DOCUMENT RESUME

ED 268 009 SE 046 537

AUTHOR Faletti, Joseph
TITLE Systematic Representation of Biology Knowledge.
PUB DATE Oct 85
NOTE 7p.; Paper presented at the Annual Meeting of the National Association for Research in Science Teaching (59th, San Francisco, CA, March 28-April 1, 1986). For Related papers presented in the same set, see SE 046 538-540.
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Artificial Intelligence; *Biology; Cognitive Processes; *Cognitive Psychology; College Science; Computer Oriented Programs; *Concept Formation; Higher Education; *Knowledge Level; *Science Education
IDENTIFIERS Science Education Research

ABSTRACT

A major goal of science education is to turn novices (students) into experts (scientists or science literates) with a minimum amount of pain, effort, and time. However, transfer of biology knowledge from instructor to student usually results in a loss of the rich interconnections that an expert has. The papers in this set describe efforts to restore and exercise these interconnections in the knowledge of the student. This first paper provides a rationale for and overview of the general approach taken which is to: (1) design a formal and systematic representation of biology knowledge in the form of a semantic network of concepts and the relationships between them; (2) build a set of computer-based tools to support the design and construction of semantic networks for particular areas of biology; and (3) build a set of computer-based games to present semantic networks to students along with tasks which will require students to exercise the interconnections among concepts. Also discussed are: previous studies of methods for improving integration of new information; previous work in knowledge representation; and reasons for selecting the particular representation method used (which stem from practical constraints arising either from instructional considerations or from the capabilities of current computer systems). (Author/JN)

I. Systematic Representation of Biology Knowledge

Joseph Faletti

Department of Computer Science
University of California — Davis
Davis, CA 95616

October 1985

Abstract

A major goal of science education is to turn novices (students) into experts (scientists or science literates) with a minimum amount of pain, effort and time. However, the transfer of biology knowledge instructor to student usually results in a loss of the rich interconnections that an expert has. The papers in this set describe efforts to restore and exercise these interconnections in the knowledge of the student.

This paper first provides a rationale for and overview of the general approach taken which is to (A) design a formal and systematic representation of biology knowledge in the form of a semantic network of concepts and the relationships between them; (B) build a set of computer-based tools to support the design and construction of semantic networks for particular areas of biology; (C) build a set of computer-based games to present semantic networks to students along with tasks which will require the students to exercise the interconnections among concepts.

Previous studies of methods of improving integration of new information are briefly reviewed. Previous work in knowledge representation is then discussed. Since the representation method chosen is somewhat simpler than the current trend in knowledge representation, the reasons are discussed, all stemming from practical constraints arising either from the goal of use in instruction or from the capabilities of current computer systems.

I. Systematic Representation of Biology Knowledge

Joseph Faletti

Department of Computer Science
University of California — Davis
Davis, CA 95616

October 1985

1. Rationale.

A major goal of science education is to turn *novices* (students) into *experts* (scientists or science literates) with a minimum amount of pain, effort and time. In the biological sciences, the bulk of the knowledge to be acquired is *descriptive* and less easily formalized than many physical sciences. Without a formal or systematic representation, this transfer of knowledge from the instructor's head to the student's head is likely to be inaccurate.

The accepted methods for this transfer require the expert to translate knowledge into a linear sequence of pictures and language. The students must then process these pictorial and verbal representations and attempt to reconstruct within their heads the facts that were in their instructor's head. The students often consider this type of learning to be a matter of memory and approach it by trying to memorize individual pieces of knowledge one at a time. After such an approach, they can usually comply when asked for a fact, but are seldom able to demonstrate understanding of the interconnections and broader organization of the knowledge. That is, *something gets lost in the translation.* Somehow the rich interconnections among the individual pieces of knowledge in the instructor's head seldom re-appear in the student's head.

The papers in this set describe efforts to formalize the biological knowledge of experts, focusing on these often unexpressed and lost interconnections. The overall goals of the project are (A) to develop a formal method of representing the descriptive content of biology; (B) to develop instructional materials to use this formal representation to communicate and exercise the organization of ideas in such a way that the students acquire richly connected biology knowledge; and (C) to develop tools to assist instructors in developing representations for new collections of knowledge for their students' use.

