

Evolution of Knowledge Representation and Retrieval Techniques

Meenakshi Malhotra

Dayanand Sagar Institutions, RIIC, Bangalore, 560078, India
Email: uppal_meenakshi@yahoo.co.in

T. R. Gopalakrishnan Nair

Saudi Aramco Endowed Chair, Technology and Information Management, PMU KSA
Email: trgnair@yahoo.com, trgnair@ieee.org

Abstract— Existing knowledge systems incorporate knowledge retrieval techniques that represent knowledge as rules, facts or a hierarchical classification of objects. Knowledge representation techniques govern validity and precision of knowledge retrieved. There is a vital need to bring intelligence as part of knowledge retrieval techniques to improve existing knowledge systems. Researchers have been putting tremendous efforts to develop knowledge-based system that can support functionalities of the human brain. The intention of this paper is to provide a reference for further research into the field of knowledge representation to provide improved techniques for knowledge retrieval. This review paper attempts to provide a broad overview of early knowledge representation and retrieval techniques along with discussion on prime challenges and issues faced by those systems. Also, state-of-the-art technique is discussed to gather advantages and the constraints leading to further research work. Finally, an emerging knowledge system that deals with constraints of existing knowledge systems and incorporates intelligence at nodes, as well as links, is proposed.

Index Terms— Informledge System, Knowledge-Based Systems, Knowledge Graphs, Ontology, Semantic Web

I. INTRODUCTION

Information sharing has been one of the important aspects of human interactions. From cave paintings to the current World Wide Web (WWW) the need to share information has led to technological changes. WWW has emerged as a huge information storage that accumulate immense information from numerous domains. Initially only data was stored and used in its raw form, subsequently it was structured to provide data as useful information. This information has strewn on the web for a long time leading to the quest for knowledge to be retrieved through defined reasoning from the stored information [29].

Studies in the field of knowledge have put forward a differentiation among data, information, knowledge and wisdom as data-information-knowledge-wisdom (DIKW) hierarchy [55]. DIKW hierarchy is also referred to as knowledge hierarchy or information hierarchy or more commonly as knowledge pyramid [3, 15].

The knowledge pyramid represents data in its raw form that can further exist in any form and can be recorded. Information is referred as relationally connected data and

knowledge as structured information whereas wisdom is knowledge in use [7,38]. The hierarchical model depicts four components of the pyramid that are linked linearly, along with interconnectivity among the four components as shown fig. 1. Volume of content involved reduces towards the vertex of the DIKW pyramid, shown in fig.1. However the usability of the content provided at each level, increases towards the vertex [57]. Increased wisdom implies more relational connections within the content resulting in an increase of available useful information.

Data that forms the basis of human information system has no meaningful existence of its own without its ability to inter-connect. Strength of connectivity between data points distinguishes data from information. Connected information, when used to perform a task or provide a solution to a given problem, is treated as knowledge. Knowledge in turn coupled with experiences embodies wisdom to the system.

In addition to four components of DIKW, intelligence and innovation also belong to the pyramid. Knowledge is referred as intelligence when applied to derive solutions to problems in an efficient way. Intelligence, when applied to a new task, is said to be innovation and lies between knowledge and wisdom in the DIKW pyramid. It is this intelligence, which needs to be incorporated into the knowledge and information retrieval system in hand.



Fig. 1. DIKW Pyramid.

Section II provides an overview over related work done for the preliminary knowledge systems and section

III discusses some of the knowledge representation and retrieval schemes used later. Section IV discusses state-of-the-art knowledge system and its techniques. Section V briefs about the future scope and upcoming intelligent knowledge systems and finally section VI provides the conclusion.

II. RELATED WORK

The need to utilize data effectively and retrieve substantial results has been the focus since computer systems were invented. Knowledge representation has arisen as a major discipline of Artificial Intelligence (AI) in computer science. There is no universally accepted definition of AI. AI is a combination of other fields namely machine learning, knowledge representation, ontology-based search, Natural Language Processing (NLP), neural networks, image processing, pattern recognition, robotics, expert systems, and many others [52]. As defined by Barr & Feigenbaum, “*Artificial Intelligence (AI) is part of computer science concerned with designing intelligent computer systems, that is, systems that exhibit characteristics we associate with intelligence in human behavior – understanding language, learning, reasoning, solving problems, and so on*” [4].

Knowledge representation has been the main component involved in constructing intelligent knowledge systems and knowledge-based systems. Knowledge has been the main focal point for knowledge representation [13, 29, 56]. Knowledge, as possessed by human brain, has been classified broadly into two types namely tacit and explicit knowledge [4]. Tacit knowledge, also known as informal knowledge, is defined as knowledge that is hard to share as the same cannot be put across completely through vocabulary. It is gained through experiences, intuition, insights and observations. It is said to be within the subconscious human mind. Contrary to tacit knowledge is explicit knowledge that is easy to share, communicate and store by means of a combination of different vocabularies. It is also referred as articulated knowledge [52, 63]. The existing information and knowledge system deals with explicit knowledge that is further categorized as shown in fig. 2.

- **Domain knowledge:** It represents knowledge pertaining to a specific group.
- **Declarative knowledge:** It describes what is known about the problem.
- **Procedural Knowledge:** This knowledge provides direction on how to do a particular task or provide a solution.
- **Commonsense knowledge:** General purpose knowledge supposed to be present with every human being.
- **Heuristic Knowledge:** Describes a rule-of-thumb that guides the reasoning process. Heuristic knowledge is often called shallow knowledge.
- **Meta Knowledge:** Describes knowledge about knowledge. Experts use this knowledge to enhance the efficiency of problem solving by directing their reasoning in most promising area.

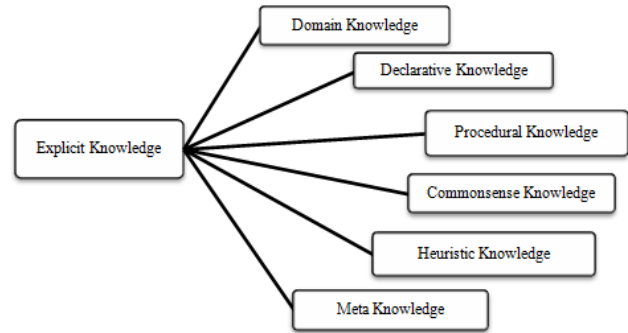


Fig. 2. Classification of Explicit Knowledge.

