CAIA 87:
Special Feature

The Evolution and Performance of the GRANT System

Rick Kjeldsen and Paul R. Cohen

University of Massachusetts

Photograph by Alnes

Pepperdine University, Malibu, California, 1987.

Matching is a basic task for AI problem solvers.[1,2] It underlies such varied tasks as classification, analogy, generalization, and rule application. Matching algorithms have been designed that tally shared and unshared properties of concepts[3] so that concepts match well if they share relatively many common properties. We will refer to this as syntactic or keyword matching, since the properties that describe concepts must exactly match, without regard to their meaning, for them to contribute to the overall match. A problem with this approach is that related concepts may not match if they are described in semantically related terms. For example, if a job description requires "knowledge of a text editor," and an individual's resume says "10 years experience with EMACS," then the two should match because EMACS is a text editor. The job description and resume match not on a common keyword but on the terms EMACS and text editor, which are semantically related by a subclass relation. We have developed a *semantic matching* algorithm that finds matches between concepts based on semantic relations between their properties. It has been implemented in an expert system, called GRANT, for finding sources of funding for researchers based on semantic matches between research proposals and descriptions of funding agencies' research priorities. In the past, GRANT has performed far better than keyword matching, but recently performance has declined. This article explores the causes of the decline and suggests how performance may be improved.

0885-9000/87/0500-0073 $01.00©1987 IEEE

The GRANT system

The goal of the GRANT system is to find funding sources that are likely to fund a given research proposal. Success in this venture depends on what one means by "likely." In fact, GRANT was built primarily to explore the interpretation of "likelihood" in classification problems,[4] problems where hypotheses are represented by typical cases, and are considered conditionally likely if data are a good "fit" or match to the hypotheses.[5-7] In GRANT, the "hypotheses" are the descriptions of funding agencies, especially their research interests, and the data are the research topics in a grant proposal.

The likelihood that an agency will fund a proposal depends on the degree of fit between their respective research interests. Syntactic or keyword matching is insufficient for this task for two reasons: Researchers and agencies may use different terms to describe their interests even though the terms may have the same or similar meanings, and interest in one research topic often implies interest in another slightly different but semantically related topic. The first problem can be partly solved by using a large table of synonyms (and GRANT has one), but the second requires knowledge of what the descriptions of research interests *mean*. GRANT searches in a large semantic network for agencies whose research interests are related to those of proposals. Its search is constrained by knowledge about which semantic relations lead to agencies that are likely to match a proposal and thus fund the research.

Once GRANT has found an agency by this search process, it calculates a total degree of match between the agency and the proposal. A simple mechanism, based on the tallying algorithms mentioned above, compares all properties of the agency with all those of the proposal. Both common keywords *and* semantically related properties are included. The results are used to determine a ranking on all the agencies found during a search. The emphasis of our research, however, is on the search itself; that is, on the process of finding agencies that are likely to match the proposal. The ranking process is not discussed here.

Search for funding agencies takes place in a large semantic net of research topics under the guidance of a set of heuristic rules. This net has been designed and built explicitly to suit the needs of the domain; for example, one

