Think Tank Information Organization Strategies and Their Challenges in the Big Data Environment: Postprint

Authors: An Nan, Zhu Zhongming

Date: 2017-11-05T00:00:00+00:00

Abstract

[Purpose/Significance] High-level new-type think tanks rely on sophisticated information support mechanisms. In the era of big data, traditional information organization mechanisms for think tanks can no longer accommodate current data characteristics and decision-making requirements, making the construction of knowledge bases that support the decision-making process an inevitable trend in think tank development.

[Method/Process] This study selects foreign think tank institutions with reference value from the "Global Think Tank Report 2016" as research subjects, employs literature review and case analysis methods to summarize common information organization approaches in current think tanks, analyzes the data value chain under big data and its requirements for organizational processes, and accordingly proposes the necessity of constructing think tank knowledge bases.

[Results/Conclusion] The study ultimately proposes a general decision-making-process-oriented framework for think tank knowledge bases and constructs a knowledge organization model within the knowledge base using semantic ontology methods, aiming to provide reference for think tanks to gradually achieve semi-automated to automated decision research processes in the big data environment.

Full Text

Think Tank Information Organization Strategies and Their Challenges in the Big Data Environment

An Nan[1,2], Zhu Zhongming[1]

[1]University of Chinese Academy of Sciences, Beijing 100049, China

Abstract

[**Purpose/Significance**] High-level new-type think tanks cannot function without sophisticated information support mechanisms. Against the backdrop of the big data era, traditional think tank information organization mechanisms have become inadequate for current data characteristics and decision-making requirements. Building knowledge repositories that support the decision-making process has become an inevitable trend in think tank development. [**Method/Process**] This study selected foreign think tank institutions with reference value from the *2016 Global Go To Think Tank Index Report* as research subjects, applying literature review and case analysis methods to summarize common information organization approaches currently used by think tanks. The study analyzed the data value chain under big data and its requirements for organizational processes, and accordingly proposed the necessity of constructing think tank knowledge repositories. [**Result/Conclusion**] Finally, a general decision-making-process-oriented knowledge repository framework for think tanks is proposed, and a knowledge organization model within the repository is constructed using semantic ontology methods, aiming to provide references for think tanks to gradually achieve semi-automated to automated decision research processes in the big data environment.

Keywords: think tank; knowledge repository; big data; information organization; organizational strategy; decision research; ontology

Classification Number: G359

Think tanks are research, analysis, and engagement institutions for public policy that conduct policy-oriented research, analysis, and consultation on domestic and international issues, enabling policymakers and the public to make decisions based on reliable information [1]. Their primary function is to provide decision-makers with timely, comprehensive, and accurate supporting information. The scope, quantity, quality, service content, and service methods of this supporting information directly affect the effectiveness of decision-making [2]. Therefore, possessing a well-developed information support mechanism is crucial for think tanks to produce high-quality decision consultation outcomes.

The big data era is characterized by the 4V features of information: massive volume, diverse formats, rapid update speed, and low value density [3][4]. In this context of data explosion, all research processes exhibit a data-driven trend. The ability to promptly discover and extract valuable knowledge from vast amounts of information will become the key factor influencing the efficiency of think tank decision research processes and outputs.

Big data-driven think tank decision research must address two fundamental problems: First, how to construct a unified data model that enables any big data resource to be processed into intelligent data that can support decision research, gradually achieving semi-automated to automated decision research processes. Second, how to semantically process relevant information from various sources and in various forms for the decision research process, strengthening data interconnections to enhance knowledge discovery capabilities and provide decision-makers with more valuable policy reference information. Therefore, this paper attempts to construct a multi-source heterogeneous data integration framework that supports big data intelligence processing and analysis, providing a unified conceptual model for building intelligent datasets.

2 Current Research on Think Tank Information Organization Mechanisms

How think tanks organize and process data content directly affects the efficiency with which researchers and intelligence experts utilize information resources. Scientifically rational organization methods not only improve data access efficiency but also help uncover potential value information in data, generating added value.

2.1 Development Status of Foreign Think Tank Information Organization Mechanisms

Modern think tanks first emerged in Western countries during World War II. Compared with China, Western countries have developed relatively mature systems in both think tank research and think tank self-construction, making influential Western think tanks more representative as research subjects. Based on the comprehensive ranking and various field rankings in the University of Pennsylvania' s *2016 Global Go To Think Tank Index Report*, this study selected over a dozen top-ranked foreign think tanks as research subjects (see), analyzing their information organization strategies through investigation of how information resources are presented on their official websites and through various types of think tank products available.

