

# Index Generation in the Construction of Large-scale Conversational Hypermedia Systems

Richard Osgood and Ray Bareiss

Northwestern University

The Institute for the Learning Sciences

1890 Maple Avenue

Evanston, Illinois 60201

internet: osgood@ils.nwu.edu, bareiss@ils.nwu.edu

## Abstract

As aids to human problem solving, CBR systems typically rely on feature matching to retrieve cases that are likely to be useful. Human experts in contrast provide much richer problem solving assistance. Experts not only can recall relevant experiences, they can provide a helpful interpretive framework of information. Because these kinds of information are highly interrelated, there are memory advantages to indexing them relative to one another—chiefly, making them more easily retrievable by human users. Therefore, we are studying the relative indexing of stories (elaborate cases) in the context of a type of hypermedia system called an ASK System. ASK Systems are designed to simulate conversations with experts. These systems provide access to manually indexed, multimedia databases of story units. Indexers (knowledge engineers) link these units together to form conversationally coherent threads. This paper discusses the theory of relative indexing employed in ASK Systems and the practical index-construction process that we have devised. However, as these systems grow in size finding appropriate relative indices manually becomes increasingly difficult. We call this the *indexer saturation problem*. Our solution is to provide automated assistance to indexers. We describe an approach that uses a theory of conversation to propose relative links between units, eliminating the need for exhaustive manual unit-to-unit comparison. Initial results suggest that the approach provides a practical solution to the saturation problem balancing the strengths of humans (*e.g.*, feature extraction and categorization) and machines (*e.g.*, rapid search and record keeping).

## Introduction

When confronted with a problem solving situation, the typical case-based reasoning system relies on matching potentially predictive situational features to features of known cases (or categories of cases) to retrieve a case

that is likely to be useful for problem solving. In contrast, a human expert can provide much more problem solving assistance. The expert not only can recall relevant experiences and, perhaps, adapt them to accommodate situational differences, but also can provide a rich context in which to interpret those cases. He or she can provide advice on which questions are best to ask in a problem solving situation, why to ask those questions, which questions not to ask, the likely results of different courses of action, how to interpret data, alternatives to a recommended approach, *et cetera*. These kinds of information are highly interrelated and, hence, there are memory advantages to indexing them relative to one another, so they can be easily retrieved in the internal context of the problem solving process.

CBR systems represent knowledge as absolute descriptions of the cases they know about. These representations are developed *a priori* to fulfill the extrinsic functional needs for the proposed system and without reference to the full complement of intrinsic relationships that might exist between cases. The idea is to perform only similarity assessments using weak methods at the feature level to retrieve the most applicable case in a new situation. The use of *absolute indexing* in CBR systems limits their ability to represent and reason about aspects of situations that require the more fully developed view of expertise implied above.

Some case-based reasoners have made secondary use of *relative indexing* but the semantics of relative links between cases have been impoverished by the failure to distinguish the many and varied kinds of relationships that exist in human memory. For example, Protos's memory [Bareiss, 1989] contained difference links between cases annotated with their key featural differences. These links were used in a hill-climbing search to find the case in memory most similar to a problem situation. Mediator [Simpson, 1985] employed failure links, the presence of which indicated that a case had been involved in a problem solving failure. The failure link pointed to a case which resolved the failure. Strictly for problem solving, this degree of relative indexing may be marginally adequate. But even then, no semantics are available to provide answers to ques-

tions like: In a failure situation does this solution always fail? If not, then in what kinds of situations does this failure situation arise? What kinds of things do I need to know in order to understand this failure well enough to apply it in my situation? We want semantics for links to stories which answer questions as rich and varied as these. This requirement is especially crucial in case-based teaching or direct user exploration systems.

We are studying the relative indexing of stories and other kinds of expert knowledge using as a vehicle a type of hypermedia system called an ASK system.<sup>1</sup> An ASK system provides access to a database of short video clips extracted from interviews with experts, as well as archival video and textual material. Its goal is to enable users to gain access to information which answers their questions as they arise and to structure their interaction with the system in a way which realizes the most important benefits of conversation with a human expert. In particular, an ASK system provides user-directed browsing of information within the framework of a coherent model of the domain under investigation. This framework enables users to assimilate answers to their questions in such a way that they can then use those answers to solve problems. We have constructed ASK systems in domains as diverse as trust bank consulting, global industrial policy making, tax accounting services, managing corporate change, and contemporary American history education [Ferguson *et al.*, 1992; Bareiss and Slator, 1993; Slator and Riesbeck, 1992].

This paper discusses the theory of relative indexing employed in ASK systems, provides an example of that indexing used in a large system, and then discusses the index-construction process and some issues that arise in practice. The paper concludes with a description of our current research on automating the process of constructing relative indices in a corpus of stories. In particular we describe a method that uses a naive model of intentionality, a simple representation, and inference procedures based on a conversational model to propose links between stories.

## The Theory of ASK Indexing

ASK systems are based on a simple, intuitive theory of human memory organization. As a new story is acquired during problem solving, it is indexed in human memory relative to recalled stories. These indices can be thought of as questions. The recalled stories raise questions that can be answered by the new story. Conversely, there are also reverse indices, where questions

<sup>1</sup>We use the term *story* to refer to an individual case in the database. The term came into usage in our work because many cases in our systems capture the firsthand experience of experts. However, many of these elements also provide general advice and are not stories in the traditional narrative sense.

the new story raises are answered by other stories in the data base [Schank, 1977].

This general theory argues that coherence in a conversation comes from the connectivity of human memory, *i.e.*, there is alignment between expression and thought (*cf.* [Chafe, 1979]). Our ASK systems are based on a model of the general memory organization that might underlie a conversation about problem solving [Schank, 1977]. We hypothesize that after hearing a piece of information in such a conversation, there are only a few general categories of follow-up information that represent a natural continuation of the thread of the conversation rather than a major topic shift. The categories can be thought of the poles of four axes or dimensions. These eight poles represent the most general kinds of questions that a user is likely to have in a conversation about problem solving. The browsing interface of an ASK system reifies this model of conversation by placing each relative link between stories in one of these eight general categories [Ferguson *et al.*, 1992]. Users can find their specific questions in the category that best describes the question. If users have only a very vague idea of what question to ask, they can simply browse through the questions in a category that looks promising.

