TWO-TIERED CONCEPT MEANING, INFERENTIAL MATCHING AND CONCEPTUAL COHESIVENESS

by

R. S. Michalski

ISG 86-21, UIUCDCS-F-86-968, University of Illinois, Urbana, IL, June 12-14, 1986.

86-24

TWO-TIERED CONCEPT MEANING, INFERENTIAL MATCHING and CONCEPTUAL COHESIVENESS

Ryszard S. Michalski

Abstract

Human concepts, the building blocks of our knowledge, rarely have precise and context independent meaning. In most cases, they are structures whose meaning can be flexibly modified by the context of discourse and/or the interpreter's background knowledge. This paper outlines several ideas about how such flexibility may be achieved and why it is useful.

The computational method proposed postulates that the meaning of any concept is defined by two components: the *base concept representation* (BCR) and the *inferential concept interpretation* (ICI). The base concept representation is a store of assertions and facts about the concept. It may include statements describing a general and easy-to-define meaning of the concept, its typical purpose or use, known examples and counter-examples, as well as observed exceptions from the typical meaning. It captures the principle or intention behind the concept when it is defined, or taught to others.

The inferential concept interpretation dynamically assigns meaning to a concept by conducting inference - deductive, analogical or inductive - on the BCR using the context of discourse and the interpreter's background knowledge. It does this by employing methods and rules of inference relevant to the given concept or its generalizations. This way the actual meaning of any concept can be flexibly modified or extended to fit many different contexts and purposes.

Experiments with a simple form of such two-tiered concept representation have revealed a surprising phenomenon, that by shifting a larger part of the concept meaning to the inferential concept interpretation, the size of the knowledge base representing a class of concepts can be significantly reduced, without affecting its performance accuracy.

Introduction

Suppose we asked someone how to get to some place in the city we were visiting and received needed instructions in response. Clearly, we would say that this person *knew* the answer, no matter whether the person knew the place personally, or just had to figure out its location on the basis of general knowledge of the city, i.e., by conducting inference. We would say this, of course, only if the answer was given to us in a reasonable amount of time.

The above example illustrates a general principle: one knows what one remembers, or what one can infer from what one remembers within a certain time constraint. Thus, our knowledge can be viewed as a combination of two components, recorded knowledge and inferential extension, i.e., knowledge that can be created from recorded knowledge by conducting inference within a certain time limit.

The main thesis of this paper is that individual concepts - elementary components of our knowledge - parallel such a two-tiered nature of knowledge. We hypothesize that processes of assigning meaning to concepts recognized in a stream of information, or retrieving them from memory to express an intended meaning are intrinsically inferential, and involve, on a smaller scale, the same types of inference - deductive, analogical and inductive - as processes of applying and constructing knowledge in general. This hypothesis reflects an intuition that the meaning of most concepts cannot be, in principle, defined in a crisp and context-independent fashion. Specifically, the concept meaning cannot be completely defined by stating some necessary or sufficient features, defining a prototype or a set of representative examplars. Rather, the meaning of a concept is a dynamic structure built each time anew, in the course of an interaction between some initial base meaning, and the context of discourse together with the interpreter's background knowledge.

This view leads us to the proposition that the meaning we assign to a concept in any given situation is a result of an interplay between two parts: the *base concept representation* (BCR), and the *inferential concept interpretation* (ICI). The base concept representation is a structure residing in memory that records both specific facts about the concept, and general characteristics of it. The specific facts may include representative examples, exceptions, and counter-examples. The general characteristics are teacher-defined, or inferred by induction from examples or by analogy. They include the typical, easily-definable, and possibly context-independent assertions about the concept. These characteristics tend to capture the principle, the ideal or intention behind a given concept. If this principle changes to reflect a deeper knowledge about the concept involved, the base concept representation is redefined. To see this, consider, for example, the changes of our understanding of concepts such as whale (from fish to mammal) or atom (from the smallest undivisible particle to the contemporary notion of a dual wave-matter form).

The inferential concept interpretation is a process of assigning the meaning to a concept using the base concept representation and the context of discourse. This process involves the interpreter's background knowledge and relevant inference methods that allow one to recognize, extend or modify the concept meaning according to the context. These methods are associated with the concept or its generalizations.

The main goal of this paper is to sketch ideas and underlying principles for constructing an adequate cognitive model of human concepts. It is not to define such a model precisely nor to present specific algorithms. It is also hoped that the proposed ideas will suggest better computational methods for representing, using, and learning concepts in artificial intelligence systems.

Inference Allow Us to Remembver Less and Know More

This section will attempt to show that the two-tiered representation of concept meaning, outlined above, can be justified on the basis of cognitive economy, i.e., economy of mental resources - memory and processing power, and that it reflects some general aspects of the organization of human memory. For a discussion of issues concerning cognitive economy see Lenat, Hayes-Roth and Klahr, 1979.

Let us start by assuming that the primary function of our knowledge is to interpret the present and predict the future. When one is exposed to any sensory inputs, one needs knowledge to interpret them. The more the knowledge and the stronger the inferential capabilities (i.e., roughly the number of production and inference rules one possesses), the greater the amount of information one can derive from a given input.

Interpreting observations in the context of the available knowledge makes it possible to derive more information from the input than presented on the surface. It also allows one to build expectations about the results of any action, and to predict and/or influence future events. The latter is possible because events and objects in our world are highly interrelated. If our world consisted of totally unrelated random events, one following the other, our knowledge of the past would be of no use for predicting the future, and this would obviate storing any knowledge. Moreover, this would presumably obviate the need for having intelligence, as the primary function of intelligence is to construct and use knowledge.

On the other hand, if our world were an eternal repetition of exactly the same scenes and events, knowledge once acquired would be applicable forever, and the need for its extension and generalization would cease. No wonder that in old, slow-changing traditional societies, the elderly enjoyed such high status. The slower the rate of change in an environment, the higher the predictive value of past specific knowledge, and the lower the need to extend and generalize knowledge. This suggests a hypothesis that the degree to which our innate, subconscious capabilities generalize any input information corresponds to the rates of change in our environment.

From the myriad sensory inputs and deluge of information received, one's mind selects and stores only a minuscule fraction. This selection is due to the goal-dependent filtering of inputs by our mind. The fraction actually stored contains a spectrum of structures representing different levels of abstraction from reality, and different beliefs in their correctness. This spectrum spans the low-level, highly-believed facts and observations, through partial plausible abstractions and heuristics, to high-level and highly hypothetical abstractions. The highest belief usually is assigned to our own personal sensory experiences, and the lowest belief to vague abstractions made by people whom we do not especially trust. These assertions are being automatically memorized as they are received or generated by inference. They then undergo the processes of forgetting, but cannot be consiously erased.

