Journal of Data Mining and Digital Humanities

EPI sciences overlay journals

Construction of Chinese Traditional Cultural Nouns Knowledge Base

Ziyang Lin[1], Xiaoru Wu[1], Minxuan Feng[1]*, Yi He[1]

1 Center of Linguistic Big Data and Computational Humanities, School of Literature, Nanjing Normal University, Nanjing, Jiangsu, China

*Corresponding author: fennel_2006@163.com

Abstract

The fine traditional culture of the Chinese nation is often embedded in cultural nouns. Building a knowledge base of them helps to better transmit these outstanding cultural heritages and provides support for compulsory education. This paper clarifies the definition of Chinese traditional cultural nouns and selects 3,512 representative nouns of them. Hownet is utilized to categorize these nouns and construct classified thesaurus. Furthermore, using CIDOC-CRM, the interpretations of these nouns are processed, and the associations between different categories of nouns are presented in a knowledge map. On this basis, a knowledge base of Chinese traditional cultural nouns has been constructed. Quantitative analysis of Chinese traditional cultural nouns reveals that the number and difficulty of Chinese traditional cultural nouns to be mastered by students show an increasing trend with the growth of the school year, which is basically in line with the cognitive characteristics of adolescents. This means that the knowledge base constructed based on the digital humanities approach contributes to cultural education at the compulsory level.

keywords

traditional culture, cultural nouns, classified thesaurus, knowledge base, compulsory education

I INTRODUCTION

Fine traditional Chinese culture, embodying the deepest spiritual pursuits of the Chinese nation, boasts a long and unbroken history. It not only stands as a testament to the Chinese nation's unique spiritual identity but also serves as a rich source of nourishment for its development. The appreciation of fine traditional Chinese culture holds profound significance for individuals, society, and the nation.

Within the linguistic system, the vocabulary subsystem is particularly vibrant and sensitive, often directly mirroring societal culture. Notably, nouns predominantly carry cultural content. Hence, defining and organizing Chinese traditional cultural nouns offers an intuitive way to understand fine traditional Chinese culture.

In recent years, the field of digital humanities has seen significant growth, with data visualization and structuring emerging as prevailing trends. Employing innovative methods to organize and present traditional culture, along with constructing a corresponding traditional cultural nouns knowledge base, is crucial for its promotion and preservation.

Our efforts to define traditional cultural terms, organize Chinese traditional cultural nouns in tree and graph structures, and develop a system of semantic categories for Chinese traditional cultural nouns, culminate in the creation of a traditional cultural noun knowledge base. This initiative not only structures traditional cultural knowledge but also lays a foundational resource for further research endeavors, such as information extraction and semantic analysis.

II DEFINITION AND SELECTION OF TRADITIONAL CULTURAL NOUNS

2.1 Review of Cultural Words

In foreign countries, scholars who study translation paid attention to "cultural words" earlier. In order to better translate words with special meanings in a certain nation, [Nida, 1988] defined these words as "terms". It is only in [Newmark, 1988] that the term "cultural words" is really introduced.

In China, scholars define cultural words from different perspectives. For example, [Chang, 1995] clearly points out that cultural words are different from general words in that they carry national cultural information, covering material, institutional, psychological and other aspects of culture. [Chen, 2006] and [Li, 2007] explain cultural words from the perspectives of teaching Chinese as a foreign language and translation, respectively. Among them, [Chen,2006] explicitly defines cultural words as "culture-specific words".

In this study, the construction of the knowledge base of Chinese traditional cultural nouns mainly serves the cultural heritage and compulsory education, so it is necessary to define and select traditional cultural nouns according to the purpose of the study.

2.2 Definition of Cultural Words

To streamline and ensure clarity in the promotion of compulsory education and the popularization of traditional culture, traditional cultural words are defined as words that arose between the epoch of the Three Sovereigns and Five Emperors and to the New Culture Movement of 1919, and the objective reality represented by the concepts are words that denote things, phenomena, behaviors, or ideas of Chinese culture.

2.2.1 Conceptual Meaning

Primarily, our reference sources include dictionaries and Wikipedia entries. Terms whose definitions are intrinsically linked to traditional culture or embody significant traditional cultural elements peculiar to the Chinese nation qualify as traditional cultural words. For example, "quadrangle" is delineated in The Modern Chinese Dictionary as a type of residential architecture featuring houses on all four sides enclosing a central courtyard. Wikipedia further elucidates that its architectural design and spatial arrangement are emblematic of traditional Chinese hierarchical concepts, alongside the principles of Yin Yang and the Five Elements. Such detailed explanations earmark it as a traditional cultural word based on its conceptual meaning.

2.2.2 Time of Creation

Given the varied research lenses and focal points, traditional Chinese culture can be dissected in numerous ways. Considering the overall traits of traditional Chinese culture and its evolutionary trajectory, only terms originating within the timeframe from the epoch of the Three Emperors and Five Emperors to the May Fourth Movement in 1919 are designated as traditional cultural words.

