conduct in-depth investigation of celebrity backgrounds and interpretation of work content, and combine official career experiences, historical status, key events, and personal interests for manual judgment and screening to ensure reasonable and accurate heterogeneous text relationships.

5. **Obvious cross-domain nature.** Although GLAM institutions belong to resource preservation organizations, their research focuses and directions differ, involving multiple disciplines including library science, archival science, history, archaeology, and computer science. Therefore, when building celebrity resource organization schemes, multiple opinions should be 采纳 (adopted), professional talent teams should be formed, and division of labor should be used to refine knowledge granularity and correlation relationships across domain resources.

3.2 Figure Resource Description Frameworks

Currently, no directly reusable figure knowledge framework or literature knowledge framework has emerged in the local celebrity resource research field. Understanding general or commonly used characteristic frameworks can provide clues for building Western Hunan local figure resource knowledge frameworks. Reference figure resource description frameworks include:

1. **CBDB Database.** The China Biographical Database focuses on historical figure biographies as core resources. Information description can be divided into official career paths, social identities, kinship associations,

and regional migration. Biographical information combined with era backgrounds and figure social relationships constructs a relatively complete figure knowledge organization scheme, providing support for local figure and related historical figure knowledge classification and attribute screening.

2. **FOAF Model.** As an ontology model for organizing and describing user information in online communities and social networks, its term vocabulary includes 13 commonly used basic classes and 55 attributes that can summarize elements such as individual users, social groups, organizations, individual relationships, and related figures, education, work, and achievements.

3. **CERIF Management Standard.** The basic entities, achievement entities, facility entities, and annotation entities involved in academic research information management systems can intuitively describe figures' full-cycle research participation and reveal their information in education, work, research, and honors.

4. **Shanghai Library Open Data Platform.** The platform provides ontology vocabularies for ancient books (37 classes, 160 attributes), genealogies (38 classes, 109 attributes), manuscripts and archives (44 classes, 195 attributes), historical figure biographies (9 classes, 35 attributes), and name authority libraries (22 classes, 68 attributes), providing guidance for knowledge element extraction from literature resources such as genealogies, manuscripts, archives, and local gazetteers in local celebrity resources.

---

4. Initial Construction of Local Celebrity Literature Resource Ontology Model

The core step in local celebrity resource knowledge organization is ontology model construction, which should revolve around celebrity literature integration and resource structure standardization, following principles of reasonable organization, orderly correlation, standard adaptation, open sharing, detailed content, and rich characteristics. Based on this, referencing common ontology vocabularies and knowledge frameworks, combined with Western Hunan figure and resource characteristics, we used the seven-step method to self-build the local celebrity literature ontology CLO (Celebrity & Literature Ontology), comprising figure and literature knowledge frameworks. Key steps are summarized below:

4.1 Knowledge Element Extraction

After organizing heterogeneous resource content, reference multiple ontology vocabularies and use system standards and ROST software to extract knowledge elements that can summarize and standardize literature content from

electronic texts, providing support for subsequent unstructured text processing and initial knowledge ontology construction. Using local literature such as *Fenghuang County Gazetteer*, *Western Hunan Historical and Cultural Materials*, and *Fenghuang: Those People, Those Events* as foundations, extract knowledge elements needed for celebrity knowledge ontology construction. Using initially organized celebrity literature such as *Kuxuezhai Diary* and *Kuxuezhai Poetry Manuscripts* as foundations, extract content elements needed for literature knowledge ontology construction, including catalog, events, style, emotion, and role.

During specific figure and literature ontology construction, core concepts and descriptive attributes must be repeatedly adjusted according to figure data completeness and discriminability, literature genre, and content characteristics.

Table 1 shows an example of literature knowledge element extraction.

4.2 Basic Class Vocabulary Construction

Compare, organize, and differentiate extracted knowledge elements, select elements with universality, recognizability, and augmentability for inclusion in term vocabularies, and supplement and adjust basic classes (Class) with incomplete or unclear definitions. Define descriptive classes for figure knowledge ontology including basic profiles (name, birthplace, ethnicity, etc.), education (schools attended, majors, enrollment time, etc.), achievements (literature, diaries, poetry), and work (institutions, positions, etc.). For literature knowledge ontology, define catalog (number of pieces, pages, fonts, etc.), events (location, involved objects, etc.), role (creator, collector, developer, etc.), genre (field, origin, etc.), style (toward fonts, literature, figures, etc.), and emotion (toward events, roles, literature, etc.). Supplement general classes such as time (start time, end time, etc.) and place (involved regions, terrain, etc.).

