



OPEN Reuse oriented information analysis methodology study on information analysis componentization

Hongyu Cai^{1,3}, Xiangfeng Wang^{1,3}, Jin Shi^{1✉} & Xiaoyang Zhou^{2✉}

Nowadays, the trend of computerization in information analysis is becoming increasingly evident. However, with the development of computer technology, the engineering and reusability of information analysis work has not improved. To solve this problem, we propose the concept of componentization to guide information analysis. Based on theories related to knowledge management, guided by Taylor's scientific management, and landed on component technology, we introduce the theory of information analysis componentization. Information analysis components include computer, human, and interactive components, while emphasizing the development and reuse of human components to enhance the reusability of information analysis work. To demonstrate the component-based approach to information analysis, we improved the DIKW (Data, Information, Knowledge, Wisdom) model and presented the complete concept of component-based information analysis based on the improved DIKW model. Finally, we summarize the contributions of this work, discuss its limitations, and outline directions for future research. Our study is expected to offer both theoretical foundations and practical guidance for enhancing the efficiency of information analysis in future applications.

In recent years, big data has garnered considerable attention within the field of information science. Scholars broadly concur that big data analytics provides the capability to uncover valuable insights from vast and otherwise inaccessible datasets¹. With the rapid advancement of big data technologies, it has increasingly become a critical instrument for information analysis². However, the intrinsic characteristics of big data—namely its massive volume, multi-source heterogeneity, high velocity, and low value density³—pose substantial challenges to effective information analysis. To address these challenges, researchers are actively exploring innovative methods and techniques, including sentiment analysis⁴ human-computer collaboration⁵ and other computational science approaches⁶ aimed at enhancing the processing of both structured and unstructured data. Simultaneously, the research paradigm in information analysis is undergoing a significant transition toward data-intensive methodologies, placing greater emphasis on data-centric approaches⁷.

Despite the rapid technological advancements and the growing dominance of data, human intelligence remains indispensable in the process of information analysis⁸. Advances in big data analytics, artificial intelligence, and related technologies notwithstanding, certain analytical tasks continue to exceed the capabilities of computational systems⁹. For example, the analysis of dialects, ancient manuscripts, or fragmented documents demands profound cultural and linguistic understanding, which current computational methods cannot adequately replicate¹⁰. Consequently, it is imperative to delineate those tasks that necessitate human execution, thereby reinforcing the irreplaceable role of human expertise within the domain of information analysis¹¹. Moreover, even in tasks that can be executed computationally, the expertise of human analysts remains critical. Although activities such as planning analytical workflows, selecting appropriate frameworks, and designing algorithms may not appear excessively complex to domain experts, they nonetheless require substantial professional experience and intellectual engagement¹². The accelerating speed and diversification of data sources further compound this challenge² often leading to analytical delays that struggle to meet the real-time demands of information consumption. Thus, improving the efficiency of information analysis to align with the dynamic and fast-paced nature of contemporary data landscapes has become an urgent research priority¹³.

¹Nanjing University, CN, Nanjing 210033, China. ²China Mobile Zijin (Jiangsu) Innovation Research Institute, Nanjing 210033, China. ³Hongyu Cai and Xiangfeng Wang contributed equally to this work and should be considered co-first authors. ✉email: shijin@nju.edu.cn; zxyjobs51@163.com

Although big data analytics has achieved considerable success in practical applications¹⁴ its contributions to the theoretical development of information science remain relatively limited. Some scholars contend that current big data methodologies are predominantly application-driven, with insufficient focus on constructing comprehensive theoretical frameworks to guide information analysis research¹⁵. In response to these challenges, we propose the theory of componentization in information analysis, which constitutes the primary theoretical innovation of this study. This theory aims to emphasize the crucial role of human involvement while enabling the rapid deployment and reuse of information analysis processes¹⁶. Simultaneously, it seeks to address the relative lack of theoretical advancements in information science arising from the proliferation of big data technologies.

Our theory is rooted in the principles of systems engineering and draws upon several foundational theoretical frameworks, including knowledge transfer¹⁷ knowledge distance¹⁸ and knowledge coupling theories¹⁹ from knowledge management; Taylor's scientific management²⁰ principles from the field of management; and component-based development practices from software engineering²¹. It advocates for the standardization and decomposition of both human and computational tasks within information analysis, encapsulating them into modular units termed Information Analysis Components (IACs). Through the standardization and management of human actions, the theory aims to reduce knowledge distance among stakeholders, enhance knowledge coupling, and improve the efficiency of knowledge transfer. Moreover, by leveraging component technology to encapsulate the work of both humans and machines into reusable modules, the efficiency and scalability of information analysis can be significantly improved.

To comprehensively illustrate the theory of information analysis componentization, this paper proceeds as follows: First, we provide an overview of the current landscape and progress in information analysis. Second, we introduce the relevant theoretical foundations underpinning our approach. Third, we elucidate the conceptual framework and connotation of information analysis componentization. Subsequently, we propose an improvement to the traditional DIKW (Data-Information-Knowledge-Wisdom) model²² representing the second major innovation of this study, and employ the improved DIKW model to demonstrate how modularized information analysis processes can be implemented. We then present a case study that illustrates the workflow of information analysis under the componentization framework. Finally, we discuss future research directions and conclude by summarizing the contributions of this study.

Related works

Engineering research for information analysis

In the era of big data, the study of information analysis methods and models carries substantial theoretical significance and practical value. However, the inherent sparsity of value in big data, coupled with the rapid evolution of analytical tools and research paradigms, demands an urgent shift towards an engineering-oriented approach to information analysis²³.

Information Analysis Engineering is defined as the systematic consolidation of information analysis methods and models into standardized processes, formalizing accumulated experience, skills, and knowledge into reproducible frameworks that consistently generate high-value knowledge outputs²⁴. With the emergence of the big data service paradigm, information analysis must increasingly leverage cloud computing infrastructures to achieve automation and intelligent processing through the correlation and mining of distributed information resources¹¹. Yet, to date, the provision of such services remains largely dominated by internet enterprises, while systematic academic efforts are notably lacking. To advance the engineering transformation of information science, it is imperative to actively integrate emerging technologies such as artificial intelligence²⁵ transitioning from traditional reliance on individual literacy and experiential heuristics toward standardized, scalable engineering systems²³.

Although automated and semi-automated systems can assist in pattern discovery and analytical processes²⁶ the essence of information analysis remains deeply reliant on human expertise and judgment. Human-computer collaboration and innovation are indispensable drivers of continuous advancement²⁷. Moreover, the limited interpretability¹⁰ of AI technologies necessitates the integration of expert knowledge into decision-making outputs to mitigate inherent risks⁷. Expert wisdom thus plays a critical, irreplaceable role across the entire information analysis lifecycle, encompassing data acquisition, information organization, knowledge generation, and product development¹⁰. Despite their extensive experience and profound domain knowledge, experts' true capabilities are often embodied in their deep understanding of complex issues, their acute perception of ambiguous information, and their accurate forecasting of future trends²⁸. These forms of expertise are inherently qualitative and resist straightforward quantification through conventional metrics²⁹ rendering the measurement of expert intelligence a persistent and formidable challenge.

In response to this problem, this study introduces a novel methodology centered on the componentization of information analysis. By employing component-based techniques, each stage of the information analysis process—including, but not limited to, expert-driven actions—is encapsulated into discrete, reusable components. The execution of information analysis tasks is thereby achieved through the dynamic orchestration of these components. This paradigm not only facilitates the reuse and rapid deployment of analytical tasks but also significantly accelerates the operationalization of information analysis projects. Within this framework, the information analysis project offers a high-level strategic blueprint, whereas the componentization process provides a concrete, systematic pathway for implementation. Through the construction of Information Analysis Components, the proposed approach fosters standardization, enhances operational efficiency, and lays the groundwork for the development of scalable, replicable, and intelligent information analysis systems.