2.3 Selection of Cultural Words

The foundational data for this research is sourced from *the National Textbook for Primary and Junior School Education: Fine Traditional Chinese Culture* (hereinafter referred to as the "Textbook"), published by the People's Education Press, and articles from the *China Culture Daily* spanning from 1 July 2011 to 28 July 2021. Utilizing this data, a corpus of approximately 600,000 words was compiled.

To evaluate the effectiveness of tools for segmentation and part of speech tagging, a subset of 1,000 words from this corpus was selected for testing. The testing outcomes revealed that the F1-score for segmentation by Jieba, at 98.5%, surpasses that of HanLP, which stood at 96.3%. This comparison underscores the superior performance of Jieba for the tasks at hand, thereby justifying its selection as the preferred tool for further analysis.

Following the segmentation and part-of-speech tagging process, a total of 3,512 nouns were identified that align with the previously established definition of traditional cultural words.

Words	Origins
《史记》 The Records of the Grand Historian	Textbook
柳宗元 Zongyuan Liu	Textbook
神仙 Immortal	China Culture Daily
屈原 Yuan Qu	Both

Table 1. Examples of selected Chinese traditional cultural nouns

III CLASSIFIED THESAURUS FOR CHINESE TRADITIONAL CULTURAL NOUNS

The content of traditional culture is so diverse that even if we were to reflect the main content of it through Chinese traditional cultural nouns, there is a lack of clarity. Using Hownet, it is possible to categorize Chinese traditional cultural nouns. Based on this, a thesaurus of these nouns can be built.

3.1 Review of Classified Thesaurus

[Kilgarriff and Yallop,2000] indicated that a thesaurus is "a resource in which words with similar meanings are grouped together". Early thesauruses were constructed to serve information retrieval, and the words in the thesaurus tended to be arranged in alphabetical order, with little emphasis on subject overview or systematic display. However, [Aitchison, 1970] claims that these helped to reveal the intrinsic associations of terms better. Since the 1950s, scholars have tried to apply the classification scheme to the construction of thesauruses. [Aitchison, 1970] developed THE THESAUROFACET, which was compiled by facet analysis, covering scientific, technological, managerial, and other subject. With classification schedules, THE THESAUROFACET promoted the research of the classified thesaurus. There are also many representative thesauruses like UNESCO Thesaurus, Medical Subject Headings (or MESH), Thesaurus of Scientific and Technical Terms, and Information Retrieval Thesaurus of Education Terms, IRTE).

In China, Aeronautical Scientific and Technical Terms Thesaurus was the first classified thesaurus, which conducts in-depth research on the terms of aeronautical science and technology. It is a significant exploration of the development of classified thesaurus. [Bu and Wang, 2005] build the Chinese Classified Thesaurus, which is one of the most representative classified thesauruses, drawing on the Chinese Library Book Classification, combining classified citation and subject citation. With the help of computer technology, it also makes progress in the processing of conceptual relationships. Since the 1990s, the research of classified thesaurus in China has been gradually transformed from the construction of thesaurus to the ontology, semantic network, and knowledge graphs.

Generally speaking, there are few classified thesauruses studies at home and abroad that are specifically oriented to the culture. Creating a thesaurus for Chinese traditional cultural nouns will allow for better categorization of this knowledge.

3.2 Method

In constructing the classified thesaurus for traditional culture, it is pivotal to establish a structured system of semantic categories reflective of traditional cultural elements. This endeavor necessitates a detailed examination and categorization of vocabulary not only by thematic relevance but also considering the intricate relationships between terms, particularly in terms of hierarchy. Such an approach ensures that the thesaurus not only serves as a repository of terms but also as a navigable map of the semantic terrain of traditional Chinese culture, highlighting the connections and distinctions among the concepts it comprises.

Using Hownet, presented by [Dong, Z and Dong, Q, 2003], the Chinese traditional cultural words can be easily categorized. Hownet is a general knowledge base that reveals the relationships between concepts and the properties possessed by concepts. As the inaugural computable semantic knowledge base, HowNet offers a wealth of semantic knowledge invaluable for applications ranging from lexicography to natural language processing. It accomplishes the depiction of word concepts through the delineation of sememes, which are defined within HowNet as the most fundamental, indivisible units of meaning. Each sememe, of which there are 2,089 identified in HowNet, is categorized into one of seven groups, such as Entity and Event, providing a systematic and unambiguous representation of meanings. The hierarchical organization of sememes within HowNet facilitates a nuanced and precise method for constructing semantic categories for traditional culture, enabling the classified thesaurus to accurately reflect the depth and complexity of traditional Chinese cultural lexicon.

