PROCEEDINGS OF THE SIXTH INTERNATIONAL WORKSHOP ON MACHINE LEARNING

Cornell University
Ithaca, New York

June 26–27, 1989

EDITOR/WORKSHOP CHAIR

Alberto Maria Segre, *Cornell University*

ORGANIZING COMMITTEE

Jaime Carbonell, *Carnegie Mellon University*
Gerald DeJong, *University of Illinois at Urbana-Champaign*
Tom Dietterich, *Oregon State University*
John Laird, *University of Michigan*
Pat Langley, *University of California at Irvine*
Ryszard Michalski, *George Mason University*
Tom Mitchell, *Carnegie Mellon University*
Jack Mostow, *Rutgers University*
J. Ross Quinlan, *University of Sydney*
Paul Rosenbloom, *University of Southern California*

PROGRAM COMMITTEE

Steve Chien, *University of Illinois at Urbana-Champaign*
Gerald DeJong, *University of Illinois at Urbana-Champaign*
Doug Fisher, *Vanderbilt University*
Allen Ginsberg, *AT&T Bell Laboratories*
John Grefenstette, *Naval Reserach Laboratory*
Michael Pazzani, *University of California at Irvine*
Claude Sammut, *University of New South Wales*
Jeff Schlimmer, *Carnegie Mellon University*
Devika Subramanian, *Cornell University*
Rich Sutton, *GTE Laboratories*
Paul Utgoff, *University of Massachusetts at Amherst*

SPONSORS

Office of Naval Research, Artificial Intelligence and Robotics Program
Office of Naval Research, Cognitive Science Program
National Science Foundation, Knowledge Models and Cognitive Systems Program
Cornell University Department of Computer Science

MORGAN KAUFMANN PUBLISHERS, INC.
SAN MATEO, CALIFORNIA

CONTROLLING SEARCH FOR THE CONSEQUENCES OF NEW INFORMATION DURING KNOWLEDGE INTEGRATION[1]

Kenneth S. Murray and Bruce W. Porter
Department of Computer Sciences
University of Texas, Austin, TX 78712

ABSTRACT

Adding new information to an existing knowledge base can have significant consequences. For example, new information might contradict existing knowledge or reveal a "gap" in the knowledge base. Most approaches to knowledge-base refinement either ignore these consequences or compute them exhaustively. Our approach, formalized in a task called *knowledge integration*, is to partially elaborate the consequences of new information. A form of domain knowledge called *views* controls the search to identify non-superficial consequences of new information. A prototype knowledge integration program has been implemented and demonstrated with a complex extension to a large knowledge base.

INTRODUCTION

Knowledge integration is the task of incorporating new information into a knowledge base. It requires elaborating new information and resolving inconsistencies with existing knowledge. The complexity of knowledge integration is due to the numerous, subtle consequences of new information for existing knowledge. Controlling the search for these consequences is the focus of this paper.

The importance of researching knowledge integration has become apparent while building a task-independent knowledge base in the domain of botany. The complexity of adding new information grows with the size of the knowledge base. For example, adding the new information that carbon dioxide is a raw material in photosynthesis has significant consequences for the existing knowledge. Some consequences identify missing information, such as the mechanism for acquiring the raw material. Others reveal anomalies in the knowledge base, such as the conclusion that plant embryos in seeds must die because they lack carbon dioxide. Determining these consequences reveals inconsistencies and "gaps" in the knowledge base.

Controlling the search for the consequences of new information has received little attention in research on knowledge-base refinement. Some approaches simply add new information and ignore its consequences, *e.g.,* [1]. At the other extreme, some approaches compute the complete reduction of the knowledge base to detect inconsistencies [5, 6]. This is an exponential calculation and is not feasible for many knowledge bases. Between these extremes, approaches have been limited to detecting "surface" inconsistencies [4, 12]. FIE [2] improves on these approaches by using stylized resolution to determine the "shallow" interaction between new information and existing beliefs. However, this approach lacks sufficient control to integrate extensions into a large knowledge base or to identify the "deep" consequences of new information.