The the filtering of the input information is done by conducting inference - deductive, analogical and inductive - that engage the input information and the goals and the knowledge of the person. The idea that a person's knowledge is involved in the processes of interpreting inputs is, of course, not new. An interesting illustration of it is presented, for example, by Anderson and Ortony (1975). They conducted experiments showing that the comprehension of a sentence depends heavily on the person's knowledge of the world and his/her analysis of the context.

The ability to conduct the above-mentioned types of inference seems to come from a naturally endowed mechanism that is automatically activated in response to any input of information. One may ask why this is so. As our memory and information processing powers are limited, it seems natural that the mind should tend to minimize the amount of information stored, and maximize the use of that which is already stored. Consequently, one may hypothesize that the inferential processes that transfer any input information to stored knowledge are affected by three factors:

1. what is important to one's goals
2. what knowledge will be maximally predictive, and
3. what knowledge will allow one to infer the maximum amount of other knowledge.

The first factor reflects the known phenomenon that facts considered very important tend to be remembered before other facts. The second factor is important because the predictive power of knowledge enables us to develop expectations about the future, and thus to prevent or avoid undesirable courses of actions, and to achieve goals. The third factor relates to cognitive economy: if we can infer B from A without much cognitive effort then it is enough just to remember A.

The second and third factors have interesting consequences. They suggest a memory organization that is primarily oriented toward storing analogies and generalizations, but facilitates the process of efficiently performing deduction on the knowledge stored.

The above factors explain the critical role of analogical and inductive inference in the process of transforming information received from the environment to knowledge actually memorized. This is so because it is analogical inference that transfers knowledge from known objects or problem solutions to new but related objects or solutions. And it is inductive inference that produces generalizations and causal explanations of given facts (from which one can deduce original facts, and predict new ones). Strict deductive inference and various forms of plausible inference (plausible deductive, analogical and inductive) are means for extending/deriving more knowledge from our base knowledge, though such derived knowledge may be of lesser certainty.

The relationship between different types of inference is shown in Figure 1. The types of inference are divided according to two dimensions: 1. mode of inference: deductive vs. inductive, and 2. strength of inference: crisp vs. plausible. "Crisp" deductive inference is the truth-preserving inference studied in formal logic. "Soft" deductive inference uses approximate rules of inference, and produces probable rather than strict consequences of given premises. This type of inference is, for example, implemented in various expert systems that generate advice together with an estimate of its certainty.

Inductive inference produces hypotheses (or explanations) that crisply or softly imply original facts (premises), i.e., original facts are deductive consequences of the hypotheses. "Crisp" inductive inference is a falsity-preserving inference. For example, if from the premise that "all professors of the Computer Science Department are bright" one generates an inductive hypothesis that "all professors of the University are bright," then it is a falsity-preserving inference. (If the premise is true, the conclusion can be true or false; but if the premise is false, the conclusion must be false also.) "Soft" inductive inference produces hypotheses that only plausibly imply the original facts. For example, seeing a smoke one may hypothesise that there is a fire somewhere. It is a soft inductive inference, because a fire does not necessarily imply smoke.

Analogical inference is placed in the middle, because it can be viewed as inductive and deductive inference combined (Michalski, 1986a). The process of noticing analogy and performing analogical mapping between two systems is intrinsically inductive; the process of deriving inferences from analogy, once noticed, is deductive. This view, derived by purely theoretical speculations, seems to be confirmed by the experimental findings of Gentner and Landers (1985). In order to explain difficulties people have in noticing analogies, they decomposed analogical reasoning into three parts, which they call "access", "structure-mapping" and "inferential power". They found that the "access" and "inferential power" are governed by different rules. The "access" is facilitated by literal similarity or mere appearance, and "inferential power" is governed by similarity of higher-order relations. Analogical "access" and "structure-mapping" are inductive processes, as they produce a structure that unifies the base and the target systems. "Inferential power" corresponds to deduction.

The view of analogy as induction and deduction combined explains why it is more difficult for people to notice analogy than to use it once it is observed. This is so because inductive inference, being an underconstrained problem, typically consumes significantly more cognitive power than deductive inference, which is a well-constrained problem.

Figure 2 illustrates levels of knowledge derived from the base knowledge by conducting various types of inference (the "trumpet model"). The higher the type of inference, the more conclusions can be generated, but the certainty of conclusions decreases. A core theory and a discussion of various aspects of human plausible inference are described in Collins and Michalski [1986].

Let us now return to the discussion of the third factor influencing inferential processes, i.e., what knowledge allows us to infer the maximum amount of other knowledge. This issue, obviously, has special significance for achieving cognitive economy. The need for cognitive economy implies that it is useful for individual words (concepts) to carry more than one meaning, when considered without any context and without inferential extension of their meaning. By allowing that the meaning of words be context-dependent and inferentially extensible, one can greatly expand the number of meanings that can be conveyed by individual words. This context-dependence, however, cannot be unlimited again because of cognitive economy. To be economical, context dependency should be employed only when the context can be identified with little mental effort. Inferential extensions also have natural limits, which are dictated by the mental power available, and the decreasing confidence in conclusions as the levels of inference increase.

Concept Meaning is Distributed Between Representation and Interpretation

Concepts are mental structures representing classes of entities united by some principle. Such a principle might be a common use or goal, the same origin or behavior, or just similar perceptual characteristics. In order to use concepts, one must possess efficient methods for recognizing them in streams of sensory signals. To do so, one needs to have appropriate mental representations of concepts.

The traditional work on concept representation assumes that the whole meaning of a concept resides in a single stored structure, e.g., a semantic network that captures all relevant properties of the concept (e.g., Collins and Quillian, 1972; Minsky 1975; Sowa, 1984). The process of recognizing a concept involves simple matching between the stored representation and perceived facts. Such matching may include tracing links in a network, but has not been assumed to involve any complex inferential processes.

