

How to organize and reveal knowledge of cultural heritage digital resources? ---A case study of Dunhuang Grottoes

Xiaoguang Wang

School of Information Management,
Wuhan University, Wuhan,430072,
CHN, wxguang@whu.edu.cn

Wanli Chang

School of Information Management,
Wuhan University, Wuhan,430072,
CHN, wanlichang@whu.edu.cn

Hongyu Wang*

School of Information Management,
Wuhan University, Wuhan,430072,
CHN, wanghongyu@whu.edu.cn

ABSTRACT

In this paper, we take Dunhuang grottoes as an example to show the graph-based knowledge organization of cultural heritage digital resources. Through the construction of knowledge graph, we further reveal the rich domain knowledge contained in Dunhuang digital resources, presenting a novel method for knowledge organization and revelation of cultural heritage digital resources as well as an idea for the construction of digital humanities infrastructure.

CCS CONCEPTS

Information systems-Digital libraries and archives

KEYWORDS

Cultural Heritage; Knowledge Organization; Knowledge revelation; Graph Model; Dunhuang Grottoes

ACM Reference format:

Xiaoguang Wang, Wanli Chang, Hongyu Wang. 2019. How to organize and reveal knowledge of cultural heritage digital resources?---A case study of Dunhuang Grottoes. In *Proceedings of The 19th ACM/IEEE Joint Conference on Digital Library (JCDL'19)*. ACM, Urbana-Champaign, IL, USA, 2 pages. <https://doi.org/10.1145/1234567890>

1 INTRODUCTION

Cultural heritage carries abundant cultural connotation and knowledge. The advancement of digital technologies, such as big data and surveying and mapping, provides an essential data-thinking and technical support for the digitalization and management as well as application of its associated information. Digital resources of cultural heritage, which are multi-source, heterogeneous and cross-modal, blesses humanists with a data-driven research paradigm, enabling the development of long-distance, quantitative and tempo-spatial visualized research in the fields of arts and history, and expediting the development of emerging and interdisciplinary digital humanities. Advancing the research in digital humanities becomes a new mission of libraries, so as to provide high-quality digital resources and accessible tools for their organization, processing, retrieval and application[1].

In the context of digital humanities, the discovery and revelation of knowledge implied in cultural heritage, in virtue of semantic, networked and expandable organization of digital resources of cultural heritage, will efficiently improve the inheritance, application and protection of cultural heritage. Currently, meta-data framework and ontology model are the two most commonly used models knowledge organization. Meta-data framework is inefficient on its own in representing the relationship between the knowledge units, and while ontology model and/or RDF is capable of doing so, yet incapable of recording the attributes associated with the relationships (Inability to Qualify Instances of Relationships), and therefore it can not record the evolution of relationships. In addition, when adding or associating new resources, it requires to redesign ontology model when adding or associating new resources,

hurdling a convenient and efficient expansion of knowledge.

[2]Attribute graph model, which can reflect the deep-seated semantic relationship between data objects through the representation of entities, relationships and their attributes, is conducive to multi-dimensional knowledge mining and deep-seated evolution analysis, and thus provides a solution to the organization, analysis and disclosure of cultural heritage knowledge.

As widely recognized, Dunhuang Grottoes as a whole is a treasure house of Chinese culture and art. Dunhuang Studies have accumulated a wealth of digital resources as a result of the development of "Digital Dunhuang" and other projects. The digital resources, however, scattered and unrelated as they are, can not underpin the deep-seated knowledge mining and analysis, thus impeding to some extent the progress of Dunhuang Studies. This paper put forwards an attribute map model according to the characteristics of Dunhuang digital resources, and then uses the knowledge map in organizing and storing domain knowledge contained in Dunhuang digital resources semantically, builds a knowledge discovery platform for Dunhuang digital resources, and provides new methods and examples for knowledge organization and discovery of cultural heritage digital resources.