Theoretical basis

The purpose of information analysis is to provide strategically significant information in the form of advanced intelligence conclusions and knowledge to meet the specific needs of users and demands. This is accomplished through logical thinking processes such as data collection, analysis, judgment, and inference based on a wide collection and accumulation of relevant experience and knowledge³⁰. In the era of big data, information analysis has become a complex and collaborative work, and it is not realistic to complete it independently. Collaborative information analysis has become a common phenomenon, involving many theories that this section will briefly introduce.

Knowledge management related theory

Knowledge transfer refers to the process of transferring knowledge from one carrier to another, including the sending and receiving processes of knowledge¹⁷. The transferred content usually involves valuable knowledge resources that have transfer and dissemination value, and information is used as a carrier for transfer. In the process of information analysis, the knowledge obtained by the information analysts from the raw data will be transferred as an information product to the information demand object (collaborator or end-user). The probability and efficiency of knowledge transfer are influenced by various factors¹⁷. In the process of information analysis, we believe that the probability and efficiency of knowledge transfer are mainly influenced by the knowledge distance between the transferer and the recipient.

Knowledge distance refers to the difference or dissimilarity between the sender and the receiver in terms of their existing knowledge level, structure, content, or other dimensions. Knowledge distance has two attributes: knowledge depth and knowledge breadth¹⁸. Knowledge depth refers to the knowledge content in a particular field of expertise and reflects the degree of difference in knowledge level among different knowledge subjects in the field. Knowledge breadth can be used to represent the diversity of knowledge, usually referring to the differences in knowledge structure among subjects in different fields. The overlap of different subjects on knowledge breadth produces a knowledge coupling phenomenon¹⁹ and the greater the strength of the relationship between knowledge, the greater the strength of knowledge coupling¹⁹. The greater the strength of knowledge coupling, the smaller the cooperation barriers between the subjects. In any organization or society, knowledge distance is objectively present, determined by the degree of difference in knowledge level between the two parties of knowledge transfer from different backgrounds and fields¹⁸. In the current knowledge transfer under the big data context, as the amount of data explodes and develops towards multi-source and heterogeneous directions, information analysis work is developing towards interdisciplinary and collaborative directions involving multiple people, and information demanders are developing towards the general public. The knowledge distance between personnel involved in the entire information analysis process is constantly increasing. To improve the efficiency of knowledge transfer, it is necessary to mitigate the impact of knowledge distance on information analysis work.

Scientific management

To reduce knowledge distance, it is easy to think of improving individuals' depth and breadth of knowledge while enhancing the knowledge coupling strength among the subjects in collaboration. While, in the context of big data, this method is inefficient and futile because the people and demands involved in each information analysis task are different. Scientific management theory provides us with an approach to offset the impact of knowledge distance. The Taylor's scientific management system is a management system composed of various elements such as human-machine collaboration, labor-capital cooperation, and social harmony²⁰. Taylor established a "scientific labor process" by precise job analysis and summarizing past labor processes, which standardized the production tools and operating procedures, thereby increasing output with the same labor.

From the early 20th century through the 1930s, Taylor's scientific management theory was extensively disseminated and applied within enterprises across the United States and Europe. As industrialization accelerated and organizational complexity increased, Taylor's emphasis on efficiency, standardization, and labor specialization offered timely and practical solutions to emerging management challenges. With the evolution of managerial practices, scholars and practitioners continued to refine and extend Taylor's framework. Over time, scientific management theory became integrated with broader theoretical paradigms, including systems theory and contingency theory, thus giving rise to more comprehensive and sophisticated management concepts and methodologies.

Despite criticisms regarding its mechanistic tendencies, the influence of scientific management persists. Techniques originally rooted in Taylor's principles—such as job analysis, workflow optimization, and performance measurement—have been widely adopted and adapted across diverse fields, notably in human resource management, operations management, and organizational design.

Fundamentally, Taylor's scientific management can be regarded as an early manifestation of system-oriented management thought. It emphasizes the establishment of standardized work processes, the formalization of work tasks, and the specialization of labor—all of which resonate closely with the structural characteristics of information analysis activities. From the perspective of the information analysis process, information analysis may be conceptualized as a sequential and structured progression: extracting relevant information from raw data, synthesizing and organizing it into systematic knowledge, and ultimately transforming this knowledge into actionable insights or wisdom³¹. Each stage in this progression follows a strict logical sequence and demands a high level of domain-specific expertise and analytical rigor.

In this context, the theoretical foundations laid by scientific management continue to inform and support modern approaches to information analysis, highlighting the enduring relevance of standardized, systematized, and specialized work methodologies in contemporary knowledge-driven environments. Therefore, we believe that using scientific management theory to guide information analysis work is feasible, at least worth a try.

Software component

Considering that most information analysis work is intellectual labor, it is difficult to decompose it into a “scientific labor process” at present. Apparently, in the era of big data, information analysis work and computers are already inseparable³⁰ so we can use the concept of software components to develop information analysis work. The component concept originated in the field of architectural engineering, referring to the various unit bodies that make up the building structure¹⁶. In actual construction projects, construction personnel do not need to understand the composition and production methods of each component, but only need to master the performance and usage methods of each component to complete the construction of the building. To solve the “software crisis” in software engineering³² the concept of “software component” was introduced into the computer field¹⁶. Software components refer to program bodies with certain functions that can work independently or coordinate with other components, and by combining existing components, new services or software can be developed²¹. In a broad sense, components have properties such as independent configuration, strict encapsulation, high reusability, and universality. Introducing the concept of components into the information analysis field can help optimize the information analysis process and counteract the impact of knowledge distance on information analysis efficiency²⁴.

Connotation of information analysis components

Introduction to information analysis components

In information science research, the concept of “component” is not a new idea, as it was already mentioned by scholars in the late 1980s³³. At that time, it was used by information scientists to develop information systems after the concept of software components was introduced³⁴. Nevertheless, the “component” discussed in this paper are not the same concept. The information analysis componentization proposed in this paper are a guiding concept specifically for the information analysis process or building information analysis models. The goal is to assemble the various parts of the information analysis process in the form of IACs, rather than assembling them into an information system. It is a product that integrates the concept of software components and the “scientific labor process” concept²⁰ with a greater emphasis on human participation.

In the current era of big data, using computers for information analysis is an undisputed approach. Whereas, as mentioned earlier, expert knowledge is an indispensable part of information analysis work⁸. Therefore, the IACs proposed in this paper can be classified into computer components, human components²⁷ and interaction components according to their survival forms. Computer components are mainly stored in the form of code and are responsible for the processing functions of big data²¹. Human components combine the characteristics of information analysis and integrate the “scientific labor process”, stored in the form of operating instructions documents, and standardize the operational actions of information analysts²⁷. They are responsible for decomposing information needs, assisting computers in completing semi-automatic information analysis processes, and verifying the information analysis results from the computer components. Interaction components are stored in the form of external devices or communication protocols in the component library. Their primary function is to display visual interfaces and facilitate the two-way transfer of information between human and computer components. All components will be stored in a specific location called the component library³² for easy access. The component library can be classified into computer component library, human component library, and interaction component library based on the form of survival.

