

Knowledge Element Ontology Model for Knowledge Representation of Scientific Literature (Postprint)

Authors: Qin Chunxiu, Zhijuan Yang, Zhao Pengwei, Liu Jie

Date: 2023-08-26T00:00:00+00:00

Abstract

[Purpose/Significance] With the exponential growth of scientific literature resources, users are overwhelmed by vast amounts of information. Providing rapid and precise fine-grained knowledge element services will become the future trend in literature knowledge retrieval.

[Method/Process] Based on analyzing the textual structure of scientific literature, this study progressively explores the content to construct a knowledge element ontology model for scientific literature knowledge representation, thereby transforming fine-grained knowledge points with complete semantic meaning into uniformly structured knowledge elements.

[Result/Conclusion] Using a scientific literature as a case study, this paper demonstrates the proposed knowledge representation method for scientific literature based on the knowledge element ontology model. However, the example only presents knowledge points from the introduction section, and further validation of the model's effectiveness is required.

Full Text

Preamble

Vol. 62 No. 3, February 2018

ChinaXiv Cooperative Journal

A Knowledge Element Ontology Model for Scientific Literature Knowledge Representation

Qin Chunxiu, Yang Zhijuan, Zhao Pengwei, Liu Jie

School of Economics and Management, Xidian University, Xi'an 710071

Abstract

[Purpose/Significance] With the explosive growth of scientific literature resources, users are drowning in a sea of documents. Providing fast and accurate fine-grained knowledge element services will become the future trend of literature knowledge retrieval. **[Method/Process]** Based on analyzing the text structure of scientific literature, this study delves progressively into the content of scientific literature to construct a knowledge element ontology model for scientific literature knowledge representation. This model represents fine-grained knowledge points with complete sentence meanings in scientific literature content as knowledge elements with a unified structure. **[Result/Conclusion]** Using a scientific literature document as an example, this paper demonstrates the proposed knowledge representation method based on the knowledge element ontology model. However, the example only presents relevant knowledge points from the introduction section of the scientific literature, and the effectiveness of the knowledge element ontology model requires further verification.

Keywords: scientific literature; knowledge representation; knowledge element; ontology model

Classification Number: G302

DOI: 10.13266/j.issn.0252-3116.2018.03.012

Scientific literature represents the crystallization of scientific exploration by researchers and contains rich specialized scientific achievements and important research discoveries, making it highly valuable for reference and application. The sharing and exchange of scientific literature knowledge resources not only facilitate the discovery of potential value in scientific literature but also promote knowledge learning, innovation, and discovery, thereby advancing scientific development. In the era of big data, scientific literature is not only massive in volume but also growing explosively at an average annual rate of 4.7% across academic journals in various fields, leaving researchers overwhelmed by the vast ocean of scientific literature [1].

When acquiring needed scientific literature knowledge resources, people currently rely on traditional information retrieval technologies to obtain document-level scientific literature, browsing through knowledge points contained in each document and manually screening for required information. This approach is extremely time-consuming and inefficient. The root cause of such inefficient literature knowledge acquisition lies in the fact that current retrieval tools typically describe, represent, and organize literature at the coarse-grained document level rather than refining the knowledge control unit to the knowledge point level. Consequently, information retrieval systems cannot quickly and accurately match users' actual knowledge needs, resulting in a situation of "resource abundance yet knowledge scarcity."

To address these problems, scholars have conducted in-depth research from two perspectives: the characteristics of knowledge points and the knowledge structure of scientific literature. (1) Starting from the characteristics of certain knowl-

edge points, they construct extraction rules for that type of knowledge point and use rule-matching algorithms to extract single types of knowledge points such as definitions [2], methods [3], or innovation points [4] from scientific literature text content. However, these knowledge points are generally scattered across various chapter sections of the literature, making it difficult to comprehensively locate all relevant knowledge points and reducing the efficiency and accuracy of knowledge point extraction. (2) To more fully mine knowledge resources in scientific literature, attempts have been made to analyze the knowledge structure of certain regions or modules in scientific literature and provide a resource description framework [5] for mining knowledge point sets in the main text area that carry complete scientific research ideas, such as research domain, background, problems, theories, methods, and evaluation. However, this method lacks description of the external characteristics of scientific literature and cannot comprehensively describe the various knowledge points and knowledge structures contained in scientific literature content, nor can it uniformly represent the internal attribute structures of these knowledge points, which is 不利于 knowledge reasoning and discovery.

Therefore, this study attempts to start from the basic link of literature knowledge organization—the knowledge description and representation stage—by systematically analyzing the text structure of scientific literature and using knowledge element ontology theory to research fine-grained knowledge representation methods for scientific literature. The aim is to provide a comprehensive and standardized scientific literature knowledge representation method from a fine-grained perspective, which can uniformly describe the knowledge structure of scientific literature and the academic knowledge points contained within, providing theoretical support for achieving fine-grained and precise knowledge point services.

