EXTRACTING VIEWPOINTS FROM MULTIFUNCTIONAL KNOWLEDGE BASES

Liane Acker
Bruce Porter
Department of Computer Sciences
University of Texas at Austin
Austin, TX 78712-1188

HUMAN RESOURCES DIRECTORATE
TECHNICAL TRAINING RESEARCH DIVISION
Brooks Air Force Base, TX 78235-5000

September 1992

Approved for public release; distribution is unlimited.
NOTICES

This technical report is published as received and has not been edited by the technical editing staff of the Armstrong Laboratory.

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

KURT W. STEUCK
Contract Monitor

HENDRICK W. RUCK, Technical Director
Technical Training Research Division

RODGER D. BALLENTINE, Colonel, USAF
Chief, Technical Training Research Division
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 of the View Retriever indicates that it provides most of the types of viewpoints that people use, and its viewpoints are comparable in coherence to those constructed by people.
1 Introduction

Knowledge bases have traditionally been built to competently perform a single task, such as diagnosis or scheduling, and many knowledge bases have met this objective [6]. After a decade of such projects, building knowledge bases to perform individual tasks is fairly routine.

Reflecting the growing maturity of artificial intelligence technology, the potential of competently performing many tasks is the new reason for building a knowledge base [1, 12, 11]. Rather than encode just the knowledge required for one task, a multifunctional knowledge base encodes general knowledge that supports diverse tasks within the domain. For example, a multifunctional knowledge base for a new aircraft might support expert programs for assembly, maintenance, instruction, and design modification.

Building a single knowledge base that supports multiple tasks has two significant advantages over building separate knowledge bases for each task. First, the effort of building a multifunctional knowledge base can be amortized over many expert system projects. Using existing technology (e.g., [27, 3]), multifunctional knowledge bases can be compiled into efficient expert systems for performing disparate tasks within the domain. In contrast, reusing a knowledge base built for a single task is typically infeasible because the knowledge is overly specific. For example, Clancey [4] documents the difficulties in reusing the Mycin medical diagnosis knowledge base for tutoring. The second advantage of multifunctional knowledge bases is a significant reduction in the brittleness of expert systems. Multifunctional knowledge bases contain fundamental domain knowledge that can help solve problems that are beyond the range of task-specific expert systems. For example, Fink [7] uses fundamental knowledge of the structure and function of complex mechanisms to supplement surface-level heuristics for diagnosing faults. Applying the principle on a large scale, the CYC knowledge base is intended to provide a comprehensive body of task-independent knowledge "to provide assistance for expert systems, natural language understanders, and so on, as they get 'stuck' on problems" [13].

Despite these advantages, multifunctional knowledge bases are considerably harder to use than single-task knowledge bases. The problem is extracting the information relevant to solving the problem at hand.

The objective of this research is to develop computational methods for extracting viewpoints from multifunctional 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.
2.1 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 as produced by the View Retriever.

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

\((\text{primary concept}) \text{ as-kind-of } (\text{reference concept})\)

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 \((\text{slot}, \text{filler})\); it is considered relevant if some more general fact appears on the frame for the reference concept. The fact \((\text{slot}', \text{filler}')\) is more general than \((\text{slot}, \text{filler})\) 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, (product, Glucose) appears on the Photosynthesis frame, (product, Substance) 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”).

Various explanation-generation systems extract knowledge structures similar to as-kind-of viewpoints. The TEXT system [18] 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 [26]. 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 [19].

2.2 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 [10], 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:
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 [26] and as used by Romper [17]. Unlike our basic dimensions, however, Romper's perspectives are domain-specific and include only facts about the primary concept; interconnection relations are omitted.

2.3 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:

\[(\text{primary concept}) \text{ as-having } (\text{slot, filler})\]

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:

\[(\text{Squirrel as-having } (\text{agent-in, Seed-Dispersal}))\]
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 [16, 9]. However, these theories are as yet descriptive rather than prescriptive, so the View Retriever cannot use them directly.