IACs have the general properties of software components. Firstly, they are independent configurable units with independent functions and structures, and must be self-contained and reusable³³. Secondly, IACs emphasize separation from the environment and other components, and do not display internal details to the outside world³⁵. Thirdly, IACs should not be persistent and should not have individual attributes³³. For Human components, they should also be simple, easy to understand, and interactive⁹. Standardized manual operations should simplify complex problems to reduce implementation difficulty, and the definition of actions should be easy to understand for relevant personnel. Their design should consider human-computer interaction requirements, focus on interoperability, and increase their reusability.

Information analysis project development under the guidance of componentization thinking should follow the principle of Plug-and-Play (PNP) to save time in information analysis. PNP is a technology proposed by Microsoft that automatically matches physical and software device drivers³⁶ and allows hardware to be used in a system without manual intervention. When introduced into software development, it means that some people are dedicated to producing software components, while others are responsible for constructing software structures and inserting software components into them for quickly complete the development of large software projects³⁵. In the componentization thinking of information analysis, PNP refers to using IACs for connection to construct information analysis programs. Under this concept, IACs need to have some common interfaces to meet PNP functionality. Figure 1 shows a visual display of an IACs’ interface. The external interface of the information analysis component is mainly composed of six parts²⁷: the Component Interface, which is the management interface of the component library used to create, call, and destroy IACs; Attributes, which are the working properties of the IACs and can be used to adjust the component’s working mode. The other four parts can be divided into the offer area and the required area. The offer area is the interface that provides functionality to other components, including the information process and need reception interfaces. Among them, the information process is the function it can provide externally, and the need reception is the interface for listening to information analysis requirements and receiving information demand content. The interfaces in the required area make requests to other components, including the processing receptacle and Need iteration interfaces. Among them, the processing receptacle is the auxiliary function required when providing information processing functions, and the Need iteration is the demand decomposition content obtained after requirements analysis.

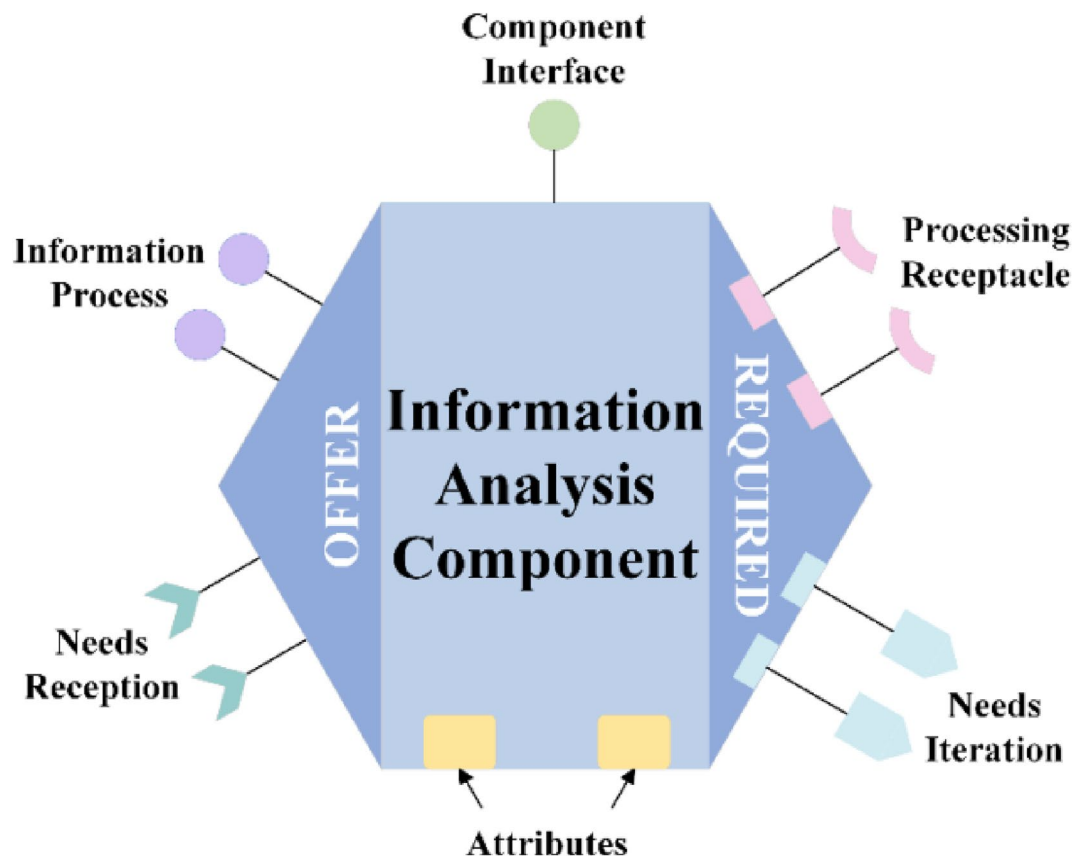


Fig. 1. Visual display of intelligence components.

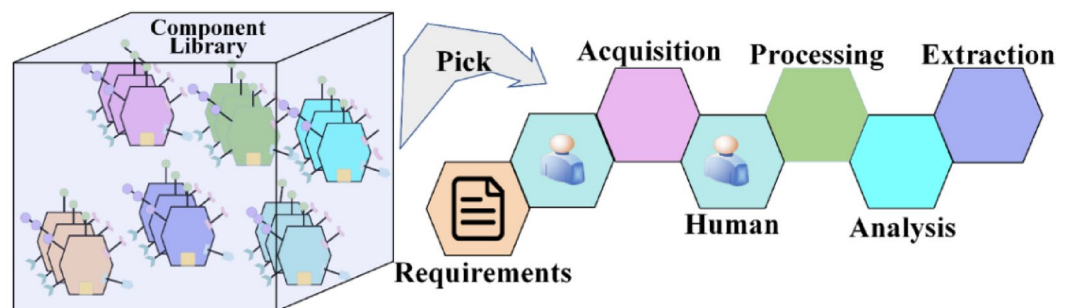


Fig. 2. Process design under the concept of information analysis componentization.

Process design under the Idea of information analysis componentization

Under the idea of componentization, the information analysis model is mainly constructed in the form of IACs assembly. To implement this idea, various information analysis functions should first be encapsulated into IACs and stored in the component library. For instance, programs with information collection functions are encapsulated as information collection components, and various information collection programs for text, websites, videos, pictures, etc. can be encapsulated as various information collection components. The algorithm with data analysis function is encapsulated as data analysis components, so that different data analysis algorithms can be encapsulated as different data analysis components. Store all the components in the component library, and when information analysis is required, consider assembling IACs with specific functions from the component library into the information analysis process. When it is found that the existing IACs cannot meet the needs, then consider developing new components. In this way, the efficiency of information analysis will be greatly improved. For example, consider a general information analysis process that includes data collection, data processing, data analysis and knowledge extraction. And its development process under the idea of information analysis componentization is shown in Fig. 2. So as to clearly introduce the design of information analysis process under this idea, this section assumes that each step of this analysis process can be completed

by only one component. In reality, each step may require multiple component to cooperate to complete. This section is only a brief introduction to the design idea, a formal example of the development process will be presented in a later section.

In this idea, when a new information analysis requirement is encountered, the requirements document is first written and modified using a domain-specific language²⁷ allowing the requirements to be normalized and placed in a requirements component. This facilitates the transfer of knowledge between people and computers at each step of the process. Domain-specific languages are computer languages that are distilled from general-purpose programming languages to meet the special requirements of information analysis. Then, data acquisition component are selected from the component library according to the requirements. Next, the expert takes the collected data through manual components and selects the corresponding data processing components for data cleaning according to their characteristics. Subsequently, the data analysis component is selected for data analysis, and finally the knowledge extraction component is selected to mine the data value. Since each IACs has its own attributes, when a building block is selected, it prompts what IACs are available for selection in the next step to increase the ease of assembly.

