Semantic Explication and Organization Methods for National History Knowledge: A Study

Authors: Wang Ying, Zhang Zhixiong, Sun Hui, Lei Feng

Date: 2016-01-25T00:00:00+00:00

Abstract

The in-depth development of education and research on the history of the People's Republic of China has consistently been a priority for all stakeholders. The semantic revelation and organization of national history knowledge is of great significance to its education and research. Building upon relevant research, this paper proposes a semantic revelation and organization methodology for national history knowledge, termed "downward mining, upward organization". This methodology is grounded in a national history ontology; based on the semantic mining and revelation of national history knowledge objects and related facts hidden in the text entries of national history resources, it constructs a national history knowledge network through the association of these objects, performs higher-level multi-dimensional organization and presentation of national history knowledge content based on temporal, categorical, hierarchical, and statistical relationships, and following this approach, develops a corresponding system platform that realizes the revelation, reorganization, and other innovative applications of national history knowledge. The semantic revelation and organization methodology for national history knowledge proposed in this paper holds reference value for the development and utilization of other types of knowledge.

Full Text

Preamble

A Method of Semantic Representation and Organization of Historic Knowledge on Contemporary China

WANG Ying, ZHANG Zhixiong, SUN Hui & LEI Feng

This article is an outcome of the project "History Education Network of the People's Republic of China," supported by the Chinese Academy of Social Sciences' Philosophy and Social Sciences Innovation Engineering Initiative.

Corresponding author: ZHANG Zhixiong
Email: zhangzhx@mail.las.ac.cn
ORCID: 0000-0003-1596-7487

Abstract

China possesses a huge volume of historic resources on its contemporary history. However, the organization of these historic resources is not satisfactory, and lots of valuable knowledge remains hidden and cannot be easily discovered or utilized. It is an urgent problem to mine the implicit semantic knowledge scattered across numerous historic resources and to organize this knowledge based on its internal relations. The authors believe that semantic representation and organization of these historic resources will benefit the education and research of contemporary Chinese history.

Based on the historic encyclopedia of contemporary China, this paper proposes a method called "Mining down, Organizing up" to semantically represent and organize historic knowledge hidden in text resources. This method helps users represent historic knowledge in a fine-grained manner, reorganize historic knowledge in a more semantically enriched way, and develop more innovative applications using historic knowledge. Based on contemporary Chinese history ontology, this method extracts knowledge objects and facts from unstructured historic text items, forms a historic knowledge network of contemporary China, and realizes multidimensional knowledge organization at a higher level through relations such as time, subclass, hierarchy, and statistics.

Aiming at key problems in this method, this paper puts forward specific procedures for representing and organizing historic knowledge of contemporary China. First, the paper builds an ontology conceptual model as the basic organization mode describing the skeleton of historic knowledge. Second, the paper determines core objects to populate the ontology instance library served as corpus for automatic identification. After that, the paper extracts facts automatically from text items by utilizing text mining technology, then ensures the reliability of these facts through judgment and complement of domain experts. Moreover, the paper builds a historic knowledge network based on association among knowledge objects. Finally, the paper realizes multidimensional knowledge organization on the basis of the network.

Based on this method, the authors represented and organized historic knowledge of contemporary China from text resources, and developed a system to implement historic knowledge visualization, reorganization, and other new applications including knowledge maps, relatedness analysis, and national history facts reconstruction. The limitations of this research are: (1) The accuracy of recognizing knowledge objects and relevant facts from text should be improved, especially the identification and recommendation of relevant national history facts—if improved, it will further reduce the workload of domain experts; (2)

The association calculation method of historic knowledge network is simple and has not fully applied current semantic similarity calculation and graph mining methods. These are key problems that need to be solved in future studies.

Studies show that the "Mining down, Organizing up" method can realize fine-grained representation of historic knowledge of contemporary China and innovative application of knowledge organization based on historic knowledge objects. It can be used as a new kind of knowledge representation and organization method applicable in other fields.

KEYWORDS: Chinese national history; Ontology; Text mining; Knowledge organization.

0 Introduction

Deepening education and research on the history of the People' s Republic of China has always been a priority. The popularization and research of national history knowledge is significant for promoting patriotism and strengthening national cohesion. With social digitalization and networking development, the scale of information resources is increasingly massive and diverse, yet their organization and management remain relatively simple. Valuable information often becomes submerged and cannot be easily revealed or utilized. Extracting fine-grained national history knowledge and facts scattered across numerous information resources, and presenting important knowledge points and their interrelationships since the founding of the nation, has become an urgent research problem.

