Preliminary Study on Semantic Retrieval in the Mongolian Music Domain: Postprint

Authors: Bao Yulai, Bi Qiang

Date: 2017-11-08T00:00:00+00:00

Abstract

【Objective】 Against the backdrop of rapid growth in Mongolian information resources, this study explores the application of ontology-based semantic retrieval to the Mongolian domain to enhance the effectiveness of Mongolian information resource retrieval. 【Method】 By leveraging the reasoning and interconnection advantages of ontology technology and utilizing the rule inference engine of the semantic parsing and reasoning tool Jena, we designed and implemented a Mongolian semantic retrieval system based on a Mongolian music domain ontology. 【Results】 Compared with keyword matching retrieval, the semantic retrieval system achieved a recall rate of 95.6% and a precision rate of 3.2%, significantly higher than keyword matching retrieval. 【Limitations】 The study focuses solely on Mongolian polyphonic music as the research object, with limited experimental data, which imposes certain constraints on retrieval. 【Conclusion】 This research presents a comprehensive study on ontology-based semantic retrieval in the Mongolian domain, establishing a solid theoretical and technical foundation for Mongolian Semantic Web application research.

Full Text

A Preliminary Study on Semantic Retrieval in the Mongolian Music Domain

Bao Yulai[1,2], *Bi Qiang*[2]
[1](Inner Mongolia University Library, Hohhot 010021, China)
[2](School of Management, Jilin University, Changchun 130022, China)

Abstract

[**Objective**] This paper aims to improve the retrieval performance of the rapidly growing Mongolian information resources by exploring the application of ontology-based semantic retrieval in the Mongolian domain. [**Methods**]

Leveraging the advantages of ontology technology in reasoning and interconnection, we designed and implemented a Mongolian semantic retrieval system based on a Mongolian music domain ontology, utilizing the rule inference engine of the semantic parsing and reasoning tool Jena. [**Results**] Compared to keyword matching retrieval, the semantic retrieval system achieved a recall rate of 95.6% and a precision rate of 93.2%, significantly outperforming keyword-based approaches. [**Limitations**] The study focused solely on Mongolian polyphonic music with limited experimental data, which constrains the generalizability of the retrieval system. [**Conclusions**] This research presents a comprehensive study on ontology-based semantic retrieval for Mongolian domains, establishing a solid theoretical and technical foundation for Mongolian semantic web applications.

Keywords: Mongolian Domain Ontology, Semantic Reasoning, Semantic Retrieval
Classification Number: TP182 G353

Introduction

The evolution of the Internet exhibits two prominent trends: the semanticization of information resources and the multilingualism of users and information resources [1]. In practice, network information resources in languages such as English, Chinese, and Russian are growing exponentially. Concurrently, with the advancement of Chinese minority language documentation and information resource construction, web resources in various minority languages are also increasing rapidly. In particular, after decades of dedicated efforts, Traditional Mongolian (hereinafter referred to as Mongolian) information resource construction has achieved remarkable success, evolving from nonexistence to substantial repositories such as the Inner Mongolia Mongolian Newspaper and Website Alliance Platform, the Yeliba Mongolian Literature Management System [2], the Mongolian Journals Network [3], and the China Mongolian Studies Information Network [4]. Additionally, the Inner Mongolia University Library participated in the "Sino-American Million Book Digital Library Cooperation Project," digitizing over 6,000 modern Mongolian books and establishing the "Mongolian Full-text Resource Database" [5]. Against this backdrop, the critical challenge facing Mongolian information resource digitization is how to efficiently access valuable Mongolian information resources and provide more effective retrieval services for Mongolian users.