Table 2 shows the basic class term vocabulary summary.

4.3 Attribute Vocabulary Construction

Use a hybrid method (top-down and bottom-up) to clarify local celebrity resource basic class hierarchies and intra-class attributes. For basic class hierarchy division, build domain upper-lower class frameworks according to hierarchical distribution and vocabulary concepts, then generalize universal classes upward and refine descriptive classes downward through instance domains and contained knowledge to ensure applicability of the overall structure of the Western Hunan celebrity literature resource ontology model and comprehensiveness of revealed elements.

For attribute classification, judge based on descriptive class facts: individual-to-individual associations are object properties (op), which have attributes explaining value types and can reveal common features between classes or between classes and instances; individual data associations are data properties

(dp), which can optimize instance knowledge organization schemes and enrich knowledge revelation perspectives based on characteristics such as categorical, ordinal, discrete, and continuous features.

Table 3 shows attribute vocabulary construction.

4.4 Ontology Model Revision

4.4.1 Revision Rules and Overview Ontology revision is a crucial step for adjusting basic classes and adding/deleting attributes based on initial models, referencing core concept extensions and revealing instance content to improve model matching degree. During revision, core concept extension definition, instance knowledge granularity refinement, and ontology model-domain instance matching degree must be evaluated based on consultation and feedback from local literature domain experts.

Based on the collation status of Western Hunan local celebrities and diary genre works, we designed consultation questionnaires from four dimensions: relevance, importance, adjustability, and operability, following judgment criteria of personal experience and independent cognition (pass=1; fail=0). Experts evaluated each basic class and attribute across the four dimensions and provided suggestions. The revision consultation involved 15 experts, with 13 questionnaires returned (response rate 86.7%). Participants came from libraries, universities, and software companies: 7 males (53.85%) and 6 females (46.15%); 11 with bachelor' s degrees or higher (84.62%), 2 with other qualifications (15.38%); 5 with senior titles (38.46%), 8 with intermediate titles (61.54%). Average work experience in information organization, knowledge management, and database construction was 4+ years. Expert evaluations, issues, and suggestions were summarized in **Table 4** . After revising CLO according to suggestions, unanimous approval was obtained.

4.4.2 Basic Class Revision Since basic classes depend on text content, class adjustments mainly affect object properties describing figure profiles, education, achievements, and work. In figure description, based on historical background and personal data gaps, remove data properties such as posthumous title, party affiliation, specialty, language, major, terrain, and page numbers; add object properties such as identity, kinship, and instructor based on expanded materials. In literature description, remove data properties about domain and origin in genre classes based on diary genre characteristics; add object properties for event class including coordinates, involved countries, and involved locations; add event emotion property in emotion class; temporarily remove external properties such as collector and developer in role class; add volume number and record time in catalog according to diary format; add object properties such as communicating figures, political figures, and historical figures. Additionally, add exchange roles in role class to annotate letter correspondents due to frequent letter exchanges in diaries. Summarize work experience-related institution transfers into separate Institution (institution name, transfer location) and Position (position

name) basic classes to refine organization and description of figure work changes. In general classes, broaden place-related class description types based on literature content, adding a Country subclass (197 countries) for annotation and description of objects involved in international events.

In summary, the adjusted figure knowledge framework includes 10 basic classes (6 major classes, 4 subclasses) and 26 attributes (12 object properties, 14 data properties). The adjusted literature resource knowledge framework includes 6 basic classes (1 major class, 5 subclasses) and 21 attributes (15 object properties, 6 data properties). **Table 5** shows the basic classes and attribute descriptions (using diary as example), with CLO basic class prefixes omitted, superscript C for basic classes, op for object properties, dp for data properties, simplified instance relationship numbering, and *** representing any of Profile, Education, Work, or Achievement classes.

