Chapter 16

Using Pathfinder as a Knowledge Elicitation Tool: Link Interpretation

Nancy Jaworski Cooke

Over the last two decades cognitive psychologists have accumulated evidence that suggests that domain-specific knowledge is central to expertise (e.g., Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981). During this same period, the concept of "knowledge-based" systems has surfaced in the field of artificial intelligence. These systems incorporate and rely heavily on specific facts and rules about the problem-solving domain in which they specialize (Hayes-Roth, Waterman, & Lenat, 1983). This emphasis on specific knowledge can be contrasted to previous approaches to human and machine problem solving that emphasized general strategies that could be applied across various domains (e.g., Newell & Simon, 1972). Today most cognitive scientists do not deny that such general strategies exist, but they emphasize the importance of domain-specific knowledge.

This knowledge-intensive trend in cognitive science has also influenced several applied areas. Recent developments in intelligent tutoring systems (Sleeman & Brown, 1982) capitalize on differences between the student's knowledge (i.e., the student's model) and some ideal knowledge base that an expert might be expected to possess. The idea is that this information will provide clues to student misconceptions and direct the tutoring session accordingly. In the area of human-computer interaction, the knowledge that a computer user has about a system (i.e., the mental model) is thought to guide learning and interaction with the system (Norman, 1983). Therefore, it is beneficial for system designers to be aware of this mental model so that the system interface can be designed in accordance with user expectations. An example of this application is discussed by McDonald and Schvaneveldt (1988). Finally, in several domains such as medical diagnosis, computer configuration, and chemistry, expert or knowledge-based systems have been developed that perform some of the tasks of human experts (Waterman, 1986). These systems embody numerous facts and rules about the domain in question. The success of all of these applications depends greatly on knowing what the student, user, or expert knows. Unfortunately, the process of eliciting a person's knowledge is a difficult and time-consuming task.

The bulk of the recent work on knowledge elicitation has been done for expert system development by knowledge engineers who are saddled with the task of uncovering the facts and rules that a human-domain expert uses and the task of transferring this information to the knowledge base of the system. This process is not well-specified, but typically involves behavioral observations, interviews with experts, and the collection of thinking-aloud protocols (Ericsson & Simon, 1984) in which the experts are asked to verbalize their mental processes while performing a domain-related task. Although these techniques are successful at uncovering some basic facts and rules, they suffer from some serious limitations.

Of particular concern is the fact that experts have difficulty verbalizing their knowledge. Nisbett and Wilson (1977) reviewed a large amount of psychological literature that concluded that verbal reports are often inaccurate and incomplete. Experts in particular may have difficulties because, by virtue of their extensive experience, much of their knowledge is automatic (Shiffrin & Schneider, 1977) or compiled (Anderson, 1982). In addition to problems with introspection and verbal reports, these typical knowledge elicitation techniques are very subjective in that they require extensive interpretation on the part of the knowledge engineer in order to transform the verbalizations or observations into a computer-useable format. These techniques are also limited in the sense that they provide virtually no information regarding the appropriate organization of knowledge. Studies on human expertise have indicated that experts differ from novices not only in the facts and rules that they possess, but also in the way that those facts and rules are organized in memory (Adelson, 1981; Chi, Feltovich, & Glaser, 1981; Murphy & Wright, 1984). These limitations must be overcome in order to make advances in knowledge-intensive applications, such as expert systems.

Several investigators (Butler & Corter, 1986; Cooke & McDonald, 1986, 1987; Gammack & Young, 1985) have proposed that a formal knowledge elicitation methodology based on psychological scaling techniques like Pathfinder would overcome some of these limitations. It is assumed that the relatedness estimates that are required by such techniques reflect information about the subject's knowledge structure and in this sense, the scaling techniques are knowledge elicitation techniques. Because scaling techniques generate structural representations of a concept set, they address the knowledge organization issue. In addition, the procedures for collecting data and generating structural representations are more constrained than techniques such as protocol analysis. Most importantly, experts are required to make simple relatedness judgments about concepts rather than difficult verbalizations of the mental processes underlying such judgments.

Given the potential advantages of scaling techniques over traditional knowledge elicitation methods, advances could be made in knowledge-intensive applications by employing such techniques. However, there are several issues that arise concerning the application of these techniques to knowledge elicitation, such as selection of an appropriate method to obtain distance estimates and selection of a particular scaling technique (Cooke & McDonald, 1987). Furthermore, scaling methods empirically capture the structure of knowledge, but they do not capture the content of that knowledge; where do the concepts that are to be represented by these techniques come from? Cooke (1987) addressed this issue by comparing four concept elicitation techniques that could be used as an initial step in a scaling-based methodology. The techniques included: (1) concept listing, (2) step listing, (3) chapter listing, and (4) transcription of concepts from a 20-minute dialogue. Comparison of the concept elicitation techniques revealed that they differ in the number of concepts elicited as well as in the form of that knowledge (e.g., fact, rule, explanation). Cooke suggested that several techniques be used in combination to elicit a relatively complete set of concepts.