TABLE:1 lists the foreign think tanks involved in this research, including Bruegel, SIPRI, WRI, CEIP, RAND, CIA, SWP, Hoover Institution, Brookings Institution, Cato Institute, Chatham House, Max Planck Society, and JIIA.

Through comparative investigation and analysis, this study summarized common information collection and organization strategies among Western think tank institutions (see [Figure 1: see original paper]) and assessed the current development status of their information support mechanisms. In terms of information collection strategies, manual collection and semi-automatic collection requiring significant human participation remain dominant. Among these, collecting publicly available data has become one of the most commonly used information collection methods due to its strong operability, broad data scope, and relatively low cost. Almost all the aforementioned think tanks consider obtaining public data via the Internet as the most conventional approach to data collection. Additionally, due to the real-time and novel nature of think tank research, think tanks often have special data requirements or involve specific projects such as war situations, behavioral science, or drug pathology, where no completely applicable data exists or previous data has limited reference value. Therefore, think tank researchers must supplement indirect collection methods with direct production and creation approaches, among which literature surveys have become the most frequently used direct data acquisition method due to their low cost and ease of implementation. Traditional think tanks such as the Brookings Institution, Hoover Institution, and Carnegie Endowment for International Peace are typical representatives in the application of traditional research methods. Of course, during the research process, think tank experts often do not confine themselves to a specific method but rather apply them flexibly and interchangeably.

Based on content organization forms, collected information resources can be organized into three types: databases, information retrieval systems, and knowledge repositories (see [Figure 1: see original paper]).

2.1.1 Databases (Datasets) One way to organize data collected directly or indirectly by think tanks is to structure it into databases or datasets. The advantages of this structured data format include ease of management, high shareability, low redundancy, and easy expansion.

Bruegel, a think tank focusing on international economic policy research, provides open access to seven specialized datasets on policy economics [5], including: (1) global and regional Gini coefficients; (2) Divisia indices for euro area monetary aggregates; (3) real effective exchange rates for 178 countries; (4) European enterprises in the global economy: internal policies under external competition; (5) sovereign bond holdings; (6) Eurosystem liquidity; and (7) regression-based record linkage on PATSTAT applications. The Stockholm International Peace Research Institute (SIPRI) is world-renowned for its authoritative assessments of global security issues. All of SIPRI' s research bases and sources are completely open, making its research results an authoritative source frequently used by international politicians, researchers, and media personnel. SIPRI maintains four specialized databases: (1) multilateral peace operations database; (2) military expenditure database; (3) arms transfers database; and (4) arms industry database. Additionally, SIPRI comprehensively maintains datasets on arms control and disarmament [6], including arms embargo reports, national arms reports, global arms trade value reports, etc. These specialized databases provide strong information support for SIPRI' s research activities.

2.1.2 Information Retrieval Systems The sophistication of an institution' s information retrieval system directly reflects the quality of its information organization. For think tanks, a powerful retrieval system not only provides efficient data support for internal research experts but also enables users to quickly and accurately obtain required information.

The Carnegie Endowment for International Peace provides a simple and user-friendly site search system where users can choose exact matching or any matching to search titles, author names, or full text. Search results can be further filtered by document type, publication year, region, topic, and project, and sorted by date or relevance. RAND Corporation' s retrieval system has relatively comprehensive functionality, allowing users to limit searches through various conditions including keyword matching, additional attributes, and document characteristics. Additional attributes include page titles, affiliated RAND departments, content types, and start dates, while document characteristics cover titles, authors, topics, ISBNs, etc., for rapid location of relevant resources. The U.S. Central Intelligence Agency (CIA), as one of the world' s most famous intelligence agencies, strives to scientifically manage massive amounts of intelligence to maximize its effectiveness, making it a model for information resource management and value-added services among intelligence agencies. When digitizing archival materials, the CIA's declassified archives retrieval system employs metadata methods (see). Unified metadata standards scientifically categorize massive information resources and enable the organic integration of different media types such as text, audio, and video, allowing unified retrieval within the same access system and greatly improving information utilization efficiency.