4.4.3 Attribute Adjustment After screening figure and literature ontology basic classes, related attributes need adjustment. Since data property input values mainly depend on text content, class adjustments primarily affect object properties describing figure profiles, education, achievements, and work. In figure description, remove data properties such as posthumous title, party affiliation, specialty, language, major, terrain, and page numbers based on historical background and personal data gaps; add object properties such as identity, kinship, and instructor based on expanded materials. In literature description, remove data properties about domain and origin in genre classes based on diary genre characteristics; add object properties for event class including coordinates, involved countries, and involved locations; add event emotion property in emotion class; temporarily remove external properties such as collector and developer in role class; add volume number and record time in catalog according to diary format; add object properties such as communicating figures, political figures, and historical figures. Additionally, add exchange roles in role class to annotate letter correspondents due to frequent letter exchanges in diaries. Summarize work experience-related institution transfers into separate Institution (institution name, transfer location) and Position (position name) basic classes to refine organization and description of figure work changes. In general classes, broaden place-related class description types based on literature content, adding a Country subclass (197 countries) for annotation and description of objects involved in international events.

5. Ontology Model Editing and Presentation Using Protégé

Protégé accelerates the transformation of domain knowledge from structured to formalized through knowledge extraction, fusion, and reorganization of local celebrity literature resources based on ontology models. Since ontology construction and revision require substantial time and repeated iteration, commonly used ontology languages and automated construction tools should be selected to

ensure and improve ontology editing efficiency. In terms of ontology languages, XML, OWL, and RDF(S) are recognized as standard languages for sharing semantic web content. In automation software, tools such as Ontolingua, OntoSaurus, WebOnto, and Protégé can accept multiple description languages for ontology model development. Among them, Protégé is an ontology automation editing and development tool developed by Stanford University' s Center for Biomedical Informatics Research, featuring secondary development, extended models, rich plugins, and multiple input/output formats, providing researchers with a customizable open-source environment [29]. Based on this, this local celebrity literature knowledge ontology editing task uses OWL language and Protégé tool support.

CLO basic classes and attribute input are shown in **Figure 2** [**Figure 2: see original paper**].

In the Protégé tool page, during the process of making the ontology "from virtual to real," first create Person (CLO_{Person}) and Diary (CLO_{Riji}) classes in Protégé's "classes"module according to the basic class description term table, then add subclasses such as Person Profile (CLO_{{{Person}}*{{Bas}}}) and Diary Catalog (CLO*{{{Riji}}_{{Lis}}}) through the "hierarchy" module according to hierarchical structure, and add GLO prefix in "prefix."After editing, the right-side ontology basic class hierarchical structure view shown in **Figure 3** [**Figure 3: see original paper**] can be obtained.

Additionally, based on the basic class relationships and attribute description terms, define each basic class, object property, and data property according to OWL/XML format encoding. For example, the left-side RDF/XML revelation box in **Figure 3** lists encoding definitions for Person Achievement (CLO_{{{Person}}*{{Ach}}}), Person Education (CLO*{{{Person}}*{{Edu}}}), and Person Profile (CLO*{{{Person}}_{{Bas}}}).

Simultaneously, input CLO object properties and data properties into corresponding automation operation bars according to **Figure 2**, and edit domains and ranges respectively. Input 32 object properties such as is_{about}, date_{{{of}}*{{birth}}}, place*{{{of}}*{{birth}}}, and date*{{{of}}*{{death}}} in the "*Object properties*" module. Add 16 data properties such as *Person*{courtesyname}, Person_{sex}, Person_{ethnicgroup}, and Person_{adschool} in the "Data properties" module. After defining classes, subclasses, object properties, and data properties, use the HermiT reasoner to correct erroneous definitions and improper associations to improve ontology hierarchical structure, and provide a basis for accurately extracting celebrity knowledge and diary metadata and semantic content. Finally, use Protégé' s "OntoGraf" module to display relationships between local celebrity literature ontology basic classes and attributes in "Radial" form, as shown in **Figure 4** [**Figure 4: see original paper**].

6. Instance Addition and Research Using Western Hunan Poet Tian Mingyu' s *Kuxuezhai Diary*

Instance addition and research are key steps in ontology revision and important methods for testing knowledge organization scheme applicability and developing local literature resources. Building a celebrity resource knowledge organization scheme driven by ontology can fully integrate heterogeneous literature resources, acquire domain fine-grained knowledge and draw figure correlation clues through attribute definition, logical reasoning, semantic querying, and humanities graphs, providing a 抓手 (grip) for shaping local cultural identity and deeply mining figure resource connotations.