1. Research Status of Literature Knowledge Representation

Literature knowledge representation is the foundation of literature knowledge organization, retrieval, and application, and its completeness and applicability affect the level of literature knowledge services. Currently, according to whether researchers describe the internal structural relationships of knowledge objects, literature knowledge representation methods are divided into two categories: feature-based knowledge representation methods [6] and structured knowledge representation methods [7].

Feature-based knowledge representation methods treat knowledge objects as independent individuals, typically using “attribute-value” pairs, n-tuples, feature vectors, and other forms to clearly describe the attribute characteristics of knowledge objects, facilitating their identification and indexing. Representative methods in this category include: (1) The nanopublication model, which uses simple RDF triples to generate named graphs describing the source, background, and contextual information of scientific conclusions in literature [8], represent-

ing scientific conclusions in literature as minimum publication information units with a unified structure but ignoring the potential scientific laws among these information units. (2) The knowledge element representation method, which uses n-tuples to describe the ID, source, classification, content, and other features of knowledge points in journal literature [9], representing fine-grained knowledge points in literature as knowledge elements with relatively complete content and consistent structure. These knowledge elements are knowledge units with relatively complete semantics in certain contexts, having diverse forms and relatively independent and complete content [10]. When the content of a knowledge element increases to a certain extent, it can be decomposed into smaller-grained knowledge elements. Therefore, “knowledge elements should have functions such as small content granularity, linkability into knowledge networks, evolutionary development, reconfigurability, and support for semantic retrieval” [11]. However, when describing a specific argumentative process in scientific literature, this knowledge element representation method can only describe relatively independent knowledge elements such as arguments, evidence, argumentation methods, and conclusions at a fine-grained level, but cannot clearly represent the semantically complete argumentative relationship chain constructed by the literature based on these knowledge element sets.

To describe the complex semantic relationships among knowledge objects within literature in detail, scholars have proposed structured knowledge representation methods, such as graph-based representation [6], object-oriented representation [12], production framework representation [13], and ontology representation [14]. (1) Graph-based representation methods represent knowledge or concept nodes that cannot be further divided in literature as nodes, with semantic edges between nodes jointly building a graph structure to display semantic associations among knowledge points within literature. Representative methods include knowledge networks [7], semantic networks [15], concept maps [16], and other knowledge representation methods. These methods have strong knowledge expressiveness in representing semantic relationships among literature knowledge points but cannot comprehensively describe the internal features of literature knowledge or concept nodes. Moreover, as the number of knowledge or concept nodes grows exponentially, the difficulty and cost of retrieving literature knowledge points also increase significantly. (2) Object-oriented representation methods encapsulate the attributes and methods of knowledge objects into structured modules, achieving inheritance and evolution among objects for describing basic concepts in literature content but lacking dynamic representation and reasoning for semantic relationships beyond inheritance among knowledge objects. (3) Production framework representation methods use slot-filling frameworks to describe the static knowledge structure of knowledge objects within literature and production rules to represent reasoning rules between knowledge objects, but they are not suitable for complex reasoning relationships. For example, when using production framework representation to describe the static knowledge structure and phonetic-semantic relationships in *Shuowen Jiezi*, complex reasoning relationships among knowledge objects lead to reduced derivation ef-

iciency. (4) Ontology representation methods have powerful logical reasoning capabilities, using concepts, attributes, relationships, instances, and other forms to describe the semantic logical relationships among knowledge objects in literature resources across various subject domains, enabling dynamic reasoning among knowledge objects, such as the “Jianghai Culture” knowledge ontology [14] and patent ontology [17]. However, literature has many subject classifications, and ontologies representing different subjects lack a unified structure, which is 不利于 knowledge representation, sharing, and exchange.

Among the above methods, the knowledge element representation method in feature-based knowledge representation can describe the attribute characteristics of knowledge points in scientific literature in detail but cannot comprehensively represent the systematic semantic relationships among knowledge points. Among structured knowledge representation methods used to sort out complex semantic relationships among knowledge points, ontology representation can standardize and describe the systematic semantic logical relationships among knowledge points in scientific literature across different subject domains but is difficult to break through domain limitations to uniformly represent the structure of knowledge points within scientific literature.

Knowledge element ontology construction defines the organizational skeleton model of knowledge elements and establishes the connection between knowledge elements and ontology elements. It comprehensively describes the internal structure and semantic associations of semantic content in knowledge objects in the form of concept-attribute-method-relationship quadruples, revealing the attribute characteristics of knowledge elements and the semantic logical relationships among knowledge elements. It represents a standardized abstraction and description of domain knowledge, facilitating semantic reasoning for knowledge and enabling knowledge organization and retrieval [18-19]. Knowledge element ontology has been widely applied in scenarios such as wiki knowledge element annotation [18], semantic indexing of disciplinary knowledge [11], classification and combination linking of knowledge elements in “Yue Customs” literature [20], logical reasoning among knowledge elements in diet and disease domains [21], and knowledge discovery based on literature topic causation [22]. Therefore, based on knowledge element ontology theory and incorporating the ideas of knowledge elements and ontology representation methods, this study proposes a knowledge element ontology model for scientific literature knowledge representation to solve the inefficiency of knowledge acquisition caused by coarse-grained knowledge representation methods for scientific literature.

