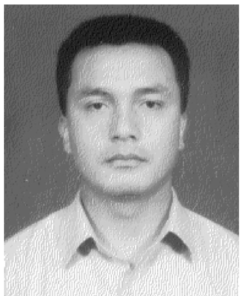
Knowledge Representation: Concept, Techniques and the Aanalytico-Synthetic Paradigm

Jagdish Chandra Binwal and Lalhmachhuana

Department of Library and Information Science,
North Eastern Hill University, Shillong, India

Dr. Jagdish Chandra Binwal is a professor in the Department of Library and Information Science and the Dean of the School of Economics, Management and Information Sciences at the North Eastern Hill University, Shillong, India. He has also taught library and information science at the Sana'a University, Republic of Yemen during 1996-1998. He served as a university librarian of the North Eastern Hill University from 1978-1991. He has been a Senior Fulbright Fellow to the Department of Information Science, University of Pittsburgh in 1988. He is the author of 3 books and 54 articles. His areas of research include structure and development of the universe of knowledge and knowledge representation.



Lalhmachhuana is a research scholar in the Department of Library and Information Science, North Eastern Hill University, Shillong and is currently working in the Central Library of the university. His areas of research interests include structure and development of the universe of knowledge, organisation of knowledge, and information retrieval.

Binwal, J.C., Lalhmachhuana (2001). **Knowledge Representation: Concept, Techniques and the Analytico-Synthetic Paradigm.** *Knowledge Organization, 28(1).* 5-16. 25 refs.

ABSTRACT: Defines the concept of knowledge representation. Explains the major AI-based knowledge representation techniques developed so far. Critically examines the strengths and weaknesses of such AI techniques. Argues that the analytico-synthetic approach advocated by Ranganathan is really a powerful knowledge representation technique containing in itself the epistemological foundations required for real progress in the field.

0. Introduction

Knowledge representation has been the focus of attention for library and information science since its origin. Today the notion of knowledge representation has pervaded many other disciplines. These include computer science, artificial intelligence, linguistics and psychology. In computer science, the issue of knowledge representation arises when data structures and the structure of records and files in databases are to be decided. In linguistics, the problem of knowledge representation arises when dealing with the syntactic and semantic structure of natural language. Artificial intelligence is concerned with the creation of a knowledge base which, when programmed, results in machine-based intelligence. In psychology, the representation system of cognitive theory and models of human memory are very much concerned with the issues of knowledge representation. There are three main approaches to knowledge representation in cognitive theory. First is the computational model represented most systematically by Fodor (1975). Second is a combination of a biological approach and a computational model as represented by Pylyshyn (1988). Third is the constructivist tendency as represented by Piaget (1962).

1. Concept of Knowledge Representation

Let us now attempt to determine what is meant by knowledge representation. Winston (1984) has described it as "*a set of syntactic and semantic conventions that make it possible to describe things.*" Brachman and Levesque (1985) state that its concern is "*to write descriptions of the world in such a way that an intelligent machine can come to new conclusions about its environment by formally manipulating these descriptions*". According to what has been stated above, it appears that the notion of knowledge representation is an easy one. It is merely writing down, in some language or communications medium, descriptions that correspond in some salient way to the world or a state of the world. However, the concept of knowledge representation is not so simple.

Knowledge representation implies some systematic way of codifying domain knowledge. It views knowledge as consisting of "facts" and "heuristics". Facts represent widely shared, publicly available, and generally agreed upon knowledge in the domain. Heuristics represent knowledge of good judgement that characterise expert level decision-making in a field. Knowledge needs to be organised in a manner that allows easy access, quick processing and is easily built-upon. It must produce intelligent results, congruent with the way people think (Dubey, 1987). Minsky (1990) is of the opinion that there are many approaches to thinking. Consequently, it results in many approaches to representing knowledge. According to him, there are two extremes, connectionist and symbolic. But, neither is adequate.

There are two basic dimensions to knowledge representation. These are: knowledge structures and reasoning mechanisms. Knowledge structures help to organise knowledge into predetermined structures. That way, they are passive, whereas reasoning mechanisms are active. They manipulate the structures to produce useful outputs such as inferences and answers (Liebowitz & Beckman, 1998).

It is not possible to represent knowledge without knowing its properties. Parsaye and Chignell (1988) have identified five elementary properties of knowledge that can be used to represent objects and their interactions. These are:

a) Naming (Proper nouns)
b) Describing (Adjectives)
c) Organising (Categorisation and possession)
d) Relative (Transitive verbs and relationship nouns)
e) Constraining (Conditions)

Davis, Shrobe and Szolovits (1993) argue that the notion of knowledge representation can be best understood in terms of five distinct roles. These are:

i) A knowledge representation (KR) is most fundamentally a surrogate, a substitute for the thing itself, used to enable an entity to determine consequences by thinking rather than acting, i.e., by reasoning about the world rather than taking action in it.

