Information Retrieval by Constrained Spreading Activation in Semantic Networks

Paul R. Cohen and Rick Kjeldsen
Department of Computer and Information Science
Lederle Graduate Research Center
University of Massachusetts, Amherst 01003

COINS Technical Report 87-66
EKSL Report 87-14

Acknowledgements

This research is funded by DARPA/RADC Contract No. F30602-85-C-0014, by National Science Foundation grant IST 8409623, and by Office of Naval Research on University Research Initiative grant N0001486K0764.. GRANT was built in collaboration with the Office of Research Affairs at the University of Massachusetts. The director of the office, Bruce McCandless, served as our expert. Over the last two years, many undergraduates, graduate students, and staff have worked on the GRANT project. In alphabetical order, these are: Marg Burggren, Alvah Davis, David Day, Jeff Delisio, Mike Freed, Mike Greenberg, Tom Gruber, Bruce McCandless, Sue Lander, Cynthia Loiselle, and Phil Stanhope.

Abstract

GRANT is an expert system for finding sources of funding given research proposals. Its search method – constrained spreading activation – makes inferences about the goals of the user, and thus finds information that the user did not explicitly request but that is likely to be useful. The architecture of GRANT and the implementation of constrained spreading activation are described, and GRANT's performance is evaluated.

1 Introduction

How does one match up several billion dollars of research funding with several thousand researchers? The researchers want the money, but they are not always sure where to apply for it. Researchers and funding agencies are participants in an intellectual matchmaking game, a process of identifying one's partner based on his or her research interests. It is such a difficult task that many research institutions employ professional funding advisors to tell faculty where to send their research proposals. We have built an expert system called GRANT that simulates the performance of a funding advisor. The technology for this project is suited to other information-retrieval tasks in which an individual relies on associative memory and a semantic representation of a request to find resources.

When funding advisors rely on memory they can offer advice relatively quickly. The same expert who listens to a research idea and in five minutes suggests several appropriate funding agencies may require 2 or 3 hours to input the idea in terms of keywords, run a database program, and sort through the results – most of which will be inappropriate. But while an expert can be much faster than a clumsy database program, he or she remembers much less. The Federal Government alone offers hundreds of funding programs, and the opportunities in the private sector fill volumes. Both are continually changing. Human memory, though fast for familiar items, is both ponderous and unreliable for unfamiliar ones [1].

From the standpoint of matching researchers with funding agencies, the major advantage of human memory is that it encodes the *meaning* of concepts by their associations with other concepts [2,3]. Since agencies are willing to fund not only the research they describe in their statements of interest but also *related* research, semantic memory is able to find agencies that keyword search methods would miss.

The GRANT system relies on a semantic memory of research concepts to help it find agencies that are likely to fund proposals. If it cannot find an agency to support research on a specific topic, then it finds agencies that support work on related topics. GRANT finds these agencies as quickly as a human funding advisor relying on his or her memory, but it keeps many more agencies permanently accessible. Its accuracy appears much higher than keyword search algorithms often used for this task.

In outline, this paper surveys the relevant psychological and AI literature on semantic memory, then in Section 4 describes the architecture of GRANT. Section 5 discusses experiments with

GRANT. We measure its performance in several ways and show how it can be improved. Section 6 is a discussion of the feasibility of other GRANT-like systems.

2 Relevant Literature

Philosophers and psychologists have long recognized that human memory is associative. The first computer model of associative memory was Quillian's semantic memory system [4]. It was based on a *semantic network* of concepts that were defined by their relationships with other concepts in the network. For example, the meaning of "expert system" is defined by its relations to concepts such as *expert, expertise, computer program* as well as more general concepts such as *problem solving, computer, knowledge* and so on. Quillian's system allowed for arbitrary relations between concepts, so long as they could be represented as binary predicates; for example, two arguments can be related by the binary predicate KIND-OF, as in KIND-OF(expertise,knowledge), but Quillian's semantic net had no representation for the ternary predicate required to express, say, "expertise is a specialized kind of knowledge" — KIND-OF(expertise,knowledge,specialized).

Many other models of semantic memory have been developed since Quillian's, and are discussed from the perspective of cognitive psychology in Anderson and Bower [2] and Anderson [5] and as AI knowledge representations in Cohen and Feigenbaum [6], Findler [7], and Brachman and Levesque [8]. A closely-related associative knowledge representation is *frames* [9, 8]. Frames have named *slots* that contain pointers to other frames; for example, the KIND-OF slot for the expertise frame contains a pointer to the knowledge frame.

While the semantic network idea does not restrict the relationships that can hold between concepts, one can define classes of concepts by the kinds of relationships they are permitted with other concepts. This idea proved very powerful in natural language processing, where it is referred to as *case semantics* [10] or *semantic primitives*. Verbs are easily defined this way; for example, many actions are defined by their relationships to an *actor*, an *object*, and an *instrument*. The meaning of a verb such as "open" in the sentence, "John opened the door with a key" is derived from the words related to the verb by its case relations: John is related to the verb by the *actor* relation, the door by *object*, and the key by *instrument*. Other work on case relations resulted in Schank's *conceptual dependency* representation of verb meanings, a system on which several generations of natural language understanding programs have been constructed [11]. Verbs are not the only objects that can be defined by case semantics: in GRANT we define all research topics and activities this way.

Previous work on semantic memory and case semantics provides the basis for the GRANT system. The key assumption of the system is that if no agency can be found to support research on a specific topic, then one might be found to support work on a semantically-related topic, and the likelihood of support depends on the relationship between the topics. Imagine a researcher is interested in dandelions, but GRANT cannot find any funding agencies in its memory that mention dandelions. GRANT may, however, find an agency to fund research on a *related* topic, say plants. The likelihood that the agency will fund work on dandelions depends, in part, on the nature of the relationship between dandelions and plants. Once GRANT has found an agency to fund a given topic or a related one, it then calculates how well all aspects of the agency description fit those of the research proposal. These two phases, finding an agency and computing overall match, are the main components of GRANT. Since the novel aspect of GRANT's architecture is how it finds agencies, that will be the focus of this paper.

3 GRANT Architecture

GRANT's architecture includes a large semantic network of research topics, a set of funding agencies, a user interface for specifying proposals and presenting results, and a control structure for finding agencies given a proposal. These are illustrated in Figure 1. The semantic network is in effect an *index* to the agencies, since each agency is linked into the network at those nodes of the network that represent its research interests. Proposals, once elicited from researchers, are linked into the network in the same way.

