Classification by Semantic Matching

Paul R. Cohen Philip M. Stanhope
Rick Kjeldsen

Experimental Knowledge Systems Laboratory
Department of Computer and Information
University of Massachusetts, Amherst 01003

September 8, 1986

AI Topics: Reasoning, uncertainty, learning, information retrieval
Application: Business

Abstract

We describe experiments with a knowledge system that finds sources of research funding based on semantic matches between research proposals and the interests of funding agencies. The GRANT system relies on domain-specific knowledge about semantic matching and a domain-independent partial matching algorithm to search for funding agencies in a semantic network. The semantic matching algorithm implements a model of uncertainty in classification problem solving. Though GRANT is used daily, we analyze cases of poor performance that illustrate how GRANT-like systems are built and refined. Finally, we summarize an algorithm for learning domain-specific semantic matching knowledge.

This a revised version of an earlier paper with the same name, COINS TR86-13.

This work is supported by NSF Grant IST-8409623 and DARPA-RADC Contract F30602-85-C-0014.

GRANT was designed and developed by many people, including Alvah Davis, David Day, Michael Freed, Mike Greenberg, Rick Kjeldsen, Susan Lander, Cynthia Loiselle, and John Luebke. The GRANT project is an ongoing collaboration with Bruce McCandless, Director of the Office of Research Affairs at the University of Massachusetts, Amherst, and Marg Burggren, also of ORA. We are grateful to all these contributors for their support.

1. Introduction

Classification problem solving involves matching data with pre-established prototypes (Clancey, 1984). Often the match is not exact: it may be partial because some aspects of the prototype lack matches in the data. This paper describes another kind of partial matching and the role it can play in classification problem solving. Semantic matches hold between concepts that are linked in characteristic ways in a semantic network. We have found that the degree of fit between data and a prototype depends on these semantic matches. Moreover, the likelihood of a prototype given the data (in the conditional sense) depends on these matches. In another paper we argued that degrees of belief in classification problem solvers should be interpreted in terms of semantic matches (Cohen et al. 1985). We have developed a program called GRANT that exploits semantic matching to find sources of research funding that are likely to support particular research proposals.

The semantic matching algorithm that underlies GRANT was developed to test a model of uncertainty in classification problem solving. Psychologists have found that we judge the likelihood of a class, given data, by the degree of fit or *representativeness* between the data and the prototype for the class (Tversky and Kahneman, 1982). For example, we judge the likelihood of a disease by the representativeness or degree of fit between the symptoms and a prototype for the disease. In an earlier paper, we described how representativeness might underlie judgments of uncertainty in classification tasks, and showed how a computational model of these judgments provided a semantics for the subjective degrees of belief that are often found in expert systems (Cohen, et al., 1985). This paper describes extensions to the model and reports on the development and testing of GRANT since 1985.

The concept of representativeness is described only informally in the psychology literature. An obvious implementation of representativeness, discussed in Section 2.2, calculates the degree to which an instance is representative of a class by a weighted sum of their common properties. For example, we say a person is likely to be suffering flu if he or she has relatively many flu symptoms (properties) and relatively few non-flu symptoms [1]. This intuitive approach — counting common properties — fails if an instance shares semantically-related, but nonidentical properties with a prototype. Imagine that the prototype for flu includes the property "nausea," but the patient reports "loss of appetite"; or the prototype may include "aching limbs," and the patient reports "pain across the neck and shoulders." In these cases, we are obliged to look at the degree of semantic match between properties before we can calculate the total degree of match between two concepts. Both processes underlie judgment by representativeness, so both are implemented in GRANT.

---

[1] Clearly, the representativeness interpretation of likelihood is not probabilistic in the frequentist or Bayesian senses, since it does not account for the prior probability of flu — only the number of shared and unshared symptoms — in assessing the liklihood of flu. See Tversky and Kahneman (1982) for other examples.

2. GRANT

GRANT is a knowledge system that finds sources of funding for research proposals. The user builds a representation of a research proposal and instructs GRANT to search for funding agencies that are likely to provide support. GRANT first constructs, then ranks, a *candidate list* of agencies. An agency is added to the candidate list if a single topic in its statement of interests is a good semantic match to a topic in the research proposal. Semantic matches exist between topics that are the endpoints of particular *paths* through a semantic network. Agencies on the candidate list are ranked by the number of semantic matches between all the topics in the proposal and all the topics in each agency's statement of interests. The best-ranked agencies are thus those that support the largest number of topics that are semantically related to the proposal.