The four dimensions are *Refocusing*, *Causality*, *Comparison*, and *Advice*. The *Refocusing* dimension concerns both adjustments to the specificity of topic under consideration as well as relevant digressions like clarifying of the meanings of terms or describing situations in which the topic arises. In a conversation, Refocusing serves the purpose of clarifying just what the topic of the conversation is. This is accomplished by providing conceptual or process generalizations and specializations, whole and part descriptions, examples, definitions, *et cetera*. One pole, *Context*, points to the big picture within which a piece of information fits. The other, *Specifics*, points to examples of a general principle, further details of a situation, definitions of terms, or descriptions of parts of the whole, *et cetera*.

The *Causality* dimension arises directly out of the human desire to understand a situation in terms of its antecedents and consequences. We group temporal order and the causal chain because people typically collapse the distinction. The *Causes* (or earlier events) pole points to how a situation developed. The *Results* (or later events) pole points to the outcome of a situation.

The *Comparison* dimension concerns questions of similarity and difference, analogy and alternative, at the same level of abstraction as the reference story. So beyond causality, the quest for understanding is also expressed transitively, as a concern to understand one thing in terms of another, by comparing the reference story and its elements with similar and dissimilar alternatives. The pole, *Analogy*, points to similar situations from other contexts or from the experiences of other experts. The *Alternatives* pole points to

different approaches that might have been taken in a situation or differences of opinion between experts.

Finally, the *Advice* dimension captures the idea of carrying away a lesson, either negative or positive, for use in the problem solver's situation. The *Opportunities* pole points to advice about things a problem solver should capitalize upon in a situation. The *Warnings* pole points to advice about things that can go wrong in a problem solving situation.

### An Example: The Trans-ASK System

How this model of conversation works in practice is best understood by considering an example from a specific domain. However, we believe the memory organization and the indexing method the example illustrates to be domain independent. Trans-ASK is an interactive video system designed to serve as a job aid, training tool, and reference for officers assigned to the United States Transportation Command (USTRANSCOM). USTRANSCOM is a joint military command which has the responsibility for planning, coordinating, and scheduling military transportation. During a fast-breaking crisis, such as Desert Shield and Desert Storm (D/S/S), a tremendous amount of planning and coordination must be accomplished effectively in a very short period of time. Unfortunately, because of the short term of many military assignments and other factors, many of the action officers who are assigned have little actual experience, and, in the frantic atmosphere of a crisis, they cannot always be given adequate assistance by their co-workers.

Trans-ASK captures the experiences in D/S/S of 25 expert transportation planners at USTRANSCOM. Their expertise is most often conveyed in the form of first person stories which are vivid and memorable and provide the viewer with analogies that more effectively guide the appropriate application of that expertise than would a set of decontextualized general principles, such as the "lessons learned" documents typically generated by military commands after each operation. For more information on the domain and implementation of Trans-ASK, see [Bareiss, 1992].

To understand relative indexing within Trans-ASK, consider the story of "Working on Christmas" told by the USTRANSCOM Operations Chief during D/S/S:

The best example of personality driven agreements or lack of agreements was Christmas. Our director of operations had a very interesting phrase he used: "looking good in the shower." It means you have to have a good image regardless of whether or not you're effective. The people had planned not to move large numbers of troops and equipment out of Europe between the 24th and 26th of December, not because they wanted Americans to have the day off but because host-nation support (German bus drivers, German truck drivers) would not be readily available over the Christmas holiday. ... This was perceived as TRANSCOM not wanting to provide transportation over the Christmas holiday even though we had

not planned it that way. And consequently, discussions were conducted at a very high level, ...it became an attitude of "Well my people are tough enough to work over Christmas and show up at the airport. Why aren't your people tough enough to be there with the airplane to pick them up and take them to Saudi Arabia?" In that kind of environment, a decision was made that the airplanes would be there and that those people would be there. Now this was completely without consideration for the planning or the long term effect of those decisions. ...Problems arose in that the host-nation support in fact was not there. ...Consequently, airplanes arrived "no load available." Very ineffective use of airlift. Again, agreements or disagreements or perceptions among senior leaders based on phone calls and handshakes in some cases had a negative effect on the operation.

This story provides a vivid introduction to the impact of image-driven decision making on the war effort. It is likely to raise a number of questions in the mind of an interested viewer. A key feature of Trans-ASK is its browsing interface which anticipates the viewer's most immediate questions at this point and provides answers to all of those for which video is available.

The system can provide answers to a number of questions raised in some of the eight general conversational categories. Here are 6 example questions from the 14 raised by this story that are answered in the system:

**Context:** Who had final control over transportation planning?

**Specifics:** How much control does USTRANSCOM leadership exercise over day-to-day operations?

**Results:** What was the effect of the Old Boy Network on transportation planning?

**Analogy:** What is another example of planning failure during D/S/S?

**Opportunities:** How can a leader be made to change his mind when he's wrong?

**Warnings:** What are the disadvantages of the Old Boy Network?

There are no questions answered in the **Causes** and **Alternatives** categories. Again, questions are only displayed in the interface if the system can provide answers.

### Indexing the Trans-ASK System

In this section we describe how Trans-ASK was indexed, then we address some of the issues for relative indexing that the experience has raised. The first step in the indexing process was to catalog the raw interview video, identifying good material and segmenting it into short clips (i.e., 1-2 minutes) that seemed to have single primary topics. This step has proven to be driven largely by common sense, and a formal methodology is only beginning to emerge. The clips are also edited and pressed to videodisk at this point in the process.

After bookkeeping information about a set of clips has been entered into an indexing tool, a portion of the ASK system is constructed by the process of *question-based indexing*. Indexers doing this must adopt the job role and, hopefully, the degree of sophistication of the intended user of the system. They view each clip in isolation and enumerate the questions that the clip is likely to raise in the mind of prospective users working on particular tasks. The general question categories of the ASK theory (illustrated in the previous section) provide guidance in this process. Next, indexers enumerate questions for which the clip is likely to provide a good answer. Each question raised and each question answered is categorized by topic (*e.g.*, high priority shipments) and by general conversational category (*e.g.*, causes).

During relative indexing, links are made between stories to enable user-directed browsing of the ASK system. The questions (raised and answered) provide an abstraction of story content that is easier to manipulate during the relative indexing process than the stories themselves. Indexers compose a set of questions raised, identifying them by topics and, optionally, by conversational category. They similarly compose a set of questions answered. Each question is entered into an indexing tool. Links between stories are made by graphically selecting a question raised and a semantically equivalent question answered. The process is performed manually, rather than via natural language processing (*i.e.*, parsing the questions and matching the resulting computer-generated representations), because of the tremendous variability possible in the statement of semantically similar questions.