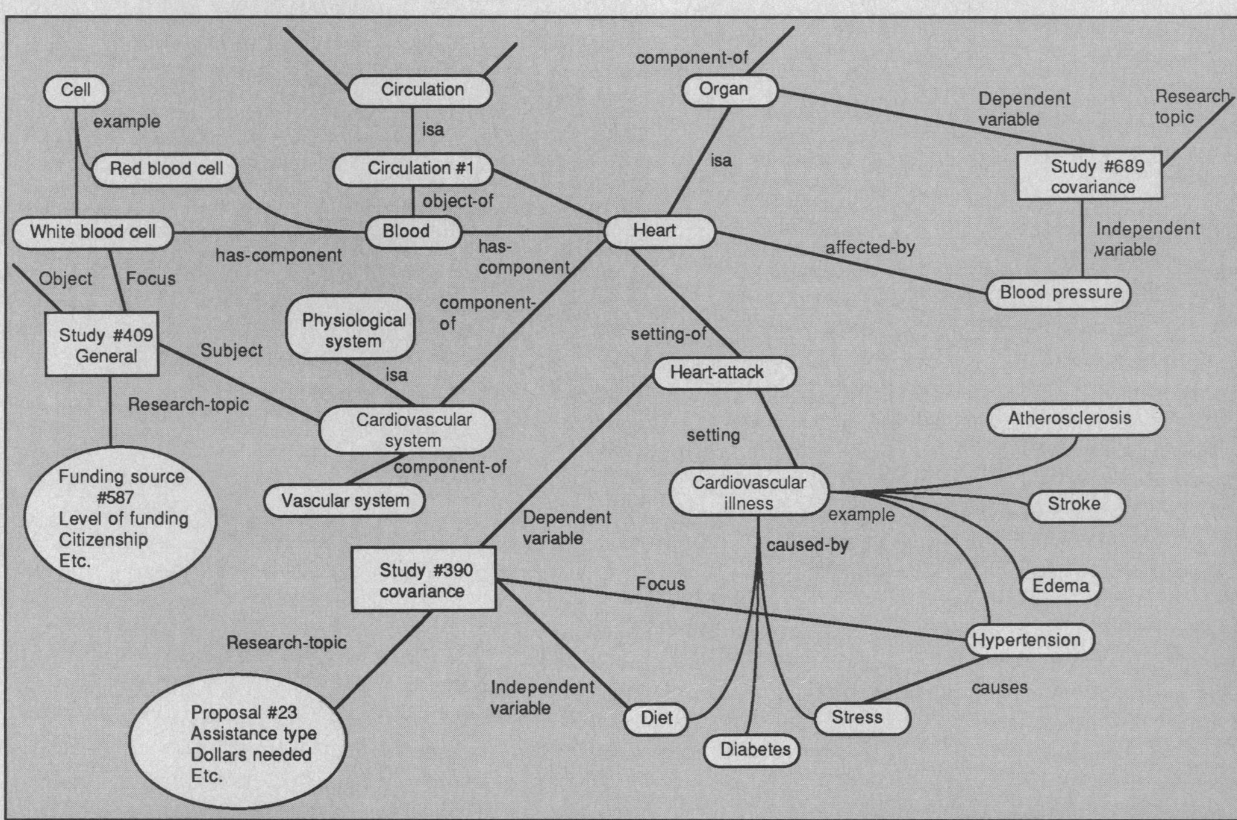


Figure 1. A portion of the GRANT knowledge base.

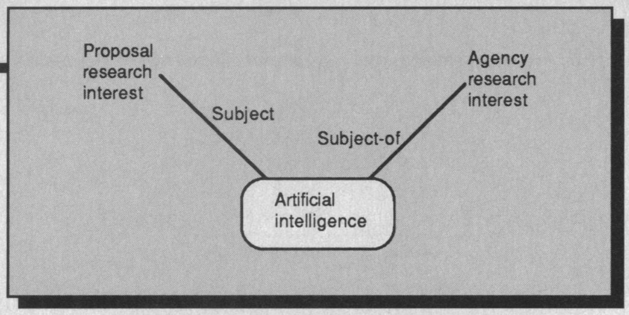


Figure 2. Case 1.

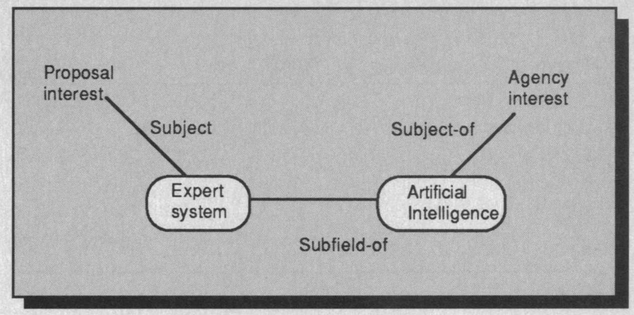


Figure 3. Case 2.

