ChinaRxiv

AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202410.00013

Postprint: Research on the Construction of Standards Systems for Scientific Literature Mining, Analysis, and Services

Authors: Cao Xiaoli, Li Hanyu, Zhang Zhixiong

Date: 2024-10-06T00:00:00+00:00

Abstract

[Objective] By constructing a standard system for scientific literature mining, analysis, and services, this study aims to enhance the quality and efficiency of knowledge acquisition, integration, and utilization from scientific literature, promote exchange and cooperation among scientific journal publishing institutions, and improve the efficiency of fine-grained knowledge mining and analysis of scientific literature in research institutions.

[Methods] Following a research framework of "hierarchy-content-system-optimization strategy", and proceeding from six levels including national, industry, local, association, enterprise, and international standards, this research covers the entire process of "data processing, knowledge mining, knowledge service, intelligence analysis, and integrated application". It thoroughly examines and analyzes the current status and characteristics of the standard system for scientific literature mining analysis and services, and proposes corresponding optimization strategies.

[Results] The standard system for scientific literature mining analysis and services demonstrates characteristics of multi-level standards with separate governance yet mutual integration, multi-dimensional content standards integrated into a unified whole, and hierarchical-content co-construction of the standard system.

[Conclusion] Optimization strategies are proposed for the standard system of scientific literature mining analysis and services, focusing on systematically advancing standard formulation across all levels, enhancing the development of the full-process content system, and concentrating on the construction of standards for both general and domain-specific knowledge objects.

ChinaRxiv 中X

Full Text

Research on Standard System Construction for Mining, Analysis, and Service of Scientific and Technological Literature

Cao Xiaoli[1],[2]), Li Hanyu[1],[2]), Zhang Zhixiong[1],[2])
[1]) National Science Library, Chinese Academy of Sciences, No. 33 Beisihuan Xilu, Zhongguancun, Haidian District, Beijing 100190, China
[2]) Department of Information Resources Management, School of Economics and Management, University of Chinese Academy of Sciences, No. 33 Beisihuan Xilu, Zhongguancun, Haidian District, Beijing 100190, China

Abstract

[**Purposes**] By constructing a standard system for mining, analysis, and service of scientific and technological literature, this study aims to improve the quality and efficiency of knowledge acquisition, integration, and utilization in scientific literature, promote communication and cooperation among scientific journal publishing institutions, and enhance the efficiency of fine-grained knowledge mining and analysis of scientific literature in research institutions.

[**Methods**] Adopting the research framework of "hierarchy-content-system-optimization strategy," this study examines six levels of standards—national, industry, local, group, enterprise, and international—covering the entire process of "data processing, knowledge mining, knowledge services, intelligence analysis, and integrated application." It thoroughly reviews and analyzes the current status and characteristics of the standard system for scientific and technological literature mining, analysis, and service, and proposes optimization strategies for the system.

[**Findings**] The standard system for mining, analysis, and service of scientific and technological literature exhibits three key characteristics: multi-level standards that are separately governed yet mutually integrated, multi-dimensional content standards that blend into a cohesive whole, and a hierarchical content co-construction approach to standard system building.

[**Conclusions**] Optimization strategies are proposed for the standard system from three perspectives: systematically promoting the formulation of standards at all levels, improving the development of the full-process content system, and focusing on the construction of universal and domain-specific knowledge object standards.

Keywords: Scientific and technological literature; Standard system; Data annotation; Knowledge mining; Optimization strategy

1 Introduction

Standards serve as the technical cornerstone of economic activities and social development, with standardization playing a fundamental and leading role in advancing the modernization of national governance systems and capabilities [1]. In March 2024, the Government Work Report emphasized the need to "implement standard upgrading actions and accelerate the construction of a standard system adapted to high-quality development requirements" [2]. Currently, AI technologies represented by AIGC (Artificial Intelligence Generated Content) have brought unprecedented opportunities and challenges to the mining, analysis, and service of scientific literature, and the efficient application of AI technologies heavily depends on the quality of data and annotation in scientific literature corpora [3]. Therefore, studying the standard system construction for scientific literature mining, analysis, and service from a standardization perspective can facilitate knowledge acquisition, integration, and utilization in scientific journals, promote academic exchange and cooperation among publishing institutions, strengthen collaborative innovation and common development between publishing institutions and AI enterprises, and enhance the efficiency of fine-grained knowledge mining and analysis of scientific literature in research institutions.

