Research, Design, and Application Practice of an AI+ Smart Knowledge Service Ecosystem: A Case Study of the Smart Service Platform Development at the National Science Library, Chinese Academy of Sciences (Postprint)

Authors: Qian Li, Liu Xiwen, Zhang Zhixiong, Liu Huizhou

Date: 2023-04-01T16:02:54+00:00

Abstract

[Purpose/Significance] Artificial intelligence (AI) is triggering cascade scientific breakthroughs, driving a new round of scientific and technological revolution and industrial transformation. How the library and information science field can leverage AI technology to provide intelligent knowledge services and intelligent information systems is currently the central focus and hot topic of the industry. [Method/Process] This study comprehensively analyzes, from both within and beyond the library and information science industry, the new platforms, new services, new opportunities, and new challenges that AI technology and big data bring to the knowledge service paradigm. It proposes the overall approach for constructing an "AI technology + big data"-driven intelligent knowledge service ecosystem, co-constructing a "Sci-Tech Brain" from three levels: intelligent data, intelligent middle platform, and intelligent services, to provide an open intelligent knowledge service ecosystem covering scientific and technological management, scientific and technological innovation, and social academic information environments. [Results/Conclusion] Through exploratory construction in six aspects at the National Science Library, Chinese Academy of Sciences—including the literature and information data lake, intelligent knowledge service engine, intelligent knowledge discovery, intelligent knowledge management, intelligent information analysis system, and intelligent perception environment—significant results have been achieved. Looking ahead, the study clarifies that AI technology still requires further enhancement in data, technology, and service models for big data governance, fine-grained knowledge identification, precise service provision, and other related aspects.

Full Text

Preamble

Volume 65, Issue 15, August 2021
ChinaXiv Cooperative Journal

Research Design and Application Practice of an AI+ Intelligent Knowledge Service Ecosystem: A Case Study of the Intelligent Service Platform Construction at the National Science Library, Chinese Academy of Sciences

Qian Li[1],[2], Liu Xiwen[1],[2], Zhang Zhixiong[1],[2], Liu Huizhou[1],[2]
[1] National Science Library, Chinese Academy of Sciences, Beijing 100190
[2] Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190

Abstract:
[Objective/Significance] Artificial Intelligence (AI) is triggering chain-reaction-like scientific breakthroughs and leading a new round of scientific and technological revolution and industrial transformation. How to leverage AI technology to provide intelligent knowledge services and intelligent information systems has become a focal point and hot topic in the library and information science field. [Method/Process] This paper comprehensively analyzes AI technology and big data from both within and outside the library and information industry to identify new platforms, new services, opportunities, and challenges they bring to knowledge service paradigms. It proposes an overall approach to building an intelligent knowledge service ecosystem driven by "AI technology + big data," constructing a "Science and Technology Intelligence Brain" from three layers: intelligent data, intelligent middle platform, and intelligent services. This provides an open intelligent knowledge service ecosystem covering science and technology management, scientific and technological innovation, and the social academic information environment. [Result/Conclusion] The National Science Library of Chinese Academy of Sciences has conducted exploratory construction in six areas: literature and information data lake, intelligent knowledge service engine, intelligent knowledge discovery, intelligent knowledge management, intelligent information analysis system, and intelligent sensing environment, achieving meaningful results. Looking to the future, the paper clarifies that AI technology still needs further improvement in data, technology, and service models for big data governance, fine-grained knowledge recognition, and precision service provision.

Keywords: Artificial Intelligence (AI); Scientific and Technological Big Data; Intelligent Knowledge Service; Deep Learning
Classification Number: G25
DOI: 10.13266/j.issn.0252-3116.2021.15.010

Artificial intelligence is currently triggering chain-reaction-like scientific breakthroughs and leading a new round of scientific and technological revolution and industrial transformation. As the foundation supporting AI development, scientific and technological big data records scientific truth verification processes, experimental observations, research conclusions, and online exchanges—providing the data foundation for AI algorithmic models used in scientific and technological innovation discovery. Zhang Dongrong and Qian Li have conducted forward-looking research on several important issues for intelligent information services, including the concept of intelligent services and necessary knowledge fusion technologies [**?**]. Ke Ping and others have proposed that the future library will evolve into a new form characterized by networked operation, digitization, virtualization, mobility, and intelligent services [**?**]. Su Xinning has suggested that future libraries will develop into a new morphology of networked, digital, virtualized, mobile, and intelligent services [**?**]. Building on these cutting-edge research perspectives, intelligent knowledge services have become a focus and hotspot in the library and information field. The core elements of knowledge services based on semantic knowledge data include space, data, knowledge, technology, platform, users, and ecology, and its "human-machine-thing" ternary computing system has become an important deployment for companies like Google and Microsoft to seize future big data AI services.

This paper explores the application and practice methods of intelligent knowledge services. Based on the above research, it further defines the concept of intelligent knowledge services: intelligent knowledge services fully utilize "AI + big data" information technology to build intelligent literature and information systems that enable information workers to become flexibly operating "data cleaning plants," "information processing plants," "knowledge generation plants," and "decision-making plants" centered around intelligent literature and information systems. This process allows information work to quickly perceive changes, refine problems, focus on targets, and form solutions, greatly compensating for human intelligence limitations and enhancing people's ability to address complex problems and tasks.

