DCPAPERS
Proc. Int'l Conf. on Dublin Core and Metadata Applications

Short Paper——Research on Ontology Design for Protection and Utilization of Great Sites

Si Li
Wuhan University, China
lsiwhu@163.com

Zhou Jing
Wuhan University, China
zhoujingwinky@whu.edu.cn

Liu Yao
Wuhan University, China
2016302480135@whu.edu.cn

Abstract

In the process of rapid urbanization in China, the protection and utilization of great sites are facing unprecedented pressure. The effective knowledge organization of great sites is a prerequisite for their protection and utilization. Ontology provides realization paths for organizing knowledge of great sites. In this paper, firstly, CIDOC- CRM and Time ontology are reused to build the top-level ontology with the results of user interviews. Secondly, top-down concept extraction and bottom-up concept expansion are adopted to gain the knowledge concepts and instances of great sites. Finally, data properties and object properties are defined. The designed ontology can provide the knowledge modeling and representation of great sites, laying the foundation for knowledge sharing.

Keywords: great site ontology; linked data; knowledge organization

1. Introduction

The great site is an important concept put forward from the perspective of heritage protection and management in China since 1990. It refers to the large site and their cultural landscapes that reflect the political, religious, industrial, agricultural, historical and cultural information at various stages of development in ancient Chinese history, with large scale, significant value and far-reaching impact(National Cultural Heritage Administration and Ministry of Finance of the People's Republic of China, 2005). Its number accounts for about 1/4 of the national key cultural relics departments, but its scale far exceeds other immovable cultural relics.

For the protection and utilization of great sites, China National Cultural Heritage Administration(2020) pointed out that the value utilization activities of great sites should be carried out based on sites, environments and information resources. On the basis of ensuring the safety of the great site itself and its environmental features, making full use of their information resources, and giving full play to the social and cultural value have become the focus of the society.

As cultural heritage (CH), great sites have intrinsic value, but only when documented and interpreted can them 'tell a story' of a culture or history (Lanzi 1999). These intellectual links (e.g., people, places, time, themes) exist in information resources of great sites which cover a wide range of fields and contain a wealth of knowledge. Knowledge organization methods, especially ontology, can promote knowledge modeling and representation of great sites. So, this paper intends to use ontology technology to extract and organize the knowledge of great sites and construct the great site ontology. In this way, the efficiency of knowledge dissemination and sharing in protection and utilization of great sites can be improved.

2. Methods

2.1. Top-level Ontology

(1) Users' understanding of great sites

A user-centered ontology can provide a vocabulary common to different stakeholders and thus optimize the interaction between practitioners and the expert system (Basu 2019). Thus, we conduct user interviews to figure out their knowledge needs of great sites. Considering the feasibility and breadth of the interviews, our interviewees include: 6 students majoring in archaeology, 3 curators

https://doi.org/10.23106/dcmi.953105403

and practitioners in cultural relics departments, 4 researchers, and 3 site museum enthusiasts. These interviewees. These respondents have a strong need for knowledge of great sites and a willingness to organize them. The interview questions and results are shown in Table 1.

TABLE 1: Interview questions and results

Interview Questions	Interview Results	
Q1: When you study or visit the great site, what information do you most want to know?	The site and its relics (6) Function (3) Area Temporal (9) Location (2) Environment (1) Historical events (7)	Cultural value (7) Academic value (2) Archaeological excavations (7) Site utilization (6) Site research (5) Site protection (8) Site Management (5)
Q2: Do you pay attention to the protection and utilization of great sites? If yes, what information are you concerned about?	Excavation technology (2) Protection measures (6) Protection Cases (3) Protection Achievements (7) Academic Achievements (2)	Utilization (4) Community impact (1) Departments and practitioners (1) Policies and plans (4)
Q3: What do you think are the components of the Great Site?	Relics (9) Remains (10) Natural environment (11) Cultural landscape (11)	Practitioners (1) Facilities (2) Management and research departments (3) Derivative products and services (1)

The interviews find that the interviewees are most concerned about the basic information, environmental, historical and cultural background of great sites. They also pay attention to their historical evolution in different life cycles. The interviewees generally believed that the elements of a great site include relics, remains, natural environment and cultural landscape. Staffs, curators and practitioners of museums and cultural relics departments pay more attention to the excavation, protection, management, utilization and research activities of great sites, especially the cases, technologies, measures, achievements, impacts, and planning.

(2) Reusing formal ontologies

According to the results of existing researches and interviews, the great site knowledge includes its basic characteristics such as relics, function and area, and knowledge about their persons, places, events, objects, time, etc. These concepts are consistent with CIDOC-CRM and Time Ontology.