   Viewing representation as a surrogate leads naturally to two important questions. The first question is: How does it correspond to its intended referent in the world? It means that there must be some form of correspondence specified between the surrogate and its intended referent in the world. This correspondence constitutes the semantics of the representation. The second question relates to fidelity: How close is the surrogate to the real thing? Which attributes of the original does it capture and make explicit, and which does it omit? Perfect fidelity is impossible, both in practice and in principle.

ii) It is a set of ontological commitments, i.e., an answer to the question: In what terms should I think about the world? It is impossible to represent the world in its full detail. It is, therefore, necessary to restrict the attention to a small number of concepts which are meaningful and sufficient to interpret the world and provide a representation adequate to a certain task or goal at hand. As a consequence, a central part of knowledge representation consists of elaborating a conceptualisation: a set of abstract objects, concepts and other entities as well as the relations that may hold between them. The commitments which are implied by the choice of one set of concepts instead of another to describe a certain phenomenon are called ontological commitments (Valente and Breuker, 1996).

   Further, all representations are imperfect approximations to reality, each approximation attending to some things and ignoring others. It involves a set of decisions about how and what to see in the world. In other words, selecting a representation means making a set of ontological commitments. This helps in focussing attention on aspects of the world we believe to be relevant. The focussing effect is an essential part of what a representation offers, because the complexity of the natural world is overwhelming. We need guidance in deciding what in the world to attend to and what to ignore.

Ontology is, therefore, required to have the following characteristics (Valente and Breuker):

(a) Parsimony;
(b) Clear theoretical basis;
(c) Categories; and
(d) Coherence

iii) It is a fragmentary theory of intelligent reasoning, expressed in terms of three components:

(a) the representation's fundamental conception of intelligent reasoning;
(b) the set of inferences the representation sanctions; and
(c) the set of inferences it recommends.

There are five distinguishable notions of what constitutes intelligent reasoning. They have been derived from the fields of mathematical logic, psychology, biology, statistics and economics. The first view, derived from mathematical logic, considers intelligent reasoning as formal calculation based on deduction. The second, rooted in psychology, sees reasoning as a characteristic human behaviour. The third, derived from biology, is based on the theory of stimulus/response behaviour. It takes reasoning to be the outcome of the architecture of the machinery that accomplishes it. The fourth view, derived from probability theory, considers intelligent reasoning as that which satisfies the axioms of probability theory. The fifth view, derived from economics, defines intelligent reasoning by applying the tenets of utility theory.

iv) It is a medium for pragmatically efficient computation, i.e., the computational environment in which thinking is accomplished.

v) It is a medium of human expressions, i.e., a language in which we describe things about the world. As a language, it should be:

(a) Easy;
(b) Precise;
(c) Expressive; and
(d) Functional

These five roles make it clear that knowledge representation incorporates a theory of intelligent reasoning. It cannot be viewed in purely epistemological terms. In fact, epistemology and reasoning are inextricably intertwined in knowledge representation. We cannot talk about one without discussing the other.

2. Knowledge Representation Techniques

There are a number of systems of knowledge representation available today in the field of Artificial Intelligence. These include Logical Systems, Production Rule Systems, Semantic Networks, Frames, Scripts, and Conceptual Dependency. These systems can be broadly grouped into three categories: Logical Systems, Production Rule Systems, and Structured Objects Systems comprising Semantic Networks, Frames, Scripts and Conceptual Dependency. Many other systems are still evolving.

2.1 Logical Systems

Logic has been used as a method of knowledge representation. There are several forms of logical representation. However, the most common are propositional logic and predicate logic. Propositional logic represents knowledge in the form of statements such as 'Socrates is a man' and 'Socrates is mortal', that are either true or false. Propositions can, of course, be compound statements linked together with connectives such as AND (&), OR (V), NOT (7), IMPLIES (⇒), and EQUIVALENT (≡).

Propositional calculus has the properties of completeness, soundness, and decidability. However, it has its limitations. One cannot deal properly with general statements of the form 'All men are mortal', e.g., one cannot derive from the conjunction of this and 'Socrates is a man' that 'Socrates is mortal'. To do this, one needs to analyse propositions into predicates and arguments and deal explicitly with quantification (Jackson, 1986).

Predicate logic provides a formalism for performing this type of analysis. The term 'predicate logic' derives from the fact that propositions are analysed into predicate argument compositions. It also uses quantifiers. These are of two kinds: existential and universal. The statement 'All feathered creatures are birds' will be expressed as:

∀x [feathered (x) - bird (x)]

On the other hand, the statement 'Some swans are black' will be expressed as:

∃x [swan (x) & black (x)], because it is not a universal statement.

Inferences are derived by using principles of *modus ponens* and *universal specialisations*.

Predicate calculus works well in case of problems having an essentially deductive nature. However, the majority of problems that humans face are of a much more inductive nature. Such problems require the handling of data that are uncertain. Solutions require complex reasoning processes, such as assuming a hypothesis to be confirmed gradually, making many cross-checks to detect error, and so on.

Ontological commitments of logical systems are based on the belief that logic is a calculus with syntactic rules of deduction. In other words, one can say that the meaning, value or outcome associated with an expression in logical systems depends solely upon its external form and not upon any extraneous associations or ideas that might attach to the symbols in the mind of someone reading or writing them. Thus, logical systems involve a fairly minimal commitment to viewing the world in terms of individual entities and relations between them.