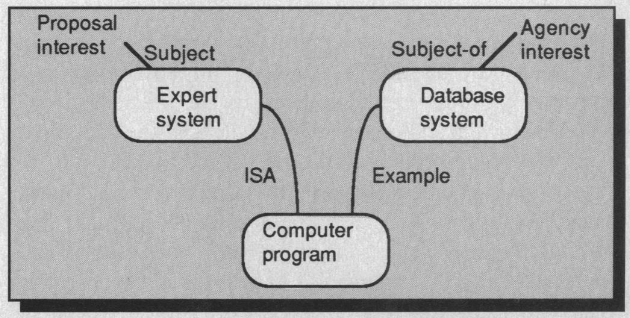


Figure 4. Case 3.

field 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 links represent simple relations with no inheritance.

The nodes in the net are represented as frames, where slots represent the links to other nodes. Some of the nodes represent funding agencies and the research topics they support. Agency frames have slots for level of funding, citizenship restrictions, and so on, as well as links to one or more research topics. Research topic nodes tie into the network through the terms that describe them (see Figure1).

The frames that describe research interests of both agencies and proposals are created by classifying the goal(s) of the research into one or more of the following 10 classes:

Design	Educate	Improve	Intervene	Manage
Supply	Promote	Protect	Study	Train

Each class is 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 identify the topic and a *focus* slot that describes which aspect of the topic will be investigated. A portion of GRANT's network, including several research topics, is shown in Figure 1. The entire network includes over 700 funding agencies, with interests described by over 4500 nodes, linked by 48 distinct relations.

A search of the network for agencies begins at the frame that represents the research topic of the proposal. All nodes one link away from the research topic nodes are added to the *search agenda*. These are ranked as will be described below. Any nodes that are ranked below a preset threshold are pruned from the agenda. Nodes with high rank are expanded; that is, all nodes one link away from *them* are added to the agenda and ranked. When a node that represents a research interest of an agency is discovered, the agency is added to a list. This continues until a prespecified limit on the depth of search is reached, or a desired number of funding agencies is found. The agencies that are found are evaluated using the full matching process we mentioned earlier.

By ranking nodes on its agenda, GRANT performs a *best-first* search, focusing on those areas of the network most likely to yield agencies that will fund the research proposal, and avoiding other areas. These areas are identified by evaluating the semantic associations that lead to them from the research interest of the proposal. Associations are represented by links in the network; thus, semantically important sequences of links called pathways are used to guide the search. These are identified by *path endorsements* set up by the knowledge engineer. During search, the ranking mechanism favors nodes that are found along *positive* endorsed pathways and avoids those on *negative* endorsed pathways. If nodes representing the research interests of a proposal and an agency anchor a positively endorsed pathway at its beginning and end, then the agency is likely to fund the proposal. Conversely, an agency found via a negatively endorsed pathway is unlikely to fund the proposal.

The simplest path endorsements represent keyword matches; for example, in Case 1 both the proposal and the agency state that AI is a research interest, as Figure 2 illustrates. But the path endorsements that account for GRANT's superiority over keyword matching are those that find semantic matches (see Figure 3). Here, the SUBJECT:SUBFIELD:SUBJECT-OF path is used to find an agency to fund research on expert systems, an agency whose stated interest is not expert systems but AI.

During search, GRANT prunes from the agenda any nodes found along negatively endorsed pathways. The most powerful negative endorsement prevents GRANT from finding agencies if their research interests and those of the proposal are different specializations of a common node, as

Table 1. Previous tests of GRANT.

	HR	FPR
Fall 1985:		
2000 nodes, 200 agencies	80%	26%
Winter 1986:		
4500 nodes, 700 agencies	98%	61%

Figure 4 exemplifies. Both *expert systems* and *database systems* are specializations of a common node, *computer program*. (Links are directional in GRANT: expert systems and database systems are each related to computer program by ISA and by EXAMPLE. However, this pathway from the proposal to the agency follows just the ISA and EXAMPLE links.) An agency that will fund research on database systems will not, in general, fund research on expert systems unless there is a reason to do so other than that they are both computer programs. In general, agencies will not fund specializations of generalizations of their interests. GRANT captures this knowledge in the negative path endorsement ISA:EXAMPLE. Any node arrived at via a path that contains this subpath (such as SUBJECT:ISA: EXAMPLE:... , above) is pruned from the agenda.