Existing research primarily focuses on specific aspects such as data annotation specifications [4,5], mining tool status [6,7], metadata standards [8,9], and JATS data standard applications [10,11] for scientific literature. For instance, Zeng Jianxun [4] pointed out the need to establish a data annotation standard system in the literature and information field and accelerate the formulation of relevant national and industry standards. However, scientific literature mining, analysis, and service involve multiple stages including data annotation, knowledge services, and intelligence analysis. Current research typically concentrates on standards and norms for a single stage or the application of a specific standard, without systematically and holistically reviewing standards across the entire process. This prevents a comprehensive view of the standard system and limits the ability to propose feasible recommendations for its improvement.

This study follows the "hierarchy-content-system-optimization strategy" research framework, examining six levels of standards—national, industry, local, group, enterprise, and international—covering the full process of "data processing, knowledge mining, knowledge services, intelligence analysis, and integrated application." It thoroughly reviews and analyzes the current status of relevant standards for scientific literature mining, analysis, and service. Based on this analysis, the study examines the characteristics of the standard system and proposes future optimization strategies, aiming to construct a more comprehensive standard system from a systematic and holistic perspective, thereby improving the quality and efficiency of knowledge utilization in scientific journals and promoting exchange and cooperation among publishing institutions.

2 Standard Hierarchy System for Scientific and Technological Literature Mining, Analysis, and Service

As normative documents, standards provide practical guidance for each stage of scientific literature mining, analysis, and service. With the exponential growth of scientific literature data and rapid iteration of large AI models, various industries and institutions increasingly emphasize standardization work in this field to continuously strengthen foundational support for data annotation, model training, and knowledge mining, further promoting the development of intelligence analysis in China and facilitating global scientific and technological innovation and international exchange [3,4].

Using keywords related to the theme of "scientific and technological literature mining, analysis, and service," this study retrieved published standards across six levels—national, industry, local, group, enterprise, and international—and manually reviewed the results. Ultimately, 44 important standards were selected for systematic summary and organization, forming a clearly structured standard system covering the entire process and all domains.

2.1 National Standards

National standards are crucial for guiding localities and industries in modernizing scientific literature mining and analysis standardization, and for achieving unified specifications for scientific information resource sharing and intelligent analysis technologies [1]. They play a significant role in promoting technological innovation and industrial development, as well as facilitating the rapid dissemination and wide application of research achievements. Based on their regulatory objects, national standards can be divided into five subsystems: "metadata standards," "knowledge mining standards," "knowledge service standards," "intelligence analysis standards," and "integrated application standards," as detailed in Table 1 .

The national standards system includes foundational standards such as "Standard Literature Metadata," "Journal Article Tag Suite," "Information and Documentation—Resource Description," and "Press and Publication—Knowledge Services—Knowledge Object Identifier." These standards address metadata description models, literature tagging, and knowledge object identification. Notably, GB/T 3792 provides the first national standard for standard literature data processing, replacing the previous GB/T 22373-2008 and defining core and common metadata for standard literature. The "Journal Article Tag Suite" fills a gap in China's full-text structured description of literature [12]. Additionally, standards for machine learning data annotation procedures and knowledge graph technical frameworks have been established to support AI applications and financial risk prevention.

2.2 Industry Standards

Industry standards provide foundational support for various sectors—including press and publication, medicine, public safety, and telecommunications—to participate in scientific literature mining, analysis, and service research in an orderly manner, offering references for cross-domain knowledge sharing, innovative applications, and intelligent decision-making. Based on their regulatory objects, industry standards can be divided into four subsystems: "metadata standards," "knowledge mining standards," "knowledge service standards," and "integrated application standards," as detailed in Table 2 .