Not all links between clips made via this method actually seem to maintain conversational coherence when the clips are viewed in juxtaposition. This is due to a combination of incompletely elaborated questions and faulty question matches. So until they gain more experience, indexers often shift to a browsing mode to evaluate the quality of links. Like expert system building, ASK system indexing relies on rapid prototyping and iterative refinement.

A problem we encountered during the construction of Trans-ASK was that the sheer magnitude of the system (hundreds of clips and several thousand questions) made question matching cumbersome. To deal with this problem we constructed Trans-ASK as a set of single expert, mini-ASK systems, each of which contained the results of a single interview. The content of such a single expert ASK system necessarily raises many interesting questions which are not answered by clips from the interview. Some of these are questions which could only be answered by the original interviewee (for example, "What were you thinking when you did that?") and, hence in our case, were discarded. However other questions, which call for general background or advice, could be answered satisfactorily by any planning expert. The indexer marked such ques-

tions as "exportable." After the set of single expert ASK systems was constructed, they were merged and a second question linking process was carried out using only the exportable questions. The full Trans-ASK system, containing multiple inter-connected experts, is the result of this process. The system has been well-received by its military sponsors; and an even larger follow-on system has been commissioned. Evaluations of the Trans-ASK system confirm the effectiveness of this methodology for content analysis and indexing of conversational hypermedia systems.

## Discussion

Our experience in Trans-ASK and other ASK Systems has begun to yield some practical guidance for the content analysis of stories by the question-based indexing method. The clip segmentation problem and the question matching problem are discussed here.

Before relative indices are built, the interview footage must be segmented into short video clips. Each clip is chosen so that it makes a single main point. In solving the granularity problem this way, a careful balance must be struck. The smaller the clip size the more the local context is narrowed such that more questions are raised and fewer questions are answered. The limiting case on the small end is of course simply the piece of video that directly answers a question raised. Because users are skipping around in the content as they navigate the system, such terse answers do not preserve enough local context for most users to make sense of the answer. Conversational coherence is lost, because users will have too many background follow-up questions to pursue just to interpret the answer. At the other end of the spectrum, if a clip is too large it provides too much context information and takes too long to get to the point where it answers the user's question. So there is a trade off between small clip size in which local context questions are answered by traversing a link to another clip and larger clip sizes where more are answered right within the clip itself. Indexers are asked to use their knowledge of the prospective user's need for local context when segmenting footage into clips.

Question vagueness, superfluity, parallelism, applicability, similarity, and insufficiency have profound effects on the quality of links found by question matching. The *vagueness* issue arises in many question matching situations. Should a link be made when a question raised in one story is about general transportation problems while the question answered by another is about one specific kind of transportation problem? How should matching be conducted when one question is vague or general and the other is specific? The *superfluity* issue arises when questions answered by a story are never raised in any other stories. Was the content analysis incomplete or is the material in these stories actually unnecessary? The *parallelism* issue arises when questions raised in one topic

area (e.g., Under the D/S/S topic: Who was responsible for transportation planning?) are also raised in parallel topic areas (e.g., Vietnam War). How should this parallelism be represented in the browsing interface? Next, the *applicability* issue applies when certain questions answered are raised in virtually every story within a given wider context. For example, "What was the role of USTRANSCOM in D/S/S?" provides background for nearly every story in Trans-ASK. How should these more globally raised questions be distinguished in the browsing interface? The *similarity* issue arises when a question raised in one story is matched by similar questions answered from several different stories. Should one or all of these alternative answers be linked? The *insufficiency* issue arises when a question raised in a story is not answered by any single story but requires multiple stories to construct a good answer, if indeed a good answer exists at all. How many outstanding questions raised with no answer can be tolerated before the utility of the network of stories is compromised? When an answer can be constructed, how are all these pieces of the answer identified and linked? Space does not permit addressing these issues here. See [Osgood, 1993] for more detail.

We single out the problem of question vagueness during matching for further discussion because it illustrates the power of the question-based indexing method. There are two kinds of vagueness: *Topical vagueness* and *conversational vagueness*. We deal with each type of vagueness for both questions raised and questions answered.

The first kind of vagueness in a question concerns the *topical* span of coverage of that question. There are many interesting stories an expert could tell that would answer a question raised like "What is another example of a planning failure during D/S/S?" Such a question typically defines a space of acceptable answers. In the example story, that question was matched to the question answered, "Did TRANSCOM ever move things by air that should have been moved by sea?" Finding a specific question answered to match a vague question raised depends on both appearing in the same categories for potential matching. Two means are available to help with this. First questions must be put into *all* of the general conversational categories that make sense. This insures that the question will appear everywhere it belongs, so linking will be complete. Second, the topical classification groups must be either very inclusive or be relaxable so that general questions and the specific questions will be in the same matching category. This approach provides indexers, who must ultimately make the matching judgments, with opportunities to find appropriate ones.

When a question answered is posed vaguely, it too is meant to define a span of *topical* applicability. A vague question defines an upper bound on the applicability of a particular story, but it does not necessarily define a lower bound (i.e., how specific a question

raised the corresponding story can answer). Therefore, a vague question answered cannot not be used by itself to match a specific question raised. Consider the mismatch between a question answered: "What role did the U.S. military play in the liberation of Kuwait?" versus the question raised: "What role did USTRANSCOM play...?" When a question answered is about a more general topic, e.g. the US Military, indexers cannot judge whether or not the story will cover USTRANSCOM, the more specific topic of the question raised. Therefore, they cannot, in general, match a topically specific question raised with a topically general question answered. As indexers develop more skill representing stories as questions answered, two kinds of questions emerge: questions that capture the most general kinds of answers the story can provide and questions that capture the most specific answer the story can provide. Defining these boundaries improves matching efficiency. In situations where this still does not resolve the vagueness of a question answered, the story itself must be examined to determine if the match is acceptable.

