

# THE ROLE OF INFORMATION AND KNOWLEDGE ENGINEERING IN AI, BIG DATA AND CLOUD SYSTEMS

## Abstract

Information and knowledge engineering includes techniques for data processing, storage, and retrieval that incorporate artificial intelligence, neural networks, and the internet of things. They support expert systems, improve judgment, and facilitate effective data management. ANNs, decision trees, fuzzy logic, and rule-based systems are examples of knowledge acquisition techniques. Rule-based systems, semantic networks, and frames are used in knowledge representation. AI, big data, and cloud systems have been influenced by these concepts, opening the door for cutting-edge data-driven solutions.

**Keywords:** Knowledge engineering, expert systems, rule-based systems, decision trees, ANNs, fuzzy logic, knowledge representation, semantic networks, frames, big data, cloud systems, and data-driven solutions.

## Authors

### **Dr. Poornima Tyagi**

Assistant Professor  
Department of Computer Science and Engineering  
Noida Institute of Engineering & Technology  
Uttar Pradesh, India.  
poornima.tyagi@gmail.com

### **Ms. Pragya Singh**

Assistant Professor  
Department of Computer Science and Engineering  
Noida Institute of Engineering & Technology  
Uttar Pradesh, India.  
pragyasinh015@gmail.com

### **Dr. Ritesh Rastogi**

Associate Professor & Head  
Department of Computer Science  
Noida Institute of Engineering & Technology  
Uttar Pradesh, India.  
ritesh.rastogi@niet.co.in

## I. UNDERSTANDING INFORMATION AND KNOWLEDGE ENGINEERING

Information and knowledge engineering is a field that focuses on the development of methods and tools for representing, storing, retrieving, and processing information and knowledge. It is closely related to current technologies like AI, Big data, and Cloud.

It mainly involves the development and implementation of technologies that facilitate the creation, sharing, and management of information and knowledge within and across organizations. These systems are designed to optimize knowledge acquisition, representation, reasoning, and retrieval.

In the talk presented by Feigenbaum on “The Art of Artificial Intelligence” at the 5th International Conference on Artificial Intelligence(IJCAI’77) the concept of knowledge engineering is introduced as the foundation of expert systems[ Rudi].

Artificial Intelligence (AI) is a broad term that refers to machines that mimic human intelligence and cognitive functions like problem-solving and learning. AI uses predictions and automation to optimize and solve complex tasks that humans have historically done, such as facial and speech recognition, decision making, and translation.

Neural Networks are a subset of machine learning and are the backbone of deep learning algorithms. They are called “neural” because they mimic how neurons in the brain signal one another. Deep learning is a subfield of machine learning that uses neural networks with multiple layers to improve the accuracy of predictions and decision-making.

For example, an ontology represents knowledge as a set of concepts within a domain and the relationships between those concepts. Knowledge representation and knowledge engineering allow AI programs to answer questions intelligently and make deductions about real-world facts.

## II. THE ROLE OF IKE IN AI SYSTEMS

Knowledge engineering involves creating rules for data to mimic human thinking and approach issues. AI can deduce how an issue was resolved by breaking down the problem into components. This knowledge is fed into a system as raw data, allowing it to diagnose and provide solutions without human input. Knowledge representation is crucial, as it involves designing formal structures for AI systems to represent knowledge about the world, including concepts, relationships, and reasoning rules.

**1. Knowledge Acquisition:** Knowledge acquisition is the process of acquiring knowledge from experts or other sources and representing it in a form that can be used by AI systems. It is the process of extracting, structuring and organizing knowledge from human experts for use in software systems.

**There are Several Methods of Knowledge Acquisition, Including:**

- **Rule-Based Systems:** Rule-based systems, also known as production systems or expert systems. These systems use a set of rules or heuristics to make decisions.The

rules are typically derived from the expertise of subject-matter specialists and can be used to describe their knowledge in a way that can be used in AI