Another issue concerns the mapping of the scaling representations onto the application (e.g., student model, interface, knowledge base). It is usually necessary to interpret the representation in order to do this. For instance, multidimensional scaling arranges the concepts along dimensions, but does not identify the dimensions. The dimensions are typically interpreted by someone familiar with the domain. Similarly, in cluster analysis solutions, the cluster cutoff point needs to be identified along with category labels. Pathfinder also requires further interpretation because the links in the graph are not labeled or differentiated semantically. The focus of this chapter is on the interpretation of links in Pathfinder networks.

The link-labeling issue concerns using information obtained by Pathfinder in applications such as knowledge engineering. More specifically, how do networks of linked concepts relate to a knowledge base of facts and rules? In this respect, Pathfinder is limited in that the links in the networks are semantically impoverished. That is, the links are weighted with the relatedness estimate value, but they contain no labels[1] indicating the specific nature of the relation. In Pathfinder networks, the links are only indicative of a general association between the concepts represented by the linked nodes.

Without link labels, interpretation and comparison of links must be done cautiously. The fact that the same pair of concepts is linked in novice and expert networks does not necessarily mean that the novices understand the relation at the expert level, but instead, the two links might have quite different meanings. Likewise, in the knowledge engineering application it is not enough to know *that* two items are related, but it is also necessary to know *how* they are related. Currently there is no formal methodology for labeling links in Pathfinder networks. Typically the experimenter or domain expert assigns labels to the various links, but this strategy is subject to the same criticisms relevant to introspection and verbal reports. The remainder of this chapter describes a methodology for labeling the links in Pathfinder networks.

In summary, psychological scaling techniques can be used as knowledge elicitation tools, and thus have applications in education, training, interface design, and development of knowledge-based systems. These techniques also have advantages over traditional knowledge elicitation techniques, particularly in their reliance on data in the form of judgments rather than introspections. On the other hand, there are several aspects of the scaling methodology that pose difficulties for application to knowledge elicitation. This chapter addresses one of these issues that is particularly relevant to Pathfinder network scaling—the interpretation of links.

Interpreting Links in Pathfinder Networks

Attaching meaning to a link might simply involve labeling or naming the link, but the meaning is fully revealed only through all the inferences that are associated with it. For example, the inheritance properties associated with the typical *is-a* link give it meaning. This type of link implies that the subordinate concept inherits properties of the superordinate concept. Consequently, labels are useful to the extent that they suggest specific inferences; however, links associated with different inferences may be given the same name, or links associated with the same inference may be named differently. In short, labels are often ambiguous and imprecise and as a result they do not reliably differentiate among links. Therefore, the methodology discussed here approaches link interpretation first as a classification problem. Classification of links is achieved by asking subjects to sort linked items into groups according to type of relation. These data are then submitted to cluster analysis. The assumption is that the links in each cluster share a set of inferences. After classification, labels can be elicited for a cluster of linked items, instead of a single pair. This classification and labeling methodology should help differentiate links based on meaning that is deeper or more abstract than the meaning expressed in the labeling of individual links.

Chaffin and Herrmann (1984) also used a sorting procedure combined with cluster analysis to empirically derive a taxonomy of relations. Their subjects sorted pairs of items according to their relations. The pairs were selected to correspond to a taxonomy of

[1]"Label" is used in this chapter to refer to the meaning of the link, and "weight" refers to the numerical value associated with the strength of a particular link.

relations which was derived from an extensive literature review of natural language relations. The taxonomy contained five major categories: (1) contrast, (2) similars, (3) class inclusion, (4) case relations, and (5) part-wholes. Results of the sorting and cluster analysis supported this taxonomy. In addition, Chaffin and Herrmann's results suggested three properties or dimensions (contrasting/noncontrasting, pragmatic/logical, inclusion/noninclusion) that distinguished relations and also indicated that subjects could distinguish among many more relations than those typically used in the literature. Chaffin and Herrmann claimed to have captured most of the important families of relations, but they selected pairs of items so that they corresponded with the the taxonomy to be verified. It would be interesting to determine whether the relations identified in empirically generated networks would be accommodated by the taxonomy. Perhaps empirically derived networks could be used to extend the taxonomy. Alternatively, Chaffin and Herrmann's taxonomy could be used to distinguish between relatively common, easy to identify links and links that are either rare or spurious.