The second kind of vagueness is conversational vagueness. Consider the specific question "How was USTRANSCOM established" versus the vaguer question "What is the history of USTRANSCOM?" If the former question was raised while the latter was answered, the indexer could be reasonably certain, by common sense conversational convention, that the vague question answered would satisfactorily answer the specific question raised, i.e., history should include origins. The converse is not necessarily true: a satisfactory answer to the specific question about the establishment of USTRANSCOM could not be necessarily expected to trace the history of the organization. Conversationally vague questions raised cannot be reliably matched to conversationally specific questions answered. However, as the reliability of indexer question-based representation grows, the absence of the more general question answered can be trusted to mean that the more specific one is indeed not answered. Otherwise, indexers must examine the underlying stories to make a linking decision.

Finally, since indexers are (or should be) posing questions at the "right" level of abstraction for the eventual user, it is tempting to believe that the above types of mismatches should rarely occur. This is untrue. On the one hand, indexers are trying to ask naturally occurring questions raised—that is, they are trying to model the user's questions. On the other, they are trying to represent stories with "covering" questions answered, specifying the upper and lower bounds on the kinds of answers the story can provide. Therefore, handling mismatches in question abstraction level is fundamental to the question-based indexing method.

|   |
|---|
| <b>AgentRole:</b> <i>athlete</i> —the role the agent plays in the story                       |
| <b>BeliefType:</b> <i>strong doesn't mean dumb</i> —the agent's belief inducing the goal      |
| <b>IntentionLevel:</b> <i>actually did</i> —the level of intentionality(goal, plan, or act)   |
| <b>IntentionType:</b> <i>get good grades</i> —the goal, plan or action of an agent            |
| <b>OutcomeTypes:</b> <i>positive emotional</i> —the results of the IntentionType              |
| <b>SituationType:</b> <i>conflict with others</i> —a name linking multiple interacting frames |
| <b>TimeOfOccurrence:</b> <i>after reference</i> —sequencing information for frames            |
| <b>StoryType:</b> <i>literal example</i> —story application information                       |

Figure 1: A representational frame for describing one scene of a story

## Automated Relative Indexing

The recent emphasis of our research is the development of a more formal knowledge representation for stories than questions and the creation of associated heuristic mechanisms for inferring potential relative links between them. In using representations for stories, we have uncovered a new application for absolute indexing, one that brings together both absolute and relative indexing. With it, we are now engaged in the construction of very large interconnected archives of stories at ILS (*e.g.*, extended Trans-ASK).

We are developing tools which provide automated assistance for indexers who are building large ASK systems. Our experience shows that as the size of a single ASK system's story base (*e.g.*, one of Trans-ASK's "single expert" systems) grows beyond 70 to 100 stories (depending upon the degree of interrelatedness), the process of identifying relevant links between stories becomes prohibitively difficult for indexers.

We call this phenomenon the *indexer saturation problem*: an indexer cannot remember enough about the contents of the database to make all appropriate connections, and the prospect of exhaustive search for all connections is onerous (*cf.*, cognitive overhead problem [Conklin, 1987]).

We are beginning to provide automated assistance for our indexers to expand the number of stories they can effectively index before the saturation problem is encountered. The contents of each story are represented as input to a computerized search process which compares simple representations of the input story to that of other stories in the story base and proposes connections between them to an indexer. Although fully automated indexing of stories would be ideal, we do not believe it to be practical, given the current state of the art of knowledge representation. It will require a more complete representation of story content as well as large amounts of commonsense knowledge to infer automatically the same set of connections typically made by human indexers. Given our desire to build a practical tool today, we have decided to employ a partial representation of story contents and very limited auxiliary knowledge. The cost of this decision is the requirement to keep a skilled human "in the loop" and to maintain a story representation that can be easily processed by both machines and humans (*cf.*, semiformal

knowledge structures [Lemke and Fischer, 1990]). The program proposes a superset of the relevant relationships between stories. Its inferences are defeasible, so the human makes the final determination of relevance. This decision balances the strengths of humans (*e.g.*, feature extraction and categorization) and computers (*e.g.*, rapid search and record keeping), enabling us to build a useful tool and solve a problem intractable to either machine or human alone.

The remainder of this paper discusses our representation of stories, the procedures for inferring links between stories, and our ongoing research. Examples are drawn from experiences indexing with GroupWrite: School Stories, an ASK system authoring tool.<sup>2</sup> However, we are now engaged in utilizing the approach with Trans-ASK, as well.

## The Partial Representation of Stories

Our approach to devising a representation for stories has been to provide a domain-independent representational frame that is instantiated with domain-specific fillers (See Figure 1). A primary purpose of the frame is to enforce consistency of feature selection by an indexer. The representation is simple, indexical, and natural for human indexers to employ and is just detailed enough to support the types of inference needed to recognize relationships between stories. In this and subsequent sections, we will describe a model of naive intentionality expressed in this frame structure and inference procedures specific to the conversational categories. We will offer examples of each from the School Stories application.

<sup>2</sup>GroupWrite is a collaborative hypermedia authoring and browsing tool developed at ILS [Schank and Osgood, 1993]. GroupWrite supports three kinds of collaboration: supervised authoring of a shared hypertext, unsupervised evolutionary authoring of a shared hypertext, and reminding-based conversational story-telling. School Stories is an application of GroupWrite of the latter sort for telling and interconnecting stories about grade K-12 experiences in US public schools that ILS faculty and graduate students used over a summer. Experience with it showed that users typically authored a new story in response to a single reference story. There was no tractable way to find the links between the new story and the rest of the story base—hence the need for inferring links automatically.

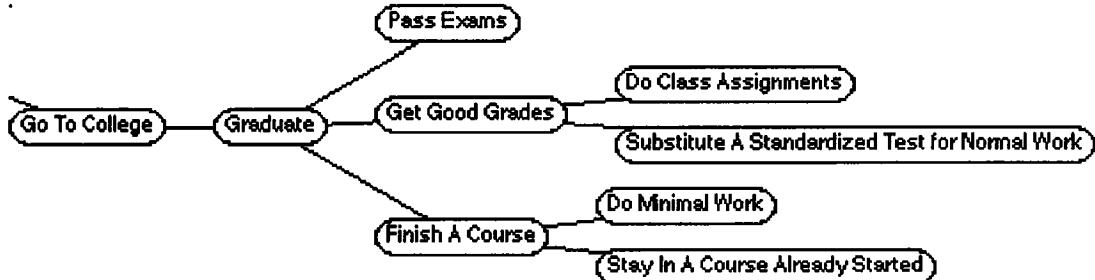


Figure 2: Fragment of the Concept Hierarchy for the IntentionType Slot Fillers Near Get Good Grades

Because all of the stories of interest in the School Stories domain (K-12 school experiences), as well as other domains in which we are working (Trans-ASK), concern human intentional behavior, our representation is based upon the intentional chain [Schank and Abelson, 1975]. Figure 1 shows the frame structure. The simple model implicit in the design of the upper section of the frame is this. First, agents play roles and have beliefs that influence their selection of a course of action. Second, to play out those roles, agents establish goals and plans to achieve them. Finally, actions based on those plans and goals yield both intended and unintended results.