The industry standards system includes the "Literature Fragment Identifier," which enables identification of fragmented literature content and provides a methodological foundation for recognizing, locating, and analyzing document segments, allowing researchers to freely adjust the granularity of literature analysis. The "Journal Full-text XML Description Tag Set" provides detailed descriptive tags and structural specifications for XML files of journal articles. Standards for AI medical devices unify specifications for diverse data modalities, annotation objects, technical tools, and platform architectures. Knowledge association service standards regulate models and metadata rules for knowledge association analysis, while AI-based knowledge graph construction standards specify processes, technical requirements, and functional specifications. Big data standards address data management platforms, integration tools, and data mining platforms.

2.3 Local Standards

Guided by national standards, different regions have established local standards according to their actual needs. Based on their regulatory objects, local standards can be divided into two subsystems: "knowledge mining standards" and "intelligence analysis standards," as detailed in Table 3 .

Local standards from Shanxi Province, such as DB14/T series, regulate general technical requirements and work procedures for AI data annotation. Shanghai's DB 31/T standards include the "Technical Transfer—Competitive Intelligence Analysis Service Specification," which establishes unified norms for data collection, processing, and analysis. As the first standard in China's competitive intelligence analysis field, it holds significant importance for technology transfer and enterprise innovation [13]. Shanghai also issued the "Technical Transfer—Technology Evaluation" standard, which provides unified norms for information resource acquisition, evaluation, and analysis, serving as Shanghai's first technology evaluation specification [14].

2.4 Group Standards

The 2018 revision of the Standardization Law of the People's Republic of China officially clarified the status of group standards in China's new standard system [15]. Driven by market demand, group standards serve technological innovation

and industrial upgrading. Based on their regulatory objects, group standards can be divided into three subsystems: “metadata standards,” “knowledge mining standards,” and “integrated application standards,” as detailed in Table 4 .

Group standards include the “Chinese Academy of Sciences Academic Journal Full-text XML Description Tag Set (CAS JATS),” which defines structured description specifications for CAS academic journals. The “NSTL Unified Literature Metadata Standard” supports data integration, fusion, analysis, and mining for NSTL business systems. Data annotation specifications for machine learning clearly define annotation processes, tools, and platforms for multi-type data resources including text, images, audio, and video. Content security-focused AI data annotation guidelines aim to improve model accuracy in identifying false information. Internet-based data mining processing system specifications regulate technical requirements and system architecture. Standards for scientific big data platform data warehouse development and construction define development processes and data models.

2.5 Enterprise Standards

To implement the National Standardization Development Outline and further improve enterprise standard formulation procedures, the State Administration for Market Regulation revised and issued the “Enterprise Standardization Promotion Measures,” which came into effect on January 1, 2024. This initiative aims to stimulate enterprise innovation vitality, encourage active participation in standard formulation, effectively combine market demand with practical experience, promote mutual enhancement between standardization work and enterprise practice, and boost healthy industry development [16]. Based on their regulatory objects, enterprise standards can be divided into two subsystems: “knowledge mining standards” and “intelligence analysis standards,” as detailed in Table 5 .

Enterprise standards include specifications for machine learning data annotation processes from companies such as CASCO (Keda Guochuang Yunwang Technology) and Hangzhou Yaogan Digital Technology. Shandong New Engineering Information Technology’s standard addresses intelligent data annotation service systems, specifying functional requirements and system architecture. Data mining process specifications from CASCO regulate modeling procedures. Hangzhou Chenggu Technology’s standard covers big data storage and mining analysis systems. Shenzhen Guohaizhifeng Intellectual Property Agency and Guangdong Huazhong Yuechuang Intellectual Property Operation Management have established standards for patent technology intelligence analysis, high-value cultivation, and technology transfer, as well as technical intelligence analysis service specifications.

2.6 International Standards

International standards provide a global perspective and reference for scientific literature mining, analysis, and service research. Their formulation and implementation enhance interoperability and compatibility across different countries and regions, facilitating international cooperation projects, strengthening international exchange and sharing of scientific achievements, and promoting global scientific innovation and knowledge sharing. Based on their regulatory objects, international standards can be divided into two subsystems: "metadata standards" and "knowledge service standards," as detailed in Table 6 .