In contrast, our view is that such a matching may involve a significant amount of deductive, analogical or inductive inference that takes into consideration the context of discourse and background knowledge of a person. Therefore, we postulate a two-tiered representation of concept meaning, which draws a distinction between the base concept representation and inferential concept interpretation. The base concept representation is a stored knowledge structure associated with the concept. It specifies the most common, typical properties of the concept, and the principle unifying different instances of the concept. It may also include representative examples, counter-examples, exceptions and other known facts about the concept. The inferential concept interpretation uses methods, relevant background knowledge and rules of inference for interpreting the base concept representation according to various contexts. The methods incorporate meta-knowledge about the concept, i.e., which properties of the concept are crucial and which are not for a given situation, how they vary among instances of the concept. They also contain procedures for matching the base concept representation with observations. Figure 3 illustrates the two-tiered concept meaning. The rectangular area denotes the scope of a concept as defined by the base concept representation. The shaded area depicts the changes in the concept meaning due to the inferential concept interpretation. For example, the rectangular area may represent all animals sharing typical physical characteristics of fish, and the dotted-line area may represent animals that can be considered fish in various contexts.

It is easy to see that to recognize an object, i.e., to assign it to a concept, one may need to match only a small portion of properties observed in the object with properties stated in the base representation. The properties that need to be matched depend on the context in which the recognition process occurs.

For example, one may recognize a given person just by some of this person's face features, the silhouette, voice, handwriting, medical record, or by one of a host of other characteristics. Thus, if the concept recognition process were based on a direct match of a fixed number of features of the target concept with properties of an observed object, then one would need to store representations for all these possibilities. Such a method would be hopelessly memory-taxing and inefficient. It is practical only in simple cases, such as those considered in most of the current expert systems.

In the proposed theory, the process of relating the base concept representation to observations is done by inferential concept interpretation. This process "matches" the base concept representation with observations by conducting inference involving the contextual information and relevant background knowledge. This inference determines what features are needed or sufficient to be matched in order to recognize a concept among a context-dependent set of candidates, and what kind of match is required. Thus, the degree of match between a concept representation, CR, and an observed entity, OE, is not just a function of CR and OE, as traditionally assumed, but rather a four argument function:

Degree_of_match(CR,OE)=f(CR,OE,Context,Background_knowledge)

The context is computed dynamically in the process of using or recognizing concepts. Thus, the proposed view requires an efficient method for representing and using contexts for any given concept. A simple introspection of our mental processes appears to confirm this: we seem to have little difficulty in determining and maintaining the context in any discourse.

There is no unique way of distributing the concept meaning between BCR and ICI. We expect that the actual distribution of the concept meaning between these two parts in our mind represents a desired tradeoff between the economy of concept representation, and the economy of inferential concept interpretation. Thus, learning a concept involves not only acquiring the base concept representation, but also the methods for inferential concept interpretation.

Let us illustrate the proposed approach by a few examples. Consider the concept of fish. Typical and general characteristics of fish are that they live in water and they swim. These and other typical physical properties of fish, as well as representative examples would be stored in the base concept representation.

Suppose someone found an animal that matches many characteristics of fish, but which does not swim. Suppose that this animal appears to be sick. The inferential concept interpretation would involve background knowledge that sick animals may not be able to move, and that swimming is a form of moving. By deductive reasoning from these facts one concludes that lack of ability to swim should not be taken as negative evidence for the animal being a fish. To the contrary, the fact that the animal does not swim might even add to the confidence that it is a fish, once the animal was recognized as being sick.

Suppose that we learned the concept of fish by reading a general description, and seeing a few examples of fish. The base concept representation consists of this general description and the memorized examples. Suppose that we visit a zoo and see an animal defined as fish that is of a shape never seen in the examples nor stated in the general description, say, of a horse-like shape. We may add this example to our base concept representation without necessarily modifying our general notion of fish. If we see another horse-shaped fish, we may recall that example to recognize the new instance of a fish without evoking the general notion of fish. This explains why we postulate that the base concept representation is not just a representation of the general, typical or essential meaning of a concept, but includes also examples of a concept.

The rules used in the above reasoning about sick fish and horse-like shape of fish would not be stored as the base concept representation for fish. They would be a part of the methods for inferential concept interpretation. These methods would be associated with the general concept of animal, rather than with the concept of fish, because they apply to all animals. Thus, we postulate that the methods for inferentially interpreting a concept can be inherited from those applicable to a more general concept.

As another example, consider the concept of sugar maple. Our prototypical image of a sugar maple is that it is a tree with three- to five-lobed leaves that have V-shaped clefts. Some of us may also remember that the teeth on the leaves are coarser than those of red maple, that slender twigs turn brown, and the buds are brown and sharp-pointed. Being a tree, a sugar maple has, of course, a trunk, roots and branches.

Suppose now that while strolling on a nice winter day someone tells us that a particular tree is a sugar maple. Simple introspection tells us that the fact that the tree does not have leaves would not strike us as a contradiction of our knowledge about sugar maples. This is surprising, because, clearly, the presence of leaves of a particular type is deeply embedded in our typical image of a maple tree. The two-tiered theory of concept representation explains this phenomenon simply: the inferential concept interpretation associated with the general concept of tree evokes a rule "in winter deciduous trees lose leaves." By deduction based on the subset relationship between a tree and a maple tree, the rule would be applied to the latter. The result of this inference would override the stored standard information about maple tree, and the inconsistency would be resolved.

Suppose further that when reading a book on artificial intelligence we encounter a drawing of an acyclic graph structure of points and straight lines connecting them, which the author calls a tree. Again, calling such a structure a tree does not evoke in us any strong objection, because we can see in it some abstracted features of a tree. Here, the matching process simply involves inductive generalization of the base concept representation. Once such a generalized notion of a tree is learned in the context of mathematical concepts, it will be used in this context.

These examples clearly show that the process of relating observations with concept representations is much more that matching features and determining a numerical score characterizing the match, as done in various mechanized decision processes, e.g., expert systems.

It should be noted that the distribution of the concept meaning between the representation and interpretation parts is not fixed, but can be done in many ways. Each way represents a tradeoff between the amount of memory for concept storage and computational complexity of concept use. At one extreme, all the meaning can be expressed by the representation. In this case the representation explicitly defines all properties of a concept, including any concept variations, exceptions and irregularities. It states directly the meaning of the concept in every possible context. It stores all known examples of the concept. This results in a very complex, memory-taxing concept representation. The concept interpretation process would, however, be relatively simple. It would involve a straightforward matching of the properties of the unknown object with information in the concept description.

At the other extreme, the concept is explicitly represented only by the most simple description characterizing its idealized form. The process of matching a concept description with observations might be in this case significantly more complex.