CIDOC-CRM, developed by the International Committee for Documentation, is a theoretical and practical tool for information integration in the field of CH (https://www.cidoc-crm.org/). CIDO-CRM is based on an event-centric information modeling, which means other classes like persons, concepts, and places are connected to each other via events (Araújo et al. 2018; Ranjgar et al. 2022). The top-level knowledge concepts of the great site ontology can correspond to the classes of CIDOC-CRM, such as E21 Person, E74 Group, E53 Place, E5 Event, and E70 Thing.

Time ontology provides a vocabulary for expressing facts about ordering relations among instants and intervals. The concept of time in great site ontology can be represented by the Temporal class of the Time ontology. And its subclasses Instant and Interval are reused.

So, this study reuses 5 classes, and customizes 6 classes: Great Site, Historical Event, Historical Actor, crm: Thing, crm: Place, crm: Person, crm: Group, Protection and Utilization Activity, Information Resource, time: Temporal.

2.2. Knowledge Concepts

(1) The top-down approach

The top-down development process was utilized to build the hierarchy with the most general classes and specialize afterward. This paper analyzes thesaurus, dictionaries, laws and regulations and consults ten experts from the Institute of Archaeology, China Academy of Social Sciences, Wuhan University, Hunan Provincial Museum, etc., the top-level structure of the knowledge concepts is determined. It is divided into two dimensions: great sites and protection and utilization. Some concepts are shown in FIG. 1 (the third-level concept has not been fully developed).

https://doi.org/10.23106/dcmi.953105403

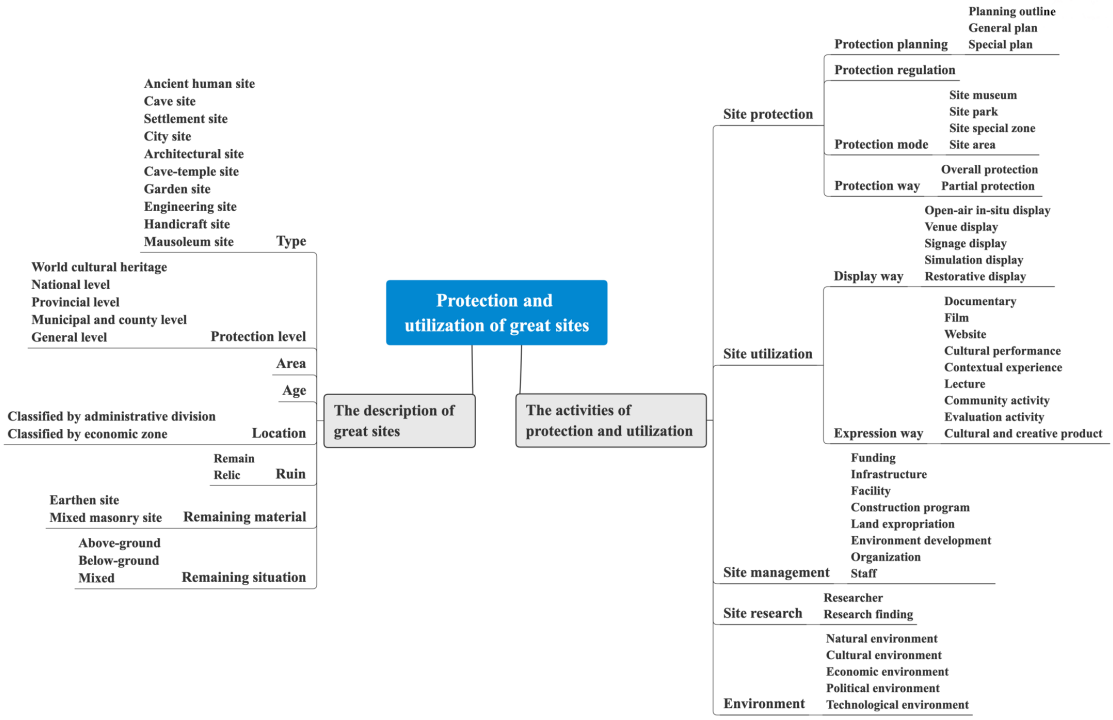

FIG. 1. The top-level structure of knowledge concepts

(2) The bottom-up approach

The bottom-up approach is applied in order to enrich and populate the structure by extracting low-level knowledge concepts.

First, building the initial corpus. Taking great sites in Xi'an Area as an example, this paper collects 24 archaeological reports and briefings, 106 journal papers, 3 books, 7 site management planning and regulations, 7 official websites of great sites, 9 project documents and reports as data sources. With a combination of OCR technology and manual proofreading, an initial corpus protection and utilization of great sites has been built.