2.1 Knowledge Representation

GRANT depends on a knowledge base (KB) of research topics and a set of rules for searching it. The latter is described in the next section. The KB is a semantic network of approximately 4500 research topics. Figure 1 shows a fragment of GRANT's knowledge about the heart, cardiovascular illness, and related topics. Nodes in the network are defined in terms of their relationships with others; for example, the heart is something with the *purpose* of circulation, the *setting* of cardiovascular illness, and an *example* of an organ[2]. Appendix 1 lists the most common relations between topics in the GRANT KB.

The GRANT KB acts as a semantic index to funding agencies. Nodes are added to the semantic network as necessary to define the research interests of agencies. An agency is represented as a frame with slots for stated research interests, average award size, citizenship restrictions, geographic preferences, and so on. The **research-interest** slot holds pointers to instances of one or more *activities* that are linked with topics in the KB. GRANT recognizes 10 activities:

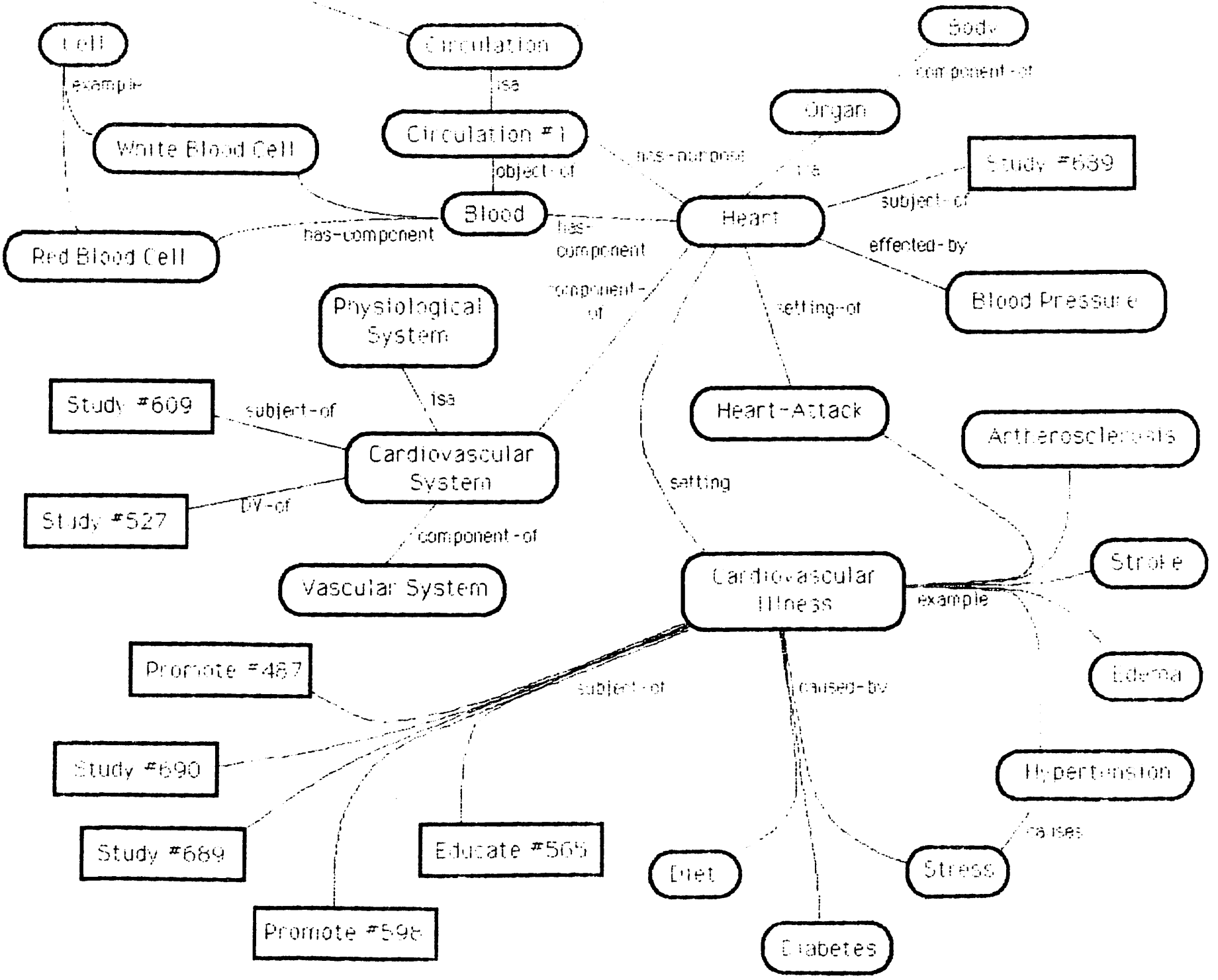
Design	Educate	Improve	Intervene	Manage
Plan	Promote	Protect	Study	Train