We are exploring an approach to controlling the search for the consequences of new information that uses a form of domain knowledge called *views*. Each view defines a segment of the knowledge base comprised of concepts that interact in some significant way. Views are used to heuristically guide search during knowledge integration by identifying the inference paths worth pursuing when the representation of a concept is extended with new information.

The "laboratory" for this research is the Botany Knowledge Base [11] which contains task-independent knowledge about plant anatomy, physiology, and development. Our group has been constructing the knowledge base for two years; it currently contains 4000 frames.

In this paper we describe the knowledge integration task and our approach to controlling search. Section 2 describes our learning program, KI, with an example of a complex extension to the Botany Knowledge Base. Section 3 describes our proposed use of views and a task agenda to guide search.

[1] Support for this research was provided by the Army Research Office under grant ARO-DAAG29-84-K-0060 and the National Science Foundation under grant IRI-8620052.

Figure 1: Learning About Nonendospermic Seeds

Teacher: There is a class of seeds that have no endosperm.

ProtoKI: These seeds must not give rise to healthy seedlings.

Teacher: Explain

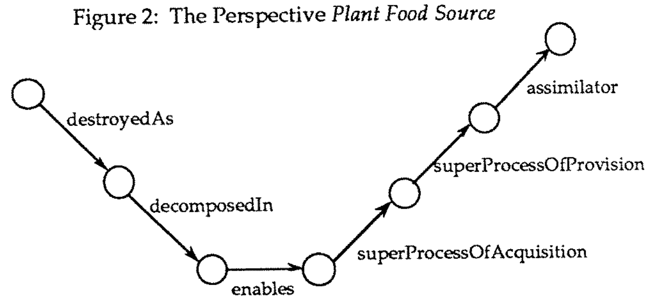
ProtoKi: The embryo starves since the endosperm provides nutrients, and nutrients are essential for survival.

Teacher: Yes, nutrients are essential; however, the embryo survives.

ProtoKI: Does this embryo get nutrients from photosynthesis?

Teacher: Yes, the shoots of some plant embryos emerge from the seed during germination and start producing photosynthate.

Figure 2: The Perspective *Plant Food Source*



This semantic-net template defines the view relevant to an object in its role as a plant food source: a plant food source must have a stage when it is destroyed and decomposed into nutrients. This decomposition enables the nutrients to be assimilated by the plant. Nutrient assimilation involves the provision and acquisition of nutrients.

KI: A TOOL FOR KNOWLEDGE INTEGRATION

We are developing KI, a knowledge acquisition tool that helps a knowledge engineer extend a knowledge base. When new information is provided, KI uses the existing knowledge to critique the new information and determine its consequences. Our computational model of knowledge integration includes three prominent activities:

1. Recognition: identifying the knowledge relevant to new information.
2. Elaboration: applying the expectations provided by relevant knowledge to determine the consequences of the new information.
3. Adaptation: modifying the knowledge base to accommodate the elaborated information.

Figure 1 presents an implemented example that involves extending the Botany Knowledge Base with new information about plant seeds. PROTOKI is a prototype implementation of KI that has been successfully tested with this example.[2] The knowledge base already has information that plant seeds contain nutritive tissue called endosperm, which is analogous to an egg's yolk. The plant embryo consumes the endosperm during its development inside the seed. A knowledge engineer wishes to extend the knowledge base with a representation of *nonendospermic seeds*, a type of seed that contains no endosperm. The task of PROTOKI is to interact with the knowledge engineer to integrate this new information.

As this example illustrates, PROTOKI integrates new information by determining its consequences. When conflicts are encountered, PROTOKI searches for alternative explanations to resolve them. The computational issues that arise during knowledge integration include identifying knowledge relevant to new information, relating the relevant knowledge to new information, and adapting the knowledge base to accommodate the new information. The following three sections discuss these issues in greater detail.