Key international standards include the TEI (Text Encoding Initiative) metadata standard, which defines markup definitions, conceptual features, description methods, and record structures for electronic text exchange. The Dublin Core metadata element set establishes 15 core metadata elements for cross-domain resource description. DATS (Data Tag Suite) defines elements, properties, and relationships for scientific dataset description. JATS (Journal Article Tag Suite) version 1.3 provides unified descriptions of journal literature formats and has been widely adopted by publishers, libraries, and scientific journals [17], with applications on platforms such as PMC and the Chinese Medical Association [18]. ISO 26324 specifies the Digital Object Identifier (DOI) system syntax, description, and resolution components, while ISO 17316 defines the International Standard Link Identifier (ISLI) for identifying relationships between information entities, promoting research on knowledge entity relationship mining in scientific literature [19]. The P2807.4 standard addresses scientific knowledge graph specifications for data models, construction processes, and sharing applications.

3 Standard Content System for Scientific and Technological Literature Mining, Analysis, and Service

The standard content system for scientific literature mining, analysis, and service constitutes the core foundation for promoting the development of intelligent scientific information processing and service ecosystems. To implement the National Standardization Development Outline and accelerate scientific and technological innovation, this content system should closely align with national S&T development strategies, fully consider the technological frontiers and industry demands of scientific literature analysis, and promote the improvement and widespread application of the standard system. Based on the standard content and classification from the previous section, combined with the actual workflow of scientific literature mining, analysis, and service [20], this study constructs a comprehensive standard content framework covering the entire process from data processing, knowledge mining, and knowledge services to intelligence analysis and integrated application (see Figure 1 [Figure 1: see original paper]), aiming to enhance the efficiency and quality of scientific literature analysis and promote knowledge discovery and innovation.

3.1 Data Processing Standards

Under the rapid development of AI technology, data processing quality directly affects the efficiency and accuracy of scientific literature mining, analysis, and service. Reviewing standards across all six levels reveals that data processing standards mainly include: (1) Metadata standards: focusing on metadata description models, identification, and structural metadata models for literature or texts to enhance data accessibility, understandability, and interoperability; (2) Literature tagging standards: regulating literature structure, format, element tags, and inter-element relationships to support journal article editing, publishing, and management; (3) Knowledge object identifier standards: specifying identification objects, coding structures, and metadata for knowledge object identifiers to mark knowledge within scientific literature.

Overall, existing standards generally focus on basic metadata specifications for scientific literature, with few standards addressing fine-grained knowledge identification at the content level, resulting in a lack of uniformity and standardization. Scientific literature knowledge exhibits diversity and complexity, encompassing not only text but also images, charts, formulas, and other modalities, with different domains having varying norms for content expression, knowledge objects, and literature tags. There is an urgent need to develop unified and standardized fine-grained knowledge object identification standards for scientific literature [21,22].

3.2 Knowledge Mining Standards

Knowledge mining standards aim to regulate the processes and technical requirements for fine-grained knowledge mining, analysis, storage, and management in scientific literature, providing unified methodologies and process specifications to improve efficiency and quality. These standards primarily include: (1) Data annotation standards: specifying annotation processes, frameworks, objects, and requirements to guide AI-oriented data annotation practices; (2) Data mining standards: defining technical requirements, system architecture, process specifications, and storage and management requirements to provide operational guidelines for internet-oriented data mining practices.

However, current standards remain insufficient in regulating knowledge object annotation, processes, technical requirements, and storage and management specifications for fine-grained knowledge mining. Moreover, different standards regulate inconsistent knowledge annotation objects with significant variations in annotation granularity. There is an urgent need to develop fine-grained knowledge annotation standards for both general and vertical domains in scientific literature to extract high-quality, trustworthy, cutting-edge, rigorous, systematic, and reliable knowledge and scientific data, thereby empowering large scientific models and AI-driven research tasks.

