Extracting Viewpoints from Knowledge Bases*

Liane Acker
IBM Corporation
11400 Burnet Road
Austin, Texas 78758
acker@austin.ibm.com

Bruce Porter
Department of Computer Sciences
University of Texas at Austin
Austin, Texas 78712
porter@cs.utexas.edu

Abstract

Viewpoints are coherent collections of facts that describe a concept from a particular perspective. They are essential for a wide variety of tasks, such as explanation generation and qualitative modeling. We have identified many types of viewpoints and developed a program, the View Retriever, for extracting them from knowledge bases, either singly or in combinations. The View Retriever provides a general solution to the central problem in extracting viewpoints: determining which facts are relevant to requested viewpoints. Our evaluation indicates that viewpoints extracted by the View Retriever are comparable in coherence to those people construct.

1 Introduction

The objective of this research is to develop computational methods for extracting viewpoints from knowledge bases. Intuitively, a viewpoint is a coherent collection of facts that describes a concept from a particular perspective. For example, three viewpoints of the concept "car" are: the viewpoint "car as-kind-of consumer durable," which describes a car's price and longevity; the structural viewpoint, which describes a car's parts and their interconnections; and the viewpoint "car as-having metal composition," which includes facts, such as a car's propensity to dent and rust, that are related to its composition.

The need for viewpoints by knowledge-based programs is widespread. For example, many explanation-generation systems require viewpoints to produce explanations that are complete and coherent (Suthers 1991; McKeown 1988; Lester & Porter 1991; McCoy 1989; Moore & Swartout 1988). Qualitative modeling systems use viewpoints to increase efficiency and to make consistent modeling assumptions (e.g.,the *model fragments* of (Falkenhainer & Forbus 1991), the *views* of (Forbus 1984), and the *ontological perspectives* of (Liu & Farley 1990).) Finally, KI (Murray & Porter 1989), a learning program, uses viewpoints to constrain its search for the consequences of adding new information to a knowledge base.

Conventional methods for accessing knowledge bases do not provide direct access to viewpoints. Some methods extract individual facts, such as the filler of a particular frame-slot. Others extract collections of facts, such as all the slots and fillers of a particular frame or those satisfying a Prolog-like query. Indisputably, these access methods can be used to extract viewpoints through a sequence of invocations. However, they ignore the central problem in extracting viewpoints: determining *which* facts to include in a viewpoint. The advantage of our access methods is that they provide a general solution to this problem (as described in Section 2), and the viewpoints extracted by our methods are comparable in coherence to those people construct (as described in Section 3).

2 The View Retriever

Our methods for accessing viewpoints are implemented in a program called the *View Retriever* (a term first proposed by Suthers (Suthers 1988)). The input to this program is a viewpoint specification and the output is a collection of facts. The task of the View Retriever is to determine which facts constitute the specified viewpoint and to request them from the knowledge base. Whether the knowledge base returns cached facts or computes them (using deduction, abduction, or induction) is irrelevant to the View Retriever. Those facts that the knowledge base cannot provide are not included in the viewpoint.

The View Retriever is used currently with the Botany Knowledge Base, a large system of over 13,000 frames and 160,000 cached facts, where a fact is a slot-filler of a frame. However, it is designed to work for any physical domain and to be easily extended to work in non-physical domains, such as those involving abstract concepts or mental processes.

*Support for this research was provided by an IBM Graduate Fellowship to Liane Acker, a grant from the National Science Foundation (IRI-9120310), a contract from the Air Force Office of Scientific Research (F49620-93-1-0239), and donations from the Digital Equipment Corporation. This work was conducted at the University of Texas at Austin.

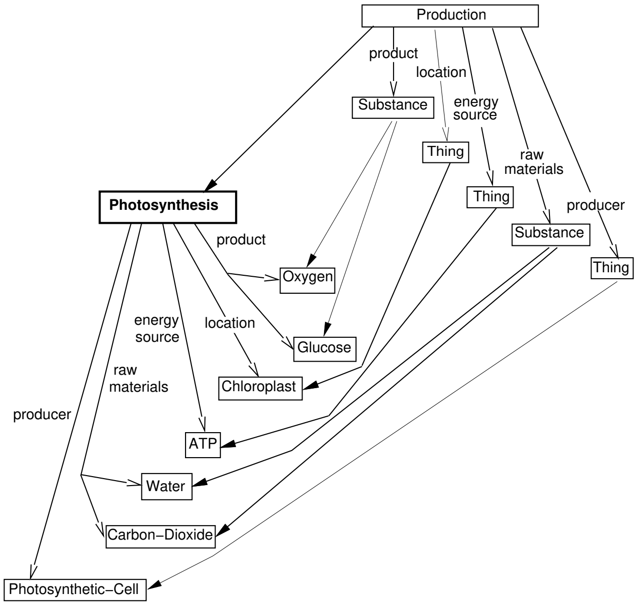

Figure 1: The viewpoint of "photosynthesis *as-kind-of* production", as extracted from the Botany Knowledge Base by the View Retriever.