---

2. Analysis of Current Problems in Knowledge Services and Important Industry Development Directions

2.1 Main Problems Facing Current Knowledge Services

In the big data era, massive amounts of scientific and technological big data are generated daily, and user demands are becoming more personalized, customized, and flattened. Professor Zhang Xiaolin has proposed that disruptive changes and the post-library era will drive supply-side structural reforms in knowledge services [**?**], directly reflecting that the imbalance between the supply and demand of scientific and technological resources and information has become the main contradiction. This asymmetry between users and scientific and techno-

logical information resources has become a core problem that urgently needs to be solved using advanced data technologies and service platforms [**?**].

The imbalance between scientific and technological knowledge resource supply and demand is mainly manifested in: difficulties in retrieving and discovering scientific knowledge resources, inability to precisely and proactively push massive knowledge resources to users, insufficient mining of data value, lack of design and activation of knowledge computing engines, failure to effectively form a scientific and technological knowledge flow ecosystem, inability to rapidly supply specialized data/thematic data/scientific research entity knowledge resources, and slow supply of scientific and technological knowledge resources needed for urgent topics and pressing knowledge innovation demands in the scientific community. The imbalance between scientific and technological information supply and demand is mainly reflected in: the failure to establish a scientific and technological big data center with knowledge computing as its core engine. Consequently, when facing scientific and technological information demands with different characteristics—such as emergency, thematic, and routine needs—information services still rely primarily on manual work for data source selection, data collection, data analysis, and report writing, resulting in relatively slow response speeds and very limited capacity to handle information service tasks. The "big data + big platform + expert wisdom" engineering-oriented information service model has still not been effectively formed due to data and platform limitations.

2.2 Analysis of Important Industry Development Directions

To address the above-mentioned supply-side imbalances in data, platforms, and information services, the industry has conducted relevant explorations and applications in both theory and engineering. By fully utilizing big data and AI technology, and employing text deep learning, structural analysis, knowledge object mining, and structural clustering, it has become possible to discover research design fingerprints regarding specific methods, processes, parameters, and results in scientific research, supporting the mining and comparative analysis of solutions [**?**]. Information analysis is trending toward human-machine intelligence fusion [**?**], and literature and information knowledge services are facing significant development opportunities.

(1) **Scientific and technological information institutions adopt AI technology as a core means for conducting information work.** IARPA has continuously deployed AI-related projects since 2011, using a hybrid intelligence information analysis model that highly integrates "machine intelligence + expert wisdom." Through AI technology learning from massive scientific data, it rapidly discovers scientific knowledge and potential scientific hypotheses. Corresponding technologies are directly applied to open innovation products. For example, Polypplexus [**?**] can propose research hypotheses and generate innovative ideas, while simultaneously providing an innovation environment for feasibility discussion,

demonstration, and market incubation of innovative ideas, creating a new path for problem-solving or scientific and technological innovation in an open network environment.

(2) **Professional publishing institutions leverage data advantages, use AI technology for data value-added, and launch new knowledge services.** For instance, Taylor & Francis has developed the knowledge graph tool Wizdom.ai [**?**] using machine learning, with a total data volume of 150TB, providing a knowledge computing-based full-value-chain intelligent information analysis service model. Digital Science [**?**] has developed intelligent tools from four major data dimensions—researchers, research institutions, funding projects, and publications—creating a new model of research information services for the entire research process. Elsevier [**?**] has developed digital and knowledge-based tools such as Scopus, HiveBench, Mendeley, Pure, SciVal, FundingSolutions, ExpertLookup, and AnalyticalServices, effectively meeting researchers' needs and basically covering a new research ecology from data, evidence, and tools to intelligent services.

(3) **Traditional academic evaluation analysis methods achieve breakthrough innovations based on AI technology.** For example, Semantic Scholar [**?**] uses AI technology to computationally analyze the semantic content of academic literature, automatically identifying "the most influential modern biomedical researchers" [**?**], discovering new talent evaluation models. It can also select the most important keywords and phrases from text without relying on author or publisher input, help scientists understand paper content (an area where Google search engines need improvement), identify truly influential references cited in papers, and use AI to help users filter large numbers of scientific papers and understand their content to some extent.

(4) **AI technology for fine-grained knowledge mining in professional fields has been applied.** In basic research fields such as new materials [**?**], chemistry [**?**], and physics [**?**], AI technology has created new models of information analysis services for intelligent semantic content recognition and knowledge computing. The intelligent knowledge lake [**?**] solution has gradually matured and been applied, promoting the formation of knowledge-intensive businesses. Semantic Scholar provides intelligent academic search engines and intelligent impact evaluation. Entellect [**?**] integrates drug, target, and disease data to provide computational services for life science companies using AI methods. BenchSci [**?**] achieves search speeds 24 times faster than traditional manual screening, reducing literature costs for antibody selection by 75%. IRIS.AI [**?**] uses AI to help corporate R&D departments or university researchers filter academic papers. Academic publisher Springer Nature has published the first book generated by machine learning [**?**]. Yewno [**?**] mimics the operation of the human brain, extracting literature meaning through full-text analysis and

computer semantic analysis.