Component-based information analysis framework

While several models can serve as examples to display our proposed theory, we firmly assert that the DIKW model is the most suitable choice. Primarily, it is a renowned and well-established model within the domains of information management and knowledge management. It has garnered wide acceptance and has been refined through numerous iterations. Additionally, it effectively illustrates the fundamental process of information analysis, making it an ideal exemplification for theoretical research. We firmly believe that the simpler it is, the more effective it is, thereby reinforcing our rationale for selecting the DIKW model. In this section, we first build a simple improved information analysis framework based on the classical DIKW chain. In the next section, we will introduce how to use the modularization approach to construct this framework to demonstrate the architecture of a component-based information analysis framework.

DIKW chain

In the discipline of information management and knowledge management, there is a very classic model called DIKW hierarchy model, also known as the knowledge pyramid²². The DIKW model shows us how data is gradually transformed into information, knowledge, and wisdom³¹ as shown in the right figure of Fig. 3. Data is at the bottom of the model, and it can be processed into general information with certain meaning. Information is carefully processed, and knowledge that reflects its internal connections is deduced through induction and deduction. Knowledge is internalized and sublimated to form wisdom that serves decision-making needs³⁷.

This knowledge mining process of DIKW is actually a simple linear information analysis workflow. As shown in Fig. 3, the “data-information-knowledge-wisdom” in the DIKW model on the right corresponds to the “information collection-information processing-information analysis-service” in the simple linear information analysis process model, where information collection is the process of obtaining, collecting, and describing information about specific problems or goals²². Information process refers to the classification, organization, processing, and handling of information data for subsequent analysis and use. Information analysis involves in-depth research, analysis, and interpretation of collected information data to gain deeper and more accurate understanding and insight. Services provide information support and consulting services to decision-makers and stakeholders.

Yet, in reality, the relationship between each stage of the information analysis workflow is not a simple linear relationship. The information and data between each stage need to flow freely, and effective communication

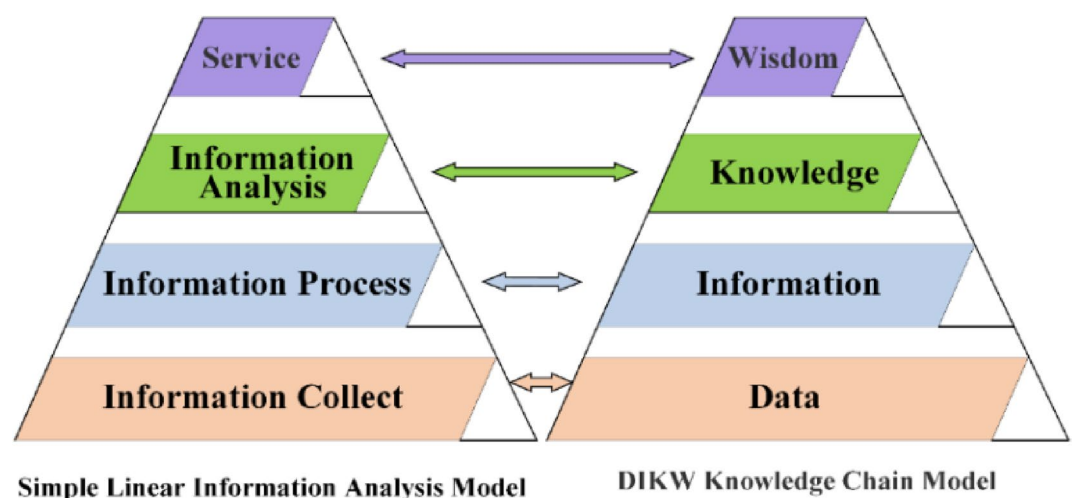


Fig. 3. Comparison of DIKW and simple linear information analysis model.

between personnel is required to achieve efficient collaboration. So as to better illustrate the information analysis workflow under the componentization concept, we have made improvements to the DIKW model.

Improved DIKW information analysis model

While the traditional DIKW model captures the general trajectory of knowledge generation, it does not adequately account for the dynamic and uncertain nature of user requirements, which is a defining characteristic of information analysis activities. In the context of big data, such uncertainty is significantly magnified, thereby exerting a profound influence on the accuracy, relevance, and value of analytical outcomes. To mitigate this uncertainty, we propose an enhanced DIKW framework—hereafter referred to as the Improved DIKW Model—which explicitly incorporates a “requirements document” into each stage of the DIKW hierarchy. This augmentation is coupled with a cyclic, iterative, and spiral-up progression mechanism, designed to ensure the comprehensive extraction of information value while maintaining continual alignment with evolving analytical objectives. The conceptual structure of the Improved DIKW Model is illustrated in Fig. 4.

Within this framework, the information analysis process is initiated by consolidating user inputs into a standardized “initial requirements document”, which serves as a foundational artifact embedded into the entire analytical workflow. This document guides the identification of data sources and informs subsequent analytical decisions. The data acquisition phase is conducted through an iterative, human–machine collaborative process aimed at verifying that the data collected is both relevant and sufficient to address the articulated requirements. Through continuous refinement, the initial document evolves into an “iterative requirements document”, which is propagated through each layer of the DIKW model. Both the data and the associated requirements documentation undergo parallel iterative refinement until the final information product is produced.

Human involvement plays a critical role in this process, primarily through the iterative refinement of the requirements document and the qualitative assessment of the congruence between system-generated outputs and user expectations. Concurrently, the computational system delivers automated analytical services, supported by visualization technologies that facilitate expert interpretation and domain-specific evaluation.