The Xiangxi Tujia and Miao Autonomous Prefecture, located at the junction of Hunan, Hubei, Chongqing, and Guizhou, is a minority settlement with long historical culture and characteristic folk customs, costumes, dance, and diet, having produced cultural and scientific celebrities such as Shen Congwen, Xiong Xiling, Tian Mingyu, Huang Yongyu, and Peng Sixun [30]. Current research on Shen Congwen, Xiong Xiling, Tian Mingyu, and Huang Yongyu focuses on literature and art fields, emphasizing text content analysis and verification in resource organization and utilization. No scholars have integrated and developed related celebrity literature resources through digital humanities perspectives and technologies. As a genre that directly reflects figures' observations, thoughts, and feelings, diaries—with their rich thematic variety and casual expression—have attracted literati and intellectuals throughout history. Compared with other literature genres, diaries more easily express authors' personalities and real situations, with recorded experiences and events possessing more natural life 气息 (flavor), making them particularly valuable for literature verification, artistic appreciation, and ideological education.

Based on this, this paper presents an instance using Western Hunan poet Tian Mingyu and his manuscript *Kuxuezhai Diary* (1961-1962) as a case study. Through "OCR + human proofreading" digitization of related books and manuscripts, combined with database figure data and preset editorial conventions, we completed heterogeneous literature resource content organization. Using LTP semantic analysis platform and NLPIR language processing system, we extracted figure and literature elements, classifying entity elements according to figure profiles, education, work, achievements, institutions, identity classification, and literature catalog, events, emotions, style, role, and place classification. After element extraction and classification, we resolved and disambiguated non-standard expressions (nicknames, abbreviations, dialects, wrong characters), manually resolved conflicts between ancient figure names and courtesy names with identical meanings (e.g., Tao Yuanming/Tao Qian), and eliminated referential ambiguity for entities with identical names but different meanings (e.g., *Shijing* actually referring to *Shijing Shuolue*). We imported 489 extracted and organized instances into the CLO local celebrity literature ontology model according to preset knowledge frameworks, assigned object properties and data properties to instances based on figure basic classes

and diary contact classes, and used Protégé' s "OntoGraf" module to visualize local celebrity literature resource entity element distribution, as shown in **Figure 5** [**Figure 5: see original paper**].

6.1 Tian Mingyu' s Figure Knowledge Correlation Analysis

Due to display limitations in the "OntoGraf" module, we used SPARQL query language to retrieve "Tian Mingyu" and formed associations using the "Grid-Alphabetical" module, as shown in **Figure 6** [**Figure 6: see original paper**]. Figure knowledge from different sources can achieve knowledge aggregation and structured distribution through identical basic classes and similar attribute features in the self-built CLO ontology model. The CLO model can clearly and intuitively organize figures' profiles and experiences across dimensions and conduct knowledge correlation.

In terms of figure profiles, knowledge such as Tian Mingyu' s courtesy name (Geshi), birth year (1890), birthplace (Fenghuang County), death year (1981), and identities (Nanshe poet, Tujia poet, Western Hunan writer, Tongmenghui member) are co-presented. In terms of figure achievements, Tian Mingyu's works (*Kuxuezhai Diary*, *Kuxuezhai Poetry Manuscripts*, *Zaohong Ci*, *Forty Years of Western Hunan Major Events*, *Western Hunan Miao Records*, *Shijing Shuolue*, *Chu Travels Trivial Records*, *Miscellaneous Poetry Manuscripts*, and *Western Hunan Tribute Collection*) are integrated, providing frameworks for subsequent literature content organization and correlation. In terms of work experience, Tian Mingyu' s appointment locations and positions (Fenghuang Lecture Institute director, Wenchang Pavilion teacher, *Yuanxiang Daily* editor and general manager, Western Hunan National Protection Army secretary, Dayong County magistrate, Yuanling County magistrate, Qianyang County magistrate, Tenth Group Army secretary, Hunan Provincial Government secretary, Fenghuang County magistrate, Hunan Provincial Cultural Relics Protection Committee member, and State Council Literature and History Research Institute member) are also correlated, providing a basis for drawing figure career portraits. Additionally, through kinship and instructor attribute correlations, multiple relationships between Tian Mingyu and his uncle Tian Xingliu can be revealed, and combining correlations among appointment locations, figures, and times helps verify whether Tian Mingyu taught the famous writer Shen Congwen during his tenure at Wenchang Pavilion.