The way a user (or application program) specifies a viewpoint and the way the View Retriever extracts it depends on the type of viewpoint. *As-kind-of* viewpoints describe concepts by relating them to more general concepts. Viewpoints constructed along *basic dimensions* describe concepts using a cluster of their attributes, such as functional, structural, or perceptual attributes. *As-having* viewpoints include the facts pertinent to a given attribute.

As-kind-of Viewpoints

An *as-kind-of* viewpoint describes a concept in terms of a more general concept. For example, the viewpoint "photosynthesis *as-kind-of* production" consists of those facts that explain how photosynthesis is a special case of production, such as its raw materials and products. Figure 1 shows a portion of this viewpoint.

The specification of an *as-kind-of* viewpoint is of the form:

$$(\langle\text{primary concept}\rangle\ \textit{as-kind-of}\ \langle\text{reference concept}\rangle)$$

where the primary concept is the one the viewpoint will be taken of and the reference concept is a generalization of the primary concept (although not necessarily an immediate generalization).

The View Retriever extracts *as-kind-of* viewpoints by selecting relevant facts about the primary concept. A fact is a tuple of the form $\langle slot, filler\rangle$; it is considered relevant if some more general fact appears on the frame for the reference concept. The fact $\langle slot', filler'\rangle$

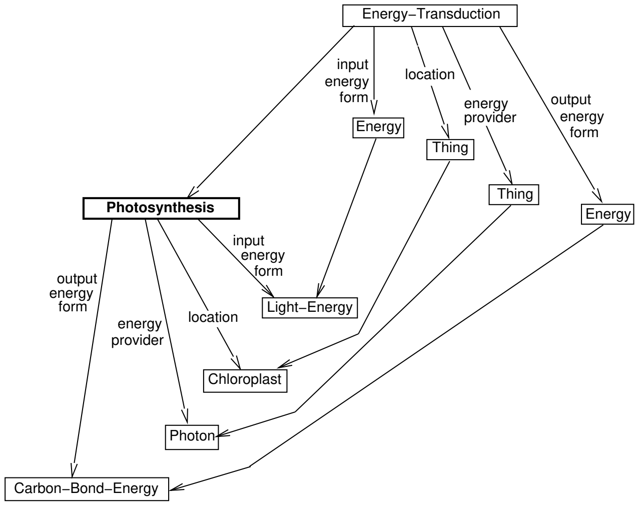

Figure 2: The viewpoint of "photosynthesis *as-kind-of* energy transduction", as extracted from the Botany Knowledge Base by the View Retriever.

is more general than $\langle slot, filler\rangle$ if any of the following conditions hold:

1. $slot = slot'$ and $filler'$ is a generalization of $filler$.
2. $filler = filler'$ and $slot'$ is a generalization of $slot$.
3. $slot'$ is a generalization of $slot$ and $filler'$ is a generalization of $filler$.

For example, the viewpoint shown in Figure 1 contains the fact that photosynthesis produces glucose, because it is known that production processes typically produce some substance and glucose is a special kind of substance. That is, $\langle product, Glucose\rangle$ appears on the *Photosynthesis* frame, $\langle product, Substance\rangle$ appears on the *Production* frame, and *Substance* is a generalization of *Glucose*. The resulting viewpoint includes the links between facts about the primary concept and the more general facts about the reference concept (see Figure 1).

The View Retriever excludes many facts about the primary concept from the viewpoint. For example, although it is true that photosynthesis converts light energy into carbon bond energy, this fact is excluded because it is irrelevant to our concept of production (although it *would* be included in "photosynthesis as-kind-of energy transduction", as shown in Figure 2).

