Representation Issues in Multimedia Case Retrieval*

Thomas R. Hinrichs and Ray Bareiss and Brian M. Slator

The Institute for the Learning Sciences
Northwestern University
Evanston, IL 60201
{hinrichs, bareiss, slator}@ils.nwu.edu

Abstract

Video stories are an effective way to teach complex subjects, but video is opaque to a computer program. To what extent must the content of such stories be represented? We present three issues in story indexing that affect the distinctions that must be drawn in story representation and illustrate these issues with examples from ORCA, a pedagogical case-retrieval system in the domain of organisational change.

Introduction

One powerful method of teaching complex subjects is to present a student with case studies in the domain. To make these cases more compelling, they may be presented as stories in video form. An implication of video presentation is that the cases are opaque to the system; a computer program can neither interpret nor adapt the story content directly. Therefore, a natural question that faces designers of such systems is "To what extent must the content of these stories be represented?" We have explored a minimalist representation scheme to support the task of multimedia case retrieval. Along the way, we have identified several key issues that affect case representation. This paper will talk about these issues in the context of ORCA, a case-based program for teaching "change management" in business.

ORCA: A Change-Management story base

One example of a multimedia story-retrieval system is ORCA, the ORganizational Change Advisor [Bareiss and Slator, 1993; Slator and Bareiss, 1992]. ORCA is designed to teach consultants how to identify problems that may face a business, and to expose them to potential solutions to those problems. To do this, ORCA presents business 'war stories' about organisational change in response to economic and technological pressures. As the user works with a client, the client becomes a new story in the system, thus extending ORCA to serve as a 'corporate memory'.

The system's task is therefore to construct a sufficient description of the user's client such that relevant stories can be retrieved. The system must elicit this description from a novice who presumably does not know the domain as an expert would. Consequently, the system must do everything it can to guide the user in describing his client. This guidance takes three main forms in ORCA:

1. ORCA queries the user in terms of a vocabulary of domain features with which he is likely to be familiar. While ORCA hypothesizes about abstract thematic categories, the user is never directly confronted with them. Instead, categories are confirmed by asking questions that are specific to the business domain.

2. The featural description of the client is elaborated by asking the user a series of questions, rather than permitting unrestricted text input or a large form-filling interface. Each question can be answered through buttons labeled "Yes", "Probably", "Maybe" "No", or "Don't Know". This question-asking interface is sometimes called a sounding-board [Kass, 1991].

3. To make it easier for the user to decide whether a feature applies to the client, the sounding board is augmented with an interface that lets the user compare and contrast the client with the stories just seen. For example, rather than deliberating about whether the nature of work in a company is "heavily dependent on information" it may be easier to decide if it is more or less so than for another company. This can be viewed as case-based elaboration of the problem description [Slator and Bareiss, 1992].

Once a client problem is sufficiently described, the system retrieves and presents an analogous story to the user. These stories are embedded in a multimedia ASK Network [Ferguson et al., 1992] that allows the user to...
view the story and to browse through related stories by traversing relational links.

The following dialog illustrates a typical session with ORCA:

ORCA: Orca (the Organizational Change Advisor) will assist you in analyzing a client's situation as you hear stories and answer questions.

Is this a new client or an existing one?

User: New Client.

ORCA: What is the client's name?

User: RustBelt Mfg.

ORCA: Choose one or more change drivers, and then press continue to answer questions on behalf of the client.

User: (select Competitive Threats and Restructuring/Reorganization from a menu.)

ORCA: Choose an industry type that best describes your client.

User: (select Manufacturing from a menu.)

ORCA: Now it is time to answer some questions about your client.

Is it time that the organization considered updating its processes?

User: Yes

ORCA: Is planning considered an important and worthwhile effort in aiding employees with their job requirements?

User: Yes

ORCA: Does the organization allot extensive training time to its employees?

User: No

ORCA: Does the organization spend sufficient time and effort researching its change options?

User: Yes

ORCA: Is a large project effort planned?

User: Yes

At this point ORCA tells a story about a previous consulting job in which an aerospace defense contractor reorganized its factory floor in order to be more efficient and competitive. An overview of the story is presented first, with follow-up questions leading to more specific stories about how the shop floor was organized, how jobs were redesigned, and how the manufacturing process flow was refined to eliminate unnecessary steps. After viewing the story, the user is asked to compare the client to the airframe manufacturer on a feature-by-feature basis in order to more accurately retrieve additional stories.