Knowledge representation involves different schemes namely logical schemes, procedural schemes, networked schemes, structured schemes [56]. Networked scheme has seen continuous growth over time. On the other hand, logical and procedural scheme works with a fixed set of symbols and instructions that get limited with the increase in information to be encoded. Structured schemes utilize a complex structure for node in the graph thereby restricting its wide usage [46]. Networked schemes have simple nodes in the graph, which stores data and allows an enormous amount of data to be embedded into the system in the form of nodes.

Tolman had introduced the concept of cognitive maps [66]. Cognitive maps provide mental representation of spatial information as knowledge [30]. However, cognitive map does not possess any of the cognitive processing of its own. Kosko introduced a fusion of fuzzy logic and the cognitive map as Fuzzy Cognitive Maps (FCM) [2]. FCM utilizes fuzzy logic to compute the strength of the relations. In 1976, Sowa developed conceptual graphs (CG) to represent the logic based on the semantic network using a graph [60, 61]. A CG is a finite, connected bipartite graph where a node either represents concept or conceptual relationship. Arcs are only allowed between concept and the conceptual relationship and not between two concepts or two conceptual relationships. The Conceptual Graph Interchange Format (CGIF) is a dialect specified to express common logic provided in CG [62]. However, CG was merely a structured representation of given information that was difficult to scale up and also lacked intelligence.

Concept Maps (CM), designed and established by Novak and Gowin [48], is a hierarchical structure that depicts hierarchy of concepts through the relationship between concepts. Here the concepts are represented by words that are linked through labeled arc. CM is used to understand the relationship between words as concepts [47]. CM has found its usability in learning as well as in assessing learning for a small number of connected concepts. It is also used to depict structuring of organizations and help administrators to manage organizations [12]. However, CM does not provide any structure for knowledge retrieval. Some of the important issues faced while dealing with the above mentioned systems were as follows:

- What primitive to be defined and how to use them to structure knowledge?
- Knowledge and concepts to be represented at what level?
- How to represent sets of objects? With so far available representation schemes, it was difficult to represent sets of objects.
- How to define different types of explicit knowledge using the same representation scheme?
- How to retrieve knowledge partially or fully when required?

Later, Sowa proposed Semantic network as another field of knowledge representation. It is a graphical structure of interconnected nodes where nodes are connected by an arc to represent knowledge [62]. AI applications for knowledge representation and manipulation utilize computer architectures that support semantic network processing [14]. Semantic network is based on the understanding that relationships between concepts define knowledge. In semantic networks, nodes represent concepts, objects, events, and time and so on, whereas an arc represents relationship as a directed arrow between nodes with a label. Semantic networks are classified broadly into following six categories [62]:

- *Definitional network*, also called as generalization hierarchy, supports the inheritance rule with or is-a between two defined concepts. Information in these networks is assumed to be inevitably true.
- *Assertional network* like relational graphs, conceptual graphs, asserts prepositions using first-order logic. Unless explicitly marked, information in an assertional network is assumed to be conditional true.
- *Implicational networks* are propositional networks where primary relation between nodes is of implication. They are also called as belief networks, causal networks, Bayesian networks, or truth-maintenance systems depending on the reasoning applied to the connectivity.
- *Executable networks* are network that commonly use mechanism like message passing, attached procedures, graph transformations to cause change to the network by itself.
- *Learning networks* are put together by acquiring new knowledge that may add or delete nodes and arcs or even update the weight of arcs. This change in the network enables knowledge system to respond effectively.
- *Hybrid networks* work with closely interacting networks by combining two or more techniques, either in a single network or separate networks.

There has been continuous research to build a system that can model properties of the human brain either partially or fully. Emulation refers to a system model that inhibits all the relevant properties of the actual system, and simulation refers to a model that includes only some of the properties [58].

Researchers from the field of knowledge systems have built Artificial Neural Network (ANN) as an effort to simulate the functionality of the human brain. Neural

network incorporates learning into their system by changing the weights assigned to the nodes or arcs of the network. The principles of ANN were first put together by neurophysiologist, Warren McCulloch and young mathematical prodigy Walter Pitts in 1943 [19]. ANN has been found useful in various fields of knowledge processing and learning. LAMSTAR (LARGE Memory STORAGE And Retrieval) neural network uses Self-Organizing Map (SOM)-based network modules. LAMSTAR is specially designed for problems with a large number of categories for storage, recognition, comparison and decision [19].

ANN has been useful in providing a solution to complex numerical computations although it has not been beneficial for solving simpler problems like balancing checks and many others. It is required to tailor ANN to a specific problem that needs to be solved [35], which makes it specific in nature. Additionally, users need to get proper training to select a starting prototype as proper paradigm from multiple potential ones.

Blue Brain team, jointly with European company, EPFL attempts to emulate the human brain under the Human Brain Project (HBP). HBP aims at understanding human brain processing mainly to reform the fields of neuroscience, medicine and technology [65]. In order to acquire an enormous amount of data to cover all possible levels of brain organization, the project involves an IBM 16,384 core Blue Gene/P supercomputer for modeling and simulation. However, HBP faced challenges in replicating the brain model as a single system that involves collaborating neurological data from varying biological organizations. The developments in the field of knowledge-based system (KBS) in AI and ANN have stimulated the emergence of another field in knowledge representation, termed as knowledge-based neurocomputing (KBN). KBN involves a neurocomputing system with methods to provide an explicit representation and processing of knowledge [11]. However, not much has been done in the field of KBN. There is a need to develop an intelligent system that can combine the neurocomputing and knowledge representation techniques. Next section analyzes many more knowledge systems.

III. KNOWLEDGE REPRESENTATION AND RETRIEVAL TECHNIQUES

In early 1980's, remarkable developments in various fields of AI had brought the need of expert systems. Initial KBS were primarily expert systems. KBS and expert systems include two major components; one is knowledge base, and other is an inference engine. However, the two systems can be distinguished based on how and for what the system is being utilized. The expert system had been usually build to substitute or assist a human expert in efficiently resolving a complex task with less complexity to save time [56]. On the other hand, KBS provides a structured architecture to represent knowledge explicitly [23].

KBS evolved as the knowledge base became structured, representing information using classes and subclasses where relations between classes and assertions were represented using instances [67]. MYCIN was the first KBS developed for medical diagnosis, under lisp platform. The inference engine EMYCIN was extrapolated from MYCIN, so that it can be made available for other researchers [67]. However, it was not practically used, as it would take approximately thirty minutes of time to respond to system queries where the program was run in large time –shared system.

