ASK TOM:
An Experimental Interface for Video Case Libraries

Roger C. Schank, William Ferguson, Lawrence Birnbaum, Jorn Barger, and Mathew Greising

Northwestern University
The Institute for the Learning Sciences and
Department of Electrical Engineering and Computer Science
Evanston, Illinois

Abstract

ASK TOM represents a new approach to structuring access to a newly emergent kind of knowledge base, the video case library. It is based on two premises: First, that cases and stories in the form of video clips can provide much of the viscerality and memorability that is lacking from textual forms of presentation; and second, that AI theories, specifically the approaches to memory organization derived from work in case-based reasoning, can provide the structure that is essential to achieving the shared context that makes communication possible. The aspect of AI research that is crucial in providing this structure does not concern algorithms. Rather, it is the *content* of domains and tasks that is paramount.

1 Introduction

Wouldn't it be nice if, whenever you had a question about something, an expert were available for you to talk with? An expert can answer your tough questions. Perhaps more importantly, an expert can help you to *ask* the right questions, and can fill in exactly the background you need to understand the answers. An expert can help you organize the information you are getting in a way that will be useful in solving your specific problem. And an expert can deliver all of this information in a way that is natural, memorable, and if you are lucky, perhaps even entertaining: An expert can tell you *stories*.

Unfortunately, expertise is expensive, and so access to experts is severely limited.[1] It therefore makes sense to explore the possibility of constructing computer-based systems that would enable users to gain access to expertise in a way that is as natural, memorable, and efficient as a real conversation with an expert. In the ideal case, such a system would be the ultimate artificial intelligence program: It would be able to understand spoken natural language, it would be a true expert—armed with powerful reasoning capabilities, a diverse stock of useful models of the domain of expertise, and a large memory of specific cases indexed and organized intelligently—and it would be able to communicate its answers to you in both verbal and visual form. In the meantime, though, can anything be done with existing technology to produce something that is close enough to the cognitive equivalent of a conversation with an expert to be useful?

We think the answer is yes. We have developed an experimental hypermedia system, called ASK TOM, for accessing a videodisc database of short (one- to four-minute) film clips culled from interviews with an expert consultant in the trust banking industry. Our goal was to construct a system that would enable naive users to gain access to clips containing information pertinent to their questions, and to structure their interaction with the system in a way that realized the most important benefits of a conversation with a human expert. In particular, we wanted the system to provide users with a coherent model of the domain, one that would enable them to assimilate the answers to their questions in such a way that they could then use those answers to solve problems.

In order to achieve this coherence, the video clips in the system's data base are indexed via nodes in an associative network that is organized to reflect both the objects and relationships in the domain and the task of solving problems in that domain. The network includes both hierarchical links for quickly pinpointing potentially relevant neighborhoods within the data base—what we call *zooming*—and a dense set of links between cases which facilitate *browsing* by providing context and managing follow-up questions. A user interacts with ASK TOM by navigating through a hypertext-like network whose links and nodes are labelled with useful questions and

[1] This is, of course, the rationale behind the construction of expert problem-solving systems.

important features, and viewing video clips that seem likely to provide answers to his or her questions.[2]

2 Using content-based AI to structure human-computer interaction

The basic interface principle underlying the construction of ASK TOM is that even a rudimentary AI theory of a task or a domain provides constraints that can be used to structure human-computer interactions in a way that may be more meaningful to the human user.[3] To the extent that an interaction so constrained corresponds to the user's understanding of the task and the domain, the context provided by the system at any given point will agree with the context that the user naturally expects at that point. This shared context provides the fundamental basis for meaningful and useful communication. Thus, the information provided by the system will correspond better to the information needed by the user; the choices made available to him or her will be more relevant, more useful, and easier to resolve; and the information conveyed in the interaction will be associated with the appropriate context in the user's memory.