Recognition

PROTOKI begins knowledge integration by identifying relevant knowledge structures. In the previous example about seeds with no endosperm, PROTOKI must determine which among the thousands of frames in the Botany Knowledge Base may be affected in some interesting way.

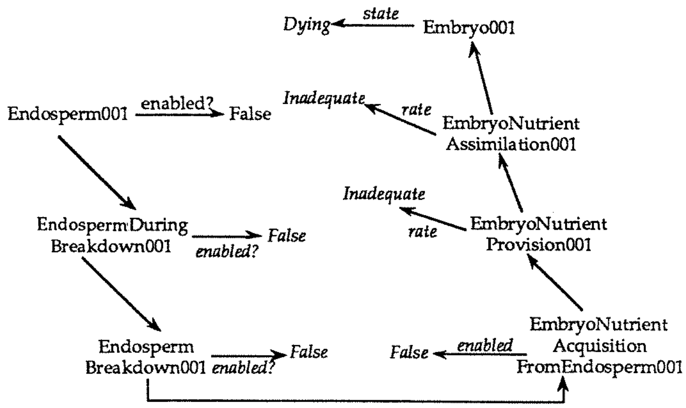
To focus the search for knowledge relevant to new information, the representation of each object in the Botany Knowledge Base is structured with *views*. Each view is a segment of the knowledge base that identifies concepts which interact in some significant way. Perspectives are a common type of view that represent concepts in particular roles. For example, one perspective of endosperm is *Plant Food Source*, as shown in Figure 2. Other perspectives include: endosperm as a *Product Of Reproduction*, endosperm as a *Contained*

[2] PROTOKI does not generate and parse natural language; this example has been converted from a language of frames, slots, and values. This example is also described in [8] with implementation details in [9].

Figure 3: Heuristic Rules Relevant to Endosperm as a Plant Food Source

1. When an entity is disabled, all of its developmental stages are disabled.
2. When an entity is disabled, all the processes involving the entity are disabled.
3. When a process is disabled, all the processes that its completion enables are disabled.
4. When the known methods of acquiring some essential resource are disabled, the rate of "provisioning" the resource is inadequate for survival.
5. When the assimilation rate for some resource is unknown, it is the same as the provision rate.
6. When nutrient assimilation is inadequate for survival, the assimilator is dying.

Figure 4: The Elaborated Context



The hypothetical endosperm is disabled, triggering the inference rules of Figure 3 which propagate the effects of this assertion throughout the context. The predicted consequences of seeds' lacking endosperm are presented in italics.

Object, and endosperm as a *Taxon Defining Part.* PROTOKI collects the views defined for objects referenced by new information and prompts the knowledge engineer to select which are appropriate.

A view is a semantic-net template that can be instantiated for hypothetical objects. PROTOKI instantiates the views selected by the knowledge engineer. Collectively, these instantiated frames comprise a *context* representing an endosperm in its role as a plant food source; this context is used to simulate the effects of the new information about endosperm.

Elaboration

During recognition, PROTOKI creates a context by instantiating concepts in the knowledge base most relevant to the new information. Next, during elaboration, PROTOKI determines how the new information interacts with existing knowledge within this context. Elaboration involves applying inference rules to propagate the effects of the new information throughout the context.

In the endosperm example, elaboration begins when PROTOKI asserts that the endosperm is absent from the context by assigning value *False* to the slot *enabled?* of *Endosperm 001.* This assignment triggers inference rules that determine the consequences of seeds lacking endosperm. For example, without the endosperm, the embryo cannot get enough nutrients to survive. The inference rules applicable to this example are listed in Figure 3, and the elaborated context is presented in Figure 4.

