Construction of Tang Poetry Knowledge Graph and Design of Intelligent Knowledge Services: Postprint

Authors: Zhou Lina, sonorous, Gao Ziyang

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] Grounded in the current demands for Tang poetry knowledge services within the big data environment, this study constructs a Tang poetry knowledge graph based on large-scale Tang poetry data and provides intelligent knowledge services, thereby promoting innovation in Tang poetry knowledge management and service approaches under the artificial intelligence environment. [Method/Process] Based on an investigation of domain knowledge service requirements, this paper designs a domain knowledge service-driven ontology model for Tang poetry. Subsequently, utilizing multi-source heterogeneous data crawled from the Web, it employs technologies such as knowledge extraction, knowledge fusion, and knowledge reasoning to automatically construct a Tang poetry knowledge graph, uniformly representing and organizing Tang poetry domain data to achieve semantic processing of large-scale Tang poetry data. [Result/Conclusion] This paper designs an intelligent knowledge service platform named KnowPoetry based on the Tang poetry knowledge graph, providing intelligent knowledge services including knowledge exploration, spatiotemporal trajectory analysis, and semantic querying in the Tang poetry domain, thereby promoting the innovative transformation of digital humanities research methods for Tang poetry under the artificial intelligence environment.

Full Text

Volume 63, Issue 2, January 2019

Construction of Tang Poetry Knowledge Graph and Design of Intelligent Knowledge Services

Zhou Lina[1], Hong Liang[1], Gao Ziyang[2]

[1]School of Information Management, Wuhan University, Wuhan 430072
[2]School of Computer Science, Wuhan University, Wuhan 430072

Abstract

[**Purpose/Significance**] Addressing the demand for Tang poetry knowledge services in the current big data environment, this study constructs a Tang poetry knowledge graph based on large-scale Tang poetry data and provides intelligent knowledge services, thereby promoting innovation in Tang poetry knowledge management and service methods in the artificial intelligence era.

[**Method/Process**] Based on an investigation of domain knowledge service requirements, this paper designs a domain knowledge service-driven Tang poetry ontology model. Using multi-source heterogeneous data crawled from the Web, it employs knowledge extraction, knowledge fusion, knowledge reasoning, and other technologies to automatically construct the Tang poetry knowledge graph, which uniformly represents and organizes Tang poetry domain data and achieves semantic processing of large-scale Tang poetry data.

[**Result/Conclusion**] This paper designs an intelligent knowledge service platform called KnowPoetry based on the Tang poetry knowledge graph, providing intelligent knowledge services such as knowledge exploration, spatiotemporal trajectory visualization, and semantic querying in the field of Tang poetry, thereby promoting the innovative transformation of Tang poetry digital humanities research in the AI environment.

1. Related Research

1.1 Advances in Digital Humanities Research on Tang Poetry

Digital humanities research related to Tang poetry is flourishing, covering not only the poetry itself but also related fields such as history, geography, and art. Its development has progressed through three main stages: from digital electronic text libraries to structured thematic databases, and then to semantic linked knowledge bases.

Early digital humanities projects focused primarily on database construction and digitization, producing electronic literature libraries such as the Complete Tang Poems Database and Chinese Poetry Database. For example, Harvard University's "Chinese Text Project" (CText) focuses on the digitization of Chinese philosophical literature, collecting electronic versions of many classic works from the Tang and Song dynasties as well as scanned materials of rare Chinese books. Consequently, electronic literature libraries of this period mainly consisted of unstructured electronic text data and image data, offering only browsing and simple keyword matching retrieval functions without achieving semantic knowledge services or intelligent retrieval. Tang poetry research still followed traditional methods of collecting, selecting, and comparing to construct logical

relationships among small amounts of poetic knowledge, making it difficult to break through the limitations of conventional qualitative research methods.