For example, the agency associated with *study-689* in Figure 1 is interested in funding studies of cardiovascular illness and the heart. GRANT's KB currently includes the 690 agencies that together provide most of the research monies at the University of Massachusetts.

When GRANT's user creates a research proposal, it is linked into the KB through its research interests just as funding agencies are. The frames that represent agencies and proposals have the same slots, illustrated in Figure 2.

---

[2]And thus, by a plausible inference, a *component-of* the body.


Cell
Circulation
Soda
example
isa
component-of
White Blood Cell
Circulation #1
Organ
has-purpose
Study #689
object-of
is
subject-of
Blood
Heart
Red Blood Cell
has-component
has-component
effected-by
Blood Pressure
component-of
Physiological System
setting-of
Study #609
subject-of
isa
Heart-Attack
Arteriosclerosis
Cardiovascular System
setting
Study #527
DV-of
component-of
Stroke
Vascular System
Cardiovascular Illness
example
Edema
Promote #487
subject-of
caused-by
Study #690
Hypertension
Study #689
Educate #565
raises
Diet
Stress
Promote #598
Diabetes

The ABC Foundation is interested in providing both grants and direct loans in order to help promote education and control of cardio-vascular illness.Funds are available for the management and maintenance of clinics ...

Funding-source*4:

is-a	: funding-source
title	: "ABC Foundation"
descr	: "... promote education and control of cardio-vascular illness ..."
topic	: manage*4

Manage*4:

is-a	: manage
topic-of	: funding-source*4
object	: clinic
subject	: cardio-vascular illness
focus	: cardio-vascular system, blood pressure
purpose	: control educate

Figure 2: The ABC Foundation is represented by the frames FUNDING-SOURCE*4 and MANAGE*4

2.2 Search Algorithms

GRANT finds agencies to fund a research proposal by finding *paths* between the nodes that represent the proposal's research interests and nodes associated with agencies. A *blind search* of the network in Figure 1 would begin, say, at the node *study-527* and extend to its associated node *cardiovascular system*, then to the associations of this node *physiological-system, vascular-system, heart, study-609* and so on, like ripples in a pond. If a node is found that represents a research interest of an agency, then a path has been established between the proposal and that agency. The GRANT KB includes so many agencies and is so highly connected that, on average, blind search finds 245 agencies within 4 links of any proposal. But according to our expert, on average 93.1% of these agencies are *unlikely* to fund the proposal. For GRANT to be useful, this *false-positive* rate must be reduced. One method is to avoid finding unlikely agencies, and the other is to discard them once they are found. These methods are discussed in turn.

Best-first Search. One can avoid finding unlikely agencies by pruning the paths that lead to them during search. Figure 3 shows three kinds of paths. The first is an *atomic match* between the proposal and the agency: the *object* of the proposed *study-418* is *vascular-disease*, which is also the *object* of *study-297*, a research interest of the agency. With few exceptions an atomic match indicates that the agency is likely to fund the proposal.

Since the links in GRANT are directional, and searches proceed from proposals to agencies, the path between the proposal and NHLBI is

$$\text{study} - 418 \xrightarrow{\text{object}} \text{vascular} - \text{disease} \xrightarrow{\text{object-inverse}} \text{study} - 297$$

A *path endorsement* is a generalization of a set of paths, obtained by dropping intermediate nodes and preserving only the relations. The path above is thus an instance of a general *(object, object-inverse)* path endorsement.