In overview, the system works by *spreading activation* from a proposal through the network until one or more agencies are activated. First, the research topics in a proposal are activated, followed by all topics that are directly related (i.e., one link away) in the network, followed by *their* related topics, and so on, as activation spreads across relations in the network like ripples in a pond. Ordinary spreading activation can quickly touch every topic in a network, which means that it can find pathways from any research proposal to any agency description. Since most agencies found this way would not fund a given proposal, GRANT uses a modified search algorithm, called *constrained spreading activation*. This algorithm is constrained by a set of rules to favor particular pathways through the network, and terminate search along other pathways. The rules lead GRANT to agencies that cannot be found by keyword search, and allow it to avoid the numerous, irrelevant agencies that are found by ordinary spreading activation.

3.1 GRANT's Knowledge Base

GRANT's semantic network of research topics was constructed specifically to represent the interests of funding agencies. Currently, the network contains over 4500 nodes that represent the research interests of 700 funding agencies. Nodes are added to the network by linking them to other nodes with one or more of 48 distinct relations. For example, we can define a *heart disease* node by linking it to *heart* with the *has-setting* relation and to the *disease* node with the *isa* relation (see Fig. 2). All relations are directional and have inverses (not shown in Fig. 2); for example, the inverse of *has-setting* is *setting-of* and the inverse of *isa* is *has-instance*. GRANT adds inverse links between nodes automatically.