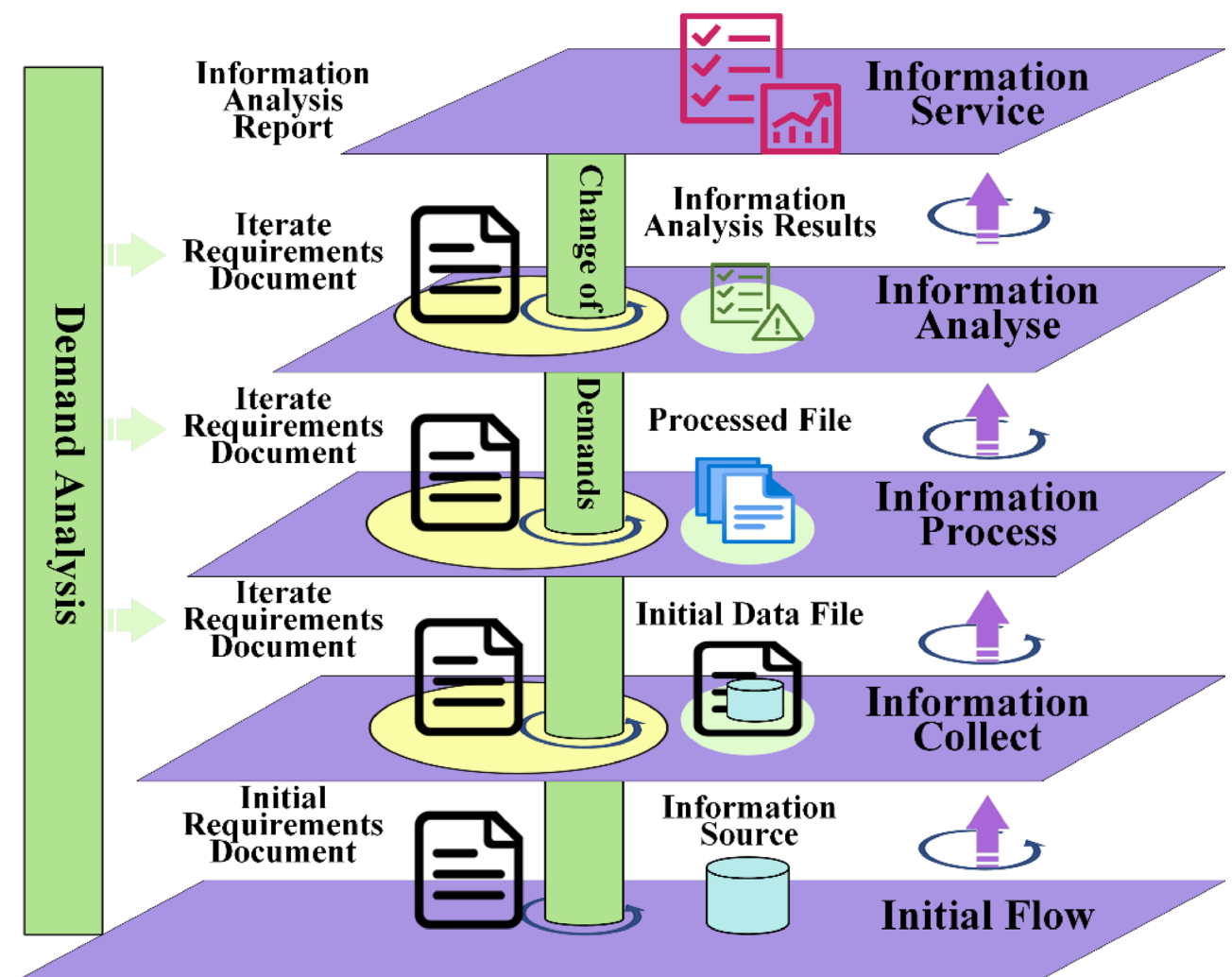


Fig. 4. Improved DIKW model.

In the requirements management phase, the document is first translated into a machine-readable format via a translation component, after which it is stored in a centralized document center for versioning and future reference. The content of the document is also rendered via a visualization component, enabling experts to conduct requirements validation and analysis with visual support. During this stage, salient features of the relevant data are also visualized to assist in decision-making. Manual revisions are performed through the document center interface. If updates are deemed necessary, a requirements change component is invoked to implement modifications; otherwise, the process continues to the subsequent analytical phase. A result evaluation component is responsible for assessing whether the analytical outcomes satisfy the requirements, and if discrepancies are identified, feedback is routed back to the document center to trigger further revisions.

From the perspective of data flow, the raw data is initially subjected to preprocessing by the data management component before being transmitted to the computational model component for analysis. Analytical results are then either stored, visualized, or transferred to higher-order stages within the DIKW hierarchy for further processing. Throughout this procedure, a human-assist component monitors the analytical process via the visualization interface and, where necessary, adjusts the parameters of the computational model through the data center to enhance output quality. This continuous human-in-the-loop mechanism ensures that data analysis remains tightly coupled with the evolving requirements landscape.

Example of component-based information analysis model development

In the previous section, we introduced an improved DIKW model for information analysis workflow guided by componentization thinking. In this section, we will discuss how to implement the Improved DIKW model with componentization thinking. Based on different levels of abstraction, we define the information analysis model under the componentization thinking as a three-level structural model, as shown in Fig. 5, with the levels ranging from high to low abstraction: the conceptual, logical, and supporting layer. The conceptual layer is mainly responsible for decomposing the workflow of information analysis on the work level, which is the information analysis model diagram. It provides a design basis for the combination of components at the lower structural layer. The logical layer describes the specific business logic of the IACs combinations in the information analysis process, while the supporting layer provides practical support for the information analysis work. The following is a specific description of the three layers.

Conceptual layer

The conceptual layer is the representation of the information analysis model, Which in this study refers to the representation of the improved DIKW model. Under the componentization thinking of information analysis, the information analysis workflow begins with a fixed-format requirements document and data source (referred to as “initial flow” below), and the final output is the analysis result and iterative requirements document. The specific workflow is shown in Fig. 6.

In the big data environment, data presents a multi-source and heterogeneous nature. Therefore, after obtaining the initial requirements and data, the corresponding computer components should first be used to process various data into standard requirements document and data source files for subsequent information analysis. Before inputting the initial flow, it is necessary to decompose the information analysis process, i.e.,

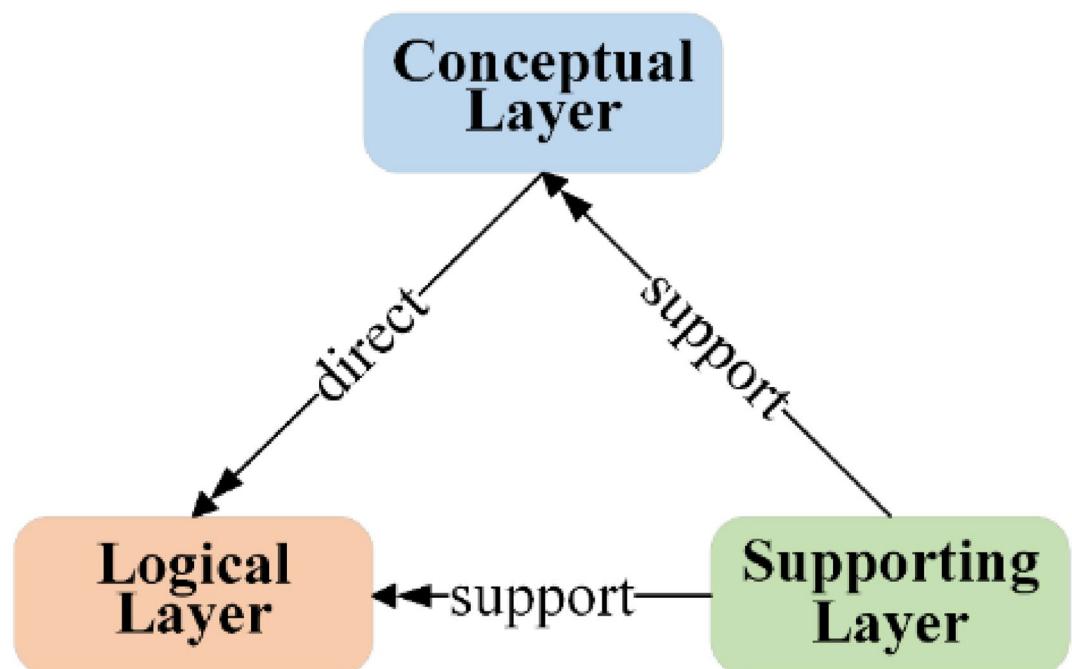


Fig. 5. Three-layer structure relationship diagram.