When representing a story, an indexer must instantiate the slots of this domain-independent frame with fillers representing the key domain concepts of the story. To achieve representational consistency, fillers are chosen from pre-enumerated taxonomies—one for each slot. Each filler exists in a domain specific hierarchy (Figure 2). The semantics of the hierarchies are intentional for the IntentionType slot, for example, *getting good grades* is a way to *graduate* and categorical for the rest, e.g., for the AgentRole slot, *a teacher without leverage* is a kind of *teacher*. Figure 1 also shows examples of fillers drawn from the School Stories domain.

*A priori* enumeration of all slot fillers is not intended. Rather our idea is to provide an indexer-extensible set for which the initial enumeration serves as an example. Indexers can enter a new term in the hierarchy by determining its similarity to pre-existing fillers. Assessment of the similarity of fillers during representation works because it is conducted by indexers in the target system's task context—the same one in which they would have judged the appropriateness of hand-crafted relative links between stories. In effect, the similarity of concepts is represented in equivalence classes, not computed from features [Porter, 1989], i.e., similar concepts have a common parent in the hierarchy. To infer links, these hierarchies of equivalence classes are processed by inference procedures described in the next section. Figure 2 displays a section of the concept hierarchy for the IntentionType slot filler, *get good grades*, used in Figure 1.

The representational frame or *scene* captures the intentionality of a single agent. The upper portion of Figure 1 frame says: *an athlete actually did get good grades* by believing that *being strong doesn't mean being dumb* and this had a *positive emotional impact* on him/her. The concept hierarchy (Figure 2) elaborates the athlete's intentions. The athlete *got good grades* as a way to *graduate* because he wanted to *go to college*.

In the frame's lower part in Figure 1 we include three additional slots. The SituationType slot functions both to group frames together and to describe the kind of agent interaction in those frames. Since most stories involve interactions among multiple agents, sometimes with conflicting goals, indexers employ multiple frames—one or more for each agent, filling just the slots in each that they feel apply (*cf.*, the Universal Indexing Frame, [Schank and Osgood, 1991]).<sup>3</sup> Frames for multiple agent interactions, called situations, are grouped by assigning a common value to the SituationType slot of each frame. In this way, the representation can support three levels of detail. First is the scene—a description of a single agent's activity. Second is the situation—a group of interacting scenes. Last is the story—the composition of a number of situations over time or through inferred causality. For example, an infatuation situation is captured by selecting *Student Infatuation* to fill the SituationType slots of two frames of the same story, one about a *student* who *wants a love relationships with a teacher* and the other about a *teacher* who *just wants to teach subject matter to students*.

The vocabulary term which fills a SituationType

<sup>3</sup>This representational issue is, in fact, even more complicated because, in addition to an intrinsic representation of each agent in a story, the complete description of interactions among agents may require representation of (possibly flawed) models of the intentionality of an agent upon which other agents in the story are acting. Dealing with stories that require this depth of representation is beyond the scope of our current work. But because we are interested in discovering the relationships that occur to people spontaneously, we do not need deep psychological models of human intentionality. All we need is enough specificity and sophistication in the fillers to support a proposed link. Indexers will discard any inappropriate suggestions.

|                         |                              |                         |                                    |
|-------------------------|------------------------------|-------------------------|------------------------------------|
| <b>AgentRole</b>        | <i>Student</i>               | <b>AgentRole</b>        | <i>Teacher</i>                     |
| <b>BeliefType</b>       |                              | <b>BeliefType</b>       |                                    |
| <b>IntentionLevel</b>   |                              | <b>IntentionLevel</b>   |                                    |
| <b>IntentionType</b>    | <i>Show Lack of Interest</i> | <b>IntentionType</b>    | <i>Assign Independent Activity</i> |
| <b>OutcomeTypes</b>     | <i>Successful</i>            | <b>OutcomeTypes</b>     | <i>Successful Positive</i>         |
| <b>SituationType</b>    | <i>Being Bored</i>           | <b>SituationType</b>    | <i>Being Bored</i>                 |
| <b>TimeOfOccurrence</b> | <i>At Reference</i>          | <b>TimeOfOccurrence</b> | <i>At Reference</i>                |
| <b>StoryType</b>        | <i>Opportunity</i>           | <b>StoryType</b>        | <i>Opportunity</i>                 |

Figure 3: Two scenes for the story Entertaining the Troublemaker

slot is descriptive of a situation, as well. The term's location in the concept hierarchy determines its similarity to other types of situations. This raises an important issue about the interdependencies among slots. There are many unrepresented relationships between the slots in the frame, *e.g.*, fillers for the top five slots must be carefully chosen so their composition into a statement of agent activity makes sense.<sup>4</sup> The use of **SituationType** for similarity assessment is another way in which the slots are not orthogonal. In this case the representation is purposefully overspecified to minimize the work needed to infer links while preserving the necessary distinctions for accuracy in inference, *e.g.*, the semantics of the **AgentRole**, *A Teacher without Leverage*, overlaps the semantics of the **SituationType**, *Taking Advantage of a Teacher*. Over-specification means that two stories about *Taking Advantage of a Teacher* can be similar even though one involves a *Teacher* and the other involves a *Teacher without Leverage*. Using the **SituationType** slot in this way is our approach to maintaining this balance between efficiency (representing similarity) and accuracy (making necessary distinctions).

The **TimeOfOccurrence** slot supports sequencing of scenes in stories to establish intrastory causal/temporal relationships. For example, the term *at reference* indicates the relative point in time of the main action of the story, while drawing a lesson from the story happens *after reference*, another time designation. When filled with an absolute time designation, *e.g.*, a date, **TimeOfOccurrence** can also be used to perform temporal ordering of whole stories.