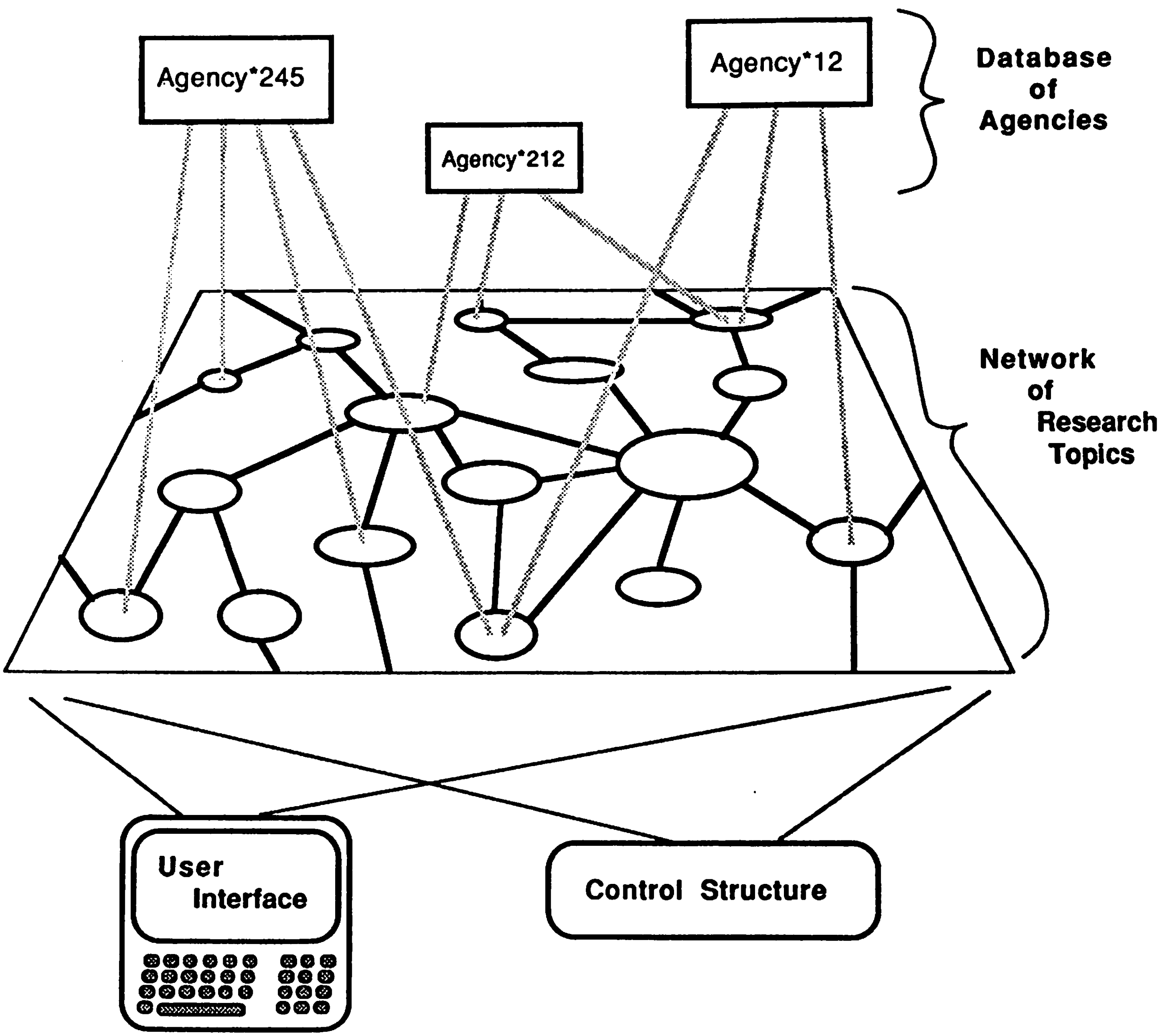

Figure 1
Overview of the GRANT system

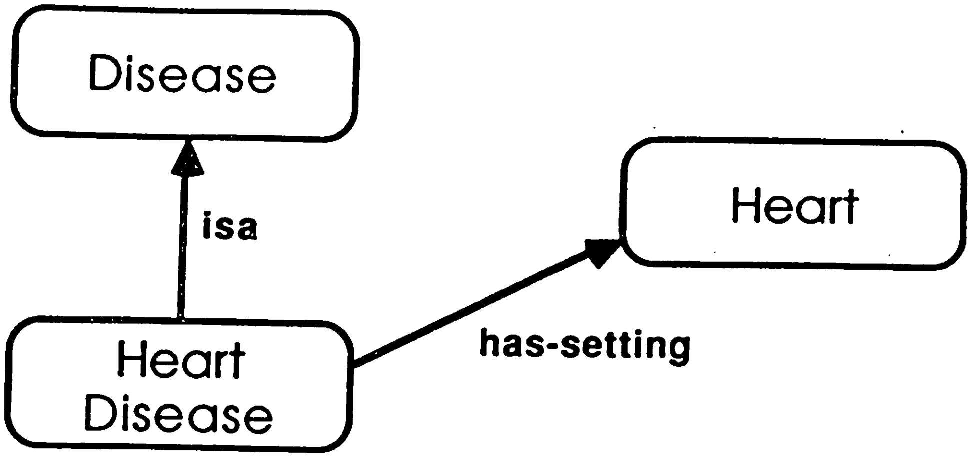

Figure 2

Sometimes the nodes that would define a new node do not exist in the network and must themselves be defined. For example, to add *mitral valve prolapse* to Figure 2 we need to say it is a *heart disease* but we also need to say its setting is the *mitral valve*, which is part of the *heart*. Figure 3 shows how adding *mitral valve prolapse* also involves adding *mitral valve*. Nodes are added only as needed to define research topics; GRANT's knowledge base is not an encyclopedia of science, medicine, and the arts, but is a highly cross-referenced index of research topics, represented from the perspective of funding agencies[1].

The relationships that define concepts are similarly tuned to GRANT's domain; for example, one field of research is a *subfield* of another, a phenomenon is an *effect* of a process, something is a *dependent variable* of a study, and so on.

All nodes in the network are represented as frames. Slots represent links or relations with other nodes. Some nodes represent funding agencies and the research topics they support. Agencies have slots for level of funding, citizenship restrictions, and so on, as well as links to their research interests (Fig. 4).

The frames that describe research interests, both for agencies and proposals, are created by classifying the goal(s) of research into one or more of ten classes:

[1]See Lenat, Prakash, and Shepard, [12], for a fascinating description of an encyclopedic knowledge base.

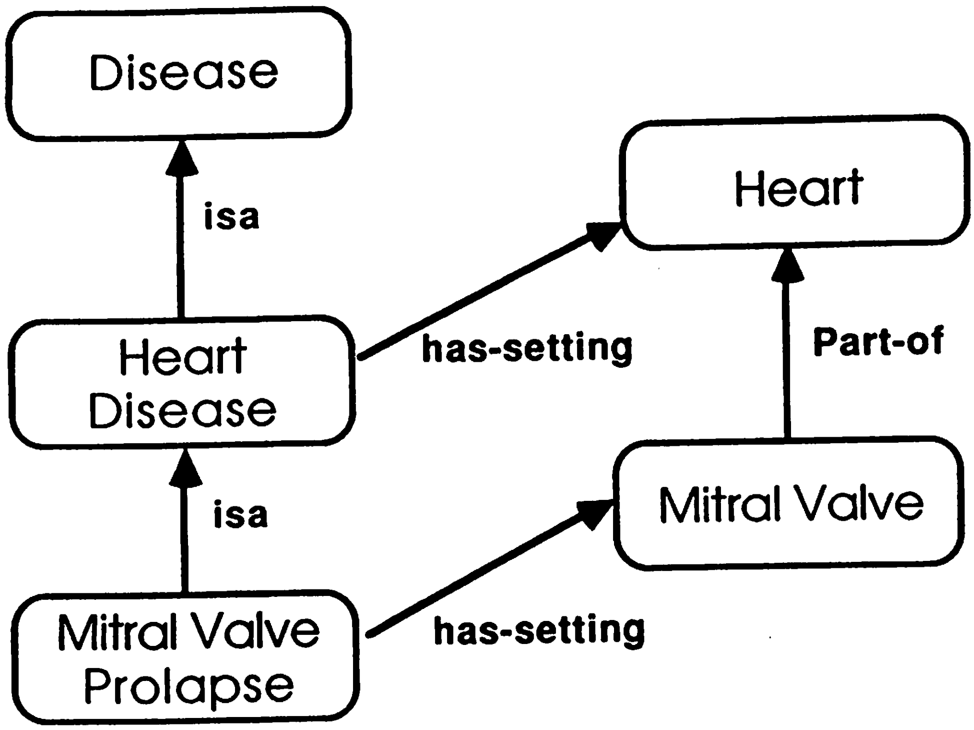
Disease
isa
Heart
Heart Disease
has-setting
Part-of
isa
Mitral Valve
Mitral Valve Prolapse
has-setting