Second, discovering and categorizing knowledge concepts and instances.

a. Using Jieba Chinese word segmentation tool and NLPIR-Paeser tool to perform dictionary word segmentation and new word discovery on the corpus and obtain multiple "word/part-of-speech tags";

b. Enriching the user dictionary and identify specific nouns such as Chinese dynasty, ancient official position, ancient people's name, etc. with the Bigrams function which can calculate the co-occurrence frequency of binary word pairs in the corpus;

c. Outputting entities such as persons, things, events, time, locations through named entity recognition;

d. Artificially categorizing and supplementing knowledge concepts and instances.

Finally, generalizing and clustering knowledge concepts and instances.

3. Results

(1) Classes and data properties

Defining data properties of each class provides a specific understanding of the knowledge concepts in great sites, as shown in Table 2. Ontologies and metadata sets such as DC, FOAF, SHLNames, GeoNames, VRA Core are reused.

TABLE 2: Classes, data properties of the great site ontology

Class	Subclass	Data Property

https://doi.org/10.23106/dcmi.953105403

Great Site	Ancient human site Cave site Settlement site …	site_name, protection_level, remaining_material, remaining_status, area
Historical Event		event_title, event_process, event_effect
Historical Actor		actor_name, schema:alternateName, gender, shl:briefBiography, shl:speciality, shl:courtesyName, shl:pseudonym, shl:nativePlace, shl:birthday, shl:deathday
crm:Thing	Remain Relic	vra: title, protection_level, vra:description, vra:measurements, vra:material, vra:technique
crm:Place		shl:country, shl:province, shl:city, shl:county, shl:town, village, zone, gn:name, gn:alternatename, gn:postal code, gn:longitude, gn:latitude
crm:Person		foaf:name, foaf:gender, foaf:mbox_shalsum, schema:telephone, shl:officialPosition, degree, shl:officialExperience, shl:createdWork, shl:birthday, shl:deathday
crm:Group		group_name, group_location, group_description, group_web, schema:telephone
Protection and Utilization Activity	Site protection	protection_planning, protection_regulation, protection_mode, protection_way
	Site utilization	display_way, expression_way
	Site management	funding, infrastructure, facility, construction_program, land_expropriation, environmental_development
	Site research	research_finding
Environment	Natural environment	geologic_feature, topographical_feature, climate, disaster
	Cultural environment	cultural_industry, religion, tradition, educational_level
	Economic environment	city_area, city_population, gdp, financial_revenue, financial_expenditure, tourism, disposable_income
	Political environment	law, policy
	Technological environment	internet_technology, digital_technology, cultural_relics_restoration_technology
Information Resource	bibo:Book bibo:Article …	dcterms:title, dc:subject, dc:creator, dc:description, dc:publisher, dc:language, dcterms:date
time: Temporal	time:Time instant	shl:temporal, time:year, time:month, time:day
	time:Time interval	time:years_duration, time:months_duration, time:days_duration

(2) Object properties

Defining the object properties can accurately represent the internal relationship of knowledge concepts and provide a basic logical structure for semantic network. The object properties in this ontology are shown in Table 3. We divide object properties into 4 types: inheritance relation, spatio-temporal relation, subordinate relation, action relation. The relationships and class hierarchy structure followed in the ontology are shown in FIG. 2.

FIG. 2. Object properties of the great site ontology

Type	Object Property	Type	Object Property
Inheritance relation	is-a	Action relation	isDescribedBy/describes
	isPartOf		isDamagedBy/damages
	isInstanceOf		wasDestroyedBy/destroyed
Spatio-temporal relation	time:after/before/hasTemporalDuration/hasTime		isDiscoveredBy/discovers
	crm: isLocatedOnOrWithin/tookPlaceAt		isDiggedBy/digges
Subordinate relation	crm:contains		isExpressedBy/expresses
	crm:isCurrentOrFormerMemberOf		isExpropriatesBy/expropriates
	shl:creatorOf/directorOf		isExcavatedAt/excavates
	outcomeOf		isFundedBy/funds
	isSubordinateTo		wasGivenBy/gave
Action relation	foaf:knows		isInfluencedBy
	hasCulturalEnvironment/ hasEconomicEnvironment/…		crm: wasModifiedBy/modified
	hasCulturalInfluence/hasEconomicInfluence/…		isManagedBy/manages

https://doi.org/10.23106/dcmi.953105403

	hasSpouse/hasFellow		wasOwnedBy/owned
	hasActivity		crm: wasProducedBy/produced
	hasRemain/Relic		isProtectedBy/protected
	hadEvent		isPlannedBy/plans
	crm: participatedIn		isRepairedBy/repairs
	wasBoughtBy/bought		isStudiedBy/studies
	isDisplayedBy/displays		isUsedBy/uses