It is important to make clear that ASK TOM contains only a trivial algorithm: The key to its functionality and the substance of our results is the *content* of the network contained in the system. The theoretical basis underlying the system's organization is provided by *case-based reasoning* (CBR), a paradigm for understanding and problem-solving which takes memory to be the foundation of both human and artificial cognition. CBR is based on the idea that solving a problem or understanding a situation involves retrieving similar prior experiences, and adapting them to provide guidance in the current situation (see, e.g., Schank, 1982; Riesbeck and Schank, 1989). It posits that a significant portion of human expertise derives from specific experiences, and is best represented in that form in memory to provide guidance in future problem-solving situations.

Two major teaching principles follow from such an account of human cognition: First, if expertise is best represented in terms of specific experiences, then a natural way to facilitate the assimilation of new information by humans is to present that information in the form of cases or *stories*. The great majority of video clips in ASK TOM's database therefore consist of an expert telling stories. Second, because the usefulness of information depends crucially on the ability to retrieve it when appropriate, such cases are best presented in a context that enables the listener to determine where they should be indexed in his or her memory, and how they should be connected to other relevant cases. In ASK TOM, this context is provided by the network that enables access to particular video clips, and by the clips that the user has viewed along the way.

The two basic processes that underlie case-based reasoning are *indexing*, or finding appropriate cases in memory, and *adaptation*, or altering the specific information in those cases to fit the current circumstances. In ASK TOM, we rely heavily on the user to carry out these processes. In particular, the human user guides the traversal of the indexing network in directions that seem likely to yield information relevant to his or her needs. The system itself knows virtually nothing about the contents of the clips other than the links that connect them to each other and to the network as a whole. However, the user's choices are constrained by a model of link types that relate cases to each other in a way that is intended to reflect their potential uses in planning and problem-solving, such as background, results, warnings, and so on.

A key issue in the approach we are pursuing here is whether it will be possible to anticipate all, or at least most, of the questions that novices would ask of a human expert in a domain such as trust bank consulting. This becomes especially critical in the case of follow-up questions, where the absence of the desired information would be extremely irritating to users, and where the design of the system interacts with physical constraints of screen layout to limit the number of choices that can be offered. Two observations provide some basis for optimism on this issue. First, experts, when consulted, repeatedly offer their favorite "war stories" as carriers of advice in response to a wide variety of different questions (Schank, 1990). Second, the paths people follow in conversations are, in many respects, predictable given the appropriate level of analysis (Schank, 1978). Not surprisingly, in conversations about planning and problem-solving tasks, the level at which we can make such predictions is precisely the level at which it is possible to represent abstract planning knowledge, as exemplified by Sussman's (1975) *planning critics* or Schank's (1982) *thematic*

---

[2] A good overview of hypertext systems and concepts may be found in Conklin (1987).

[3] This view is gaining increasing acceptance in AI and cognitive science; see, for example, Winograd and Flores (1986), or the announcement for the 1990 AAAI Spring Symposium on "Knowledge-Based Human-Computer Communication" organized by Gerhard Fischer, Clayton Lewis, Jim Miller, and Elaine Rich.

organization points.[4] It is precisely at such a level of analysis, therefore, that we strove to label and link stories in ASK TOM's indexing network.

The design of ASK TOM was also greatly influenced by the *lack* of a usual constraint, in that we did not need rigor, specificity, or rigid consistency in our representations. They were not, after all, going to be manipulated by any natural language generator or parser, nor by any automated reasoning system. The mind of the user could supply the knowledge of language and of the world needed to weed out minor inconsistencies and, more importantly, to specify vague denotations appropriately. In fact, in this system *vagueness* is a resource we used liberally to enable our representation system to function in many contexts and across many categories in what is, after all, a large and rich domain.

3 Indexing system description

ASK TOM consists of a case-base of video clips about consulting in the trust banking industry, together with an indexing system as described in general terms above. The scenario for using ASK TOM is that the user, having a question in mind, will navigate through the network to an appropriate clip and then play it. He or she may then go back to the beginning and find the answer to further, unrelated questions, or else move on through the network to locate and play clips which answer followup questions.