Figure 3

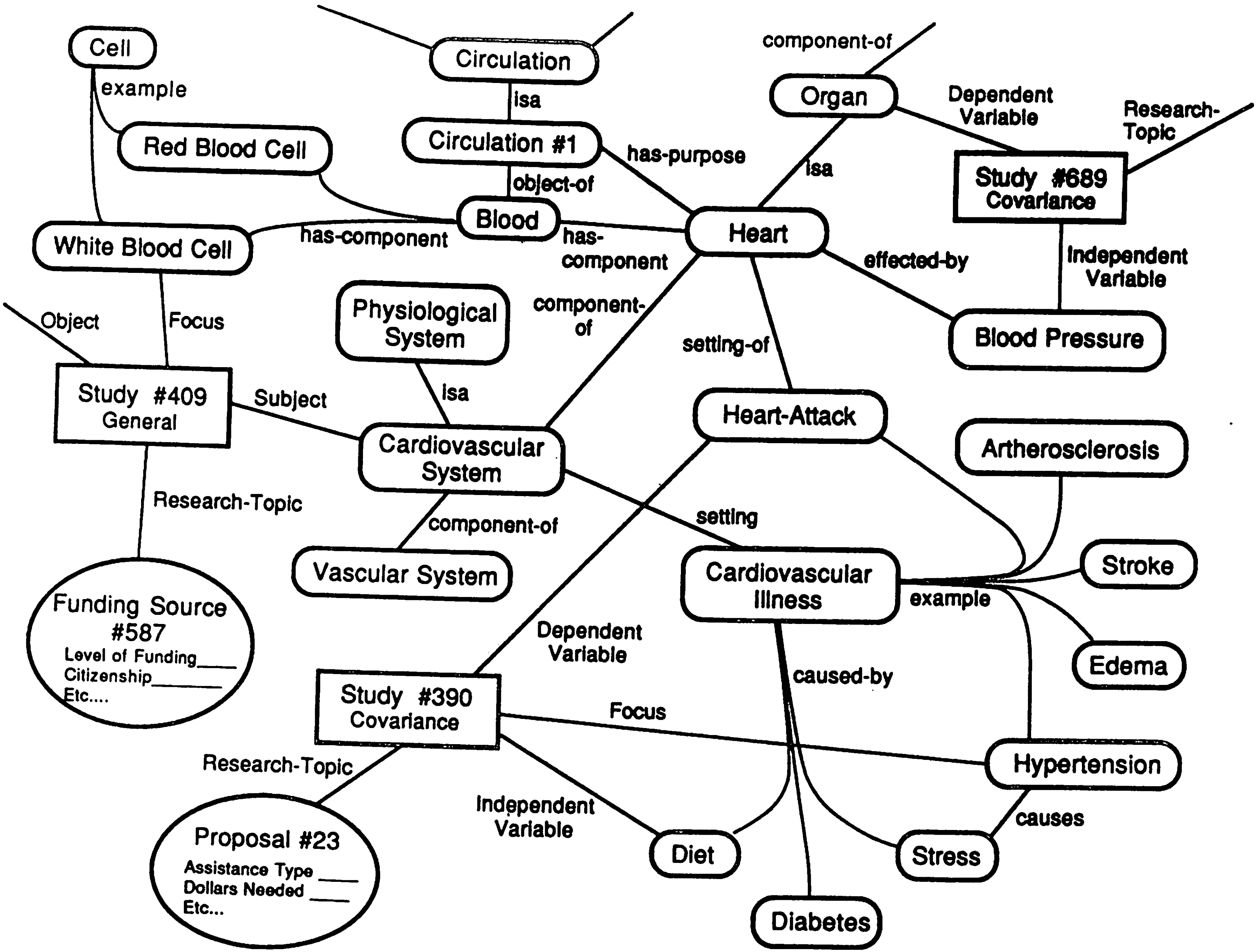

Figure 4: A portion of the GRANT knowledge base

Design	Educate	Improve	Intervene	Manage
Supply	Promote	Protect	Study	Train

Each class is represented by a case frame with a set of obligatory and optional slots. For example, a *study* frame represents exploration of some topic, and so has *subject* and *object* slots that represent the topic, and a *focus* slot that describes which aspect of the topic will be studied.

3.2 Constrained Spreading Activation

During a run of the GRANT system, activation spreads from the topics stated in a proposal, through the network, to agencies via their stated interests. Some constraint on the spreading activation is required, otherwise all agencies linked into the network would eventually be activated. Three kinds of constraints have been imposed. The *distance* constraint says that activation should cease at a distance of 4 links (i.e., 5 nodes) from any research topic mentioned in the proposal. This is an extremely weak constraint. A second *fan-out* constraint says that activation should cease at nodes that have very high connectivity or fan-out. Examples of these nodes include *science, disease*, and *person*. Two research topics may be semantically related by both being sciences, but this does not guarantee that an agency will fund one if it will fund the other.

The third kind of constraint captures the idea that the likelihood of an agency funding a proposal depends on the nature of the relationships between the agency's interests and those of the researcher. Formally, GRANT is an inference system that applies repeatedly a single inference schema:

$$\text{request-funds-for-topic}(x) \text{ and } \mathbf{R}(x,y) \rightarrow \text{request-funds-for-topic}(y) \quad (1)$$

for "paths" **R**. (Note that **R** can be thought of as a single link, such as ISA, or more generally as a path of n links connecting n + 1 nodes, as described below.) If one would ask an agency to fund research on dandelions, request-funds-for-topic(dandelions), and dandelions are a kind of plant, then one stands a reasonable chance of obtaining funding from an agency that supports research on plants.

$$\text{request-funds-for-topic(dandelions) and ISA(dandelions,plants)} \rightarrow$$

$$\text{request-funds-for-topic(plants)} \quad (2)$$

If we replace the constants with variables, leaving just the relationship ISA, we get a rule of inference of the form described in (1) that we call a *path endorsement*:

request-funds-for-topic(x) and ISA(x,y) →

request-funds-for-topic(y) (3)

Associated with each path endorsement is a score denoting how likely it is that an agency would fund research on topic x if they would fund research on topic y. The rule above has a high score because funding agencies often support work on specializations of their stated interests; an agency may specify *plants* but support *dandelions*, may specify *transportation* but support *air travel*, may specify *heart disease* but support *mitral valve prolapse*. On the other hand, agencies typically state their interests at the most general level possible, so proposals that request funding for more general topics are likely to be denied. One cannot approach the National Heart, Lung, and Blood Institute with a proposal to study anatomy, since that agency is interested in much more specialized topics. This reasoning is represented by giving the following path endorsement a low score, and calling it a *negative* path endorsement.

request-funds-for-topic(x) and INSTANCE-OF(x,y) →

request-funds-for-topic(y) (4)

Negative path endorsements constrain spreading activation by disallowing particular transitions through the network. The example in (4) says that if we are searching for funding from the *heart-disease* node in Figure 3, we should not allow activation to spread to the *mitral valve prolapse* node over the *instance-of* relation because any agency associated with that node would be unlikely to fund the proposal.

The relationship **R** in (1) need not be a single link, but could be a chain of links. Referring again to Figure 3, one can imagine that a funding agency interested in the heart might support work on mitral valve prolapse; that is, spreading activation from mitral valve prolapse to its *setting*, the mitral valve, then to the heart, which *has-part* mitral valve, may find an agency that is likely to fund the original proposal. This is denoted by giving a high score to the positive path endorsement

request-funds-for-topic(x) and HAS-SETTING:PART-OF(x,y) →
request-funds-for-topic(y) (5)

Negative path endorsements like (4) constrain search by disallowing spreading activation. Since GRANT follows high-scoring endorsed paths before lower-scoring ones, positive endorsements like (5) order search. Path endorsements are *heuristic*: (3) and (5) could lead to agencies that will not fund the proposal, and (4) could lead to a willing one[2]. Currently, GRANT uses about

[2]GRANT engages in *best-first search* [13] through a search space defined by its network. The heuristic evaluation function is not computed dynamically at each node by lookahead, but is rather a precompiled list of endorsed paths to search and prune.