Traditional Mongolian belongs to the Altaic language family and is predominantly used in China' s Inner Mongolia Autonomous Region, Liaoning, Heilongjiang, Jilin, Gansu, Qinghai, and Xinjiang Uygur Autonomous Region, representing one of the most widely used, socially influential, and internationally recognized minority languages in China [6]. The Mongolian script comprises 34 letters, including 7 vowels and 27 consonants. Mongolian writing is vertical at the word level, with letters within a word connected continuously and words separated by spaces. The overall writing rule proceeds from top to bottom within a

line and from left to right across lines. Each Mongolian letter has different variants for initial, medial, and final positions; some variants have multiple forms, resulting in inconsistent written forms and phenomena of one character with multiple shapes, multiple characters with the same shape, and same shape with different pronunciations. Moreover, Mongolian features an extremely complex and rigorous system of orthographic and grammatical rules [7]. These characteristics render alphabetical pronunciation-based retrieval impossible, earning Mongolian the reputation as one of the most difficult scripts for information retrieval.

With the emergence of the Semantic Web, ontology has become a new research direction and hotspot. Domestic researchers have primarily focused on ontology construction and semantic retrieval, developing several systems with reasoning capabilities. Li et al. [8] proposed a compositional reasoning method based on description logic. Huang et al. [9] utilized semantic similarity computation to extract hidden information from rough ontologies, achieving associated information mining. Meng et al. [10] improved literature retrieval accuracy through semantic annotation and similarity computation of important concepts in the literature domain ontology. Zhou et al. [11] designed a patent semantic retrieval system based on Jena reasoning and the Lucene framework, enhancing patent retrieval efficiency. Wu [12] analyzed the limitations of traditional information retrieval methods and constructed a retrieval system for computer science journal literature by leveraging ontology structural characteristics and semantic annotation techniques. Li [13] developed a semantic retrieval system for the patent domain, improving precision through semantic reasoning. In Mongolian semantic web research, Su et al. [14] established a small-scale Mongolian place name ontology for Inner Mongolia to facilitate geographic information semantic retrieval. Additionally, Su et al. [15] constructed a Mongolian domain ontology for the "Artificial Intelligence ( )"curriculum and developed a learning system for Mongolian education by computing semantic similarity of retrieved knowledge points. Ta et al. [16] proposed a preliminary design method for a Mongolian-Chinese bilingual computer terminology semantic dictionary using domain ontology construction for cross-language information retrieval. Overall, Mongolian semantic web research remains relatively underdeveloped compared to Chinese semantic web research.

Therefore, this paper introduces Semantic Web technology, selecting representative Mongolian polyphonic music forms—Choor ( ) and Morin Khuur ( )—as the ontology research objects. The study focuses on developing reasoning rules applicable to the Mongolian music domain ontology, constructing and implementing a prototype semantic retrieval system for Mongolian music to lay a foundation for Mongolian semantic web applications and enrich the research scope and content of the semantic web.

2. Mongolian Music Domain Ontology Construction

Currently mature ontology construction methods include TOVE [17], IDEF-5 [18], Skeleton [19], and the Seven-Step method [20]. This study adapts the main procedures of the Seven-Step and Skeleton methods, incorporating characteristics of Methontology and IDEF-5, to establish a construction approach and workflow tailored to the Mongolian music domain, as illustrated in [Figure 1: see original paper]. Ontology research lacks support for Chinese, particularly minority languages. This investigation of Mongolian music domain ontology construction methods validates the feasibility of building Mongolian domain ontologies, facilitating the organization of Mongolian digital resources and enriching research in Mongolian information processing and the semantic web.

The research objects are representative forms of traditional Mongolian music: Choor and Morin Khuur. While the shareable nature of ontologies enables data reuse, no reusable ontology data currently exists for the Mongolian music domain. We collected terminology and conceptual knowledge about Choor and Morin Khuur through domain experts and relevant literature, establishing relationships among core concepts. The ontology model was constructed using Protégé, with the Mongolian music domain ontology validated against construction principles including conceptual clarity, consistency, completeness, and extensibility. During the evolution phase, leveraging the extensibility and flexibility of ontologies—particularly important given Mongolian's phenomena of one character with multiple shapes and one word with multiple meanings—some conceptual relationships may initially remain ambiguous. The ontology can undergo continuous maintenance and refinement to address these challenges.