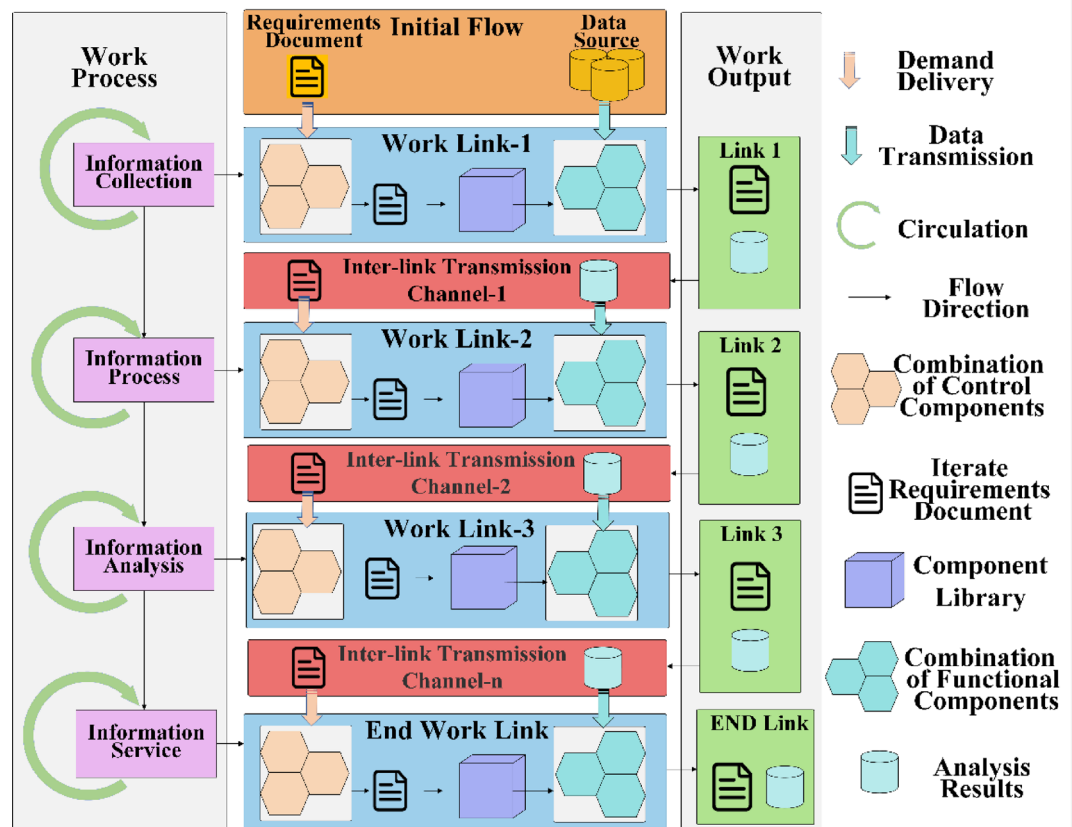


Fig. 6. Process design under the concept of information analysis componentization.

to divide the information analysis work based on the information analysis process model, forming a series of interlocking and clearly bounded work sections. After the division is completed, the initial flow will be input into the first work section for subsequent transmission and analysis. In each work section, the control component assembly composed of control components will first perform requirement decomposition on the requirements document of that section, forming new requirement content. This process is called “requirement iteration”. Then, the component assembly will call the functional components in the component library that can meet the corresponding requirements and assemble them into a functional component assembly that can complete the information analysis. Finally, data will be input into the functional component assembly for analysis and the analysis results will be obtained. After the results are obtained, the human components in the combination of functional components determine whether they meet the expectations, and if they are far from each other, the refinement will be repeated in this layer. If they meet the predefined expectations, the iterated requirements document and analysis results in the data will be transmitted between the sections through the transmission channel for further “requirement iteration” and data analysis work until the information analysis work is completed. In the last section, the iterative requirements document and analysis result of the information service section will be infinitely close to the actual needs of the user.

Logical layer

The logical layer performs logical mapping of specific businesses in the conceptual layer and can fully reflect the idea of componentization. In the process of logical mapping under the idea of componentization, the logical layer consists of control module, view module, preparation module, and function module. The control module provides various interactive services to ensure smooth information exchange between people, people and computers, and computers and computers in the information analysis process. The view module mainly provides information visualization services, which can visually display the information processing process and results to assist humans in controlling the information analysis process. The function module provides information processing services and undertakes the function of automatically analyzing and processing data. The preparation module provides logistical services, such as data preparation, requirement analysis, and intermediate product storage. The relationship among the modules is shown in Fig. 7. The function module provides technical support to other modules, the control module adjusts and controls other modules, the view module visualizes the results of other modules, and the preparation module provides data materials for other modules.

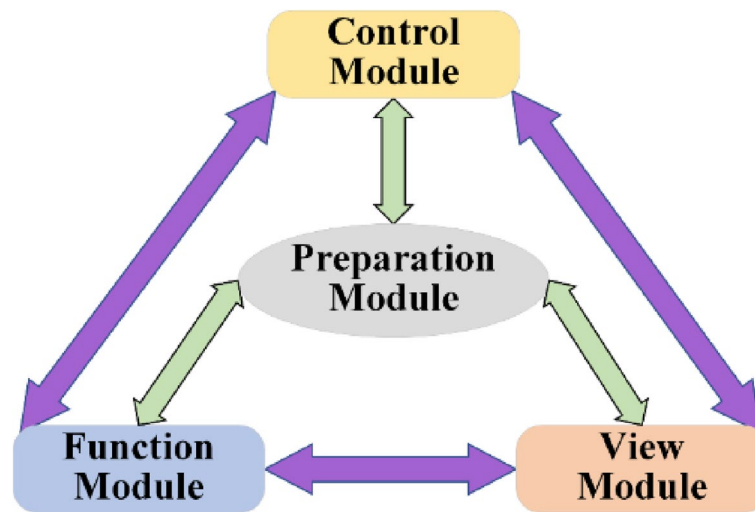


Fig. 7. Logic layer module relationship.

Supporting layer

The supporting layer is the guarantee for the implementation of the logical and conceptual layers. It consists of data sources and component libraries. The data source is the origin of data for conducting the analysis, and in the cause of improve the effectiveness of information analysis, it should be as comprehensive as possible, containing a wide variety of relevant information from various sources¹.

The component library is where various IACs are stored. According to the classification of IACs, it contains computer components, human components, and interactive components. It can also be subdivided according to other criteria. Computer components are components that are executed directly by the computer and can be classified into visualization components, data management components, data acquisition and analysis algorithm components, etc. according to the functions provided. Human components mainly encapsulate human activities and specify the knowledge background¹¹ professional skills²⁴ operational guidelines²⁷ etc. required for engaging in a certain activity. Interaction components⁹ can be classified into human-to-human interaction components (defining the ways, guidelines, templates, etc. of human communication), human-to-machine interaction components (data acquisition equipment, virtual reality equipment, etc.), and machine-to-machine interaction components such as communication protocols (reference literature), etc.

Case study of information analysis componentization

Information analysis can be broadly categorized into two modes: result-oriented analysis and problem-oriented analysis, based on the level of understanding of the target. Result-oriented analysis entails the analyst having a preconceived notion of the target and needing to gather information from a wide range of sources to substantiate their hypothesis. For instance, the assertion that “war destabilizes the world” exemplifies the result-oriented analysis model. To validate this claim, an approach based on component-based information analysis is employed, wherein the problem is broken down into sub-problems. Such as, the impact of war on the political, economic, cultural, and social aspects is examined, and within each sub-problem, further subdivisions are made. For example, social problems can continue to be divided into sub-sub-issues such as refugees, ethics, riots, and so forth, until no further breakdown is possible. Consequently, information collection, processing, analysis, and services are conducted for the most specific issues, gradually working back up the hierarchy until the assertion that “war destabilizes the world” is substantiated.

Conversely, the problem-oriented analysis model takes the opposite approach. Initially, the information analyst possesses limited knowledge about the target under analysis and only has a single question to explore. The answer to this question must be analyzed using a diverse range of information sources. For example, a question such as “What will be the outcome of public opinion regarding the COVID-19?” is an inquiry lacking a definitive answer. To address such a question, we constructed a simplified opinion analysis model based on the improved DIKW framework, showcasing the basic process of the component-based information analysis method. This model is depicted in Fig. 8.

