Efficient Question Answering in a Hybrid System

Joachim Diederich[1,2] **Debra L. Long**[2]

[1]German National Research Center for Computer Science (GMD)
Schloß Birlinghoven, P.O. Box 1240
D-5205 St. Augustin 1, Germany

[2]Department of Psychology
University of California, Davis
Davis, CA 95616, U.S.A.

Abstract

A connectionist model for answering open-class questions in the context of text processing is presented. The system answers questions from different question categories, such as "How," Why," and "Consequence" questions. These question categories have been identified in several empirical studies (Graesser & Clark, 1985; Graesser, 1990). The system responds to a question by generating a set of possible answers that are weighted according to their plausibility. Search is performed by means of a massively parallel, directed spreading activation process. The search process operates on several knowledge sources (i.e., connectionist networks) that are learned or explicitly built-in. Spreading activation involves the use of signature messages (Lange & Dyer, 1989). Signature messages are numeric values that are propagated throughout the networks and identify a particular question category (this makes the system *hybrid*). Binder units that gate the flow of activation between textual units receive these signatures and change their states. That is, the binder units either block the spread of activation or allow the flow of activation in a certain direction. The process results in a pattern of activation that represents a set of candidate answers based on available knowledge sources.

1. Introduction.

Question answering is one of the most important functions for high-level connectionist systems. It requires fast and efficient search and inference procedures, as well as knowledge about question categories. Knowledge about different types of questions allows a system to adjust search and inference mechanisms in order to retrieve the best answer based on available knowledge. Connectionist models allow fast, parallel search plus evidential reasoning, and are a promising and natural method for efficient question answering.

In this paper, question answering in the context of text processing is investigated. It is assumed that a system has knowledge about a particular text and that it uses this knowledge as well as additional knowledge sources (i.e., generic world knowledge, knowledge about question categories) to answer open-class questions. Question answering is realized by a directed spreading activation process and the propagation of signature messages throughout a connectionist network. It allows the user to specify a *question category* (by activating a question unit) and a *question focus* (by "clamping on" one or more units that encode explicit knowledge from the text). The system responds with a set of plausible answers within 15-30 parallel updates of the network. These answers are weighted according to their plausibility; the unit with the highest activation value is considered to be the best legal answer based on the available knowledge sources.

In many cases, question answering involves explanations of "how" a certain action was performed, "why" it was performed, and what the "consequences" of an action or event were. These kinds of explanations require the highly selective activation of network units in order to generate explanation structures; a task that is extremely difficult for connectionist systems in general (Diederich, 1989a; 1990b). Our system gives semantically meaningful responses to "Why" questions, based on the connectionist representation of causal relations among propositions, and hence offers a solution to the explanation problem in connectionist networks.

The connectionist system has been implemented and tested in the context of a single text, a brief fairy-tale that has been investigated extensively by Graesser and his colleagues (Graesser & Clark, 1985; Long, Golding, Graesser, & Clark, 1989). The text can be represented as a set of explicit and implicit propositions with labeled links (as described in section 2). The system answers questions about the following story:

The Czar And His Daughters

Once there was a Czar who had three lovely daughters. One day the three daughters went walking in the woods. They were enjoying themselves so much that they forgot the time and stayed too long. A dragon kidnapped the three daughters. As they were being dragged off, they cried for help. Three heros heard the cries and set off to rescue the daughters. The heros came and fought the dragon and rescued the maidens. Then the heros returned the daughters to their palace. When the Czar heard of the rescue, he rewarded the heros.

This paper is organized as follows. First, we briefly describe a model of human question answering called QUEST. QUEST was developed by Graesser and Clark (1985) and is a symbolic model of text comprehension and question answering. Second, we introduce a connectionist system loosely based on the QUEST model. Third, a propositional network representation of The Czar and His Daughters text is described (the propositional network is compiled into the connectionist system). Fourth, the structured connectionist network is introduced in detail, the question answering mechanism is outlined and first results are presented. Finally, advantages and shortcomings of the approach are discussed.