With the development of structured relational database technology, numerous structured thematic databases emerged in the digital humanities field. A typical representative is the "China Biographical Database Project" (CBDB) created through collaboration between Robert M. Hartwell of Harvard University and Peking University, which contains extensive biographical information on Tang poets and literary figures, with its structured format facilitating large-scale analysis by scholars. Additionally, Harvard University and Fudan University collaborated to create the "China Historical Geographic Information System" (CHGIS), which establishes continuous time series to describe changes in place names, administrative systems, and other basic geographical elements over time, providing retrieval and query functions for fundamental historical geographic information in China. In China, Wang Zhaopeng built the "Souyun" platform for chronological maps of Tang and Song literature, visualizing poets' spatiotemporal trajectory distributions and providing keyword-matching retrieval interfaces. Meanwhile, the establishment of thematic databases provided research platforms and tool support for scholar communities, leading to numerous research outcomes using digital tools to understand Chinese history and culture. For example, Hilde De Weerdt from Leiden University created the semi-automatic annotation platform MARKUS for ancient Chinese texts, which supports automatic tagging of tasks, locations, and user-defined concepts in classical Chinese texts. Yale University's "Ten Thousand Rooms Project" utilizes the community attributes of the internet to provide scholars with a collaborative research platform for public domain literature, such as creating annotated editions of Tang poetry collections. J.W. Chen conducted text mining research on Tang poetry and the Shishuo Xinyu, while P. Jason performed GIS/network analysis on Song dynasty Buddhist poets.

Currently, with the development of linked semantic technologies and artificial intelligence, semantic linked knowledge bases have emerged, such as the global Linked Open Data (LOD) and Wikipedia knowledge graph (DBpedia), which contain partially semantic data related to ancient Chinese literature and historical geography. Libraries, archives, and museums have strong data advantages in digital humanities research, and in recent years, many such institutions have developed a series of linked semantic knowledge bases using their collection resources. For instance, the Shanghai Library has established knowledge bases for Sheng Xuanhuai, genealogical knowledge bases, ancient text evidence-based platforms, and foundational knowledge bases (including historical figure authority files, historical geography knowledge bases, historical chronology knowledge bases, and historical event knowledge bases). When constructing the genealogical knowledge base, the Shanghai Library designed a genealogical ontology based on BIBFRAME using ontology modeling methods to interpret the documentary characteristics and content attributes of genealogical resources and enhance semantic associations between contents. In building the historical geography knowledge base, they adopted a knowledge ontology approach to design

a spatiotemporal model for historical geographic data to meet the open application needs of library digital humanities construction projects.

1.2 Research on Intelligent Knowledge Services for Tang Poetry

Currently, knowledge service applications in the Tang poetry domain represent a hot topic for researchers in both digital humanities and computer science. Applications of intelligent knowledge services for Tang poetry are mainly divided into three categories: poetry robots, intelligent question-answering systems, and poetry knowledge graphs.

In the field of intelligent poetry composition, foreign researchers proposed and implemented intelligent poetry software applications earlier, but these were primarily based on automatic poetry creation in foreign languages. For example, as early as 1959, Lutz in Germany created the first poetry-writing software, followed by systems like RACTER and PROSE. Typical foreign poetry software includes Desktop Poet, a tool supporting English poetry creation; Google's LED interactive device Poetrics, which uses voice retrieval technology and voice platforms to convert voice materials into poetry; and SwiftKey's automated poet "Fluent," which creates works in the style of Shakespeare's sonnets. With the continuous development of AI technology, numerous poetry creation systems have emerged in China, such as Tsinghua University's "Jiuge" automatic poetry-writing robot based on deep learning and meter rules, online poetry software using template splicing, IBM's "Oude," Baidu's "Write Poems for You," Microsoft's "Jueju," and "Microsoft Xiaoice." However, machine-generated poetry suffers from poor meaning and artistic conception, failing to meet the deep knowledge service needs of the Tang poetry domain.