To describe the structure of the indexing scheme which lies at the heart of ASK TOM, three indexing levels need to be distinguished. They are (1) the *big-picture models*, (2) the *themes*, and (3) the *low-level links*. We will describe these three levels and their function by walking through a prototypical interaction with the system.

3.1 Big-picture models

When ASK TOM is first started, the user is presented with a diagram showing the major agents in the world of trust bank consulting—e.g., the trust department of a bank, a consulting firm, the individual consultant, the trust industry as a whole, and so on—with links between them showing how they interact. This is the top-level big-picture model (see figure 1). Clicking on the nodes or links in this diagram causes the screen to display either a more specific big-picture diagram, describing, e.g., the temporal layout of a typical consulting engagement at a trust bank, or else a screen showing a set of themes and the clips that they organize (discussed below). The set of all big-picture models in the system forms a conceptual map of the domain, organized around its most important agents, relationships, and processes. These big-picture models mediate the first (and often the second) choices made by the user during the initial process of zooming in on a region of specific interest. When the user makes choices on these diagrams, he or she is implicitly categorizing the information desired in terms of the model denoted by the diagram. Thus, the user is implicitly preparing to incorporate that information into his or her own memory in a way that reflects its proper relationship to the rest of the model.

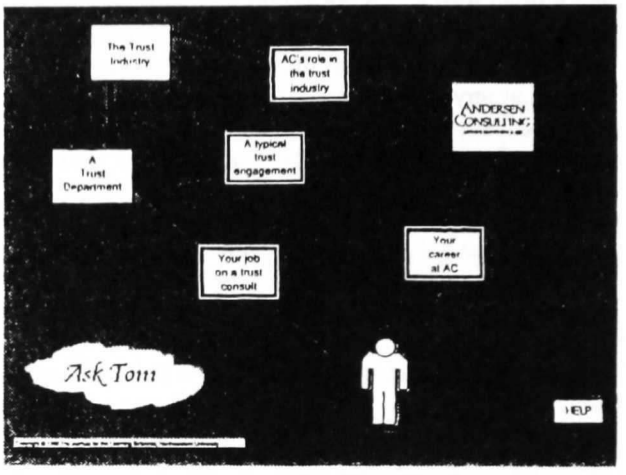

Figure 1: The opening screen of the ASK TOM system.

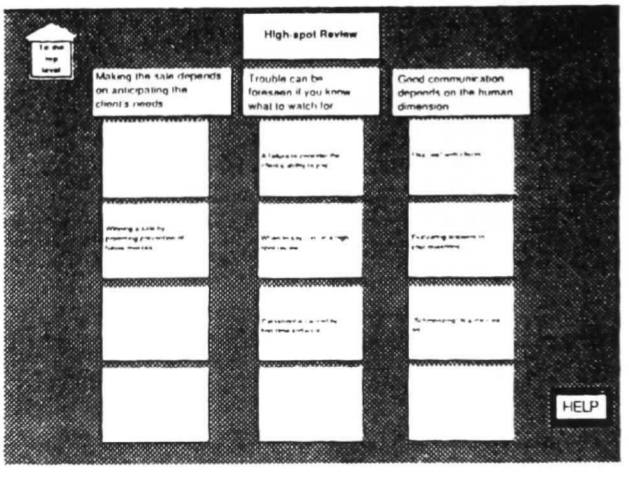

Figure 2: The theme screen under the High-spot review button.

3.2 Themes

After one or two choices, the user will arrive at a theme screen (Figure 2). This screen displays the clips relevant to the previously selected elements of the big-picture models, partitioned by the themes which they embody. A theme, as the term is used in

[4] See also Sacerdoti (1977), Wilensky (1983), and Hammond (1989), among others.