2. The QUEST Model of Human Question Answering

Graesser and his colleagues have developed a model (called "QUEST") that describes how adults answer open-class questions in the context of text comprehension, such as "why," "how," "when," and "what-happened-next" questions (Graesser & Franklin, 1991). These researchers have investigated the processes involved in human question answering and the knowledge structures that supply answers. QUEST can be segregated into three major components. The first component involves knowledge representation and information sources. Graesser assumes that a network of propositional-like units (called a "conceptual graph structure") is constructed during reading. The propositional-like units or "statement nodes" are categorized into node types (i.e., state, event, goal, style specification) and are connected by directed relational arcs (i.e., Consequence, Implies, Reason, Initiate, Outcome, Manner). The statement nodes are constructed on the basis of a variety of information sources. These information sources include the explicit text and both general and specific world knowledge structures. The world knowledge structures are networks that contain goal/plan hierarchies, causal networks, taxonomic hierarchies, and descriptive structures. When comprehending a text, the reader continually expands the conceptual graph structure by adding explicit text information and information that is inferred from the world knowledge structures.

The second component involves question identification and categorization. That is, the model translates the question into a logical form and assigns it to a question category (e.g., why-event, why-action, how-event, how-action). Graesser assumes that there are a finite number of question categories and that each question category has its own unique question answering procedure. Each question category elicits a different type of information. For example, why-action questions elicit information about reasons and motives for intentional actions, whereas why-event questions elicit information about causal antecedents.

The third component involves convergence mechanisms. Convergence mechanisms compute the subset of nodes in the information sources that serve as relevant answers to a question.

3. The text and its representation.

Nodes in the QUEST-like propositional network representation are categorized as Goals, Events, or States. Nodes are connected with directed and categorized links representing the relationships Consequence, Implies, Initiate and Reason. Explicit actions in the text are represented by a goal node connected to an event node via an Outcome arc. It should be noted that the connectionist realization of the propositional network does not include all of the propositions that are necessary for a complete question answering system. Although the system does not include all propositions, the model has approximately 3000 units and 100,000 links. Thus, the model certainly includes a sufficient number of propositions to enable it to answer a variety of questions about the text.

At the connectionist level, the system currently has no representation of predicates (i.e., labeled links), *but instead uses empirical information on the strength of associations* for both, text comprehension and question answering. It turns out that this weaker representation is sufficient for the generation of semantically meaningful activation patterns that represent possible answers (see section 4.3 for further details).

4. Connectionist text processing.

4.1 The general architecture of the network.

The connectionist network has three modules. First, a module for the representation of text-based knowledge is constructed by compilation. In addition, a network for generic world knowledge that uses both built-in and learned knowledge is incorporated (see section 4.2 and Diederich (1989a, 1990a) for a full description of this architecture). Finally, the network contains a module with single units representing question categories. Although most of the knowledge in the connectionist network is built-in, the system has the ability to adapt and modify structure by recruitment learning. Diederich (1989a) describes a recruitment learning approach by use of the generic world knowledge module.

4.2 The connectionist representation of generic knowledge.

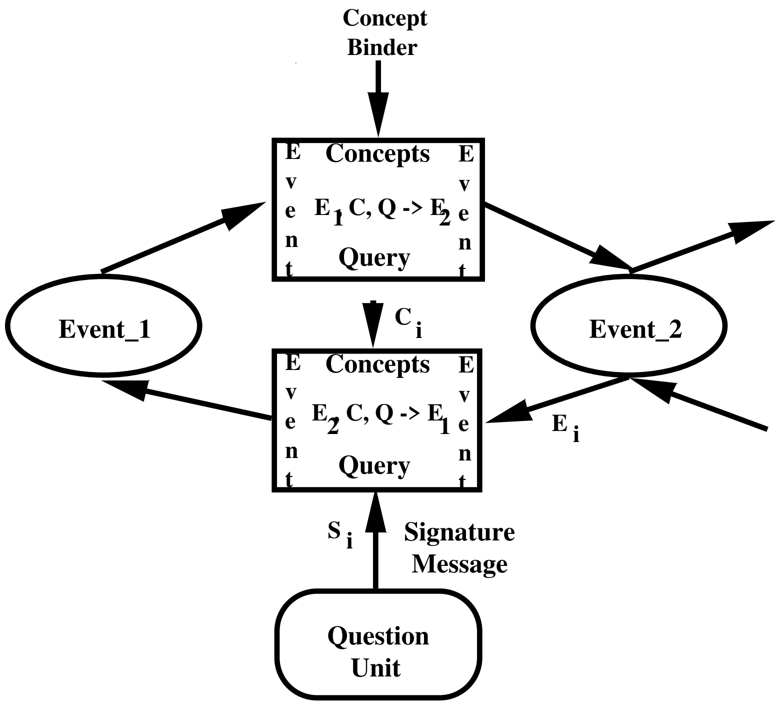
The generic world knowledge module has four "spaces." Specifically, there is a space for the representation of structured objects, a space for the representation of attributes, a space for the representation of values of attributes, and a space that represents instances.