3.3 Knowledge Service Standards

Knowledge service standards aim to provide unified technical specifications and service guarantees for big data platforms and systems, regulating basic requirements for database interfaces, software interfaces, and operation and maintenance service interfaces to provide solid technical support for data resource sharing and knowledge application [23]. These standards mainly include: (1) Knowledge graph construction standards: specifying conceptual models, technical frameworks, construction processes, and sharing applications for knowledge graphs to guide structured organization and integrated management of diverse knowledge from scientific literature; (2) Knowledge association service standards: regulating encoding structure models, encoding rules, and relationship identification for knowledge association to promote interconnectivity of knowledge and technical information resources and enhance knowledge service levels.

Nevertheless, corresponding standards are lacking for knowledge base interfaces and data interfaces in scientific literature knowledge service systems, tools, and platforms. The design, implementation, and usage requirements of knowledge bases and data interfaces should satisfy cross-system and cross-platform compatibility, which is crucial for knowledge association, description, analysis, and application services, as well as for the sharing and application of knowledge and technical resources [24].

3.4 Intelligence Analysis Standards

The intelligent intelligence analysis standard system includes specifications for intelligence data collection, analysis methods, result output, quality assessment, technology transfer, and patent technology transformation, aiming to improve the efficiency and accuracy of intelligence analysis, promote intellectual property management and technology achievement transformation [25], and realize the application value of scientific literature mining in intelligence analysis. These standards include: (1) Technology transfer standards: regulating technical requirements, data collection and processing, service evaluation requirements, and technology evaluation indicators for technology transfer intelligence analysis to guide technology transfer and evaluation services; (2) Technical intelligence analysis standards: specifying patent cultivation processes and methods, patent technology cultivation and transformation, and technical intelligence analysis workflows to provide guidelines for intellectual property management and intelligence analysis.

Overall, existing standards cover limited intelligence analysis scenarios, focusing only on key tasks such as technology transfer, technology evaluation, and patent cultivation. Identifying typical intelligence analysis scenarios for various stakeholders—including universities, research institutes, and enterprises—and extracting domain-specific and scenario-specific intelligence analysis workflows can provide decision support for intellectual property management and risk pre-

vention.

3.5 Integrated Application Standards

Integrated application standards play a normative role in data resource management and technical integration applications for scientific big data platforms, ensuring that scientific literature mining and analysis technologies can be effectively implemented in practical scenarios. These standards include: (1) Data management platform standards: regulating technical requirements for big data management platforms, data warehouse models, development processes, and construction plans to promote effective big data governance; (2) Data mining platform standards: specifying development requirements for data integration tools, technical requirements for data mining tools, and platform specifications to provide guidelines for better data mining services.

However, existing standards still lack unified specifications for integrated interoperability, interface norms, process specifications, input-output specifications, and data representation methods for scientific literature data intelligence analysis tools. Integrated application standards and intelligence analysis standards complement each other and should provide unified and standardized storage and management for scientific literature data across different domains to satisfy cross-platform interoperability requirements for intelligence analysis tools.

4 Construction of the Standard System for Scientific and Technological Literature Mining, Analysis, and Service

The hierarchical system and content system of standards for scientific literature mining, analysis, and service are interwoven, jointly constructing the complete standard system, as shown in Table 7 .

4.1 Multi-Level Standard Separation and Integration

The standard system exhibits characteristics of multi-level standard separation and integration. At the national level, relevant standards have been formulated for the entire process of scientific literature mining, analysis, and service, playing a leading role in practice. However, national standards primarily focus on data processing and knowledge service stages, with knowledge mining, intelligence analysis, and integrated application standards yet to be perfected and enriched. Due to different industries and application scenarios, national standards are generally macro-level and require further refinement and improvement. Industry standards concentrate on data processing, knowledge mining, knowledge services, and integrated applications in sectors such as press and publication, medicine, and telecommunications, and should continuously expand to new industries to develop adaptive standards that meet evolving needs. Local standards serve regional development plans with distinct focuses and characteristics, but currently remain limited in number, requiring further top-level planning to guide relevant standard formulation. Group and enterprise standards focus

more on specific domains and organizational needs, offering greater flexibility to satisfy particular business scenario requirements. International standards, primarily addressing metadata, journal article tags, and knowledge services, provide an international perspective for China's standard system construction and constitute an important component of the three-dimensional standard system. Although standards at each level have different emphases, they mutually support each other in content. Promoting the integrated development of national, industry, local, group, enterprise, and international standards represents an important goal for modern scientific literature mining, analysis, and service standard system construction.