The theory of intelligent reasoning in these systems is derived from mathematical logic, which proceeds with the assumption that intelligent reasoning is some variety of formal calculation, typically deduction. In other words, reasoning intelligently means reasoning in the fashion defined by first order logic. However, a great deal of common sense knowledge cannot be obtained using formal logic. Humans often resort to abduction, which is not permissible in formal logic. Abduction is a legitimate and widely practised problem solving technique. It enables one to make progress in solving a problem by assuming informed "guesses" when little is known with certainty. Similarly, another useful technique is induction.

As a language, logical reasoning cannot be said to be simple. However, it possesses the characteristics of soundness and expressiveness.

2.2 Poduction Rule Systems

Production rules represent knowledge in a situation- action couple. It means that whenever a certain situation given as the left side of the rule is encountered, the action given on the right is performed.

A system based on production rules usually has three components:

i) The Rule base;
ii) Facts base, containing the known facts relevant to the domain of interest.
iii) The Interpreter that decides which rules to apply (Bonnet, 1985).

When certain facts or knowledge about a situation are presented to the system, the interpreter starts checking the facts against sets of rules. When the "If" portion of a rule is satisfied by the facts, the action specified by the "Then" portion is performed. In the process, the facts base may be modified by adding new facts to the base. The new facts added to the base can later on be used to form matches with the "If" portion of rules. Matching of rule "If" portions to the facts produces inference chains. These inference chains can be displayed to the user to help explain how the system reached its conclusions.

The inference engine applies the rules in two important ways: one is called "forward chaining" and the other "backward chaining". If the inference engine looks first at the established data or facts to decide if these satisfy the left side of a rule (the premise), it is said to work in a forward direction. If, on the other hand, it looks first at the aims to be attained as given on the right side of the rule (the action part) and then tries to satisfy only those rules which have these aims, it is said to work backwards (Bonnet, 1985). In other words, forward chaining is bottom-up or event-driven reasoning and backward chaining is goal-directed reasoning or top-down reasoning.

One major issue to be sorted out in a production rule system is conflict resolution. It arises when several rules are applicable to a given situation. In such a case, the interpreter has to decide whether or not to apply them in some particular order and whether to apply all that are applicable or only some selection. This problem is resolved by adopting conflict resolution strategies, some of which are:

(a) Performing the first rule;
(b) Sequencing technique: adopting the rules in the sequence they are;
(c) Performing the most specific;
(d) Most recent rule;
(e) Ordering in decreasing order of the strength of the conclusions or of the premises; and
(f) Priority to those that have been most frequently used.

Here one is reminded of the Basic Laws enunciated by Ranganathan. They also pertain to conflict resolution. They are to be invoked when there is conflict between two or more laws. The Basic Laws are:

i) Laws of Interpretation

ii) Law of Impartiality
iii) Law of Symmetry
iv) Law of Parsimony
v) Law of Local Variation
vi) Law of Osmosis

Potentiality of these laws for conflict resolution in production rule systems will be known only when used in some expert systems.

As far as ontological commitments are concerned, rule-based systems view the world in terms of attribute-object-value triples and the rules of plausible inference. Intelligent reasoning process involves a combination of logic and psychological traditions. Rule based systems reflect psychological tradition in the sense that they capture guesses of the type that human experts make, guesses that are not necessarily sound or true. They reflect logical tradition in the sense that propositional and predicate logic are used to derive inferences. These deductive methods pose problems in real-life situations where the data as well as inferences, which reflect expert opinion, are usually uncertain. Thus, several methods that employ measures of the uncertainties associated with plausible inferences (Bonnet, 1985) have been developed to take these uncertainties into account. One of the methods is known as the Bayesian approach. The basis of this method is to assign to each statement a measure of likelihood or confidence lying between 1(True) and -1(False), where O indicates complete uncertainty. Another important method uses the concept of the "fuzzy set". It takes into account the degree of the imprecision of the data rather than the uncertainty of the data.

As a language, the production rule system is flexible. It can be easily modified due to its considerable degree of modularity. Thus, it can evolve. It is simple and expressive.

However, production rule systems suffer from a number of significant disadvantages. First, there is no mechanism for showing context dependency. Second, conflict resolution strategies often cause some strange side effects in the firing order of the rules. Third, production rule systems fail in reflecting the structure of the domain in terms of taxonomic, part-whole, or cause-effect relationships between objects and between classes of objects. Fourth, no discipline is imposed on the ordering of the rules. As a result of these disadvantages, the system tends to become complex, slow, and inefficient as it grows in size.

2.3 Structured Objects Systems

These systems have been developed around the concept of structured objects. A structured object is regarded as a prototype, meaning an ideal model of the object or situation with which objects being studied are compared. An object is considered to have a number of aspects or attributes. Each attribute can have either default values or possible values. Most representations by structured objects allow the specification of default values to be used when no other relevant information is available and a default value can be replaced by an actual value if this becomes known at some later stage.