Retrieval in ORCA

The principal mechanism in ORCA is reminding, based on association among features; ORCA was built as an experiment in minimal representation. The mechanism can be characterised in terms of the following design decisions:

1. The basic retrieval mechanism is a simple, single-step spreading activation in an associational network. When a question is answered, the feature corresponding to the question propagates activation to its nearest neighbors. The next most active feature is then chosen and presented to the user as a question.

2. Features may be confirmed at different levels of activation, and links between features have different strengths. When the user answers a question through one of the five reply buttons, the corresponding feature is confirmed at the specified activation level. For example, "Yes" confirms a feature with activation = 4, "Probably" confirms a feature with activation = 2, and so forth. When activation is propagated to neighboring features, it is attenuated by the strength of the link.

3. The network is homogeneous. All links are essentially 'reminding' links with different strengths. The idea is that if one feature is associated with the client, then it should remind the system of other features that are also likely to be relevant. The reminding links between features can be thought of as a feed-forward network that helps the system to ask relevant questions to elaborate the problem description.

4. All reminding links are bi-directional. While, conceptually, reminding links might be asymmetric (e.g., dogs remind you of fur, but not vice-versa), in working towards a simple representation, we deemed it impractical for the system builders to make these judgements. One exception to this rule is links from a type of feature called Change Drivers, which are essentially vital statistics about the client and general problem areas of concern to the client, that the user fills in initially in order to 'prime the pump'.

5. Features are propositional. In order to be compatible with spreading activation, features are represented as simple propositions, rather than as predicates, or attribute-value pairs.

6. Features are subdivided into several types. For example, Change Drivers represent reasons why a client is undergoing organizational change (e.g., merger / acquisition, new technology). Surface features represent directly observable properties of the client (e.g., "The organisation is heavily dependent on information"). In addition, we use proverbs as client-independent labels for abstract categories representing deep thematic situations (e.g., "A pig with two masters will starve") [Owens, 1990].

These design decisions have imposed certain limitations on the system, and the trade-offs have affected the indexing of ORCA. In the next section we describe the representation and indexing implications of this design.
Three issues in story indexing

Our experience in building and working with ORCA has led us to identify three critical issues in story indexing: the accuracy, efficiency, and difficulty of indexing. These criteria imply that a system should maximize the relevance of the questions it asks and the stories it tells; it should minimize the number of features it must know in order to retrieve a good story, and it should minimize the effort required to construct and index the story base. We explore each of these issues below.

Accuracy

In any story-retrieval system, the most critical issue is selecting relevant, on-point stories. In a complex, weak-theory domain such as business management there may not be a single 'best' story or a strong criterion for relevance. In fact, the criteria for relevance are probably determined by the task that the system supports. For a pedagogical problem-solving system such as ORCA, our working definition of relevance is that a story will be relevant if:

- It shares an abstract problem with the current situation.
- It provides some kind of advice or moral.
- It shares some surface similarity with the current situation (in order to facilitate mapping and transfer in the mind of the human user), and
- It is not redundant (i.e., not too similar to what’s been seen before).

Each of these criteria contributes to the quality of story retrieval, and each criterion has implications for story indexing. In ORCA, our assumption has been that the activation level of a story would be a measure of its similarity to the current client, and that the similarity of a story would serve as a measure of its relevance. The above definition provides a way to evaluate this assumption.

The first criterion suggests that there may be qualitative differences between features; features that are abstract or thematic may be more important for retrieval than other types of features. In ORCA, for example, a story tends to be on-point or analogous if it is a good exemplar of a proverb that it shares with the client’s problem description. (We treat proverbs as abstract categories of ‘business diseases’.) In ORCA, a story must bear both surface and thematic similarity to the client before that story will be told. This rule suppresses the telling of stories that are off-point, either due to vacuous surface similarity in the absence of thematic coherence, or because of opaque abstract similarity. If the telling of a story depended simply on the linear sum of the activation of its features, then there would be no way to distinguish between stories whose activation was due to a central primary proverb and those with no proverb at all.