(5) **New NLP technologies based on BERT pre-training models provide new methods for knowledge computing.** Due to the emergence of big data and powerful computing support, deep learning language model technologies have gradually matured and been practically applied, with machine translation approaching human levels [**?**]. In particular, BERT pre-training model-based natural language processing technologies have achieved significantly better results than traditional methods in knowledge extraction and information computing, such as the all-domain pre-training model SciBERT [**?**] and domain-specific pre-training models like BioBERT [**?**]. Domestic institutions like Harbin Institute of Technology and Baidu have trained and released models for open Chinese information data [**?**], and the National Science Library of Chinese Academy of Sciences has also developed a Chinese scientific and technological pre-training model based on the CSCD Chinese dataset [**?**].

The above demonstrates that the scientific and technological literature and information service field is indeed facing significant development opportunities, but also major challenges. There are still deficiencies in the above-mentioned advanced technologies and applications that require further breakthroughs. For example, IARPA's relevant results still require more expert wisdom intervention and participation in the semantic understandability and process interpretability of AI technology application results. While professional publishing institutions launch intelligent applications, they need to ensure data update timeliness, openness, and precision, and consider how to embed themselves into users' research spaces to prepare for future open scientific ecosystems—particularly for closed and conservative publishers who should plan ahead for new knowledge service ecological architectures. Meanwhile, AI technology, especially deep learning methods based on language pre-training models, represents a major breakthrough over traditional natural language processing technologies. Using big data and powerful computing power, it has achieved initial success in deep understanding of professional domain knowledge. However, further research and breakthroughs are still needed on how to embed domain professional knowledge into deep learning models, optimize and improve pre-training objectives according to different application scenarios, and thereby achieve more professional, precise, and comprehensive learning features and patterns.

Faced with these development opportunities and major challenges, this paper mainly conducts research design and application practice from two aspects: On the one hand, it fully utilizes the substantive breakthrough advantages brought by big data and AI technology to establish a literature and information science and technology big data center—the literature and information data lake—and studies its construction model. By opening up channels linking users with users, users with knowledge, and knowledge with knowledge, it builds an open data ecosystem that enables different user roles (academic research, management services, and decision-making) to participate in and contribute to data services at

different stages. On the other hand, it strengthens the construction of intelligent literature and information service platforms and tool systems supporting intelligent knowledge services, and enhances the construction of information research service systems oriented toward science and technology management decisions, disciplinary development, and economic and social development. This achieves an engineering and tool-oriented approach for multi-source, multi-scenario scientific and technological information services, accelerating knowledge extraction speed and improving information response capabilities. Additionally, it designs an intelligent and collaborative intelligent knowledge service ecosystem supported by "one data center, one data middle platform, and one user authentication system," establishing an open and linked sustainable data development mechanism. By providing a "big platform, small tools" approach, it effectively integrates and embeds into users' scientific and technological information service processes while also establishing a Chinese Academy of Sciences independent intellectual property rights scientific and technological information platform, ensuring autonomous control and information security of researchers' scientific activity information and scientific and technological big data.

---

3. Framework Design of the Intelligent Knowledge Service Ecosystem

3.1 Design Ideas and Methods

Big data and AI technology serve as the new engine and driving force for building the intelligent knowledge service ecosystem. With a data-driven approach, we create a "Scientific Innovation Knowledge Base" supporting scientific and technological innovation and development. By fully utilizing AI technology to build intelligent information systems, we enable information work to become flexibly operating "data cleaning plants," "information processing plants," "knowledge generation plants," and "decision-making plants" centered around intelligent information systems. This process allows information work to quickly perceive changes, refine problems, focus on targets, and form solutions, greatly compensating for human intelligence limitations and enhancing people's ability to address complex problems and tasks.

3.2 Ecosystem Framework Design

Based on the above design ideas and research methods, this paper designs an overall framework for an intelligent knowledge service ecosystem with the "Science and Technology Intelligence Brain" as the command center, as shown in Figure 1 [Figure 1: see original paper]. This creates an open ecosystem covering the entire lifecycle of data flow and knowledge flow from science and technology management and decision-making, Chinese Academy of Sciences scientific and technological innovation, other units of the scientific and technological innovation system, and the social academic information environment. The system

builds three layers around the "Science and Technology Intelligence Brain": intelligent data (i.e., scientific and technological literature and knowledge big data center), intelligent middle platform (i.e., knowledge computing platform and tool system), and intelligent services (i.e., microservices-based intelligent knowledge service platform for different application scenarios). The intelligent knowledge service platform, based on the intelligent middle platform, flexibly provides diversified and personalized service functions for multi-demand scenarios and multi-user problems across the entire ecosystem. The detailed design ideas are as follows:

3.2.1 Intelligent Data Research Design Aiming to build a national-level "Scientific Innovation Knowledge Base" supporting scientific and technological innovation and to support the transformation and upgrading to knowledge computing-based data services, this paper's intelligent data architecture designs a scientific and technological literature basic database and knowledge big database (scientific innovation domain knowledge base and scientific innovation knowledge graph) to provide multi-level and in-depth intelligent data services for deep learning model training and knowledge organization computing in the intelligent middle platform.