4.2 Multi-Dimensional Content Standard Integration

The content system of standards for scientific literature mining, analysis, and service uses data processing as the foundational guarantee, with knowledge mining, knowledge services, and intelligence analysis as the core to achieve integrated application goals, promoting comprehensive upgrading of the scientific information intelligent processing and service ecosystem. Data processing standards cover metadata, journal literature tags, and knowledge object identifiers, receiving attention from diverse industries both domestically and internationally. Knowledge mining standards address data annotation, big data mining, analysis, storage, and management, representing a key focus for domestic industry, local, group, and enterprise standards. Knowledge service standards concentrate on critical aspects such as knowledge association and knowledge graph construction and application, forming important support for building intelligent information service ecosystems. Intelligence analysis standards primarily address technology transfer, competitive intelligence analysis, patent technology cultivation, and achievement transformation, though current relevant standards are limited and require strengthening. Integrated application standards involve the construction of data management systems, data mining platforms, and integration tools, representing important norms for implementing data processing and analysis technologies. The cross-complementarity and integrated development of multi-dimensional content standards constitute crucial guarantees for the comprehensive and multi-dimensional implementation of standardization in scientific literature mining, analysis, and service.

4.3 Hierarchical Content Co-Construction of the Standard System

Standards at different levels and dimensions interact and work together to construct a complete standard system for scientific literature mining, analysis, and service, demonstrating the integrity and systematic nature of the system. In this construction, national standards serve as top-level design, providing comprehensive planning and guidance for data processing, knowledge mining, knowledge services, intelligence analysis, and integrated application, playing a leading role. Industry standards focus on specific industry needs and development directions, providing norms and support for related businesses within industries

and promoting healthy industry development. Local standards provide specific guidance for local scientific literature mining and analysis based on regional needs, compensating for the insufficient local application of national standards. Group and enterprise standards are more closely aligned with specific business scenarios and practical needs, offering strong flexibility and targeted support for enterprises and organizations. International standards serve as a link between domestic and international standard systems, promoting the alignment and integration of Chinese standards with international standards and expanding the breadth and depth of the standard system.

5 Optimization Strategies for the Standard System

Based on the above analysis of the hierarchical and content systems of standards for scientific literature mining, analysis, and service, this section summarizes current deficiencies and proposes future optimization strategies.

(1) Systematically Promote Standard Formulation at All Levels. Currently, relevant national, industry, local, group, enterprise, and international standards have been issued under the theme of "scientific literature mining, analysis, and service." Among them, national standards are the most comprehensive, covering the entire process of "data processing, knowledge mining, knowledge services, intelligence analysis, and integrated application," playing a leading role in practice. However, other level standards, such as local, enterprise, and international standards, involve only two stages, indicating deficiencies and room for improvement. Future efforts should systematically promote coordinated standard formulation at all levels, with particular emphasis on national and industry standards to provide top-level planning for general-domain scientific literature mining, analysis, and service while meeting the needs of vertical domains.

(2) Improve Full-Process Content System Development. Scientific literature mining, analysis, and service involve the entire process of "data processing, knowledge mining, knowledge services, intelligence analysis, and integrated application." However, existing standards mostly remain at the data processing level, such as metadata and literature tags, and the knowledge mining level, such as data annotation. Moreover, current data annotation standards have not penetrated to the fine-grained knowledge annotation level of scientific literature content. Future efforts should formulate standards addressing specific needs and scenarios of scientific literature mining, analysis, and service, covering fine-grained knowledge representation, knowledge mining, association services, intelligence analysis, and integrated application of mining and analysis tools to improve the full-process content system.