Another potential problem is “populism”, in which a high level of activation may be due to a preponderance of weakly confirmed sources. Stories that are weakly related to everything may overshadow more specific stories that have fewer, but more strongly confirmed, sources. Such ‘promiscuous’ stories are often very general, and consequently may be less interesting or useful when they follow on the heels of more specific stories. One solution to this problem is to impose extra discipline on the indexing process, such that stories are labeled more conservatively and weak indexing links are ruthlessly excised. Another solution is to implement an algorithm for ‘aging’ the activation levels in memory, so that aggregations of weak remindings would lose reminding strength over time. A third possible solution is to avoid adding general stories to the system in the first place. However, it is not clear that general stories are useless, nor that the order of their presentation is significant. Consequently, ORCA requires that stories be thematically similar to the current problem, as well as strongly activated. Because activation levels do not distinguish between thematic and non-thematic features, we conclude that activation, by itself, is not always a good measure of similarity.

The second criterion for relevance is that a story should provide some kind of advice or moral. For example, a story about a company with an insufficiently trained workforce should either help to anticipate future problems or show how they solved their problems. What would not be as useful would be a story that merely says “Here’s another company that’s just like yours.” We have assumed that ORCA would retrieve stories that were similar to the description of the client. Because stories are indexed directly by their descriptions, this means that stories must be described in terms of the problems they address. If stories are encoded that do not explicate both a problem and its solution (as is the case, unfortunately, with some of the ORCA stories), the indexer must plausibly reconstruct the problem and describe the story in terms of that problem. If stories are encoded that do not distinguish between the problem and the solution, then the system may not be able to distinguish between stories that are relevant to the user’s problem and those that are not. For this reason, similarity, by itself, is not always a good measure of relevance.

The problem/solution distinction is a special case of the more general issue of representational granularity and scope. *Granularity* denotes the distinctions that are drawn in a representation (such as problem vs. solution). *Scope* denotes the extent of the world that is represented (such as whether or not a problem is explicitly represented in a case). For example, a story might be about a company that does an excellent job in assimilating new technology. In ORCA, this story is most likely to appear when the client is also excellent in this regard, since ORCA operates on the premise that similar stories are relevant. This is an assumption that
arises when the domain is complex, as in change management. In rich, weak-theory domains of this type, no two stories are ever identical. Stories are retrieved on the basis of two levels of similarity, but they are important because of their ability to bring new features into focus. The similarities between stories is what makes them analogous, the differences between similar stories is what makes them illustrative.

On the other hand, one could argue that such a story might best be told in response to a client whose problem is outdated or misused technology. That is, to retrieve on the basis of counter-example. This notion turns out to be difficult to capture with a homogeneous spreading activation network. A more complex predicate style representation would enable this functionality, as well as enabling the representation of mutually exclusive alternatives. A simpler, but more ad hoc solution would be to introduce into the representation a method for explicit counter-example linking between stories. Neither of these has been implemented in ORCA, and the efficacy of this trade-off is an empirical question.

One implication of this is that the truth of a propositional feature is not necessarily the same as its relevance. When associating "Yes" answers with positive activation there is danger of conflating truth with relevance. For example, when asking the question: "Does the company provide sufficient training?" a negative answer should be highly relevant, while a positive answer means that training is not a concern. This is a common problem with spreading activation and propositional features. At least two solutions are possible: The first solution is to increase the representational power (and complexity) of the system and explicitly represent the valence of a feature. Features for which truth and relevance are opposite would have an associated negative multiplier to invert the activation they pass. The ORCA solution is simply to carefully choose the feature vocabulary such that truth always corresponds to relevance. For example, a better question would be "Does the company neglect training?"? In this case, a positive reply would emphasize stories about training. We are in the process of refining our representational vocabulary to ensure this.

Efficiency

A second issue in story indexing is the efficiency of retrieval. By efficiency, we mean not merely speed, but more importantly, the number of features that must be determined in order to accurately retrieve a story. For a question-asking type of system such as ORCA, efficiency is inversely proportional to the amount of effort the user must expend to retrieve a story. The user's effort consists of the mental effort required to answer each question and the number of questions to be answered.

A question may be difficult to answer for two reasons: 1) The question may be ambiguous or vague, or 2) The answers may be too similar to discriminate easily. In either case, the user will end up deliberating over possible answers. To avoid such deliberation, the indexing features should be at an appropriate level of abstraction and should embody a vocabulary that is familiar to the user. Proverbs, for instance, are notoriously difficult to interpret. Rather than ask business consultants to describe their clients in terms of proverbs, ORCA associates each proverb with a set of domain-specific surface level features called proverb-confirming features. These features are easier for the user to recognize than the proverb itself. For example, one proverb in the system is: "The good foreman must be a good carpenter first." To confirm this proverb, ORCA asks the following questions: "Do the organization's leaders take a 'hands on' approach?" and "Do the managers understand their employees' jobs?". If the answer to either question is yes, the proverb is confirmed.