The second path in Figure 3 is a *semantic match* between a proposal and an agency. The proposal wants to study hypertension. Whereas an *atomic* match, represented by a path endorsement like *(object, object-inverse)*, guarantees that proposal and agency have a common interest, a semantic match ensures only that the interests of the proposal and agency are somehow related. *The nature of the relationship, represented by a path endorsement, determines the likelihood that the agency will fund the proposal.* For example, when an agency says it funds research on vascular disease, it means that it funds research on many or all kinds of vascular disease, including hypertension. This argument holds for agencies and topics in general: if agencies say they fund X, they are likely to fund instances of X. By this reasoning, if we begin a search at a proposal and follow a *(object, isa, object-inverse)* path to an agency, then the agency is likely to fund the proposal. Any path that is an instance of the *(object, isa, object-inverse)* path endorsement is apt to find a likely agency.

Paths Between Proposals and Agencies

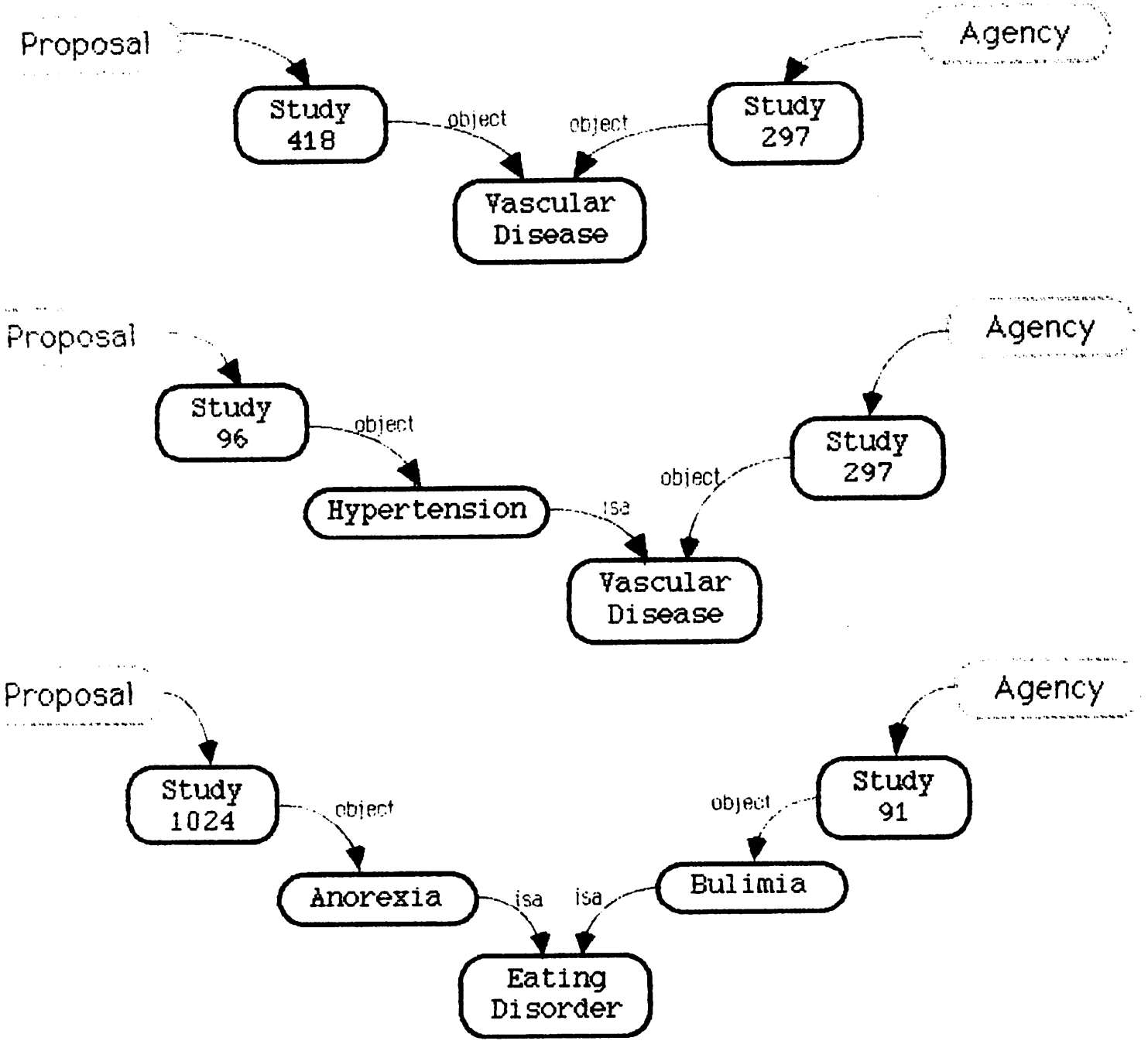


Figure 3

Just as path endorsements mark likely paths to agencies, so they mark paths to be avoided. The third path in Figure 3 is an example. The research topic of the proposal is *anorexia* and that of the agency is *bulimia*. Now bulimia is an instance of an *eating-disorder* and when an agency says it will fund the study of an instance of X it usually means that it will not fund the study of *other* instances of X. This agency is unlikely to fund the study of other eating disorders such as anorexia. In general, if a path between a proposal and an agency is an instance of the path endorsement *(object, isa, isa-inverse, object-inverse)*, then the agency is unlikely to fund the proposal and the path should be avoided.

Path endorsements thus constrain the search for agencies in GRANT. Appendix 2 lists some of GRANT's path endorsements. The complete set of path endorsements is still only a fraction of the combinatorially possible path endorsements. Any path that has not been classified as likely or unlikely is denoted *unknown*. Best-first search in GRANT proceeds as follows:

> Assume the program starts at a proposal and follows link l_i to node n_i: $\langle l_i n_i \rangle$. If a continuation of this path along link l_j to node n_j results in a path endorsement (l_i,l_j) that GRANT recognizes as poor, then n_j is pruned from the list of nodes that GRANT tries to expand. If (l_i,l_j) is a good path endorsement, then GRANT will give n_j priority to be expanded before any node n_k found by an *unknown* path $\langle l_i n_i l_k n_k \rangle$. Search from any path longer than 4 links is terminated.