The link interpretation methodology discussed here was developed and evaluated using the set of common concepts presented in Table 1. Note that these concepts have a variety of relations. Common concepts were chosen instead of concepts in a specific domain of expertise for several reasons. Most importantly, the methodology could be best developed and evaluated using concepts with relations that are relatively obvious. Furthermore, subjects who know these concepts are common, whereas domain experts are difficult to locate, especially in the numbers necessary to develop a methodology. Finally, the methodology is best illustrated using concepts and relations that are understood by the general public.

Table 1. Common concept set.

blood	mammal	color	red
bird	raking	bats	leaves
flying	rabies	chicken	hair
brushing	barking	tree	injection
milk	dog	animal	egg

A Pathfinder network was generated for the 20 concepts in Table 1, from pairwise relatedness ratings obtained from 30 introductory psychology students at New Mexico State University. Subjects were told to assign ratings on the basis of their first impression of overall relatedness. The relatedness scale ranged from one (slightly related) to five (highly related). In addition, subjects had the option of responding with a "U" indicating that the pair was unrelated. These ratings were transformed to distances by subtraction from 6 ("U" was assigned a distance of 6). In order to account for individual differences using the rating scale, each subject's set of ratings was converted to *z*-scores. The datasets were then averaged across subjects and converted back to the original scale. These average distances were submitted to Pathfinder ($r = \infty$, $q = n{-}1 = 19$). The resulting network contains 20 links and is presented in Figure 1. For each of these links, the original ratings consist of no more than 10 percent judgments of unrelated ("U").

The link interpretation methodology was applied to the links in the Pathfinder network in Figure 1. The network consists of 20 links connecting the 20 concepts. It is a tree with the exception of a single link. The weights for the links *hair-dog* and *animal-blood* are tied at 19. Removal of either of these links would leave a tree. The two phases of the

methodology are link classification, from which clusters of link types are derived, and link labeling, in which the relation represented by each cluster is labeled.

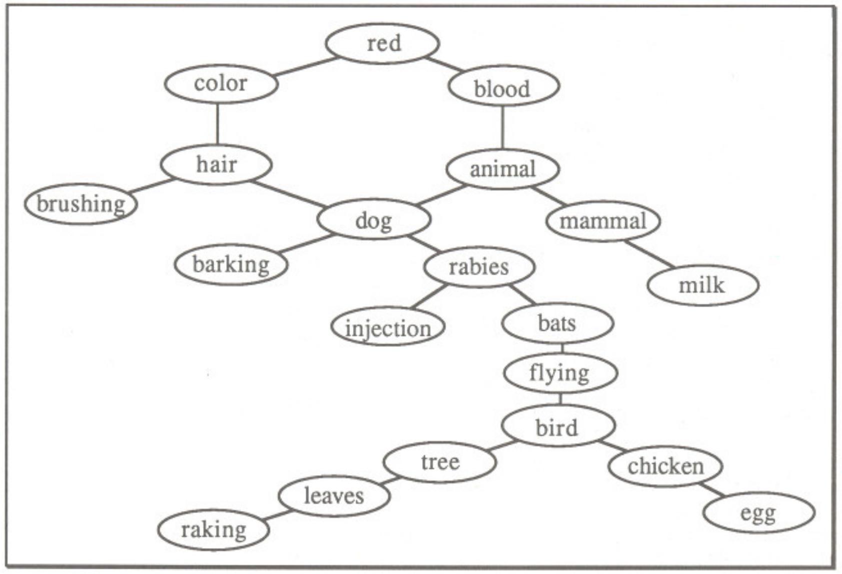

Figure 1. Pathfinder network of 20 common concepts: PFNET(∞, 19).

Link Classification

The purpose of this phase of the methodology was to classify links in the Pathfinder network according to the meaning of the relation associated with each link. Subjects sorted pairs of linked items according to the meanings of the relations, and cluster analysis was performed on the number of subjects who placed each pair in the same pile; frequency of co-occurrence was used as an estimate of distance between pairs of links. As a partial evaluation of the methodology, a control condition was included consisting of pairs that were not linked in the network. If the methodology is successful at classifying meaningful links, then these meaningless control pairs should be excluded from any major categories.

Method

Eight introductory psychology students from New Mexico State University participated in this study. Because pilot studies had indicated that the task (sorting pairs of linked items according to the meaning of the relation) was difficult, subjects who participated were required to correctly answer 15 to 20 analogy problems on a pretest.