(3) Focus on Universal and Domain Knowledge Object Standard Construction. Knowledge objects are core elements in the standard system for scientific literature mining, analysis, and service, and their standard construction plays a vital role. Given the multiple challenges currently facing scientific liter-

ature knowledge services and intelligent intelligence analysis—including rapid literature resource growth, knowledge fragmentation, and data heterogeneity—there is an urgent need to conduct multi-dimensional revelation, deconstruction, and standardization of scientific literature knowledge [27]. Additionally, existing standards and relevant scientific literature annotation models should be integrated, supplemented, and optimized to establish a more comprehensive and inclusive annotation framework and description specification for scientific literature knowledge. Furthermore, while national standards only cover "universal knowledge objects," we should also focus on representation and annotation standards for domain-specific knowledge objects, such as the UMLS standard in medicine, to provide references for industry framework models.

6 Conclusion

This study employs the "hierarchy-content-system-optimization strategy" framework to systematically analyze the hierarchical and content systems of standards for scientific literature mining, analysis, and service. It thoroughly examines the multi-dimensional characteristics of the standard system, emphasizing the integration and unification of multi-level and multi-dimensional content standards, and clarifying the interdependent and complementary relationships among standard content areas including data processing, knowledge mining, knowledge services, intelligence analysis, and integrated application. Finally, the study proposes optimization strategies for systematically promoting standard formulation at all levels, improving full-process content system development, and focusing on universal and domain knowledge object standard construction.

Standardization of scientific literature mining, analysis, and service is not only a systematic project but also an important force in promoting new quality productive forces. From a systematic and holistic perspective, this study investigates standard system construction to promote intelligent processing of scientific information, knowledge sharing and cooperative exchange among journal publishing institutions, comprehensive upgrading of the scientific publishing service ecosystem, and strong support for scientific innovation and knowledge discovery. However, in-depth discussion and analysis of standards for specific practical scenarios such as knowledge object annotation, data association, knowledge graph construction, and technology transfer remain insufficient. Future research will focus on fine-grained knowledge annotation frameworks and tool standards for scientific literature, standardizing annotation and description rules for universal and domain-specific knowledge objects to ensure annotation quality and standardized management and promote data exchange and interoperability.

[1] Xinhua News Agency. The Central Committee of the Communist Party of China and the State Council Issue the "National Standardization Development Outline" [EB/OL]. [2024-02-26]. https://www.gov.cn/zhengce/2021-10/10/content_{5641727}.htm.
[2] Xinhua Net. Two Sessions Authorized Release | Government Work Report [EB/OL]. [2024-03-20]. http://news.enorth.com.cn/system/2024/03/13/055587636.shtml.