Ranking Agencies by Partial Matching. The result of best-first search is a candidate list of agencies. Each is known to have a single research interest that atomically or semantically matches one research interest of the proposal. To the extent that the proposal and an agency share several common research interests, the agency is more likely to fund the proposal. Thus, GRANT ranks the candidate list of agencies by the degree of overlap between the research interests of the proposal and each agency. This is done by a partial matching function based on both atomic and semantic matching. Hayes-Roth (1978), Tversky (1977), and others measure the degree of overlap between sets in terms of set intersection and symmetric difference; for example, Tversky's *contrast model* (1977) calculates overlap this way:

$$S(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A).$$

The function f returns the cardinality of the set to which it is applied. If A and B are frames, then $f(\text{A} \cap \text{B})$ is the number of slot-value pairs shared by A and B, and $f(\text{A} - \text{B})$ is the number of slot-value pairs in A not shared by B. The parameters θ, α, and β are set empirically; in GRANT each is 1.0. If A and B are frames representing the research interests of a proposal and an agency, respectively, then $S(a,b)$ measures the number of research topics they have in common relative to those they do not share. Agencies for which $S(a,b)$ is higher are more likely to fund the proposal.

In GRANT, (A ∩ B) includes both atomic and semantic matches. If a path between A and B contains a single node (e.g., the first case in Fig. 3), or if the path is an instance of a likely path endorsement (e.g., the second case in Fig. 3), then $f(A \cap B)$ is incremented. Unlikely path endorsements, such as the third case in Figure 3, and unknown paths do not contribute to $f(A \cap B)$. The quantities $f(A - B)$ and $f(B - A)$ are increased when research topics in the proposal lack an atomic or semantic match to the agency, and vice versa.

In summary, GRANT searches for agencies in two stages. First it constructs a candidate list of agencies by best-first search in a semantic network of research topics, then it ranks the agencies on the list by their degree of overlap with the research proposal.