The fundamental idea is that the properties of an object that appear relevant, interesting, and so on, are a function of how one perceives the object and for what purpose. As a result, descriptions are viewed as a process of comparison in which one specifies a new entity by saying in what way it is similar to but different from existing objects.

Further, representation by objects allows the two forms of expressing knowledge, i.e., declarative and procedural, to be combined by giving the necessary procedural information with certain attributes of the objects.

2.3.1 Semantic Networks

The idea of using semantic networks to represent knowledge is attributed to Quillian(1969). A semantic network represents knowledge as a pattern of interconnected nodes and arcs. The nodes represent either concepts, attributes, states, or events and the arcs show the relationships between the nodes. Many link structures are being used today in semantic nets. Some important ones are:

a) "is-a" to represent hierarchy
b) "has" to represent attributes
c) "is in" to represent location
d) "part- of" to represent parts, organs, etc.

Of all these links, the "is-a" link is the most prominent. It generates a hierarchical structure within the network. An individual node forms a subset of a generic node, which forms a subset of another generic node, and so on. Another property of this link is inheritance. This means that items lower in the net can inherit properties from items higher up in the net. This saves space since information about similar nodes does not have to be repeated at each node. This property of inheritance is also applicable to other links.

Semantic nets have also been used successfully in Natural Language Processing. Here, a sentence is first semantically analysed into predicate and arguments. These form the nodes. Arcs define the relationships between the predicate and the arguments (concepts) associated with that predicate. Relationships are represented by specifying the role, such as Agent, Patient, Instrument, Beneficiary, Location, Time etc. played by an argument in the sentence.

Ontological commitment is evident in semantic networks to view the world in terms of individual entities and relations between them. 'Hierarchy' and 'Inheritance' are its other commitments. Intelligent reasoning has its base in the theory of connectionism. It is quite simple. All that has to be done is to specify the start node. From the initial node, other nodes are pursued using the links until the final node is reached. However, it lacks the standardisation and formalisation of reasoning process. As a language, it is quite flexible. New nodes and links can be defined and added as needed. It also shows a general commitment to simplicity and parsimony.

2.3.2 Fames

Frames use a relational table approach to knowledge representation. Structured objects form the core of frames. A frame describes an object by containing all of the information about that object in "slots". In one sense, a frame can be considered a record with fields or slots that can be filled with specific values. Marvin Minsky (1975) who originated the frame idea, describes it as follows:

> "*A frame is a data structure for representing a stereotyped situation, like being in a certain kind of living room or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed*".

Frames are used to represent declarative as well as procedural knowledge. Frames containing information about objects are called declarative/factual/situational frames. Frames containing slots, which explain how to do a thing, are called action-procedures frames. Like semantic networks, frames, if attached or associated one to the other, form a hierarchy and display the property of inheritance. This is why Minsky has stated that "*we can think of frames as a network of nodes and relations*" (1975).

Ontological commitments of frames have their roots in thought on prototypical objects, defaults, and hierarchy. The theory of intelligent reasoning in frames has been derived from psychological traditions. Minsky (1981) is very clear about this, when he states that "*this is a partial theory of thinking, ... whenever one encounters a new situation (or makes a substantial change in one's viewpoint), he selects from memory a structure called a frame; a remembered framework to be adopted to fit reality by changing details as necessary*".

Reasoning in frames is done by the instantiation process. This process begins when the given situation is matched with frames already in existence. When a match is found values are assigned to the slots needing it, resulting in the depiction of a particular situation. Generally, if a given slot characteristic is not present, the slot provides a default value for that characteristic. Reasoning process also allows one to move from one frame to another to match the current situation.

As a language, frames are easier to understand. They are simple and able to express the situation adequately.

2.3.3 Scripts

Scripts, which are similar to frames, are another means of knowledge representation. Whereas frames are used to represent all kinds of knowledge, scripts are used to represent stereotyped events that take place in day-to-day activities. Some such events are:

i) Going to a hotel
ii) Going to the library
iii) Going to the bank

The idea of scripts was introduced by Roger Schank and Robert Abelson (1977). Their main point is that knowledge of the scripts for many situations is necessary for understanding the way in which the different actions comprising an event are interlinked. Scripts describe the causal relationships between the different episodes and make it possible to draw inferences, to guess things that are implied but not stated, and, to some extent, to fill in the blank spaces (Bonnet).

Ontological commitments of scripts have their bases in thinking in terms of stereotyped events. Scripts can only represent stereotyped knowledge. Intelligent reasoning embedded in scripts has its roots in psychological traditions. In this sense, scripts are similar to frames. The values of the slots are instantiated. Scripts, in some respects, have advantages over Frames. They allow one to identify preceding and succeeding scenes. An event can be described to the minutest detail. However, knowledge cannot be shared across scripts. What is happening in a script is true only for that script. It is difficult to arrange for processes or items of information that are common to a number of scripts to be shared if these scripts do not form a hierarchy in which, for example, each may represent a specialisation of the one above (Bonnet).

As a language it is simple and expressive, but limited to stereotyped events or situations.