Leveraging HowNet's robust sememe system, the classified thesaurus can thus be developed with a clear and logical structure that not only categorizes traditional cultural words thematically but also aligns them according to their semantic relationships. This ensures that the thesaurus is both a comprehensive catalog of traditional cultural vocabulary and a dynamic tool that elucidates the semantic web underpinning traditional Chinese culture, offering scholars, educators, and students alike a rich resource for exploring and understanding the vast expanse of China's cultural heritage.

Figure 1. Hierarchical structure of sememe in Hownet

From Figure 1, it can be seen that the hierarchical structure of sememe in HowNet is not simply a single-level hierarchical structure, but rather a multi-level relationship. Taking the sememes

in Entity as examples, "beast|走兽" and "bird|禽" are both at the seventh level of the hierarchical structure of sememes in Entity and are both subordinate to "AnimalHuman|动物".

However, this system is not completely applicable to Chinese traditional cultural nouns. Therefore, this article adds some subordinate sememes. For example, under the sememe "human|人", four labels are added: "true person | 真人", "identity | 身份", "literary character | 文学人物", and "official position | 官职". Under the sememe "facilities|设施", seven labels are added: "cultural landscape | 文化景观", "defense facilities | 防卫措施", "religious facilities | 宗教设施", "daily buildings | 日常建筑", "transportation facilities | 交通设施", "burial facilities | 墓葬设施", and "technical facilities | 科技设施".

By adding, deleting, and revising the sememe in the system of HowNet, we successfully built a classified thesaurus of 3,512 Chinese traditional cultural nouns.

Categorization(OpenHownet)	Words	Origins
human|人	柳宗元 Zongyuan Liu	Textbook
human|人	屈原 Yuan Qu	Both
humanized|拟人	神仙 Immortal	China Culture Daily
publication|书刊	《史记》 The Records of the Grand Historian	Textbook

Table 2. Classfied thesaurus of Chinese traditional cultural nouns using Hownet

IV KNOWLEDGE BASE FOR CHINESE TRADITIONAL CULTURAL NOUNS

4.1 Related Work

With the advancement of natural language processing technologies, the traditional classified thesaurus has increasingly been regarded as insufficiently granular. [Gracia et al. 2017] advocate for moving beyond the hierarchical constraints of traditional tree structures by employing graph-based representations in dictionary construction, allowing for a more accurate depiction of the many-to-many semantic relationships among lexical items. Similarly, [Speer et al. 2017] adopt a graph structure to represent the relationships between various concepts in the development of ConceptNet. If the relationships among Chinese traditional cultural nouns could likewise be represented through a graph-based approach, the underlying structure and interconnections of traditional culture would become more transparent and comprehensible.

[Yuan et al, 2006] indicated that knowledge base, also called knowledge warehouse, is the product of combining traditional database and artificial intelligence technology. From [Hamid et al. 2002], while traditional databases emphasize data storage, knowledge bases focus more on the organization, management, and representation of data. A database presented in the form of a graph structure can be referred to as a knowledge base. Therefore, this section will focus on the construction of a knowledge base for Chinese traditional cultural nouns.

4.2 Method

The method of constructing a knowledge base based on ontology is more common. [Wei and Liu, 2020] has tried to apply it to the cultural field. The construction of Chinese traditional cultural nouns knowledge base is also mainly based on ontology.

[Yan and Chu, 2022] claims that ontology is the study of how human knowledge system is represented and organized. It is usually a triple containing knowledge, concepts, and relationships. The classified thesaurus has already categorized Chinese traditional cultural nouns. Therefore, the focus of the following sections lies primarily in how to represent the relationships among these cultural nouns.

4.2.1 Annotation in Definition

To indicate the relationship between different kinds of cultural nouns, it is first necessary to look for an association between them. This association can be found in the definition of the noun. A new column "definition" has been added to the classified thesaurus. The main references for the definition of words are the Modern Chinese Dictionary and Wikipedia.

Categorization (OpenHownet)	**Words**	**Origins**	**Definiton**
human|人	柳宗元 Zongyuan Liu	Textbook	柳宗元(773-819)唐代文学家、哲学家。字子厚，河东解县(今山西运城西南)人，世称柳河东。……有散文《捕蛇者说》、《封建论》、《三戒》、《永州八论》以及诗歌《江雪》、《渔翁》等名篇。哲学著作《天说》、《天对》等，具有朴素的唯物主义思想。有《河东先生集》。
human|人	屈原 Yuan Qu	Both	屈原(约前 340-约前 278)。战国楚诗人。芈姓，屈氏，名平，字原；又自称名正则，字灵均。屈原受楚怀王信任，且有楚国第一诗人的美称，初任左徒、三闾大夫，主张推行“美政”，改革政治。后遭旧贵族谗言攻击，被迫去官。楚顷襄王时，被放逐沅湘流域，终因理想无从实现，遂投汨罗江自杀。后世端午节吃粽子即是为纪念屈原。作有《离骚》、《九章》、《天问》、《九歌》等篇，开创了“楚辞”这一诗歌样式，对后世文学具有极大影响。