(1) **Construct a comprehensive, authoritative, and timely scientific and technological literature basic database**, namely the Scientific Innovation Basic Database (see Figure 2 [Figure 2: see original paper]): From five dimensions—research subjects (experts, scholars, research institutions, academic journals, research teams, publishing platforms, technology enterprises, funding agencies), research activities (research projects, academic conferences, training exchanges, technology competitions, data sharing, news information, social activities, science and technology policies), research achievements (papers, patents, reports, awards, monographs, standards, software, products, data), research facilities (large scientific facilities, instruments, consumables, research methods), and scientific data (research data, etc.)—we have built the "Scientific Innovation Basic Knowledge Base" through aggregation and fusion, with deep indexing from discipline classification, industry classification, theme classification, and STKOS (Science and Technology Knowledge Organization System) category classification, providing high-quality foundational data for knowledge classification computing.

(2) **Intelligently build domain knowledge bases through content mining and fine-grained knowledge recognition**: Based on the Scientific Innovation Basic Database, using BERT pre-training model-based new NLP technology methods, we have demonstrated the construction of domain knowledge graphs in two fields: artificial intelligence and chemical bond energy. The artificial intelligence domain knowledge base includes four types of fine-grained knowledge: research problems, research methods, research data, and implementation indicators. The chemical bond

energy domain knowledge base includes seven types of fine-grained knowledge: compounds, solutions, methods, PKA, PKA-VALUE, Bond, and reactions.

(3) **Intelligently construct a scientific innovation knowledge graph through knowledge association computing** (see Figure 3 [Figure 3: see original paper]): On the basis of the Scientific Innovation Basic Database, we have formulated data fusion and association rules, using big data and AI technology to govern, integrate, and associate multi-source heterogeneous scientific and technological resources, connecting various innovation entities and innovation resource entities (papers, journals, scholars, institutions, projects, themes, etc.). We have built a scientific innovation knowledge graph with 21 types of knowledge relationships and over 10 billion knowledge relationships. The "Scientific Innovation Knowledge Base" has achieved a breakthrough from traditional expert think tank-based decision support models to machine intelligence-based knowledge computing models, providing scientific methods for solving scientific problems. As essential infrastructure supporting data-intensive innovation, the Scientific Innovation Knowledge Base can perform knowledge computing, association reasoning, and deep mining of innovation elements, thereby promoting data sharing and deepening data application, which will have a positive impact on China's scientific and technological innovation. Currently, the Scientific Innovation Knowledge Base has played an important exemplary role in talent identification, institutional evaluation, project assessment, technology analysis, and innovation discovery.

3.2.2 Intelligent Middle Platform Research Design The construction goal of the intelligent middle platform is to build a new-generation knowledge intelligence technology tool platform, providing rich knowledge intelligence analysis tools integrated and encapsulated through microservices to uniformly provide technical support and data knowledge output for multi-scenario, multi-demand intelligent service applications in the upper layer, thereby improving the quality and efficiency of knowledge service product construction. The research and development mainly focus on three types:

(1) **Middle platform services centered on knowledge organization tools**: Facing various resources and data in the 全院 knowledge service big data center system, we establish a knowledge organization application system, strengthen technical methods and tool systems for semantic enhancement, and achieve semantic enrichment and mining of knowledge content in data, such as pattern mining, relationship reasoning, user profiling, community computing, and team identification.

(2) **Middle platform services centered on knowledge computing tools**: Combining blockchain, intelligent mining, social computing, deep learning, and other technologies, we encapsulate knowledge mining and intelligent analysis-related tools and models. Through loosely coupled

methods, we build business-oriented application systems, accelerate the construction speed of knowledge computing application systems by quickly configuring resources and accessing existing technologies, reduce human and resource costs, and provide solid guarantees for rapid business innovation: intelligent analysis such as semantic retrieval, intelligent recommendation, predictive visualization, technology mining, and topic evolution.

(3) **Middle platform services centered on professional domain knowledge organization and information computing analysis**: Specialized, vertical knowledge services are necessary ways to support scientific and technological innovation. Focusing on domain characteristics, we expand and condense tools for specific professional domains or data resources, such as extraction and recognition of domain entities or concepts, explore professional knowledge organization tools for the research process, provide domain knowledge fusion services, and discover and develop domain-related technologies, methods, and tools to more effectively support researchers' scientific activities and provide more comprehensive knowledge services such as domain strategic layout, competitiveness analysis, development trend analysis, research opportunity discovery, and research hotspot analysis.

3.2.3 Intelligent Services Research Design With a data and service scenario-driven intelligent service design concept, we achieve seamless integration and embedding of key knowledge service functional components, support intelligent navigation and guidance of user needs, and provide both proactive discovery of user needs and intelligent recommendations, as well as personalized exploration of knowledge in different application scenarios to solve problems. Specifically, we mainly provide four application scenarios for four types of user roles, with unified authentication, integrated data resource management, and synchronized messaging mechanisms among scenarios, ensuring continuous knowledge service provision according to users' associative thinking to a certain extent. The detailed functional implementation effects are described in Section 4.

(1) **AI data services for knowledge management**: For research institution objects, we achieve active and precise distribution of knowledge achievements, precise institutional profiling, and real-time institutional information analysis, providing functions such as institutional profiling, achievement management, project management, talent management, expert management, activity management, intelligent project review, intelligent fund topic selection, institutional competitiveness comparative analysis, and institutional research dynamics analysis.