Addressing this situation, this paper proposes a "Mining down, Organizing up" approach for semantic revelation and organization of national history knowledge. This method, based on national history ontology, semantically mines and reveals knowledge objects and related facts hidden in national history resource text entries. Through association of national history knowledge objects, it constructs a national history knowledge network and enables higher-level multidimensional organization and display based on time, hierarchy, statistics, and other relationships. This approach can help researchers effectively acquire and utilize information to produce research outcomes, promote the popularization of national history knowledge, and facilitate further processing, research, and open sharing of information resources.

In the historical domain, researchers have begun exploring knowledge organization and semantic content revelation. Some have constructed historical ontologies such as the KMT-CPC Cooperation History Ontology [1], the Anti-Japanese War History Ontology [2], and the Zizhi Tongjian Historical Ontology [3]. Professor Dong Hui' s team from Wuhan University [4-5] used the Twenty-Four Histories as a foundation, employing semantic technologies to build a semantic knowledge base for Chinese historical records and establishing an analytical plat-

form for basic Chinese historical texts, providing semantic knowledge services to users. Hyvönen et al. [6] constructed a Finnish historical events ontology and applied it in the CultureSampo semantic portal. Corda et al. [7] proposed a logical model of event ontology for discovering connections between historical events. Ide and Woolner [8] proposed historical ontology models reflecting relationships between entities in different time periods.

Drawing on these domestic and international research achievements, this paper proposes the "Mining down, Organizing up" approach and develops a corresponding system platform based on this concept.

---

1 Main Approach

The "Contemporary Chinese History Education Network" project aims to popularize national history knowledge and assist national history education. The key challenge is how to utilize automatic information processing technology to achieve deeper organization of national history knowledge and text entries. Original national history information resources were only organized by simple time periods (e.g., "Basic Completion of Socialist Transformation Period," "Comprehensive Socialist Construction Period," "Cultural Revolution"), which cannot effectively reveal deeper knowledge relationships within text entries.

The main problems to address are: (1) Which core national history knowledge objects need to be identified? (2) How to extract knowledge object-related facts from text resources? (3) How to further achieve deep-level organization?

To address these issues, this paper proposes specific methods for semantic revelation and organization of national history knowledge. Based on the national history ontology, this approach extracts important knowledge objects and facts from text resources through text mining technology, forms a complex network, and realizes multidimensional knowledge organization through association calculation on the network. The "Mining down, Organizing up" approach transforms national history knowledge from text entries into a national history knowledge network composed of knowledge objects, important facts, and text entries. This downward mining process is a deconstruction process. Through association and reorganization based on the generated knowledge network, it becomes a construction process, enabling innovative applications such as visualization, reorganization, and other new applications.

[Figure 1: see original paper] shows the overall approach for semantic revelation and organization of national history knowledge.

---

2 Specific Methods for Semantic Revelation and Organization of National History Knowledge

2.1 Building a National History Ontology Conceptual Model as the Basic Organization Pattern

National history materials feature detailed content with rich and complex historical details. To achieve effective organization, it is necessary to clarify which knowledge needs focused attention and which historical details should be highlighted—essentially determining the basic organization pattern for national history knowledge. This study uses ontology as the core framework for organizing national history knowledge, utilizing abstract classes to reveal core knowledge objects, and employing attributes and relations to reveal characteristics and relationships between knowledge objects. This foundation reveals the skeleton of national history knowledge and enables overall understanding.

Referencing the skeleton method [9] and seven-step method [10], and with assistance from national history experts, this study proposes a conceptual model for national history ontology based on domain characteristics. By analyzing text resources, the development 脉络 of national history mainly revolves around major historical events, important figures, organizations, documents, and concepts. Therefore, the national history ontology first defines core classes such as Event, Person, Organization, Document, and Concept/Term. To represent descriptive information involved in these knowledge objects, the ontology defines 15 numerical properties and 28 object properties to represent attribute information and relationships. To show historical event details, it defines numerical properties such as name and textual description, and object properties such as location with their value ranges. Based on attribute descriptions and relationship characteristics in national history knowledge, the ontology defines property constraints. For example, "hasSubEvent" and "isSubEventOf" are inverse relations with transitive properties.

[Figure 2: see original paper] shows the core classes of the national history ontology. [Figure 3: see original paper] shows the properties and relations of the Event class.

2.2 Determining Core Objects as the Basis for Automatic Identification

After determining the basic organization pattern, it is necessary to collect and organize knowledge objects as basic instances of the national history ontology to support further revelation and organization. This study extracted metadata and structural information from national history information resources such as historical dictionaries and encyclopedias, using standardized event names, person names, etc., as titles for text entries. In this process, direct associations between text entries and knowledge objects were established, while existing thesauri for persons, places, and names were fully utilized. Through automatic processing methods such as name rule conversion and manual revision, data standardiza-

tion was achieved. More importantly, national history experts screened and organized major historical events and meetings since the founding of the nation, 梳理 ed hierarchical relationships between events and meetings, and determined standardized names.