6.2 *Kuxuezhai Diary* Knowledge Correlation Analysis

Using SPARQL query language to retrieve "*Kuxuezhai Diary*" and forming associations through the "Grid-Alphabetical" module, the result is shown in **Figure 7** [**Figure 7: see original paper**]. The diary content recorded in *Kuxuezhai Diary* can be divided into six categories according to the CLO ontology model: catalog, events, roles, emotions, style, and places. Knowledge in each dimension achieves co-presentation and correlation under the guidance of basic classes and attribute features, enabling more intuitive presentation of diary core events, fig-

ure correlations, and hotspot regions, and providing samples for further 演绎 (deducing) figure behavioral details, revealing style evaluations, and reasoning emotional tendencies.

In the ontology model, relationship reasoning criteria such as Inverse functional, Transitive, Symmetric, Asymmetric, and Reflexive can be set for knowledge reasoning and evaluation. By setting existing basic classes and attributes as corresponding relationships, new associations can be inferred and erroneous associations corrected based on existing assignments, thereby improving knowledge correlation accuracy. After knowledge reasoning, correlations among roles, events, works, and emotions can be further achieved. After textual knowledge correlation, it can be observed that when Tian Mingyu mentioned relatives (Chun' er, Fusun, Zhensun, Zhaisun), high-frequency events (hospitalization, going to the countryside, delivering letters) mostly contained caring emotions (deep care, deep concern, attentive inquiries). When mentioning communicating figures (Shen Congwen and Zhang Zhaohe), high-frequency events (inscriptions, sending pictures, home visits) reflected sincere emotions (peach and plum spring breeze, earnest teachings), further confirming the teacher-student relationship between Tian Mingyu and Shen Congwen. When mentioning historical figures (Tao Yuanming, Ouyang Xiu, Li Bai), he often used poems and artistic conceptions from poets' works (*Returning to the Farm*, *Li Taibai Collection*) to express feelings about beautiful scenery and landscapes, and frequently used them to verify his own poems and express emotions (autumn sorrow, nostalgia, expressing feelings through landscapes). Additionally, political figures (Nehru, Kennedy, Khrushchev) and international events (Sino-Indian border self-defense counterattack, Apollo program) recorded in the diary also formed emotional correlations (fearless, battle-hardened, unyielding). It can be seen that knowledge reasoning through rule frameworks such as "role-event (work)-emotion" can effectively organize and correlate various diary elements, assist humanities scholars in multi-dimensional fine-grained knowledge 梳理 (organization) and discovery, and enable research on figure and event emotional tendencies through frequency conversion.

6.3 Knowledge Organization Process

Through the above operations, this study used Tian Mingyu and his work *Kuxuezhai Diary* as examples to complete knowledge organization and correlation research on local celebrity resources through four steps: heterogeneous literature resource content organization, local celebrity resource ontology construction, entity and relationship fusion, and local celebrity resource knowledge application. The process diagram is shown in **Figure 8** [**Figure 8: see original paper**]. Subsequent work can extract various elements according to the ontology, use visualization tools such as Gephi, QGIS, and Cytoscape for spatiotemporal network revelation, paragraph emotion classification, and interactive behavior analysis, and integrate diary content into graph databases such as Neo4j to facilitate the construction of local celebrity literature resource knowledge bases

and the development of open access services.

7. Conclusion

The purpose of conducting knowledge organization and correlation of local celebrity literature resources is to integrate heterogeneous literature resources, reveal figure and literature knowledge features, and further refine the accuracy and granularity of celebrity literature resource description. This paper proposes a knowledge organization and correlation scheme for developing and utilizing local celebrity resources based on four steps: heterogeneous literature resource content organization, local celebrity resource ontology construction, entity and relationship fusion, and local celebrity resource knowledge application. Combining local celebrity literature resource characteristics and figure resource description frameworks, we integrated figure and literature knowledge frameworks into the local celebrity literature ontology model CLO. In the case study, using Western Hunan poet Tian Mingyu' s materials and works, we achieved fine-grained knowledge retrieval and characteristic revelation for Tian Mingyu and his unpublished manuscript *Kuxuezhai Diary*, verifying the operability and practicality of the CLO ontology model in local celebrity literature resource organization and development. This study also explored the advantages of local literature resource research and development from a digital humanities perspective, summarized as follows:

1. **Mutual thinking and complementary advantages.** Compared with traditional philological research thinking and methods, tools, technologies, and algorithms involved in digital humanities can expand local literature research from "text" to "text + data," breaking barriers between liberal arts and sciences while combining linguistic description advantages with mathematical operation advantages to achieve integration of qualitative and quantitative thinking and enrich output forms and fields of local literature research results.