(2) **AI integrated services for knowledge discovery**: For public users, we provide intelligent retrieval and discovery services, including knowledge retrieval, precise push, book retrieval, full-text acquisition, AI con-

sultation, academic cards, data sharing, academic exchange communities, cultural communication, and intelligent tools (research travel, paper format review, intelligent topic selection, project evaluation, scientific and technological novelty search, etc.).

(3) **AI analysis services for information analysis**: For information analysts, we provide data management and intelligent analysis services, including thematic trend analysis, disruptive technology identification services, competitive situation analysis, institutional analysis, talent attraction analysis, discipline analysis, and industry analysis.

(4) **Leadership dashboard for scientific and technological decision-making**: For decision-makers, we provide dynamic scanning and perception analysis services for global scientific research achievements, including global achievement distribution, global talent distribution, and comparative analysis of global institutional scientific and technological capabilities.

---

4. Application and Practice

Based on the intelligent knowledge service ecosystem architecture designed in this paper, the National Science Library of Chinese Academy of Sciences has conducted comprehensive practical application in the "13th Five-Year Plan's transformation and upgrading of scientific and technological knowledge services," achieving obvious results in automatic data collection, automatic aggregation and fusion, intelligent knowledge computing, precise services, and multi-dimensional data profiling. This has laid a solid data, technology, and platform foundation for library and information science to support knowledge services. Specifically, six areas have achieved phased application results: First, we constructed the "Literature and Information Data Lake" and "Literature and Information Knowledge Service Engine Tools," providing rich data knowledge and algorithm tools at the bottom layer of intelligent knowledge services. Second, we developed an intelligent knowledge service platform for knowledge discovery, an institutional digital asset management and analysis platform for knowledge management, and an intelligent analysis platform for thematic information data management, providing tool platforms for universal knowledge discovery, knowledge management, and information analysis. Finally, we designed a leadership dashboard for scientific and technological decision-making.

4.1 Implementation of AI+ "Literature and Information Data Lake" and Knowledge Service Infrastructure

We have built a global scientific and technological big data knowledge resource, forming a "Scientific Innovation Knowledge Base" for scientific and technological innovation, including global scientific and technological literature, patent data, scientific and technological talents, project data, graph data, and monitoring data. We have also established the National Digital Resource Long-term

Preservation Center [**?**], achieving full-text local preservation of 65 foreign resources and 3 domestic resources, providing strategic resource guarantees for scientific and technological innovation.

Key features include: The data governance and computing intelligence degree of the "Literature and Information Data Lake" (see Figure 4 [Figure 4: see original paper]) reaches over 90%, achieving intelligent standardization of institution names, intelligent classification, intelligent summarization, and intelligent extraction of keywords. We have established a human-machine fused, standardized, precise, and structured "data lake" cloud governance service platform [**?**], enabling online real-time governance of 19 types of data entities. We provide rapid thematic database construction functions, supporting the quick extraction and aggregation of authoritative thematic databases from the literature and information "data lake" according to preset thematic knowledge structures, such as the Novel Coronavirus Thematic Knowledge Service and Research Collaboration Platform [**?**]. Using big data and AI technology, we have built an academic knowledge graph [**?**] including 400 million+ dataset records, 1 billion+ scientific and technological entities, and 10 billion+ knowledge relationships, forming a national-level Scientific Innovation Knowledge Base that can support scientific and technological innovation. We have built 200 industry domain knowledge bases [**?**] to support information monitoring and scientific and technological decision-making. Data security, software security, and platform security are currently important issues. In our literature and information "data lake" and knowledge infrastructure construction, infrastructure software such as distributed storage, distributed computing, and distributed indexing are primarily self-developed.

4.2 Construction of AI+ Literature and Information Knowledge Service Engine and Tools

Based on the "Literature and Information Data Lake" + AI technology, and fully leveraging the advantages of large-scale scientific and technological literature datasets, we have developed a scientific and technological literature AI service engine using BERT pre-training models.

Key features include: Using deep learning technology on massive metadata from the "Literature and Information Data Lake," we have trained a series of knowledge computing service engines [**?**] (see partial screenshots in Figure 5 [Figure 5: see original paper] and Figure 6 [Figure 6: see original paper]), including literature classification, keyword recognition, concept sentence recognition, text tag generation, reviewer recommendation, named entity recognition, intelligent identification of technologies and problems, and intelligent standardization of entity names, supporting semantic mining research and exploratory discovery. These results were officially released on December 5, 2020. Using the above knowledge service engines and further applying deep learning technology based on the "pre-training + fine-tuning" NLP pre-training model pattern, we have formed a methodological system for professional domain knowledge graph con-

struction. In the artificial intelligence domain [**?**] and chemical bond energy domain, we have achieved fine-grained knowledge identification and extraction to form domain knowledge graphs. For example, we can intelligently identify fine-grained knowledge of research problems, research methods, research data, and research indicators in the artificial intelligence domain, and automatically identify scientific data and construct databases from full-text literature in the chemical bond energy domain.