With the deepening development of semantic association technology, general search engines have begun to provide intelligent question-answering functions for poetry knowledge by integrating and processing massive poetry information and directly providing answers based on natural language queries. For example, searching "raise head to gaze at the bright moon" on Baidu returns the answer "lower head and think of hometown." Additionally, specialized poetry intelligent Q&A systems like "Poetry Xiaoqiang" have emerged, providing semantic retrieval services based on associated knowledge graphs. However, these intelligent Q&A systems still rely on traditional keyword matching technology for complex natural language queries and sometimes struggle to return desired answers directly.

The emergence of knowledge graph technology provides new ideas for knowledge services in the Tang poetry domain, and the construction of domain knowledge graphs related to Tang and Song poetry culture has become mainstream in academic circles. For example, Beijing Normal University developed the "Tang Poetry Bieyuan" platform based on knowledge graphs for complete Tang poetry retrieval and visualization, achieving semantic retrieval and knowledge graph visualization but lacking theoretical model framework research on the in-

terpretability, associativity, and historicity of Tang poetry domain knowledge, making it difficult to support deep and complex Tang poetry knowledge association mining and reasoning research. Peking University's KVision Laboratory developed a knowledge graph of Song dynasty academic inheritance, extracting academic inheritance and partial kinship relationships among Song figures from CBDB, using knowledge graphs to display and query data and providing dynamic, visual historical knowledge exploration and discovery. However, the current graph is limited to academic inheritance and kinship relationships, with general academic and social relationships not yet ontologized and semanticized.

Reviewing domestic and international research, topics related to Tang poetry knowledge graphs and services have been addressed. However, numerous shortcomings exist in Tang poetry knowledge graph construction and domain intelligent knowledge service applications: there is no systematic representation and modeling of Tang poetry domain knowledge; no automated construction scheme for large-scale knowledge graphs from multi-source heterogeneous data; and no design scheme for intelligent Tang poetry knowledge service platforms implementing semantic retrieval, association analysis, and knowledge reasoning from quantitative, objective, and dynamic perspectives. Therefore, this paper aims to overcome these research challenges, using Tang poetry knowledge graph construction and domain knowledge services as a model to explore the transformation and innovation of digital humanities research methods in the AI era.

2. Domain Knowledge Service-Driven Tang Poetry Ontology Modeling

The semantic Web-based Tang poetry domain knowledge representation model is the foundation for constructing the Tang poetry knowledge graph, the prerequisite for machine-understandable Tang poetry semantic knowledge, and the basis for supporting the construction of intelligent Tang poetry service platforms. Existing Tang poetry semantic knowledge bases generally lack systematic representation and modeling of Tang poetry domain knowledge. Therefore, this paper designs a domain knowledge service-driven Tang poetry ontology by investigating domain knowledge service requirements, combining multi-source heterogeneous data characteristics crawled from the Web for the Tang poetry domain.

2.1 Tang Poetry Domain Knowledge Service Requirements

The construction of the Tang poetry knowledge graph must be based on scholar-oriented research needs and intelligent knowledge service needs in the current big data environment, guiding the design and construction of an intelligent Tang poetry domain knowledge service platform. Current Tang poetry research theories and methods are mature, with Tang poetry knowledge involving the intersection of three major fields: poetics, philology, and historiography. Different research fields have both correlations and distinctions in their knowledge graph

construction needs: (1) Poetics research focuses on meter, word usage, sentence structure, grammar, and parallelism in form, and emphasizes emotional imagery attributes in content. However, dictionary construction for Tang poetry texts remains blank, making automated establishment of poetic imagery-emotion associations a major challenge. (2) Philology research focuses on version origins, compilation, collation, and authenticity of ancient texts containing Tang poetry, but many relevant documents are lost. Different dynasties after the Tang varied in their compilation of Tang poetry, making it difficult to establish associations between poets' stylistic attributes and poetry's content attributes like imagery and emotion, and to judge authenticity for supporting textual research. (3) Historical research mainly explores Tang dynasty politics, economy, culture, and folk geography related to Tang poetry, with the most important issues being the mapping and conversion of historical geographic data, such as the evolution of ancient and modern place names and conversion between Gregorian and historical chronologies, to mine poets' life experiences and creative development trajectories.