The second reason for user deliberation is that the answers may be too similar to discriminate easily. The user should not have to make difficult decisions to distinguish between alternatives that are either very similar or don't count much. For example, a question such as "Has productivity, efficiency, or quality emerged as a major concern?" will be true to some extent of any organization. Rather than forcing a fine-grained distinction, it might be better to either re-word the question or to disable the "Probably" and "Maybe" buttons. In general, small differences in judgement should not lead to large differences in behavior. For this reason, the difference between "Probably" and "Maybe" in terms of activation is kept quite small in ORCA.

The overall efficiency of case retrieval is also determined by the number of questions that the user must answer. The user should not have to answer questions that appear self-evident. The system should, at the very least, draw rudimentary inferences. To combine inference and association, we define three additional types of features and their corresponding inferences:

- **Censors** are features which, if confirmed true, completely inhibit other features. For example, the feature "Many confusing, overlapping programs compete for the organization's attention" censors the proverb "Two heads are better than one".

- **Confirmatory features** are features which, if true, automatically confirm other features. Proverb-confirming features such as "Do the managers understand their employees' jobs?" are an example of confirmatory features. This feature confirms the proverb "The good foreman must be a good carpenter first."

- **Gatekeepers** are features which, if confirmed false, inhibit other features. For example, disconfirming a gatekeeper such as "A change in leadership or top management has occurred" would automatically disconfirm it is linked to, such as "A leadership
change has resulted in a marked difference in employee motivation."

We have provided this functionality in ORCA by adding new types of non-reminding links that may connect any pair of features. The new links are triggered when the source feature is either confirmed or disconfirmed, and in turn, they either confirm or disconfirm the destination feature to which they are linked. This is perhaps the simplest means of combining inference and association, from which more complex relationships between features can be constructed.

In discussing efficiency, we have assumed that the objective is to make it easy for users to describe their problem. A case could be made, however, that deliberating over the description could be a learning experience equally valuable to hearing stories. Forcing users to think is usually a good idea in pedagogical systems. Nevertheless, we believe the user should invest his effort into interpreting his problem, rather than interpreting the meaning of features or second guessing the system, and this places a particular premium on posing questions at a level the user can easily understand and answer.

Difficulty of Indexing

A third important issue in story indexing is the difficulty of creating and indexing the case-base such that stories can be accurately and efficiently retrieved. This difficulty depends in large part on the degree to which the process can be automated and on how systematic the manual part of the process is.

The entire ORCA story base was constructed manually, on a story by story basis. Although there was a pre-existing schema for characterising cases in terms of surface features, there was no theory of how these features related to each other, nor any notion of the abstract categories exemplified by proverbs in ORCA, nor how these abstract categories related to the surface features. Thus, indexing each story was done in discrete steps. The first step involved reading the story and assigning it labels from the domain. In general, this could be done at a rate of 2-4 stories an hour. However, building the memory at the same time added considerably to the indexing effort, particularly at first. Later stories were added with less theory building, as the memory matured, but nonetheless, the effort of building this story base turned out to be quite high. This was partly because indexing stories and designing the representation were combined in the same process. Each step contributes to the difficulty of indexing. This was also because of the lack of good tools for manipulating memory. It goes without saying that cases should be described with as small a vocabulary as possible. It is a much harder thing to say how small that is. An over-large vocabulary forces the indexer to wade through many similar but often subtly different features. However, in complex domains there are quite often important differences that can seem subtle but are not. The feature set in ORCA was quite large by some measures, containing hundreds of features, but it had been adapted from a pre-existing vocabulary of organizational change, and it was feared that pruning could only be accomplished at the risk of loss in generality. In general, the vocabulary of the domain is imposed by the domain itself, and can only be modified with expertise of that domain near at hand.

To the extent possible, the case-base should also be monotonic, or additive. The addition of new cases should not interfere with the retrieval of previously indexed cases. However, this property is difficult or impossible to preserve when the domain theory is being built incrementally. In ORCA, for example, most of the difficulty of indexing consisted of defining the reminding links between features and deciding how strong these links should be.