2 GRAPH MODEL CONSTRUCTION

Graph model-based knowledge representation model of Dunhuang Grottoes provides the basis for integrating knowledge maps of Dunhuang grottoes and premises the fine-grained semantic revelation of Dunhuang digital resources. Dunhuang Grottoes and the cultural elements therein are complicated in structure, for example, Grotto 061 comprises of an archway and a main chamber with four walls, with each wall containing bountiful objects of cultural elements. To facilitate the knowledge organization of grottoes and their cultural elements, relevant digital resources are categorized into three types in this paper, namely, cave, component and object; and an all-round analysis is conducted of the relationships between and attributes of the types. Afterwards, on the basis of introduction on "Digital Dunhuang" website and with reference to *Metadata Specification of Grottoes and Monasteries* published by Peking University, an attribute graph model is built for the knowledge organization of Dunhuang Grottoes, as shown in Figure 1.

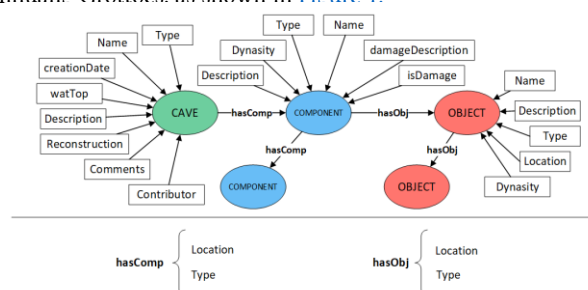


Figure 1: Design of the property map model of Dunhuang Grottoes

The attribute graph model enables us to extract structured association knowledge from data resources of Dunhuang Grottoes for the purpose of building a knowledge map. Due to its rich content, multitude of objects in Dunhuang Grottoes, and the varied description in different literature and documents (for example, "Sakyamuni" and "Rulai" as different representations

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

JCDL'19, June 2-6 2019, Urbana-Champaign, Illinois, USA

?2019 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-0000-0/19/22

<https://doi.org/10.1145/1234567890>

actually refer to the same entity), it's necessary to use such tools as thesaurus to disambiguate the entities and to realize knowledge integration. With reference to the category structure of AAT vocabulary, a vocabulary has already been produced, which has established an initial thesaurus with collections from relevant Chinese dictionary documents of Dunhuang, such as the *Dictionary of Tunhuangology*[3]. On this basis, a semi-automated method is used to process the *Dictionary of Tunhuangology* and over 700 papers from authoritative journals of Tunhuangology, and a vocabulary is thus established, realizing the hierarchical storage and management of subject keywords and example words in Dunhuang Studies[4]. An efficient and accurate knowledge organization of Dunhuang digital resources is achieved in this paper with the aid of the vocabulary.

3 DATA PROCESSING

Dunhuang Grottoes Content Catalogue, compiled by Dunhuang Research Institute and published in 1996, provides the source of data used in the knowledge map building. Based on the knowledge organization model of Dunhuang Grottoes, a fine-grained text semantic annotation tool is developed, so as to handily and accurately perform in-depth semantic annotation on data and to extract structured knowledge of Dunhuang Grottoes. The tool is shown in Figure 2.

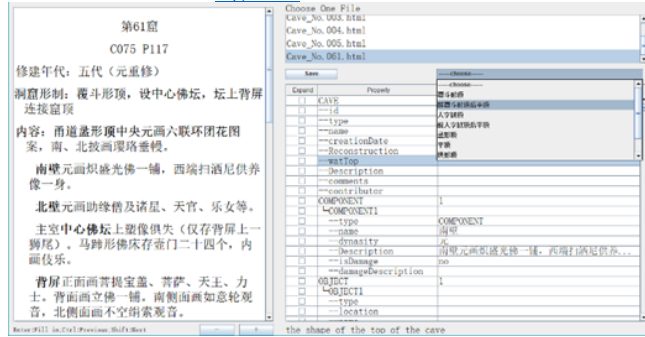


Figure 2: Text semantic annotation tool for knowledge graph

Three types entities of CAVE, COMPONENT and OBJECT as well as their relationships are extracted from *Dunhuang Grottoes Content Catalogue*, and their attributes are annotated, linking the knowledge organization and semantics of Dunhuang Grottoes in the form of attribute graph. As a step further, these knowledge are stored together with relevant multi-media resources in Neo4j graph database. Taking Grotto 61 for example, the contents of its west and south walls are extracted according to the Dunhuang Grottoes Map Model, as shown in Figure 3.