Various explanation-generation systems extract knowledge structures similar to *as-kind-of* viewpoints. The TEXT system (McKeown 1985) uses a function (called the *identification rhetorical predicate*) to differentiate a concept from a more general concept. TEXT determines what facts to include using a type of knowledge called *focus constraints*: facts are selected incrementally based on their connection with previously selected facts, rather than a global coherence criteria. Suthers's system uses a *genus-and-differentia* function similar to TEXT's identification predicate (Suthers 1991). McKeown's ADVISOR system constructs knowledge structures similar to *as-kind-of* viewpoints by restricting to predefined partitions of the knowledge base the superconcepts from which a concept can inherit slot fillers (McKeown 1988).

Viewpoints Constructed Along Basic Dimensions

In addition to viewpoints that describe concepts in terms of more general concepts, the View Retriever can extract viewpoints along *basic dimensions*, which are general types of facts, such as facts about an object's structure, function, or appearance. (We have borrowed the term from *Metaphors We Live By* (Lakoff & Johnson 1980), a work that has significantly influenced our characterization of viewpoint types.) Below we describe the basic dimensions used by the View Retriever.

Basic dimensions for objects:

- **Structural**, which includes the parts or substances that make up the object. It also includes the connections and spatial relations among them, what we call *interconnection relations*. The structural dimension also includes the relative sizes or number of the parts.
- **Perceptual**, which includes information regarding how humans perceive (see, hear, etc.) the object. This includes the shape, symmetry, size, color, and temperature of the object.
- **Functional**, which includes what the object "does" (the processes in which it is an actor). The functional dimension also includes properties suggestive of some unspecified process in which the object is involved, such as *life span* and *metabolic rate*.
- **Temporal**, which includes the temporal parts of an object (its stages or states). It also includes as interconnection relations the temporal ordering constraints among the stages or states.

Basic dimensions for processes:

- **Behavioral**, which includes the types and roles of the actors in the process and the changes that the process effects upon them. Initial and final conditions of the process are included as well.
- **Procedural**, which includes the steps (subevents) of the process and (as interconnection relations) any temporal ordering constraints that exist among the steps.

Basic dimensions for both objects and processes:

- **Taxonomic**, which includes the taxonomic breakdown of a class of objects or processes into subclasses. The taxonomic dimension also includes the relative sizes of the subclasses, the criteria for the breakdown, and (as interconnection relations) information about which subclasses are disjoint.
- **Modulatory**, which includes information about how one object or process affects other objects or processes. This includes causal relationships (e.g.,causes, enables, prevents, facilitates) and qualitative influences between quantities (e.g.,directly-affects, inversely-influences, correlated-with).

The specification for a viewpoint constructed along a basic dimension simply names the primary concept and the basic dimension desired:

$$(\langle \text{primary concept} \rangle \ \textit{dimension} \ \langle \text{basic dimension} \rangle)$$

The View Retriever constructs the viewpoint first by extracting facts about the primary concept that belong to the basic dimension, then by adding to the viewpoint any interconnection relations for the basic dimension. For example, to construct a structural viewpoint of a plant seed, the View Retriever first selects those slots and fillers from the Seed frame that belong to the structural dimension, including ⟨part, Seed-Coat⟩, ⟨part, Embryo⟩, and ⟨part, Endosperm⟩. The View Retriever then selects interconnection relations among the selected parts (seed coat, embryo, and endosperm). For the structural dimension, interconnection relations include *connected-to, contains, surrounds*, etc. Thus, the resulting viewpoint contains the information that the seed is made up of a seed coat containing an embryo and an endosperm.

To construct viewpoints along basic dimensions, the View Retriever uses knowledge of which slots in the knowledge base are within each dimension. Based on our experience with the Botany Knowledge Base, this knowledge is easily encoded because the distinctions made by the basic dimensions are reflected in the top levels of the slot hierarchy.

Viewpoints created by the View Retriever along basic dimensions are similar to *perspectives* as suggested by Suthers (Suthers 1991) and as used by Romper (McCoy 1989). Unlike our basic dimensions, however, Romper's perspectives are domain-specific and include only facts about the primary concept; interconnection relations are omitted.

As-Having Viewpoints

An *as-having* viewpoint contains all and only the information about a concept that is relevant to some specified fact about the concept. Its specification has the following form:

$$(\langle \text{primary concept} \rangle \ \textit{as-having} \ \langle \text{slot, filler} \rangle)$$

To our knowledge, general methods do not exist for extracting *as-having* viewpoints. Therefore, unlike for the other types of viewpoints, the View Retriever depends on *a priori* knowledge of relevance to select the facts that constitute *as-having* viewpoints.