3. Analysis of GRANT Performance

GRANT's performance has been tested at all stages of its development. The basic method is to run samples of proposals and compare the agencies selected by GRANT with the choices of our expert. Sample sizes have ranged between 20 and 30 proposals. We compute many statistics for each search from a proposal, but two are broad indicators of GRANT's performance:

$$\text{hit-rate} = \frac{\text{number of agencies judged good by GRANT and by the expert}}{\text{number of agencies judged good by the expert}}$$

$$\text{false-positive rate} = \frac{\text{number of agencies judged good by GRANT and bad by the expert}}{\text{number of agencies judged good by GRANT}}$$

We average these statistics over the searches from the individual proposals in a sample.

When we first tested GRANT (Cohen et al., 1985) its knowledge base contained approximately 700 nodes and 50 agencies. We contrasted blind and best-first search as follows: for each of 23 proposals the system searched blindly for agencies until it reached a predetermined stopping criterion. On average, blind search found 15.1 agencies per proposal. We gave our expert the list of agencies found for each proposal by blind search and asked him to rank each agency as likely or unlikely to fund the proposal. On average, only 2 agencies per proposal were considered likely; that is, the false-positive rate for blind search was (15.1 - 2)/15.1 = 86%. In contrast, best-first or path endorsement constrained search found on average just 2.78 agencies per proposal, of which 1.48 were judged likely to fund the proposal. The false-positive rate was 32%, a big improvement over blind search. The downside was a hit rate of 80%, indicating that GRANT had pruned away one likely agency in five. We have tested all subsequent versions of GRANT this same way, using blind search to find candidate agencies and an expert to rank them, then comparing best-first search with the expert's rankings. Table 1 shows best-first

search statistics for several versions of GRANT. Blind search statistics are not represented; in all tests blind search had a false positive rate greater than 80%, and as the knowledge base increased in size this figure increased dramatically.

Grant, Spring 85 (700 nodes, 50 agencies)	
Hit Rate	80%
False Positive Rate	32%
Grant, Fall 85 (2,000 nodes, 200 agencies)	
Hit Rate	80%
False Positive Rate	26%
Contrast Model	
Hit Rate	76%
False Positive Rate	22%
Grant, Winter 86 (4,500 nodes, 700 agencies)	
Hit Rate	98%
False Positive Rate	61%
Contrast Model	
Hit Rate	96.1%
False Positive Rate	57%
Grant, Winter 86 (4,500 nodes, 700 agencies)	
Modified Path Endorsements	
Hit Rate	96.3%
False Positive Rate	55.8%
Contrast	
Hit Rate	96.4%
False Positive Rate	53.4%

Table I.

The differences between GRANT today and the version we tested in Spring, 1985 are its size and the incorporation of Tversky's contrast model for summing the total degree of overlap between proposals and agencies. The false positive rate of the early version, 32%, decreased during the subsequent months as the knowledge base increased to 2000 nodes with 200 agencies. At that time we introduced the contrast model, described above, and realized a further small decrease in the false positive rate, which was offset by a decrease in the hit rate. In the last two months we have again more than doubled the size of the knowledge base and more than tripled the number of agencies from the Fall, 1985 level. As a result, performance has decreased substantially. The hit rate of best-first search is 98%, but the false positive rate is 61%: the

system finds virtually all the agencies it should, but nearly two-thirds of the agencies it finds are not likely to fund the proposal.

Why did the increase from Spring, 1985 to Fall, 1985 not decrease GRANT's performance, while the latter one did? Many factors are involved. First, the density of agencies is increasing. In the early version, 700 nodes supported 50 agencies – a ratio of 14:1. In Fall, 1985, the ratio was 10:1. The most recent knowledge base has a ratio of 6.4:1. It is much easier to find many agencies close to a proposal in GRANT's semantic net than it was in the past. Indeed, we have evidence to suggest that as the density of the knowledge base increases, the hit rate goes up and the false positive rate down: An intermediate version of the Winter, 1986 knowledge base included approximately 600 *orphans*, nodes used to define another node but disconnected from all other nodes. In this version, the density of nodes per agency was 5.8:1. There were too many agencies and too few associative paths to differentiate good agencies from bad ones.

A second contributor to the high false positive rate in the Winter, 1986 version is the kinds of agencies being represented. Roughly 200 of the new agencies were for the arts and humanities. Their descriptions of research interests were fairly broad and gave little basis for differentiation. Consequently, when GRANT searches in that part of the knowledge base, its false positive rate increases dramatically. A related problem is that in the most recent version of GRANT, new agencies were not represented in as much detail as old ones. Necessarily, this meant viable distinctions between agencies were lost.