- **Decision Trees:** Decision tree is a type of algorithm that can be used for classification or regression tasks. It uses a series of if-then-else statements to arrive at a decision. In the field of AI, supervised learning algorithm uses decision trees for tasks like classification and regression. They are trained using datasets that have been labelled, where the input features are associated with their corresponding target labels or numerical values.
- **ANN:** ANN systems are excellent tools for learning from data, making predictions based on extensive relationships between inputs and outputs, and knowledge acquisition. They are useful for pattern identification, and detecting complex patterns, and may be applied to learn from labelled datasets. Expert systems that use ANNs may mimic expert decision-making in real-time applications by extracting knowledge and data from the past. ANNs, which assist in deducing conclusions and learning through data-driven rules, may also be used to do automated reasoning. ANNs are able to derive hypotheses from observed patterns and generate innovative conclusions from previously understood data.
- **Fuzzy Logic :** Particularly in expert systems, fuzzy logic is a powerful technique in AI for knowledge acquisition. In circumstances when decisions must be made utilizing unclear or inaccurate data, it aids in determining linguistic variables and membership functions to express ambiguity. On the other hand, rule-based systems use fuzzy logic to improve how well they can handle contradictory inputs and outputs. For instance, by assigning varied degrees of membership to distinct diagnostic categories, fuzzy logic helps manage ambiguity in medical diagnosis. The type of data that is available and the nature of the problem determines which approach is most appropriate for a particular application. Rule-based systems, for instance, are better suited to issues with obvious rules, whereas artificial neural networks are more suited to problems with intricate interactions between inputs and outputs.

### III. KNOWLEDGE REPRESENTATION

This is the process of representing knowledge in a form that can be used by AI systems. The various methods to represent knowledge:

1. **Semantic Networks:** These are graphical representations of knowledge that use nodes to represent concepts and links to represent relationships between concepts. It visually depicts how concepts are related to one another and illustrates the connections between them.

In AI and Big data, semantic networks are a potent tool for knowledge representation. They promote data integration, support structured information organization, improve search and retrieval, and aid in the comprehension and use of intricate linkages found in data. This supports the creation of increasingly sophisticated

and clever AI systems. They also successfully address the difficulties brought forth by big data's diversified and expansive datasets.

- 2. Frames** A frame might be compared to a data structure that contains several attributes and actions related to a specific notion or subject. It consists of fillers, which are the values assigned to certain spaces, and slots, which reflect various features or attributes of the notion. Pointers to other frames are another feature of frames that enables the modelling of hierarchical connections and inheritance.

A few advantages of AI frames include conceptual modelling, characteristics and values, hierarchical relationships, inheritance, contextual relationships, reasoning and inference, learning and generalization. Frames express concepts and entities in a systematic manner. It is possible to represent taxonomies and classification systems using hierarchical relationships, which show categories and subcategories. AI systems may represent specialized characteristics at lower levels and shared features at higher levels because frames can inherit attributes and behaviours from other frames. They can effectively record contextual relationships, making them crucial for AI applications like natural language understanding. They also facilitate reasoning and inference, which aids learning and generalization. In AI systems frames are used to understand user behavior and make individualized suggestions based on similar characteristics.

- 3. Rule-based systems:** Rule-based systems are essential for representing knowledge in AI. They provide transparency and enable logical reasoning, decision-making, and problem-solving. AI practitioners should utilize this tool to its fullest potential.

#### IV. RULE-BASED SYSTEMS IN AI

By using conditional statements or rules to create relationships between certain situations and actions, rule-based systems may capture expert knowledge. These systems contain a rule structure in which the prior set conditions and that follows the actions. These systems are employed in industries such as recommendation systems, expert systems, financial analysis, and medical diagnostics, depend heavily on inference and reasoning. When logic is required and clear information can be transformed into rules, these systems work effectively.

AI rule-based systems provide adaptability, understanding, and transparency in knowledge management. In dynamic domains like law and medical, where AI decisions are essential, these systems are appropriate. They are constrained, in how they deal with ambiguity, convoluted connections, and missing or inconsistent data. To resolve these difficulties, hybrid methods that integrate rule-based systems and machine learning are used. Hybrid techniques can improve the efficiency of AI in managing knowledge in dynamic environments through the combination of data-driven learning and structured reasoning.

**Ontologies:** Structured representations of knowledge, known as ontologies, are essential for facilitating the reasoning process of AI and Big data systems. They provide an organized framework for complex domains, depicting relationships and concepts in a particular field. Without ontologies, these systems would struggle to make sense of complex information and would be less effective overall.

Ontologies are organized frameworks that define concepts and relationships while representing knowledge in a domain or subject area. They are essential to AI systems' knowledge representation, reasoning, and semantic comprehension. Logical inference, semantic interoperability, natural language processing, structured data extraction, and data integration are all become accessible by ontologies. Additionally, they improve information extraction, relevance, and natural language comprehension, which makes them versatile and adaptive in the fast-paced AI environment.

## V. THE ROLE OF IKE IN BIG DATA SYSTEMS

"The availability of information and knowledge engineering can provide meaning to an enormous amount of unstructured data." - Jatin Thakkar

Information and knowledge engineering enables the development of big data systems that can effectively manage, process, and analyze large volumes of disparate data. By integrating advanced data mining and knowledge discovery techniques into big data systems, organizations can gain valuable insights into their data and make data-driven decisions.