KBS has been an active area of research which has faced a limitation referred as knowledge acquisition bottleneck. The information available in a particular domain is so vast that it creates a bottleneck to extract all the necessary information from the human experts in the form of rules into the inference engine.

KBS is different from the database. Databases answer query based on what is explicitly stored. However, the knowledge in KBS is stored primarily to reason and retrieve information from this system [70]. The field of knowledge representation has seen the development of knowledge management (KM) products which involve multi-disciplinary organizational knowledge as a repository of manuals, procedures, reusable design and code, error handling documents and many more. The objective of KM is mainly to store and reuse the information where KBS is an automated system that involves reasoning. In the current context, the discussion is restricted to KBS, which intends to depict some of the complexities of human information storage and processing.

With the advancement in the techniques used to represent knowledge in knowledge bases, it was possible to use the structure of classes and subclasses, and their instances more commonly implemented today as ontology, whose foundation is laid down from the field of philosophy. It was established in the field of computer science by Thomas R Gruber and is defined as “ontology is formal, explicit specification of a shared conceptualization” [20]. Ontology provides a formal specification explicitly for the terms and their relationships associated with other terms in a particular domain. This specification builds up a vocabulary to define the terms in the domain [21]. Ontology finds its usage over the earlier existing methodologies as it allows the following:

- The domain users, as well as the software agents, share the understanding of the domain structure and thus analyze the domain knowledge.
- Domain assumptions and the conceptualization are made explicit.
- The above stated features enable sharing and reuse of domain knowledge.

Data acquisition in accordance to the ontology builds the knowledge base which, when used by domain applications and software agents, provides problem-solving methods [49]. One of the earliest knowledge bases that utilize ontology for its construction is Cyc [51]. Cyc, which started in 1984, intends to construct a single

knowledge base to represent the complete commonsense knowledge. Cyc has been there for last three decades, and one of the biggest limitations it faces is in structuring and adding new knowledge into knowledge base. The other major challenge faced by Cyc was to provide completeness to the knowledge base in depth.

These limitations have led to the classification of ontologies. Guarino classified ontologies on the basis of information, which has been put together as concepts and structuring of the concepts.

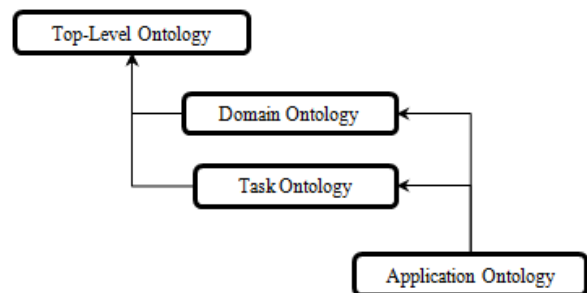


Fig. 3. Guarino's kinds of Ontology

Fig. 3 depicts Guarino's ontologies and their inter-relationships [21]:

- Top-level Ontologies include concepts that do not belong to any domain, but are general concepts like time and object.
- Domain Ontologies and Task Ontologies include concepts that are related to a specific domain like neuroscience and medicine, or to a specific task like diagnosing.
- Application Ontologies include concepts that are usual specializations of both domain and task ontologies.

This classification has been further fine-grained into six categories namely [51]:

- Top-level Ontology, more commonly referred to as upper ontologies, provide structure for the top-most general concepts. However, structuring and defining boundaries for top-level basic concepts has been a difficult task. It had been a controversy to make it either too narrow or too broad [64].
- Domain Ontology provides vocabulary to represent conceptual structure of a particular domain. They are usually connected to some top-level ontology so that there is no requirement to include the common basic knowledge.
- Task Ontology provides a conceptual structure to define the most general concepts for basic tasks and activities.
- Domain-Task Ontology provides a conceptual structure for domain centric concepts pertaining to the domain specific tasks and activities [64].
- Application Ontologies include concepts that are usual specializations of both domain and task ontologies.
- Method Ontology provides a structuring of concepts definitions to stipulate the reasoning process in order to accomplish a given task
- Application Ontology provides structuring of concepts at the application level.

The errors in the ontology development model have an influence over the information retrieved from the ontology-based system [10]. Ontological modeling of concepts faces numerous challenges. Some of these challenges are [50, 32]:

- Challenge faced in classification of concepts and relationships.
- Difficulty in representing concepts across domains unambiguously.
- How to represent instances of a concept in hierarchical structuring of concepts?
- How to define the relationship between similar concepts?
- Complexity in representing a single concept, having multiple meanings.
- Intricacies of representing multiple concepts with similar meaning.
- Complexity in assimilating new concepts in-between two or more concepts

Top-level Ontology aims at building an upper ontology, which provides support for semantic interoperability. Some of the upper ontologies developed are, namely, Cyc, Suggested Upper Merged Ontology (SUMO), Basic Formal Ontology (BFO), WordNet, DOLCE, COmmon Semantic MOdel (COSMO), General Formal Ontology (GFO), Unified Foundation Ontology (UFO) and many others [37]. Developing an upper ontology involves a proper interpretation of primitive concepts [45]. The biggest challenge faced in the development of the upper ontology is involved in building a completely compatible concept structure of top-level concepts that can be integrated with multiple domains [8]. Majority of the knowledge systems developed today incorporate the ontological structure, which have proved to be beneficial in various domains, but are strongly limited by ever increasing content and its coherent integration with information from multiple domains.

IV. STATE-OF-THE-ART: KNOWLEDGE REPRESENTATION SYSTEMS

The field of knowledge system has seen enormous development over the last two decades. In order to use information for automation, integration, decision-making and reuse across various applications, there has been a need to define and link the information that flooded the WWW. Semantic Web is large semantic network that commits in providing a common framework that allows data to be shared and reused across application, enterprise, and community boundaries [22].

In order to manage and automate the huge volume of data on the web, Semantic Web has defined standard web ontology language (OWL). OWL provides a common syntax to define vocabulary that defines terms and relationships to represent the knowledge in the domain [59]. Semantic Web provides semantic interoperability of data through technologies such as RDF, OWL, SKOS, and SPARQL as given by W3C standards [5].