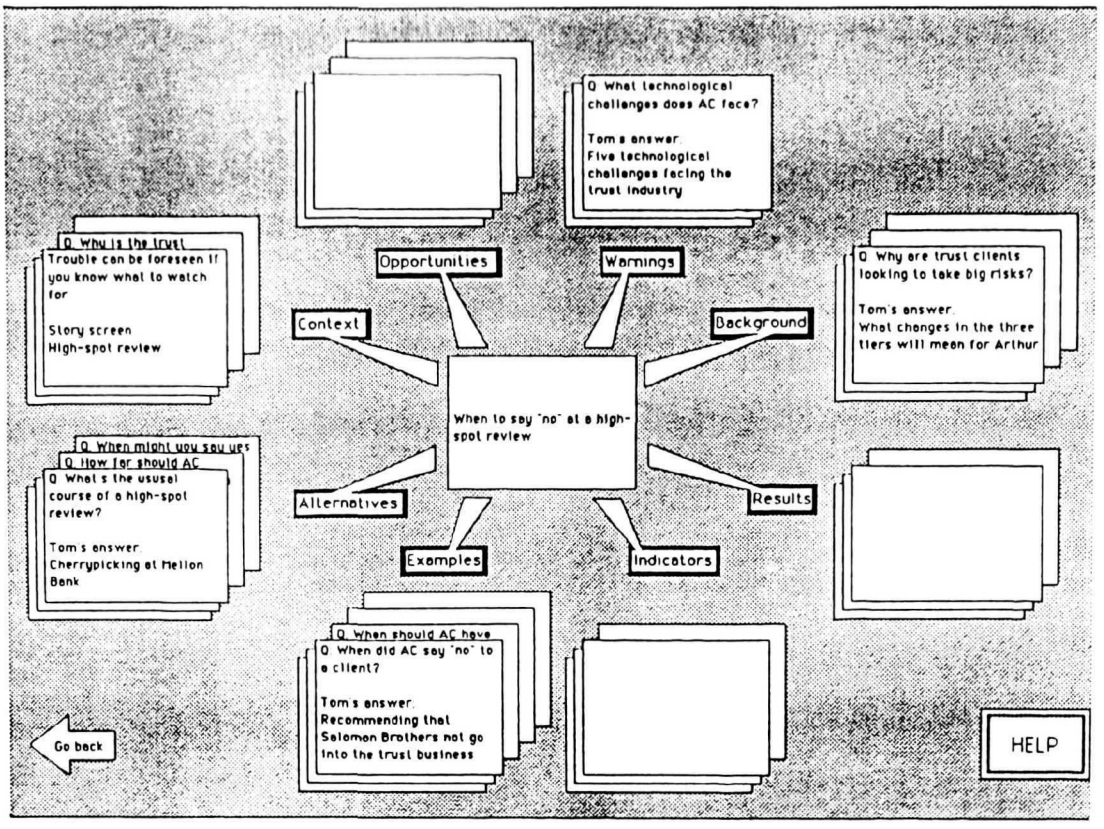

Figure 4: The clip screen reachable from the theme screen above.

ASK TOM, is a lesson or point relevant to the big-picture node under which it is found, e.g., "Correct trouble at its source." Themes thus serve as a bridge between domain models and case knowledge. They are not, however, specializations of, nor are they components of, the big-picture nodes under which they are indexed. Rather, they serve to categorize the contents of the system's clips according to how the information they convey is useful in problem-solving. The selection of a clip from a theme screen concludes the zooming process. (Themes are discussed in more detail below.)

3.3 The low-level links

When a clip is selected from a theme screen the system displays a clip screen (see Figure 4). In the center is the selected clip, which can now be clicked on and run, at which point it will appear on the screen. From the central clip node, the low-level links radiate toward eight "stacks" of clips which are the neighbors of the center clip. Each clip node displayed on the screen is tagged with a question that relates it to the information presented by the central clip, along with a one sentence summary of its contents. The eight link types are the same throughout the network of clips, and always appear in each clip screen. As discussed earlier, they are intended to relate the system's cases to each other in ways that are relevant to the tasks of planning and problem-solving, and thus to suggest, and offer answers to, useful follow-up questions. The eight link types are:

Background: This links a clip to antecedent information about that clip, e.g., the history behind the story it contains, or the causes of the situation it describes.