2. Scientific Literature Text Structure Analysis

A scientific literature document typically has both physical and logical structures. The physical structure presents the subjective cognitive structure of the literature, including title, author, affiliation, chapters, sections, paragraphs, sentences, words, and citations. On this basis, hierarchical division of the text is performed to mine the themes of content at different levels, obtaining the log-

ical structure of the text, including document-level theme, hierarchical theme, paragraph theme, sentence theme, theme words, identifier words, and classification numbers, aiming to represent the knowledge structure contained in the text content [23]. Using only a single thematic classification method to describe the knowledge structure within literature ignores the simple, intuitive, and unambiguous non-thematic literature features, making it impossible to accurately exclude a large number of irrelevant documents during literature retrieval [24]. Therefore, integrating the above literature features, the features of scientific literature are uniformly divided into external features and content features. External features generally include symbolic identifiers, title, author, publication time, publisher, journal, conference, and document type [25], while content features include theme words, classification numbers, etc. [25], facilitating clear, effective, and comprehensive description of the characteristics of fine-grained knowledge points in scientific literature.

In practical applications, research literature itself has a specific text structure composed of title, author, abstract, keywords, and main text [26]. Delving deeper into the main text content reveals a universal structure: introduction, method, result, and discussion [27]. This universal structure has certain sectional semantic division functions in the expression of scientific literature content. Y. Ding et al. used literature analysis methods to refine this universal structure, forming a coarse-grained functional structure composed of abstract, introduction, related research, method, experiment/result, and conclusion [28]. Wang Peng et al., following a hierarchical segmentation text processing method [29], further subdivided the knowledge blocks in each part of the above functional structure to obtain a hierarchical functional structure composed of fine-grained knowledge points within scientific literature. In the abstract section of scientific literature, the background, conclusion, method, object, and result of scientific research are generally summarized, with the most frequent summaries being research conclusions, methods, and results [30]. In the main text of scientific literature, the introduction contains three types of information: background knowledge, problem analysis, and main work [31], while before problem analysis, it is necessary to introduce the research problem and describe the origin, motivation, and purpose of the research [32]. Methods are divided into scientific research methods and problem-solving methods. Scientific research methods include questionnaire surveys, expert interviews, case analysis methods, etc. [3], while problem-solving methods are divided into models, algorithms, and indicators according to their forms of expression, with models including framework models and mathematical models [33]. Experiments mainly discuss experimental data, experimental process, experimental results, evaluation, experimental findings, and discussion, where the experimental process includes system design and implementation [32]. Additionally, literature research shows that evaluation practices mainly study issues such as evaluation indicators, evaluation methods, and evaluation procedures [34]. Conclusions mainly include the elaboration of major contributions, explanations of (un)expected results, promotion of results, and future research directions [35].

Therefore, this study will, based on the above hierarchical functional structure of scientific literature, divide the knowledge resources in scientific literature into fine-grained functional knowledge points. This hierarchical functional structure meets people's usage needs for knowledge points in scientific literature, facilitating quick understanding of research background and keeping up with academic frontiers. Building a hierarchical information model of scientific literature from its text structure can provide more accurate basic information for the discovery of research hotspots and the similarity and classification comparison of research content. Therefore, this study will delve into the text content of scientific literature, mine the hierarchical functional structure within scientific literature content, and aim to build a knowledge element ontology model for representing scientific literature knowledge.

3. Scientific Literature Knowledge Element Ontology Model

Knowledge element ontology structure is typically represented as a quadruple [18], i.e., $K = (C, P, M, R)$, which describes the concept C characteristics of literature knowledge elements through the attribute set P and method set M , and describes the internal system structure of knowledge elements and the networked structure among knowledge elements through the semantic relationship set R among concepts, providing guarantee for the construction of knowledge element semantic links. Existing knowledge element ontology structures only represent the functional relationships among concepts and semantic relationships in the method set and relationship set, respectively, and fail to fully describe the functional relationships among concept attributes and the semantic relationships among concepts, attributes, and methods.

