Lib. Sc. 24; 1987; Paper T

DESIGN OF CONCEPT CODES IN INFORMATION RETRIEVAL: An analysis of trends

M A GOPINATH, *Documentation Research and Training Centre, Indian Statistical Institute, 8th Mile, Mysore Road, R.V. College P.O., Bangalore-560 059.*

Presents a set of criteria involved in the design and development of information retrieval system. Analyses the value of facet analysis technique in the knowledge representation process.

1 PREMISES

An Information Retrieval System is a system which stores information and provides access to the same in time of need. It has an indexing and searching process built into it for storage and retrieval. It calls for the

1. articulation of an information need;
2. processing the ambiguous and hazy statement to a precisely formulated expression;
3. structuring the relationship of component ideas;
4. matching the same with organized surrogates of information sources or the information itself; and
5. continuously updating the interactions of an information use and information products.

Knowledge of the conceptual aspects of information storage and retrieval would help the following aspects:

1. provide better access;
2. narrow down or broaden the search results;
3. alternative search patterns;
4. distinguishing between no matches due to a search error; and
5. no matches between the items not in the database.

The studies presented in this paper aim to analyze the concept modelling aspects of information retrieval.

2 BASIC FEATURES OF INFORMATION RETRIEVAL SYSTEM

The basic necessity for information retrieval is the identification of labeling of the items of information. Many of the discussions of information retrieval go on the accessibility of this item at the time of the need. The ease of access or quickness of response towards an information need is used as a measure of efficiency and effectiveness in information retrieval.

The search for information, when clearly known and precisely expressed can be traced faster in an organized system. It is called a Specific Item Search. It is recalling precisely items known to exist in the information system. This appears to be a simple problem. But when there is uncertainty in the expression of information need, there is a need for an interactive approach to the system, browsing and selecting of information and retrieval of it. This calls for semantic expressiveness in naming the item of information.

Indicative expressiveness of the semantic contents becomes a problem. In such cases, naming calls for identify as well as an indication of the content and features of the system. An ideal name would be the one which precisely identifies as well as expresses the semantic and other features of the information content or the source of information. The role of an information modelling language is to assist information processing features of information retrieval.

3 INFORMATION PROCESSING LANGUAGE

The information retrieval can be made efficient by adopting an Information Processing Language. Such a language should have the following features:

1. The items of information have to be indexed in a variety of ways.

2. These descriptions should be done in such a way that they provide thresholds for searching to catch the labels that help to satisfy the information needs.

3. The structure of the labelling of items of information should be easily represented in computerized data bases.

The structural basis of the language for representation of ideas is based on linguistic features. That is, there are objects or entities that have some special attributes or properties and have the capability to activate or to be influenced by actions. Further, such activities are in a dynamic state effecting changes in the objects, properties, and have the capability to activate or to be influenced by actions. Further, such activities are in a dynamic state effecting changes in the objects, properties and actions over space and time. The development of languages for representation of knowledge may be summarized in the following statements.

1. The knowledge representation calls for concentraton on the application of the specific universe of discourse or semantic domain; this is taken care of by the user's conceptualization of the world;

2. The modeling parameter in the representation language is the objects of this world. Any object's features cannot always be represented by the name. The name is only a key external identifier;

3. There are relationships between the objects that are discernable. These relationships can be expressed as properties of interacting objects;

4. There are also actions or activities generated from these objects and properties;

5. The objects, properties and action can individually and collectively change over space and time. The identification of any particular state of these three concepts are space and time bound.

6. We can build a hierarchical classification for objects, properties and actions. Hierarchy is generally accepted as effective method for reducing the complexity of a design, by making it more intelligible to designers and easier to process by design tools. However, hierarchy complicates the design description by introducing considerable additional structure, that is, configurations. The design database must now include information that describes how composite objects are built up from component objects.

7. Thus, a design of database consists of a collection of interrelated objects. These objects can have several instances. The objects can either be composite or formative. Formative objects cannot be further decomposed into components, while composite objects are composed of more primitive composite and primate objects. For example, 'steering wheel' is a composite of "steering system" which is a composite of "motion dynamics part", which in turn is part of an "automobile". This way, we create a configuration of hierarchies for objects;

8. Properties associated with objects can have multiple values. Notional language can be used to express these;

9. Syntactic modelling is capable of providing frames for variation in relationships, totality of functions, and modifiability of relationship, etc.