4.3 Development of AI+ Intelligent Knowledge Service Platform for Knowledge Discovery

We have achieved a series of functions from multi-source data aggregation, integrated knowledge retrieval and discovery, intelligent summarization, thematic analysis, precise proactive knowledge push, academic cards, and automatic navigation (download/document delivery/reference consultation/thematic analysis/graph discovery scenarios), as well as an integrated service and operation system supporting PC, WeChat, and APP multiple terminals—forming the main service portal for literature and information services [**?**]. Updated and launched on November 11, 2020, December 2020 data shows that the average daily 访问量 is more than three times that of the old service system (see Figure 7 [Figure 7: see original paper]).

Key features include: Unified retrieval and association discovery of multi-type knowledge based on academic knowledge graphs, from unified retrieval of paper, patent, and report literature data, to unified retrieval of scholar, institution, and theme data, and then to association discovery linking a piece of literature to themes, scholars, literature, and institutions (relevant data association discovery is being added), breaking through traditional functions that only provide metadata display. Intelligent summarization functions for massive retrieval results, enabling quick understanding of retrieval result knowledge. Calculation of "semantically related themes" for retrieval objects, enabling intelligent analysis by theme to quickly understand development trends and annual core papers of the theme. Provision of user ID authentication download service integration, supporting users to download full-text papers anytime, anywhere according to their rights. Seamless integration of all functions of "Huikexue - Intelligent Portable Research Assistant" [**?**] (see Figure 8 [Figure 8: see original paper]), especially providing features such as "automatic creation of scholar academic cards, personal academic achievement verification and management, automatic push of high-value knowledge, personalized knowledge subscription, open data sharing, and innovation communities for rapid frontier problem solving and idea collision," as well as intelligent functions such as intelligent project selection for fund project guidelines, research travel, scientific and technological novelty search, intelligent evaluation of application projects, and intelligent review of paper formats, striving to create the "Today's Headlines" for the scientific community. The platform seamlessly integrates other tool platforms, expert teams, and data resources of the National Science Library, achieving one-stop resource

discovery and acquisition on one platform. The platform also provides a user unified authentication center, supporting not only full-text downloads based on ID accounts according to institutional subscription rights but also enabling cross-system single sign-on in the main service system.

4.4 Development of AI+ Institutional Digital Asset Management and Analysis Platform for Knowledge Management

Based on the "Literature and Information Data Lake" and using big data and AI technology, we have developed an institutional digital asset management and analysis platform for knowledge management (https://inst.scholarin.cn/). This platform provides comprehensive, multi-perspective analysis and evaluation of institutions, assisting institutional managers in managing institutional knowledge assets (see Figure 9 [Figure 9: see original paper]). The platform also provides objective data support for institutional research achievement recognition and personnel identity verification, effectively contributing to academic ethics to a certain extent.

Key features include: Intelligent and precise distribution of research achievements by institution dimension, automatically constructing institutional research knowledge resource databases, breaking through traditional data management methods that rely on manual data entry. Real-time institutional panoramic profiling functions, providing real-time panoramic profiles from dimensions such as institutional research achievement output trends, cooperative institution networks, high-output scholars, high-output patent inventors, high-output project leaders, paper inclusion statistics, and annual research hotspot themes. Research team management functions, providing real-time analysis from dimensions such as professional titles, age, gender, departments, research directions, and expert teams, with online data management functions.

Full-process project management functions, providing data management services for the entire process from project application, approval, initiation, mid-term review, to completion. Research achievement analysis functions, real-time perception and analysis of institutional research achievement type distribution, output trends, real-time statistics of SCI and other inclusion data, and retrieval and discovery functions. It supports integrated management functions for localized research achievement knowledge assets, enabling batch upload of localized knowledge assets according to standardized data formats for integrated management, thereby revitalizing localized knowledge assets. It also provides management of institutional subscription database usage data and user authorization management.

4.5 Construction of Topic Intelligence Data Management and Intelligent Analysis Platform Using a Data-Driven Approach

Based on the "Literature and Information Data Lake" and using big data and AI technology, we have developed an "AI+ Intelligent Analysis Platform for Information Analysis" (http://ai.scholarin.cn/) (see Figure 10 [Figure 10: see

original paper]) using a "data + platform + expert" hybrid intelligence model, attempting to solve problems of data dispersion, inability to accumulate and reuse, inability to share, and inability to quickly analyze and provide data services.

Key features include: Supporting rapid integration of expert wisdom in a platform-based manner to quickly construct knowledge structures for thematic information directions, laying a foundation for obtaining more precise and comprehensive analysis datasets for subsequent information analysis. For a thematic information analysis scenario, providing functions to import local datasets and standardized data (countries, institutions, names, keywords, etc.) into the platform, achieving integrated management of online and offline data to meet goals of data accumulation, reuse, and centralized management. Providing online data automatic cleaning functions based on standardized database data from the literature and information data lake. Providing statistical analysis service functions based on traditional scientometrics, including development trend analysis, institutional analysis, regional analysis, journal analysis, keyword analysis, and hotspot analysis for paper data; development trend analysis, technology analysis, patentee analysis, inventor analysis, regional analysis, keyword analysis, and research hotspot analysis for patent data; and development trend analysis, funding unit analysis, principal investigator analysis, undertaking institution analysis, regional analysis, keyword analysis, and research hotspot analysis for fund project data. Exploring and attempting content analysis services based on semantic computing: research problem analysis, key technology analysis, problem-technology association analysis, and research hotspot technology identification, providing rapid generation of "data-based information analysis reports" to assist information analysts, accelerate information production speed, and improve response speed to research information demands.