As far as memory representation of concepts is concerned, we assume that their base concept representations are stored as a collection of assertions and facts. These collections are organized into *part* or *type* hierarchies with inheritance properties. The methods used by inferential concept interpretation are also arranged into hierarchies. For example, the rule that a sick fish may not swim is not stored with the ICI methods associated with the concept of fish, but rather with the concept of animal.

As mentioned earlier, the process of inferential concept interpretation may involve performing on the base concept interpretation not just truth-preserving deductive inference, but various forms of plausible inference. In particular, it may create an inductive generalization of the base concept representation, draw analogies, run mental simulations, or envisioning consequences of some acts or features. The background knowledge needed for inferential interpretation includes information about methods for relating concept representations to observations, about which properties are important and which are not in various contexts, information about typicality of features, statistical distribution of properties and concept occurrences, etc. An inferential interpreter may produce a "yes-no" answer, or a score representing the degree to which the base representation matches given observations. Extending the meaning of a single concept by conducting inference corresponds on a small scale to extending any knowledge by inference.

When an unknown entity is matched against a base concept representation, it may satisfy it directly, or it may satisfy some of its inferential extensions. The type of inference performed to match the description of the entity with the base concept representation determines the type of match (Figure 4). If the description of an entity strictly satisfies the base concept representation, i.e., matches it directly or its specialization (in other words, falls into its deductive extension), then we have a *strict match*; if it satisfies an approximate deductive extension, then we have an *approximate* match; if it matches an analogical extension, i.e., satisfies a generalization that unifies the base concept representation with the description of the entity, then we have an *analogical match.*

The above mentioned analogical match is not to be confused with the analogical mapping discussed in structure-mapping theory of analogy by Gentner (1983). The analogical match is related to what Gentner and Landers (1985) call "analogical access." It involves finding semantic correspondences between attributes and relations of the entity to be recognized, and the base knowledge representation.

As mentioned earlier, when recognizing an entity in the context of a finite set of candidate entities, usually only a small part of the properties of the entity will need to match the properties in the base representation of candidate concepts. This set is defined by the *discriminant concept description*, which can be determined by conducting inductive inference on the base representation of the candidate concepts. A method for an efficient recognition of concepts in the context of candidate concepts, called *dynamic recognition*, is described in Michalski (1986b).

The process of inferential concept interpretation can be viewed as a vehicle for extending the base concept meaning into a large space of variations by the use of context and general knowledge. This process is an important means for achieving flexibility of concepts, and thus leads to cognitive economy. Later, in the section describing experimental results, we present an example of an inferential interpretation of a simple logic-style base concept representation.

Some Other Views on Concept Representation

There seems to be a universal agreement that human concepts, except for special cases occurring predominantly in science (concepts such as a prime number, a triangle, a vertebrate, etc.), are structures with flexible and/or imprecise boundaries. They allow a varying degree of match between them and observed instances, and have context-dependent meaning. Flexible boundaries make it possible to "fit" the meaning of a concept to changing situations, and to avoid precision when it is not needed or not possible. The varying degree of match reflects the varying representativeness of a concept by different instances. According to the theory presented, this is accomplished by applying the inferential concept matching which takes into consideration the context and background knowledge of the interpreter.

Instances of a concept are rarely homogeneous. Among instances of a concept people usually distinguish a "typical instance", a "non-typical instance", or, generally, they rank instances according to their typicality. By the use of context, the meaning of almost any concept can be expanded in a multitude of directions that cannot be predicted in advance. An interesting illustration of this is given by Hofstadter (1985; ch.24), who shows how a seemingly well-defined concept, such as "First Lady," can express a great variety of meanings depending on the context in which it is applied. For example, it might even include the husband of Margaret Thatcher.

Despite various efforts, the issue of how to represent concepts in such a rich and context-dependent sense is not resolved. Smith and Medin (1981) distinguish between three approaches: the *classical view*, the *probabilistic view*, and the *exemplar view*. The classical view assumes that concepts are representable by features that are singly necessary and jointly sufficient to define a concept. This view seems to apply only to very simple cases. The probabilistic view represents concepts as weighted, additive combinations of features. It postulates that concepts should correspond to linearly separable subareas in a feature space. Experiments indicate, however, that this view is also not adequate (Smith and Medin, 1981; Wattenmaker et al, 1986). The exemplar view represents concepts by one or more typical exemplars, rather than by generalized descriptions. While it is easy to demonstrate that we do store and use concept exemplars for some particular purposes, it seems clear that we also store certain abstract concept representations. Many important novel ideas on concept representation and organization from the computational viewpoint are in (Minsky, 1980; Sowa, 1984; and Lenat, Prakash and Shepherd, 1986).

The notion of typicality can be captured by a measure, called *family resemblance* (Rosch and Mervis, 1975). This measure represents a combination of frequencies with which different features occur in different subsets of a superordinate concept, such as furniture, vehicle, etc. The individual subsets are represented by typical members. Non-typical members are viewed as corruptions of the typical, differing from them in various small aspects, as children differ from the parents (e.g., Wittgenstein, 1921; Rosch and Mervis, 1975).

Another approach uses the notion of a *fuzzy set* as a formal model of a concept (Zadeh, 1976). Members of such a set are characterized by a gradual numerical set membership function, rather than by the in/out function seen in the classical notion of a set. This set-membership function is defined by people describing the concept, and thus is subjective. This approach allows one to express the varying degree of membership of entities in a concept, but does not have appropriate mechanisms for expressing and handling the context- and background knowledge-dependence of the concept meaning. It does not explain what are the computational processes that determine the set membership functions.

The idea of two-tiered representation of concept meaning, described here, first appeared in a simple form in the experiments conducted by Michalski and Chilausky (1980) on inductive knowledge acquisition. In these experiments, two-valued logic-based decision rules were learned by induction form examples. These rules, however, were tested on new facts by interpreting them not as two-valued logic expressions, but by applying to them various many-valued logic evaluation schemes. For example, the logical disjunction was interpreted either as the maximum function and as the probabilistic sum. The logical conjunction was interpreted as the minimum function, the average or the probabilistic product. The experiments showed that such modifications of rule interpretation can lead to an improvement of the rule performance.