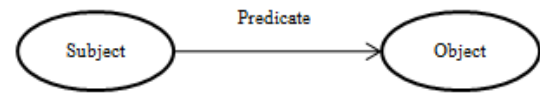


Fig. 4. RDF Triple

The Semantic Web architecture is a layered stack of its standardized languages, at the bottom of this stack lays the Resource Description Framework (RDF) [24, 33]. RDF is the foundation for processing metadata and its data model consists of three object types [31]:

- *Resources*: RDF expression describes things like a part or an entire webpage or an entire website. These things are termed as resources and are named by Uniform Resource Identifiers (URI).
- *Properties*: A resource is described using properties that can be an explicit characteristic, attribute or relation, with a specific meaning, the types of resources it can describe and its relationship with other properties.
- *Statements*: The RDF statement, also referred as RDF triple, consists of three components namely, subject, predicate and object as shown in fig. 4. RDF graph consists of a set of such triples or statements, where subject and objects are represented by nodes and predicate by an arc.

RDF uses URI references to identify resources and properties. Each URI represents a concept. Fig. 5. represents RDF statement: Richard Cygniak de is editor of rdf11-concepts identified by URI given for the subject. RDF schemas (RDFS) is RDF vocabulary description language provide generalized hierarchical structure of classes and properties [9]. OWL is layered on top of RDF and RDFS as it is widely used and is more expressive [17]. SPARQL, the Semantic Web query language, helps to query and extract data from RDF graphs instead of tables.

Semantic Web, which came into the picture around a decade ago, is undergoing research and development with an aim to organize the vast amount of information available on the internet, for faster access. Users from diverse fields have submitted case studies to Semantic Web, some of them include, healthcare, sciences, financial institution, automotive, oil and gas industry, public and government institutions, defense, broadcasting, publishing and telecommunications [25, 5]. These case studies comprise of the following usage areas namely, data and B2B integration, portals with better efficiency.

Few of the applications that have been benefitted by Semantic Web technologies in the recent past are listed as [68, 27]:

- The Health Care and Life Science Interest Group (HCLS IG) is one of the first application group, set up in 2005, to demonstrate the usability of Semantic Web technologies to the HCLS. The main objective behind the Allen Brain Atlas, HCLS demo, was the access and integration of public datasets via the Semantic We in areas like



Fig. 5. RDF statement represented by a triple.

drug discovery, patient care management and reporting, publication of scientific knowledge, drug approval procedures, etc. [68].

- Biogen Idec manufactures pharmaceutical products. It is famous for manufacturing drugs that are used to treat multiple sclerosis. This industry utilizes Semantic Web technologies in managing its global supply chain.
- Chevron oil and gas industry has been experimenting over a range of applications with Semantic Web technologies. One of the experiments in the field of data integration has provided a better understanding and ability to predict daily oil field operations, to the engineers and researchers by integrating random data in arbitrary ways [27].
- Web Search engines and Ecommerce websites have incorporated the metadata with their existing content, which have proved to be useful in returning relevant results. These search engines allow search based on local ontology and ontological reasoning. Facebook has developed the Open Graph Protocol, which is very similar to RDF. Similarly, search engines like Microsoft, Google, and Yahoo use Schema.org, which has an RDFa representation. Some of the popular websites using Semantic Web technologies behind the scenes are: Elsevier's DOPE browser, intelligent search at Volkswagen, Yahoo! Portals, Vodafone live! GoPubMed, Sun's White Paper and System Handbook collections, Nokia's S60 support portal, Oracle's virtual press room, Harper's online magazine, and many more.
- British Broadcasting Corporation (BBC) utilized Semantic Web technologies for the maximum public usage till date. Semantic Web technologies have created ontology for the world cup model that loaded concepts and relationships from various sources to create RDF-triple store, to run the World Cup 2010 website. Semantic Web technologies have benefitted the processing media content that observes constant change in usage patterns, as well as significant cross-document relatedness.

With the growth in the applicability of Semantic Web technologies, there have been demands for development of tools that help in building Semantic Web and linked data applications. The Semantic Web development tools need to be there for various categories namely, triple stores, inference engines, converters, search engines, middleware, CMS, Semantic Web browsers and development environments. Apache Jena is a free and open source Java framework is one such tool [16]. Few more tools for the above mentioned categories are AllegroGraph, Mulgara, Sesame, flickurl, TopBraid Suite, Virtuoso, Falcon, Drupal 7, Redland, Pellet, Disco, Oracle 11g, RacerPro, IODT, Ontobroker, OWLIM, RDF

Gateway, RDFLib, Open Anzo, Zitgist, Protégé, Thetus publisher, SemanticWorks, SWI-Prolog, RDFStore and many more. Protégé has been identified as the most widely used ontology development [71].

Big datasets have been created or linked to existing datasets to provide vocabulary to the Semantic Web applications. These datasets include IngentaConnect, eClassOwl, Gene ontology, GeoNames, FOAF (Friend of a Friend) ontology [26].

The DBpedia and Freebase are the two main linked data sources that have emerged in the recent past and have been able to connect various other datasets. Connecting various datasets was the main objective defined for Semantic Web. DBpedia is a community effort to build a knowledge base that can be used to extract structured data using expressive queries [69, 34]. The knowledge base of DBpedia is constructed automatically from Wikipedia, which is one of the biggest encyclopedias on the web. DBpedia extracts information from the Wikipedia page either in dumps or live. DBpedia-live extracts continuous streams updates from Wikipedia [1]. Extracting data from Wikipedia has been a challenge as its search is restricted to keyword matching and inconsistent data over multiple pages [40]. DBpedia is a step towards fulfilling the objectives of Semantic Web. In doing so, it faces following challenges [27, 28, 40]:

- Identifying the triples to be extracted or deleted when the article changes is challenging.
- There is a limitation due to inaccurate and incomplete data stored with infobox.
- Temporal and spatial dimension have not been considered
- DBpedia heavy-weight release process, which involves monthly dump-based extraction from Wikipedia, does not reflect the current state and requires manual export.
- Its large schema also poses a challenge during extraction.

Though there were continuous developments in Semantic Web technologies over the last decade, yet more need to be done to make all the data on the web to be integrated together. In doing so, researchers have come across many challenges as listed below [6, 24]:

- It requires technical experts' time and knowledge to construct RDF for all the existing web pages.
- Some of the web browsers do not support the web pages written in RDF.
- It is difficult to implement for non-technical users.
- There are issues with query performance because of the distributed structure.
- Web page creation for Semantic Web requires more time as it needs to specify the metadata along with the usual web page development.
- One of the main challenge faced by Semantic Web development is surrender of privacy and anonymity of personal web pages
- The authenticity of the content provided is questionable. Semantic Web development faces deceit when the information provider intentionally misleads

the user. This risk is alleviated to a certain extent by incorporating cryptography techniques.