One approach to this problem is to declare that links have clear, precise semantics. This could reduce the amount of deliberation the indexer devotes to interpreting the meaning of features and links. There are several ways that links could be interpreted:

- **Causality:** Two features could be strongly linked if one causes the other. (e.g., an "aging workforce" causes "high health-care costs".)

- **Prototypicality:** A story could be strongly linked to a category (i.e., proverb) if it is a highly prototypical exemplar of the category. (e.g., a story about the effective use of employee workgroups might be a typical exemplar of the proverbial category "Two heads are better than one.")

- **Certainty:** A link between two features could be an estimate of the indexer's subjective belief that one implies the other. (e.g., "New technology is introduced" probably implies "Employee training is critical.")

- **Frequency of co-occurrence:** A link strength could be an empirical measure of the frequency with which two features occur together in stories in the story base.

- **Proportion:** A link could represent the proportional importance or composition of a feature with respect to a category or a story. (e.g., "This company's problems are 95% training and 5% management.")

If indexers pick a single consistent interpretation, they can more readily assign link strengths between features and thereby determine what questions follow from other questions. Note that a second benefit may accrue from having a single interpretation. If the story base is sufficiently large and representative of the domain, then one of these interpretations, the frequency of co-occurrence, suggests a way to automatically construct links between features: In this case, the strength of a link would simply be the conditional probability of their co-occurrence in the story base. Under this interpretation, the next question asked would always...
correspond to the feature most likely to be true with respect to the case base. Of course, the most likely feature is not necessarily the most relevant question to ask. It may be that the best question is the one that provides the most information. In that case, discrimination might be a more natural indexing scheme.

Any “upgrading” of the representational machinery, beyond statistical frequency data which could be automatically generated, would involve considerable overhead in terms of representational complexity and would come at the expense of the simple associative properties of the memory. As originally conceived, the ORCA representation was based on association and reminding, rather than any semantic properties. This minimalist experimental assumption was an attempt at managing the complexity of a domain that does not admit of concretized rules or simple solutions. No representational upgrade has been implemented in ORCA, and the efficacy of this trade-off is another empirical question.

Future Work

Given the simplicity of its underlying mechanism, ORCA has proven to be a remarkably effective story-based training system. It tends to behave as if it were generating hypotheses and testing them systematically, although no such control strategy is built into the system. In fact, the control effectively emerges from the indexing structure of the case base. Indexing a case base in this way has proven to be very difficult and time consuming, and we are now trying to capitalize on our experience with ORCA by experimenting with alternative retrieval mechanisms and automated knowledge acquisition tools to find better indexing methodologies.

We hope to determine how the issues of accuracy, efficiency, and difficulty manifest themselves for different methods of case retrieval. To do this, we are in the process of building case retrievers based on a more general spreading activation model that integrates inference and association, another retriever based on the PROTOS model [Bareiss, 1989], and a third retriever based on variations of discrimination nets, as implemented in CYRUS [Kolodner, 1984]. We expect that a better understanding of the interaction between a control strategy and an indexing strategy will help us to more easily build accurate and efficient case bases.

We have also been extending ORCA to support knowledge acquisition through failure-driven refinement of the case base. This process involves fine-tuning the indexing structure based on user feedback during problem solving To enable the user to provide feedback, ORCA must provide a means by which the user can interrupt the system at any point and criticize its behavior. Currently, ORCA allows an expert to reject a story or question as “irrelevant”. It then takes the expert into a knowledge acquisition dialog that explains the reasons why it was reminded of a story or question and proposes strategies to repair the case base. Much work remains to be done to refine this capability and to extend the coverage of indexing problems that can be repaired.

Conclusions

One of our reasons for exploring spreading activation in story retrieval was that we expected it to be easier to build a case base since we could avoid representing the content of stories. In such a system, the representation of a story and its indexing would be one and the same. Beginning with this hypothesis, we have used the criteria of accuracy, efficiency and difficulty to examine what further representational distinctions may be needed to retrieve stories in a problem-solving domain. These distinctions include thematic vs. surface features, problem vs. solution, truth vs. relevance, and association vs. inference. In a homogeneous spreading activation network, most such distinctions need not be explicit in the representation of stories, but there are costs, in terms of representational power, associated with leaving them out.

Acknowledgements

We would like to thank Larry Birnbaum, Gregg Collins, Paul Brightbill and Laura Carpenter for their contributions to this research.

References