GRANT's search for agencies is heuristically constrained by about 60 positive endorsements and an equal number of negative endorsements. During search, it first prunes negatively endorsed paths, then expands positively endorsed ones, and then may optionally expand pathways for which no endorsement exists.

At its best, GRANT has performed well, finding 4/5 of the agencies selected by an expert; and of the agencies it recommended, 2/3 were judged good. The Office of Research Affairs (ORA) at the University of Massachusetts is enthusiastic about the performance and utility of the system, which has been in use since its completion. As GRANT has evolved, however, its performance has decreased; for example, when we quadrupled the size of its semantic network, the false positive rate increased. In the past, GRANT underwent limited testing that raised more questions about its performance than it answered. In the next section, we summarize the results of these early tests, and present recent, more extensive testing done in an effort to identify the reasons for decreased performance. We conclude with new and substantially better performance figures from a modified version of GRANT.

Experiments with GRANT

Ideal performance would be to find all agencies that are likely to fund a proposal and none that are unlikely to fund it. GRANT's *hit rate* (HR) measures the proportion of agencies it finds that are judged by an expert as likely to fund the proposal. Its *false positive rate* (FPR) is the proportion of agencies it found that are judged unlikely to fund the proposal. That is,

$$\text{HR} = \frac{\text{agencies judged good by GRANT and the expert}}{\text{agencies judged good by the expert}}$$

$$\text{FPR} = \frac{\text{agencies judged good by GRANT and bad by the expert}}{\text{agencies judged good by GRANT}}$$

A test set of 23 proposals was used to assess these figures for two versions of GRANT. The results are shown in Table 1. The first test was run when GRANT was relatively small (roughly 2000 nodes in the network, and 200 funding agencies). The same test set was used to assess the system after it had been scaled up to over 4500 nodes with some 700 funding sources.

The most disturbing aspect of these results is that, as the knowledge base expanded, the false positive rate began to rise. From an acceptable 26 percent initially, FPR rose to 61 percent. Admittedly, GRANT was finding 98 percent of the agencies it should, but almost two out of three agencies it reported were judged unlikely to fund the proposal. We speculated on several causes for the increased FPR: First, there was a large push to make the system bigger and more useful to the ORA. Over 2000 general nodes and some 500 agencies were added to the network. During this time, the ratio of general nodes to agencies decreased from 10:1 to 6.4:1. Moreover, in the rush to add the general nodes needed to describe the new agencies, nodes may have been given only very basic, incomplete definitions. We suspected that GRANT was less able to discriminate bad agencies from good because it had fewer general nodes per agency to support the distinction, and those it had were not well linked into the rest of the network. A second change was that GRANT was expanded to include many arts and humanities, in addition to its original science-dominated network. The link set, however, was not expanded and may not have been sufficient to encode (and thus differentiate) the new topics of interest. Finally, the semantics of the net were not rigidly enforced. The many individuals who worked on the network may have had different ideas about how nodes should be defined. This could have caused links to be used in different ways at different times, so that path endorsement could not rely on a consistent interpretation for links.

These results raised questions about the engineering and utility of GRANT-like systems. Do the path endorsements help, or would unconstrained search do as well? What proportion of the hits and false positives were found on the basis of keyword matching; that is, does semantic matching help? These questions and others were probed in a second set of experiments.