120 path endorsements to prune and order search paths. These were determined empirically during the early days of the GRANT project and have not been changed appreciably since. Given that 48 different links are used in GRANT's network, many more than 120 different pathways can be traversed. The set of path endorsements is not complete, except in the weak sense that unendorsed pathways are treated as if they are negatively endorsed – that is, they are pruned during search.

3.3 Full matching

Constrained spreading activation finds a single semantic pathway between a proposal and each agency it reports as a potential funding source. But what if the proposal and agency share just a single interest – discovered by the search – but are otherwise completely different? For example, an agency may support research on reproduction in plants, while a proposal requests funding to study the economic impact of dandelions on landscaping. These seem to be a poor match, yet according to (2) above, the agency is likely to fund the proposal based on the semantic match between *dandelions* and *plants*. It appears that GRANT needs a way to calculate the full match between all aspects of a proposal and an agency, once it has found a partial match based on single pathway between them. In fact, we have not focused on full matching algorithms because GRANT currently performs adequately without one, and because its performance was not significantly improved when we added one to an earlier version of the system. Looking to the future, however, the analyses of partial matching presented in this paper have convinced us that GRANT will eventually require full matching to achieve major reductions in its fallout rate.

4 Evaluation of GRANT

GRANT's evolution from a small, prototype system [14] to the present has given us the opportunity to compare performance as the system has been scaled up, and to consider the potentials and pitfalls of developing other GRANT-like systems. This section discusses a battery of tests on the current system.

The primary measures of GRANT's performance are *recall* and *fallout rate*. (A third statistic, *precision*, is 1.0 - fallout.) Recall is the percentage of all the agencies accepted by the expert that GRANT found, and fallout is the percentage of all the agencies found by GRANT that were judged good by GRANT but bad by the expert:

$$\text{fallout} = \frac{\text{num. of agencies judged good by GRANT, bad by expert}}{\text{num. of agencies judged good by GRANT}}$$

$$\text{recall rate} = \frac{\text{num. of agencies judged good by GRANT, good by expert}}{\text{num. of agencies judged good by expert}}$$

To calculate recall and fallout for a proposal, we need to generate a list of agencies from which the expert can select the ones that are likely to fund the proposal. One method would be to have the expert rank all 700 agencies in the network for each proposal, but this would be exhausting. Instead, GRANT is run in a minimally-constrained, spreading activation search that reports all agencies found within a given "distance" from each research topic in the proposal. This is called breadth-first (BF) search[3]. For each proposal, we first run a BF search then ask our expert to classify the agencies it finds as good or bad. Since the search is blind, many of the agencies are bad; that is, unlikely in the expert's judgment to fund the proposal. Then we run GRANT in an *endorsment constrained* mode called EC search, avoiding negatively-endorsed pathways and favoring positively-endorsed ones. It finds a subset of the agencies discovered by BF search. Ideally, it should find all and only the agencies ranked as good by the expert, but in practice it fails to find some of the good agencies (called *misses*) and finds some bad ones (called *false positives*). GRANT's miss rate tends to be very low, so we will be concerned primarily with the relationship between the fallout rate and recall rate.

The following tests were all performed on a set of 27 proposals, representing the interests of a diverse group of first-year faculty at the University of Massachusetts. The first test was designed to probe the utility of endorsement-constrained search. We compared EC and BF search with a third mode called *unconstrained keyword search* (UKW). It finds all agencies that share a common research interest with a proposal. It is implemented as a search for all agencies exactly 2 links distant from the proposal. For example, if a proposal and an agency share the common interest *dandelions*, then each will be linked to that node by, say, a SUBJECT link. The two-link

SUBJECT : *dandelions* : SUBJECT-OF

path connects the agency and the proposal via the common term dandelion; and, in general, any two-link path between an agency and a proposal indicates a shared term. UKW search is thus a simple *keyword* search, since it finds only those agencies that share terms with proposals. The relevant statistics for UKW, EC, and BF searches are shown in Table 1.

	UKW	EC	BF
fallout rate	64%	71%	94%
recall rate	44%	67%	100%
number of agencies found	164	406	2145
number of false positives	106	207	2013
number of hits	58	88	132
number correctly rejected	0	111	0

Table 1. Statistics from UKW, EC, and BF searches.

[3]Completely unconstrained BF search finds all agencies in the network, each by dozens of different paths, and requires hours of CPU time on a TI Explorer Lisp Machine. The data presented here are for a modified version of BF search that avoids nodes with extremely high fan-out and prunes paths longer than 4 links.

EC search has a higher recall than UKW and a lower fallout rate than BF. Its fallout rate is typically higher than UKW because it subsumes UKW: it finds all the agencies that UKW finds, then finds some more by exploiting semantic relations. Let us consider the utility of this additional search.

Of the agencies found by GRANT for the 27 test cases, the expert thought that 132 would be likely to fund their respective proposals. UKW found just 44% of these. To find the rest, it is necessary to exploit semantic relationships between the terms used in research proposals and agency descriptions. EC search found 67% of the agencies judged good by the expert. It found 242 more agencies than UKW search: 30 hits, 101 false positives, and 111 correctly rejected. So in the regions of the network that cannot be explored by keyword UKW search, EC search found 40% of the agencies it should, and incorrectly accepted 101 agencies, for a "marginal" fallout rate of 42% . In contrast, BF search found almost all the agencies judged good by the expert, but at a cost of a 94% fallout rate.

In practice, GRANT's mode of operation is EC search. It is preferred to UKW search because it finds more agencies, and to BF search because it has higher precision. BF search finds about 80 agencies per proposal at a precision of 6% — only 1 agency in 20 is truly worth pursuing. EC search reports fewer agencies (15 per proposal), has a better level of precision (29%) than BF search, and has an acceptable, intermediate recall rate (67%).

Since EC search subsumes UKW search, it also inherits a significant fallout rate. The fallout rate for agencies found by keyword UKW search is 64%, but the marginal rate for those agencies found by additional semantic matching is just 42%. Clearly, path endorsements can increase precision. But their utility is obscured to some extent by the fact that EC search "starts off" with the 106 false positives found by UKW search. With this proviso stated, we now explore how to increase the recall and precision of EC search.

Our experiments are designed to address two general hypotheses:

- GRANT's performance is due to its path endorsements.
- GRANT's performance is affected by the *structure* of its network, including the lengths of pathways between proposals and agencies, and the degree of interconnection between nodes.