The relations we use to represent agencies have not changed appreciably since the early version of GRANT, but the number of things they are required to represent is greatly increased. Combined with the fact that GRANT was developed to represent "hard science" topics and now includes arts, humanities, and social sciences, this suggests that the relations must be augmented and perhaps reworked. This also requires reworking the set of path endorsements. In fact, an experimental set of path endorsements gave somewhat better performance for the Winter, 1986 version. The hit rate remained very high but the false positive rate dropped to 55.8%.

The partial matching algorithm, based on Tversky's contrast model, was not as effective as we had hoped in pruning agencies based on the total degree of overlap between proposals and agencies. In general, the false positive rate can be reduced but not without a corresponding reduction in the hit rate. The algorithm contributes little because in most cases, a proposal shares only one research topic with an agency. Since this overlap is usually found by semantic matching, best-first search will continue to be the heart of GRANT's problem-solving method, and path endorsements will receive more attention than tuning the partial matching algorithm. The next section describes an algorithm for learning path endorsements.

4. In Prospect: Learning Path Endorsements

The likelihood that an agency will fund a proposal depends on the path endorsement that characterizes the semantic match between them. Path endorsements as discussed above either

support the proposition that the agency will fund the proposal, or detract from it, or their support for the proposition is unknown. In practice, GRANT's path endorsements are empirically ranked into six classes: *very likely, likely, maybe, unknown,* and *trash*. Detracting path endorsements belong to the class *trash*. The class *very likely* is reserved for atomic matches. Thus, semantic matches that support the proposition that an agency will fund the proposal are differentiated only by the classes *likely* and *maybe*.

We have developed an algorithm to assign a continuous weight to path endorsements, based on whether they find likely agencies or false positives. The algorithm learns from examples presented by a human tutor. Each example is a pair of nodes for which the tutor expects GRANT to-find a semantic match. The algorithm generates a set of paths between these nodes from GRANT's knowledge base, and adjusts the weight of each path to favor short paths over long ones. After many iterations, short paths that are commonly found between training examples have high weights, relative to other paths.

The algorithm has been tested on small samples of examples and it has not yet been integrated with GRANT. In prospect, however, its principle advantage is that it learns the *empirical* worth of path endorsements, in contrast to our a priori efforts to categorize path endorsements as *likely* or *maybe*. Kjeldsen (1986) describes the algorithm in detail.

Two other extensions to GRANT should be mentioned. First, we have developed an "empty" version and will be experimenting with semantic matching in other domains. Second, we are generalizing the inference rule that underlies GRANT — "if an agency is interested in X then they will be interested in Y = R(X)" — to a logic for plausible inference in associative knowledge bases.

5. Appendix 1

Relations for funding agencies:

1. The TITLE slot should contain a text string with full title that will include the Parent Agency, Department, and Program Name.
2. The UNIQUE-ID slot should contain a text string that is the unique number assigned by the Catalogue of Federal Domestic Assistance (CFDA).
3. The FUNDING-TYPE slot should contain the type of funding that is available, e.g., *project-grant, large-grant, small-grant, direct-loan, fellowship,* or *scholarship.*
4. The CONTACT slot should contain the name, address, and phone number of the person to contact for more information and applications.
5. The DEADLINES slot should contain the application and renewal deadlines for the program.
6. The DESCRIPTION slot should contain the abstract that is provided by the agency and describes their interests and motivations.

7. The TOPIC slot should contain one or more instances of the STUDY, MANAGE, EDUCATE, or ENGINEER frames.
8. The PURPOSE slot is optional for the top-level of a *funding-source* frame since it might be present in one of the values for the TOPIC slot.

Relations for defining research interests:

1. The OBJECT slot contain the person, place, process, or thing that is being studied.
2. The SUBJECT slot contain the particular filed of study that is to be applied to the *object*.
3. The FOCUS slot should contain the particular aspect of the *subject* that is being considered.
4. The DV slot should contain the *object* that is being studied.
5. The IV slot should contain the variables that whose effect upon the dependent variable are being studied.
6. The RV slot should contain one or more variables that are being studied.
7. The PURPOSE slot should contain the overall goal of the funding source.
8. The WHO-FOR slot should contain an instance of a social-group that will benefit from the proposed research and funding.
9. The SETTING slot should contain the place in which the *object* will be studied.
10. The LOCATION slot should contain a geographical place to which funding is restricted.