As previously mentioned, our assumption is based on the existence of a comprehensive component library that contains all the necessary components. However, in practical scenarios, it is possible to encounter new requirements that cannot be fulfilled by the components available in the current library. In such cases, the development of new components becomes necessary in order to expand the library and facilitate reusability. It is important to note that while we acknowledge the existence of more efficient problem-specific information analysis models within the academic community, our proposed models, namely the simple opinion analysis model and the improved DIKW model, are primarily designed to illustrate the concept of a component-based approach to information analysis, rather than pursuing alternative objectives.

In addressing the problem of public opinion analysis, as shown in Fig. 8, a crucial step is the identification of data sources. Following the concept of componentization, experts carefully choose relevant data sources,

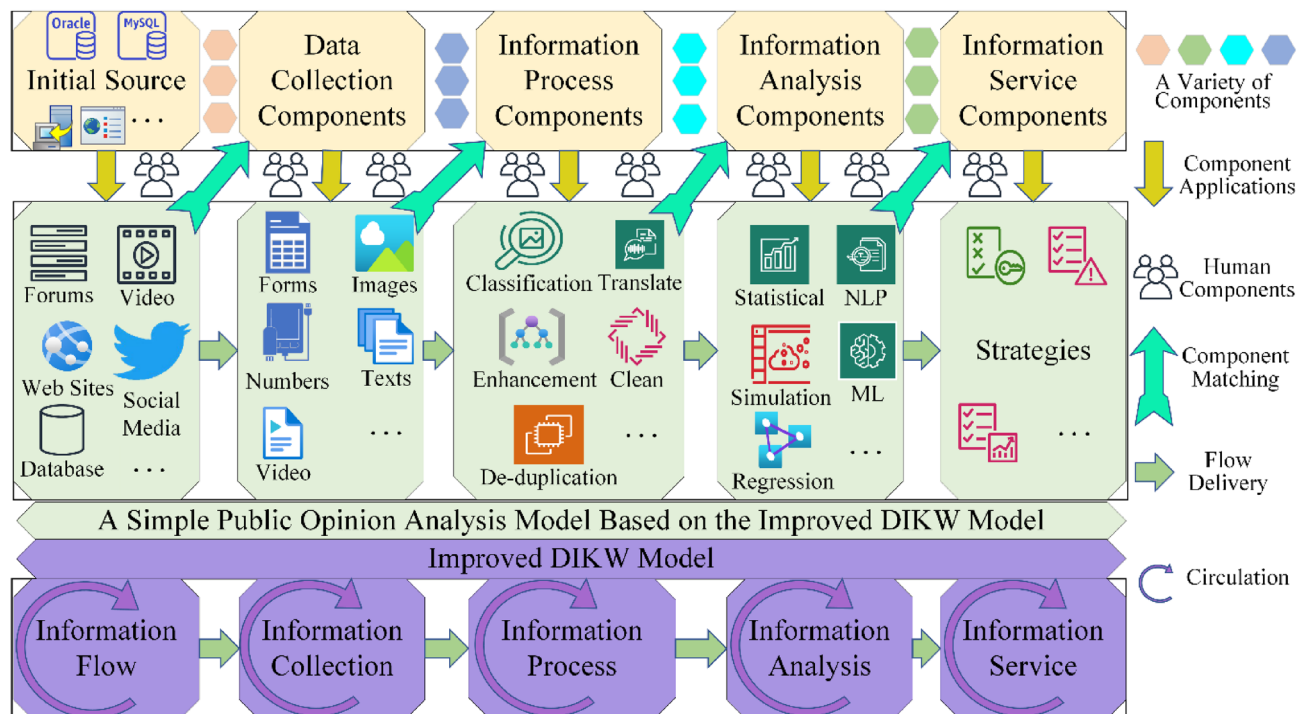


Fig. 8. Flow chart of public opinion analysis case.

which may include the World Wide Web, specific databases, forums, video sites, Twitter, and other social media platforms. Subsequently, experts consider the specific characteristics of each data source and select the corresponding data collection components from a library of such components. Each data collection component within the library is designed for collecting a particular type of data, enabling experts to conveniently choose the appropriate one based on the data source. These data collection components effectively capture diverse information related to the Covid-19 outbreak, encompassing spreadsheets, images, numbers, text, videos, and more.

Next, human experts employ data processing components from an information processing component library to handle the various types of data. These components facilitate operations such as translation, classification, cleaning, de-duplication, and data enhancement. The aim of these operations is to transform the data into a standardized format that can be readily identified and utilized by subsequent information analysis components. Once the data has been pre-processed, it becomes suitable for information analysis.

In the subsequent step, experts evaluate the characteristics of the data and the specific problem at hand. Based on this evaluation, they select appropriate information analysis methods, such as statistical analysis, natural language processing, simulation computing, machine learning, or regression analysis. Additionally, they choose the corresponding information analysis components from an information analysis component library to initiate the information analysis process.

Finally, experts identify suitable information service components based on the results of the information analysis. These components facilitate the presentation of the information analysis results to users and offer decision support for controlling epidemic opinions. Throughout the entire process, any involvement of experts is integrated into a human component that incorporates their expertise, thus enabling the reuse of their wisdom to achieve the desired goals.

The development process of IACs is undoubtedly more complex compared to their utilization. For instance, the development of information collection components necessitates the establishment of distinct standards tailored to different data types, ensuring universality. Furthermore, information processing components must consider the requirement to standardize data upon completion for the subsequent utilization by information analysis components. Similarly, information analysis components must address both forward compatibility with information processing components and backward compatibility with information service components. Additionally, all component types should consider the integration with human components, while the development of human components must account for the invocation of various components and the characteristics specific to the information domain. By doing so, the aim is to attain efficient and accurate information analysis.

Discussion and future work

In this study, we propose a theoretical framework of Information Analysis Componentization, drawing upon theories of knowledge management, Taylor's scientific management, and software component technology. This constitutes the primary theoretical contribution of our research. The framework advocates for the construction and assembly of IACs to accomplish information analysis tasks, aiming to enhance the reusability, modularity,

and engineering level of information analysis work. At the same time, it underscores that, even in the era of rapid advancements in artificial intelligence, human expertise remains the indispensable core of information analysis. Therefore, while promoting the computerization and automation of information analysis, it is equally crucial to systematically summarize, standardize, and codify human analytical behaviors to further improve overall efficiency and quality.

Information analysis guided by the idea of componentization exhibits characteristics such as ease of development, high reusability, and operational simplicity, significantly reducing both development and operational costs. Building on this foundation, we further introduce the concepts of cyclic iteration and spiral growth, and propose an Improved DIKW Model based on the traditional Data-Information-Knowledge-Wisdom hierarchy, representing the second major innovation of this study. The Improved DIKW Model integrates iterative refinement of requirement analysis and data value mining, aiming to maximize the effectiveness and practical value of information analysis outcomes through continuous optimization.

Subsequently, we take the Improved DIKW Model as an example to systematically elaborate on the methods for decomposing, assembling, and reconstructing the information analysis process under the guidance of the componentization framework, thereby demonstrating the feasibility and applicability of the proposed approach.

It is important to note that the case study conducted in this research primarily focuses on theoretical process analysis, providing preliminary explanations and validations of the feasibility of applying the componentization framework to information analysis tasks. However, it does not yet include empirical validation based on real-world data. In future work, we plan to undertake comprehensive empirical studies or simulation-based experiments to further validate and refine the theory of Information Analysis Componentization.

Data availability

All data generated or analyzed during this study are included in this published article.