A third hypothesis is that GRANT's performance is affected by how its language of links is used to encode the interests of agencies. Since many people worked on GRANT's knowledge base, we were concerned that knowledge was encoded inconsistently. We calculated several statistics that measure consistency, but we did not find significant or even suggestive correlations of these measures with fallout rates. We cannot conclude that inconsistencies have no affect on GRANT's performance, because our measures of consistency may not be sufficiently sensitive. But we have found much stronger evidence for the other two hypotheses.

Structural Factors in Recall and Precision. We first calculated the recall and fallout rates as a function of the distance between proposals and agencies in EC search (Table 2). As noted, at distance = 2 EC has the same fallout rate as UKW search, which finds all agencies within two links of the proposal. Extending the search one more link increases the recall rate substantially (from 42% to 70%) and also raises the fallout rate somewhat. Interestingly, extending the search further has almost no affect on the recall rate but does increase the fallout rate. This suggests that endorsement-constrained search as implemented here offers most advantage when finding agencies based on a single semantic relationship between a term used in the proposal and a term used in the agency description. Increased fallout limits the utility of longer chains of relations.

length	fallout rate	recall rate
less than 3	64	42
less than 4	73	70
less than 5	78	69

Table 2. Recall and fallout rates for searches along pathways of different lengths.

The structural feature of GRANT's network that accounts for most variance in recall rate and fallout rate is the *branching factor* of nodes, that is, the number of links that connect nodes. In an experiment reported in Kjeldsen and Cohen [15] we found that the fallout rate was correlated with the average branching factor of pathways to agencies. Average branching factor is the average of the number of links emanating from each node on a pathway. It is a measure of the "density" of the network in the vicinity of the pathway. We expected dense areas of the network to have low fallout rates relative to recall rates, since there are more nodes per agency in dense areas, and thus more basis for discriminating good agencies from bad ones. Table 3 shows the percentage of the false positives found along pathways with low, medium, and high branching factors.

EC Search	average branching factor		
	2 - 7	8 - 15	> 16
% hits	20.3	40.6	39.1
% false positives	8.4	36.9	54.6

UKW Search	average branching factor		
	2 - 7	8 - 15	> 16
% hits	30.7	55.1	14.1
% false positives	8.4	37.3	51.8

Table 3. Hits and false positives for EC and UKW search, distributed by average branching factor.

Contrary to our expectations, the majority of false positives were associated not with low branching factors but rather with high ones. For EC search, 54% of the false positives were found on paths with an average branching factor greater than 16. For UKW search, 51% of the false positives were associated with high branching factor; furthermore, only 14% of the hits were found in these areas. We looked at the test cases individually to try to explain this result. Many of the false positives were associated with nodes with high fan-out, such as "animal" and "location." We believe that such nodes are relatively general, that their fan-out is due to their many specializations. To say an agency is associated with one of these general nodes is to say very little about its interests, so agencies found via these nodes are more likely to be false positives.

These data seem to suggest that we could increase GRANT's precision by pruning agencies associated with general nodes. In fact, this is an artifact of the way we calculate precision. We could certainly reduce the number of false positives this way, but we would also reduce the number of agencies GRANT finds, and so would have little effect on the fallout *rate*. Moreover, since the denominator of the recall rate is constant — the number of agencies judged good by the expert — pruning agencies can only reduce the recall rate. Clearly, false positives are associated with higher branching factors. However, the key to improving precision is not to prune agencies, but to restructure the network so that it has fewer pathways with high branching factors, that is, fewer nodes that represent very general concepts. For example, the current network defines *dandelion* and *tomato plant* as instances of the *plant* node, though they are obviously different kinds of plants. The distinction could be made by defining dandelion as an instance of a *weed* and tomato plant as a *domestic plant,* but because these nodes are not in the network, the fan-out of *plant* is higher than it should be and *dandelion* and *tomato plant* are not adequately discriminated.

The statistics in Table 3 suggest that the "ideal" branching factor is less than 16. Another experiment was needed to pinpoint the ideal more precisely. Starting with the list of agencies found by the EC search and reported in Table 1, we ranked the agencies by their branching factors, and recalculated the recall rate and fallout rate for each successive level of the ranking. That is, we superimposed a ranking by branching factor on the list of agencies found by EC search and asked about the recall rate and fallout rate of all agencies that had, first, low branching factor, then those that had higher branching factor, and so on. (For reasons discussed below, we used the branching factor of the last node on a pathway instead of the average branching factor over all nodes on a pathway.) The results are shown in Table 4.

Agency is counted as "good" if the branching factor is:	fallout rate	recall rate	number of FPs	% change number of FPs	number of hits	% change number of hits
any number	73	63	219	2	82	1
16 or less	73	62	215	14	81	17
13 or less	73	53	188	55	69	15
10 or less	67	46	121	157	60	140
7 or less	66	19	47	81	25	25
3 or less	58	15	26		20	

Table 4. Fallout and recall rates from ranking agencies by branching factor.

These data suggest that disproportionate numbers of false positives are associated with low and moderately high branching factors. At the lowest level (branching factor of 3 or less) there are few false positives (26) and hits (20) because few nodes have such low branching factors. At the next level we consider agencies found via nodes with branching factor of 7 or less. 47 are false positives, an increase of 81%, and 25 are hits, an increase of 25%. Thus, fallout rate increases faster than recall rate for nodes with relatively low branching factors. When nodes with higher branching factors (10 or less) are considered, fallout rate increases by 157% and recall rate by a comparable 140%. However, adding agencies that are found by nodes at the next level of branching factor (13 or less) increases fallout rate by 55% but increases recall rate by only 15%. The rates then increase proportionately for higher levels of branching factor.

The greatest increase in recall and fallout occurs when we add the agencies found via nodes with branching factors between 8 and 10. Moreover, the numbers of hits and fallouts increase by roughly the same amount in this area (about 150%). In contrast, false positives increase more rapidly than hits at low (3 - 7) and moderately high (11 - 14) branching factors. This suggests that the "ideal" branching factor is between 8 and 10, and supports the hypothesis that recall and fallout rate are correlated with the generality – as measured by branching factor – of nodes. As mentioned above, we used the branching factor of the last node on a pathway – the one "nearest" to the agency and "furthest" from the proposal – to produce the data in Table 4. We reasoned that very specific nodes, those with low branching factor, would rarely be part of an agency description, and so would not be associated with many hits. On the other hand, as we argued above, nodes with very high branching factors are too general to represent the interests of an agency unambiguously, and so would be associated with high fallout rates.