[3] Sun Tan, Zhang Zhixiong, Zhou Lihong, et al. AI-driven Fifth Research Paradigm (AI4S) Transformation and Observations [J]. Journal of Library and Information Science in Agriculture, 2023, 35(10): 4-32.
[4] Zeng Jianxun. Building Data Annotation Bases in the Literature and Information Field [J]. Journal of Library and Information Science in Agriculture, 2024, 36(4): 102-103.
[5] Lai Maosheng. Knowledge Organization in the Digital Intelligence Era [J]. Journal of Science and Technology Intelligence Research, 2024, 6(2): 1-10.
[6] Bai Rujiang, Zhao Mengmeng, Zhang Yujie, et al. Review of Scientific Literature Mining Tool Platforms and Key Technologies [J]. Frontiers of Data and Computing, 2021, 3(6): 60-80.
[7] Luo Wei, Tan Yushan. Content-based Scientific Literature Big Data Mining and Application [J]. Information Studies: Theory and Application, 2021, 44(6): 154-157.
[8] Gao Yuwei, Zhu Yanhua, Kong Lihua, et al. Research on Metadata Standards for Data Papers and Associated Scientific Datasets [J]. Chinese Journal of Scientific and Technical Periodicals, 2023, 34(10): 1270-1282.
[9] Yu Qianqian, Zhang Jianyong, Huang Yongwen. Analysis of Design Characteristics of Literature Metadata Standards in Big Data Environments [J]. Library Journal, 2018, 37(11): 35-39, 46.
[10] Cho Y. Open-source code to convert Journal Article Tag Suite Extensible Markup Language (JATS XML) to various viewers and other XML types for scholarly journal publishing [J]. Science Editing, 2022, 9(2): 162-168.
[11] Kang Hongyu, Li Jiao, Hou Zhen, et al. Processing and Transformation of Medical Journal Electronic Resources Based on JATS Standard [J]. Chinese Journal of Scientific and Technical Periodicals, 2018, 29(8): 822-827.
[12] Lin Jiale. National Standard "Journal Article Tag Suite": Providing Guarantee for Building World-class Journals in China [EB/OL]. [2024-02-28]. https://www.magtech.com.cn/CN/news/news119.shtml.
[13] Chen Feng. Analysis of the Implementation of Shanghai's "Technical Transfer—Competitive Intelligence Analysis Service Specification" [J]. Competitive Intelligence, 2021, 17(4): 47-51.
[14] Dai Lixin. Shanghai Issues Local Standards for Technology Transfer Services of Scientific and Technological Achievements [EB/OL]. [2024-03-29]. https://www.sast.gov.cn/content.html?id=kjb214792&cid=1372366026574102530.
[15] Han Dan, Zhang Qingchun, Du Weijun. Analysis and Suggestions on the Development of Group Standards [J]. China Standardization, 2024(4): 25-27.
[16] Zeng Wei, Hu Xiaoping. Analysis of Enterprise Standardization Development Trends Based on the "Enterprise Standardization Promotion Measures" [J]. China Standardization, 2024(2): 51-53, 63.
[17] Randall L, Ubnoske S. Manuscript exchange common approach and journal article tag suite (JATS) compatibility: a case study utilizing the JATS compatibility meta model [J]. Science Editing, 2021, 8(1): 93-97.
[18] Li Jiao, Kou Yuantao, Huang Yongwen, et al. Research on Semantic Publishing Practices at Home and Abroad [J]. Digital Library Forum, 2017(12): 25-31.

[19] Zhu Jiang, Li Xinyi, Jiang Enbo, et al. Association of Scientific Literature and Scientific Data Based on ISLI Standard [J]. Library Theory and Practice, 2020(5): 80-83, 91.
[20] Li Mengli, Wang Ying, Qian Li, et al. Construction of Scientific and Technological Talent Data Infrastructure for New Quality Productive Forces Development [J]. Journal of Library and Information Science in Agriculture, 2024, 36(2): 15-25.
[21] Luo Tingting, Zhao Ruixue, Li Jiao, et al. Construction of Metadata Standard System for Multi-source Heterogeneous Scientific Information Governance [J]. Digital Library Forum, 2021(4): 58-67.
[22] Intelligent Computing Network Alliance. China Academy of Information and Communications Technology Launches Large Model Series Standard Formulation Work [EB/OL]. [2024-05-29]. https://zslm.openi.org.cn/index.php?m=content&c=index&a=show&catid
[23] Zhang Can, Xiong Lijun, Wang Qing, et al. Research on Big Data Knowledge Service Model for Scientific and Technological Innovation [J]. Journal of Library Science, 2024, 46(2): 55-59, 65.
[24] Cao Shujin, Zeng Yingying, Liao Saiyuan, et al. Research on the Development of Standards and Specifications in the Knowledge Organization Field [J]. Journal of Library Science in China, 2023, 49(2): 105-120.
[25] An Ran, Lu Xiaobin, Zheng Yanning. Research on Industrial Technology Intelligence Analysis Method System from the Perspective of Digital Intelligence Convergence—Based on Analysis of Papers in Library and Information Science [J/OL]. Information Studies: Theory and Practice, [2024-05-29]. http://kns.cnki.net/kcms/detail/11.1762.G3.20240313.1318.003.html.
[26] Zheng Chunyu, Wang Lipeng, Chen Jie. Analysis of Current Standard System Status for Chinese Scientific Journals and Related International Progress [J]. Chinese Journal of Scientific and Technical Periodicals, 2021, 32(7): 851-858.
[27] Huang Yongwen, Sun Tan, Zhao Ruixue, et al. Research and Practice of New Knowledge Services Under the Background of Big Data and Artificial Intelligence [J]. Library and Information Service, 2022, 66(19): 36-46.

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

Source: ChinaXiv — Machine translation. Verify with original.