Based on systematic analysis of scientific literature text structure, this study extends and expands the above knowledge element ontology structure to implement the design of a scientific literature knowledge element ontology model. Using the hierarchical structure of the concept set in the ontology, the multi-level text structure in scientific literature is clearly described. Each concept can abstractly describe each type of knowledge element in scientific literature, and by meticulously depicting the attribute characteristics of concepts, the internal structure of various knowledge elements within scientific literature is formally represented. The method set of the ontology model is defined to describe the functional relationships existing among concepts, between concepts and attributes, and among attribute sets, clarifying the semantic rules of concepts and attributes to which knowledge elements in scientific literature belong, and standardizing the representation of the internal semantic associations among knowledge elements in scientific literature. The relationship set among concepts, attributes, and methods is defined to systematically describe the logical relationships among knowledge elements in scientific literature, accurately presenting the implicit scientific research ideas in scientific literature.

Therefore, based on existing knowledge element ontology theory and following

the above design 思路 for scientific literature knowledge element ontology, this study defines a scientific literature knowledge element ontology model and its four constituent elements—concept, attribute, method, and relationship—by analyzing the external and content features of scientific literature. The aim is to provide a scientific literature-oriented knowledge representation method for uniformly describing the internal fine-grained knowledge point features and knowledge associations within scientific literature.

3.1 Mathematical Description of the Scientific Literature Knowledge Element Ontology Model

A scientific literature document is a small knowledge base with relatively complete semantics composed of multiple knowledge elements that are relatively independent in content and semantically interrelated. To facilitate users' precise and fine-grained retrieval of required knowledge elements, this study constructs a scientific literature knowledge element ontology model from a micro perspective to uniformly describe and represent the composition of scientific literature knowledge elements and the semantic relationships among knowledge elements. This knowledge element ontology model can be formally represented as:

$$\text{KEOSL} = \{C, PC, MC,P, RC,P,M\}$$

Where KEOSL represents the scientific literature knowledge element ontology, C represents the concept set in scientific literature knowledge resources, PC represents the attribute set of concept C, MC,P represents the method set related to concept C and its attribute P, and RC,P,M represents the semantic relationship set among concept C, attribute P, and method M.

For any scientific literature knowledge element e, the structure of its concept set Ce, attribute set Pe, method set Me, and relationship set Re, as well as the relationships among them, are represented by mathematical formulas as follows:

- (1) Attribute set of concept Ce: $Pe = (PDe, POe, DP)$ $PDe = (DCe, Ve, DPD)$ $POe = (PFe, Te, DPO)$

Where DP indicates whether the concept's attributes are describable or measurable, which can be divided into DPD and DPO. In the data attribute PDe structure, $DCe = \phi$ and DCe indicates that the domain to which data attribute PDe belongs is the concept set Ce, Ve indicates the value type of data attribute PDe, and DPD indicates whether data attribute PDe is describable or measurable. In the object attribute POe structure, PFe indicates the parent attribute to which object attribute POe belongs, Te indicates the characteristics of object attribute POe, and DPO indicates whether object attribute POe is describable or measurable.

- (2) Method set related to concept Ce and its attribute Pe: $Me = (MFe, MOCGe, DM)$

Where MFe represents the formula description of the method; MOCGe repre-

sents the combination of object types involved in the method, including combinations of concepts and their attributes, and combinations of concepts; DM indicates whether the method is describable or measurable.

- (3) Relationship set among concept C_e , attribute P_e , and method M_e : $Re = (RFMe, POe, RLMe, ROCGe, DR)$

Where $RFMe$ represents the former part of the semantic association; POe represents the semantic relationship between objects, i.e., the object attribute set; $RLMe$ represents the latter part of the semantic association; $ROCGe$ represents the combination of semantic association object categories, including concept-concept combinations, attribute-attribute combinations, method-method combinations, concept-attribute combinations, concept-method combinations, and attribute-method combinations; DR indicates whether the semantic relationship is describable or measurable.

DX indicates whether data attribute DPD , object attribute DPO , method DM , or relationship DR is describable or measurable, where $DX = 0$ indicates the relationship is not describable, $DX = 1$ indicates the relationship is describable but not measurable, and $DX = 2$ indicates the relationship is both describable and measurable.

From the above scientific literature knowledge element ontology structure $KEOSL$, the constituent elements of the ontology structure of scientific literature knowledge element e can be specifically represented as concept set $C_e = \{C1, C2, C3, \dots\}$, attribute set $P_e = \{P1, P2, P3, \dots\}$, method set $M_e = \{M1, M2, M3, \dots\}$, and relationship set $Re = \{R1, R2, R3, \dots\}$. This scientific literature knowledge element ontology structure can clearly describe the knowledge element structure of scientific literature and the semantic relationships among knowledge elements.

3.2 Element Definitions of the Scientific Literature Knowledge Element Ontology Model

Based on the formal mathematical model of the scientific literature knowledge element ontology given above, this study defines the specific content of the concept set, attribute set, method set, and relationship set of scientific literature knowledge elements. Simultaneously, according to T.R. Gruber's ontology design principles [36], each design stage of the scientific literature knowledge element ontology model is standardized to describe the semantic-level knowledge in scientific literature, ensuring the ontology model has clarity, objectivity, consistency, extensibility, and minimal ontological commitment [37].