Therefore, based on a thorough analysis of domain knowledge service requirements and addressing the above research hotspots and difficulties, combined with multi-source heterogeneous data characteristics crawled from the Web for the Tang poetry domain, we design a Tang poetry ontology oriented toward the three specific domain research needs of poetics, philology, and historiography, including poetry-poet modeling for poetics and philology applications and spatiotemporal experience modeling for historiography applications.

2.2 Poetry-Poet Ontology Model for Poetics and Philology Applications

As a form of traditional Chinese classical literature, Tang poetry texts have special grammar and conventions, containing content attributes such as imagery, artistic conception, emotion, allusions, and themes, with often non-explicit relationships between these attributes. Therefore, this paper proposes establishing a "Poetry" entity model to support imagery association analysis for poetics. Meanwhile, the relationship between poetry and poet is authorship, for which we establish a "Poet" entity model. Judging the origin and development of Tang poetry versions and authenticity information through poetry and poet data constitutes an important aspect of philology research. The RDF triple-based representation method is commonly used in knowledge graph construction, providing flexibility in knowledge modeling. We need to expand poetry-related entities, attributes, and entity relationships based on the RDF poetry entity model to support Tang poetry authenticity reasoning applications. While RDF provides modeling flexibility, its extensibility is limited, whereas RDF Schema (RDFS) provides a way to define terms and concepts and specify which properties can be applied to which objects on top of RDF. Therefore, this paper designs a poetry-poet ontology for poetics and philology applications based on the RDF/RDFS triple representation method, as shown in Figure 1 [Figure 1: see original paper].

In this binary ontology model, poetry serves as the central entity, with properties divided into Inc: content properties and Des: descriptive properties, among which there are hierarchical associations constrained by RDFS vocabulary. For example, the genre property of poetry "seven-character quatrain" is an rdfs:subClassOf the rdfs:Class "quatrain." Furthermore, various imagery elements constitute artistic conception, which expresses the emotional and thematic properties of poetry. Poetry-poetry relationships include imagery association, version association, and citation relationships. From the perspective of "poet" as the central entity, we define poet properties and poet-poet relationships, including citation, school, kinship, friendship, and official position relationships.

RDFS enables reasoning mechanisms after adding constraints to RDF, mainly manifested in poet citation relationships, school relationships, and poetry version relationships. Poet citation relationship extraction is primarily reflected through Range: "Poetry." First, we establish definite mutual citations through the rule of mutual mention, then define that as long as a poet is mentioned in a poem's title or text, a citation is added between them. Simultaneously, we create a corresponding citation entity or citation event connecting the two poets and the relevant poem. If a poem mentions the other party multiple times, it counts as only one citation. Through this rule, we ultimately obtain citation relationships between poets, such as "Liu Yuxi" rdfs:author "In Response to Letian at Yangzhou" and "Bai Juyi" rdfs:style name "Letian," thus deriving "Liu Yuxi" rdfs:rel "Bai Juyi" (citation relationship). Poetry version relationships are determined through the documentary sources that compile the poetry, while poet school relationships are established by defining and constraining the types to which poetry themes belong.

The Complete Tang Poems version attributes "Lanxi Zhao Ge" to Dai Shulun (732-789, styled Yougong, from Jiangsu), whose poetic style is fresh and fluent, mostly belonging to the pastoral landscape school. This version's "Lanxi Zhao Ge" has the thematic property of praising nature and expresses the emotional property of leisurely reclusive life. However, in the Ming Dynasty Three Hundred Poems version, "Lanxi Zhao Ge" is attributed to another poet, Wang Guangyang (?-1380, named Chaozong, from Jiangsu), without recording Dai Shulun's poem, thus creating a 疑点 (point of suspicion). Based on the constructed poetry-poet binary ontology model, we can set reasoning rules for version authentication: all works by the same poet should have consistent properties, especially relatively fixed poetic styles or schools, with poet school relationships established through defining and constraining poetry theme types. If another poet's work is mixed into a poet's collection, it can be judged by style. That is:

IF (Poet_{School}: "Pastoral Landscape School")
THEN (Poetry_{{{WroteByPoet}}_{{Theme}}}: "Pastoral Landscape")

Through the association between poetry's thematic and emotional properties and poet's school properties in the poetry-poet binary ontology model, we can

conclude that the Complete Tang Poems version of "Lanxi Zhao Ge" is correct. Meanwhile, extracting the poet's spatiotemporal experience information reveals that Wang Guangyang had been to Lanxi and wrote another "Lanxi Zhao Ge" with themes depicting fishermen's life and military life. Therefore, through the above rule examples, we can discover the authenticity of different poetry collection versions and reason about and judge erroneous attributions of poems existing under multiple poets' names, as shown in Figure 2 [Figure 2: see original paper].

2.3 Spatiotemporal Experience Ontology Model for Historical Studies

A poet's life experiences are not only important data for analyzing creative development but also reflect the changes in Tang dynasty society, holding significant meaning for studying the historical features of Tang society. Therefore, this paper treats poet experience as a third type of entity, dividing it into four normative classes—person, location, time, and event—to describe a complete experience and construct a poet's spatiotemporal experience ontology. First, location corresponds to spatial entities that exist or have existed in the real world, while time includes the correspondence and conversion between Chinese historical chronology and Gregorian calendar, thus building location and time models to achieve access, location, and association of data objects in the internet environment. Second, events have chronological characteristics, with relationships of inclusion, causality, continuity, and inheritance between event concepts. Moreover, the occurrence of events affects the content of poets' creative works. Therefore, we construct a poet experience spatiotemporal sequence model to reflect the temporal change characteristics of poet experiences, as shown in Figure 3 [Figure 3: see original paper].

Dictionary construction for Tang poetry texts remains blank, so we establish an information entropy-based segmentation model to solve the word segmentation problem in Tang poetry texts and achieve automatic extraction of Tang poetry texts, such as experience extraction, imagery-emotion extraction, and citation relationship extraction.

3. Construction of Tang Poetry Knowledge Graph

Based on the investigation and analysis of domain requirements and according to the domain knowledge service-driven Tang poetry ontology model in Section 2, we construct the Tang poetry knowledge graph as shown in Figure 4 [Figure 4: see original paper]. The main process involves using acquired multi-source Tang poetry data, employing knowledge extraction and fusion technologies to build the Tang poetry knowledge graph, storing it in a graph database in RDF triple form, and building an intelligent Tang poetry knowledge service platform on this foundation.

3.1 Data Acquisition

We crawl multi-source heterogeneous data for the Tang poetry domain from the Web, including encyclopedia websites, Chinese poetry websites, name and place dictionaries, and spatiotemporal coordinate data, then perform data conversion and processing.

3.2 Knowledge Extraction

Poets, poetry, and experiences as entities each have their own properties while maintaining various relationships among them, which are primarily supplemented and enriched through knowledge extraction. Therefore, Tang poetry text knowledge extraction mainly includes entity property extraction and relationship extraction. However, current dictionary construction for Tang poetry texts remains blank, so we establish an information entropy-based segmentation model to solve the word segmentation problem and achieve automatic extraction of Tang poetry texts.

3.3 Knowledge Fusion

The fusion process of the Tang poetry knowledge graph employs natural language processing, machine learning, and other AI technologies to distinguish and disambiguate obtained multi-source data, establishing different entity sets. Based on extracted relationship data, we link various entities in the main database and embed different information types into the database according to different patterns (properties, virtual entities, predicates).

3.4 Graph Management

After performing Tang poetry knowledge fusion through entity disambiguation and entity linking methods, we store the extracted knowledge base in RDF triple form and manage it using the graph database system gStore.