A more advanced inferential matching was proposed in the method of conceptual clustering described by Michalski and Stepp (1983). The method utilized the idea of *conceptual cohesiveness*. Suppose that a given observed object does not match any concept description precisely. There are, however, several concepts that are candidates for an imprecise, or, generally, an inferential match. The proposed solution is to generalize each concept so that it includes the object under consideration. The resulting generalized concepts are then evaluated from the viewpoint of conceptual cohesiveness. Such a measure attempts to minimize the degree of generalization that the concept represents over the known facts, and maximize the simplicity of the obtained description of the enlarged set. The concept that receives the highest score after being generalized to include the unknown object is viewed as the right "home" for the object. The concept of conceptual cohesiveness is graphically illustrated in Figure 5.

Most related to our ideas is the work by Murphy and Medin (1985), and Barsalou and Medin (1986). Computational techniques of using knowledge for interpreting observations via deductive inference are presented in the work by DeJong (1986), DeJong and Mooney (1986), and Mitchell (1986). Collins and Gentner (1986) present an illustration of some of the issues involved in processes of creating a mental representation of concepts.

Next section describes our recent experimental study investigating a simple form of two-tiered concept representation done in the context of learning decision rules from examples in the area of medicine.

The Two-Tiered Representation Can Reduce Memory Needed: An Experiment

We will describe here the results of an experiment investigating a simple form of two-tiered concept representation. Concepts under consideration were four different types of lymphography. In the experiments the base concept representation is a disjunction of conjunctive expressions called a *cover*. Individual conjunctive expressions are called *complexes*; each complex is a conjunction of relational statements, called *selectors*. The selectors characterize some aspect of the described entity by stating a value or a set of values that an attribute takes on for a set of entities representing a concept. For example, here are two examples of relational statements (selectors):

[Blood type = A or B]

(Read: The blood type is A or B.)

[Diastolic blood pressure = 65 .. 90]

(Read: The diastolic blood pressure is between 65 and 90.)

Each complex (a conjunction of selectors) in the representation (cover) is associated with a pair of weights: t and u, representing respectively the *total* number of known cases that it covers or explains, and the number of cases that it covers *uniquely*. Thus, statements with high t-weight may be viewed as characterizing typical cases of a concept, and statements with low u-weight can be viewed as characterizing rare, exceptional cases, or errors.

Complexes in the disjunction (serving as the base concept representation) were ordered according to the decreasing values of t-weights. The distribution of the concept meaning between the base concept representation and inferential concept interpretation was varied by applying the so called TRUNC method. First, the lightest complexes were cut off (i.e., complexes with the smallest t-weight), then next lightest, etc., until only one, termed the *best complex*, remained in the base concept representation (Figure 6). Each so truncated cover was used to diagnose a number of new cases of the disease it describes, and the performance score was determined. The diagnosis was determined by simple inferential matching (called "flexible matching") of the cover with a disease case, which took into consideration which properties were matched completely, which were not matched, what was the relative range of values of a property in the base concept representation with regard to the possible range, and the t-weights of the complexes to be matched (that serve as prior probability estimates). Technical details on the matching function and the method are in [Michalski et al., 1986].

A summary of results are shown in Figure 7. Only three cases of cover reduction are presented:

* "no" (no cover reduction) - when the base concept representation included all complexes that were needed to represent all known cases of the given disease, i.e., the *complete* description.

* "unique > 1" - when the cover included only complexes with the u-weight greater than 1,

* "best cpx" - when the base concept representation was reduced to the single complex with the highest t-weight (the "heaviest").

The system's performance was evaluated by counting the percentage of the correct diagnoses (defined the diagnosis that receives the highest degree of match and is considered correct by an expert - see column "Accuracy 1st choice"). For comparison, the columns "human experts" and "random choice" show the estimated performance of human experts, and the performance representing random choice.

As shown in Figure 7, the best performance (82%) was obtained surprisingly when the base concept representation consisted of only **one conjunctive statement** (complex) per concept ("best cpx"). This representation was also, of course, the simplest, as it required approximately one-forth the memory of the complete description.

These results show that by using a very simple concept representation (here, a single conjunction), and only a slightly more complex concept interpretation (as compared to the one that would strictly match the complete concept description), one may significantly reduce the amount of storage required without affecting the performance accuracy of the concept description. Further details and more results from this experiment are described in [Michalski, et al., 1986]. Amoung interesting topics for further research are to study and experiment with more advanced methods for base concept representation and inferential matching, and to test whether similar results can be obtained in other domains of application.

Summary

The two-tiered concept representation postulates that the total concept meaning is distributed between a base concept representation and an inferential concept interpretation. The base representation covers the typical, easily explainable concept meaning, and contains a store of facts about the concept. The inferential concept interpretation is a vehicle for using concepts flexibly and adapting their meaning to different contexts. This is done by conducting inference for matching the base concept representation against observations. This inference process involves contextual information and relevant background knowledge. It can involve all types of inference, from truth-preserving deductive inference, through

approximate deductive and analogical inference, to falsity-preserving inductive inference.

Experiments testing some of the ideas on a simple medical example showed that distributing concept meaning more toward inferential concept interpretation than toward the base concept representation (as compared with a *complete* concept representation and simple, direct matching) was quite beneficial. It has lead to a significant reduction of the needed size of memory for storing concept descriptions without decreasing the diagnostic performance.

Acknowledgements

The author thanks Doug Medin, Andrew Ortony, Gail Thornburg and Maria Zemankova for comments and useful criticisms on the earlier versions of this paper. Remarks of Intelligent Systems Group members P. Haddawy, L. Iwanska, C. Kadie, B. Katz, and H. Ko were also helpful.

This research was supported in part by the National Science Foundation under grant No. DCR 84-06801, in part by the Office of Naval Research under grant No. N00014-82- K-0186, and in part by the Defense Advanced Research Projects Agency under grant N00014-K-85-0878.

References

Anderson, R. C. & Ortony, A. (1975) On Putting Apples into Bottles - A Problem of Polysemy, *Cognitive Psychology*, 7, 167-180.

Barsalou L. W., Medin D. L. (1986) Concepts: Fixed Definitions or Concept-Dependent Representations, Invited paper for a special issue on Concepts and Cognition of *Cahiers de Psychologie Cognitive*, 6, 187-202.

Collins, A., Gentner., D. (1987) How People Construct Mental Models, to appear in Cultural Models in Language and Thought, D. Holland and N. Quinn (eds.), *Cultural Models in Language and Thought*, Cambridge University Press, 243-265.

Collins, A. M. and Quillian, M.R. (1972) Experiments on Semantic Memory and Language Comprehension, in *Cognition Learning and Memory*, Lee W. Gregg (ed.), John Wiley & Sons, Inc.