Each space is an n-layer network with mutual excitatory connections between neighboring layers (in both directions) and inhibitory connections within layers (see Diederich 1989a for a formal description plus complete results). Units represent a concept, attribute, value of an attribute and constitute a hierarchy. That is, object, attribute, and value units are embedded in a multi-layer network and form a straight-forward spreading activation hierarchy. Units are not only connected to units in the same space, but are also connected to units in other spaces. Object, attribute and value units are connected by concept binder units. A concept binder unit has three or more "sites" that receive symmetric input-lines and has an activation function that requires input from at least two sites to generate a positive output. That is, a concept binder unit gates the flow of activation between a concept, attribute and value unit.

4.3 The connectionist representation of textual knowledge.

The connectionist representation of textual knowledge is constructed by compilation. That is, the QUEST-network is compiled into a connectionist system. Association strengths are translated to weights. Because there are no direct connections among textual units in the connectionist system, labeled arcs are translated to paths of connections among textual units in which each individual weight equals the value for the corresponding inference class.

The Consequence and Reason arcs are the strongest links because these connections are sufficient for on-line text processing. Also, only nodes with incoming Consequence and Reason arcs make good answers during question answering (Golding et al., in press). The Outcome arc is translated to a weaker connection because this link is only used to connect goal and event units in order to create an action. The Implication arc is the weakest link and plays almost no role in connectionist question answering: All important processes operate on the connectionist realization of event chains (Figure 1) and goal hierarchies, and these structures use Reason, Consequence and Outcome links.



$$\textbf{If } E_i > 0 \ \&\& \ C_i > 0 \ \&\& \ S_i \in \{s_1 \ldots s_n\} \textbf{ Then output = max Else output = 0}$$

Figure 1: The binder units structure that is used to represent an event chain. Binder units allow the flow of activation if a site has input and a signature has arrived.

Propositional units that represent textual knowledge are connected by binder units. A binder connects two textual units at the connectionist level. Binder units may be in one of several states and can enable or block the flow of activation in a particular direction (i.e., from one propositional unit to another). Furthermore, there are connections between propositional units at the textual level and concepts and actions (attributes) at the generic knowledge level (i.e., there are connections from propositional units to conceptual units). Binder units among textual units at the connectionist level may be in one of several states depending on the mode in which the network is operating. Throughout this paper, we only consider the "question answering" mode in which the state of binder units depends on the question category (the system can also operate in several on-line text processing modes plus two learning modes). If the question category is a "how-action", for instance, then binder units allow the flow of activation downward in goal hierarchies and backward in event chains. See the next section for details.

5. Question answering.

Question units are specialized nodes that allow the propagation of signature messages (Lange & Dyer, 1989). Each question unit spreads a unique signature (i.e., an integer that is not changed during propagation) which denotes the question category (i.e., why-action, how-action). Currently, the system contains five question units and therefore five signatures. Note that the number of signatures used for question answering equals the number of question categories known to the system. Because of the minimal number of signatures, the complexity of binder units is bound and very little time is required for the coding of signatures (i.e., matching input to known signature values).