Subjects individually sorted 58 index cards based on the meaning of the relation implicit in the pair of items on each card. Each pair was connected by an arrow indicating the direction of the relation (e.g., *robin* → *bird*). The 58 pairs included the 20 links in the network (see Figure 1) presented in both directions (e.g., *robin* → *bird*, *bird* → *robin*), because specific relations are dependent on order (e.g., *robin* is-a *bird,* but it is not the case that *bird* is-a *robin*). The remaining nine pairs (18 in both directions) were not linked in the network, but were included in order to determine how the methodology deals with spurious links. Spurious links can occur in a Pathfinder network because of noise in the relatedness estimates. Also, if the graph is to be completely connected, then some links may appear simply because a node that is not related to any other concept must be connected to the network. Fortunately, Pathfinder has an option that allows for unconnected graphs. Three of the nine unlinked pairs were linked in slightly denser networks (*blood-injection, chicken-animal, mammal-bats*) and three were linked in very dense networks (*bats-chicken, bats-tree, leaves-flying*). The three remaining pairs were not linked (*raking-rabies, brushing-egg, red-flying*) in even the most dense networks.

The sorting procedure was a variation on the method of controlled association discussed by Miyamoto, Oi, Abe, Katsuya, and Nakayama (1986). A card was randomly selected from the deck of 58 by the experimenter and the subject went through the remainder of the deck selecting any cards for which the relation was the same or nearly the same. Subjects were asked to concentrate on the relation for each pair, keeping in mind the direction of the relation. This procedure was repeated with the size of the deck decreasing on each successive sort until all of the cards had been sorted. After the cards had been sorted, subjects were allowed to make any changes to the piles that they wished. They were also permitted to use blank cards to duplicate a pair if they wished to place it in more than one pile.

Results and Discussion

The data for each subject took the form of a vector of 1163 zeros and ones, each corresponding to a different pair of linked items (58 taken two at a time). A one in this vector indicated that the subject placed both pairs of linked items in the same pile and a zero indicated that they were placed in separate piles. The vectors were summed across the eight subjects, resulting in frequencies of co-occurrence for each pair of linked items. These values were converted to distances by subtraction from nine. Thus, a value of one indicated that all eight subjects placed the pair in the same pile and a value of nine indicated that no subjects placed the pair in the same pile. These distances were submitted to a single-link cluster analysis procedure (Johnson, 1967) and the results are presented in Figure 2.

The analysis resulted in 36 clusters that ranged in size from 2 to 12 pairs. The cluster analysis discriminated among several very cohesive groups; 86% of the 36 clusters are based on the classification of five or more of the eight subjects. In most cases, clusters consisted of pairs ordered in one direction only, and each cluster tended to have a mirror-image cluster that contained the same pairs, but in the opposite direction.

It is interesting to note the fate of the nine unlinked pairs. The three pairs that would never be linked either never clustered with any other pair (e.g., *brushing* → *egg*) or clustered late with other unlinked pairs (e.g., *raking* → *rabies* and *flying* → *red*), or their inverses (e.g., *flying* → *red* and *red* → *flying*). Most of the other six unlinked pairs that would be linked in denser networks did cluster in the analysis and some clustered early (e.g., *chicken* → *animal*). It seems that unless the relation was bizarre, subjects were able to find a relation that fit the pair. However, it is reassuring that pairs that would never be linked by Pathfinder did not come together until later in the cluster analysis. On the other hand, some pairs that were linked in the network (e.g., *chicken* → *egg*, *hair* → *color*) also clustered late, suggesting that these links may also be spurious. Alternatively, late clustering might suggest that the link was unique and therefore unrelated to any other link in the graph.

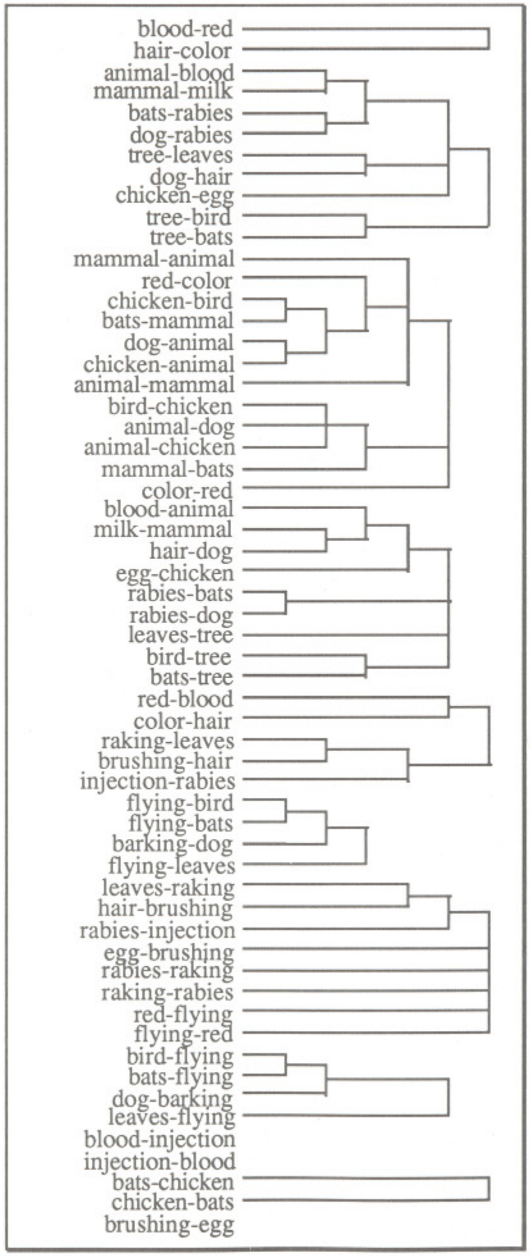

Figure 2. Cluster analysis (single-link method) based on link sorting.