2.3.4 Conceptual Dependency

Conceptual Dependency, though mainly a theory of natural language processing, is used as a knowledge representation technique for the following reasons:

a) It enables the development of computer programmes that can understand natural language
b) It provides a means of representation which is language independent.
c) It enables the system to participate in question-answer dialogues.
d) It allows the derivation of inferences from statements.

The concept of conceptual dependency was developed by Schank (1972). The main ideas of this system can be summarized as follows (Bonnet):

i) Two phrases or sentences having equivalent meanings must have the same internal representation.
ii) Every action should be expressed in terms of certain primitives. For example, a primitive for 'drink' could be 'ingest', which could also be used for 'swallow' and 'eat'.
iii) All the information implied in a phrase or sentence must be made explicit in the internal representation.

The meaning of a phrase is represented by a schema. This consists of four kinds of nodes or categories:

a) PP (Picture Production) equivalent to nouns.
b) ACT (Action) equivalent to verbs
c) PA (Picture Aider) equivalent to adjectives (Modifiers of PP)
d) AA (Action Aider) equivalent to adverbs (Modifiers of ACT)

Dependencies or relations are represented by:

<==> : This means mutual dependence between two concepts
----> : This means one way dependence between an ACT and a PP or between a PP and a PA
<== : This means one way dependence between two PPs.

Thus, the conceptual dependency schema for "Arthur revised his latest book" will be:

```
Arthur  <==>  revised  <----  book
                              ↑  ⇑    (possessed
                                          by)
                          Latest  Arthur
```

In other words, we can say that Schank has developed the use of a set of four categories of concepts in a sentence. These are Objects, Actions, Modifiers of Actions, and Modifiers of Objects. Some of the major primitive actions are:

Primitive Action	**Explanation**
a) ALTRANS	Transfer of abstract relationship (e.g., give)
b) PTRANS	Transfer of the physical location of an object (e.g., go)
c) PROPEL	Application of physical force to an object (e.g., push)
d) MOVE	Movement of a body part by its owner (e.g., kick)
e) GRASP	Grasping of an object by an actor (e.g., throw)
f) INGEST	Ingesting of an object by an animal (e.g., eat)
g) EXPEL	Expulsion of something from the body of an animal (e.g., cry)
h) MTRANS	Transfer of mental information (e.g., tell)
i) MBUILD	Building new information out of old (e.g., decide)
j) SPEAK	Producing of sounds (e.g., say)
k) ATTEND	Focussing of a sense organ toward a stimulus (e.g., listen)

Similar work has also been done by Wilks (1975) with his theory of preferential semantics. He has developed a set of up to 100 primitive semantic elements. He has grouped them into five classes: Entities, Actions, Type Indicators, Sorts, and Cases.

As far as ontological commitments are concerned, conceptual dependency believes in semantic primitives, relations/dependencies among concepts and language independent representation. Intelligent reasoning is derived from psychological traditions. As a language, it is simple and expressive.

2.3.5 Structured Objects Systems: Evaluation

As a whole, structured objects systems lend themselves well to model-fitting approaches. However, their representation of hierarchy and inheritance is not fool-proof, especially when an object tends to inherit properties from many classes. In addition, the mechanism of exceptions and defaults is a strong component of object systems. These systems do not usually support backtracking. The way inherited properties are cancelled and altered in actual practice, makes it difficult to properly define one thing in terms of another.

3. Analytico-Synthetic Paradigm

Having analysed the main features of knowledge representation languages in the field of artificial intelligence, it will not be out of place to refer to the remarks of Peter Jackson (1986) who is of the opinion that "*knowledge representation languages are really just high-level programming languages, and that their epistemological foundations are in fact, quite shallow. A more charitable interpretation is that such languages do provide data and control structures which are more flexible than those associated with conventional languages, and they are therefore more suited to the simulation of human reasoning than anything we had before. These two interpretations are not mutually exclusive, in that one can hold them both to be true at the same time.*"

Thus, it cannot be assumed that contemporary knowledge representation issues have somehow been resolved by the development of highly sophisticated tools which, despite the best of efforts, still exhibit inherent shortcomings when considered from the epistemological point of view. Further work on the epistemological foundations is, therefore, essential for real progress. A knowledge representation language needs epistemological foundations in the following areas:

a) **Conceptualisation of Concepts, Objects and Relations;**
b) **Classification**;
c) **Inheritance**;
d) **Intelligent Reasoning and Heuristics**; and
e) **Conflict Resolution**.

In this context, Ranganathan's ideas enshrined in his classic work *Prolegomena to Library Classification*, as well as in the works of other researchers who have extended his ideas, are quite promising. Epistemological foundations for three crucial dimensions of knowledge representation, viz., 1) Conceptualisation of concepts, objects and relations, 2) Classification, and 3) Inheritance are available in the form of Basic Laws, Canons, Postulates and Principles.

3.1. Conceptualisation of Concepts, Objects and Relations:

The conceptualisation of concepts, objects and their relations involves the application of abstract categories and their formalisation. Broadly speaking, conceptualisation occurs whenever the human mind recognises an entity and, presumably, creates an abstraction of that entity which is stored in the memory as a concept for later use. In the realm of KR, conceptualisation involves the identification and analysis of the several attributes/facets of the objects/entities/concepts constituting the domain in relation to users' interest and arranging these attributes/facets within the domain in some meaningful sequence.