A new set of 27 proposals was obtained from the ORA, representing the interests of a diverse group of new faculty at the University of Massachusetts. GRANT was run in on this data using three types of search. The first was an endorsement-constrained (EC) search that found agencies with interests semantically related to the proposal using the path endorsements as described above. The second was an unconstrained keyword search (UKW) that found *all* agencies that shared a common research interest with the proposal (see Case 1, above). UKW is our standard of comparison, since it is the kind of search one would get from a keyword search in a database system and the kind of search the ORA had been using until GRANT. It is implemented as a search for all agencies exactly two links distant from the proposal. The third search was like EC, except that it allowed unendorsed and some negatively endorsed paths to be expanded. We called it breadth-first search (BF), although it was not blind and exhaustive. An unconstrained BF search would, of course, find all the agencies in the network. Ours was limited in several ways: Extremely general nodes such as "thing" or "action" were not expanded, certain negative endorsements were used, and we imposed a depth limit of five links from the start node. BF was intended to probe the utility of path endorsements in constraining search for semantic matches.

Our expert classified the 403 agencies found by BF search on the 27 proposals. Of these agencies, 115 were classified as good, 288 as bad. Sixty-five percent of the agencies found by BF search were judged bad. This data was used to evaluate the agencies found during EC and UKW search. The results are shown in Table 2. As expected from our previous experiments, the false positive rate for EC search was quite high (63 percent), but we were surprised by the relatively low HR (68 percent), especially since BF search has about the same false positive rate and a 100-percent HR. UKW had a false positive rate roughly comparable with EC and BF (59 percent) but a lower HR (55 percent). UKW found *all* keyword matches between the proposals and the agencies, yet it only found 55 percent of the agencies judged good by the expert. The other 45 percent must be found by semantic matching. This clearly argues for GRANT's potential. Unfortunately, EC search did not realize that potential; it found only 76 of the 115 agencies.

In order to determine the explanations for the high FPR, we followed our suspicion that the expansion of GRANT's knowledge base had introduced a large number of incompletely defined nodes; that is, nodes with few defining links. This would reduce the power of the network to discriminate good agencies from bad. We measured the average branching factor along pathways found to hits and false positives, expecting false positives to cluster in "sparse" areas of the network; that is, those with low average branching factor. The results are shown in Table 3.

Contrary to our expectations, false positives were not correlated with low branching factors but rather with high ones. For EC search, 54 percent of the false positives were found on paths with an average branching factor greater than 16. For UKW search, 51 percent of the false positives were associated with a high branching factor; furthermore, only 14 percent 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.

Our next goal was to determine why endorsements seem to lower the HR as much as they do. The data in Table 2 suggest that EC search is too "conservative," expanding too few nodes to significantly improve on UKW search. This implies that the set of path endorsements is too restrictive, either because it has too few positive endorsements or too many negative ones. Looking at the test data, only eight agencies were incorrectly pruned by EC search (that is, found along a negatively endorsed pathway), compared with 51 that were never found because we had no positive endorsement to lead GRANT to them. We looked at the individual cases again to see which path endorsements were responsible for these results. Remarkably, just three path endorsements accounted for about 85 percent of the hits, but the same three led to 42 percent of the false positives. The culprits are (1) SUBJECT:SUBJECT-OF, (2) SUBJECT:SUBJECT-OF:SUBJECT-OF, AND (3) OBJECT:SUBJECT-OF.

Table 2. Statistics from UKW, EC, and BF searches.

	UKW	EC	BF
False-positive rate	59%	63%	65%
Hit rate	55%	68%	100%
Number of agencies found	175	272	403
Number of false positives	104	130	288
Number of hits	64	76	115

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

EC search	Average branching factor		
	2-7	8-15	>16
Percentage of hits	20.3	40.6	39.1
Percentage of false positives	8.4	36.9	54.6
UKW search	**Average branching factor**		
	2-7	8-15	>16
Percentage of hits	30.7	55.1	14.1
Percentage of false positives	8.4	37.3	51.8

Table 4. Search with new path endorsements.