The **StoryType** slot allows the indexer to advise the inferencing mechanism to identify what the story might be useful for and what the level of abstraction of the story content is. For example, if a story contains useful cautionary advice this slot will contain the value *Warnings*. If a story is a good explicit example of something, *Literal Example* would fill this slot.

This frame representation works in conjunction with

<sup>4</sup>The inference procedures described in the next section depend on these implicit relationships between slots. Our work to date has not made it necessary to represent and process these relationships explicitly—that is, we have not had to explicitly represent the relationships among the slots in the system. This is one of the ways in which our representation is deliberately incomplete.

the domain concept hierarchies, forming the first two components of our assistance to indexers. The third is the inferencing.

|                         |                                   |
|-------------------------|-----------------------------------|
| <b>AgentRole</b>        | <i>Student</i>                    |
| <b>BeliefType</b>       |                                   |
| <b>IntentionLevel</b>   |                                   |
| <b>IntentionType</b>    | <i>Actually Did Disrupt Class</i> |
| <b>OutcomeTypes</b>     | <i>Successful</i>                 |
| <b>SituationType</b>    | <i>Being Bored</i>                |
| <b>TimeOfOccurrence</b> | <i>At Reference</i>               |
| <b>StoryType</b>        | <i>Literal Example</i>            |

Figure 4: A Scene from A Different Bag of Tools

### An Example: Indexing School Stories

Automated inference helps indexers find appropriate links between stories. While our work in this area is ongoing, the examples below taken from GroupWrite: School Stories illustrate the kinds of links between stories that can be inferred from the simple representation of stories described above. One story entitled *Entertaining the Troublemaker* begins:

One problem for smart kids is to keep from boring them in school. Each year that I was in school, my teachers had to find some way to keep me out of trouble since I was both bored and rambunctious. In the second grade I ran messages for the teacher. In the third I built baseball parks out of oak tag. In the fourth I wrote songs. These events turn out to be most of what I remember from those years. School for me was one long attempt to avoid boredom and trouble.<sup>5</sup>

Its representational frame consists of the two scenes in Figure 3. Now suppose the inference procedure for finding examples links is being run and the frame in Figure 4 is encountered. The process for examples inference first specializes the fillers for each of the slots of the frame for the reference story (Figure 3). For instance, one causal specialization of the **IntentionType** slot with filler, *Show Lack of Interest* is *Disrupt Class*, *i.e.*, one way to show lack of interest is to the disrupt class (Figure 5). Because the candidate follow-up story frame in Figure 4 has the **StoryType** slot filler *Literal Example*, the system proposes an examples link to story *A Different Bag of Tools*, the story

<sup>5</sup>This story was written by Roger Schank for GroupWrite: School Stories.

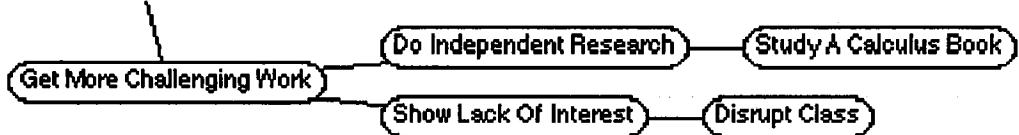


Figure 5: Concept Hierarchy for IntentionType Fillers Near *Disrupt Class*

represented partially by the frame in Figure 4, which reads:

I had learned to do integrals by various methods shown in a book that my high school physics teacher, Mr. Bader, had given me.

One day he told me to stay after class. "Feynman," he said, "you talk too much and you make too much noise. I know why. You're bored. So I'm going to give you a book. You go up there in the back, in the corner, and study this book, and when you know everything that's in this book, you can talk again."<sup>6</sup>

|                         |                              |
|-------------------------|------------------------------|
| <b>AgentRole</b>        | <i>Student</i>               |
| <b>BeliefType</b>       |                              |
| <b>IntentionLevel</b>   |                              |
| <b>IntentionType</b>    |                              |
| <b>OutcomeTypes</b>     |                              |
| <b>SituationType</b>    |                              |
| <b>TimeOfOccurrence</b> |                              |
| <b>StoryType</b>        |                              |
|                         | <i>Actually Did</i>          |
|                         | <i>Leave Class</i>           |
|                         | <i>Successful</i>            |
|                         | <i>Getting What You Want</i> |
|                         | <i>At Reference</i>          |
|                         | <i>Literal Example</i>       |

Figure 6: A Scene from the story *A Deal's a Deal*

In this simple case, our representation was sufficient to infer a possible *examples* link.<sup>7</sup> If we had looked for additional ways in which to connect these same two stories, we would have also found a *similarity* link as well through **SituationType**: *Being Bored*. The human indexer can accept one or both of these links for inclusion in the GroupWrite: School Stories ASK system, although the more specific *examples* link is preferred (presented first) by the system.

Our representation does exclude some close yet still inappropriate links. The frame for the story *A Deal's a Deal* in Figure 6 does not qualify as an *examples* link for our original story because, while it is has the **StoryType** slot filler *Literal Example*, the IntentionType filler *Leave Class* is not a specialization of *Show Lack of Interest* (Figure 5).

<sup>6</sup>This story was extracted by Ian Underwood for GroupWrite: School Stories from Feynman, R (1985) *Surely you're joking, Mr. Feynman: adventures of a curious character*. New York: W. W. Norton.

<sup>7</sup>In a group story-telling environment authors did not maintain strong causal/temporal threads by telling a sequence of related stories. Therefore the conversational categories have analogical semantics. In the case of an examples link, one story is an example of the kind of thing discussed in general terms by a story which is probably by another author.

How well the approach excludes near misses depends on the assignment of filler terms to equivalence classes in the concept hierarchies, e.g., the intentionality hierarchy captures a path of reasoning not a single action, plan or goal. So equivalence class assignment means that agents do similar things for the same reasons. This kind of similarity limits inadvertent feature matching, because similarities are derived within the context of a specific unambiguous hierarchy locale. In the above example, one construal of *Leave Class* could conceivably be to *Show Lack of Interest*, but that is not the reason in *A Deal's a Deal* (Figure 7). In that story the agents *Leave Class* as a way to *Refuse to Cooperate with a Teacher*. *Showing Lack of Interest* is a weaker reason and is not represented as similar, i.e., not placed in the same local context of the intentional hierarchy (Figure 5).