The sorting and cluster analysis procedures generated intuitively reasonable clusters of linked pairs in the sense that there is a cluster that could be labeled *is-a* (*chicken* → *bird*, *bats* → *mammal*) and one that could be labeled *has* (*dog* → *hair*, *tree* → *leaves*). However, these clusters derived from the frequency data were not explicitly labeled by the subjects, but when subjects were asked informally to label the groups that they had sorted or to label individual links, they had difficulties. For instance, some relations were hard to verbalize (e.g., *bird* → *chicken*). It is possible that the clusters resulting from the frequency data would be easier to name than individual links because the clusters of links suggest more alternative labels—one or more of which may be "good." Therefore, in the following study subjects were asked to label the groups of links generated by the cluster analysis.

Link Labeling

The purpose of this phase of the methodology was to determine the nature of the relation represented by each cluster of linked items obtained in the classification phase of the methodology. Subjects were asked to assign a rating to each cluster indicating the goodness or cohesiveness of the cluster and to assign a label to cohesive clusters. The purpose of the ratings was to provide empirical support for the intuitive claim that the clusters of links generated in the classification phase were meaningful. Thus the ratings provided a means of evaluating the methodology. Additionally, the ratings and names should help to differentiate meaningful clusters from those that occur because of noise or because everything must ultimately come together in a cluster. The cluster diagram can be simplified by eliminating clusters that are not meaningful and by merging meaningful clusters that have the same label. Determining the appropriate cutoff level for a cluster is not peculiar to this methodology, but it is an issue relevant to cluster analysis in general.

Method

Twelve introductory psychology students from New Mexico State University participated in this study. As before, these subjects were required to pass an analogy pretest. Stimuli consisted of the 36 clusters of linked items obtained in the classification study, with an additional 15 groups of pairs that did not cluster until the very end. Unclustered groups included (*color* → *hair*, *blood* → *animal*) and (*egg* → *brushing*, *chicken* → *bats*, *tree* → *bats*, *chicken* → *animal*). These groups served as a control condition against which the clustered groups were compared.

Subjects were seated in front of a Sun Microsystems terminal and presented with instructions. They were told that for each trial two or more pairs of items would be displayed on the terminal screen. They were to first determine the relation for each pair in the group and then enter a rating (1, 2, 3, 4, or 5) by pressing the corresponding key on the keyboard. The ratings were to indicate the similarity of the relations present in the group of pairs with 1 indicating that the relations were not very similar and 5 indicating that the relations were very similar. The rating was meant to reflect the goodness or cohesiveness of each cluster of linked items. In addition, subjects were informed that after the rating was entered they might be required to enter a name that best described the relations in that group. Subjects were asked to name a group if they had given it a similarity rating of 3 or greater. The groups of linked items were presented to each subject in random order.

Results and Discussion

Average ratings for clustered and unclustered groups were calculated for each subject. The mean rating across all subjects was 3.42 for the clustered groups and 1.52 for the

unclustered ones. This difference was significant, $t\,(22) = 8.33$, $SE = .229$, $p < .001$, supporting the claim that the clusters produced by the cluster analysis were meaningful. Groups of linked items that were clustered in the analysis were judged more cohesive than groups that did not come together until the end.

The average cohesiveness ratings, along with a cluster label (the name most frequently assigned to the cluster by the 12 subjects), are presented for each cluster in Figure 3. At least half of the subjects agreed on the same or nearly the same label for each of the clusters that had average ratings greater than three. The judgments about what constituted an "identical" label were subjective, but the decisions seemed straightforward and were easily made. The cohesiveness ratings were low for clusters that contained spurious pairs (i.e., those never linked in a Pathfinder network). Also, subjects did not agree on labels for these clusters. These results were expected, given that the pairs in these groups were judged as being unrelated.