The primary implication of these results is that knowledge engineers for GRANT-style systems should ensure that the definitions of new terms are as specific as possible. For example, the knowledge engineer should define a new plant in terms of the most specific possible subclass of plants, or perhaps create a new subclass, rather than linking the new plant to the general *plant* node. Currently, GRANT is programmed to avoid nodes with extremely high fan-out. An alternative would be to alert the knowledge engineer to them during the development of the knowledge base, to fix the problem before it arises. Then, any remaining nodes with high fan-out almost certainly denote concepts that are too general to be useful, and endorsements could be designed to avoid them, or to give them a low rank.

Endorsements as Factor in Recall and Precision. Our second hypothesis is that although the representation language for the network is probably sufficient to encode the meaning of research proposals and agency descriptions, these representations are not being exploited by endorsement-constrained search. Several findings support this hypothesis. In Kjeldsen and Cohen [15] we reported that just three path endorsements accounted for 85% of the hits but the same three led to 42% of the false positives. The culprits were:

- SUBJECT : SUBJECT-OF
- SUBJECT : SUBJECT-OF : SUBJECT-OF
- OBJECT : SUBJECT-OF

Despite the fact that 48 distinct relations are used in the network to connect concepts, just 3 (SUBJECT, OBJECT, and SUBJECT-OF) were sufficient to find the majority of hits and a sizeable portion of false-positives. This is partly due to the relative frequency of these links in the network: they are very common and so support a disproportionate number of path traversals. However, our data suggest that the reliance on these links is not due entirely to their frequency, and that intelligent use of other links could increase recall rate.

We measured the frequency with which different links were used to represent agency descriptions. These data are shown in Table 5. As expected, SUBJECT, OBJECT, and FOCUS were most common, but WHO-FOR and LOCATION were not infrequent. However, these latter links were almost never traversed to find agencies: Table 6 shows the results of using the last link in a pathway (the one closest to the proposal) to rank the agencies found by EC search. If SUBJECT and OBJECT are the only links that GRANT is allowed to traverse, then it finds 74 hits and 179 false positives. It finds an additional 15 hits when it is also allowed to traverse FOCUS. But, remarkably, allowing it to traverse *any* link results in only 2 more hits: Most of GRANT's hits are found by following SUBJECT, OBJECT, and FOCUS links into an agency. Although WHO-FOR and LOCATION are used quite often to define the interests of agencies, they are not used to find the agencies. This is not surprising, since WHO-FOR and LOCATION are the final link in only 2 path endorsements. But it does suggest that using these and other links judiciously could increase GRANT's recall rate. In general, these results stress that path endorsements must reflect the conventions for representing concepts.

Link	Number of uses in agency definitions	Number of uses as last link of endorsements
subject	513	19
object	258	10
focus	238	17
who-for	124	2
location	80	0
dv	30	8
iv	20	5
rv	18	5

Table 5. Number of times each link is used to define agency interests, and number of times it is the final link in an endorsement.

Agency is counted as "good" if it is found by an endorsement classified as:	fallout rate	recall rate	number of FPs	% change in number of FPs	number of hits	% change number of hits
very-likely	55	18	28	425%	23	78%
likely or very-likely	73	42	147	41%	54	59%
maybe, likely, or very-likely	71	67	207	4%	86	0%
unlikely, maybe, likely, or very-likely	72	67	216		86	

Table 6. Fallout and recall rates from ranking agencies by class of path endorsements.

To get a more complete picture of the utility of GRANT's path endorsements we would perform "ablation studies" — removing path endorsements one at a time to see how they affect recall and precision. Unfortunately, an exhaustive analysis of all endorsements would require weeks of computer time. Instead, we grouped the path endorsements and assessed the effects on performance of removing these classes. Every path endorsement is assigned to one of five classes that reflects the subjective probability that an agency found by that endorsement would fund the proposal. The classes are trash, unlikely, maybe, likely, and very-likely. We used these classes to rank as "good" or "bad" the agencies found by EC search, then recalculated recall and fallout rates for each rank. The results are shown in Table 7.

Agency is counted as "good" if the last link in a pathway is:	fallout rate	recall rate	number of FPs	number of hits
SUBJECT or OBJECT	71	57	179	74
SUBJECT, OBJECT, or FOCUS	72	68	228	89
ANY LINK	73	70	251	91

Table 7. Fallout and recall rates from ranking agencies by final link.

When only *very-likely* endorsements are allowed, the numbers of hits and false positives are low (23 and 28, respectively). Adding in agencies that are found via paths with *likely* endorsements increases the number of false positives by over 400% to 147. This seems an excessive price to pay for the 78% increase (from 23 to 54) in the number of hits. In contrast, adding in agencies with *maybe* endorsements increases the number of hits by 59% and increases false positives by a significantly lower amount, 41% . (The main reason for the increase in recall is that FOCUS links are used in a preponderance of *maybe* endorsements, and are infrequently used in *likely* or *very-likely*. We saw in Table 5 that the FOCUS link is used frequently in defining agencies, and in Table 6 that inclusion of the FOCUS link increases GRANT's recall rate.)

Clearly, GRANT's fallout rate could be improved by refining its *likely* endorsements. The improvement in performance due to adding *maybe* endorsements — specifically those dealing with FOCUS links — convinces us that it is possible to add endorsements that will increase recall and precision simultaneously. Table 5 suggests that these endorsements should exploit WHO-FOR and LOCATION links, which are used to define agencies but are rarely traversed to find them. We are currently designing new endorsements, though they will have to be tested on a new set of proposals to ensure that they are not simply "tuned" to the current test cases.