4.6 Creation of AI+ Leadership Dashboard for Scientific and Technological Decision-Making

We have designed and created an immersive perception environment for decision analysis (see Figure 11 [Figure 11: see original paper]), integrating core technologies such as the literature and information data lake, intelligent knowledge engine, and visualization. It can perform image recognition, intelligent voice interaction, information dynamic perception visualization, and scientific and technological information dynamic perception, providing more intuitive, convenient, precise, and intelligent environments for scientific and technological decision-making and information analysis and judgment, offering new models and methods for scientific and technological decision-making.

---

5. Conclusion and Outlook

Currently, big data and AI technology have brought significant development opportunities for industry and economic development, as well as opportunities for transformation and upgrading of scientific and technological knowledge service models, while also presenting many challenges.

Against this background, this paper reviews current applications and practices of big data and AI in knowledge services, combines development opportunities in the literature and information field, proposes an architecture for building an AI+ intelligent knowledge service ecosystem, conducts in-depth research design and practice in data aggregation, knowledge computing, tool development, and knowledge service platform development, and publicly releases six major intelligent knowledge service systems. These have been practically applied to the Chinese Academy of Sciences, provincial academies of sciences, and some research institutions, receiving widespread user praise and laying a good foundation for the future formation of a literature and information knowledge service ecosystem. In the future, this research will continue to focus on the AI+ intelligent knowledge service ecosystem architecture, further improving and enhancing intelligent knowledge service capabilities and levels in terms of data precision, knowledge depth, service professionalism, information timeliness, and the intelligence degree of the intelligent knowledge service system.

Acknowledgments

The research results of this paper have received strong support from the Literature and Information Capability Special Breakthrough Project. We especially thank relevant team members and all colleagues at the center for their support, and thank the information system department's product design, data computing, system development, and network support personnel.

References

[1] Zhang Dongrong, Qian Li. Special Topic on "Construction of Scientific and Technological Big Data and Intelligent Knowledge Service Platform" [J]. Data Analysis and Knowledge Discovery, 2019, 3(1): 3-3. [2] Ke Ping, Zou Jinhui. Library Transformation in the Post-Knowledge Service Era [J]. Journal of Library Science in China, 2019, 45(1): 4-17. [3] Luo Liqun, Li Guangjian. Intelligent Information Services and Knowledge Fusion [J]. Information and Documentation Services, 2019, 40(2): 87-94. [4] Su Xinning. Reflections on the Mission of Libraries in the New Era and Future Library Science Education [J]. Journal of Library Science in China, 2020(1): 53-62. [5] Zhang Xiaolin. Disruptive Changes and the Post-Library Era: Promoting Supply-Side Structural Reform of Knowledge Services [J]. Journal of Library Science in China, 2018, 44(1): 4-16. [6] Qian Li, Xie Jing, Chang Zhijun, et al. Research Design of Intelligent Knowledge Service System Based on Scientific and Technological Big Data [J]. Data Analysis and Knowledge Discovery, 2019, 3(1): 4-14. [7]

Qian Li, Zhang Xiaolin, Wang Qian. Research on Research Design Fingerprint Description Framework Based on Scientific and Technological Literature [J]. Journal of Academic Libraries, 2015, 33(1): 14-20. [8] Li Guangjian, Jiang Xinyu. On Computational Information Analysis [J]. Journal of Library Science in China, 2018, 44(2): 4-16. [9] About us - polyplexus.com [EB/OL]. [2020-10-06]. https://start.polyplexus.com/about-us/. [10] wizdom.ai - intelligence for everyone [EB/OL]. [2020-10-06]. https://www.wizdom.ai/#about. [11] Digital Science [EB/OL]. [2020-10-06]. https://www.digital-science.com/. [12] Research Intelligence [EB/OL]. [2020-10-06]. https://www.elsevier.com/research-intelligence. [13] A free, AI-powered research tool for scientific literature [EB/OL]. [2020-10-06]. https://www.semanticscholar.org/. [14] Who's the most influential biomedical scientist? Computer program guided by artificial intelligence says it knows | Science | AAAS [EB/OL]. [2020-10-06]. http://www.sciencemag.org/news/2017/10/who-s-most-influential-biomedical-scientist-computer-program-guided-artificial. [15] Tshitoyan V, Dagdelen J, Weston L, et al. Unsupervised word embeddings capture latent knowledge from materials science literature [J]. Nature, 2019, 571(7763): 95-98. [16] Kim E, Huang K, Saunders A, et al. Materials synthesis insights from scientific literature via text extraction and machine learning [J]. Chemistry of Materials, 2017, 29(21): 9436-9444. [17] Zhou Q, Tang P, Liu S, et al. Learning atoms for materials discovery [J]. Proceedings of the National Academy of Sciences, 2018, 115(28): E6411-E6417. [18] Kiyohara S, Miyata T, Tsuda K, et al. Data-driven approach for the prediction and interpretation of core-electron loss spectroscopy [J]. Scientific Reports, 2018, 8(1): 1-12. [19] Stokes J M, Yang K, Swanson K, et al. A deep learning approach to antibiotic discovery [J]. Cell, 2020, 180(4): 688-702. [20] Segler M H, Preuss M, Waller M P. Planning chemical syntheses with deep neural networks and symbolic AI [J]. Nature, 2018, 555(7698): 604-610. [21] Granda J M, Donina L, Dragone V, et al. Controlling an organic synthesis robot with machine learning to search for new reactivity [J]. Nature, 2018, 559(7714): 377-381. [22] Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design [J]. Science Advances, 2018, 4(7): 7885. [23] Radovic A, Williams M, Rousseau D, et al. Machine learning at the energy and intensity frontiers of particle physics [J]. Nature, 2018, 560(7716): 41-48. [24] Mathuriya A, Bard D, Mendygral P, et al. CosmoFlow: using deep learning to learn the universe at scale [C]//SC18: International Conference for High Performance Computing. Dallas Texas: IEEE, 2018: 819-829. [25] Zhang Y G, Gajjar V, Foster G, et al. Fast radio burst detection and periodicity: a machine learning approach [J]. The Astrophysical Journal, 2018, 866(2): 149. [26] Beheshti L A, Benatallah B, Sheng Q Z, et al. Intelligent knowledge lakes: the age of artificial intelligence and big data [C]//WISE2020. Singapore: Springer, 2020. [27] Connecting life science data to enable impactful analytics - Entellect [EB/OL]. [2020-10-06]. https://www.elsevier.com/solutions/entellect. [28] AI-assisted reagent selection and experiment design - BenchSci [EB/OL]. [2020-10-06]. https://www.benchsci.com/. [29] Your science assistant [EB/OL]. [2020-10-06]. https://iris.ai/. [30] Academic publisher Springer Nature published the first book authored by AI [EB/OL]. [2020-10-06].