TABLE:2 shows the metadata used in the CIA' s declassified document retrieval system, including document type, collection, document number, release decision, page count, original classification, file attachment, sequence number, case number, and publication date.

2.1.3 Knowledge Repositories In an information environment, a knowledge repository can be defined as a knowledge collection established by an organization for specific application purposes (such as supporting research, education, or management processes). Generally, there are two basic types of knowledge repositories: domain/thematic knowledge repositories and institutional knowledge repositories. The former collects, organizes, and disseminates knowledge content in specific disciplinary fields or topics, while the latter primarily provides services for preserving and managing knowledge output produced by an institution. As a mechanism for storing, organizing, and managing digital knowledge, knowledge repositories have been widely applied in scientific research fields, but their application in decision-making consultation institutions such as think tanks is still immature. A considerable number of think tanks, due to funding, resources, or lack of awareness, have not yet built internal knowledge repositories and remain at the stage of information "storage repositories."

Building on previous research [7], this study conducted a comparative analysis of the knowledge repository construction processes of RAND Corporation and the German Institute for International and Security Affairs (SWP) from three aspects: collection development, intelligence gathering, and technical support (see). The analysis reveals that both RAND and SWP attach great importance to information resource development, with rich internal collections and extensive database content. In terms of knowledge organization, both adopt classification-based approaches, establishing thematic knowledge repositories according to research topics while also using regions as a basis for research project classification. Both think tanks have developed information support systems through technical means and actively attempt to cooperate with other institutions and departments for information resource sharing and collaborative construction to compensate for their own professional deficiencies while reducing data redundancy.

TABLE:3 presents the comparison between RAND and SWP knowledge repositories, showing their internal libraries, project reports, personnel departments, sub-institutions, and developed systems.

2.2 Challenges of the Big Data Era to Think Tank Information Organization

The investigation reveals that although current global think tank information support mechanisms have developed comprehensively, many deficiencies remain. On one hand, information organized in database/dataset form is relatively independent. Even though databases may broadly categorize stored information by theme or other characteristics, at finer granularities the data remains independent and lacks necessary interconnections, which is not conducive to potential knowledge discovery during think tank data mining and analysis. Additionally, this independent data organization approach lacks context-based connections, resulting in weak contextualized and situational knowledge application—that is, limited ability to understand the same concept in different contexts.

On the other hand, among think tanks that organize information through information retrieval systems, most have only annotated the extension of information without conducting deeper semantic annotation of information content. This is not conducive to computer understanding of data information or the development of automated think tank decision-making processes. Under the big data era' s increasing demands for think tank response speed to international situations, this has clearly hindered decision output efficiency. To transform meaningless data in big data into intelligent data that can support decision research, various types of information must be processed from unstructured, independently existing coarse-grained data into structured, computer-operable, interrelated, context-aware fine-grained data.

McKinsey & Company first proposed the arrival of the "big data" era. In this era, various digital resources are growing exponentially and gradually becoming the mainstream of information resources. Faced with the 4V characteristics of big data, traditional think tank information support mechanisms can no longer efficiently process such massive amounts of heterogeneous data. How to effectively obtain valuable information from complex data, how to scientifically manage and organize collected massive information, and thereby provide users with rapid and accurate services—these requirements demand that new-type think tanks must promptly adjust their data collection, storage, and organization strategies to adapt to current big data characteristics.

T. Gustafson and D. Fink proposed the concept of "big data value chain" in 2013

[8], arguing that every big data value chain should at least consist of four basic stages: data acquisition, data storage, data analysis, and data application. As knowledge organization institutions, think tanks' decision research and output processes are essentially knowledge value-added processes. Based on this, this paper proposes the think tank data value chain in the big data environment (see [Figure 2: see original paper]), which reflects the data-centered activities in various stages of think tank decision research and output, while big data poses requirements for each stage.

Unlike commercial research data value chains that rely on computer-automated extraction, processing, and analysis of big data to obtain results, the knowledge value-added process in the think tank data value chain is based on the full processing and organization of big data in the former stage, which then serves as supporting data for decision research to think tank experts and, together with their tacit knowledge, finally forms think tank products. In this data-driven decision process, the timeliness, comprehensiveness, accuracy, and degree of organization of the supporting data provided to think tank experts will directly affect the efficiency and quality of final think tank product output.