Relations for organizing knowledge in GRANT's knowledge base:

1. The CAUSES slot should contain a concept that has a causal association with the node.
2. The EFFECTS slot is used to represent relationships that are not necessarily causal but nonetheless present.
3. The HAS-COMPONENT slot should contain those things that make up the node. For example, one could say that a earthquake has-component shock-wave.
4. The HAS-MECHANISM slot is used to represent those processes that a concept might have. For example a seismology has-mechanism seismometer.
5. The HAS-PURPOSE slot is used to hold an instance of an action. For example, a seismometer has-purpose measure, with the object of the measure being shock-wave.

6. Appendix 2

Path Endorsements for the Knowledge Base in the rule set that is used in a bottom-up data driven search from proposal to funding source. Many of these traversal rules are effectively used to prune the number of potential nodes to expand. A SUCCESS-NODE is any node that can be found as a value for wither the TOPIC or PURPOSE slot of a fundinf-source.

- The class SELF has 1 traversal rule
  - Self - basically an identity rule for paths of length 0
- The class VERY-LIKELY includes 7 path endorsements, all atomic matches. For example,
  - X→ subject→ Y→ subject-of→ SUCCESS-NODE
  - X→ focus→ Y→ focus-of→ SUCCESS-NODE
- The class LIKELY has over 50 path endorsements representing semantic matches between a proposal and an agency that is likely to fund it. For example,
  - X→ subject→ Y→ isa→ Z→ subject-of→ SUCCESS-NODE
  - X→ subject→ Y→ component-of→ Z→ focus-of→ SUCCESS-NODE
  - X→ done-by→ Y→ does→ object-of→ SUCCESS-NODE
- The class MAYBE has 18 path endorsements. These represent semantic matches between a proposal and funding agencies that are somewhat less likely to fund the research, for example:
  - X→ focus→ Y → subject-of→ Z→ subject-of→ SUCCESS-NODE
  - X→ object→ Y→ focus-of→ Z→ subject-of→ SUCCESS-NODE
  - X→ object→ Y→ object-of→ Z→ focus-of→ SUCCESS-NODE
- The class UNKNOWN accepts any path less than 6 links long
- The class of UNUSABLE paths prunes GRANT's search. Among these paths are any that contain a node with an extremely high branching factor (e.g., science, education). Specific pathways of the kind listed above include
  - STEP*→ isa→ example→ Y
  - STEP*→ subfield-of→ has-subfield→ Y
  - NOT(new-investigator)→ STEP*→ new-investigator
  - NOT(minority-student)→ STEP*→ minority-student
  - X→ object→ Y→ subject-of→ Z→ focus-of→ SUCCESS-NODE
  - X→ rv→ Y→ dv-of→ SUCCESS-NODE
  - X→ subject→ Y→ isa→ Z→ dv-of→ SUCCESS-NODE

REFERENCES

[1] Cohen, P.R., Davis, A., Day, D., Greenberg, M., Kjeldsen, R., Lander, S., and Loiselle, C. 1985. Representativeness and uncertainty in classification systems. *AI Magazine*, Fall 1985.

[2] Clancey, W.S., 1984. Classification problem solving. *Proceedings of the AAAI,* p.49.

[3] Hayes-Roth, F., 1978. The role of partial and best matches in knowledge systems. *Pattern Directed Inference Systems*, Waterman, D., Hayes-Roth, D., and Lenat, D. (Eds). Academic Press.

[4] Kjeldsen, Rick, 1986. Learning traversal rules for semantic nets. *EKSL Working Paper.*

[5] Tversky, A., 1977. Features of Similarity. *Psychological Review.*

[6] Kahneman, D. and Tversky, A. 1982. *Subjective probability: a judgment of representativeness, Judgment Under Uncertainty: Heuristics and Biases,* Kahneman, D., Slovic, P., and Tversky, A. (eds.), Cambridge University Press, 1982.