Received: 10 May 2024; Accepted: 30 May 2025

Published online: 09 July 2025

References

1. Agarwal, R. & Dhar, V. Big data, data science, and analytics: the opportunity and challenge for IS research. *Inf. Syst. Res.* **25**, 443–448 (2014).
2. Chen, D. Q., Preston, D. S. & Swink, M. How big data analytics affects supply chain Decision-Making: an empirical analysis. *J. Assoc. Inf. Syst.* **22**, 1224–1244 (2021).
3. Qiu, J., Chai, Y., Tian, Z., Du, X. & Guizani, M. Automatic concept extraction based on semantic graphs from big data in smart City. *IEEE Trans. Comput. Soc. Syst.* **7**, 225–233 (2019).
4. Raviya, K. & Vennila, M. S. An implementation of hybrid enhanced sentiment analysis system using spark ML pipeline: a big data analytics framework. *Int. J. Adv. Comput. Sci. Appl.* **12**, 323–329 (2021).
5. Pan, M. et al. Dynamic human preference analytics framework: a case study on taxi drivers' learning curve analysis. *ACM Trans. Intell. Syst. Technol.* **11**, 8:1–19 (2020).
6. Chai, Y., Du, L., Qiu, J., Yin, L. & Tian, Z. Dynamic prototype network based on sample adaptation for Few-Shot malware detection. *IEEE Trans. Knowl. Data Eng.* **35**, 4754–4766 (2022).
7. Grover, V., Lindberg, A., Benbasat, I. & Lyytinen, K. The perils and promises of big data research in information systems. *J. Assoc. Inf. Syst.* **21**, 268–291 (2020).
8. Mithas, S., Chen, Y., Liu, C. W. & Han, K. Are foreign and domestic information technology professionals complements or substitutes?? *MIS Q.* **46**, 2351–2366 (2022).
9. Wang, Y., Liu, Y. & Zang, J.S. Research on component-based information analysis. *J. Intell.* **41**, 47–54 (2022).
10. Qiu, J. Analysis of human interactive accounting management information systems based on artificial intelligence. *J. Glob. Inf. Manag.* **30**, 294905 (2022).
11. Wang, Y., Shi, J., Li, M. & Liu, Y. Research on the information analysis model based on the component. *Libr. Inf. Serv.* **66**, 92–101 (2022).
12. Liu, Y., Alzahrani, I. R., Jaleel, R. A. & Al Sulaie S. An efficient smart data mining framework based cloud internet of things for developing artificial intelligence of marketing information analysis. *Inf. Process. Manag.* **60**, 103121 (2023).
13. Singh, S. S., Srivastava, D., Verma, M. & Muhuri, S. Information diffusion analysis: process, model, deployment, and application. *Knowl. Eng. Rev.* **39**, e11 (2025).
14. Qiu, J. et al. A survey on access control in the age of internet of things. *IEEE Internet Things J.* **7**, 4682–4696 (2020).
15. Kar, A. K. & Dwivedi, Y. K. Theory Building with big data-driven research - Moving away from the what towards the why. *Int. J. Inf. Manag.* **54**, 102205 (2020).
16. Mili, H., Mili, F. & Mili, A. Reusing software: issues and research directions. *IEEE Trans. Softw. Eng.* **21**, 528–562 (1995).
17. Teece, D. J. Technology transfer by multinational firms: the resource cost of transferring technological know-how. *Econ. J.* **87**, 242–261 (1977).
18. Turner, S. F., Bettis, R. A. & Burton, R. M. Exploring depth versus breadth in knowledge management strategies. *Comput. Math. Organ. Theory.* **8**, 49–73 (2002).
19. Yayavaram, S. & Chen, W. R. Changes in firm knowledge couplings and firm innovation performance: the moderating role of technological complexity. *Strateg. Manag. J.* **36**, 377–396 (2015).
20. Taylor, F.W. *Scientific Management* (Routledge, 2004).
21. Crnkovic, I. & Larsson, M. Challenges of component-based development. *J. Syst. Softw.* **61**, 201–212 (2002).
22. Ackoff, R. L. From data to wisdom. *J. Appl. Syst. Anal.* **16**, 3–9 (1989).
23. He, D. The research paradigm of science and technology intelligence under the engineering thinking style—tentative discussion on intelligence engineering. *J. China Soc. Sci. Tech. Inf.* **33**, 1236–1241 (2014).
24. Cai, H., Xu, Z., Zhang, W. & Shi, J. Research on the resource sharing mechanism of information component based on blockchain thinking. *J. Intell.* **43**, 176–182 (2024).
25. Bawack, R., Bawack, R. & Hey Librarian What can AI and analytics do for you: a systematic literature review and sociotechnical perspective. *Aslib J. Inform. Manage. (ahead-of-print)* (2025).
26. Neigel, A. R., Caylor, J. P., Kase, S. E., Vanni, M. T. & Hoye, J. The role of trust and automation in an intelligence analyst decisional guidance paradigm. *J. Cogn. Eng. Decis. Mak.* **12**, 239–247 (2018).
27. Wang, X., Cai, H., Zhang, W. & Shi, J. Research on the componentization of Portfolio-based information analysis. *J. Mod. Inf.* **44**, 143–152 (2024).

28. Trstenjak, M., Benešova, A., Opetuk, T. & Cajner, H. Human factors and ergonomics in industry 5.0—a systematic literature review. *Appl. Sci.* **15**, 2123 (2025).
29. Galvez, J. A. et al. Measuring Human Trainers' Skill for the Design of Better Robot Control Algorithms for Gait Training after Spinal Cord Injury. In *Proceedings of the 9th International Conference on Rehabilitation Robotics ICORR 231–234* (2005).
30. López-Robles, J. R., Otegi-Olaso, J. R., Gómez, I. P. & Cobo, M. J. 30 years of intelligence models in management and business: a bibliometric review. *Int. J. Inf. Manag.* **48**, 22–38 (2019).
31. Rowley, J. The wisdom hierarchy: representations of the DIKW hierarchy. *J. Inf. Sci.* **33**, 163–180 (2007).
32. McIlroy, M. D., Buxton, J., Naur, P. & Randell, B. Mass-produced software components. In *Proceedings of the 1st international conference on software engineering*, Garmisch Pattenkirchen, Germany. 88–98 (1968).
33. Kruse, J. The Automatic-indexing component of a German patent Information-system. *Nachr. Doc.* **40**, 99–101 (1989).
34. Ilk, N., Zhao, J. L., Goes, P. & Hofmann, P. Semantic enrichment process: an approach to software component reuse in modernizing enterprise systems. *Inf. Syst. Front.* **13**, 359–370 (2011).
35. Lau, K. K. In *Proceedings of the 28th International Conference on Software Engineering* 1081–1082 (2006).
36. Raza, S. K., Pagurek, B. & White, T. Distributed computing for plug-and-play network service configuration. *Network Operations & Management Symposium* 933–934 (2000).
37. McDowell, K. Storytelling wisdom: story, information, and DIKW. *J. Assoc. Inf. Sci. Technol.* **72**, 1223–1233 (2021).

Acknowledgements

Financial support from the National Social Science Found of China(NO.21BTQ012 and NO.19ZDA347) and the Nanjing University China Mobile Joint Research Institute Project is gratefully acknowledged.

Author contributions

Hongyu Cai: Present the research framework, write and revise the paper. Xiangfeng Wang: Collaborate to improve research framework, write and revise the paper. Jin Shi: Collaborate to improve the research framework and revise the paper. Xiaoyang Zhou: revise the paper.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.S. or X.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025