FIGURE:2 illustrates the data value chain of think tanks in the context of big data. The figure shows that big data has impacted all aspects of traditional think tank research processes, with the most critical being the data organization stage in the middle. As an intermediate link connecting preceding and following stages, think tank data organization must adapt to the complex data types collected from various sources while also preparing for subsequent storage and data analysis technologies and tools. In summary, think tanks in the big data era need an information support system that can effectively organize and manage captured and preserved information through management methods and information technology—that is, a think tank knowledge repository.

3 Ontology-Based Data Integration Framework for Think Tank Knowledge Repositories

Think tank knowledge repositories [9] generally refer to knowledge repository systems that support and serve think tank operations, representing an important mechanism for think tank knowledge capacity building. Collecting, preserving, organizing, and providing services for relevant knowledge content around decision-making domains served by think tank research is the primary task of think tank knowledge repositories. Simultaneously, publishing and disseminating decision consultation products output by think tanks themselves is also an important function of these repositories. Therefore, think tank knowledge repositories possess dual attributes and functions of both domain knowledge repositories and institutional knowledge repositories—they are not only important information support tools for normal think tank operations and decision product output but also effective mechanisms for think tanks to manage and utilize their knowledge assets.

Achieving information semantics is the primary goal of think tank data organization in the big data era. Through analysis and synthesis of relevant literature [10][11], this paper constructs a decision-support-ontology-based data integration framework for think tank knowledge repositories (see [Figure 3: see original paper]). The framework describes and organizes big data resources from coarse to fine granularity across four levels: data resources → datasets → documents → entities, depicting how different types of data information from various sources undergo information extraction and further semantic processing in think tank knowledge repositories, ultimately being transformed into "intelligent data" that can support decision research. Intelligent data refers to valuable information and knowledge extracted from massive data through processing and analysis, endowing data with "intelligence." Compared with the "bigness" of big data, intelligent data possesses higher data value and is more worthy of in-depth mining. Models can be built to seek solutions to existing problems or make predictions —the "beer + diapers" case being a typical intelligent data application.

The information processing in this data integration framework mainly includes five stages: information extraction, data storage, data preparation, semantic data modeling, and application.

FIGURE:3 presents the data integration framework based on decision support ontology.

3.1 Information Extraction Stage

This is the first step in the semantic processing of information resources under big data for think tanks, primarily involving the extraction of relevant data from appropriate information sources. Sources can be external information sources where data is automatically crawled, or internal information sources where outputs are collected from institutional personnel. Based on different information resource formats, the extraction process can be divided into structured and unstructured extraction, with data stored in different types of repositories.

3.2 Data Storage Stage

This stage involves preserving data collected from internal and external sources in the previous stage. Based on data types, storage methods are divided into structured and unstructured data storage. Storage tools range from traditional database management systems (DBMS) such as MySQL and PostgreSQL, to enterprise data warehouses (EDW) and massively parallel processing databases (MPP) such as PADB and SAND. Additionally, distributed file systems like HDFS and HBase, and NoSQL databases such as MongoDB and CouchDB, are also frequently used for unstructured data storage.

3.3 Data Preparation Stage

In the data preparation stage, structured and unstructured information resources in storage devices are organized into datasets according to think tank institutions' service objects and research goals, based on topics, subject domains, or projects. Within each dataset, resources are further subdivided into documents composed of text.

This process involves data cleaning to make data conform to target schemas. Typical processing methods include data normalization, deduplication, integrity constraint violation checking, regular expression-based data filtering, sorting, and grouping.

3.4 Semantic Modeling Stage

The semantic data model is the core component of the data integration framework and the key to achieving semantic processing of decision research-oriented information resources. In the semantic data model, an ontology supporting decision research is first constructed, followed by domain-specific ontologies based on the specific fields involved in think tank decision research. Existing ontologies and various general ontologies are also reused to expand the ontology model for continuously updated information. Finally, through processes such as mapping data fields between ontologies and aligning similar data fields, a general semantic data model oriented toward the decision research process is formed. By this stage, the construction of the think tank knowledge repository' s data organization model is essentially complete. Any data collected from any source under big data can be semantically processed for decision research through the above steps, becoming interconnected instances that better facilitate potential knowledge mining for decision-makers.