To construct an *as-having viewpoint*, the View Retriever first looks for a cached *as-having* viewpoint that is based on the same fact (slot and filler), or a more general fact, as the requested viewpoint, but with a different primary concept. For example, to extract the viewpoint:

(Squirrel *as-having* ⟨agent-in, Seed-Dispersal⟩)

the View Retriever first looks in the knowledge base for a related, cached viewpoint such as one of the following:

1. (Animal *as-having* ⟨agent-in, Seed-Dispersal⟩)
2. (Bird *as-having* ⟨agent-in, Seed-Dispersal⟩)
3. (Animal *as-having* ⟨agent-in, Transportation⟩)

If a related viewpoint is found, the View Retriever uses it to determine which facts should be included in the new viewpoint. It does this by finding for each fact of the cached viewpoint a corresponding fact that is true of the primary concept of the new viewpoint. If the primary concept of the cached viewpoint is a *generalization* of the primary concept of the new viewpoint, then finding corresponding facts between the two consists of finding facts about the primary concept of the new viewpoint that are *specializations* of facts in the cached viewpoint. If the primary concepts of the two viewpoints are *siblings*, then finding corresponding facts between the two is more difficult. It requires finding pairs of facts that share a common abstraction.

If a related, cached viewpoint cannot be found in the knowledge base, then the View Retriever constructs *as-having* viewpoints by collecting all the facts about the primary concept that are implied by the specified fact, using all the inference rules and mechanisms available in the knowledge base. This method assumes (sometimes incorrectly) that any fact implied by some other fact is relevant to it. However, it has the advantage that it does not require viewpoints to be cached in the knowledge base.

Ideally, *as-having* viewpoints would be extracted using a theory of relevance to determine what facts are relevant. As a first step toward such a theory, several researchers have analyzed texts to determine the various ways that one fact may be relevant to another (Mann & Thompson 1987; Hobbs 1985). However, these theories are as yet descriptive rather than prescriptive, so the View Retriever cannot use them directly.

Composite Viewpoints

In addition to extracting individual viewpoints as described above, the View Retriever can combine them to form composite viewpoints. This involves more than simply concatenating the contents of two individual viewpoints; it involves putting them into correspondence with one another and removing the portions that do not correspond. Despite the apparent utility of composite viewpoints, we know of no other general methods for extracting them from knowledge bases.

The specification for a composite viewpoint has the following form:

(*composite* ⟨viewpoint1⟩ ⟨viewpoint2⟩ ⟨relation⟩)

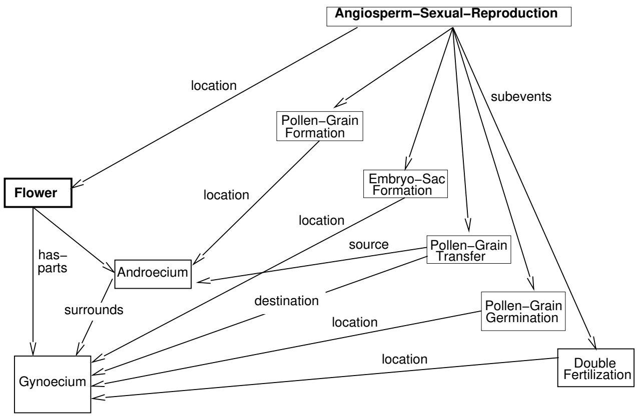

Figure 3: The composite ("structural-functional") viewpoint of a flower in its role in plant reproduction, as extracted from the Botany Knowledge Base by the View Retriever.

where *viewpoint1* and *viewpoint2* are individual viewpoints (or specifications for them) and *relation* specifies the correspondence to be established between the viewpoints.

One commonly used composite viewpoint, called "structural-functional", describes the roles an object (and its parts) play in an event (and its subevents). Its specification is the following:

(*composite* (⟨object⟩ dimension structural) (⟨event⟩ dimension procedural) actor-in)

For example, the viewpoint that describes the roles of a flower's parts in the steps of plant reproduction is specified as follows:

(*composite* (Flower dimension structural) (Plant-Reproduction dimension procedural) actor-in)

Its contents are shown in Figure 3.