- In addition Semantic Web development has to deal with huge inputs and the infinite number of pages on the WWW, which makes it difficult to purge all semantically duplicate terms.
- Different ontologies from numerous sources are pooled together to form large ontologies. However, the process usually results into inconsistencies arising from logical contradictions, which cannot be dealt using deductive reasoning.
- There are vague concepts to be dealt which are provided by content providers like big, small, few, many or old. Such vague concepts also arise when the user queries try to combine knowledge bases with similar but slightly different concepts.
- In addition to vague concepts, there are some precise concepts that hold indecisive values. Probabilistic reasoning techniques are used, to address this uncertainty.
- With the scalability of Semantic Web, there are challenges in representing the knowledge gathered from different resources with multiple formats.
- The researchers of Semantic Web development have been continuously questioned about the feasibility of a complete or partial accomplishment of Semantic Web.

Freebase a collaborative community-driven web portal which includes structured information, which is built to address some of the challenges faced during Semantic Web development where large scale information from different sources needs to be integrated [18]. Freebase contains data integrated from various sources such as Wikipedia, ChefMoz, Notable Names Database (NNDB), Fashion Model Directory (FMD) and MusicBrainz, it also includes data contributed data from its users. Freebase was initially developed by Metaweb Company and later acquire by Google in 2010. Freebase is a type system that includes the following objects [18]:

- *Topic*: Each topic represents exactly one concrete and specific entity where each topic is identified using globally unique identifier (GUID). It may have multiple names.
- *Literal*: It can be a scalar string, a numeric value, Boolean, or timestamp.
- *Type*: It groups entities i.e. topics semantically. Topics linked to a type are considered to be instances of the type. A particular topic can belong to multiple types e.g. a teacher named 'XYZ' belongs to both person and teacher type.
- *Property*: Properties defines the qualities for the type. They can be literals or relationship to another topic.
- *Schema*: Schema for an object type specifies a collection of zero or more properties of that type. If a topic is an instance of type, then all the properties specified in the schema of the type would be applicable to that topic.
- *Domain*: Types are grouped into domains, which specifies a particular knowledge domain

Freebase utilizes graph model to structure the data. Every fact present in Freebase is contained as a triplet RDF and N-triples RDF is referred as a graph. Google uses Freebase and has made it open source product. Freebase is referred as the free knowledge graph with Google. It is inevitable for freebase to contain inaccurate data, being an open source product. Also, knowledge retrieval is restricted by quota limit set for Freebase API calls.

Freebase RDF dump have been used by many products, one of them is ':BaseKB' which is a product from the company, Ontology2. :BaseKB is an imminent knowledge base enabling interactions of web applications by utilizing information supplied by Dbpedia, Facebook open graph and other semantic-social systems. The integrated data provided from numerous domains are queried using SPARQL. :BaseKB eliminates invalid and unnecessary facts from the content that makes it more efficient than the Freebase. However, knowledge retrieval in both the systems is restricted by the content structuring where there can be a single connectivity between concepts.

Managing connected data within the upcoming business applications have been a concern with the already existing relational databases. The traditional relational database systems are good at processing related tabular data. However it is easier and efficient for the current knowledge-based retrieval systems to represent the interrelated and rich data using a graph database. A graph database incorporates graph structures with nodes, edges and relationships to store and represent the interrelated data for efficient retrievals. The graph traversal provides answers to user queries. Many graph database projects have been under development like AllegroGraph, BigData, Oracle Spatial and Graph, Sqrrl Enterprise, Neo4j and many more.

Some of these are RDF graphs, and others are property graphs. A property graph stores data in nodes that are connected by directed, type relationships and properties resides on both the nodes and relationships. One the most popular open source graph database is *Neo4j* [54]. Graph databases have index-free adjacency property that enables a given node to find its neighbors without considering its full set of relationships in the graph [39]. Neo4j uses Cypher as the graph query language, where it need to specify explicitly the nodes key-value pair and relationship property between the nodes, as well as the directed arrow [53]. An example to create a graph shown

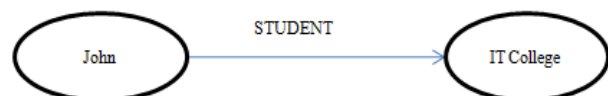


Fig. 6. A graph to represent Student Information in database.

In fig. 6 for the relationship "John is a student of IT College" using Cypher is as below:

```
Create (john),
(IT College),
(john)-[:STUDENT]->(IT College);
```

The graph databases have an advantage over the conventional relational databases whereby data insertion is not restricted by defined table design and also supports processing of unstructured data. At the same point, the directivity of relationship between nodes in the graph has to be provided explicitly. This explicit representation depends on user inputs and intelligence that poses a limitation in intelligent information storage.

V. CURRENT SYSTEMS LIMITATIONS AND FUTURE SCOPE

The developments in information representation techniques are shown in fig. 7. Knowledge representation and retrieval techniques mentioned so far deal with information as connected words at the time of input and processing. There is a need to develop new information representation technique that could incorporate innovative and intelligent knowledge retrieval properties into the system.

The usage of concepts has been restricted to representation of words. However, to represent the concept there is a need to connect with related sub-concepts e.g. Cow, as a word means nothing unless it is associated with its properties. Thus, the set of connected sub-concepts make a concept. Also, these systems fail to provide dynamic connectivity between existing nodes, wherein any new relationship needs to be specified using separate rules. Many of the social networks like Google, Facebook, and Twitter have included graph databases.

Graph databases as mentioned earlier provide explicit connectivity between nodes whereas human brain network does not provide a fixed and explicit connectivity between the neurons [36].

The researchers have believed that the information in the human brain, as well as information in knowledge and information systems, is stored as a network of inter-connected nodes. However, the human brain network and human-made knowledge systems differ considerably in the way nodes are structured, connections between the nodes are made and the efficiency with which knowledge is retrieved. In the human brain, network links have varying properties that help in their fast or slow knowledge retrieval [58]. There is a need to develop a knowledge system that can provide for an autonomous node with an ability to decide the subsequent connectivity. In addition, connectivity between nodes is not just an assigned string of relationship but where the intelligence of the network lies.

Another promising approach for the development of intelligent knowledge system is provided by Informedge System (ILS). This knowledge system provides intelligent knowledge retrieval from the stored information by virtue of ILS autonomous nodes and the multilateral links [41, 42]. It follows a distinct way of representation for its nodes with four quadrant structure to provide processing capabilities unlike the nodes provided by the other knowledge systems.