## Implementation

These simple examples illustrate how richly connected the stories in our test domain are and how, with a simple representation and processes, these links can be inferred. Given the human commitment to fill out frames for stories and to verify each system-proposed link, such a method significantly reduces the cognitive load human indexers face and thereby ameliorates the *indexer saturation problem*.

We have implemented inference procedures for all of the link types specified by the ASK model of hypermedia. In concept, inference procedures compare one of the representation frames of a reference story with all other frames in the story base. Operationally, inference is implemented as path finding, not exhaustive search and test. Links from slot fillers in the reference story frames are traversed in the concept hierarchy to identify sibling fillers which instantiate slots of other stories (cf., dynamic memory [Schank, 1982]). Inference by pathfinding is efficient because slot fillers of the scene frames of the reference story are pointers into the hierarchy of domain concept frames and each concept frame contains backpointers to all scene frames that use it as a filler. This structure incrementally compiles all relationships among fillers of scene slots in stories. Inference procedures are implemented as deductive retrieval rules which exploit these relationships. Each rule explains its own reasoning to the human user of the system and can create one of more links depending on whether or not the link type is symmetric, e.g., *analogies/analogies*, or complementary, e.g.,

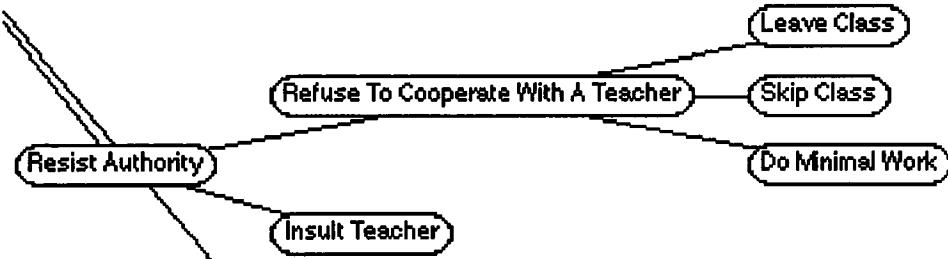


Figure 7: Concept Hierarchy for IntentionType Fillers Near *Leave Class*

#### *context/specifics.*

A particular rule finds one sense of a link type. There are many senses and, hence, rules. Summaries of each sense we have implemented are provided.<sup>8</sup> *Context, Specifics, and Examples* are the Refocusing links. In a reference story scene if the parent concept of the situation or the agent's activity occurs in a potential follow-up story scene, the context link is proposed. If on the other hand it is a child concept that is present in the follow-up story scene, then the specifics link is proposed. When a specifics link has been proposed and the follow-up story scene also has the story type of literal example, then an examples link is also proposed.

*Earlier Events, Later Events, Causes and Results* are the Causality links. When absolute temporal information is available in a reference story scene, and a potential follow-up story scene describes the same situation or similar agent activity and has an earlier absolute time designation, the earlier events link is proposed. Otherwise the later events link is proposed. When absolute temporal information is not available in a reference story scene, and a potential follow-up story scene has the same agent activity but an earlier position in the goal-plan-act sequence of intentionality, a causes link is proposed. A results link is proposed if the follow-up is later than the reference story scene in the goal-plan-act sequence. Also, when a reference story scene is missing a belief to explain an agent's activity or situation, causes links are proposed to all follow-up story scenes that can supply one. A results link is proposed if the reference and follow-up story scenes are about similar situations or have similar agent activity and the follow-up story scene can provide the reference scene with missing outcome information.

#### *Analogy and Alternatives* are the Comparison

<sup>8</sup>In summarizing the link rules we have not included descriptions of three capabilities. We omitted the mapping of the information requirements of inference to the actual sources of that information in the intentional frame (See Figure 1). Also these descriptions only cover the proposal of links from reference to follow-up story scenes. We have omitted descriptions of similar rules that propose links from existing stories back to the reference. Also omitted is the capability of the rules to propose, when appropriate, symmetric and complementary links between follow-up and reference stories.

links. If a reference and follow-up story scene have agents with similar beliefs, situations, or activities, then an analogies link is proposed between them. However, if in otherwise similar story scenes, dissimilar values are found in exactly one of the slots used above to compute similarity, then an alternatives link is proposed instead.

*Warnings and Opportunities* are the Advice links. In similar reference and follow-up story scenes, if one has a story type of one of the advice link types and the other does not, then a link of that type is proposed from the former to the latter. The indexer provides these story type values when representing the story.

When we first defined the system, the information needs of these inference procedures determined the definition of the frame as well as the parts of domain concept hierarchy vocabulary that are explicitly mentioned in the rules, e.g., a story type of, *literal example*, used in the *examples* link inference. Likewise these rules operate in conjunction with the representations of similarity built into the equivalence classes of the hierarchy. The effectiveness of machine assisted relative indexing is dependent upon the tuning of this relationship between rules and representation. Experience with tuning the School Stories system indicates that this task is within the capabilities of our indexers.

## Ongoing Research

This work raises a number of research issues: balancing a fine grained representation against the ability to do simple syntactic feature matching, extending domain concept hierarchies consistently, and testing the effectiveness of the inference rules for machine assisted indexing.

It is difficult to determine just how much detailed domain knowledge should be represented in the content hierarchies to support the kinds of inferencing we have envisioned. There is a trade-off between the coarseness of the representation and its usefulness for inferring links by feature matching. At one extreme we could have used fine grained representations that enrich expressiveness but make overall determination of similarity between stories very difficult, because the representations must be processed deeply to compensate for acceptable variation in representation. At the

other extreme we could have reified complex relationships into flat propositional features which reduces inferencing to simple feature matching. For example, we rejected the use of complex role relations as a way to represent multiple interacting agents in the AgentRole slot, *e.g., student who is infatuated with the teacher but the teacher does not respond favorably*. Use of such unduly extended filler names flattens the representation lessening the ability to infer links, because the internal structure of the filler is not accessible to inference [Domeshek, 1992]. To compensate, indexers would have had to proliferate features with diminishing returns as they try to select slot fillers from ever expanding, marginally differentiated lists. Therefore, we have tried to find an acceptable balance in our representation between flat and deep representation. Our principle is to provide just the amount of representation needed by the inference rules we have defined.