Question answering is achieved by (1) "clamping on" a question unit with an unit function that allows signature propagation, and by (2) choosing a question focus that is a connectionist textual unit. Next, the network is updated until the activation pattern is stable. The textual unit with the highest activation is retrieved as the best available answer; units with lower activation are considered as less plausible answers based on the available knowledge sources.

5.1 Answering "How"-Questions.

Figure 2 demonstrates the procedure for answering how-action questions. The question unit "How-action" is clamped on with an activation function that allows signature propagation. The question focus for a how-action question is an explicit goal proposition connected to an explicit event proposition (i.e., the text-based representation of an action). "How-action" has connections to all binder units in goal-hierarchies and event chains. These binder units receive the signature and as a response change their state. Binder units for a downward flow of activation in goal hierarchies enter a state that allows activation to spread in that direction. Binder units for an upward flow of activation in goal hierarchies enter a state that allows no activation propagation. In event chains, the forward binders are blocked and the backward binders are enabled. The spreading activation process generates an excitation pattern in which subgoals of the question focus are activated as well as antecedent events.

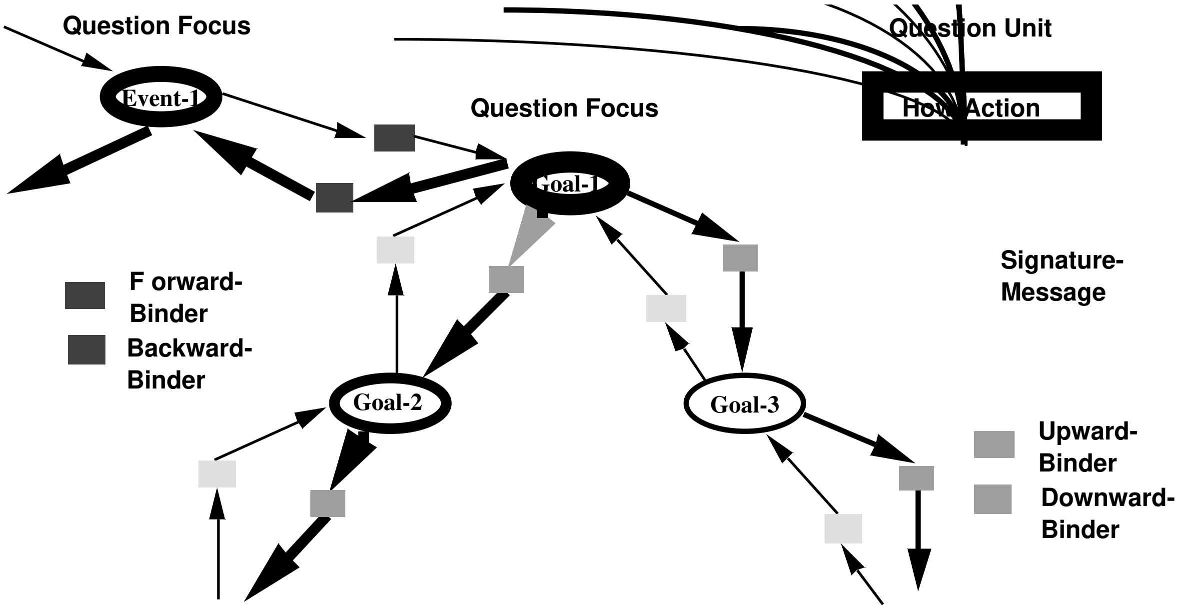


Figure 2: Spread of activation in answer to a how-action question. The question unit and the question focus units (i.e., an explicit goal proposition and an explicit event proposition) are clamped on and activation spreads downward along a goal hierarchy and backward in an event chain. As an example: Goal-1 = Daughters enjoy themselves; Goal-2 = Daughters look at wildlife; Goal-3 = Daughters get exercise; Event-1 = Daughters sing.