The average ratings and the labels provide an indication of clusters that are real, as opposed to those formed as a result of noise, or because all items must eventually come together. For instance, the cluster that contains *blood* → *red* and *hair* → *color* received an average rating of 2.67, suggesting that this cluster is not very meaningful. Also, in several cases there were two or more clusters that were assigned the same label (e.g., the *is-a* clusters). The fact that these clusters were not differentiated by labels may indicate that the smaller clusters should be merged into a single cluster. Thus, low average cluster ratings and redundant labels were used to simplify the cluster solution, clusters were dropped (unclustered) if they had an average rating of less than three, and finally clusters that shared the same label were merged. The resulting simplified solution is presented in Figure 4. These simplification procedures reduced the number of clusters from 36 to 17.

An inspection of Figure 4 reveals four major clusters of relation: *is-a/superordinate*, *has-a/comes from*, *controls/is controlled by*, and *does/is done by*. Each of these four basic relations can be placed in the Chaffin and Herrmann (1984) taxonomy, but some of the specific relations that were nested within one of the four basic types were not as easily classified within their taxonomy. The *is-a/superordinate* (or *category*) relations fit into the CLASS INCLUSION group of the taxonomy. More specifically, these relations fall under the subheading of PERCEPTUAL SUBORDINATES or "objects that are principally characterized by their visible, physical properties" (Chaffin & Herrmann, 1984, p. 135) as opposed to functional properties. The *has a/comes from* relation is similar to the PART/WHOLE relation in the taxonomy. However, the taxonomy subtypes for this category do not seem to match the aspects of *has* that were captured in this study (i.e., *contains, is covered by, carry, is home of*). Both the *controls/is controlled by* and the *does/is done by* relations fall under the taxonomy heading of CASE relations. The *controls* relation can be considered an instantiation of an ACTION-INSTRUMENT relation and the *does* relation is an instantiation of an AGENT-ACTION relation. Thus, although the basic types of relations identified in this study could be placed in the taxonomy, there were some specific distinctions made by subjects within these categories that could be used to extend the taxonomy.

In summary, the results of this study indicated that groups of clustered links were more meaningful than groups of unclustered links. Low ratings and redundant labels were used to eliminate the less meaningful clusters, thus greatly simplifying the solution. Clusters of relations that were revealed in the analysis fit nicely into the Chaffin and Herrmann (1984) taxonomy of relations, although there were some fine distinctions within categories that were not compatible with the taxonomy. The sorting and cluster procedures, in combination with the rating and naming tasks, comprise a valid and useful approach to interpreting links in Pathfinder networks.