Results: This provides a link between a clip and the consequences of, or events subsequent to, the situation it describes.

Examples: This link relates a clip to more detailed examples of the situation it describes.

Context: This links a clip to a context which enables its point to be appreciated.

Warnings: This links a clip to negative prospects that might arise within the context of the situation it describes.

Opportunities: This link relates a clip to positive prospects that might arise within the context of the situation it describes.

Alternatives: This provides a link between the situation described in a clip and alternative or contrasting situations that might arise in a similar context.

Indicators: This links a clip to features that facilitate detection or diagnosis of the situation that it describes within the contexts in which it might arise.

After the user has watched the clip in the center—or, perhaps, instead of watching that clip—he or she may move to one of the peripheral clips by clicking on its clip node. When this happens, the center clip absorbs its other neighbors and moves to a point on the periphery; then the chosen clip node moves to the center, and a new clip screen unfolds from it, displaying its neighbors. This process can be repeated as the user browses through the clip network, travelling over links representing questions arising in each context.

4 Themes as a bridge between domain and case knowledge

The ASK TOM project, because of its scope, prodded us to look at some large-scale representation problems that have often been lost in the details in existing knowledge-intensive systems. In designing the system's indexing network, we worked simultaneously from two directions: We looked at the domain information in the clips from the top down, attempting to pick out the major agents, techniques, situations, and relationships, and to construct out of these elements some high-level models that would span the entire domain. Pragmatically, at least, we succeeded at this endeavor. Simultaneously, we worked bottom-up, starting from the individual clips and the relationships that existed among them. Again we had some success.

At this point, however, we had a problem, insofar as the nodes in our high level models could not be linked, in any meaningful way, to the clips in the low level network. A fairly crude partitioning was possible—we could, for instance, pick out twenty or so clips and say that these belonged under the node denoting the relationship between a trust department and an individual at a consulting firm. But we still needed a well-motivated approach to structuring this set itself.

Our approach to this problem, as described above, was to categorize these moderately-sized sets of clips into *themes* or *lessons*. We were steered toward this answer, in part, by our investigations of story-based models of human memory. As we examined the clip database, we realized that certain lessons cropped up repeatedly, lessons like "Know enough not to ask stupid questions," "The trust banking industry is changing in ways that we must be prepared for," "Correct trouble at its source," and so on. Often, these lessons were explicitly and consciously drawn as morals by the expert telling the story. As described earlier, these themes provided a division of the clips which was independent of, and often very different than, that provided by the big-picture nodes. By placing these themes as an additional indexing level between the big-picture models and the clips themselves, we were able to introduce enough distinctions to easily distinguish and locate most clips.

5 The need for an indexing methodology

By far the biggest impediment to the construction of the first version of ASK TOM has been the sheer magnitude of the data-analysis phase. After interviewing the expert, we had nine hours of video which we edited into about 200 clips. Each clip then had to be placed properly into the indexing network. This was accomplished in two stages: First, each clip was associated with all appropriate big-picture nodes[5]; then the clip's neighbors had to be determined, and the appropriate low level links established. We looked for neighbors in four ways: First, clips indexed under the same big-picture node were likely candidates. Second, clips that were near each other in the original interview were also liable to be linked by whatever thread led from one to the other in the interview. Third, we sorted all the clips into piles based on "subject" (a term which we never tried to define), and generated possible neighbors that way. Finally, we worked backwards: For each clip, we considered each of the eight low-level links in turn, and probed our own memories of the corpus, i.e., we tried to get reminded of other clips that would attach to the current one via the given link. This last method proved especially fruitful, providing us with some confidence that we had indeed hit upon a semantically significant set of low level links.

This whole process was arduous, especially in that it required several members of the project team to internalize the contents of the entire corpus. The construction of larger video data bases along the lines of ASK TOM will necessitate the development of alternative approaches, because it will ultimately be impossible for any one person to be familiar with, let alone internalize, all of the cases in such a system.