Categories are fundamental to all cognitive activities. Categories can be understood in the sense of general concepts. Eduard Sukiasyan (1998) has aptly defined a category as " *the most generic notion encompassing the most universal and essential attributes, properties, connections and relations of objects and phenomena in the real world.*" B.C.Vickery (1953) understands "**conceptual categories**" to mean *"concepts of a high degree of generality with a wide area of application elaborated by the mind in referring directly or indirectly to empirical knowledge and utilized by the mind in interpreting such knowledge."* Thus, categories can be characterized as *"the building blocks of cognition because they permit the individual to generalize to new experiences the information associated in memory with a particular category label"*(Jacob,1994).

Categorization, simply put, is the cognitive process of dividing the world of human experience into generalized groups or broad categories comprising certain components sharing immediate similarity in terms of attributes within a given context. That this context may vary, and with it the composition of the category, is the very basis for both the flexibility and the power of cognitive categorization. Elin K. Jacob has pointed out that categorization is "*the fundamental cognitive process of constructing order out of the potential-*

ly chaotic environment in which the individual lives by dividing the world of experience into named groups of entities whose members bear some relation of similarity to each other." Without recourse to categorization, the experience of any one entity would be totally unique, requiring labeling and storage in the memory as a singular experience identified uniquely by its own set of defining characteristics. Behavior based on learning (i.e., the generalization of acquired information) would be impossible. Categorization is thus "*the fundamental cognitive mechanism that simplifies the individual's interaction with the environment by facilitating the efficient storage and retrieval of information and thereby reducing the demands for cognitive storage that would otherwise be placed on human memory.*"(Jacob, 1994)

Based on the categorical relationships, concepts constituting the objects of a domain can be hierarchically structured to express relations between them for the purpose of knowledge representation and organization. Categorization may thus be understood to include the following processes:

a) the naming of each possible characteristic of an entity/object;
b) each named characteristic (as an element in relation with other elements) forms a component of a structure;
c) Appropriate characteristics are identified and subdivided (under a particular category) into varieties which may be ordered (arranged in some meaningful/helpful sequence); and
d) provision of procedural rules as to the use and maintenance of the structure to achieve higher levels of sophistication.(McIntosh and Griffel, 1969)

It is in the area of conceptualisation of concepts/objects and their relations that Ranganathan has made the most fundamental contributions through his postulates and principles concerning categories and categorisation. To him, KR involves identifying the subjects/objects, recognising the nature of relationships among the components, and then representing them as data structures to be applied in some combinations under specific contexts.

The basic approach of Ranganathan is a shift away from typical concepts to categories. To him, the idea of categories is profoundly related to a very practical purpose – that of ordering in a uniform sequence the attributes/facets under the various objects/domains. The term "category" is used by Ranganathan to mean the broadest conceivable classes of phenomena applicable for use in the whole, or a large part, of knowledge. Ranganathan prefers the term "**fundamental categories**" to describe the category labels representing the relations of commonality between the members of a group of entities, including the particular attributes that identify group members as similar entities. The postulate of fundamental categories which states that "*There are five and only five fundamental categories – viz., Time, Space, Energy, Matter, and Personality*" is the basic postulate and all other subsequent postulates are secondary to it. Ranganathan further states that "*each facet of any subject, as well as each division of a facet, is considered as a manifestation of one of the five fundamental categories*". This framework is based on the assumption that any specific field of knowledge is formed by the interaction of the five fundamental categories in relation to the basic facet. In Ranganathan's scheme of categories, the "Basic Facet" is the context-specifying element that provides the environment within which the object/subject is studied. Owing to the difficulty of specifically defining the environment, Ranganathan refers the problem to a higher level of postulates and suggests that the basic facet (i.e., the context) needs to be postulated. Though the above categories have no claim to any scientific justification, and are based entirely on intuition and belong to the first level of abstraction, they are integral to the Analytico-Synthetic approach to knowledge classification, representation and organization. The concept of categories expounded by Ranganathan, in particular, belongs essentially at the highest level of knowledge organization, and thus sets the general direction towards the conception of a universal knowledge representation model.

The most powerful and influential among Ranganathan's ideas are those relating to facet analysis and synthesis applicable to real world knowledge structures and their representation. Ranganathan has demonstrated that facet analysis (i.e., breaking down subjects into their component parts) and synthesis (i.e., recombining these parts by appropriate connecting symbols to fit the subjects of documents) provide the most viable approach to representing the subject contents of documents (or for that matter, objects of domains). This powerful and highly flexible method of knowledge organization and representation is known as the Analytico-Synthetic (AS) approach. Although the ideas behind this approach have been known to exist much earlier, their formal enunciation must be attributed to Ranganathan who systematized and established guidelines for them.

In addition to the postulates concerning categories, Ranganathan also provides specific principles for achieving an orderly sequence of facets within the categories. These principles for facet sequence, based on the analogy of real-life objects and phenomena, seek to determine the helpful sequence of the various facets occurring in specific domains. In other words, they seek to achieve an infallible order of objects and their attributes – an area which till today remains a vexed issue in KR research.