5.2 Answering "Why"-Questions.

The question focus for a why-action question consists of an explicit goal proposition connected to an explicit event proposition. "Why-action" has connections to all binder units in goal-hierarchies and event chains. These binder units receive the signature message and in response binder units involved in upward flow of activation enable activation to spread up the goal hierarchy. Binder units involved downward flow of activation in goal hierarchies block activation propagation. In event chains, the forward binders are blocked and the backward binders are enabled. Spreading activation generates an excitation pattern in which superordinate goals and antecedent events receive activation.

6. Results.

The system is fully implemented by use of the Rochester Connectionist Simulator. All connectionist units, with exception of binder units, simply calculate the weighted sum of inputs. Asking a "How" or "Why" question results in a stable pattern of activation in 15-30 network updates. The patterns of activation show *a clear distance effect,* i.e., units close to the question focus units receive more activation than distant units (Figure 3). This distance effect is also predicted by the QUEST model. The model presented here, however, takes full advantage of the inherent massive parallelism of a neural network.

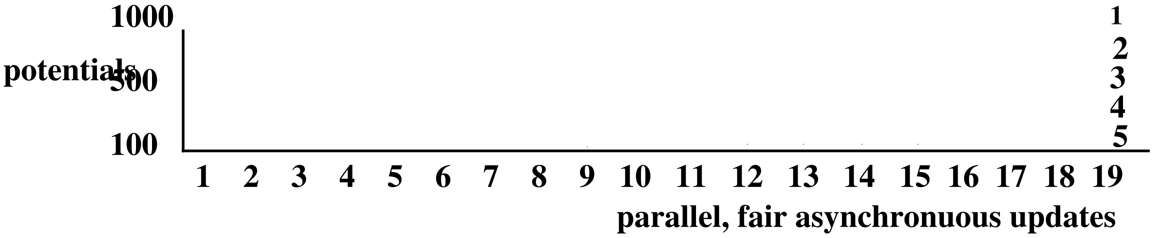


Figure 3: The potentials of goal hierarchy units with distance to source (i.e., distance from the question focus unit). Question type is "How."

7. Discussion.

The proposed solution is efficient because *it generates a set of answers in a few network updates.* In particular, the use of signature messages *reduces network complexity significantly* (i.e., saves binder units), however, their use requires that binder units match incoming signatures. That is, speed of processing is five times slower (there are 5 signatures) compared to a conventional connectionist network in which each unit calculates the weighted sum of inputs only. There is no danger that the complexity of units will increase with network size because the number of signatures known to the network equals the number of question categories. It is possible to save even more units by using coarse coding over binder units. No attempt has yet been made to save textual binders, but coarse coding is used for concept binders.

References

Diederich, J. (1989a) Instruction and High-Level Learning in Connectionist Networks. Connection Science. Vol.1, No.2, 161-180.

Diederich, J. (1990a) Spreading Activation and Connectionist Systems for Natural Language Processing. Theoretical Linguistics, Vol.16, No.1, 25-64.

Diederich, J. (1990b) An Explanation Component for a Connectionist Inference System. In: L. Aiello (Ed.): ECAI-90, 9th European Conference on Artificial Intelligence, 222-227, Pitman Publishing, London.

Golding, J. M., Graesser, A. C., & Millis, K. K. (in press) What makes a good answer to a question?: Testing a psychological model of question answering. Discourse Processes.

Graesser, A. C. (1990) Investigation of human question answering. Office of Naval Research: Technical Report 90-1.

Graesser, A. C. & Clark, L. F. (1985) Structures and procedures of implicit knowledge. Norwood, NJ: Ablex.

Graesser A. C. & Franklin, S. P. (in press) QUEST: A cognitive model of question answering. Discourse Processes.

Long, D. L., Golding, J. M., Graesser, A. C., & Clark, L. F. (1989) Goal, event, and state inferences: An investigation of inference generation during story comprehension. In: A. C. Graesser and G. H. Bower (Eds.), Inferences and text comprehension:. The psychology of learning and motivation, vol. 25. New York: Academic Press.

Lange, T.E. & Dyer, M.G. (1989) High-Level Inferencing in a Connectionist Network. Connection Science, Vol.1, No.2, 181-217.