2.1 Acquisition of Key Mongolian Music Terms and Concepts

We selected academic papers on Mongolian music, Choor, and Morin Khuur from the Mongolian Journals Network and CNKI, preprocessing the data using a Mongolian word segmentation system based on cascaded hidden Markov models [21] to extract collections of Mongolian music terminology and concepts. Following the principle of maximal representativeness in the Mongolian music domain, key terminological concepts were selected as follows:

- Mongolian Polyphonic Music ( )
- Choor ( )
- Modon Choor ( )
- Tobshuur ( )
- Khoomii Choor ( )
- Choor Duu ( )
- Khuur ( )
- Polyphony ( )
- Sustained Bass Part ( )
- Sustained Bass Part as Instrumental Performance ( )

- Melodic Part ( )
- Melodic Part as Instrumental Performance ( )
- Melodic Part as Long Song ( )
- Choor Instruments ( )
- Yikile ( )
- Khailasen Choor ( )
- Choor Music Cultural Context ( )
- Geographic Environment ( )
- Living Context ( )
- Choor Music Origin and Development ( )
- Early Choor Music Phenomenon ( )
- Natural Sound ( )
- Performance Form ( )
- Solo Performance ( )
- Group Performance ( )
- Musical Structure ( )
- Voice + Voice (( ) + ( ))
- Voice + Instrument (( ) +)
- Instrument + Instrument (+)
- Development and Distribution ( )
- Xinjiang Mongolian Modon Choor ( )
- Khorchin Mongolian Traditional Instrument Khailasen ( )
- Xilingol Region Popular Choor Duu ( )
- Morin Khuur ( )
- Three Tuning Methods ( )
- Perfect Fourth Tuning ( )
- Perfect Fifth Tuning ( )
- Inverse Fourth Tuning ( )
- Five Performance Techniques ( )
- Choor Tuning Khuur Performance Technique ( )
- Choor Tuning or Khili Tuning Oirat Performance Technique ( )
- Chakhar Tuning Harmonic Performance Technique ( )
- Borjigin Tuning Harmonic Performance Technique ( )
- Borjigin Tuning Tobshuur Performance Technique ( )
- Distribution ( )
- Inner Mongolia ( )
- Liaoning ( )
- Jilin ( )
- Heilongjiang ( )

- Gansu ( )
- Xinjiang ( )
- Representative Figures ( )
- Balgan ( )
- Burenbayar ( )
- Sanduuren ( )
- Suresh ( )
- Classic Works ( )
- Taibus Banner Asar ( )
- Demon Subjugation ( )
- Beautiful and Prosperous Alxa ( )
- Serun Saihan Hanggai ( )
- Walking Horse ( )
- Spring in Ordos ( )

2.2 Establishing Class Hierarchical Relationships in the Mongolian Music Ontology

Classes describe abstract entity objects, representing collections of instances with shared characteristics. Classes exhibit inheritance and organize in hierarchical structures, where top-level classes represent the most abstract concepts, and subclasses inherit parental characteristics while representing more specific or narrowly scoped entities. Through literature review, this study employs a bottom-up approach to clarify the meaning of each concept in the concept set and analyze both explicit and implicit relationships among concepts. Core terminological concepts in the Mongolian polyphonic music domain of Choor and Morin Khuur are established based on Mongolian music fundamentals, with their subordinate relationships reflecting conceptual hierarchical structures, as shown in [Figure 2: see original paper] and [Figure 3: see original paper].

2.3 Defining Classes and Properties in the Mongolian Music Ontology

The ontology representation process fundamentally involves defining classes and properties. In ontology models, properties represent relationships between different concepts (classes) and reveal distinctions among conceptual classes. Protégé primarily includes object properties and data properties. Object properties clarify relationships between different conceptual classes, while data properties describe the data attributes of conceptual classes. Defining properties distinguishes conceptual classes, avoids redundant class definitions, and facilitates description and generalization to gradually form knowledge hierarchies.