The View Retriever constructs this composite viewpoint by the following procedure. First it extracts the two individual viewpoints (the structural viewpoint of Flower and the procedural viewpoint of Plant-Reproduction). Then it determines which parts of the Flower that are in the structural viewpoint are related to Plant-Reproduction or one of its subevents (as given in the procedural viewpoint) by an *actor-in* relation or some more specific relation (such as *location-of*). Those parts, such as the Flower's corolla, that are not actors in the event are omitted from the composite viewpoint. Similarly, those subevents, such as Fruit-Ripening, that do not involve any of the parts in the structural viewpoint of Flower are omitted.

This procedure can extract diverse viewpoints. For example, the composite viewpoint that describes the parts of a plant ovary as related to the parts of the fruit of which it is a developmental stage can be extracted with the following specification:

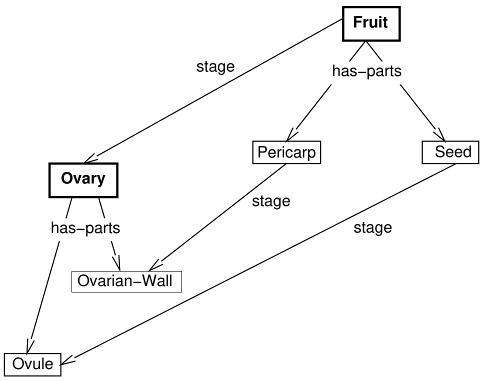

Figure 4: The composite viewpoint of the parts of a plant ovary as related to the parts of the fruit of which it is a developmental stage, as extracted from the Botany Knowledge Base by the View Retriever.

(*composite* (Fruit dimension structural) (Ovary dimension structural) stages)

This composite viewpoint, as shown in Figure 4, includes the parts of the fruit (seed, pericarp, etc.), the parts of the ovary (ovule, ovarian wall, etc.), and the *stage* relations between them, such as the facts that the ovule is a developmental stage of the seed and the ovarian wall is a developmental stage of the pericarp.

The procedure for constructing composite viewpoints can also extract the viewpoint that categorizes angiosperms (flower-bearing plants) according to the different types of flowers they have. The specification is the following:

(*composite* (Angiosperm dimension taxonomic) (Flower dimension taxonomic) parts)

This composite viewpoint includes, for example, the fact that one kind of angiosperm is the orchid, which has an irregular flower.

Evaluation of the View Retriever

The purpose of our evaluation was to measure the coherence of viewpoints the View Retriever extracts, as compared to the coherence of viewpoints found in human-generated text. For each of 12 topics in botany, sets of facts were drawn from 3 sources:

- a college-level botany textbook (Raven, Evert, & Curtis 1976),
- the View Retriever applied to the Botany Knowledge Base, and
- facts selected randomly from a particular frame in the Botany Knowledge Base.

The viewpoints ranged in size from 3 to 11 facts. For each topic, textbook passages and random sets of facts were chosen to be roughly the same size as the viewpoint on that topic. Each group of facts (including the textbook passages) was translated manually into "simple English" to normalize presentation style. The viewpoints included about equal numbers of *as-kind-of*, basic dimension, and composite viewpoints; *as-having* viewpoints were omitted from this study because they often use cached viewpoints.

Source	Coherence	
	Mean	σ
(1) Textbook Viewpoints	4.23	0.56
(2) View Retriever's Viewpoints	3.76	0.74
(3) Degraded Viewpoints	2.86	0.94
(4) Random Collections of Facts	2.62	0.86

Table 1: Ten judges rated the coherence of sets of facts from four sources (1=incoherent; 5=coherent). A statistical analysis using the T-test with 0.95 level of confidence shows no significant difference in coherence between sources (1) and (2) or between sources (3) and (4). There is a significant difference between all other pairs.

Ten subjects (senior undergraduates and graduate students from the Botany and Biology Departments of the University of Texas at Austin) judged the coherence of several passages from each source. The subjects were asked to use a scale of 1 to 5, to assign a passage a score of "1" if it seemed no more coherent than a randomly selected group of facts on the subject, and to assign a passage a score of "5" if it was as coherent as a passage of comparable length on the subject from a good textbook.

Table 1 summarizes the subjects' responses. Statistical analysis (using a T-test with 0.95 level of confidence) yields the following results:

- The mean coherence of viewpoints from textbooks did not differ significantly from the mean coherence of viewpoints extracted by the View Retriever.
- The mean coherence of extracted viewpoints *did* differ significantly from the mean coherence of random collections of facts drawn from the same frame.