4. Intelligent Knowledge Services Based on the Knowledge Graph

Based on the Tang poetry knowledge graph constructed in Section 3, this paper builds the intelligent knowledge service platform KnowPoetry (accessible at http://kg.whu.edu.cn), providing panoramic knowledge exploration, spatiotemporal trajectory visualization, and semantic retrieval of internal associations among poetry, poets, and experiences. The platform supports Tang poetry research in three major fields: poetics, philology, and historiography, including historical research based on poet spatiotemporal trajectory visualization, poetics research based on imagery association analysis, and philological research based on poetry version authentication.

4.1 Knowledge Exploration

Knowledge exploration based on the Tang poetry knowledge graph provides users with more precise, faster, and more intelligent knowledge content resources such as poets, poetry, and experiences. It supports associated knowledge acquisition and analysis services, including poet citation relationship exploration and multi-dimensional association analysis of poetry's emotion, imagery, and allusions. By providing a visual interface for user browsing and retrieval, it displays multi-dimensional knowledge graphs of poets and poetry in forms such as ring charts and force-directed graphs. Users can directly obtain Tang poetry knowledge from different dimensions and perform query interactions based on certain themes, thereby achieving knowledge navigation functions.

4.1.1 Poet Homepage By searching and locating a specific poet, users can enter the poet's homepage, which details the poet's life experiences and creative works. The homepage displays the poet's work graph in force-directed graph form and provides specific node information. For example, through Li Bai's work graph, we can understand his main representative works and link to the work homepage, as shown in Figure 5 [Figure 5: see original paper].

4.1.2 Poet Citation Relationship Graph By defining citation relationships between poets, this platform constructs a Tang dynasty poet social network, supporting quantitative statistics on mutual citation frequencies between two poets, which helps understand their closeness and measure individual poet influence. Taking poets as nodes and citation relationships as links forms an associated poet citation relationship graph. Users clicking on corresponding nodes can obtain detailed poet information and link to their images or personal homepages. Figure 6 [Figure 6: see original paper] shows poets with citation relationships to Li Bai, including Du Fu, Gao Shi, Zheng Gu, etc. Selecting Li Bai's node reveals detailed information.

4.2 Spatiotemporal Trajectory

A poet's geographical location historically may change over time due to experiences such as promotion, demotion, military campaigns, reclusion, travel, and study. Therefore, we extract person, location, time, and event attributes from corresponding poet experience data and dynamically map their trajectories onto maps. If certain experience data is incomplete, we can infer the trajectory based on spatiotemporal constraints and the poet's creative experiences.

Simply displaying all poet trajectories on a map may result in unclear trajectory directions and ambiguous start/end points. Therefore, we use different colors to identify different poets' trajectories and provide segmented cursor dynamic demonstrations of all trajectories in interactive map form to help users find poets active in specific periods and geographical areas and their specific experience descriptions. Figure 7 [Figure 7: see original paper] intuitively reveals a segment of Li Bai's spatiotemporal experience from Chengdu to Yangzhou.

4.3 Semantic Querying

Semantic querying can automatically provide answers to users' natural language questions through efficient subgraph matching algorithms. Using the RDF database gStore for query operations, SPARQL descriptions are transformed into subgraph matching using efficient algorithms for pruning and matching to return query results. For example, when processing the query "poems with Bai Juyi in the title," the SPARQL language is:

```
sparql = '''
Select ?title ?name ?content
Where {
?poet <http://www.freekkg.com/poet/name> ?name.
?poem <http://www.freekkg.com/poet/author> ?poet.
?poem <http://www.freekkg.com/poet/title> ?title.
Filter regex(?tags, \"^%s$\")
```