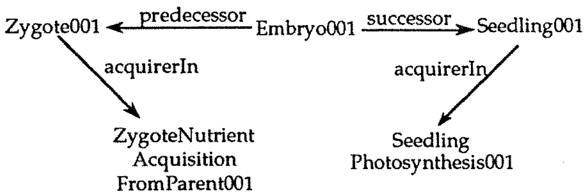
Through elaboration, PROTOKI concludes that the plant embryo is dying from lack of nutrients. This triggers the instantiation of a second view defined for plants that are starving and in danger of dying. The original context is expanded to include the plant's developmental stages immediately before and after its embryo stage and how nutrients are acquired during each of these developmental stages. This additional knowledge is presented in Figure 5. Through continued elaboration, PROTOKI concludes that the plant's seedling stage is not reached because the plant dies during its embryo stage.

An important function of elaboration is identifying confounded expectations. These occur when expectations of the knowledge base are violated by new information or when two rules reach conflicting conclusions. Resolving inconsistencies involves correcting the new information to comply with current expectations or adapting the existing knowledge structures to accommodate the new information.

Adaptation

Elaboration reveals anomalies in the knowledge base; adaptation resolves them. An anomaly can result from inconsistencies introduced either by inference rules used during elaboration or by facts the knowledge engineer

Figure 5: The Context Extension



This is the extension to the context of Figure 4, showing the developmental predecessor and successor of *Embryo 001* and their methods of nutrient acquisition

Figure 6: The Suspect Explanation

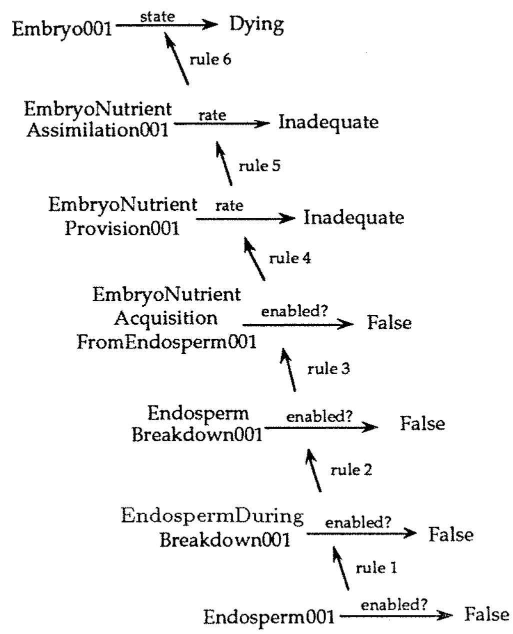
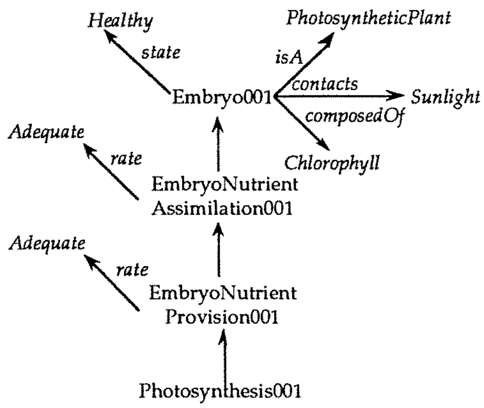


Figure 7: The Adapted Context



The context is adapted to account for adequate nutrient provision when no nutrients are acquired from the endosperm. Assuming the plant embryo acquires nutrients through photosynthesis requires that it contacts sunlight and possesses chlorophyll.

asserts. In the endosperm example, an anomaly is detected when the knowledge engineer asserts that the embryos of nonendospermic seeds survive, correcting the prediction that these embryos starve.

Resolving anomalies requires correcting explanations that support failed expectations and constructing alternative explanations to account for new information. When the knowledge engineer refutes the prediction that embryos of nonendospermic seeds starve, PROTOKI inspects the explanation for this prediction to determine its weakest premise. This suspect explanation is presented in Figure 6. Rule 4 (from Figure 3) relies on a closed-world assumption and is considered a relatively weak inference. Therefore, PROTOKI retracts its conclusion and assumes *Embryo Nutrient Provision 001* is adequate for the embryo's survival. This change propagates through the explanation, retracting the belief that the embryo starves.