3.2.1 Concept Set of the Scientific Literature Knowledge Element Ontology In the design stage of the concept set, the ontology model must have clarity, objectivity, consistency, and extensibility. The meaning of each type of concept term must be clear to avoid ambiguity, ensuring semantic consistency of concept classes. When adding new concepts, existing content need not be

modified, enabling the concept set in this knowledge element ontology model to comprehensively and abstractly describe various types of concepts at different levels in scientific literature.

From the perspective of scientific literature features, a scientific literature document is clearly divided into two major concept classes: content features and external features. Each concept class is expanded to obtain the concept set of this scientific literature knowledge element ontology. External features of scientific literature include nine categories: symbolic identifier, title, language, author, publication time, publisher, journal, conference, and document type [25], where the types of scientific literature include scientific books, scientific journals, patent literature, conference literature, scientific reports, government publications, dissertations, standard literature, product materials, and other documents [38]. Internal features of scientific literature include discipline, theme, classification number, keywords, innovation points, and functional features such as abstract, introduction, research status, core research content, experiment and evaluation, and conclusion and prospect. The concepts and hierarchical structure among concepts in the scientific literature knowledge element ontology are shown in Figure 1 [Figure 1: see original paper].

3.2.2 Attribute Set of the Scientific Literature Knowledge Element Ontology In the design stage of the attribute set, the ontology model must have clarity, objectivity, consistency, and minimal ontological commitment. Each type of concept should have a relatively unified, clear, and concise attribute set, enabling this knowledge element ontology model to uniformly describe the internal structure of various knowledge elements. Therefore, the attributes of concept classes in the scientific literature knowledge element ontology are divided into two categories: data attributes and object attributes, with certain hierarchical relationships existing among attributes of the same type. Data attributes describe the data characteristics of the inherent features of concept classes, i.e., the attribute set P of concept classes in the scientific literature knowledge element ontology, where concept subclasses inherit the data attributes of their parent classes. Object attributes refer to semantic relationships between concept classes or attributes, specifically describing part of the relationship set R in the scientific literature knowledge element ontology.

In the scientific literature knowledge element ontology, data attributes include attributes of concepts in the content feature class and attributes of concept classes in the external feature class. In the internal feature class, abstract, introduction, research status, core research content, experiment and evaluation, and conclusion and prospect all have attributes such as knowledge identifier number, knowledge element name, knowledge navigation, original text content, text length, representation form, feature words, and citation number [9]. In the external feature class, attributes of symbolic identifier include document number, patent number, standard number, and report number; attributes of title include title and article title; attributes of publisher include publisher name and pub-

lisher address; attributes of journal include journal name, volume/issue/page; attributes of conference include conference name and conference address; attributes of language include Chinese and foreign languages; attributes of author class include ID, address, affiliation, name, gender, and nationality.

In the scientific literature knowledge element ontology, object attributes include two types: coordinate relationships and hierarchical relationships [25]. Coordinate relationships can be subdivided into innovation relationships, synonym relationships (such as elaboration relationships), positioning relationships, relevant relationships, and sequential relationships. Hierarchical relationships can be subdivided into part-whole relationships, inheritance relationships, attribute relationships, and instance relationships. Among them, relevant relationships include cross relationships (such as cooperative relationships), parallel relationships, and mutually exclusive relationships. Sequential relationships can be subdivided into derivation relationships, order relationships, evolution relationships, implication relationships, and writing relationships. The specific descriptions of object attributes in the scientific literature knowledge element ontology are shown in Table 1 . Except for object attributes with symmetry, other object attributes have inverse attributes. For example, the inverse attribute of the derivation relationship is the derived-from relationship. Due to space limitations, this paper will not elaborate further.

3.2.3 Method Set of the Scientific Literature Knowledge Element Ontology In the design stage of the method set, the ontology model must have clarity and consistency, and various defined methods must be meaningful and non-contradictory. The method set of the scientific literature knowledge element ontology reflects the functional relationships among concepts and attributes in scientific literature. Since title and abstract information can better represent the theme of the text [26], the identification method for sentence-level innovation points in scientific literature is constructed based on this, which can abstractly describe the identification process of theoretical or methodological innovation points. According to the rule that technological innovation points in scientific literature appear more frequently in core research content chapters, the identification method for technological innovation points is defined. Drawing on existing research on academic literature stylistic features [24], the calculation method for text length is simplified. The specific content is shown in Table 2 .

3.2.4 Relationship Set of the Scientific Literature Knowledge Element Ontology In the design stage of the relationship set, the ontology model must have clarity, consistency, and minimal ontological commitment, avoiding making a large number of inferences about the same thing to meet the needs of knowledge sharing and exchange in the scientific literature domain. This model defines the semantic relationships among concepts, attributes, and methods in the scientific literature knowledge element ontology by analyzing the research ideas in scientific literature. The specific content is shown in Table 3 . Among them, the inheritance relationships and part-whole relationships among concepts

are shown in Figure 1 [Figure 1: see original paper].