Knowledge retrieval is not only faster than traditional relational data queries but also eliminates the cumbersome operation of using SPARQL query language to call multiple databases, shortening system response time. For example, retrieving "Bai Juyi's poems about plum blossoms," the SPARQL statement is:

```
sparql = '''
Select ?title ?name ?content
Where {
?poet <http://www.freekkg.com/poet/name> ?name.
Filter regex(?name, "%s")
?poem <http://www.freekkg.com/poet/author> ?poet.
?poem <http://www.freekkg.com/poet/title> ?title.
?poem <http://www.freekkg.com/poet/tags> ?tags.
Filter regex(?tags, \"^%s$\")
?poem <http://www.freekkg.com/poet/content> ?content.}
'''% ("白居易", "描写梅花")
```

The system quickly returns results directly, and users can click buttons to enter poetry homepages. The three returned results are shown in Figure 8 [Figure 8: see original paper].

4.4 Platform Service Effect Analysis

By constructing the Tang poetry knowledge graph, we can respond to knowledge service demands in the Tang poetry domain under the big data environment, build an intelligent knowledge service platform based on the knowledge graph, achieve three unifications of data-driven Tang poetry domain research, and form an "intelligent ecological cycle system" for knowledge creation, as shown in Figure 9 [Figure 9: see original paper]. Based on this cycle system, we mainly analyze from three aspects: research methods, research objects, and research purposes.

(1) Research Methods: Unification of Qualitative and Quantitative Approaches. Combining traditional qualitative research foundations, constructing knowledge graphs enables quantitative revelation of knowledge about poets, poetry, Tang dynasty historical chronology data, and spatial geographic coordinate data in Tang poetry, answering research questions based on statistical analysis of Tang poetry knowledge and thereby breaking through methodological limitations in Tang poetry research. For example, to answer "What is Li Bai's favorite allusion?" we can extract and quantitatively 统计 (statistically analyze) allusions in poetic texts, finding that Li Bai's works cite allusions 779 times, with "Wushan Cloud" being the most frequent at 11 times.

(2) Research Objects: Unification of Fine-Grained and Global Perspectives. By establishing a global Tang poetry knowledge ontology model, we can organically fuse multi-source heterogeneous Tang poetry data into a relatively complete and comprehensive knowledge framework. Using semantic knowledge organization forms facilitates machine reading and processing, enabling global analysis of complex relationships based on poetry, poets, and poet experiences, thereby broadening research objects in the Tang poetry domain.

(3) Research Purposes: Unification of Humanities and Intelligence. Knowledge graphs use natural language processing technology to convert complex natural language questions into relationships between entities, providing intelligent query functions that offer knowledge services for digital humanities research in the Tang poetry domain, thereby achieving the purpose of innovating Tang poetry digital humanities research paradigms.

References

[1] DONG H. Enrichment of cross-lingual information on Chinese genealogical linked data [EB/OL]. [2018-12-02]. https://www.ideals.illinois.edu/handle/2142/98870.

[2] Xu Zenglin, Sheng Yongpan, He Lirong, et al. A survey of knowledge graph technology [J]. Journal of University of Electronic Science and Technology of China, 2016, 45(4): 589-606.

[3] Chinese Text Project [EB/OL]. [2018-10-15]. https://ctext.org/zh/.

[4] China Biographical Database Project [EB/OL]. [2018-10-15]. https://projects.iq.harvard.edu/chinesecbdb.

[5] China Historical GIS [EB/OL]. [2018-10-15]. https://sites.fas.harvard.edu/~chgis/.

[6] Souyun: Poetry Portal [EB/OL]. [2018-10-15]. https://sou-yun.com/.

[7] MARKUS: Semi-automatic Marking Platform for Ancient Chinese Texts [EB/OL]. [2018-10-15]. http://www.academia.edu/11078612/MARKUS_ 中文古籍文本半自动标记平台.

[8] Li Youren, Song Yingchun. Analysis of digital humanities China studies in North America and Western Europe [J]. Shandong Social Sciences, 2018(7): 54-58, 63.