3.5 Decision Application Stage

After sufficient semantic processing, data resources have become high-value intelligent data that can provide information support for decision-makers from different perspectives according to various needs in the decision research process. Compared with traditional think tank decision information support mechanisms that rely heavily on human participation, this framework utilizes various information processing technologies and tools to deeply process big data as raw material, making it "computable information" that can be automatically processed by computers, gradually achieving semi-automated to automated decision processes. Moreover, thanks to the massive volume of big data and the semantic processing of data resources by the semantic module, deeper potential knowledge and knowledge associations can be mined and discovered. What is ultimately provided to decision-makers is intelligent decision support data rather than general data information. Consequently, this big data analysis-based decision research method can yield more scientific, reliable, and rapid decision results than traditional methods.

4 Construction of Decision Support Ontology

As a "formal, explicit specification of a shared conceptualization" [12], ontology aims to capture, describe, and represent knowledge in relevant domains, providing a common understanding of that domain' s knowledge, establishing mutually recognized vocabulary, and formally defining these terms (vocabulary) and their interrelationships from various formal schemas [13]. To address the massive unstructured data faced by think tanks in information collection and organization stages under big data, this paper chooses to use ontology methods to semantically process these complex data types, achieving an ontology-driven decision process.

The decision support ontology oriented toward the decision-making process is the core component of the semantic model in think tank knowledge repositories. The structural design of the decision support ontology includes three stages: requirements analysis, ontology modeling, and ontology implementation.

4.1 Requirements Analysis

First, key issues surrounding the decision research process are identified through literature and web surveys, then decomposed to extract entity types and relationship types. Entities and relationships involved in the decision research process have important characteristics that impose corresponding requirements on ontology design:

(1) The decision research process involves many types of entities, such as events, personnel, institutions, and geographic locations, which may be further subdivided. For example, in a water pollution incident, involved institution types may include enterprises, judicial departments, and industrial and commercial departments. Additionally, event observers may sometimes be entity personnel or institutions directly participating in the event—for instance, local residents not only observe and complain about the incident but are also victims who drank the polluted water. The ontology must have the capacity to accommodate various related entities.

(2) Entities and relationships related to decision research are time-sensitive —decision issues, related events, participants, and relationships all exist within specific timeframes. The ontology needs to describe the temporal dimension to support representation and analysis of event development processes.

(3) For each entity and relationship, substantial corresponding data materials provide rich descriptions and commentary. After processing and integration, these data materials can provide revealing quantitative or qualitative references. The decision ontology should include these foundational supporting data materials.

(4) The decision ontology should retain interfaces to other ontologies, supporting expansion of existing and newly built ontologies. For example,

science and technology domain issues will use disciplinary domain ontologies, while supporting materials will use publication ontologies.

The overall requirement for ontology design is to maintain a concise logical model while supporting the above analysis.

4.2 Ontology Modeling

After clarifying requirements, appropriate construction methods are selected for ontology modeling of decision research problems. The ontology design oriented toward the decision research process involves establishing a conceptual model based on the construction goals of the decision support ontology. The goal of building the decision support ontology is to establish a semantic retrieval system based on decision support information, providing think tank experts and policy researchers with semantic information query methods that break through traditional think tank decision information support mechanisms and offer semantic-level decision information query services.

Based on the actual situation and needs of decision research problems, this paper selects the internationally mature seven-step method as the main reference for constructing the decision support ontology in science and technology think tank knowledge repositories. The specific construction steps are as follows:

(1) **Determine the ontology's scope and purpose**: Define the scope of the decision support ontology, i.e., clarify how to describe a decision-related event or resource and to what extent. The purpose is to establish a general ontology oriented toward the decision research process to describe issues triggered by decision topics or events, responses to problems, involved projects, relevant participants, institutions, and their interrelationships.

(2) **Consider reusing existing ontologies**: This paper chooses to reuse DC and ABC ontologies. Dublin Core (DC), as one of the most widely used ontologies, has broad applicability and extensibility for semantic description of basic resource information. The ABC ontology' s ability to model changes makes it suitable for describing various entities and their relationships, including objects of all media types (text, images, video, audio, web pages, and multimedia). It can also model abstract concepts such as knowledge content and temporal entities, like performance or lifecycle events of objects. Therefore, this paper ultimately chooses to reuse relevant classes and properties from DC and ABC ontologies, while utilizing namespaces such as XML schema, RDF schema, and OWL. All reused ontologies are shown in .