Collins, A. and Michalski, R. S. (1986) Logic of Plausible Inference, Reports of the Intelligent Systems Group, 86-3, Department of Computer Science, University of Illinois, Urbana.

DeJong, G. (1986) An Approach to Learning from Observation, Machine Learning: An Artificial Intellignece Approach II, Michalski R.S., Carbonell J. G. and Mitchell T. M. (eds.), Morgan Kaufmann Publishers, pp.571-590.

DeJong, G. and Mooney, R. (1986) Explanation-based Learning: An Alternative View, *Machine Learning*, *1* (2).

Gentner, D. (1983) Structure-Mapping: A Theoretical Framework for Analogy, *Cognitive Science*, *7*, 155-170.

Gentner D., Landers, R. (1985) Analogical Reminding: A Good Match is Hard to Find, Paper prepared for the Panel on Commonsense Reasoning at the International Conference on Systems, Man and Cybernetics, Tucson, Arizona, November.

Hofstadter, D. R. (1985) *Metamagical Themas: Questing for the Essence of Mind and Pattern*, Basic Books, Inc.

Holyoak, K.J. (1985) The Pragmatics of Analogical Transfer, *The Psychology of Learning and Motivation*, 19.

Fillmore, C. (1982) Towards a Descriptive Framework for Spacial Deixis, in *Speech, Place and Action: Studies in Deixis and Related Topics*, R. L. Jarvell & W. Klein (Eds.), Chichester, England Wiley, pp. 31-59.

Lenat, D. B., Hayes-Roth, F., and Klahr, P. (1979) Cognitive Economy in Artificial Intelligence Systems, IJCAI 6, Tokyo.

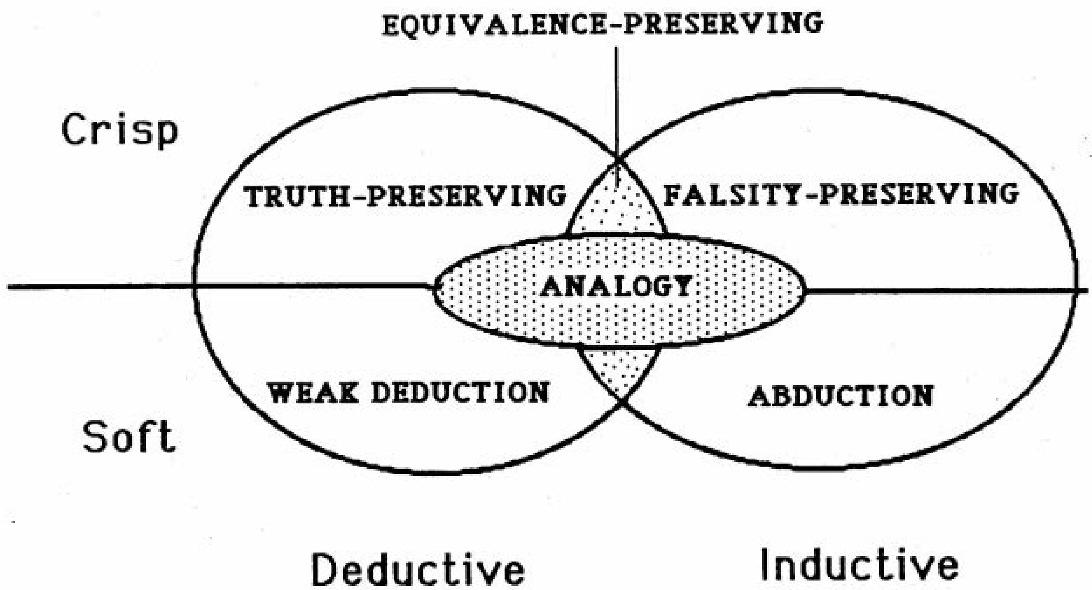
Lenat, D., Prakash, M. and Shepherd, M. (1986) CYC: Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks, AI Magazine, *6* (4).

Minsky, Marvin (1975) A Framework for Representing Knowledge, chapter in *The Psychology of Computer Vision*, P.H. Winston (ed.), McGraw-Hill, New York.

Mitchell, T. M., Keller, R. M., and Kedar-Cabelli, S.T. (1986) Explanation-Based Generalization: A Unifying View, *Machine Learning*, *1* (1), 47-80.

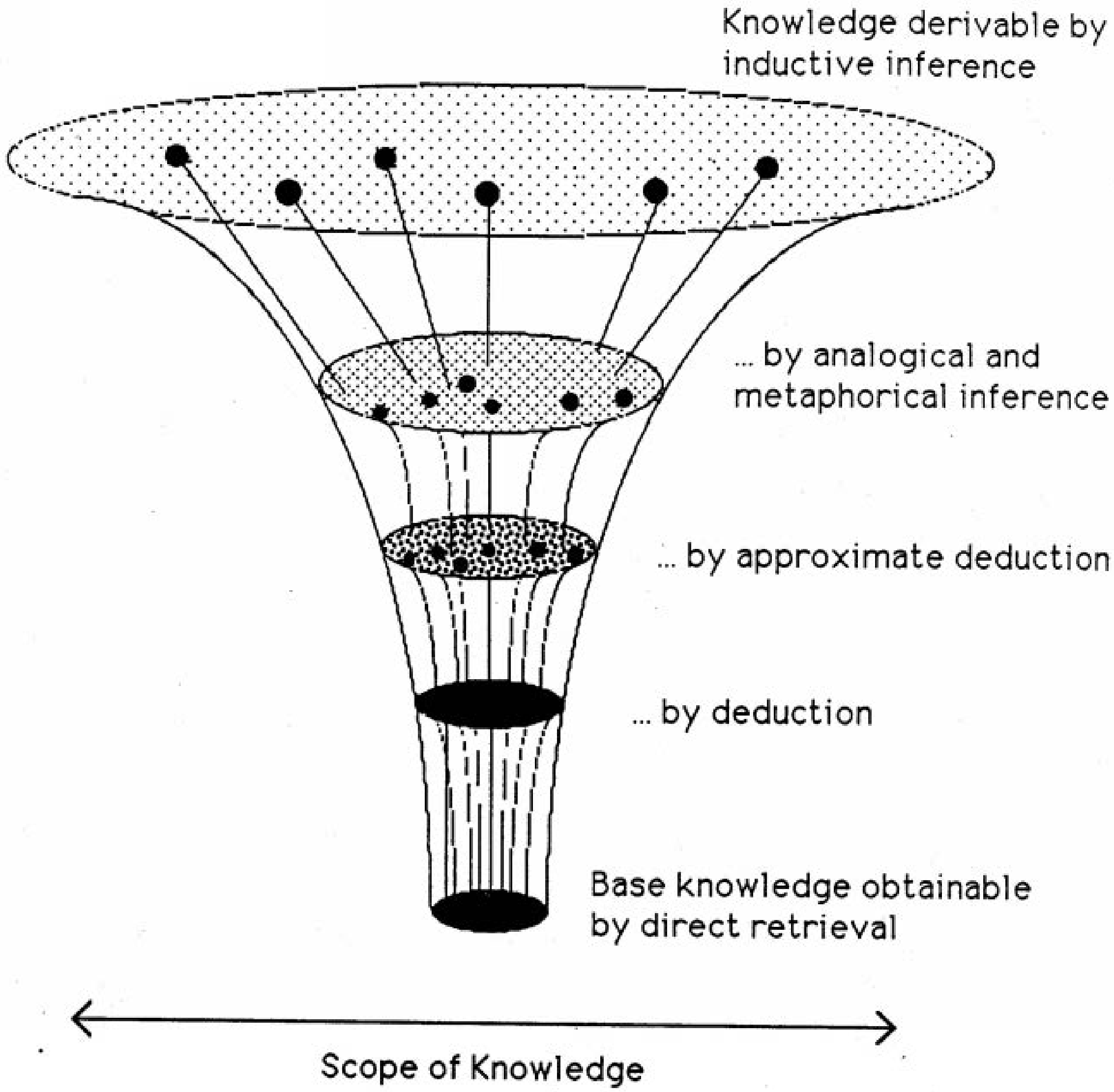
Michalski, R. S. and Chilausky, R. L. (1980) Learning by Being Told and Learning from Examples: An Experimental Comparison of Two Methods of Knowledge Acquisition in the Context of Developing an Expert System for Soybean Disease Diagnosis, *International Journal of Policy Analysis and Information Systems*, *4* (2).