It is the indexer's job to define the domain concept hierarchies and use these as fillers in frames for stories. These fillers establish equivalence classes for inferencing. Also where they are placed in the hierarchy represents a prediction about where future indexers will find fillers to describe their stories. Therefore, consistency and economy in the selection of the hierarchy vocabulary is required by both machine and human. We do not yet know how consistent the human extension of domain hierarchies will be. Our experience to date suggests that indexers sometimes overlook or misinterpret the semantics of existing fillers. In many domains, different vocabularies tend to be used in different situations. The result is the creation of synonymous categories. Indexers may also misuse the hierarchy by placing elements of widely divergent levels of abstraction at the same level in the hierarchy. Our experience so far indicates that some model revision and reindexing is necessary when one starts a project in a new domain. This effect grows in impact with the complexity of domain models. Our current solution is to use the simplest partial domain models that will support the desired inferences—a corollary of the principle governing representation for rules stated above.

Finally, we have not yet subjected the conversational category-based inference rules for machine assisted linking to a systematic comparison with the link sets derived by human indexers independently. We have however conducted some informal checks on the system's performance in one domain (School Stories) and are beginning to check it in another, *i.e.*, the military transportation planning domain of the Trans-ASK system. Initial results are promising.

## Conclusions

Our work in Trans-ASK and other ASK systems demonstrates the value of relative indexing for stories. The use of a conversational theory to organize and present these links provides effective access to more complex forms of expertise than can be offered typi-

cally by a CBR system. Experience with the indexing of large collections of stories, like those gathered for Trans-ASK, has produced an effective methodology for building conversational hypermedia systems to deliver the expertise they contain.

In the process, we have discovered a new role for absolute indexing, *i.e.*, to assist with the relative indexing process. Our approach has been to infer potential relationships between stories using representations of stories and inference procedures based on our conversational theory. We have begun to see some significant benefits to indexers from machine-assisted knowledge acquisition because it helps solve the indexer saturation problem. Ideally, as our inference procedures are improved and as our confidence grows that the indexes generated converge with those humans would produce, we may be able to grant autonomy to some of them, enabling our ASK hypermedia systems to generate some classes of relative links dynamically. Whether or not that proves possible, we are creating an optimal partnership between human and tool, enabling large-scale relative indexing which neither human nor machine can do alone.

**Acknowledgments:** The Trans-ASK system is being constructed by Bob Hooker, Jim McNaught, Mark Chung, Ann Kolb, Rich Lynch, Carolyn Majkowski, Mike Sijacic, Don Steul, and John Welch. Other faculty contributors to the project include Roger Schank, Richard Beckwith, Gregg Collins, Alex Kass, Larry Birnbaum, and Chris Riesbeck. The Trans-ASK project has been funded by DARPA (contract number MOA972-91-C-0052) as a subcontract to SRA Corporation (subcontract number SC-707-001). The dynamic indexing tool was written by Paul Brown and Paul Rowland. This research was supported, in part, by the Defense Advanced Research Projects Agency, monitored by the Air Force Office of Scientific Research under contract F49620-88-C-0058 and the Office of Naval Research under contract N00014-90-J-4117, by the Office of Naval Research under contract N00014-89-J-1987, and by the Air Force Office of Scientific Research under contract AFOSR-89-0493. The Institute for the Learning Sciences was established in 1989 with the support of Andersen Consulting, part of The Arthur Andersen Worldwide Organization. The Institute receives additional support from Ameritech and North West Water which are Institute partners, and from IBM.

## References

- [Bareiss and Slator, 1993] R. Bareiss and B. Slator. The evolution of a case-based computational approach to knowledge representation, classification, and learning. In G. Nakamura and R. Taraban, editors, *Acquisition, Processing and Representation of Concepts: The Contribution of Exemplars and Theories*. Academic Press, Orlando, 1993. Forthcoming.
- [Bareiss, 1989] R. Bareiss. *Exemplar-based Knowledge Acquisition: a Unified Approach to Concept Repre-*

- sentation, Classification, and Learning.* Academic Press, San Diego, 1989.
- [Bareiss, 1992] R. Bareiss. Trans-ASK: An ASK system for military transportation planning. Forthcoming technical report, 1992.
- [Chafe, 1979] W. Chafe. The flow of thought and the flow of language. In T. Givon, editor, *Discourse and syntax*, pages 159–181. Academic Press, New York, 1979.
- [Conklin, 1987] E. Conklin. Hypertext: An introduction and survey. *IEEE Computer*, 2:17–41, 1987.
- [Domeshek, 1992] E. Domeshek. *Do the Right Thing: Component Theory for Indexing Stories as Social Advice.* PhD thesis, Yale University, New Haven, CT., 1992.
- [Ferguson *et al.*, 1992]
- W. Ferguson, R. Bareiss, L. Birnbaum, and R. Osgood. ASK systems: An approach to the realization of story-based teachers. *The Journal of the Learning Sciences*, 2:95–134, 1992.
- [Lemke and Fischer, 1990] A. Lemke and G. Fischer. A cooperative problem solving system for user interface design. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, Menlo Park, CA, 1990. AAAI Press/The MIT Press.
- [Osgood, 1993] R. Osgood. Question-based indexing. Forthcoming technical report, 1993.
- [Porter, 1989] B. Porter. Similarity assessment: Computation vs. representation. In *Proceedings: Case-Based Reasoning Workshop*, San Mateo, CA, 1989. Morgan Kaufman Publishers.
- [Schank and Abelson, 1975] R. Schank and R. Abelson. *Scripts, Plans, Goals and Understanding.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1975.
- [Schank and Osgood, 1991] R. Schank and R. Osgood. A content theory of memory indexing. Technical Report 2, The Institute for the Learning Sciences, Northwestern University, Evanston, IL., 1991.
- [Schank and Osgood, 1993] R. Schank and R. Osgood. The communications story. Technical Report 37, The Institute for the Learning Sciences, Northwestern University, Evanston, IL, 1993.
- [Schank, 1977] R. Schank. Rules and topics in conversation. *Cognitive Science*, 1:421–441, 1977.
- [Schank, 1982] R. Schank. *Dynamic Memory.* Cambridge University Press, 1982.
- [Simpson, 1985] R. Simpson. *A Computer Model of Case-Based Reasoning in Problem Solving: An Investigation in the Domain of Dispute Mediation.* PhD thesis, School of Information and Computer Science, Georgia Institute of Technology, Atlanta, GA, 1985.
- [Slator and Riesbeck, 1992] B. Slator and C. Riesbeck. TaxOps: A case-based advisor. *International Journal of Expert Systems*, 4(2), 1992.