2.4 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:
Figure 2: The composite ("structural-functional") viewpoint of a flower in its role in plant reproduction. This viewpoint was extracted from the Botany Knowledge Base by the View Retriever.

\[(\text{composite } (\text{viewpoint1}) (\text{viewpoint2}) (\text{relation}))\]

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:

\[(\text{composite } ((\text{object}) \text{ dimension structural}) ((\text{event}) \text{ dimension procedural}) \text{ actor-in})\]

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

\[(\text{composite } (\text{Flower dimension structural}) (\text{Plant-Reproduction dimension procedural}) \text{ actor-in})\]

Its contents, as extracted from the Botany Knowledge Base by the View Retriever, are shown in Figure 2.
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:

\[ (\text{composite} \ (\text{Fruit dimension structural}) \ (\text{Ovary dimension structural}) \ \text{stages}) \]

This composite viewpoint 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:

\[ (\text{composite} \ (\text{Angiosperm dimension taxonomic}) \ (\text{Flower dimension taxonomic}) \ \text{parts}) \]

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

3 Evaluation

Two claims are central to this research and both have been evaluated. The first is that the View Retriever provides the variety of viewpoints that people use. The second claim is that viewpoints are coherent and that those extracted by the View Retriever are as coherent as ones that people generate. All the parts of this evaluation — data, questionnaire, View Retriever, and the Botany Knowledge Base — are available upon request.
3.1 Coverage of the View Retriever

Section 2 described the types of viewpoints that the View Retriever can extract from a knowledge base. To assess the degree to which these types cover the space of viewpoints that people use, we performed the following analysis. We examined 26 paragraphs on plant physiology from a college-level Biology textbook [5]. We focused on the content of each paragraph and attempted to characterize each one by the viewpoint(s) it contained.

In the text we identified 104 different viewpoints, an average of four viewpoints per paragraph. We were able to characterize roughly two-thirds of the total text using the types of viewpoints described above. When we excluded rhetorical text (e.g., figure references, reminders, organizational aids) and illustrative examples, the percentage of characterized text rose to roughly 90 percent. The remaining 10 percent was a diverse set. It included statements about modalities (e.g., "it is believed that," "It is customary to think.") and facts about etiology (e.g., "in some species the roots are highly specialized for ..."), which are outside the current scope of the View Retriever and the Botany Knowledge Base.

The textbook chapter that we analyzed was written to explain the “characteristics of roots and stems” and “the processes by which substances are transported throughout the plant body.” Thus, we expected a predominant number of viewpoints constructed along the structural, functional, and behavioral basic dimensions. (Recall that these dimensions describe the physical structure of objects, the roles of objects in processes, and the actors in processes, respectively.) This expectation was borne out by our analysis: Of the 104 viewpoints, 31 were of the behavioral dimension, 22 were of the structural dimension, and 18 were of the functional dimension. This suggests that an explanation-generation system can use the overall orientation of an explanation to bias the selection of viewpoints.

The study of instructional texts reported in [24] provides additional evidence that basic dimensions account for several common types of explanations.

3.2 Coherence of Viewpoints

In addition to evaluating the View Retriever’s coverage, we evaluated the quality of its results. We asked 10 subjects to judge the coherence of collections of facts drawn from these sources: a college-level Botany textbook [23], the View Retriever applied to the Botany Knowledge Base, and randomly selected facts all on the same topic from the Botany Knowledge Base. We manually translated them (including the textbook passages) into “simple English” to normalize presentation style. The subjects were all from the Botany or Biology Departments of the University of Texas at Austin; eight were graduate students and two were seniors.

Each subject was given 30 passages of text to judge (in about one hour). Lacking a good
<table>
<thead>
<tr>
<th>Source</th>
<th>Coherence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>(1) Textbook Viewpoints</td>
<td>4.24</td>
</tr>
<tr>
<td>(2) View Retriever's Viewpoints</td>
<td>3.75</td>
</tr>
<tr>
<td>(3) Degraded Viewpoints</td>
<td>2.85</td>
</tr>
<tr>
<td>(4) Random Collections of Facts</td>
<td>2.63</td>
</tr>
</tbody>
</table>

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 significance shows no significant difference in coherence between sources (1) and (2) or between sources (3) and (4); however, there is a significant difference between all other pairs.

definition of “coherence,” we instructed the subjects to use their own criteria and to score each passage on a scale from one (incoherent) to five (coherent). As detailed in Table 1, we found the following:

- There was no significant difference in the level of coherence between viewpoints from textbooks and those extracted from the Botany Knowledge Base by the View Retriever (based on a T-test with 0.95 level of significance).

- There was a significant difference in the level of coherence between viewpoints extracted by the View Retriever and random collections of facts all on the same topic.

A further study bolstered our conclusion that the View Retriever extracts coherent collections of facts. We presented the judges with sets of facts 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. We found a statistically significant difference in the level of coherence between viewpoints extracted by the View Retriever and these degraded viewpoints (see Table 1).

4 Discussion

In summary, 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 it provides
most of the viewpoints that people use and 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.

References