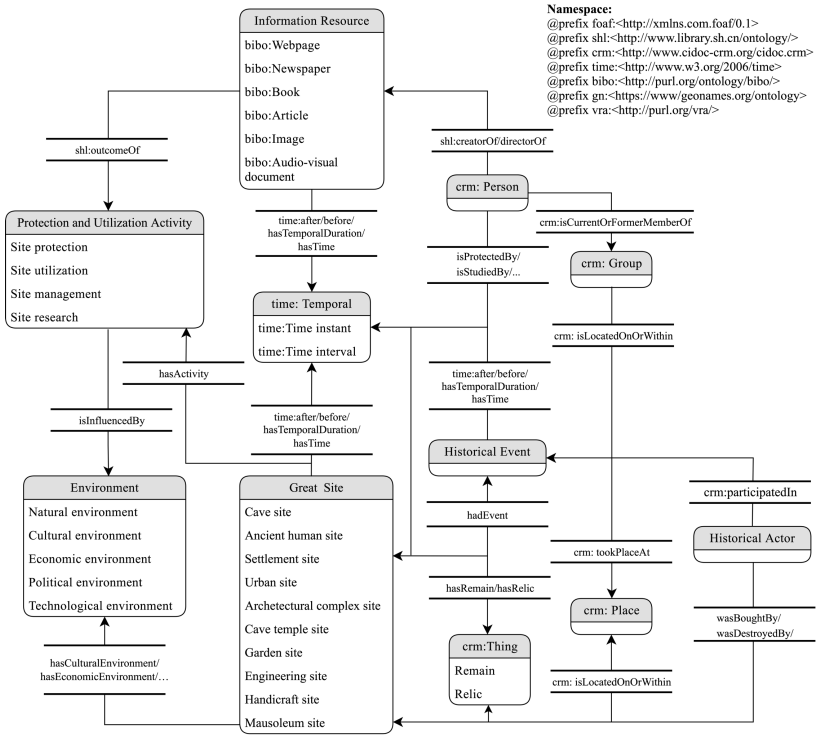

FIG.2. The relationships and class hierarchy structure of the great site ontology

4. Conclusions

The protection and utilization of great sites in China have gone through three stages of "passive protection—exploration for utilization—promoting protection by utilization"(W. Weihong 2020), which rarely involves the excavation and expression of the knowledge behind the great sites. Aiming at this research gap, from the perspective of ontology, this paper organizes, publishes and displays the diverse and heterogeneous great site knowledge. Based on this, we will further visualize and instantiate the ontology in Protégé. With the help of the good conceptual hierarchy and logical reasoning function of the ontology, we can implement semantic retrieval and reasoning in order to obtain the explicit and implicit knowledge of the great site.

The great site contains a wealth of knowledge to be revealed and utilized which is recorded in multiple types of information resources. This paper only sorts out basic knowledge elements and chooses several great sites as instances. The granularity of knowledge organization needs to be further refined. In the future work, we will expand data sources and research objects, study the organization of massive heterogeneous knowledge in great sites, and realize a wider range of online retrieval with external vocabulary.

https://doi.org/10.23106/dcmi.953105403

References

Araújo, C., Martini, R. G., Henriques, P. R., & Almeida, J. J. (2018). Annotated documents and expanded CIDOC-CRM ontology in the automatic construction of a virtual museum. Developments and advances in intelligent systems and applications (pp. 91–110).

Basu, A. (2019). Semantic web, ontology, and linked data. Web services: concepts, methodologies, tools, and applications (pp. 127–148).

China National Cultural Heritage Administration. (2020). Guidelines for the Utilization of Great Sites (Trial). Retrieved June 4, 2022, from http://www.ncha.gov.cn/module/download/downfile.jsp?classid=0&filename=2008251015511896708.pdf

Lanzi, E. (1999). Introduction to vocabularies: Enhancing access to cultural heritage information. Getty Publications.

National Cultural Heritage Administration, & Ministry of Finance of the People's Republic of China. (2005). Notice on the issuance of the "Great Site Protection Special Funds Management Measures. Retrieved June 4, 2022, from http://www.mof.gov.cn/zhengwuxinxi/zhengcefabu/2005zcfb/200805/t20080519_22511.htm

Ranjgar, B., Sadeghi-Niaraki, A., Shakeri, M., & Choi, S.-M. (2022). An ontological data model for points of interest (POI) in a cultural heritage site. Heritage Science, 10(1), 1–22.

Weihong, W. (2020). Theory, Method and Positioning: Three Questions on Soil Heritage Sites Protection and the Construction of Heritage Parks. Southeast Culture, 23–29.

https://doi.org/10.23106/dcmi.953105403