Through hierarchical classification of scientific literature concepts, the attribute features of concepts are defined, the method set among concepts and attributes is described, and the semantic relationships among concepts, attributes, and methods are represented, ultimately constructing a scientific literature knowledge element ontology model.

4. Example and Discussion

4.1 Example

This experiment uses a scientific literature document as an example to represent and describe the knowledge elements in the document based on the scientific literature knowledge element ontology model, demonstrating the rationality of the above knowledge representation model. The ontology development tool Protégé 4.3.0 was used to construct the knowledge element ontology model, and the Hermit reasoner built into Protégé was used to detect the continuity and consistency of the ontology, indicating that the model has logical reasoning capabilities. Since the reasoning results of Protégé 4.3.0 could not correctly display the Chinese labels of concept classes in the ontology model, this experiment will use English labels to present the representation and reasoning process of the ontology model.

4.1.1 Construction of the Scientific Literature Knowledge Element Ontology Model Since academic journal databases such as CNKI and Wanfang have already used the external features of the scientific literature knowledge element ontology model to classify literature, this experiment mainly constructs the content feature unit of the scientific literature knowledge element ontology model, presenting the content structure of scientific literature. Part of the scientific literature knowledge element ontology model is shown in Figure 2 [Figure 2: see original paper].

4.1.2 Representation and Reasoning of the Scientific Literature Knowledge Element Ontology Model From the Wanfang database, the literature with the highest citation frequency in the past two years (2016-2017) in the domestic information science core journal *Journal of the China Society for Scientific and Technical Information* and most relevant to this paper's theme was selected as the example literature. This is the scientific literature titled “Structural Function Identification of Academic Text—Identification Based on Chapter Content” written by Huang Yong, Lu Wei, and Cheng Qikai. This literature has a standardized writing structure and completely describes five concept subclass instance knowledge elements in the introduction class, making it highly representative.

From the hierarchical structure of the scientific literature knowledge element ontology, introduction knowledge elements can be subdivided into concept def-

inition, research background, key problem, and other knowledge elements. In the introduction section, the main concepts in a field are generally described first, then the research background of the topic area is explored in depth, and the key problems that still need to be solved are analyzed, forming a knowledge chain of “ConceptDefinition \rightarrow ResearchBackground \rightarrow ResearchProblem.” If knowledge element 10 of concept definition, knowledge element 12 of key problem, and another knowledge element 11 are identified from the introduction of a scientific literature document, and this knowledge element 11 is semantically derived from knowledge element 10 and further derives knowledge element 12, then it needs to be determined whether knowledge element 11 identified from the introduction belongs to research background. When this logical 思路 is input into the Protégé ontology editor and the HermiT reasoner is run, the result shows that knowledge element 11 belongs to the research background knowledge element in the introduction. This result is shown in the dashed box in the Type class of Description11, indicating that the above logical relationship has been verified. The detailed reasoning results are shown in Figure 3 [Figure 3: see original paper].

Additionally, when deducing the method set of the scientific literature knowledge element ontology model, the identification method for theoretical innovation points in scientific literature is: if the feature words of the knowledge element to be identified intersect with the feature words of the title (Title), theme words (Theme), and purpose overview (PurposeOverview) in the abstract, then the knowledge point is a theoretical innovation point (TheoryInnovation). From the core research content, a theoretical knowledge point (KnowledgePoint01) is found. According to the above judgment rules, instances 01 of title, theme words, and purpose overview are created respectively. Since the feature words of the knowledge point instance to be identified contain the same feature word in title instance 01 (Title01), theme instance (Theme01), and purpose overview instance (PurposeOverview01), it is verified that the knowledge point is a theoretical innovation point of the literature. The specific operation trace is shown in Figure 3 [Figure 3: see original paper].

4.2 Analysis and Discussion

By analyzing the text structure of scientific literature and building on knowledge element ontology theory, this study constructs a scientific literature knowledge element ontology model to represent the structure of knowledge elements in scientific literature and the semantic relationships among knowledge elements. This knowledge element ontology model demonstrates good performance in content presentation, construction method, and application scenarios, specifically including the following points:

- (1) **Presenting a clear knowledge organization system.** The scientific literature knowledge element ontology model constructed in this study has dual organizational structures—horizontal and vertical. First, in the horizontal structure of the model, starting from the external and content

features of scientific literature and delving progressively into chapters and paragraphs of scientific literature, functional knowledge points in scientific literature are sought, ultimately presenting the knowledge structure at different levels within scientific literature. Second, in the vertical structure of the model, from the perspective of research habits, attention is first paid to the obvious external features of scientific literature for preliminary screening of authors, document types, publishers, and publication times. Then, the content of scientific literature is explored by reading sections such as abstracts and introductions to quickly understand the general content of the literature. If relevant thematic knowledge of interest to users is found, the precise knowledge elements needed by users in that chapter are further explored. Therefore, this ontology model can clearly present the text structure and knowledge context in the scientific literature domain.