The standardized knowledge objects include 1,508 events, 2,621 persons, 1,035 organizations, 1,200 documents, and 262 special groups. These knowledge objects and relationships were populated into the national history ontology according to defined properties and relations, generating instance data and relationships. These knowledge object instance names serve as dictionaries for automatic identification and as the basis for revealing knowledge object-related facts in text entries.

2.3 Utilizing Text Mining Technology for Fact Discovery

Although automatic processing can discover potential knowledge, due to natural language complexity, the accuracy of text mining results remains difficult to guarantee and cannot be directly added to the ontology. National history experts must combine their domain knowledge to judge fact correctness through multiple reviews and revisions.

Based on the above-organized knowledge object name dictionary, automatic annotation technology performs semantic annotation on text entries. Custom named entity recognition rules achieve automatic extraction of events, meetings, and other knowledge objects. Using relation extraction technology, relevant facts about knowledge objects are automatically discovered from text entries, forming a series of fact triples. For example, the entry "Third Plenary Session of the 11th CPC Central Committee" in the *Dictionary of Chinese Communist Party History* reveals facts such as the session' s time, location, and participants. The entry in *An Outline of Modern Chinese History* reveals the same event' s time, location, and related concepts. The entry in *Encyclopedia of the People' s Republic of China* reveals the session' s time, location, related events, and other facts. These multiple source entries both verify fact accuracy and reflect close relationships between text entries.

2.4 Building Knowledge Network Based on Knowledge Object Associations

Through the above processes, text entries are transformed into objectified knowledge units that reflect knowledge content. Based on these facts and objects, national history experts establish explicit semantic associations between knowledge objects. By discovering associations between knowledge objects within the same or different text entries, direct or indirect associations between different knowledge objects and between text entries are obtained. National history knowledge is thus transformed into a national history knowledge network composed of knowledge objects, historical facts, and text entries.

[Figure 4: see original paper] shows the knowledge network construction for the "Third Plenary Session of the 11th CPC Central Committee." The network describes the session's content, participants, related events, and other knowledge objects, forming a complex network with intersecting associations between the object layer and fact layer.

2.5 Realizing Multidimensional Knowledge Organization Based on Knowledge Network

The national history knowledge network forms a complex network with entry layer, object layer, and associations between various levels. This enables standardized semantic representation of national history knowledge objects and facts, making retrieval, reorganization, and other knowledge exploration possible.

Time-based organization: Time dimension is the most direct angle for showing historical development. According to time sequences, text entries from different books can be organized by historical periods. Knowledge objects can be sorted by appearance time using the Time class in the ontology. Facts about the same knowledge object can be sorted chronologically.

Object-based organization: Through semantic associations, aggregation of the same type of knowledge objects can be achieved. Using facts about the same knowledge object, maps describing knowledge objects can be built.

Text entry organization: Knowledge objects and facts revealed within text entries provide the basis for deep organization of text entries. Based on these facts and objects, national history knowledge materials can be integrated with external resources such as other historical books, webpages, or various databases to build extended applications.

---

3 Application Effects of National History Knowledge Semantic Revelation and Organization

Based on the above method implementation, a system platform was developed to support national history knowledge revelation, association analysis, and fact reconstruction, providing rich materials for biographies, historical data compilation, and new book writing.

3.1 Realizing National History Knowledge Map Revelation

Knowledge maps enable navigation between knowledge and provide progressive guidance for domain knowledge browsing [12]. Based on the national history knowledge network, knowledge map revelation was implemented, providing associated network display of national history knowledge content. Through visualization, nodes represent knowledge objects and edges represent various semantic

relationships. Users can intuitively understand national history knowledge without reading text information and conduct extended reading, making knowledge acquisition more efficient.

For example, right-clicking the "One Country, Two Systems" concept node shows its proposal time as "1984" and that the concept was announced in the 1984 *Government Work Report*. Right-clicking again shows its knowledge source entry from the *Encyclopedia of the People' s Republic of China*. [Figure 5: see original paper] shows a fragment of the national history knowledge map. By clicking knowledge nodes, users can browse according to their associations. Clicking the "Third Plenary Session of the 11th CPC Central Committee" node allows further browsing of participants. Clicking Deng Xiaoping' s node enables browsing of the "One Country, Two Systems" concept, and clicking the *Government Work Report* (1984) node completes map-based navigation and browsing.

3.2 Realizing Association Analysis of National History Knowledge Objects

Using the complex network formed through semantic revelation and organization, graph traversal can discover associations between knowledge objects and reveal potential relationships. For example, starting from the "Third Plenary Session of the 11th CPC Central Committee," complex knowledge networks with path length 2 can be obtained, showing related meetings, organizations, and their interconnections. Although the session was held after Chairman Mao' s death, multiple indirect associations can still be discovered through the network.