A further study gives additional evidence that the View Retriever extracts coherent viewpoints. Along with passages from the three sources described above, the subjects were given passages from a fourth source: viewpoints extracted by the View Retriever and then "degraded" by replacing some of their facts with randomly selected facts on the same topic. Twenty-eight such degraded viewpoints were constructed, each with between one and seven facts replaced. Of the twenty-eight, each subject received six. Table 1 shows the mean coherence score of the degraded viewpoints. Statistical analysis shows a significant difference in the mean coherence of "pure" viewpoints and degraded viewpoints.

A final study adds more evidence that passages vary in coherence based on their source and that viewpoints extracted by the View Retriever are consistently judged to be coherent. A two-way analysis of variance, computed by Paul Cohen[1], determined that there was no significant interaction effect between:

- the variance in coherence scores assigned by different judges, and
- the variance in coherence scores for passages from different sources (e.g.,textbooks, the View Retriever).

Thus, although judges varied in their harshness, they largely agreed on relative orderings.

3 Discussion

Viewpoints are coherent collections of facts that describe a concept from a particular perspective. They are essential for a wide variety of tasks, such as explanation generation and qualitative modeling. We have identified several types of viewpoints and developed a program, the View Retriever, for extracting them from knowledge bases, either singly or in combination. Our evaluation of the View Retriever indicates that its viewpoints are comparable in coherence to those constructed by people.

The View Retriever has several known limitations, some of which we are addressing. First, viewpoint specifications use the names of frames and slots in the knowledge base. Therefore, users of the View Retriever must have extensive knowledge of the concept and slot hierarchies in order to use the View Retriever. To address this limitation, we are developing methods whereby users can specify frames and slots descriptively rather than by name. Second, our textbook analysis reveals that most explanations consist of several viewpoints used in concert. Although the View Retriever can extract composite viewpoints, we have not yet identified which combinations are commonly used. A third limitation is that the View Retriever ignores knowledge about the *a priori* importance of facts. Therefore, it cannot extract viewpoints of a concept in the order of their importance, a potentially useful ability.

The View Retriever will be evaluated more extensively when it supports our tutoring system for plant anatomy and physiology. It will be the primary method used by the tutor to access the Botany Knowledge Base to build qualitative models and generate explanations. We are currently building this tutoring system, and we have found that knowledge base access at the level of viewpoints (as opposed to the level of individual facts or frames) greatly simplifies system design and implementation.

[1] Computer Science Department, University of Massachusetts at Amherst

References

Falkenhainer, B., and Forbus, K. 1991. Compositional modeling: Finding the right model for the job. *Artificial Intelligence* 51:95–143.

Forbus, K. 1984. Qualitative process theory. *Artificial Intelligence* 24:85–168.

Hobbs, J. 1985. On the coherence and the structure of discourse. Technical Report CSLI-85-37, Computer Science Department, Stanford University.

Lakoff, G., and Johnson, M. 1980. *Metaphors We Live By.* University of Chicago Press.

Lester, J., and Porter, B. 1991. A student-sensitive discourse generator for intelligent tutoring systems. In *Proceedings of the International Conference on the Learning Sciences*, 298–304.

Liu, Z., and Farley, A. 1990. Shifting ontological perspectives in reasoning about physical systems. In *Proceedings of the 8th National Conference on Artificial Intelligence.*

Mann, W., and Thompson, S. 1987. Rhetorical structure theory: A theory of text organizations. Technical Report ISI/RS-87-190, Information Sciences Institute, University of Southern California.

McCoy, K. 1989. Generating context-sensitive responses to object-related misconceptions. *Artificial Intelligence* 41:157–195.

McKeown, K. 1985. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text.* Cambridge University Press.

McKeown, K. 1988. Generating goal-oriented explanations. *International Journal of Expert Systems* 1(4):377–395.

Moore, D., and Swartout, W. 1988. A reactive approach to explanation. In *Proceedings of the Fourth International Workshop on Natural Language Generation.*

Murray, K., and Porter, B. 1989. Controlling search for the consequences of new information during knowledge integration. In *Proceedings of the Machine Learning Workshop*, 290–295. Palo Alto, California: Morgan Kaufmann.

Raven, P.; Evert, R.; and Curtis, H. 1976. *Biology of Plants.* New York: Worth Publishers.

Suthers, D. 1988. Providing multiple views of reasoning for explanation. In *Proceedings of the International Conference on Intelligent Tutoring Systems*, 435–442.

Suthers, D. 1991. Task-appropriate hybrid architectures for explanation. In *Proceedings of the AAAI-91 Workshop on Comparative Analysis of Explanation Planning Architectures.*