Key concept sets and their relationships include:

- Mongolian Polyphonic Music: {Choor, Morin Khuur}
- Choor: {Modon Choor, Tobshuur, Khoomii Choor, Choor Duu, Khuur}
- Polyphony: {Sustained Bass Part, Sustained Bass Part as Instrumental Performance}

- Melodic Part: {Vocal Melodic Part, Melodic Part as Instrumental Performance, Melodic Part as Long Song}
- Choor Instruments: {Yikile, Khailasen Choor}
- Choor Music Cultural Context: {Geographic Environment, Living Context}
- Choor Music Origin and Development: {Early Choor Music Phenomenon, Natural Sound}
- Performance Form: {Solo Performance, Group Performance}
- Musical Structure: {Voice + Voice, Voice + Instrument, Instrument + Instrument}
- Development and Distribution: {Xinjiang Mongolian Modon Choor, Khorchin Mongolian Traditional Instrument Khailasen, Xilingol Region Popular Choor Duu}
- Morin Khuur Three Tuning Methods: {Perfect Fourth Tuning, Perfect Fifth Tuning, Inverse Fourth Tuning}
- Morin Khuur Five Performance Techniques: {Choor Tuning Khuur Technique, Choor or Khili Tuning Oirat Technique, Chakhar Tuning Harmonic Technique, Borjigin Tuning Harmonic Technique, Borjigin Tuning Tobshuur Technique}
- Distribution: {Inner Mongolia, Liaoning, Jilin, Heilongjiang, Gansu, Xinjiang}
- Representative Figures: {Balgan, Burenbayar, Sanduuren, Suresh}
- Classic Works: {Taibus Banner Asar, Demon Subjugation, Beautiful and Prosperous Alxa, Serun Saihan Hanggai, Walking Horse, Spring in Ordos}

To better describe relationships among ontology concepts, new object properties were added: **Depend** (依赖), **Depended** (被依赖), **Use** (使用), and **Used** (被使用).

2.4 Adding Instances

Instances (individuals) of classes were added through Protégé' s Individuals panel, with attribute values assigned to each instance to complete the Mongolian music domain ontology construction, as shown in [Figure 4: see original paper].

2.5 Semantic Reasoning for the Mongolian Music Ontology

This study employs the open-source Java framework Jena [22] as the reasoning tool to implement semantic inference of conceptual knowledge within the Mongolian music domain. The Jena framework includes two rule inference engines—RETE and Tabled Datalog—that can implement three types of reasoning models. This research adopts rule-based reasoning for semantic inference, which involves formalizing domain expert knowledge into systematic rules. These rules represent domain-specific problems and their corresponding solutions, enabling emulation of expert associative reasoning capabilities. Rule formulation must closely follow the practical application of the ontology; since the primary research domain is Mongolian music, reasoning rules were developed based on the

Mongolian music domain ontology and retrieval requirements.

The reasoning rules established in this study are:

1. **Inheritance Reasoning**: The hierarchical structure among concepts in the ontology model shares similarities with class inheritance. In ontology concepts, class inheritance manifests when subclass concepts inherit parent class characteristics while possessing unique attributes. The rule is formalized as:

   ```
   Rule1: (?a, rdfs:subClassOf, ?b), (?s, rdfs:type, ?a) -> (?s, rdfs:type, ?b)
   ```

   If a is a subclass of b and s is an instance or property of a, then s is also an instance or property of b.

2. **Equivalent Class Reasoning**: Ontologies may contain two concept classes with identical property values but different names, sometimes representing the same knowledge point:

   ```
   Rule2: (?a, owl:equivalentClass, ?b), (?s, rdf:type, ?a) -> (?s, rdf:type, ?b)
   Rule3: (?a, owl:equivalentClass, ?b), (?s, rdf:type, ?b) -> (?s, rdf:type, ?a)
   ```

   Concept classes a and b with different names but representing the same knowledge point share identical property values.