Table 3. Classified thesaurus with new column

In the definition, the categories of certain content are labeled. Take the nouns in {human | 人} as an example:

Words	**Labels**	**Annotation of Definition**
王献之 Xianzhi Wang	/生卒/(birth & death)	王献之**/称谓/**(344-386)，**/生卒/**东晋**/朝代/**书法家。**/身份/**字子敬，**/称谓/**琅邪临沂(今属山东)人。**/出生地/**各体皆精，行草尤佳，号“破体”，富豪迈之气，对后世影响很大。**/评价/**有墨迹《鸭头丸帖》等存世。**/成果/**逸事多见于《世说新语》与《晋书·王羲之传》中。**/参见文献/**
	/朝代/(dynasty)	
	/出生地/(birthplace)	
	/身份/(identity)	
	/地位/(status)	
	/称谓/(alternative names)	
	/事件/(events)	
	/评价/(appraisals)	
	/成果/(achievements)	
	/参见文献/(references)	

Table 4. Example of annotation in definition

In addition to labeling cultural nouns (e.g., in Table 4, 《世说新语》 is labeled as a reference), we also labeled some birth & death, identity and other character information, in order to better present the "cultural information" of a noun.

4.2.2 Using CIDOC-CRM as Reference

Conceptual Reference Model (CRM) proposed by the International Committee for Documentation (CIDOC) is used to represent the relationship of the labels. CRM is a knowledge ontology that facilitates the integration, transfer, and mutual exchange of diverse cultural heritage information sources. CIDOC-CRM (ISO/CD21127) has been accepted as the international standard for CRM for cultural assets museums consulting on cultural objects by ISOTC46/SC4 and is widely used.

In CRM, concepts are known as "Classes" and relationships are known as "Properties". CRM 7.2.1 defines 81 Classes and 148 Properties. The high-level categories are mainly Temporal_Entity, Time-Span, Place, Dimension, and Persistent_Item. Take the class Event as an example, the following will illustrate the way CIDOC-CRM describes classes and the property relationships between them.

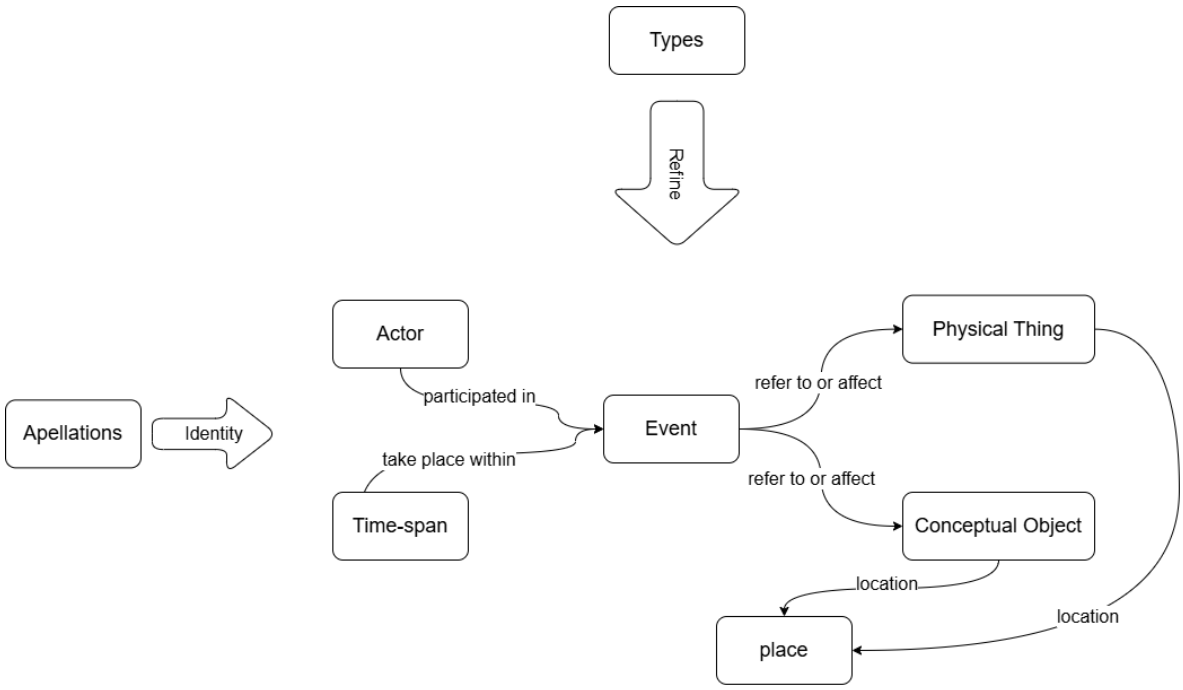

Figure 2. CIDOC-CRM description of "Event" classes and property relationships

The class "Event" needs the class "Actor", so the property relationship between them is "participated in". Similarly, the property relationship between the class Event and the class Time-span is "take place within".