[9] CHEN J W, BOROVSKY Z, KAWANO Y, et al. The Shishuo xinyu as data visualization [J]. Early Medieval China, 2014(20): 23-59.

[10] JASON P. Toward a Spatial History of Chan [J]. Review of religion and Chinese society, 2016(3): 164-188.

[11] Zeng Lei, Wang Xiaoguang, Fan Wei. Smart data in the library, archive, and museum domain and its role in digital humanities research [J]. Journal of Library Science in China, 2018, 44(1): 17-34.

[12] Xia Cuijuan, Liu Wei, Zhang Lei, et al. Genealogy ontology design based on BIBFRAME [J]. Library Tribune, 2014, 34(11): 5-19.

[13] Xia Cuijuan. Research on open application of Chinese historical geographic data in library digital humanities projects [J]. Journal of Library Science in China, 2017, 43(2): 40-53.

[14] Desktop Poet [EB/OL]. [2018-10-15]. https://www.file-extensions.org/desktop-poet-file-extensions.

[15] Tsinghua AI Jiuge Computer Poetry Creation System [EB/OL]. [2018-10-15]. https://jiuge.thunlp.cn/.

[16] Orion Online Poetry Writing [EB/OL]. [2018-10-15]. http://www.dopoem.com.

[17] Microsoft Asia Research regulated verse and quatrain [EB/OL]. [2018-10-15]. http://duilian.msra.cn/jueju/.

[18] Microsoft Sixth-Generation Xiaoice [EB/OL]. [2018-10-15]. http://www.msxiaobing.com.

[19] Tang Poetry Bieyuan [EB/OL]. [2018-10-15]. http://poem.studentsystem.org.

[20] KVision Song Dynasty Academic Semantic Network [EB/OL]. [2018-10-15]. http://dh.kvlab.org/cbdb_{kg}/.

[21] Xia Cuijuan. Research on open application of Chinese historical geographic data in library digital humanities projects [J]. Journal of Library Science in China, 2017, 43(2): 40-53.

Author Contributions

Zhou Lina: Responsible for writing the main content of the paper, including the introduction, related research, knowledge modeling, and Tang poetry intelligent knowledge service platform sections.

Hong Liang: Guided the overall framework and ideas of the paper, macroscopically grasping the progress and 思路 (thinking) of the Tang poetry knowledge graph project.

Gao Ziyang: Responsible for writing the Tang poetry knowledge graph construction section of the paper and for data management and platform implementation in the Tang poetry knowledge graph construction project.

Construction of Knowledge Graph of Chinese Tang Poetry and Design of Intelligent Knowledge Services

Zhou Lina[1], Hong Liang[1], Gao Ziyang[2]

[1]School of Information Management, Wuhan University, Wuhan 430072
[2]College of Computer Science, Wuhan University, Wuhan 430072

Abstract: [Purpose/significance] Based on the demands of Tang poetry knowledge service under the current big data environment, the knowledge graph of Tang poetry is constructed and intelligent knowledge service is provided on the basis of large-scale data of Tang poetry, which promotes the innovation of knowledge management and knowledge service mode of Tang poetry under the artificial intelligence environment. [Method/process] Based on the investigation of domain knowledge service requirements, this paper designs the Tang poetry ontology model driven by domain knowledge service, then uses multi-source heterogeneous data crawled from the Web, adopts knowledge extraction, knowledge fusion, knowledge reasoning and other technologies to automatically construct the Tang poetry knowledge graph, uniformly represents and organizes Tang poetry domain data, and realizes semantic processing of large-scale Tang poetry data. [Result/conclusion] This paper designs an intelligent knowledge service platform KnowPoetry based on the Tang poetry knowledge graph, providing intelligent knowledge services such as knowledge exploration, spatio-temporal trajectory, and semantic query in the field of Tang poetry, and promotes the innovative transformation of Tang poetry digital humanities research methods in the artificial intelligence environment.

Keywords: knowledge graph of Tang poetry; intelligent knowledge service; digital humanities; knowledge model

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