Michalski, R.S., Stepp, R.E. (1983) Learning from Observation: Conceptual Clustering, in *Machine Learning - An Artificial Intelligence Approach*, R. S. Michalski, J. G. Carbonell and T. M. Mitchell (Eds.), Morgan Kaufmann Publishers, pp. 331-364.



Types of Inference

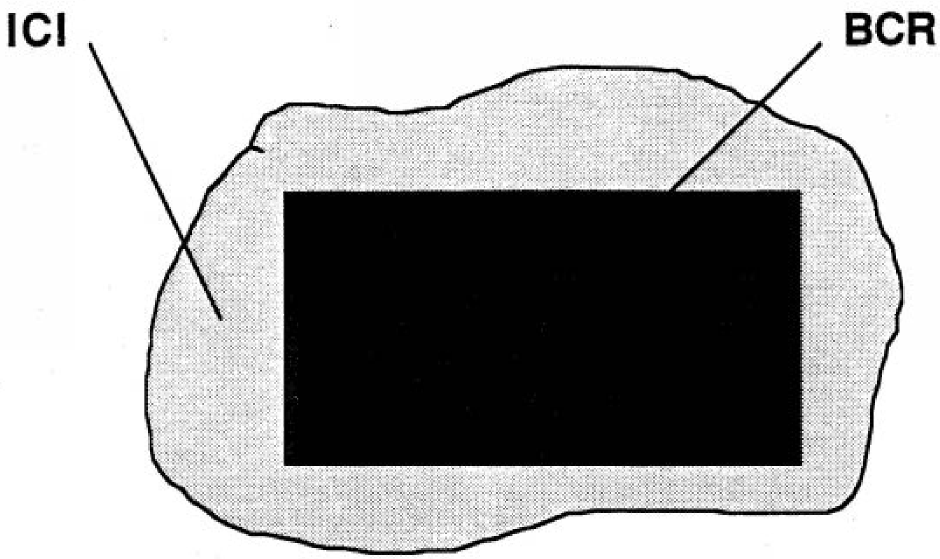
Figure 1.



Shading represents decreasing strength of belief in inferentially derived knowledge.

A "Trumpet" Model of Inferential Knowledge Extension

Figure 2.