2. Previous Work on Integration of Facts

There have been many studies of the effect of various instruction techniques on the integration of facts during learning. We mention only a few which seem most relevant to our current work. [Hayes-Roth and Thorndike, 1979], [Hayes-Roth and Walker, 1979]). and [Walker and Meyer, 1980] have shown that integration of facts can be improved by controlling the distance between facts that should be connected, by using shared identical wordin_ in related facts, and by controlling the depth of the information structure being built at any one time.

3. Previous Work on Representing Knowledge

Research on the representation of knowledge has been done within three fields with disparate goals: linguistics, artificial intelligence (AI), and cognitive psychology.

3.1. Knowledge Representation in Linguistics

Some of the earliest work in representing knowledge was done in linguistics, notably by [Fillmore, 1968, 1971] whose work focused on representing the meanings of individual sentences by making explicit the relationships between the central action represented by the verb (such as "hit") and the various *case roles* (actor, object, instrument, etc.) in the action. Since the focus was on actions by humans, this work is not specifically applicable to the representation of most descriptive biology knowledge. However it introduced the main theme that runs through all work on knowledge representation: knowledge is made of *concepts* plus their *relations* to other concepts. These relations may generally be reduced to a relatively small set.

3.2. Knowledge Representation in Artificial Intelligence.

Knowledge representation has recently become the central problem in the field of artificial intelligence ([Bobrow and Collins, 1975], [Brachman, 1985]. It has been a major component of research on tasks ranging from text understanding (e.g. [Schank and Abelson, 1977], [Wilensky, 1978], [Findler, 1981]) and problem solving (e.g. [Sacerdoti, 1977], [Wilensky, 1983], [Faletti, 1982]) to speech (e.g. [Walker, 1978]) and vision (e.g. [Minsky, 1975], [Winston, 1975], [Ballard and Brown, 1982]). However, each task requires quite different kinds of knowledge and its form and organization has generally varied significantly.

Again in most of this work, it was often possible to capture huge subsets of knowledge with limited sets of concepts and relations, although not all researchers have tried or accepted this as desirable. For example, Schank and Abelson succeeded in representing most human actions with about a dozen primitive actions plus a small set of relations including roles (e.g. actor, object, donor, recipient, direction) and other relations (e.g. instrument, causation). There has been little chance as yet to synthesize these recent results but their common agreement (in addition to the need for concepts and relations among them) has been that larger richer organizations of concepts are needed for intelligent application of knowledge. Almost every paper cited above recommends a different structure for the larger organizations of concepts however, and primary point of agreement is that the result is a collection of complex networks of concepts connected by relations.

The resulting representations of knowledge can all be viewed as *semantic networks*: collections of concepts multiply-connected in complex networks.

As in linguistics, most of this work has been involved in representing actions rather than descriptive knowledge. The principal exception is work on vision which has tried to represent the physical relations among the parts of objects but unfortunately not the functional relationships. The precise description of functional relationships in complex processes or of details about individual types of objects has generally been avoided (primarily because the tasks at hand did not require such detailed descriptive representations but also partly because of the difficulties perceived in their representation).

Only a few AI researchers have worked on knowledge representation of complex processes or the functional relationships of objects. [Rieger, 1976] has represented the operation of devices such as light bulbs, bicycles and flush toilets. Once again, a small set of relations resulted, the central ones being 8 kinds of causal relations to relate state changes in the parts of a device during its operation. However, a complete network for a bicycle begins to get unwieldy and hard to understand or use, so Rieger also suggests that richer, more structured organizations of the concepts (possibly subnetworks for various subsystems of the bicycle) would reduce the complexity of the resulting networks.