- 1. Role of ANNs:** ANNs are effective in data analysis, predictive modeling, customization, and recommendation when used in Big Data scenarios. ANNs may find hidden insights and connections by examining historical data and precisely predicting future patterns. ANNs can also analyze massive volumes of data to predict events, trends, or anomalies by extracting considerable information from the data. Additionally, ANNs employ knowledge gathered from user interactions to offer tailored suggestions based on user interests and behavior. In big data contexts, ANNs enable the development of individualized experiences and individualized suggestions.
- 2. Fuzzy Logic:** As part of the acquisition of Big Data knowledge, fuzzy logic plays an important role in data scrubbing, data deduplication, and search and retrieval processes. It helps in finding comparable but different data and reconciliation it, especially when there is conflicting data from several sources.

Fuzzy logic also aids in data deduplication which tries to determine the degree of similarity between records in datasets with duplicates or near-duplicates. It also improves search and retrieval processes by taking into account the closeness of related search terms and available data, making it crucial for responding to client requests that diverge from the information already supplied. The techniques used in Fuzzy logic improve AI systems and Big Data analytics overall by assisting in intelligent decision-making and retrieving useful information from complex real-world contexts.

- 3. Frames:** In Big Data, frames are crucial for organizing complex information, capturing relationships and context, enabling semantic comprehension, and managing hierarchical data. They present data in an organized way, allowing for the modeling of things and the qualities that link to them. Frames may also be used to create parent-child connections, which makes them advantageous for classifying data. frames are essential in Big Data systems, for adjusting to changing data and an ever-changing environment. By identifying patterns and little reconfiguration, they improve learning and generalization. In a landscape that is continuously changing, frame adjustments offer agility and relevance.

4. **Rule based systems :** Big Data relies on rule-based systems to enable data reduction and filtering to extract relevant data. These systems assist decision-making, alerting, event detection, and pattern identification. They assist in reducing dataset size while preserving important data by deciding when data should be included or removed from analysis. Furthermore, rule-based systems facilitate real-time or almost real-time event detection, which is useful for applications like fraud detection, network security, and predictive maintenance. By encoding rules, deriving insights, and capturing semantic context, rule-based systems play a noteworthy part in big data. To overcome challenges, they may be integrated with machine learning.

They can preprocess data, remove noise, and then input it into algorithms for analysis and prediction. Traditional rule-based systems are capable of dealing with the complexity of Big Data by encoding domain knowledge and modifying rules to take changing data into account. These systems make it possible to assess and explain data, which makes them vital in regulated fields or for important applications.

## VI. THE ROLE OF IKE IN CLOUD SYSTEMS

Cloud-based systems, also known as cloud computing, offer cloud services and resources over the Internet. Instead of using local servers or physical infrastructure cloud computing uses remote servers hosted in data centres to store and process the data. These systems have several advantages scalability, flexibility, cost effectiveness and ease of access. Because users can access the services provided by the cloud system from anywhere with an active internet connection. They don't need to manage the physical resources and can easily manage their resources.

In the context of knowledge engineering cloud-based systems are used to manage large volumes of knowledge or information and provide access to those knowledge-based systems from anywhere. These systems also facilitate collaborative knowledge sharing and support the computational task required in knowledge processing and analysis. In combination with knowledge engineering cloud-based systems can lead to more efficient and effective knowledge-based applications and services for users to make better decisions and gain better insights from the data. Knowledge engineering and cloud based systems are related to each other in many ways:

1. **Knowledge Storage and Management:** A large volume of data and knowledge can be stored and managed easily in cloud based systems with the help of robust and scalable infrastructure. This knowledge is stored in the cloud using various databases, file storage systems or knowledge repositories and knowledge engineering is used to capture and represent this knowledge. The cloud's distributed and flexible nature enables efficient storage and retrieval of knowledge, making it accessible to users anytime and from anywhere.
2. **Collaboration and Knowledge Sharing:** Cloud based systems allow collaboration among users by sharing their knowledge across different users. Whereas knowledge engineering allows multiple users to contribute their expert ideas to a shared knowledge

base. So this collaborative approach of knowledge engineering and cloud-based system lead to more accurate and comprehensive knowledge-based applications.

**3. Knowledge Acquisition for Cloud System:** To enhance the capability of cloud systems knowledge acquisition methods are used which provides process of gathering, creating and integrating information. This complete process involves collecting relevant data from various sources , converting it to structured data and then use it for decision making. There are different data acquisition methods in cloud computing are:

- **Batch Data Transfer:** It includes transferring data into batches means organisations upload data in bulk to the cloud system for further processing. This method is suitable where the data is needed weekly or monthly.
- **API Integration:** Cloud systems integration with API (Application Programming Interface) can also fetch the data by programmatically.
- **ETL (Extract, Transform & Load):** ETL tool can also be used to acquire the data from different sources and transform it to required format and the load on the cloud system for further use.
- **File sync and Share:** Different tools like drop box , google drive and Microsoft one drive link can be created for file sharing on cloud system.
- **Data lakes and warehouses:** Centralized storage for both structured and unstructured data is used for data acquisition for subsequent analysis and querying.