http://m.elecfans.com/article/906694.html. [31] Transforming information into knowledge [EB/OL]. [2020-10-06]. https://www.yewno.com/. [32] Google's neural machine translation system: bridging the gap between human and machine translation [EB/OL]. [2020-10-06]. https://arxiv.org/pdf/1609.08144.pdf. [33] SciBERT: a pretrained language model for scientific text [EB/OL]. [2020-10-06]. https://arxiv.org/abs/1903.10676. [34] BioBERT: a pre-trained biomedical language representation model [EB/OL]. [2020-10-06]. https://arxiv.org/abs/1901.08746. [35] Pre-training with Whole Word Masking for Chinese BERT [EB/OL]. [2020-10-06]. https://arxiv.org/abs/1906.08101. [36] Chinese tasks comprehensively surpass BERT: Baidu officially releases NLP pre-training model ERNIE [EB/OL]. [2020-10-06]. https://zhuanlan.zhihu.com/p/59436589. [37] Pre-training model based on scientific and technological literature [EB/OL]. [2020-10-06]. http://sciengine.las.ac.cn/. [38] Wang Ying, Qian Li, Xie Jing, et al. Research on Construction Model and Method of Scientific and Technological Big Data Knowledge Graph [J]. Data Analysis and Knowledge Discovery, 2019(1): 15-26. [39] National Digital Science and Technology Literature Resource Long-term Preservation System [EB/OL]. [2020-10-06]. http://ndpp.ac.cn/. [40] Scientific and Technological Big Data Multi-source Knowledge Fusion Governance Cloud Platform [EB/OL]. [2021-02-16]. http://data.scholarin.cn. [41] Scientific and Technological Big Data Multi-source Knowledge Fusion Governance Cloud Platform [EB/OL]. [2020-10-06]. http://data.scholarin.cn/. [42] Novel Coronavirus Thematic Knowledge Service and Research Collaboration Platform [EB/OL]. [2020-10-06]. https://ncov.scholarin.cn/. [43] Big Data Discovery System [EB/OL]. [2020-10-06]. http://kg-view.las.ac.cn/discover. [44] Domain Scientific and Technological Information Monitoring Service Cloud Platform [EB/OL]. [2020-10-06]. http://stmcloud.las.ac.cn/. [45] AI Engine Based on Scientific and Technological Literature Knowledge Resources [EB/OL]. [2020-10-06]. http://sciengine.las.ac.cn/. [46] Deep Semantic Mining Tool [EB/OL]. [2020-10-06]. http://finger.las.ac.cn/Page_{{{sharingTool}}*{{tm}}}*{NER}.html. [47] Knowledge Service Portal of National Science Library, Chinese Academy of Sciences [EB/OL]. [2020-10-06]. http://www.las.ac.cn. [48] Huikexue - Intelligent Portable Research Assistant [EB/OL]. [2020-10-06]. http://scholarin.cn.

Author Contributions:

Qian Li: Paper writing and revision, development of intelligent knowledge service platform.

Liu Xiwen: AI+ intelligent knowledge service ecosystem architecture design, paper content revision and guidance.

Zhang Zhixiong: Construction and guidance of literature and information data lake and knowledge service engine, paper content revision and guidance.

Liu Huizhou: Proposed the construction content of AI+ intelligent knowledge service platform, and revised and guided the paper content.

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