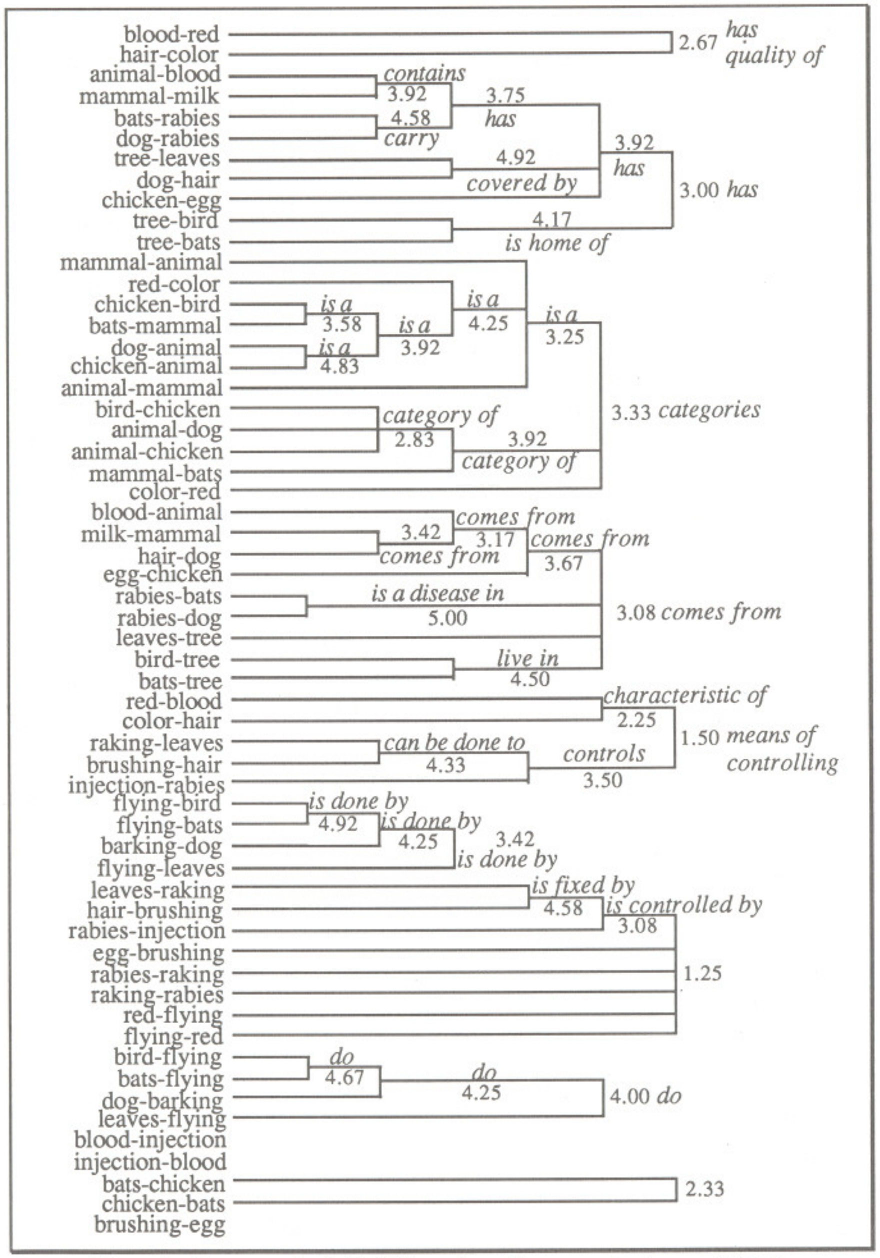

Figure 3. Cluster analysis for link sorting with average ratings and labels.

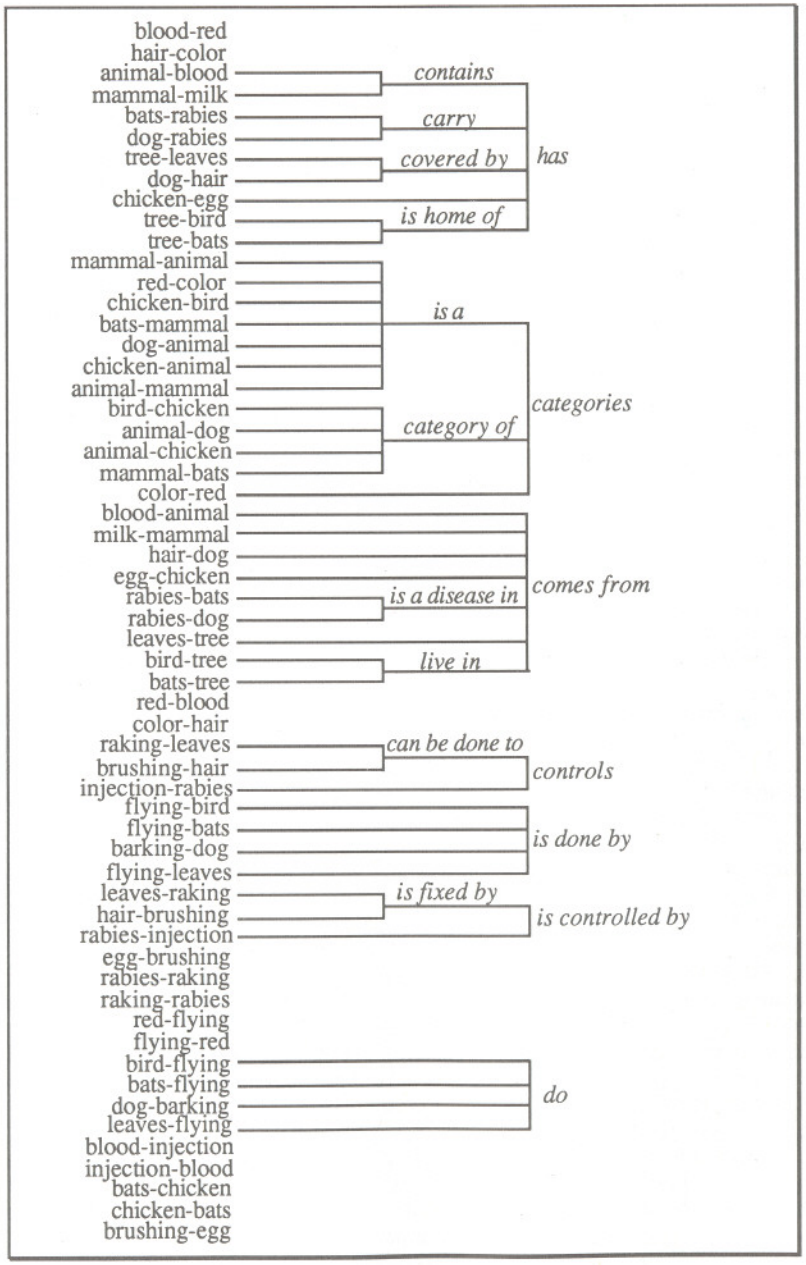

Figure 4. Simplified cluster analysis for link sorting.