The original anomaly has been resolved by assuming adequate nutrient provision by the embryos of nonendospermic seeds. However, no alternative method is known for plant embryos to acquire nutrients. PROTOKI seeks to construct an explanation for the assumed nutrient acquisition using the following inference:

If a resource provision is adequate for survival, but no acquisition method is known,
then assume the acquisition method of the developmental successor is employed.

This rule suggests the embryos of nonendospermic seeds acquire nutrients by photosynthesis, as is done by seedlings. However, this hypothesis introduces new constraints on the embryos of nonendospermic seeds. For example, to acquire nutrients by photosynthesis, the embryo must be a photosynthetic plant. Therefore, to apply this inference, PROTOKI asserts that *Embryo 001* is an instance of *Photosynthetic Plant*. As a photosynthetic plant, the embryo inherits the following features: its color is green, it contacts sunlight, and its composition includes chlorophyll. This is illustrated in Figure 7. In short, the plausibility of explaining the survival of nonendospermic embryos by assuming they engage in photosynthesis is contingent on their contacting sunlight and possessing chlorophyll. Validating these assumptions leads to the acquisition of further knowledge from the knowledge engineer.

CONTROLLING THE SEARCH FOR CONSEQUENCES

The preceding example illustrates how a tool for knowledge integration can identify consequences of new information and acquire additional knowledge to fill gaps and resolve inconsistencies. The challenge of performing knowledge integration is controlling the search for the consequences of new information for existing knowledge. Our prototype implementation uses views to determine which inference paths to pursue; however, it relies on the user to select among candidate views. We are currently exploring approaches to automate view selection. Two existing schemes to control search for plausible extensions to existing knowledge include using a notion of "interestingness" and imposing structural constraints on the path that each line of reasoning is allowed to form.

In AM [7] the property of interestingness is used to constrain search for new concepts. As new operations are proposed (e.g., finding the value of some frame's slot), a measure of how interesting the completion of the operation is likely to be is computed by combining the innate interestingness of the frame and slot with the interestingness of the context that proposes the operation. Operations are selected in order of their interestingness. Similarly, FIE [2] uses interestingness (i.e., mathematical simplicity) to restrict elaboration of clauses added to a theorem prover.

In GRANT [3] structural constraints are imposed on paths of plausible inference. Each constraint, called a path endorsement, prohibits or promotes the coupling of specified links. For example, an endorsement might promote combining *cigarette hasPart tobacco* and *tobacco causes cancer* to conclude *cigarette causes cancer*, while a second endorsement might prohibit combining *boat floats* and *boat hasPart anchor* to conclude *anchor floats*.

KI's method of controlling search involves elements of both approaches. We propose to conduct a best-first search of the space of plausible conclusions, ordered by interestingness. However, rather than assessing interestingness of candidate inferences as each inference is selected, we propose selecting a very limited segment of the knowledge base and allowing all possible inference paths to be constructed within the frames contained in the selected segment.

Each segment corresponds to a view and denotes a set of frames in the knowledge base that interact in some significant way. Views are structurally similar to endorsements, represented as a graph of paths emanating from a concept. However, a view defined for one concept need not be inherited to all other concepts.

When an extension is made to a concept, KI activates one or more of the views defined for the concept. Plausible rules of inference are limited to the frames within the activated views. An agenda determines which view, if any, will be selected for activation next. Agenda tasks are ordered by a heuristic measure of "interestingness," which is a function of the activation level of views and the conflict level. Activation level measures the degree to which a view overlaps the frames comprising the current instantiated context. Conflict level measures the confidence in conflicting beliefs and the degree to which they conflict. The processing cycle continues until the agenda is empty (i.e., no task has an interest rating above a minimum threshold) or the user suspends KI.