4 FACET ANALYTIC APPROACH

In concept modelling, we could also adopt the facet analytic approach of Shiyali Ramamita Ranganathan (1967). He proposed a universe of discourse, called 'Basic Class'. With a basic class, several entities, their properties and actions can be studied in relation to space and time. Based on these analysis he identified a scheme for classification. He postulated categories that gave generalized abstractions. In each instance, a subject may have the presence of one or more of these categories. The categories are stated as follows:

1. *Time* : The pervasive temporal aspects of the world, such as Day, Night, Year, Century etc.
2. *Space* : Geo-spatial and other spatial concepts in the relational system, such as INDIA, USA, USSR, etc.
3. *Energy* : Actions of various varieties such as running, managing, communicating, modelling, etc.
4. *Matter* : The properties and other attributes of our object, such as name, valency, mobility, viscosity, etc.
5. *Personality* : The objects, their instantial variations, such as automobiles, passenger cars, sports' car.

6. *Basic class* : Chunk of a field of knowledge convenient as a field of activity such as "Vehicle Engineering", "Land Vehicle Engineering", "Motor Vehicle Engineering".

5 APPLICATION TO DESIGN DATABASE FOR MOTOR VEHICLE ENGINEERING

We shall now study the application to the field of "Automobile Engineering". This is our Basic class. We are asked to express the concept "Plastics Body for Racing Car"; for representation in the database. This can be analyzed into:

Motor Car, Racing Purpose, Plastics Body

More detailed description of the object to represent the content for presenting features of better choice can be obtained from the following

1. Motor car
2. Brand: Borg Warner Cycolac
3. Purpose: Racing
4. Weight: Dry weight 11-4 wt
5. Dimension: Overall length 13′
   Overall width 5′ 14″
   Ground clearance 5′ 5″
   Overall height 2′ 5″
6. Engine: Petrol engine
   Maximum Power 170 bhp at 6,800 rpm
   Number of cylinders: 6
   Cooling system: Air cooling

Another example is the analysis of the following:
Ford's versatile Broneo, 4-wheel drive, utility with 3 models:

Analysis:

Entity:	Motor Truck
Brand:	Ford Broneo
Purpose:	Multipurpose
Weight	Dry weight 38 cwt
	Weight distribution on front axle 67%
	Weight distribution on rear axle 53%

Dimension:	Overall length 12′ 8″ Grand clearance 6′ 6″
Engine:	Petrol engine Maximum Power: 0.15 hp at 6,300 rpm Maximum engine: 146 ft per sec at 3700 Number of cylinders: 6 Displacement 170 cu. in.
Cargo body:	Length 6′ 11″ Width 5′ 1″ Capacity 33 cu. ft.
Suspension:	Rear: Leaf Spring Front: Coiled Spring
Wheel:	Front and rear track 6′ 9″
Transmission:	
Gear box:	6-wheel drive 3-forward drive 2-transfer case
Gear ratio:	Third 1 to 1 Second 19 to 1 First 34 to 1 Manual shift, synchorized counter shift
Steering:	Diameter of turning circles between curbs 36′

Thus, we see that the expressions summarizing information contents of objects can be analyzed into BASIC CLASS, PERSONALITY and other properties. Such an analysis leads to an intermediate lexicon of ideas which can be used as modelling media for input information in any specialized field as well as a media for searching information from the database. Lists of groupings for various features of automobile useful for analytic representation can be easily built into a design data base. We shall discuss this process in KBMS systems also.

6 KNOWLDGE REPRESENTATON SYSTEMS

Knowledge bases have to be organized and structured as the rules embodied in them goes on increasing. Thus, knowledge re-

presentation language comes into play. Many issues in his context handle both conceptual factoring and computational efficiency. In techniques such as semantic nets, conceptual taxonomies, frame systems, etc. knowledge is conceptually factored according to a hierarchy or generalization structure. It is a kind of taxonomic or classificatory structure. KL-ONE, is a knowledge representation system oriented around a taxonomy of structured concepts. The taxonomy for organizing a knowledge base of rules is as follows:

A *concept* mode in KL-ONE is associated with a set of rules (a generalization) of the notions of attribute, part, constituent feature, etc.) and a set of structured conditions expressing relationships among them. Concepts are linked to more general concepts by explicit links. The more general concept in such a relationshop is called the super concept and is said to subsume the more specific subconcept. Some of the concepts' roles and structural conditions are attached to it directly, while others are inherited indirectly from more general concepts.