While the AS approach is the outcome of the efforts of many researchers from library and information science, documentation science, computer science and related fields, the most pathbreaking ideas belong to Ranganathan, the chief architect and exponent of the AS approach. Other scholars who have made significant contributions to the ideas concerning facets and categories include Serge Tchakhotine, Z. Dobrowolski, and G. Cordonnier in France; J.E. Holmstrom, E.G. Brisch, J.E.L. Farradane, D.J. Foskett, B.C. Vickery, Derek Austin, B. Kyle, and others in Britain; and Calvin N. Mooers, J.W. Perry, H.P. Luhn, H.E. Bliss, and Mortimer Taube in the US; and many others. Prominent among present-day scholars who have further advanced the ideas of facets and categories are I. Dahlberg, Jean Aitchison, Nancy J. Williamson, Hemlata Iyer, I.C. McIlwaine, Clare Beghtol, to name only a few.

Considerable development of the AS approach has come through two schools of thought: Ranganathan's group at Bangalore in India and the Classification Research Group in London. Both schools have been active for several decades now though research has resulted in interesting, yet inevitable, deviations necessitated by the revolutionary changes in technology and the fluidity of knowledge structures. Different approaches can have different implications for categorization in particular. In the absence of any *a priori* standard method of categorization for achieving a viable knowledge representation system, surely, the AS framework expounded by Ranganathan deserves more serious consideration. Experience has shown, time and again, that Ranganathan's scheme of categories provides an adequate and acceptable framework for representing subjects/objects of any degree of complexity, irrespective of domain.

3.2 Classification

Subsequent to the initial conceptualisation of concepts, objects, their relations and attributes, the next process inevitably involves the grouping of the concepts/objects/entities/phenomena as a step towards their systematic representation. Specifically, it requires the structuring of each category into a hierarchy, subdivided stage by stage (at each level) by the application of a series of characteristics/differentiating attributes/sub-facets. It is necessary to group objects into classes in order to enable objects in a class to share attributes. Each class has its own sets of attributes, which may be numerous. Out of these, only those relevant attributes are selected. However, the process is not as simple as it appears to be. It needs guidance from sound theory. Ranganathan's work in this context is fundamental. The following specific Canons and Principles enunciated by him provide the framework for the classification of knowledge domains/subjects, concepts/objects, and entities as a step towards their representation:

a) The **Canons for Characteristics** which seek to regulate the choice and application of characteristics/attributes in terms of their inherent qualities which are vital for securing differentiation, relevance, ascertainability, and permanence;
b) The **Canons for Succession of Characteristics** which are prescriptive and aid in determining the order of succession of characteristics, i.e., attributes chosen for the purpose of classification. The order so chosen should not result in concomitance, but should be relevant to and consistent with the intended purpose;
c) The **Canons for Array** which prescribe certain conditions be fulfilled in the course of the formation, assortment, and arrangement of arrays in a sequence which is both helpful and consistent; and
d) The **Principles for Helpful Sequence** which include eight guiding principles intended to fulfil the demands of the Canon of Helpful Sequence which each array of classes should satisfy. These principles are based on considerations relating to time sequence, evolutionary sequence, spatial contiguity, quantitative measure, complexity, traditional sequence, literary warrant and, when all else fails, on alphabetical sequence.

These Canons and Principles provide the basis for the assortment and ordering of the attributes of objects in some meaningful/helpful sequence.

3.3 Inheritance

Following the conceptualisation of concepts, objects, their relations, and their grouping based on specific attributes; the next logical step involves establishing hierarchical relationships among them. Inheritance is achieved by arranging the groups, subgroups, and isolates in a helpful and filiatory sequence. It involves the formation of hierarchies of successive levels. The relationship that emerges is that of superordination and subordination expressed as genus-species, whole-part, etc.

In this regard, the Canons for Chain provided by Ranganathan prescribe that each chain of classes or of ranked isolates in a scheme for classification should satisfy the following two canons: the Canon of Decreasing Extension and the Canon of Modulation. These Canons further seek to achieve the formation of successive levels of hierarchies in an order of decreasing extension (or increasing intension) without any missing links between the first and the last links. The Canon of Modulation, in particular, implies the concept of "Resolving Power" – the power of recognizing the classes or the ranked isolates appropriate to the array of the first order of an immediate universe under consideration. Within the hierarchy established, Ranganathan further recommends due consideration of the Canons for Filiatory Sequence which demand that the classes of the universe of subjects (read facets/objects of categories/domains) should be arranged according to their degree of mutual relationship or affinity recognised as of subordinate or coordinate status.