This approach to knowledge refinement uses views for two purposes. First, views provide a coarse granularity during the search for deep consequences of new information. Each cycle of the agenda selects a view for activation and applies all inferences defined within the view. Interestingness is assessed only after fleshing-out the highly interdependent frames within separable segments of the knowledge base, rather than after firing each inference rule. This has the advantage of sufficiently developing each context to determine the most interesting direction to pursue. Second, views define local, computational environments. We are developing KI to enforce consistency of the knowledge-base within views. This policy operationalizes the adage of maintaining *local consistency* and avoids computing the deductive closure of the knowledge base.

Our current approach to representing views has limited flexibility. This raises concerns about the number of views required to structure a knowledge base and the convenience of defining them explicitly. We are researching methods to acquire new views automatically during knowledge integration, and our colleagues are researching techniques to represent views implicitly [10].

SUMMARY

Knowledge integration is the task of incorporating new information into a knowledge base. This is an important and difficult task because the interactions between new information and existing knowledge can be numerous and subtle. Current approaches to knowledge refinement ignore subtle interactions.

We are exploring an approach to controlling the search for the consequences of new information that uses a form of domain knowledge called *views*. Each view defines a segment of the knowledge base comprised of concepts that interact in some significant way. Views are used to heuristically guide search during knowledge integration by identifying the inference paths worth pursuing when the representation of a concept is extended with new information.

We are testing our approach to knowledge integration with a tool that helps with the arduous task of extending a task-independent knowledge base. The tool identifies the consequences of new information and acquires additional knowledge when a gap or anomaly is revealed. An initial prototype of this tool has been implemented and tested with a complex extension to the Botany Knowledge Base.

References

[1] E.R. Bareiss, B. Porter, and C. Wier. Protos: An exemplar-based learning apprentice. In Pat Langley, editor, *Proceedings of the Fourth International Workshop on Machine Learning*, pages 12–23. Morgan Kaufmann, Los Altos, California, 1987.

[2] D. Cohen. A forward inference engine to aid in understanding specifications. In *Proceedings of the National Conference on Artificial Intelligence*, pages 56–60, 1984.

[3] P. Cohen, A. Davis, D. Day, M. Greenberg, R. Kjeldsen, S. Lander, and C. Loiselle. Representativeness and uncertainty in classification systems. *AI Magazine*, 6(3):136–149, 1985.

[4] R. Davis. Interactive transfer of expertise: Acquisition of new inference rules. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 321–328, 1977.

[5] A. Ginsberg. Knowledge-base reduction: A new approach to checking knowledge bases for inconsistency & redundancy. In *Proceedings of the National Conference on Artificial Intelligence*, pages 585–589, 1988.

[6] A. Ginsberg. Theory revision via prior operationalization. In *Proceedings of the National Conference on Artificial Intelligence*, pages 590–595, 1988.

[7] D. Lenat. *AM: An Artificial Intelligence Approach to Discovery in Mathematics as Heuristic Search.* PhD thesis, Computer Science Department, Stanford University, 1976.

[8] K. Murray. KI: An experiment in automating knowledge integration. Technical Report AI-TR-88-90, Department of Computer Sciences, University of Texas at Austin, 1988.

[9] K. Murray and B. Porter. Developing a tool for knowledge integration: Initial results. In *Proceedings of the Knowledge Acquisition for Knowledge-based Systems Workshop*, 1988.

[10] B. Porter, L. Acker, J. Lester, and A. Souther. Generating explanations in an intelligent tutor designed to teach fundamental knowledge. In *Proceedings of the Second Intelligent Tutoring Systems Research Forum*, pages 55–70, 1989.

[11] B. Porter, J. Lester, K. Murray, K. Pittman, A. Souther, L. Acker, and T. Jones. AI research in the context of a multifunctional knowledge base: The Botany Knowledge Base project. Technical Report AI-TR-88-88, Department of Computer Sciences, University of Texas at Austin, 1988.

[12] D.C. Wilkens. Knowledge base refinement using apprenticeship learning techniques. In *Proceedings of the National Conference on Artificial Intelligence*, pages 646–651, 1988.