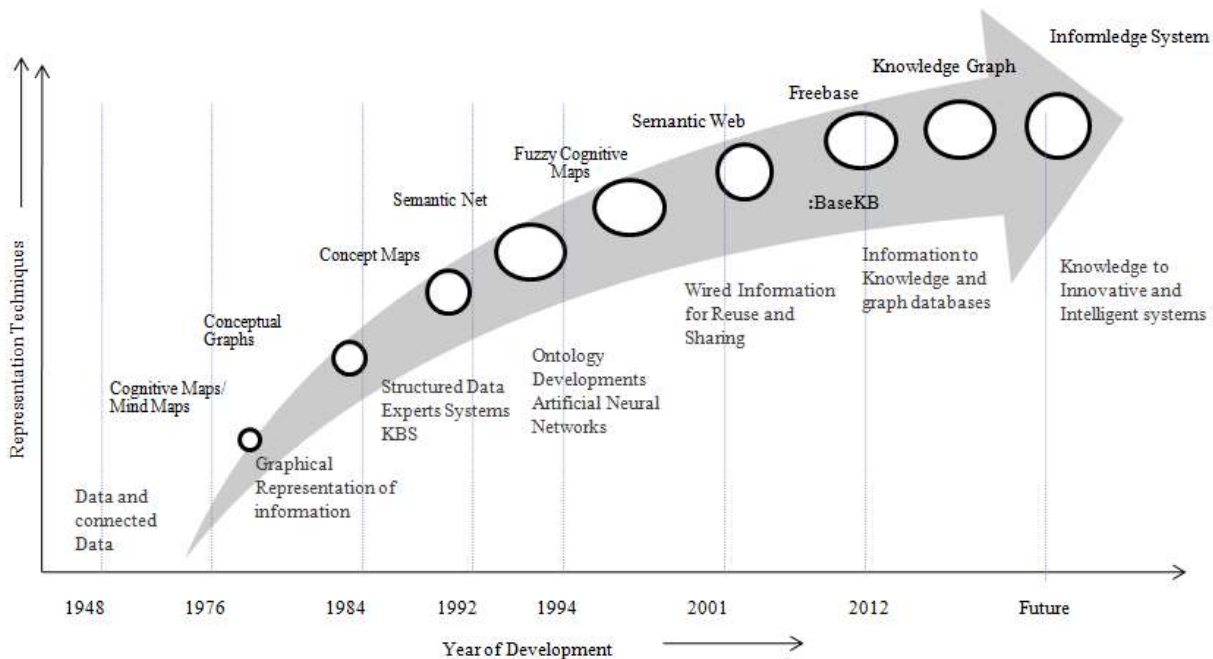


Fig. 7. Evolution of Knowledge Representation Techniques

The multilateral links provide intelligence to the system through its multi-stranded properties [43, 44]. Intelligence of the human brain processing lies with the individual neuron to connect automatically to multiple other neurons that are the basis of development of ILS.

VI. CONCLUSIONS

The advancements in knowledge retrieval systems have enabled the researchers from this field to provide better solutions for autonomous systems with information sharing capabilities. Ontology development and Semantic

Web over the last decade have greatly contributed to information sharing and ease of access. However, these systems have not been able to provide a single complete ontological structure to represent data across multiple domains. The data and information have been ever increasing as more and more data is getting inducted into accessible network systems through huge libraries and legacy organizations. This has resulted in continuous demand to extract knowledge intelligently. The proposed knowledge system, ILS retrieves knowledge intelligently. ILS nodes possess processing and self-propagation properties, and the links provide a coherent connectivity between the nodes of the network and propagate intelligently.