- (2) **More precise and efficient model construction method.** The construction method of this knowledge element ontology model integrates the dual advantages of ontology construction methods and knowledge element construction methods. First, under the guidance of ontology construction methods, the constructed ontology conceptual framework can abstractly describe the universal text structure in scientific literature, based on which document-level scientific literature can be segmented into sentence-level fine-grained knowledge points. Additionally, according to the knowledge chain network in the constructed ontology model, not only can rich knowledge points in scientific literature be accurately located and represented, but the reasoning and discovery of knowledge within scientific literature can also be effectively promoted. Second, incorporating the idea of the n-tuple construction method of knowledge elements, the attribute features of the concept classes to which knowledge elements in scientific literature belong are standardized, and a relatively unified attribute set is adopted to concisely and efficiently represent the internal structure of knowledge elements such as abstract, introduction, and research status in scientific literature, as well as finer-grained knowledge elements like research background, research problem, and research method. This is conducive to the representation, storage, sharing, and exchange of knowledge in various documents such as academic papers, journal articles, dissertations, and patents.
- (3) **Broad application prospects.** This knowledge element ontology model is applicable to knowledge representation, knowledge organization, and knowledge indexing scenarios, assisting users at different levels in learning research knowledge in different domains. First, in knowledge representation application scenarios, based on this ontology model, important knowledge points in scientific literature in a certain domain can be described and annotated, facilitating users' quick understanding of the knowledge structure of scientific literature in the domain and assisting users with other disciplinary backgrounds in quickly learning professional knowledge in the

domain. Second, in knowledge organization application scenarios, for academic researchers who have not yet formed standardized research and writing 思路, they can understand the research ideas and writing patterns of scientific literature in the domain based on the text structure presented in the model, shortening the cycle of producing high-level research results and promoting the rapid development of multi-domain disciplines. Third, in knowledge indexing application scenarios, this ontology model can be applied to online submission systems. During paper submission, editorial departments will set the main knowledge modules that submitted papers should have according to this ontology model, standardizing the content and format of submitted papers. When the paper is accepted, editorial departments should encourage and guide the authors to accurately decompose and annotate the paper content according to the knowledge structure of this ontology model, promoting collaborative indexing of knowledge in the scientific literature domain.

In summary, this ontology model can effectively represent and describe the text structure of scientific literature and the internal structural associations of knowledge elements, has broad application prospects, and can provide more efficient literature knowledge services for users. However, using only one example is insufficient to verify the effectiveness of the model, and the model still has the following shortcomings: (1) The scientific literature knowledge element ontology model only presents the universal text structure of the scientific literature domain, while the text structure of scientific literature varies across different disciplines and topics, requiring further refinement of the applicable domains of the model and improvement of the ontology structure. (2) The ontology model has not yet undergone scientific evaluation, and future work will deeply study the evaluation methods for scientific literature knowledge element ontology models.

5. Conclusion and Outlook

To provide fine-grained and precise knowledge point services, this study proposes a knowledge representation model for scientific literature based on analyzing the hierarchical functional structure of scientific literature content. Using a scientific literature document as an example, the internal and external features of fine-grained knowledge elements in scientific literature and the logical relationships among knowledge elements are uniformly represented, laying a solid theoretical foundation for multi-angle, precise, and fast knowledge point retrieval. This ontology model can serve as a standard for the representation and storage mode of knowledge resources within scientific literature, used to improve existing coarse-grained literature knowledge service models and obtain smaller-grained and deeper-level scientific and technological resources. Applying this ontology model to the organization and collaborative indexing scenarios of scientific literature knowledge resources will change the knowledge publishing, usage, dissemination, and acquisition methods of publishers and researchers, promoting the sharing of fine-grained knowledge resources and the discovery of

innovative knowledge points.

Future research will focus on: (1) Refining the ontology structure for different disciplinary domains; (2) Developing scientific evaluation methods for the ontology model; (3) Applying the model to large-scale scientific literature databases for practical validation.