2. **Broad perspective and structural convergence.** Digital humanities provides an interdisciplinary application perspective for developing and utilizing local celebrity literature resources, promoting structural convergence of multi-source heterogeneous resources of local celebrities and knowledge interconnection of external heterogeneous resources. This improves resource openness, utilization, and sharing value while clarifying directions for local literature preservation institutions to further conduct knowledge clustering research, humanities knowledge graph drawing, and characteristic resource knowledge discovery.

This study has the following limitations: the knowledge organization and correlation scheme for local celebrity literature resources emphasizes engineering operations with a relatively broad perspective; the research object is limited to diary literature, requiring empirical research on other works. Future work will

refer to this scheme for local celebrity knowledge base construction, hoping to continuously adjust scheme structure, classes, and attributes in practical applications to provide references and foundations for developing and utilizing local celebrity literature resources.

References

[1] Liu Shi, Sun Maosong, Gu Qing. *Digital Humanities* [M]. Beijing: Zhonghua Book Company, 2020.
[2] Feng Qingjun. *Theory and Practice of Modern Library Local Literature Work* [M]. Beijing: Central Party Literature Press, 2008.
[3] Xie Man. Construction of figure thematic databases and data mining exploration—Taking the construction of Chinese female figure thematic database as an example [J]. *Modern Intelligence*, 2010, 30(6): 49-53.
[4] Liu C, Huang C, Wang H, et al. Mining local gazetteers of literary Chinese with CRF and pattern-based methods for biographical information in Chinese history[C]//Proceedings of 2015 IEEE International Conference on Big Data. New York: IEEE, 2015: 1629-1634.
[5] Zhao Yufei. Research on figure information aggregation based on Chinese name authority files [D]. Taiyuan: Shanxi University, 2020.
[6] Kim J W. The role of philosophy of religion in the research of digital humanities: Through the experience of building a digital archive of protestant missionary letters from Korea (1880-1942)[J]. *The Journal of the Humanities*, 2021, 123(1): 79-105.
[7] Zorrilla N. The exclusion of early modern women philosophers from the canon: Causes and counteractive strategies from the digital humanities[J]. *Hypatia—A Journal of Feminist Philosophy*, 2022, 37(2): 177-186.
[8] Lupanova Y M. M.V. Lomonosov' s image in the historic memory of modern youths[J]. *Sotsiologicheskii Zhurnal*, 2017, 23(3): 163-182.
[9] Kun H, Jianfeng Z. A progressive web application on ancient Roman Empire coins and relevant historical figures with graph database[C]//Lecture Notes in Computer Science (LNCS 11197). *Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection*, 7th International Conference. Berlin: Springer International Publishing, 2018: 235-241.
[10] Bonch-Osmolovskaya A, Skorinkin D, Pavlova I, et al. Tolstoy semanticized: Constructing a digital edition for knowledge discovery[J]. *Journal of Web Semantics*, 2019, 59(100483): 1-9.
[11] Chen C, Chang C. A Chinese ancient book digital humanities research platform to support digital humanities research[J]. *Electronic Library*, 2019, 37(2): 314-336.
[12] Huang J, Luo T. Computing Len for exploring the historical people' s social network[C]//Proceedings of 2018 IEEE 6th International Conference on Future Internet of Things and Cloud Workshops (W-FICLOUD 2018). New York: IEEE, 2018: 95-101.
[13] Ji H H, Min M S, Cheol C G, et al. Visual analysis on the political orientation of historical characters in the Joseon Dynasty: Focusing on Seo, Geojeong[J].