4.2.3 Mapping Labels to CIDOC-CRM's System

In CIDOC-CRM, the class that directly describes characters is *E21_Person*, which includes both real persons and fictional character images. So we classified all the words under *{human | 人}* in thesaurus into *E21_Person*. The classes describing birth and death in CRM are *E67_Birth* and *E69_Death*, which refer to the birth and death of any living thing respectively. Therefore the label */生卒/* (birth & death) can be mapped to these two classes. The dynasty in which a person lives is a time-span concept that corresponds to the *E52_Time-Span* in CRM, so the label */朝代/*(dynasty) can be mapped to this class. By analogy, the label */出生地/*(birthplace) can be mapped to *E53_Place*, */称谓/*(Alternative Names) can be mapped to *E41_Applellation*, */事件/*(Events) can be mapped to *E5_Event*, */身份/*(Identity) can be mapped to *E55_Type*, and */成果*

/(Achievements) can be mapped to *E28_Conceptual Object*. Finally, the mapping between annotation in definiton and classes in CIDOC-CRM is finished

Annotation Labels	Classes in CIDOC-CRM
/人物/(people){human | 人}	E21_Person
/生卒/(birth & death)	E67_Birth/E69_Death
/出生地/(birthplace)	E53_Place
/身份/(identity)	E55_Type
/地位/(status)	E62_String
/称谓/(alternative names)	E41_Appellation
/事件/(events)	E5_Event
/评价/(appraisals)	E62_String
/成果/(achievements)	E28_Conceptual Object
/参见文献/(references)	E31_Document
/朝代/(dynasty)	E52_Time-Span

Table 5. Example of mapping for some labels

However, some labels could not correspond to the classes in CIDOC-CRM, so 8 new classes were added, starting with T, to distinguish them from the established classes. The CIDOC-CRM hierarchy table of 11 categories of nouns is shown below. The "-" in front of a class is used to indicate its hierarchy." The fewer the number of "-", the higher the level is, and the more abstract the concept is.

Examples of Hierarchy of CIDOC-CRM	New Classes Added to CIDOC-CRM
E1 -CRM Entity	T1 -Manner
E2 -Temporal Entity	T2 -Reference
E3 - -Condition State	T3 -Example
E4 - -Period	T4 -Assertion
E5 - - -Event	T5 -Qualitative
E13 - - - - -Attribute Assignment	T6 -Style
E14 - - - - - - Condition Assessment	T7 -Origin
E65 - - - - -Creation	
E63 - - - - Beginning of Existence	
E67 - - - - - Birth	
E64 - - - - End of Existence	

Table 6. Examples of hierarchy of CIDOC-CRM and new classes added to CIDOC-CRM

At this point, the classes can be connected through the properties in CIDOC-CRM. The property relationships between classes are shown in the following table:

Classes as sources	Property	Classes as goals
E21_Person	P98_brought into life(was born)	E67_Birth
E21_Person	P100_was death of(died in)	E69_Death
E39_Actor	P11_had participant(participated in)	E5_Event
E39_Actor	P11_had participant(participated in)	E5_Event
……	……	……

Table 7. Property relationships between classes

4.2.4 Visualization with Protégé

Protégé is a free, open-source ontology editor developed by Stanford University. The relationships between nouns can be visualized by adding the mapped labels and relationships between labels to protégé. After adding classes and properties into protégé, OntoGraf can be used to visualize them.

Once a sufficient number of instances have been expanded, the entire ontology can be presented as a knowledge graph. Furthermore, it is possible to switch between different views, which facilitates the observation of relationships among various entities as well as among different nouns. This is illustrated in the figure below.