5 Discussion

The main conclusion of our work is that constrained spreading activation finds agencies based on semantic relations, with reasonable recall and precision, that would not be found by simple keyword search. From a pragmatic standpoint, the Office of Research Affairs at the University of Massachusetts prefers GRANT for several reasons to the database program that it used previously. GRANT is more efficient. A session takes just a few minutes: the proposal is coded, GRANT runs a search, a list of 15 agencies (on average) is returned, and the user sorts them to find 2 or 3 that are ideal for the client. In contrast, a similar search takes about 2 hours with the old keyword database system, in part because the dozens of agencies returned by the old system must be carefully sorted (its precision is only about 5%). GRANT's performance is well-suited to the funding domain because researchers rarely send a proposal to many agencies, but several agencies will typically fund a piece of research. Thus, GRANT's relatively low precision (29%) is not bothersome because a search returns relatively few agencies — ample to find 2 or 3 for the client but few enough to sort quickly. And since a proposal can potentially be funded by several agencies, GRANT's recall rate (67%) is sufficient to find enough good candidates for the user.

GRANT was designed to have the advantages of human associate memory but to be more reliable. It is difficult to evaluate any system on such vague criteria, but the experiences of GRANT's users are suggestive: At first, they expected GRANT to accelerate their processing of "easy" cases. They found instead that easy cases were those that could be answered from memory, and that GRANT is most useful for difficult cases — those for which no agencies come to mind. Apparently, GRANT's associative memory finds plausible semantic connections between topics in proposals and agencies that human funding advisors either forgot or never knew.

We are considering other applications of constrained spreading activation. A straightforward extension of GRANT is to run the system "backwards," taking as input an agency's request for proposals (RFP) and searching for the appropriate faculty members to receive the RFP. The research interests of many of the faculty at the University of Massachusetts have been encoded for this purpose. Another goal is an intelligent index for a major reference book, since GRANT is adept at inferences of the form "if a researcher (or reader) is interested in topic **X** then he or she is likely to be interested in a related topic **Y**." Other potential applications are literature search and searching databases of news wire services.

Although constrained spreading activation is a simple algorithm, and seems widely applicable, the investment required to build GRANT-like systems is substantial. Five steps are involved. First, one must analyze the domain to design a language for representing the domain's concepts and their interrelationships. Concepts in GRANT's network are linked by 24 different relationships and their inverses. We had to interview an expert funding advisor at length to acquire this vocabulary of links. Second, a network must be constructed to represent and index the targets of search, be they agencies, bibliographic references, or people. Roughly 4 person-months of effort were required to build GRANT's 4500-node, 700-agency network. Third, path endorsements must be formulated. Fourth, the system must be tested and the path endorsements refined. Finally, for most interesting domains, one will be constantly updating information about the targets of search, adding new ones, modifying the descriptions of old ones, and so on.

Current artificial intelligence technology may streamline the process of building GRANT-like systems. In GRANT's domain, for example, it should be possible to exploit tools from natural language research to automatically parse the textual descriptions of agencies' research interests. The parsed representations would then be automatically indexed by GRANT's network. It seems unlikely that this could succeed, however, without an extensive network of research topics to begin with. A second area of potential improvement is the design and testing of path endorsements. GRANT can, in principle, learn path endorsements for itself. Several mechanisms have been proposed. The simplest keeps lists of paths that lead to good agencies and bad ones. Path endorsements are derived from particular paths by dropping the intervening nodes, leaving only the links; for example, if the path

SUBJECT : ***dandelions*** : ISA : ***plant*** : SUBJECT-OF

leads to a good agency, then one could generate the path endorsement SUBJECT : ISA : SUBJECT-OF, or increase its score if it already exists. Similarly, paths that lead to bad agencies result in lower scores for their associated path endorsements.

Constrained spreading activation has a fundamental limitation that cannot be overcome by adjusting path endorsements. It requires only one positively-endorsed pathway between a proposal and an agency to report a hit. All the statistics in this paper are based on this ***partial*** matching approach, which finds a single basis for matching a proposal with an agency. Partial matching will always have a significant fallout rate, because it will inevitably find agencies that have one thing in common with a proposal but are otherwise completely different. For example, a proposal to study dietary sodium and hypertension will be recommended to an agency that supports research on salt domes for oil reserves, simply because both are interested in salt. The only solution to this problem is full matching — calculating the match between a proposal and an agency based on all the connections between them. The constrained spreading activation algorithm can find these connections, but we have yet to explore how a total degree of match is to be calculated.

References

[1] Baddeley, A. *The psychology of memory.* New York: Basic Books; 1976

[2] Anderson, J. R.; Bower, G. H. *Human Associative Memory.* Washington D. C.: Winston; 1973.

[3] Minsky, M. *Semantic Information Processing.* Cambridge, Mass.:MIT Press; 1968

[4] Quillian, M. R. "Semantic Memory." In M. Minsky, Ed. *Semantic Information Processing.* Cambridge, Mass.:MIT Press; 1968; 216-270.

[5] Anderson, J. R. *Language, Memory, and Thought.* Hillsdale, N.J.: Lawrence Erlbaum Associates; 1976.

[6] Cohen, P. R.; Feigenbaum, E. A. *The Handbook of Artificial Intelligence, Volume 3.* Los Altos CA.: William Kaufmann, Inc.; 1982.

[7] Findler, N. (Ed.) *Associative networks: the representation and use of knowledge by computers.* New York: Academic Press; 1979.

[8] Brachman, R.; Levesque, H. *Readings in Knowledge Representation.* Los Altos, CA.: Morgan-Kaufmann, Inc.; 1985.

[9] Minsky, M. "A framework for representing knowledge." In P. H. Winston (Ed.) *The psychology of computer vision.* New York: McGraw Hill; 1975; 211-277.

[10] Fillmore, C. "The case for case." In E. Bach and R. Harms (Eds.) *Universals in linguistic theory.* New York: Holt, Rinehart, and Winston; 1968; 1-88.

[11] Schank, R. *Conceptual information processing.* New York: North Holland; 1975.

[12] Lenat, D; Prakash, M; Shepherd, M. "CYC: Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks." *AI Magazine.* 6(4): 65-85; 1986.

[13] Nilsson, N. *Principles of Artificial Intelligence.* Palo Alto CA: Tioga; 1980.

[14] Cohen, P.R.; Davis, A.; Day, D.; Greenberg, M.; Kjeldsen, R.; Lander, S.; Loiselle, C. "Representativeness and Uncertainty in Classification Systems." *AI Magazine.* 6(3): 136-149; 1985.

[15] Kjeldsen R.; Cohen, P.R. "The Evolution and Performance of the GRANT System."to appear *IEEE Expert.* Spring 1987.