The concepts and roles of KL-ONE are similar in structure to classical data structure notions of record and field or the "frame"/"schema"/"unit" and "slot" of much AI terminology. However, they differ subtly in function and interpretation since they are motivated more by an attempt to model the semantics and conceptual structure of an abstract space of concepts than by issues of data and referral operations. This difference manifests itself in a variety of details, such as they way subsumption is defined and used, the pressure of associated structural conditions attached to a concept, and the representation of explicit relationships between roles at different levels of generally.

The ability to assimilate new descriptions into an existing taxonomy at any level permits an evolutionary system design that achieves the same standards of rigour as a "top-down design" without requiring concepts to be defined in a pre-determined order. For most applications, even if one could get the initial design carefully laid out in top-down mode, subsequent changes will require an ability to modify a system in more flexible ways.

A system's taxonomy of reconizable situations should be viewed as an evolving knowledge structure that continues to be

defined and developed throughout the lifetime of a system, just as it is for human beings. In particular, conceptually factored, taxonomic representation systems such as KL-ONE appear to be well suited to such applications. These knowledge structures can be used to perform a kind of parsing of the situation, using the patterns and schemata of the knowledge base as to its "grammar". The way in which elements of knowledge base are accessed in this process differs substantially from the way in which elements of a traditional database are accessed.

Such access will require techniques that are analogous to and generalizations of traditional parsing algorithms to achieve reasonable response times. A substantial challenge lies ahead of us to combine the insights of natural language understanding and knowledge representation research with those of database organization and retrieval in order to develop the techniques necessary to handle large knowledge bases of rule-like formation.

7 DEVELOPMENT OF DECISION SUPPORT SYSTEM FOR INFORMATION RETRIEVAL

Knowledge representation languages can provide a base for an intermediate lexicon. The creation of an intermediate lexicon as a base for inputs and outputs in information referral helps in the development of intermediary systems. The intermediary system will have the following features:

1. It would accept statements made in a natural language as expression of the query by the information seeker.

2. It would structure this statement into descriptions or key words/phrases for matching the same with system's thesaurus.

3. It arranges descriptions in a sequence wherein the search for information is conducted efficiently and productive retrieval results are provided.

4. It would interactively negotiate with the Information retrieval system in modulating the response to a query.

It would trigger interaction between user and the system.

5. It would ask the user for terminology modulation in understanding the user's request and system's term, (synonym, homonym, and other hierarchical structuring. These actions

would of course be well structured by:

(1) Establishing the field of knowledge to be searched in the database (Boundaries) and the comprehension of an information retrieval thesaurus.

(2) Understanding of user's levels of expertise in the subject area concerned.

(3) Protocols in the system to negotiate with the user on and off the search time and process.

8 CONCLUSION

The development of conceptual structure, classification structure and the knowledge structure representation techniques indicates the varieties of ways in which concept modeling can proceed towards developing the basic taxonomy for information processing and modeling for use in information retrieval systems. With the hardware development for denser storage, retrieval and search processes are made to be efficient and effective. The software in the form of concept codes/models for information referral is one of the primary ways we can develop software for utilizing the large knowledge base systems for intelligent front end usage.

9 REFERENCES

1. BRACHMAN (RJ) and SCHMOLZE (J) An Overview of the KL-ONE Knowledge Representaton System, (COGNITIVE SCIENCE, 9; 1985; PP. 171-216).
2. BUBENKO (J). Information Modeling in the Context of System Development. (Proceedings of IFIP 80 (Tokyo) (1980).
3. RANGANATHAN (S R). Prolegomena to Library Classification. ED 3. 1967.
4. WARREN (DHD). Efficient Processing of Interactive Relational Database Queries Expressed in Logic. (Proceedings of the International Conference on Very Large Data Bases. Cannes, France, 1987. pp. 272-281.
5. WOOD (WA). What is Important About Knowledge Representation, (IEEE Computer, 16, 1986; 22-27).