(3) **List important terms in the ontology**: Determine the specific vocabulary and logical relationships of core concepts in the ontology. The most common approach is to directly extract subject terms and classification terms from corresponding domain thesauri and classification tables.

(4) **List key entities and classes**: Evaluate extracted core concepts, group

them according to logical rules, and design reasonable classes and hierarchical structures. By referencing relevant research papers and results [14][15] and combining them with actual characteristics of the decision research process, this paper finally determines to construct the decision ontology from five core aspects: Issue, Decision Output, Decision Suggestion, Participant, and Resource, and develops the entire class hierarchy accordingly.

- **Issue class** refers to various problems generated by events in science and technology or other related fields.
- **Decision Output class** refers to decision products finally obtained by think tank researchers and decision-makers through science and technology policy research on generated problems, as well as various intermediate data produced during the research process. The specific classification system is shown in [Figure 4: see original paper].
- **Decision Suggestion class** is the final product of the science and technology think tank decision research process and a subclass of the Decision Output class. Another subclass of Decision Output is the Mid-product class, referring to intermediate data generated from raw data processed by think tank experts and researchers during policy formulation that provides ideas for decision-makers. These intermediate data have great reference value for future research on similar projects and are usually organized and preserved by think tanks.
- **Participant class** refers to all individuals or organizations directly or indirectly involved in the decision research process. Its classification system is shown in [Figure 5: see original paper].
- **Resource class** refers to various types of information resources that support the generation of decision products during the decision research process, including various data, methods, models, tools, etc., while also providing pathways for traceability and evidence-based approaches to science and technology problems. The detailed classification system is shown in [Figure 6: see original paper].

The overall framework of the constructed decision support ontology is shown in [Figure 7: see original paper], divided into three layers: core layer, extension layer, and support layer: - **Core layer**: Issue and Decision Output - **Extension layer**: Participant, Event, Project, Task, Decision Suggestion, Mid-product, etc. - **Support layer**: Resource, Data, Model, Method, Tool, etc.

The layered structure provides a concise logical model that makes entity relationships in the core, extension, and support layers clear and orderly. Different layers store data with varying structural complexity and precision, allowing the system to balance accuracy and comprehensiveness according to query needs, meeting personalized query and analysis requirements.

(5) **Analyze entity properties**: Ontology data properties represent relationships between classes or concepts and values (e.g., the "age" property value represents a specific age of an organism), while object properties represent

non-hierarchical relationships between classes (e.g., "trigger" or "cause" properties can represent causal relationships between classes). Properly defining data and object properties can effectively reflect inter-class relationships. Due to space limitations, only partial properties of the core layer classes are shown in .

(6) **Analyze property constraints**: Apply necessary constraints to properties. Constraints on data properties include describing property value types (string, boolean, enumeration, etc.), value ranges, cardinality (single or multiple), etc. For example, the Year field of temporal properties could have a maximum value of 2017, or human gender could only be selected from "male" or "female."

(7) **Create instances**: Create specific instances based on the conceptual model established in previous steps. This paper uses Protégé 4.3 to add partial instances for a specific domain problem as a demonstration. Additionally, API tools can be used for batch instance import. The created ontology is saved locally in OWL text format for easy copying, backup, editing, and modification.

4.3 Implementation Process of Think Tank Knowledge Repositories

The overall operation process of think tank knowledge repositories is shown in [Figure 8: see original paper] (using science and technology think tanks as an example). API tools and crawlers can capture relevant events and problem content in research domains from the Internet while collecting user policy needs or proposals. After natural language processing, text processing tools analyze the above content, extract entities and types, and use the above ontology construction model and editors like Protégé for ontology construction and data maintenance. External ontologies such as DC metadata and science and technology domain ontologies can be introduced based on information source nature and data characteristics. Finally, all content collected from the real world is fully entity-ized, forming semantically interconnected intelligent data that, combined with previous user policy needs or proposals, provides corresponding matching results to offer decision support information for policymakers. The ultimate goal is to build a semantic-driven retrieval system oriented toward the decision research process based on big data information collected and processed by think tanks ([Figure 9: see original paper]).

FIGURE:8 illustrates the knowledge repository operation process of think tanks, while **FIGURE:9** shows an example of a semantic-driven decision-making retrieval system.