Conclusions

Using the methodology described, links were interpreted in a Pathfinder network. This information enhanced the Pathfinder representation, and consequently adds to its effectiveness as a knowledge elicitation tool. Link interpretation can be thought of as part of the third stage of a knowledge elicitation methodology that consists of (1) concept elicitation (see Cooke, 1987), (2) psychological scaling, and (3) interpretation. In this third stage the scaling representations are interpreted by identifying dimensions in multidimensional scaling solutions, identifying cluster cutoffs and labels in cluster analyses, and identifying link labels in Pathfinder networks. The link interpretation methodology presented here consists of two separate phases: (1) link classification using a sorting procedure and cluster analysis, and (2) cluster rating and labeling in which relations are identified for meaningful clusters of links.

One possible criticism of the study described here is that it involved pairs of items with obvious relations. That is, the common concepts used in the study were not representative of concepts in specific domains of expertise, such as physics, computer programming, or flight maneuvers. However, this methodology has also been successfully applied to the domain of computer programming in order to label links in a network of abstract programming terms, such as *algorithm, subroutine*, and *character data* (Cooke, 1987). Cooke found that 24 links relating various pairs of programming terms were each classified as one of six bidirectional relations: (1) *can be done to/is operated on*, (2) *is part of/contains*, (3) *interchangeable/interchangeable*, (4) *is a type of/superset,* (5) *is a type of/could be*, and (6) *can be done to/is used with.* Again, these link types were consistent with the Chaffin and Herrmann (1984) taxonomy; however, there were some subtleties that could be used to expand the taxonomy.

It is informative to compare the results of the link labeling methodology described here with the more straightforward approach of asking experts to simply label links in a network. Cooke and McDonald (1987) attempted to identify link labels for the same programming network that was used by Cooke (1987), but by simply asking experienced programmers to label the individual links. Twenty of the 24 links were labeled identically by three or more of five programmers. However, there were 11 different labels assigned to the same 20 links, as opposed to the 6 labels generated using the classify and label methodology.

The sorting and clustering methodology could be characterized as a generalization mechanism to the extent that it reduces the number of distinct link types. Unfortunately, subtle differences among links could be masked by this methodology. It is important to constrain the number of link types for the sake of efficiency, but it is equally important to capture the distinctions that people are able to make. The rating task used in the second phase of the methodology provides one check on overgeneralization by eliminating clusters of pairs that are given low ratings. On the other hand, the merging of clusters based on similar labels encourages generalization. Fine distinctions between clusters may exist, yet may be hard to verbalize.

The issue of the level of abstraction for link labels is tied to the problem of identifying an appropriate cutoff level in a cluster analysis solution. There are some problems raised by the rating and naming approach used here to determine cutoff. Names are subject to variation and therefore require experimenter judgment in order to determine whether two names are equivalent. Also, the naming task is difficult and subject to the usual problems

with introspection and verbal reports. Ratings might overcome these problems by enabling subjects to express subtle differences among clusters in that they are unable to convey through cluster names. Unfortunately, ratings are also plagued by variance. For instance, does an average rating of 3.2 differ significantly from an average rating of 3.1? Variance should be taken into account in cases in which fine distinctions need to be made using ratings.

In regard to the task itself, it is encouraging that the experienced programmers used by Cooke (1987), unlike the introductory psychology students, had no difficulty with the sorting task and therefore did not require pretesting. This difference might be due to the fact that the programmers had spent considerable time thinking about the programming terms or because they were better educated (they were mostly graduate students). At any rate, this difference suggests that little practice should be required to perform these tasks in a typical knowledge elicitation setting.

Link classification and labeling are important components of link interpretation, but in order to fully interpret links, it is necessary to know what inferences are associated with each one. For example, the relation *is-a* is typically associated with the inference that the subordinate concept can inherit properties of the superordinate concept, but what inferences can be made from the *controls* or *does* relations? The problem could be even greater when concepts and relations are specific to a particular domain. The research presented here does not address this issue, but there are several possible approaches to the elicitation of this information from domain experts. Experts could be asked to associate each obscure labeled relation with an analogous relation for natural language concepts—for example, "this relation is like the relation between *robin* and *bird.*" Then it could be assumed that the inferences that hold for the common relations are similar to the domain-specific relations. However, this strategy still requires knowledge of inferences associated with the common relations so future research is required.

In summary, the use of scaling techniques by themselves enables the researcher to address structural issues regarding memory organization. Without concept elicitation and interpretation methodologies, the information content of the structure is impoverished. Consequently, such enhancements combined with Pathfinder and other scaling techniques comprise a knowledge elicitation methodology that is better suited for this application than is scaling alone. Empirically driven knowledge elicitation is particularly appealing in comparison to the highly subjective and time-consuming protocol analysis and interview techniques. Ongoing research is directed toward expanding this methodology to include other techniques as well as toward validating various aspects of the methodology. In addition, there are plans to automate the methodology in the form of a knowledge acquisition and representation tool kit.