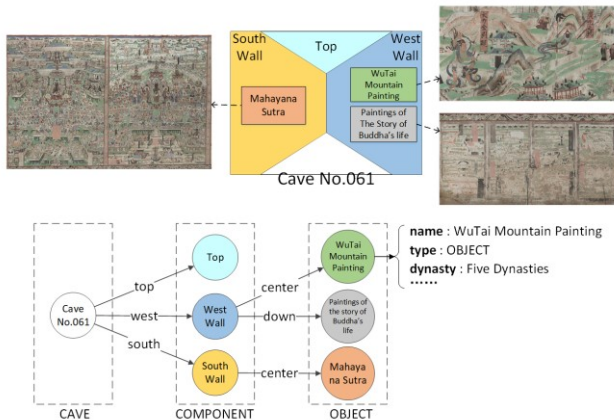


Figure 3: Example of knowledge extraction from Dunhuang Grottoes

4 RESULTS

The knowledge graph platform built in this paper of Dunhuang Grottoes, with its providence of knowledge navigation and semantic retrieval, can be applied to assist in further development of Dunhuang Studies. The platform allow users to search by inputting query words, and to view the structure and cultural elements of any specific cave by clicking on the navigation bar on the left.

For example, if the query of "Bodhisattvas appearing at the same time as the Avalokitesvara Sutra" is input, the answer will show all the eligible bodhisattvas, as shown in Figure 4. And the location of each specific bodhisattvas in the grotto is accessible according to the path. Meanwhile, on the platform, users may click on a specific node to get its metadata information. For example, by clicking on the "Bodhisattva with pot" node in Cave 123, user can see the metadata information of the node as well as the associated multimedia resources such as photos and videos.

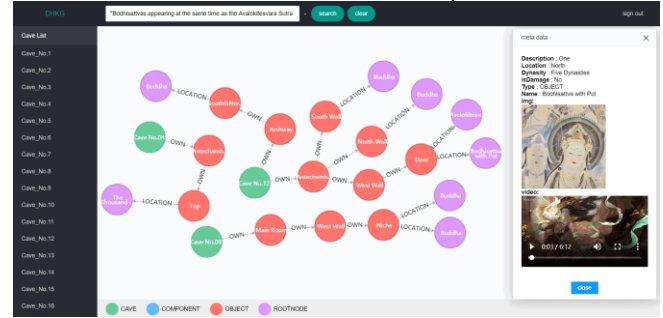


Figure 4: Semantic retrieval example of Dunhuang Grottoes Knowledge Graph

5 CONCLUSIONS

Using Dunhuang Grottoes as example, this paper illustrates the process of attribute graph model-based knowledge organization and disclosure of cultural heritage digital resources. Attribute graph model enables an handy and efficient semantic organization of entities of cultural heritage knowledge, their relationships and attributes, which further enables a fine-grained analysis and revelation of cultural heritage. It facilitates a deeper analysis of correlation between cultural heritage knowledge and cultural element objects, and provides another perspective for the organization of digital resources and the revelation of their cultural connotations. In addition, it provides a feasible way of thinking for the infrastructure development in digital humanities. In the following, the research will attempt to further standardize the metadata of Dunhuang Grottoes on the basis of international standards such as CDWA, CIDOC-CRM, etc., and work towards the associated release and sharing of datasets.

REFERENCES

- [1] Kretschmar, W. A., & Gray Potter, W. 2010. Library collaboration with large digital humanities projects. *Literary and Linguistic Computing*, 25 (4), 439-445.
- [2] Gracia, J., Liem, J., Lozano, E., Corcho, O., Trna, M., & Gomez Perez, M., et al 2010. *Semantic Techniques for Enabling Knowledge Reuse in Conceptual Modelling*. The Semantic Web – ISWC 2010.
- [3] Huinan Li, Hanghang Cheng, Xiaoguang Wang, et al. Chinese Dunhuang Mural Vocabulary Construction Based on Human-Machine Cooperation[C]. Digital Humanities Conference 2019. 2019
- [4] Han Huang, Hongyu Wang, Dawei Jin. A Low-Cost Named Entity Recognition Research Based on Active Learning[J]. *Scientific Programming*, 2018, 2018:1890683. 2018