---

[5]Each clip was stored in an average of two places, but there was considerable variance, with some clips appearing four or five times. Each location corresponds roughly to a distinct high level question that the clip was perceived to answer.

In addition, if the construction of these indexing networks is to become commercially viable on a large scale, the process must be codified and realized in a set of software tools so that it can be carried out by people who do not have extensive training in developing representations for artificial intelligence systems.[6] The successful codification and implementation of such a process in software tools would provide evidence for the transportability and generality of the indexing vocabulary used in ASK TOM.

6 Progress on the implementation

We have completed a prototype of the ASK TOM system containing a twenty-six clip database, which is a limit imposed by using a single video disc. We have also substantially completed the analysis of our entire corpus of approximately 200 clips. This process has gone far enough that we feel confident that the big-picture models we have developed will be adequate, and that the same eight low-level links can be used across the entire corpus without alteration. The hardware consists of a Pioneer LC-V330 video jukebox connected to a Macintosh II workstation. The software is written in Supercard, using Voyager VideoStack video drivers.

7 Conclusion

ASK TOM represents a new approach to structuring access to a newly emergent kind of knowledge base, the video case library. It is based on two premises: First, that cases and stories in the form of video clips can provide much of the viscerality and memorability that is lacking from textual forms of presentation; and second, that AI theories, specifically the approaches to memory organization derived from work in case-based reasoning, can provide the structure that is essential to achieving the shared context that makes communication possible. The aspect of AI research that is crucial in providing this structure does not concern algorithms. Rather, it is the *content* of domains and tasks that is paramount.

Our eventual goal is to situate systems like ASK TOM in teaching environments which also include a simulator and a tutorial planner. The scenario we envision is one in which the planner sets up the simulator to produce an interesting situation that the student must address. In response to the developing situation and the student's actions, we want to be able to provide, when appropriate, expert advice in the form of video clips of stories which bear on the problem at hand. We view this as a crucial step towards the development of a new generation of computer-based educational software systems, which takes as its fundamental principle the idea that learners must be actively engaged in realistic problem-solving tasks, embedded in realistic situations, in order to properly assimilate lessons in a way that will enable their application in appropriate situations in the future.

[6] In some ways, our problem here is similar to the knowledge-acquisition problems faced in the construction of expert problem-solving systems.

Acknowledgments:

We thank Ryan Lanham, Chris Riesbeck, Diane Schwartz (who did the graphics design for the system), and especially our anonymous experts at Andersen Consulting for their help. The Institute for the Learning Sciences was established in 1989 with the support of Andersen Consulting, part of The Arthur Andersen Worldwide Organization.

References

Conklin, E. 1987. Hypertext: An introduction and survey. *IEEE Computer*, vol. 2, pp. 17-41.

Hammond, K. 1989. *Case-Based Planning: Viewing Planning as a Memory Task.* Academic Press, San Diego.

Riesbeck, C., and Schank, R. (eds.) 1989. *Inside Case-Based Reasoning.* Erlbaum, Hillsdale, NJ.

Schank, R. 1977. Rules and topics in conversation. *Cognitive Science*, vol. 1, pp. 421-441.

Schank, R. 1982. *Dynamic Memory: A Theory of Reminding and Learning in Computers and People.* Cambridge University Press, Cambridge, England.

Schank, R. 1990. *Tell Me a Story: A New Look at Real and Artificial Memory.* Charles Scribner's Sons, New York.

Sacerdoti, E. 1977. *A Structure for Plans and Behavior.* American Elsevier, New York.

Sussman, G. 1975. *A Computer Model of Skill Acquisition.* American Elsevier, New York.

Wilensky, R. 1982. *Planning and Understanding.* Addison-Wesley, Reading, MA.

Winograd, T., and Flores, F. 1986. *Understanding Computers and Cognition: A New Foundation for Design.* Ablex, Norwood, NJ.