REFERENCES

- [1] C. Bizer, et al., "DBpedia - A crystallization point for the Web of Data," *Web Semantics: Science, Services and Agents on the WWW*, pp 154-165, 2009, doi: 10.1016/j.websem.2009.07.002.
- [2] B. Kosko, *Fuzzy Thinking The New Science of Fuzzy Logic*, Hyperion, 1993.
- [3] R. L. Ackoff, "From data to wisdom," *J. of Applied Systems Analysis*, vol. 15, pp. 3-9, 1989.
- [4] A. Barr, and E. Feigenbaum, *The Handbook of Artificial Intelligence*, vol. 1, William Kaufmann, Inc., 1981.
- [5] T. Berners-Lee, "Artificial Intelligence and the Semantic Web," *AAAI Keynote, W3C Web site*, 2006, retrieved from <http://www.w3.org/2006/Talks/0718-aaai-tbl/Overview.html>.
- [6] K. Bollacker, P. Tufts, T. Pierce, and R. Cook, "A Platform for Scalable, Collaborative, Structured Information Integration," *6th International Workshop on Information Integration on the Web, AAAI*, 2007.
- [7] F. Bouthillier, and K. Shearer, "Understanding knowledge management and information management: the need for an empirical perspective," *Information Research*, 8, 2002.
- [8] C. Brewster, and K. O'Hara, "Knowledge representation with ontologies: Present challenges—Future possibilities," *Int. J. of Human-Computer Studies*, vol. 65, pp. 563–568, 2007, doi:10.1016/j.ijhcs.2007.04.003.
- [9] D. Brickley, and R. Guha, "RDF vocabulary description language 1.0: RDF schema," *w3.org*, 2004, retrieved from <http://www.w3.org/tr/rdf-schema/>.
- [10] D. Buscaldi, and M. C. S. Figueroa, "Effects of Ontology Pitfalls on Ontology-based Information Retrieval Systems," *Knowledge Encoding and Ontology Development*, INSTICC, 2013.
- [11] Cloete, and M. J. Zurada, *Knowledge-Based Neurocomputing*, The MIT Press, ch1, 2000.
- [12] J.W. Coffey, R.R. Hoffman, A.J. Cañas, and K.M. Ford, "A Concept Map-based Knowledge Modeling Approach to Expert Knowledge Sharing," *The Int. Conf. on Information and Knowledge Sharing*, 2002.
- [13] R. Davis, H. Shrobe, and P. Szolovits, "What is a Knowledge Representation?," *AI Magazine*, 14, pp. 17-33, 1993.
- [14] J. G. Delgado-Frias, S. Vassiliadis, and J. Goshtasbi, "Semantic Network Architectures – An Evaluation," *Int. J. Artif. Intell. Tools*, vol. 01, issue. 01, 1992, pp. 57-83, Doi: 10.1142/S0218213092000132.
- [15] M. Frické, "The knowledge pyramid: a critique of the DIKW hierarchy," *J. of Inf. Sc.*, vol. 35, 2009, pp. 131-142.
- [16] "Getting started with Apache Jena," 2014, retrieved from http://jena.apache.org/getting_started/index.html
- [17] José Manuel Gómez-Pérez and Carlos Ruiz, "Ontological Engineering and the Semantic Web", *Advanced Techniques in Web Intelligence – I, in Studies in Computational Intelligence*, vol. 311, Juan D. Velásquez and Lakhmi C. Jain, Eds. Springer Berlin Heidelberg, 2010, pp. 191-224, DOI10.1007/978-3-642-14461-5_8.
- [18] Google Developers, "Get Started with the Freebase API," 2013, retrieved from <https://developers.google.com/freebase/v1/getting-started>.
- [19] D. Graupe, "Principles of Artificial Neural Networks," *Adv. Series on Circuits and Systems*, vol. 6, 2nd edn., World Scientific Publishing Co. Pte. Ltd, Singapore, 2007.
- [20] T. R. Gruber, "A translation approach to portable ontology specifications," In: *Knowledge Acquisition*, Academic Press, 1993, pp. 199-220.
- [21] N. Guarino, "Formal Ontology in Information Systems," Proc. In *Formal Ontology in Information Systems*, Nicola Guarino Ed., IOS Press, 1998, pp. 3-15.
- [22] S. Hawke, I. Herman, P. Archer, and E. Prud'hommeaux, "W3C Semantic Web Activity", 2013, retrieved from <http://www.w3.org/2001/sw/>.
- [23] F. Hayes-Roth, and N. Jacobstein, "The state of knowledge-based systems," *Communications of the ACM*, 37(3), 1994, pp. 26-39, DOI: 10.1145/175247.175249.
- [24] J. Hendler, and F. V. Harmelen, "The Semantic Web: webizing knowledge representation," in *Handbook of Knowledge Representation*, F. Van Harmelen, V. Lifschitz and B. Porter Eds., Elsevier, 2008, pp. 821-839. DOI: 10.1016/S1574-6526(07)03021-0.
- [25] J. Hendler, T. Berners-Lee, and E. Miller, 2002. "Integrating Applications on the Semantic Web," *Journal of the Institute of Electrical Engineers of Japan*, vol. 122, issue.10, 2002, pp. 676-680.
- [26] Herman, "W3C Semantic Web Frequently Asked Questions," 2009, retrieved from <http://www.w3.org/RDF/FAQ> 2009-11-12.
- [27] Herman, 2012. "Semantic Web Adoption and Applications," 2012, retrieved from <http://www.w3.org/People/Ivan/CorePresentations/Applications/>.
- [28] J. Hoart, F.M. Suchanek, K. Berberich, and G. Weikum, "YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia," *Artificial Intelligence*, vol. 194, 2013, pp. 28–61.
- [29] L. B. Holder, Z. Markov, and I. Russell, 2006. "Advances in Knowledge Acquisition and Representation," *J. Artif. Intell. Tools*, vol. 15, issue 6, pp. 867-874, doi: 10.1142/S0218213006003016.
- [30] R. M. Kitchin, "Cognitive maps: What are they and why study them?," In Proceedings of *Journal of Environmental Psychology*, vol.14, 1994, pp. 1-19, DOI: 10.1016/S0272-4944(05)80194-X.
- [31] G. Klyne, and J.J. Carroll, "Resource Description Framework (RDF): Concepts and Abstract Syntax," *W3C Recommendation*, 2004, retrieved from <http://www.w3.org/TR/rdf-concepts/>.
- [32] P. López-García, E. Mena, and J. Bermúdez, "Some Common Pit Falls in the Design of Ontology Driven Information Systems," *Knowledge Encoding and Ontology Development*, INSTICC Press, 2009, pp 468-471.
- [33] O.Lassila, and R. R. Swick, "Resource Description Framework (RDF) Model and Syntax Specification," *W3C Recommendation*, 1999, retrieved from <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>.