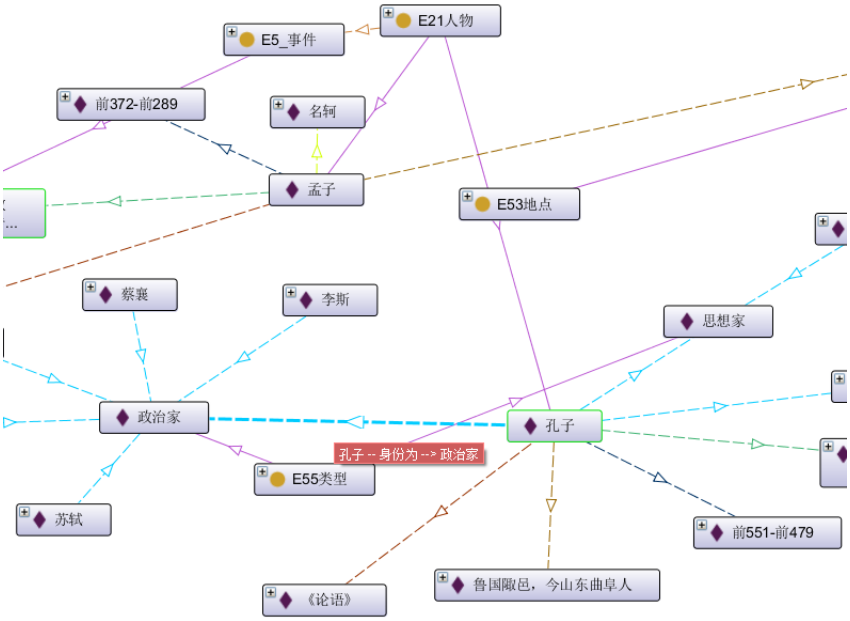

Figure 3. Visualization of Chinese traditional cultural nouns

This graph takes "Confucius" as the core node. In the graph, "Confucius" belongs to the class "E21_Person". His birthplace is associated with "Zuoyi, Lu" (present-day Qufu, Shandong Province), and his birth date is given as 551 B.C.E.. His representative work is The Analects of Confucius. In addition, the graph illustrates the connections between "Confucius" and historical figures such as "Mencius" and "Li Si".

V QUANTITATIVE ANALYSIS OF CHINESE CULTURAL TRADITONAL NOUNS

5.1 Statistics on cultural nouns in compulsory education

A total of 1385 Chinese traditional cultural nouns related to compulsory education were extracted from the knowledge base, covering 59 categories. After categorizing them by educational level into low primary (grades1-2), mid primary (grades3-4), high primary (grades 5-6), and junior high school(grades 7-9), it was found that the number of Chinese traditional cultural nouns increases from low primary to junior high school in general.

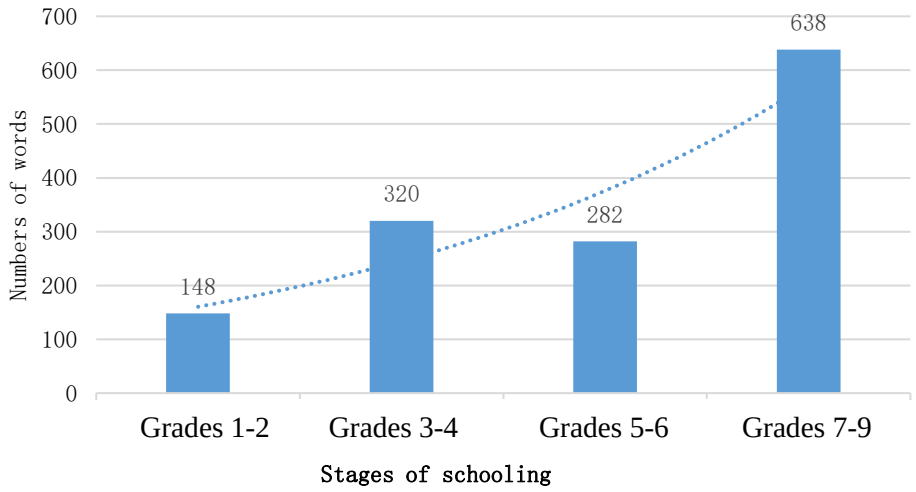

Figure 4. Number of words for Chinese traditional cultural nouns in compulsory education

Some of the categories were extracted from the knowledge base to observe their distribution across the stages of compulsory education, as shown in the figure 5:

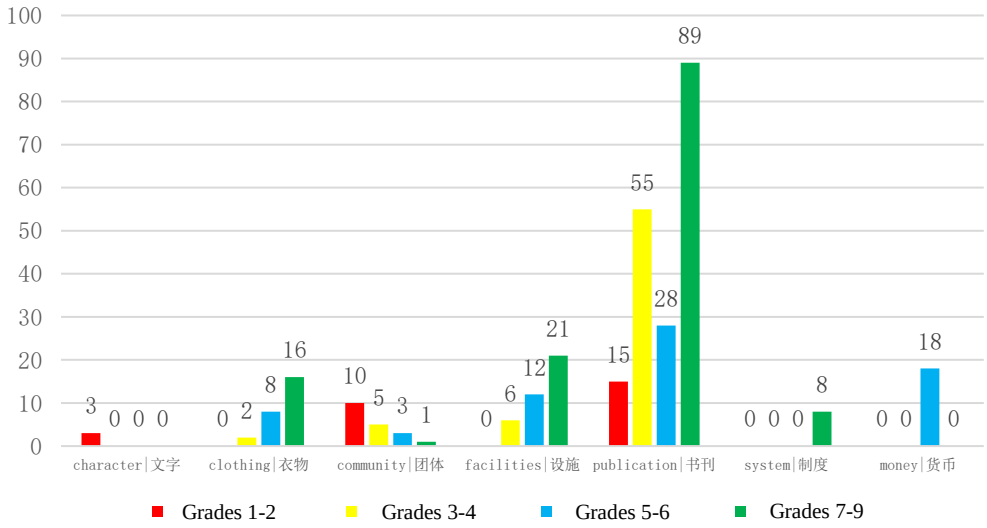

Figure 5. Distribution of categories of Chinese traditional cultural nouns in compulsory education

The "character|文字" in the image above is only distributed in grades 1-2, and there are only three words in the category: "pictographs", "associative compounds", and "pictophonetic". Literacy is the basic requirement for reading, and learners in the compulsory education stage, only need to understand the relevant knowledge of the text at the beginning of learning.