[Figure 6: see original paper] shows an association analysis example. Through associations between the Time class and Party/Organization/Meeting classes, indirect connections between Person class and Time class can be obtained. The session's holding time reveals Person 1's indirect connection to Time 1, enabling queries about relevant facts at specified time points and revealing a person' s entire life historical activities.

3.3 Realizing Reconstruction of National History Facts

The most typical application of national history knowledge is showing historical 脉络. Unlike original text entries' period metadata, time can be calculated precisely to specific dates after ontology processing, making revealed knowledge more accurate. By mining relationships between knowledge objects, important facts about specific time periods or points can be obtained, enabling applications such as chronicles and biographies.

For example, [Figure 7: see original paper] shows historical activities in December 1949. Through time-meeting associations, the first meeting of the Central People' s Government Committee on December 2, 1949 can be found, revealing historical activities of the government committee at this time point. Similarly, a person' s entire life historical activities can be revealed.

4 Conclusion

To achieve effective organization and utilization of national history knowledge, this paper proposes the "Mining down, Organizing up" approach. By combining text mining and expert judgment, important knowledge objects and facts are extracted from national history information resource text entries. Using national history ontology for semantic organization, a national history ontology for knowledge discovery and utilization was built, achieving organic organization of national history information resources and knowledge enrichment. Based on this, new applications including national history knowledge map revelation, knowledge object association analysis, and national history fact reconstruction were implemented.

This method can effectively realize fine-grained revelation of national history knowledge content and innovative organization applications based on knowledge objects, serving as a new knowledge revelation and organization method applicable to other fields. However, this research has limitations: (1) The accuracy of identifying national history knowledge objects and related facts from text needs improvement, especially in identifying and recommending relevant facts—improved accuracy would further reduce experts' workload; (2) The association calculation method for knowledge networks is simple and has not fully utilized current semantic similarity calculation and graph mining methods. These are key problems requiring future research.

References

[1] DONG Hui, YU Chuanming, YANG Ning, et al. Research on the ontology-based retrieval model of digital library—history domain ontology building[J]. Journal of the China Society for Scientific and Technical Information, 2006, 25(5): 564-474.

[2] WU Lijie. Research on knowledge organization of characterized database based on ontology[J]. Journal of Library Science, 2012(3): 41-43.

[3] PENG Weimin, SONG Jihua. Research on Zizhi Tongjian historical ontology construction and application[J]. Journal of Chinese Information Processing, 2010(2): 33-38.

[4] Research on domain ontology constructing and reasoning of the three kingdoms[D]. Central China Normal University, 2011.

[5] DONG Hui, XU Lei, WANG Fei, et al. Study on semantic analysis system (III)—implementation of Chinese historical records semantic analysis system[J]. Journal of the China Society for Scientific and Technical Information, 2014, 33(2): 204-214.

[6] Hyvönen E, Alm O, Kuittinen H. Using an ontology of historical events in semantic portals for cultural heritage[C]//Proceedings of the Cultural Heritage on the Semantic Web Workshop at the 6th International Semantic Web Conference (ISWC 2007), 2007.

[7] Corda I, Bennett B, Dimitrova V. A logical model of an event ontology for exploring connections in historical domains[C]//Workshop on Detection, Representation and Exploitation of Events in Semantic Web (DeRIVE 2011), Proceedings of the Tenth International Semantic Web Conference (ISWC' 11), 2011.

[8] Ide N, Woolner D. Historical ontologies[M]//Words and intelligence II: essays in honor of Yorick Wilks. Springer, 2007: 137-152.

[9] Uschold M, King M. Towards a methodology for building ontologies[C]//Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI-95, Montreal, Canada, 1995.

[10] Noy N F, McGuinness D L. Ontology development 101: a guide to creating your first ontology[EB/OL]. [2015-03-25]. http://wenku.baidu.com/view/30fb4b9956bec0975f465e2bf.html.

[11] SUN Hui, LEI Feng. Research on the contemporary Chinese history ontology building[J]. Journal of Modern Information, 2014, 34(2): 32-42.

[12] WANG Hao, GU Jun, SU Xinning. Research on the model and its application of ontology-driven knowledge management system[J]. Journal of Library Science in China, 2013, 39(2): 98-110.

Author Affiliations:
WANG Ying: Librarian, National Science Library, Chinese Academy of Sciences, Beijing
ZHANG Zhixiong: Research Librarian, PhD Supervisor, National Science Library, Chinese Academy of Sciences, Beijing
SUN Hui: Associate Editor, Institute of Contemporary China Studies, Beijing
LEI Feng: Senior Engineer, Institute of Contemporary China Studies, Beijing

Received: 2015-04-14

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