Overall, Ranganathan's AS approach to subject/object structuring is really a scheme of knowledge representation. It enables us to create surrogates for concrete as well as abstract entities. It is based on a set of ontological commitments provided by postulates and principles. Theoretical foundations of this ontological commitment are enshrined in *Prolegomena.* The scheme of categorisation of concepts is parsimonious, coherent, and corresponds to the syntax in which the majority of people arrange ideas in their minds. Perhaps, it is more appropriate to say that the postulates and principles guiding it conform to the absolute syntax. As a language, it is precise, expressive and hospitable. As far as intelligent reasoning is concerned, it believes in psychological tradition, which takes intelligent reasoning as a particular variety of human behaviour. It is based on the assumption that human problem solving behaviour could usefully be viewed in terms of goals, plans, and other mental structures. Psychological tradition believes that intelligence is an inherently complex natural phenomenon. It consists of a variety of mechanisms and phenomena, for which complete and concise descriptions may not always be possible.

4. Conclusion

Finally, it can be said that the three dimensions relating to the conceptualisation of concepts, objects and their relations, their grouping, and their structuring form the core of KR. In this context, it is heartening to note that Ranganathan's ideas address the very foundation of knowledge representation. The authors hope that if present day KR researchers devote more time to ponder over the profundity of the ideas propounded by Ranganathan, then, what was decades ago considered a "**paradigm shift**" in library classification theory can surely emerge as a potential "**paradigm set**" in knowledge representation. Such is the power of Ranganathan's ideas that they have transcended all barriers of space, time, technology, and cultures.

References

Bonnet, Alain (1985). Artificial intelligence: promise and performance. London: Prentice Hall International, p.108

Brachman, R. J. and Levesque, H. J.(1985). Readings in knowledge representation. Los Altos, CA: Morgan Kaufman.

Davis, Randall; Shrobe, Howard and Szolovits, Peter (1993). What is knowledge representation? *AI Magazine*, v.14(1), pp.17-33

Dubey, Yogendra P.(1987). Knowledge representation and reasoning. In Allen Kent; James Williams (Eds) Encyclopedia of Microcomputers. New York: Marcel Dekker, v.10, pp.1-18.

Fodor, J. A.(1975). The language of thought, New York: Crowell.

Jackson, Peter (1986) Introduction to expert systems. New York: Addison-Wesley, p.74.

Jacob, Elin K. (1994). Classification and Crossdisciplinary Communication: Breaching the Boundaries Imposed by Classificatory Structure. In Knowledge Organisation and Quality Management: Proceedings of the Third International ISKO Conference, 20-24 June 1994, Copenhagen, Denmark. Hanne Albrechtsen and Susanne Oernager (Eds.), Frankfurt: Indeks Verlag, p.101-108.

Liebowitz, J. and Beckman, Tom (1998). Knowledge organisations. Boston: St. Lucie Press, p.90.

McIntosh, Stuart D. and Griffel, David M. (1969).Computers and Categorisation. In DRTC Annual Seminar 7, Bangalore, p.201-220.

Minsky, Marvin (1975). A framework for representing knowledge. In The Psychology of Computer

Vision, edited by D.H. Winston. Chapter 6. New York : McGrawHill.

Minsky, Marvin (1981). A framework for representing knowledge. In Mind Design, edited by J. Hangeland. Cambridge, Mass: MIT Press, pp.95-128.

Minsky, Marvin (1990). Logical vs analogical or symbolic vs connectionist or neat vs scruffy. In Artificial intelligence at MIT: expanding frontiers, edited by P. Winston. Cambridge, Mass: MIT Press.

Parsaye, K. and Chignell, M.(1988). Expert systems for experts. New York: John Wiley.

Piaget, J. (1962). Play, dreams and imitation in childhood. London: Routledge and Kegan Paul.

Pylyshyn, Z. W.(1988). Computation and cognition:Issues in the foundations of cognitive science. *Behavioural and Brain Sciences*, v.3, pp. 111-69.

Quillian, M R. (1969). Semantic memory. In Semantic Information Processing, edited by M. Minsky. Cambridge, Mass : MIT Press, pp. 227-270

Ranganathan, S. R.(1967). Prolegomena to library classification, 3rd ed. Bombay: Asia Publishing House.

Schank, R C.(1972). Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*. v.3.

Schank, Roger and Abelson, Robert (1977). Scripts, plans, goals and understanding. Hillsdale, NJ: Lawrence Erebaum Assn.

Star, Susan Leigh (1998). Grounded classification: grounded theory and faceted classification. *Library Trends*, v.47(2), p.222.

Sukiasyan, Eduard (1998). Classification Systems in their Historical Development: Problems of Typology and Terminology. In Structures and Relations in Knowledge Organisation: Proceedings of the Fifth International ISKO Conference, 25-29 August 1998, Lille, France, edited by Widad Mustafa el Hadi, Jacques Maniez and Steven A. Pollitt. Wurzburg: Ergon-Verlag, p.72-79.

Valente, Andre and Breuker, Joost(1996). Towards principled Core Ontologies. WWW.KSI.CPSC. VCALGARY.CA/KAW/KAW/96/VALENTE/ DOC.HTML.

Vickery, B.C. (1953). Systematic Subject Indexing. *Journal of Documentation*, 9(1), p.54.

Wilks, Y. (1975). A preferential pattern seeking semantics for natural language inference. *Artificial Intelligence*,v.6(1), pp. 53-74

Winston, P. H. (1984). Artificial intelligence. 2nd ed. Reading, Mass : Addison-Wesley.