In this retrieval system, massive data resources under big data—such as social hotspot events, social issues, news reports, and research publications—are extracted by think tanks, undergo natural language processing, and are then processed and stored in think tank knowledge repositories to become intelligent data. During retrieval, user query needs also first undergo natural language processing, after which requested data is mapped to the decision ontology in the think tank knowledge repository in the same way, subsequently matching previously entity-ized data and returning intelligent data-based retrieval results to users. Compared with traditional think tank methods of data search and collection in decision research, this semantic-driven and intelligent data-based retrieval system can obtain more comprehensive, scientific, and interrelated retrieval results—in essence, context-based results. Such retrieval results can more easily facilitate potential knowledge mining and, with context, be more conveniently applied directly to decision-making. For example, when retrieving information on "haze prevention and control," users will obtain through this semantic-driven decision retrieval system a series of information including causes, start and end times, locations, relevant personnel, institutions, publication descriptions of a haze occurrence event, as well as relevant research institutions, researchers, policy implementation status, and temporal impacts of haze prevention and control policies. This achieves a context-aware, semantic retrieval mechanism oriented toward decision research, breaking traditional think tank decision information support mechanisms that rely heavily on human participation.

This paper primarily investigates current traditional think tank information support mechanisms and their challenges in the big data era, proposes a general semantic think tank decision information processing framework, and designs a data organization model for science and technology think tank knowledge repositories based on this framework, aiming to provide references for new-type think tank construction under big data.

Regarding ontology creation for think tank policy domain research, current domestic attempts are still rare, and much subsequent work remains. While continuously improving the functions of this think tank decision information processing framework and maintaining the decision ontology, future research will continue to focus on ontology mapping, ontology evaluation, and integration expansion of the decision ontology.

References

[1] James G. McGann, University of Pennsylvania, 2016 Global Go To Think Tank Index Report[EB/OL].[2016-08-10]. http://repository.upenn.edu/think_tanks/10/.

[2] Wu Yuliang. Research on Foreign Think Tank Decision Information Support and Its Enlightenment[J]. Library Theory and Practice, 2015(10):31-35.

[3] Liao Qiu, Yan Yangfan, Mo Chongju. Research on Institution-Building Academic Databases in the Big Data Era[J]. Library Science Journal, 2014(4):34-36.

[4] MOORTHY J, et al. Big data: prospects and challenges[J]. The Journal for Decision Makers, 2015,40(1):74-96.

[5] Bruegel. Datasets[EB/OL]. [2017-03-01].http://bruegel.org/publications/datasets/.

[6] SIPRI. Databases[EB/OL]. [2017-03-01].https://www.sipri.org/databases.

[7] Xu Xin, Wu Shanyan. Research on the Construction of Think Tank Knowledge Repositories[J]. Information Studies: Theory & Application, 2014(3):68-72.

[8] Studer R, Benjamins C R, Fensel D. Knowledge Engineering, Principles and Methods[J]. Data and Knowledge Engineering,1998, 25(1-2):161-197.

[9] Benjamin Horne, Tanya Torres, Jessica Mackenzie. Think Tank Management: Establishing a Knowledge Repository[EB/OL].[2016-12-30]. http://www.ksi-indonesia.org/en/news/detail/think-tank-management-establishing-a-knowledge-repository.

[10] Srividya K Bansal, Sebastian Kagemann. Semantic Extract-Transform-Load framework for Big Data Integration[J]. Computer, 2015,48(3):42-50.

[11] Francesco Corcoglioniti, Marco Rospocher, etc. The KnowledgeStore: A Storage Framework for Interlinking Unstructured and Structured Knowledge[J]. International Journal on Semantic Web and Information Systems,2015,11(2):1-35.

[12] Studer R, Benjamins C R, Fensel D. Knowledge Engineering, Principles and Methods[J]. Data and Knowledge Engineering. 1998, 25(1-2).

[13] Liao Jun. Research on Information Retrieval Based on Domain Ontology[D]. Changsha: Central South University, 2007.

[14] Leo Wanner, Marco Rospocher, etc. Ontology-centered environmental information delivery for personalized decision support[J]. Expert System with Applications.2015(42):5032-5046.

[15] Studer R, Benjamins C R, Fensel D. Knowledge Engineering, Principles and Methods[J]. Data and Knowledge Engineering. 1998, 25(1-2).

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