In the image above, the two categories of "clothing|衣物" and "facilities|设施" show an increasing number of internal words and learning difficulty as grades increase. Taking "clothing|衣物" as an example, students in grades 3-4 only need to master words such as "belt" and "crown", while students in grades 5-6 need to master more complex words such as "Ru Skirt" and "cross collar". According to prototype theory, the words learned during primary school are often at the basic level categories, and even if they are not at the basic level categories, the integrity of the words is relatively high and easy to understand. However, the words learned during junior high school are mostly located in subordinate categories, such

as "wooden clogs", "coronation robes", "funeral clothes", etc. Correct understanding and application of these words require students to have a higher level of cognition.

While "money|货币" and "system|制度" are only distributed in the upper grades of primary schools and junior high schools, as the vocabulary within these categories is often associated with specific economic and political phenomena, and the requirements for learners are relatively high.

To sum up, the system as a whole matches the learning abilities and cognitive characteristics of students in the compulsory education stage. As they grow older, learners need to master more Chinese traditional cultural nouns, which is more difficult and involves much more culture and phenomena.

5.2 Statistics on cultural nouns from different origins

Further analysis of the terms in the compulsory education stage from both *China Culture Daily* and *Textbook* revealed a total of 1989 words, of which 630 were only found in *China Culture Daily*, 1176 were only found in *Textbook*, and 183 were found in both *China Culture Daily* and *Textbook*. There were a total of 91 categories, of which 48 were only found in *China Culture Daily*, 8 were only found in *Textbook*, and 35 were found in both *China Culture Daily* and *Textbook*. The distribution of categories is shown in the following figure:

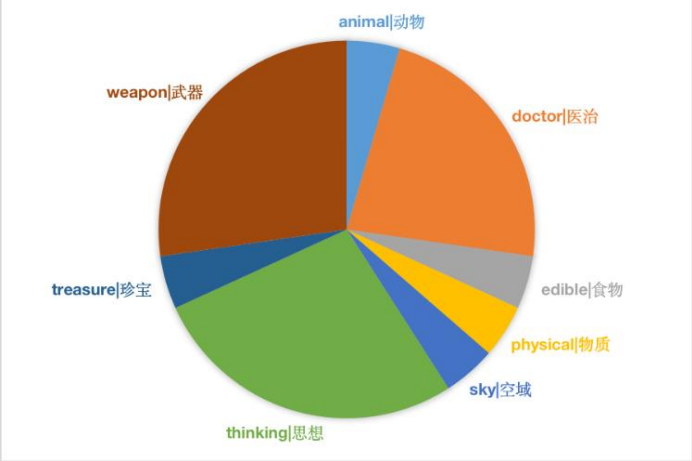

Figure 6. Categorization only in *Textbook*

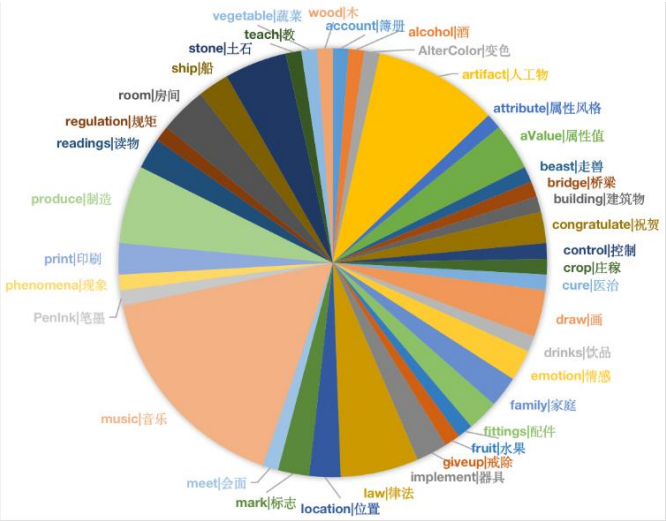

Figure 7. Categorization only in *China Culture Daily*

In terms of the unique categories covered by *China Culture Daily* and *Textbook*, the former contains a broader range of unique categories, including technology, nature, emotions, objects, animals, art, and skills, while the latter contains relatively fewer unique categories, mostly including words such as objects. It can be seen that the cultural knowledge involved in cultural words is more extensive, focusing on various aspects of public life, and paying attention to more professional scientific interpretation of a certain thing. The entries in the *Textbook* are more humanistic, coming from classic texts and humanistic allusions, indicating that school education pays more attention to the inheritance and development of classical culture. The knowledge base constructed based on these two corpora can be used for both compulsory education and popularization.