3. **Transitive Subclass Reasoning**: For ontological classes with containment relationships:

   ```
   Rule4: (?a, rdfs:subClassOf, ?b), (?b, rdfs:subClassOf, ?c) -> (?a, rdfs:subClassOf, ?c
   ```

   Concept classes a, b, and c form an object-oriented inheritance relationship, where a inherits properties from c while containing attributes unique to c.

4. **Parallel Relationship Reasoning**: Musical concepts exhibit interdependent and parallel relationships:

   ```
   Rule5: (?f depended ?a), (?f depended ?b) -> (?a parallel ?b)
   ```

   If concept classes a and b both depend on f, they are parallel concepts.

5. **Association Relationship Reasoning**: Different musical concepts exhibit correlations during usage:

   ```
   Rule6: (?a used ?b), notEqual(?a, ?b) -> (?a reference ?b)
   ```

   When retrieving concept a, information about concept b is utilized, and a and b represent different concepts, establishing an associative relationship between them.

3. Implementation of the Mongolian Music Semantic Retrieval System

This system is a semantic retrieval system based on the Mongolian music domain ontology. It employs ontology reasoning technology to semantically expand query keywords and obtains hidden associative information from ontology data through rule-based semantic reasoning, thereby implementing a prototype semantic retrieval system for the Mongolian music domain.

3.1 System Architecture

Based on functional requirements analysis and referencing mainstream web search engine architectures, we designed a semantic retrieval system model that integrates semantic web and ontology reasoning technologies, with Mongolian music domain concepts as retrieval objects. The system adopts a Browser/Server (B/S) structure comprising three layers: browser client, WebAPP server (Tomcat), and database (storing domain ontology files). The technical implementation route uses the RichFaces framework for system-user interaction, including query interfaces and result feedback. A JavaBean reasoning middleware designed with the Jena framework performs semantic parsing and query scope expansion on keywords (search conditions). The reasoning engine and rules execute the inference process, invoking ontology data to match with the ontology model and outputting search results. The Mongolian music domain ontology was constructed using Protégé, with Jena enabling persistent storage and customized reasoning rules for the Mongolian music domain. The overall system architecture is illustrated in [Figure 5: see original paper].

3.2 Semantic Retrieval Process

The semantic retrieval workflow operates as follows: (1) The system captures user-input query keywords through the search interface, constraining the search scope to concepts, properties, or instances in the ontology knowledge base to clarify retrieval requirements. (2) Upon submission of the search request, the system performs semantic expansion on the input terms, generating a set of semantically similar or related query terms based on reasoning rules to broaden the search scope at the semantic level. (3) Original information and inferred implicit information are combined to form a complete post-reasoning database. (4) User queries are matched against this inferred database to retrieve relevant information. (5) Final results are presented to users in a clear, organized, and easily identifiable format. The retrieval process is depicted in [Figure 6: see original paper].

3.3 System Validation

Without semantic reasoning enabled, searching for the concept "Mongolian Polyphonic Music" returned results including representative concepts such as "Choor," "Choor Performance Forms," and "Musical Structure," as shown in

[Figure 7: see original paper]. With semantic reasoning activated, the same query yielded not only "Choor"and related concepts but also specific Choor types such as "Khoomii Choor" and "Modon Choor," along with detailed performance attributes like "Solo Performance" and "Voice + Instrument," as illustrated in [Figure 8: see original paper].

Testing on 150 Mongolian music domain documents demonstrated that the semantic retrieval system achieved a recall rate of 95.6% and precision rate of 93.2%, substantially surpassing keyword matching retrieval (75.2% recall, 73.6% precision).