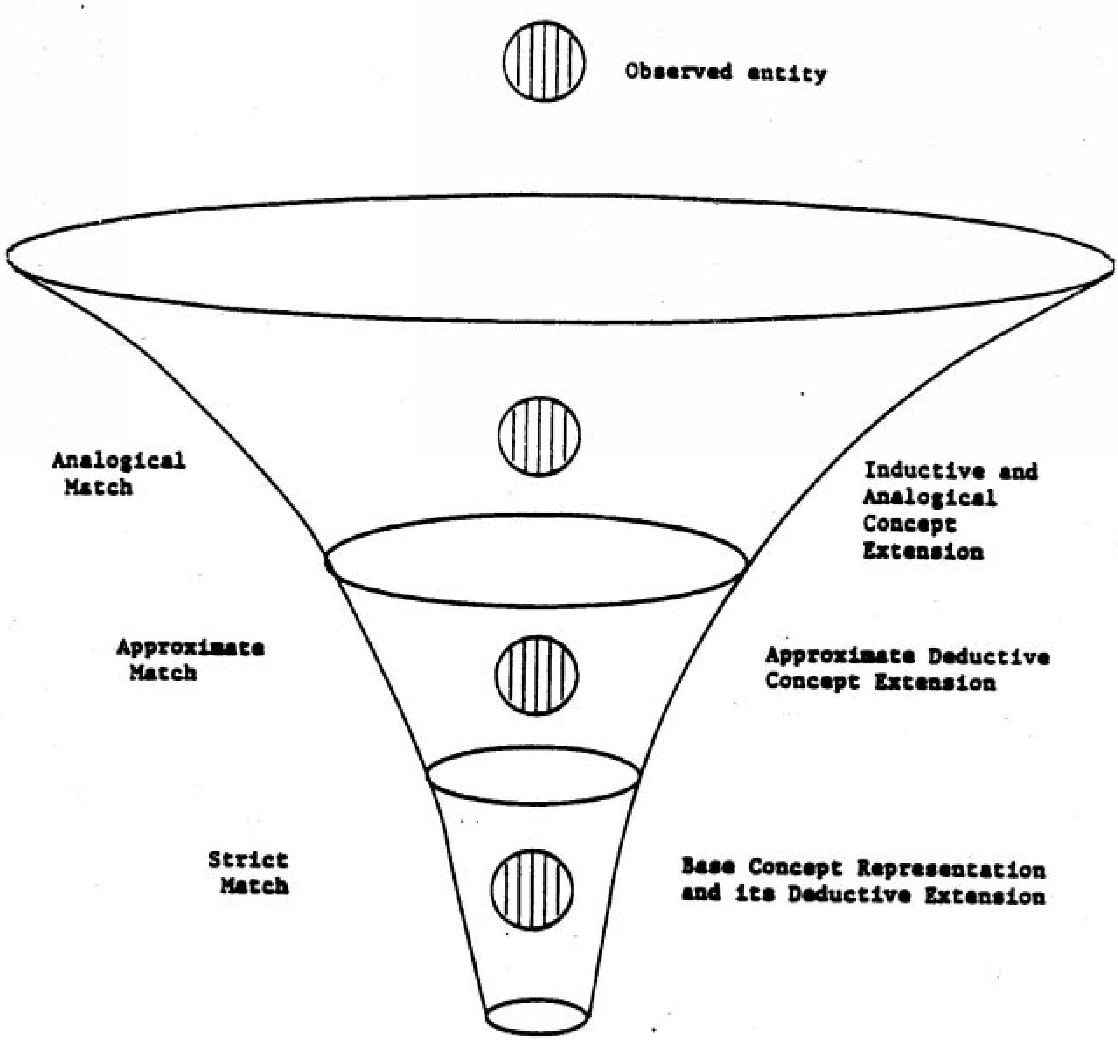


BCR - the scope of the concept defined by the Base Concept Representation

ICI - the scope of the concept as derived by Inferential Concept Interpretation for a given context and background knowledge

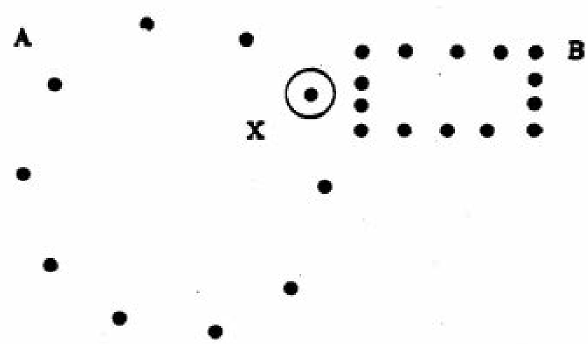
An Illustration of the Two-tiered Concept Representation

Figure 3.



Types of Inferential Concept Matching

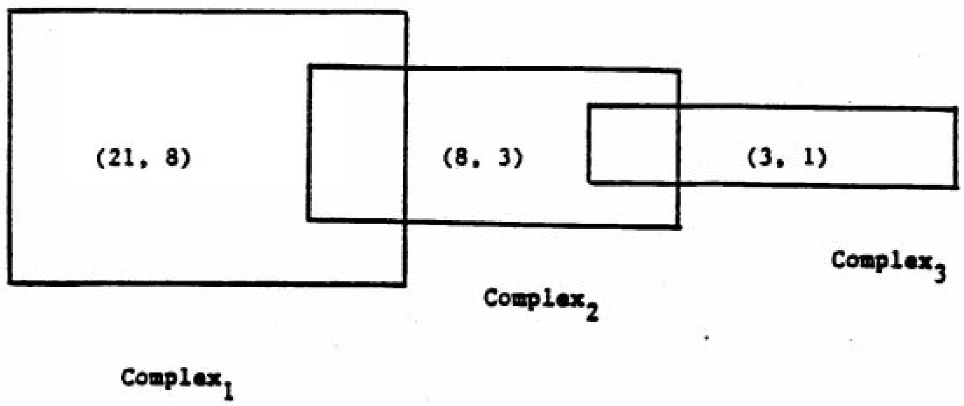
Figure 4.



Object X has higher conceptual cohesiveness with concept concept A than with concept B though it is "closer" to B.

An Illustration of Conceptual Cohesiveness

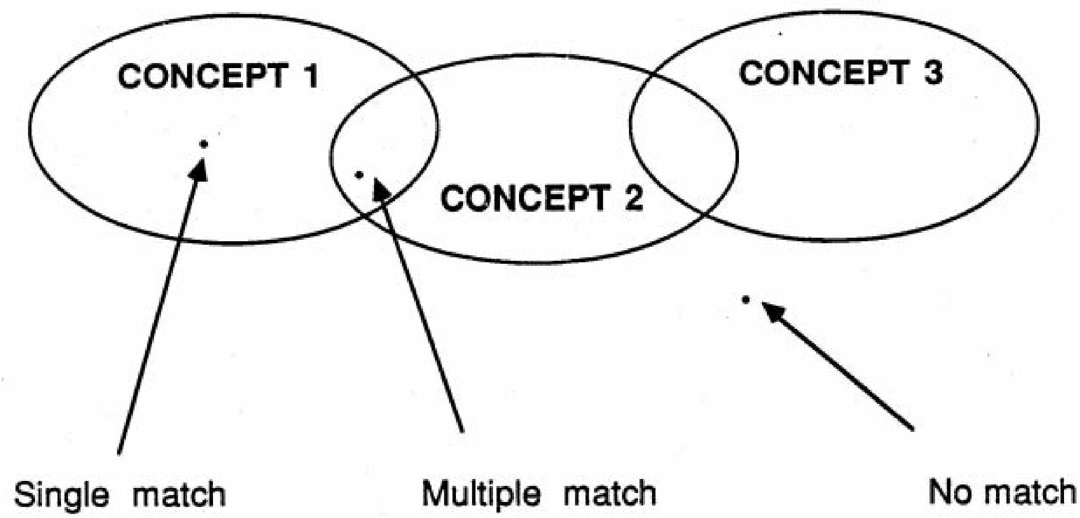
Figure 5.



Numbers in parentheses denote the t-weight and u-weight of the complexes in the cover, respectively.

An Ordered Cover as a Base Concept Representation

Figure 6.



Three Possible Outcomes of Matching an Event with the Base Concept Representation of Different Concepts

Figure 7.

Domain	Cover reduction	Complexity #Sel	#Cpx	Accuracy 1st choice	Human Experts	Random Choice
Lymphography	no	37	12	81%	60/85% (estimate)	25%
	unique >1	34	10	80%		
	best cpx	10	4	82%		

A Summary of Results

Figure 8.