3.3. Knowledge Representation in Cognitive Psychology.

Quite a few researchers in cognitive psychology have either helped to develop or adopted the various representation schemes used in AI and have proceeded to use them to describe and test models of thinking and learning. Most have settled on the "least common denominator" which is essentially a semantic network. [Stewart, 1984] has used pre-constructed semantic networks to present biology concepts to students in college courses with positive student reaction. [Novak, 1980, 1981] has asked seventh grade students to study science material by constructing their own semantic networks also with positive student reaction

4. Representation Method Chosen.

Based on this previous work and our goals and practical considerations discussed below, we have chosen to represent biology knowledge in a semantic network in which each concept is a node connected to related nodes by arcs labeled with the name of the relationship. We have worked to keep the number of relations small (the number of concepts is determined by the biological content). However, not surprisingly, as we add new areas of biology we find that each area requires new relations. We have now represented significant portions of knowledge in molecular biology, ecology, and human anatomy (the first two are described in related papers by (a) Fisher, and (b) Garb, Fisher and Faletti). The representations in molecular biology and ecology have been pilot-tested in a section of introductory biology with good response from the students. The formal assessment of their usefulness is under way. We have built three prototype computer programs for the manipulation and exercise of semantic networks. Two used textual descriptions of the relations with limited success. The third used pictorial representations with much greater success (see the related paper by Callman, Faletti, and Fisher).

5. Small Set of Relations

Our choice of representation method was affected by the constraint that it not be so hard to learn that it significantly increases the time to learn or volume of the biology knowledge involved. Hence it is particularly desirable to have as small a set of relations as possible so that the students can learn them quickly and immediately get to the task of learning biology. This was found to be a reasonable goal in molecular biology (see the related paper by Fisher), but was much more difficult with ecology where the focus of the science is on the kinds of complex relationships between (sets of) organisms (see the related paper by Garb et al).

6. Ease of Understanding and Construction.

An additional constraint was that the network must be easy to read and absorb when presented on flat surfaces such as paper or a terminal screen. Many of the more structured methods of knowledge representation are quite difficult to read and to represent on a flat surface and so had to be eliminated. However, we found that some of the features eliminated for these reasons became desirable in the representation of ecology (see the related paper by Garb et al).

Networks must be easy enough to construct that the task of construction does not discourage their use. We have experimented with paper-and-pencil methods with acceptable success, although some pain. The computer prototypes have each been difficult to use in one way or another, but we feel we have learned enough from each that useful tools with the best from each are possible.

7. References

Ballard, D. and C. Brown. 1982. *Computer Vision*. Prentice-Hall, Englewood Cliffs.

Bobrow, D. G. and A. Collins (eds.). 1975. *Representation and Understanding*. Academic Press, New York.

Brachman, R. 1985. *Readings in Knowledge Representation*. Kauffman, Menlo Park.

Faletti, J. 1982. PANDORA -- A Program for Doing Commonsense Planning in Complex Situations. In the *Proceedings of the National Conference on Artificial Intelligence*. Pittsburgh, PA. August. 1982.

Fillmore, C. J. 1968. The Case for Case. In *Universals in Linguistic Theory*. E. W. Bach, and R. H. Harms (eds.). Holt, Rinehart, and Winston, New York. 1-88.

Fillmore, C. J. 1971. Verbs of Judging: An Exercise in Semantic Description. In *Studies in Linguistic Semantics*, C. J. Fillmore and J. T. Langendoen (eds.). Holt, Rinehart, and Winston, New York. 273-290.

Findler, N. 1979. *Associative Networks: The Representation and Use of Knowledge by Computers* Academic Press, New York.

Minsky, M. 1975. A Framework for Representing Knowledge. In *The Psychology of Computer Vision*. P. H. Winston, (ed.). Addison-Wesley Publishing Company, Reading, Mass. pp. 211-277.

Rieger, C. 1976. An Organization of Knowledge for Problem Solving and Language Comprehension. *Artificial Intelligence 7*, pp. 89-127.

Sacerdoti, E. 1977. *A Structure for Plans and Behavior*. Elsevier, Amsterdam.

Schank, R.C. and Abelson, R.P. 1977. *Scripts, Plans, Goals, and Understanding*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Walker, D. (ed). 1978 *Understanding Spoken Language*. North-Holland, New York.

Wilensky, R. 1978. *Understanding Goal-Based Stories*. Ph.D. Thesis. Technical Report 140, Computer Science Department, Yale Univ., New Haven. September 1978.

Wilensky, R. 1983. *Planning and Understanding: A Computational Approach to Human Reasoning.* Addison-Wesley Publishing Company, Reading, Mass.

Winston, P. H. 1975. Learning Structural Descriptions From Examples. In *The Psychology of Computer Vision.* P. H. Winston, (ed.). Addison-Wesley Publishing Company, Reading, Mass 157-209.