One negative path endorsement: SUBJECT:FOCUS-OF:SUBJECT-OF	
Nine positive path endorsements constructed by selecting	
one of	SUBJECT, OBJECT, FOCUS
followed by	SUBFIELD-OF
followed by one of	SUBJECT-OF, OBJECT-OF, FOCUS-OF

Why, then, was the false positive rate low (20 percent) in an earlier version of GRANT? The argument is speculative, but we believe that the representation of research topics in the network has changed over time and now uses fewer link types while the set of path endorsements has remained the same. This has placed a disproportionate emphasis on just a few path endorsements that are not sufficient to differentiate all bad agencies from good. Agencies that we should find are not discovered because we lack path endorsements to lead to them given the new patterns of link use. This decreases the number of correctly endorsed agencies, which in turn increases the false positive rate. Moreover, many positive path endorsements are no longer relevant because the links they follow are now seldom used to describe research topics. They cannot contribute many hits; however, because of the way GRANT's path-ranking mechanism was implemented, they may still lead GRANT toward false positives. (When GRANT is ranking pathways, it attempts to complete positively endorsed pathways by giving subpaths of them high rank.)

To test these ideas, we made minor adjustments to the rule set, and observed the effects on performance. We designed 10 new endorsements to control search over the links that have become the most commonly used to describe research topics. These are summarized in Table 4. The results of EC search with these endorsements are promising: The false positive rate decreased to 45 percent and the HR increased to 86 percent. EC search found 35 more good agencies than keyword search, or 69 percent of the good agencies it could have found by semantic matching. This experiment was done quickly and had no control group; thus, our new endorsements *might* have been "tuned" to the test cases. We doubt this is the case for two reasons: The new path endorsements were designed to be general (each found new agencies in several test cases), and we can construct a logical argument for each in terms of our domain. We believe that further performance increases can be achieved by adjusting the rule set without sacrificing its generality. These results argue for a GRANT-style approach to semantic matching, but emphasize that the design of path endorsements must match the design of the network.

Discussion

We have shown that semantic matching can improve on keyword search for finding related topics. Could such a system ever achieve perfect performance? It may be that an absolute minimum FPR is determined by the approach itself. First, it can be argued that the basic mode of inference used in GRANT is abduction and that abduction itself can lead to false positive results. Second, path endorsements without some notion of context may not be sufficient to encode semantically meaningful associations.

Our goal is to find concepts that match our starting point by some measure. We take as an axiom that matching concepts are likely to be semantically related. In GRANT, this is the premise of the abductive inference that the existence of semantic relationships between concepts implies a match between them. That is, given the premise that matching concepts are related, if GRANT finds a relationship between concepts, it infers that the concepts match. Because this is an abductive inference, not a deductive one, it will occasionally be wrong. When it is, it will lead to false positive results, as GRANT will find semantically related concepts that do not match.

An extension of our axiom would state that the better two concepts match, the more relationships they are likely to share. Using an abductive argument once again, we can conclude that two concepts with several semantic relationships between them are likely to be a good match, and at least are more likely to be an acceptable match than concepts with a single relationship. Thus, one way to reduce false positives due to GRANT's method of inference is with an improved *full* matching algorithm, which takes into account multiple associations between concepts.

Path endorsements are based on the assumption that relationships between topics are sufficient to encode semantically meaningful associations *regardless* of the context in which they are used. For example, a component of a social group (a person) may not interest a funding source supporting investigation of interactions within such groups, while a component of a mechanism may well interest an agency willing to fund improvement of that mechanism. Thus the SUBJECT:HAS-COMPONENT: SUBJECT-OF path is good when the SUBJECT is a rotary engine, but not when the SUBJECT is a soccer team. Given a rich enough set of links we could capture such subtle differences (we could use HAS-MEMBER when referring to a social group) but this could lead to an unmanageable proliferation of links. Another alternative is to include in the rules a notion of context so that we can tell GRANT to use the above rule only when the SUBJECT node is connected by one or more ISA links to "thing." Without such improvements, path endorsements may not be able to differentiate concepts sufficiently to improve performance.

In a domain such as information retrieval, where a high

FPR is acceptable, GRANT's existing search techniques may be sufficient.[8] In a more stringent environment, however, changes may have to be made. We believe that with improvements such as better full matching, addition of context, and more careful knowledge engineering (including an improved set of links and better enforcement of network semantics) a GRANT-like system could achieve an impressive level of performance.