Archives of Design Research, 2019, 32(1): 147-161.
[14] Bae S. Digital humanities and Song Dynasty research focus on 1163' s civil service examination and the social network of LouYue[J]. *Journal of Asian Historical Studies*, 2019, 146(1): 157-191.
[15] Xu Yongming. Several visualization approaches for Chinese classical literature research—Taking Tang Xianzu research as an example[J]. *Journal of Zhejiang University (Humanities and Social Sciences Edition)*, 2018, 48(2): 164-174.
[16] Qian Chaofeng, Du Debin. Spatial structure and evolution of Northern Song bureaucratic family networks: An investigation based on CBDB and CHGIS[J]. *Historical Geography Research*, 2019, 39(2): 83-94, 161-162.
[17] Yang Haici, Wang Jun. Construction and visualization of Song Dynasty academic master-disciple knowledge graphs[J]. *Data Analysis and Knowledge Discovery*, 2019, 3(6): 109-116.
[18] Liu Jingchen. Research on Ming-Qing Jinshi families from a big data perspective—Taking CBDB and Chinese Roots Web as examples[J]. *Peking University Journal (Philosophy and Social Sciences Edition)*, 2019, 56(4): 96-108.
[19] Deng Jun, Zhong Chuyi, Wang Ruan, et al. Knowledge organization and correlation analysis of Qing Dynasty officials—Taking *Changchun County Gazetteer: Changchun Official Annotations* as an example[J]. *Library and Information Service*, 2020, 64(17): 18-26.
[20] Huo Xiaonan. Research on spatiotemporal structure and social relationship networks of exiled poets in *Complete Tang Poems* from a digital humanities perspective[D]. Changchun: Jilin University, 2021.
[21] Song Xueyan, Cui Haonan, Liang Ying, et al. Knowledge discovery research on celebrity diary resources from a digital humanities perspective—Taking Wang Shijie' s diaries as an example[J]. *Information Theory and Practice*, 2021, 44(6): 105-111.
[22] Liu Ningjing, Liu Yin, Wang Moyan, et al. Research on knowledge model construction of academic celebrity resources from a digital humanities perspective—Taking Li Zhengdao Digital Resource Center as an example[J]. *Library and Information Service*, 2019, 63(23): 113-121.
[23] Li H, Zhu L, Shen W, et al. Research on knowledge organization and visualization of historical events in the Republic of China Era[J]. *Library Trends*, 2020, 69(1): 23-45.
[24] Yao Tianhong, Chen Yanmei, Liu Ge, et al. Research on semantic knowledge organization of digital humanities historical materials based on CIDOC-CRM—Taking Zhang Xueliang historical materials as an example[J]. *Journal of Library Science*, 2019, 41(7): 35-43.
[25] Wei J, Liu R. An approach of constructing knowledge graph of the hundred schools of thought in ancient China[C]//Proceedings of 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL 2019). New York: IEEE, 2019: 335-336.
[26] Cheok A D, Edirisinghe C, Karunanayaka K. Confucius computer: A philosophical digital agent for intergenerational philosophical play[J]. *Personal and Ubiquitous Computing*, 2017, 21(2): 327-343.
[27] Niu Li, Gao Chenxiang, Liu Lichao, et al. Hierarchy and space: Research on

value mining of celebrity archives from a digital memory perspective[J]. *Archives Science Study*, 2021, 187(5): 138-144.
[28] Zhang Yunzhong, Sun Ping. Construction and visualization of historical and cultural celebrity study tour footprint knowledge graphs[J]. *Library Journal*, 2021, 40(9): 81-87, 96.
[29] Wang Haofen, Qi Guilin, Chen Huajun. *Knowledge Graphs: Methods, Practice and Application* [M]. Beijing: Publishing House of Electronics Industry, 2019.
[30] Chinese People' s Political Consultative Conference Xiangxi Tujia and Miao Autonomous Prefecture Committee, Historical and Cultural Materials Research Committee. *Xiangxi Celebrities* [M]// *Xiangxi Historical and Cultural Materials*, Volumes 34-35. Xiangxi: Xiangxi Prefecture Historical and Cultural Materials Committee, 2017.

Author Contributions:
Wang Ziyi: Proposed research topic, wrote and revised paper;
Chen Chen: Collected review literature and organized manuscript materials;
Wang Xianghua: Provided manuscript materials, guided research plan, checked paper content and proposed revisions.

Note: Figure translations are in progress. See original paper for figures.