References

- [1] GU X, BLACKMORE K L. Recent trends in academic journal growth[J]. *Scientometrics*, 2016, 108(2): 693-716.
- [2] HUA Bolin, LIU Yining, ZHENG Yanning. Research on construction method of extraction rules for academic definitions[J]. *Information Studies: Theory & Application*, 2011, 34(12): 5-9.
- [3] HUA Bolin. Research on types and description rules of method knowledge elements in academic papers[J]. *Journal of Library Science in China*, 2016, 42(1): 30-40.
- [4] WEN Youkui, WEN Hao, XU Duanyi, et al. Knowledge element mining based on innovation points[J]. *Journal of the China Society for Scientific and Technical Information*, 2005, 24(6): 663-668.
- [5] QIN Chunxiu, LIU Jie, LIU Huailiang, et al. Research on content description framework of scientific text based on knowledge elements[J]. *Library and Information Service*, 2017, 61(10): 116-124.
- [6] XIAO Quan, CAI Shuqin, YE Bo. Research on knowledge similarity calculation model based on hypergraph structure[J]. *Journal of the China Society for Scientific and Technical Information*, 2010, 29(5): 805-812.
- [7] MA Chuangxin, CHEN Xiaohe, QU Weiguang. Research and design of knowledge network for classical ancient annotation literature[J]. *Library and Information Service*, 2013, 57(9): 124-128.
- [8] WU Sizhu, LI Feng, ZHANG Zhixiong. Research on semantic representation and publishing mode of knowledge resources—Taking Nanopublication as an example[J]. *Journal of Library Science in China*, 2013, 39(4): 102-109.
- [9] WANG Yu, LIU Miao. Construction of journal literature knowledge warehouse based on knowledge elements[J]. *Information Studies: Theory & Application*, 2013, 36(8): 91-94.
- [10] SUO Chuanjun. Research on aging and innovation of academic papers from the perspective of knowledge transfer[J]. *Library and Information Service*, 2014, 58(5): 5-12.
- [11] GAO Junfang. Research on disciplinary knowledge services in university libraries under cloud computing[J]. *Library Science Research*, 2015, 33(2): 54-58.
- [12] HUA Bolin. Research on architecture of knowledge extraction system based on NLP[J]. *New Technology of Library and Information Service*, 2007, 2(10): 38-41.
- [13] SONG Jihua, LI Guoyu, WANG Ning. Production expression of phonetic-semantic relationships in Shuowen Jiezi[J]. *Journal of Chinese Information*

- of scientific literature to improve information access[EB/OL]. [2017-11-15]. http://www.aclweb.org/old_{anthology}/W/W13/W13-19.pdf#page=114.
- [31] ZHU Liping, LI Hongqi, YANG Zhongguo, et al. An information extraction method for introduction of scientific literature[J]. Journal of Shandong University (Natural Science), 2015, 50(7): 23-30.
- [32] LU Wei, HUANG Yong, CHENG Qikai. Functional framework of structural function identification of academic text and identification based on chapter titles[J]. Journal of the China Society for Scientific and Technical Information, 2014(9): 979-985.
- [33] HUA Bolin. Extraction of intelligence method terms from Chinese academic literature[J]. New Technology of Library and Information Service, 2013(6): 68-75.
- [34] LIAN Yanhua, MA Xiaoguang. Discussion on evaluation process analysis model[J]. R&D Management, 1999(3): 1-5.
- [35] ZHANG L. Grasping the structure of journal articles: Utilizing the functions of information units[J]. Journal of the Association for Information Science & Technology, 2012, 63(3): 469-480.
- [36] GRUBER T R. Toward principles for the design of ontologies used for knowledge sharing[J]. International Journal of Human Computer Studies, 1995, 43(5-6): 907-928.
- [37] WANG Hongwei, HUO Jiazhen, WANG Wei, et al. Ontology model structure design for semantic retrieval applications[J]. Systems Engineering and Electronics, 2010, 32(1): 166-174.
- [38] ZHANG Han, XU Shuo, QIAO Xiaodong. Review of topic model development integrating internal and external features of scientific literature[J]. Journal of the China Society for Scientific and Technical Information, 2014(10): 1108-1120.

Author Contributions

Qin Chunxiu: Proposed the research question and 思路, provided revision suggestions;

Yang Zhijuan: Designed the research plan, wrote and revised the paper;

Zhao Pengwei: Provided revision suggestions;

Liu Jie: Provided revision suggestions.

English Abstract

[Purpose/significance] With the rapid growth of scientific literature resources, users are surrounded by the ocean of scientific literature, and the future trend of document knowledge retrieval is to provide users with fast and accurate fine-grained knowledge element services. **[Method/process]** Based on the analysis of the content structure of scientific literature, this paper went gradually and deeply into the contents of scientific literature, with a view of constructing a knowledge element ontology model of scientific literature for the knowledge representation, to express fine-grained knowledge points owning

holonomic sentence meanings in the content of scientific literature as knowledge elements with a unified structure. **[Result/conclusion]** The paper illustrates the rationality of the model of scientific literature by means of displaying the content of a scientific literature, but this example only shows knowledge points in the introduction of scientific literature and the knowledge element ontology model needs to be further verified.

Keywords: scientific literature; knowledge representation; knowledge element; ontology model

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

Source: ChinaXiv — Machine translation. Verify with original.