- [34] J. Lehmann, et al. 2012. "DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia," *Semantic Web*, vol. 1, IOS Press, 2012, pp 1–5.
- [35] E.Y.Li, "Artificial neural networks and their business applications." *Inf. & Management*, vol. 27, 1994, pp. 303-313.
- [36] D. Lindley, "Brain and Bytes," *Communications of the ACM, Science*, vol. 53, issue 9, ACM New York, Sept. 2010, pp. 13-15, doi: 10.1145/1810891.1810897.
- [37] V. Mascardi, V. Cordi and P. Rosso, "A comparison of upper ontologies," *Technical Report*, University of Genova, Italy, 2006.
- [38] C. T. Meadow, B.R. Boyce, and D.H. Kraft, *Text information retrieval systems*, 2nd ed., 2000, Academic Press, San Diego, CA, 2000.
- [39] D. Montag, 2013. "Understanding Neo4j Scalability," White Paper, Neotechnology 2013, retrieved from [http://info.neotechnology.com/rs/neotechnology/images/Understanding_Neo4j_Scalability_\(2\).pdf](http://info.neotechnology.com/rs/neotechnology/images/Understanding_Neo4j_Scalability_(2).pdf).
- [40] M. Morsey, J. Lehmann, S. Auer, C. Stadler, and S. Hellmann, "DBpedia and the live extraction of structured data from Wikipedia," *Program: Electronic library and Information Systems*, vol. 46, issue 2, 2012, pp. 157 – 181.
- [41] T. R. Gopalakrishnan Nair, and M. Malhotra, "Informledge System- A Modified Knowledge Network with Autonomous Nodes using Multi-lateral Links," *Knowledge Encoding and Ontology Development (KEOD)*, 2010, pp. 473 -477, DOI: 10.5220/0003069103510354.
- [42] T. R. Gopalakrishnan Nair, and M. Malhotra, "Knowledge Embedding and Retrieval Strategies in an Informledge System," *International Conference on Information and Knowledge Management (ICKM)*, July 2011, pp. 351-354.
- [43] T. R. Gopalakrishnan Nair, and M. Malhotra, 2011b. "Creating Intelligent Linking for Information Threading in Knowledge Networks," *IEEE-INDICON*, December 2011, pp. 17–18, DOI: <http://dx.doi.org/10.1109/INDCON.2011.6139335>.
- [44] T. R. Gopalakrishnan Nair, and M. Malhotra, 2012. "Correlating and Cross-linking Knowledge Threads in Informledge System for Creating New Knowledge," *Knowledge Encoding and Ontology Development (KEOD)*, 2012, pp. 251-256, DOI: 10.5220/0004143302510256.
- [45] Niles, and A. Pease, 2001. "Towards a Standard Upper Ontology," *Proceedings in Formal Ontology in Information Systems*, ACM, 2001, DOI: 1-58113-377-4/01/0010.
- [46] K., Niwa, K. Sasaki, and H. Ihara, "An Experimental Comparison of Knowledge Representation Schemes," *AI Magazine*, vol.5, issue 2, 1984, pp. 29-36.
- [47] J. D. Novak, and A. J. Cañas, "The Theory Underlying Concept Maps and How to Construct Them," *Technical Report IHMC CmapTools*, Florida Institute for Human and Machine Cognition, 2008.
- [48] J. D Novak, and D. B. Gowin, *Learning How to Learn*, UK: Cambridge University Press, 1984, pp. 15-40.
- [49] N. F. Noy, and D. L McGuinness, *Ontology Development 101: A Guide to Creating*, 2001, retrieved from http://protege.stanford.edu/publications/ontology_development/ontology101.pdf.
- [50] A. Pease, *Ontology Development Pitfalls*, 2011, retrieved from <http://www.ontologyportal.org/Pitfalls.html>
- [51] D. Ramachandran, P. Reagan, and K. Goolsbey, "First-Ordered ResearchCyc: Expressivity and Efficiency in a Common-Sense Ontology," *AAAI Workshop on Contexts and Ontologies: Theory, Practice and Applications*, 2005.
- [52] E. Rich, K. Knight, and S.B. Nair, *Artificial Intelligence*, 3rd ed., 2010, Tata McGraw Hill Education Private limited New Delhi, India.
- [53] Robinson, What is a Graph Database? What is Neo4j? , 2014, retrieved by <http://www.neo4j.org/>.
- [54] Robinson, J. Webber, and E. Eifrem, *Graph Database*, Neo Technology Inc., 2013, O'Reilly Media.
- [55] J. Rowley, 2007. "The wisdom hierarchy: Representations of the DIKW hierarchy," *J. of Inf. Sc.*, vol. 33, 2007, pp. 163-180.
- [56] S. Russell, and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed., 2009, Prentice Hall.
- [57] P. Sajja, and A. R. Akerkar, "Knowledge-Based Systems for Development," *Advanced Knowledge-Based Systems: Models, Applications and Research*, vol. 1, 2010, pp. 1–11.
- [58] Sandberg, and N. Bostrom, "Whole Brain Emulation: A Roadmap," *Technical Report # 2008 - 3*, 2008, Future of Humanity Institute, Oxford University.
- [59] M. K. Smith, C. Welty, and D. L. McGuinness, 2004. *OWL Web Ontology Language Guide*, 2004, W3C Recommendation, <http://www.w3.org/TR/2004/REC-owl-guide-20040210/>.
- [60] J. F. Sowa, "Conceptual graphs for a database interface," *IBM J. of Research and Development*, vol. 20, 1976, pp. 336-357.
- [61] J. F. Sowa, *Conceptual Structures: Information Processing in Mind and Machine Reading*, Addison-Wesley, MA, 1984, ISBN 978-0201144727.
- [62] J. F. Sowa, *Encyclopedia of Artificial Intelligence*, Stuart C. Shapiro Ed., 1992, Wiley.
- [63] J. F. Sowa, *Handbook of Knowledge Representation*, Harmelen Ed., Elsevier, 2008, pp. 213-237.
- [64] H.B. Styltsvig, "Ontology-based Information Retrieval," *Ph.D. Thesis, Computer Science Section*, Roskilde University, Denmark, May 2006.
- [65] *The Blue Brain Project*, EPFL, 2012. <http://jahia-prod.epfl.ch/page-58110-en.html> Updated 19.06.2012.
- [66] E.C. Tolman, 1948. "Cognitive maps in rats and men," *Psychological Review*, vol. 55, 1948, pp. 189–208, doi: 10.1037/h0061626.
- [67] E. Turban, and J.E. Aronson, *Decision support systems and intelligent systems*, 6th ed., 2000, Prentice Hall.
- [68] *Vertical Applications*, 2013, retrieved from W3C <http://www.w3.org/standards/semanticweb/applications>.
- [69] M. Watson, *Practical Semantic Web and Linked Data Applications*, Publication: Lulu.com, 2012, ISBN-13: 9781257167012.
- [70] M. Witbrock, "Knowledge is more than Data: A Comparison of Knowledge Bases with Database," *Cycorp, Inc.*, 2001.
- [71] V. Jain and M. Singh, 2013. "Ontology Development and Query Retrieval using Protégé Tool," *I.J. Intelligent Systems and Applications*, Vol.5, number 9, August 2013, pp. 67-75, DOI: 10.5815/ijisa.2013.09.08.

Authors' Profiles



Ms. Meenakshi Malhotra received her B. Tech degree from National Institute of Technology, Kurukshetra, India in 1999 and M.S. degree in Software Systems from BITS, Pilani, India in 2005 She is currently a Ph.D. student at Visvesvaraya Technological University, Belgaum, India. She has 10 years of experience in

professional field spread over industry, research and education. Her research interests include Brain Informatics, Knowledge Representation and Processing. She is a student member of IEEE.



T.R. Gopalakrishnan Nair holds an M.Tech from IISc, Bangalore and Ph.D. in Computer Science.

Currently, he is the Saudi Aramco Endowed Chair, ARAMCO Endowed Chair – Technology, PMU, KSA Advanced Networking Research Group, and VP, Research, Research and Industry Incubation Center (RIIC), Dayananda Sagar Institutions, Bangalore, India. He has three decades of experience in computer science and engineering through research, industry and education. He has published several papers and holds patents in multi domains.

He is a Fellow of IE, senior member of IEEE & ACM and a member of AAAS. He is winner of PARAM Award for technology innovation. (www.trgnair.org).

How to cite this paper: Meenakshi Malhotra, T. R. Gopalakrishnan Nair, "Evolution of Knowledge Representation and Retrieval Techniques", *International Journal of Intelligent Systems and Applications (IJISA)*, vol.7, no.7, pp.18-28, 2015. DOI: 10.5815/ijisa.2015.07.03