References

[1] Zhang Chengzhi. *Domain Ontology Learning Research with Multi Language* [M]. Nanjing: Nanjing University Press, 2012: 1-3.
[2] Mongolian Literature Management System [EB/OL]. [2016-03-31]. http://www.surag.net/?p=8116.
[3] Mongolian Journals Network [EB/OL]. [2016-03-31]. http://journal.surag.net/.
[4] China Mongolia Information Network [EB/OL]. [2016-03-31]. http://mgl.surag.net/.
[5] Wang Fu, Kang Jian. Study on the Construction and Sharing of the System of Mongolian Digital Resources [J]. *Library and Information Service*, 2015, 59(12): 39-43.
[6] Di Dongmei. Design of Institutional Repository Traditional Mongolian Literature with DSpace [J]. *Computer Engineering and Science*, 2010, 32(10): 139-141, 149.
[7] Ao Qier. Mongolian International Standard Code [C]. In: *Proceedings of the 2011 International Symposium on Technology Standard and Distance Education*, 2001.
[8] Li Guanyu, Ma Di, Loua V. Fuzzy Ontology Based Knowledge Reasoning Framework Design [C]. In: *Proceedings of the 3rd International Conference on Software Engineering and Service Science*, 2012.
[9] Huang Y H, Li G Y, Li Q Q. Rough Ontology Based Semantic Information Retrieval [C]. In: *Proceedings of the 6th International Symposium on Computational Intelligence and Design*, 2013.
[10] Meng Hongwei, Zhang Zhiping, Zhang Xiaodan. Research on Intelligent Information Retrieval Model Based on Domain Ontology [J]. *Journal of Information*, 2013, 32(9): 180-184.
[11] Zhou Qunfang, Wu Jie, Gu Jun. Study on Ontology-based Patent Semantic Retrieval [J]. *Information Research*, 2013(9): 71-74.
[12] Wu Junli. Research and Implementation of Semantic Retrieval System Based on Domain Ontology [D]. Taiyuan: Taiyuan University of Technology, 2010.
[13] Li Bing. Research on Patent Semantic Retrieval Based on Domain Ontology [D]. Beijing: Beijing Institute of Technology, 2015.
[14] Su Yila, Sun Riwang, Tan Yanmei, et al. Study of Mongolian Place Name Recognition Based on Ontology Annotation [J]. *Computer Engineering and*

Science, 2013, 35(8): 156-162.
[15] Su Yila, Ji Yatu, Dou Baoyuan. A Semantic Web-based Reasoning and Retrieval Method in Mongolian Curriculum [J]. *Computer Engineering and Science*, 2016, 38(2): 376-384.
[16] Ta Na, Lin Min, Li Xiaoqing. Preliminary Study on Building Mongolian and Chinese Bilingual Semantic Dictionary for Cross-Language Information Retrieval [J]. *Computer and Digital Engineering*, 2010, 32(8): 42-45, 54.
[17] Ma J, Zhang Q, Zhang J, et al. A Rule-based Ontology Reasoning for Scientific Effects Retrieval [C]. In: *Proceedings of the 2012 IEEE International Conference on Management of Innovation and Technology*, 2012: 232-237.
[18] Delbru R, Campinas S, Tummarello G. Searching Web Data: An Entity Retrieval and High-performance Indexing Model [J]. *Journal of Web Semantics*, 2012, 10(1): 33-58.
[19] Gao Hongmei. Research and Implementation of Semantic Search Technology Based on Ontology in Literature Domain [D]. Shanghai: East China University of Science and Technology, 2014.
[20] Dewan H. REDIS: The Data Structure Server for Your Cloud [J]. *PC Quest*, 2011.
[21] Cong Wei. Study of Mongolian Word Segmentation System Based on the Cascaded Hidden Markov Model [D]. Hohhot: Inner Mongolia University, 2009.
[22] Wang Haichuan. Application Research of Service Semantic Retrieval and Combination Method in Cloud Computing Environment [D]. Ji' nan: Shandong University of Finance and Economics, 2014.

Author Contributions

Bao Yulai: Conceptualized the research, designed the study, collected data, constructed the ontology, built the system, and drafted the manuscript.
Bi Qiang: Designed and implemented the research framework and revised the manuscript.

Conflict of Interest Statement

All authors declare no conflict of interest.

Supporting Data

Supporting data is self-archived by the authors, E-mail: 65003846@qq.com.
[1] Bao Yulai, Bi Qiang. mongolian music.owl. Mongolian music ontology data.

Manuscript Dates

Received: June 27, 2016
Revised: October 17, 2016

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