**4. Knowledge Representation:** Knowledge representation combined with cloud computing improves the organisation's overall performance and enhances the capability of decision-making. The integration of knowledge engineering and cloud based systems offers several advantages which include efficient knowledge storage and processing, improved accessibility, scalability and improved accuracy. By combining the strengths of both we can develop a powerful knowledge based system. There are some methods which are used for knowledge representation in cloud based systems are:

- **Ontology:** Ontology is used in cloud computing due to its ability to give common access to all the information available in information layer. I cloud computing ontologies provide a standardized representation of concepts, relationships and attributes. Which enhances the decision making , communication and resource management in cloud systems.

It is very important to manage and distribute the resources very efficiently over the cloud system to utilize them efficiently. Ontologies represent the various attributes and relationship between different resources which helps in resource allocation based on customer requirement.

- **Rule Based Systems:** This is a system that contains set of rules to store and manipulate the data to solve a particular problem. In cloud computing set of predefined rules are used to evaluate the data to establish the relationship between them and to access those data effectively. These rules may apply in resource management and allocation, auto scaling , policy enforcement, workflow automation, policy- driven resource optimization and cost management.
- **Semantic Graphs:** To represent the relationships and connection between the entities and to store , represent and querying purpose semantic graphs are used in cloud based systems. So semantic graph technologies help to to capture the context and meaning behind the data and to enable more sophisticated data management. Semantic graphs also provide interoperability between different cloud based systems which is useful when integrating services, resources and data from diverse sources.

## VII. CHALLENGES AND OPPORTUNITIES

Information and knowledge engineering offers substantial challenges in alongside having huge potential to boost the capabilities of big data, AI and cloud systems. Organizations must discover methods to strike a balance between privacy and data protection concerns and the potential advantages of data-driven decision-making in order for these technologies to succeed. Furthermore, the increasing complexity of these systems poses difficulties for both design and maintenance.

## VIII. FUTURE DIRECTIONS

Information and knowledge engineering will continue to play an increasingly significant part in today's businesses as big data, cloud, and AI technologies become more commonplace. More complex AI systems will be created as a result of developments in robotics, machine learning, and natural language processing, while new big data systems will be created as a result of the continual growth of digital data.

Additionally, the Internet of Things and edge computing will be supported by the next generation of cloud systems, opening up new possibilities for information and knowledge engineering to influence the development of these technologies.

## IX. CONCLUSION

Building intelligent systems that can transform businesses and enhance their decision-making powers requires the integration of information and knowledge engineering with AI, big data, and cloud platforms. Even if there are still challenges, these technologies have a bright future ahead of them, and businesses that harness the power of information and knowledge engineering will be well-positioned to prosper in the years to come.



## REFERENCES

- [1] [Rudi 98] - Studer Rudi ,Benjamins V Richard, and Fensel.Dieter, “ Knowledge engineering: principles and methods”. Data & knowledge engineering, 25(1):161–197, 1998
- [2] “The Knowledge Engineering Review” is a journal that publishes reports of new research in symbolic AI, knowledge-based systems, knowledge engineering, and related areas<sup>1</sup>.
- [3] “Artificial Intelligence and Knowledge Engineering” is a conference article that incorporates concepts from other fields like Task (project management) and Pattern recognition<sup>3</sup>.
- [4] “Knowledge Engineering with Big Data” is a research paper that presents BigKE, a knowledge engineering framework that handles fragmented knowledge modeling and online learning from multiple information sources<sup>4</sup>.
- [5] “A Study on Big Knowledge and Its Engineering Issues” is a study that discusses the concepts of big knowledge, big-knowledge system, and big-knowledge engineering<sup>5</sup>.
- [6] “Journal of Cloud Computing” is a journal that publishes research articles on all aspects of Cloud Computing<sup>7</sup>.
- [7] “A Survey of Advanced Information Fusion System: from Model-Driven to Knowledge-Enabled” is a survey paper that summarizes an advanced information fusion system, from model-driven to knowledge-enabled<sup>8</sup>.
- [8] “Mapping Knowledge Area Analysis in E-Learning Systems Based on Cloud” is a research paper that conducts a thorough assessment of scientific production in the field of e-learning and cloud technology fields