VI DISSCUSSION

Firstly, the definition of traditional cultural words needs to be further improved. Even though many restrictions have been added to the definition, there may still be ambiguity in the specific practice process, which needs to be supplemented and revised in time.

Secondly, even though the knowledge base of traditional cultural nouns itself conforms to the cognitive pattern of adolescents and the selection of words is representative to a certain extent, whether they can adequately reflect the entire contents of traditional culture needs further study.

Thirdly, although the entity and property categories have been adjusted multiple times during the construction of the ontology model, the number of instantiations is insufficient.

VII CONCLUSION

This study constructs a knowledge base of Chinese traditional cultural nouns through the definition, classification, and semantic annotation of nouns rooted in traditional culture. By combining thesaurus construction with ontology modeling and visualizing the data via Protégé, the study effectively maps cultural concepts and their interrelations within a structured semantic framework. The integration of HowNet's sememe system and CIDOC-CRM standards allows for a fine-grained and interpretable representation of traditional cultural knowledge. Quantitative analysis further demonstrates the alignment between cultural vocabulary and the cognitive development stages of compulsory education. The resulting knowledge base not only facilitates cultural inheritance and educational application but also lays a solid foundation for further studies in semantic analysis, cultural computing, and intelligent knowledge services.

References

Nida E. Linguistics and ethnology in translation-problems. *Word*. 1945;1(2):194–208.
Nida E. *Toward a Science of Translating*. Shanghai: Shanghai Foreign Language Education Press; 2004.
Eastman C. "Culture-loaded" vocabulary and language resurrect. *Current Anthropology*. 1979;20(2):401–402.
Newmark P. *A Textbook of Translation*. Shanghai: Shanghai Foreign Language Education Press; 2001.
Chang J. *Chinese Vocabulary and Culture*. Beijing: Peking University Press; 1995. (In Chinese)
Li X. *Research on Cultural Teaching of Chinese as a Second Language*. Beijing: The Commercial Press; 2007. (In Chinese)
Chen L. The translation of the culture-specific words. *J Chongqing Vocational & Technical Institute*. 2006;15(3):113–115. (In Chinese)
Kilgarriff A, Yallop C. What's in a thesaurus? *LREC*. 2000:1371–1379.
Aitchison J. The thesaurofacet: A multipurpose retrieval language tool. *Journal of Documentation*. 1970;26(3):187–203. \
Aitchison J, Clarke S. The thesaurus: A historical viewpoint, with a look to the future. *Cataloging & Classification Quarterly*. 2004;37(3–4):5–21.
No.628 Research Institute of the Third Ministry of Machinery Industry. *Aeronautical Scientific and Technical Terms Thesaurus*. Beijing: No. 628 Research Institute; 1965. (In Chinese)
Bu S, Wang D. Chinese classified thesaurus and its development and application in the internet era. *Library and Information Service*. 2005;49(7):25–28, 75. (In Chinese)
Dong Z, Dong Q. HowNet – A hybrid language and knowledge resource. In: *Proc. of Int. Conf. on Natural Language Processing and Knowledge Engineering*. 2003:820–824.

Yuan L, Zhang H, Chen J, Lu J. Design pattern of knowledge base based on ontology knowledge model. *Computer Engineering and Applications*. 2006;30:65–68, 104. (In Chinese)
Hamid R, David M, Lakshmi S, Richard T. Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. *Decision Support Systems*. 2002;33(2):143–161.
Wei Q, Liu M. Construction of knowledge base of intangible cultural heritage. *Research on Library Science*. 2020;6:33–38. DOI:10.15941/j.cnki.issn1001-0424.2020.06.005. (In Chinese)
Yan L, Chu H. Extracting concepts and semantic associates for teaching Tang 300 poems to L2 learners. *Lecture Notes in Computer Science*. 2022;13495:261–272.
Su X. Theory and method in compiling list of common words in compulsory education (draft). *Applied Linguistics*. 2017;3:2–11. DOI:10.16499/j.cnki.1003-5397.2017.03.001. (In Chinese)
Gracia J, Kernerman I, Bosque-Gil J. Toward linked data-native dictionaries. In: *Proceedings of eLex 2017: Electronic Lexicography in the 21st Century*. Leiden, Netherlands: Lexical Computing CZ s.r.o.; 2017:550–559.
Speer R, Chin J, Havasi C. ConceptNet 5.5: An open multilingual graph of general knowledge. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. 2017;31(1).
Musen MA. The Protégé project: A look back and a look forward. *AI Matters*. 2015;1(4).