We have discussed GRANT's architecture and the utility of augmenting keyword matching with semantic matching. GRANT has evolved since it was first reported in 1985: Its knowledge base is much bigger, more agencies are being indexed by fewer general nodes, and the style of link use has changed. These changes have resulted in decreased performance. We have explored the specific causes of this reduction and have shown that with minor changes designed to counteract those causes, performance could be improved. In the coming months we will explore further the reasons for false positives in GRANT, and see how far performance can be improved without tuning the system to one set of test data. Other current research concerns algorithms for full matching of agencies, improving the rule set using learning techniques, and an examination of the semantics of the knowledge base. E

Acknowledgments

This research is funded by DARPA/RADC contract F30602-85-C-0014 and by NSF grant IST8409623.

We wish to thank Bruce McCandless, Office of Research Affairs director, and Marg Burggren, also of the Office of Research Affairs, the University of Massachusetts, for their contributions during our work with GRANT. Thomas Gruber and Michael Greenberg provided valuable technical assistance. Finally, we thank the following researchers who have worked on the GRANT system: A. Davis, D. Day, J. Delisio, S. Lander, C. Loiselle, and D. Suthers.

The authors can be reached in care of Paul Cohen, at the Experimental Knowledge Systems Lab, Dept. of Computer and Information Science, University of Massachusetts, Amherst, MA 01003.

References

1. B. Chandrasekaran, "Generic Tasks in Knowledge-Based Reasoning: High-Level Building Blocks for Expert System Design," *IEEE Expert*, Fall 1986, pp. 23-30.

2. F. Hayes-Roth, "The Role of Partial and Best Matches in Knowledge Systems," in *Pattern Directed Inference Systems*, D. Waterman, D. Hayes-Roth, and D. Lenat, eds., Academic Press, New York, N.Y., 1978.

3. A. Tversky, "Features of Similarity," *Psychological Review*, Vol. 84, No. 4, 1977.

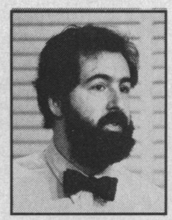
4. P.R. Cohen et al., "Representativeness and Uncertainty in Classification Systems," *AI Magazine*, Fall 1985.

5. W.S. Clancey, "Classification Problem Solving," *Proc. AAAI*, Menlo Park, Calif., 1984, pp. 49-55.

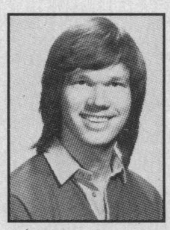
6. P. Cohen et al., "Management of Uncertainty in Medicine," to be published in the *Int'l J. Automated Reasoning*, Spring 1987.

7. P. Szolovits and S.G. Pauker, "Categorical and Probabilistic Reasoning in Medical Diagnosis," *Artificial Intelligence*, Vol. 11, 1978, pp. 115-144.

8. P.R. Cohen and R. Kjeldsen, "Information Retrieval by Constrained Spreading Activation in Semantic Networks," *Information Processing and Management*, Spring 1987.



Paul Cohen is an assistant professor of computer science and director of the experimental knowledge systems lab at the University of Massachusetts, Amherst. He did his graduate work at Stanford University, where he also edited a volume of the *Handbook of Artificial Intelligence*. Cohen's research is on managing uncertainty. Specifically, he is designing architectures for knowledge systems that know enough about the sources and consequences of uncertainty, and the cost and availability of evidence to be able to plan sequences of actions to minimize uncertainty and its consequences.



Rick Kjeldsen is a staff programmer for the AI systems group at IBM's Thomas J. Watson Research Center. He received his BSEE with distinction from Clarkson University in 1981 and served as a development engineer with IBM for three years. He received his MS in computer science from the University of Massachusetts, where he was a research assistant with the experimental knowledge systems lab. His interests include machine learning, automated reasoning, and the management of uncertainty.