Functional Categorization of Knowledge: Applications in Modeling Scientific Research and Discovery

Sakir Kocabas

The central thesis of my dissertation (Kocabas 1989)[1] is that in complex systems, descriptive and definitive knowledge can be organized into functional categories; this categorization provides clarity and efficiency in representation and facilitates the integrated use of various methods of learning. I describe a formalism for organizing knowledge into such functional categories and some of its implementations. In this formalism, descriptive scientific knowledge is classified into seven categories. The categorization formalism allows complex propositions to be analyzed into their simple constituents; in turn, these constituents can be maintained in their categories. They can then be combined using a simple transformation function to form complex constructs such as frames and schemata. The methodology facilitates the implementation of knowledge-level methods of learning such as similarity-based learning, explanation-based learning, and conceptual clustering. It simplifies the identification and resolution of conflicts in knowledge systems.

To demonstrate its viability, I implemented this knowledge organization in four computational models of scientific reasoning and discovery that operate in astronomy and particle physics, incorporating various methods of learning and theory revision. These systems are described in terms of their knowledge organization and representation, rules of action, input and output, and results. Finally, I implemented the categorization in a system, CER, which constitutes a move toward building a comprehensive computational model of scientific research. CER operates in the domain of high-temperature oxide superconductivity, using a large amount of general and specific knowledge.

Knowledge Representation and Learning

Chapter 2 examines the existing methods of knowledge representation and learning. Several knowledge representation methods are examined, followed by a review of learning. I then discuss the relationships between knowledge organization, representation, and learning.

The main conclusion drawn from this survey is that some methods of knowledge organization and representation are more suitable for certain forms of learning and discovery. Briefly, the concept-based organization of knowledge in frames facilitates formal (for example, taxonomic and inheritance-based) reasoning and learning by analogy. Logic-based methods facilitate logical and extralogical inference and the implementation of similarity-, explanation-, and exclusion-based learning and conceptual clustering. The organization of knowledge into strings of messages or matrixes facilitates classification and recognition.

Chapter 3 focuses on how various methods of representation and learning have been used in several computational models of discovery and examines the relationships between knowledge organization, representation, and learning in such systems. These systems include Dendral and Meta-Dendral (Buchanan and Feigenbaum 1978), MOLGEN (Friedland 1979), GLAUBER (Langley, Simon, Bradshaw, and Zytkow 1987), STAHL (Zytkow and Simon, 1986), STAHLp (Rose and Langley 1986), BACON (Langley et al. 1987), AM (Lenat 1979) and EURISKO (Lenat 1983), PI (Thagard and Holyoak 1985), and KEKADA (Kulkarni and Simon 1988). The systems are reviewed in several groups in accordance with the types of discovery tasks and their learning and discovery methods.

Functional Categorization of Knowledge

Chapter 4 introduces the method of categorization based on the logico-philosophical work of Wittgenstein (1974) and the empirical cognitive studies of Piaget (1971). In language, propositions can be distinguished from one another by their methods of verification, function, grammatical labels (that is, prefixes) that can meaningfully be attached to them, and comparisons and contrasts with prototypical examples.

This chapter explains how definitive and descriptive scientific knowledge can be organized into the following seven categories: (1) logical, (2) formal, (3) mathematical, (4) grammatical, (5) theoretical-hypothetical-empirical, (6) historical, and (7) factual. A set of logical and linguistic criteria is used in this categorization.

> *…some methods of knowledge organization and representation are more suitable for certain forms of learning and discovery.*

Logical knowledge includes the rules of logic, logical function definitions, and the logical relationships between concepts in language and is totally domain independent. This type of knowledge is provable only in the logical grammar of language; empirical methods of verification, such as observation and experimentation, do not apply to it. Certain prefixes, such as "by definition," are meaningful for logical propositions, but certain others, such as "according to the hypothesis," "probably," and "according to (so and so)," are not meaningful.

Formal knowledge includes concept definitions and member-class and

class-superclass relationships. Formal statements are not verified by empirical methods; their truth or falsity is implied by the conceptual structure of the language in which they are expressed. Certain prefixes are applicable to formal statements, but others are not. Formal knowledge is less abstract than logical knowledge because some formal knowledge can be domain dependent.

Mathematical knowledge involves symbolic and numeric relationships in language. In principle, mathematical statements are verified by logical, formal, and mathematical methods but not by empirical methods.

Grammatical knowledge involves metastatements about concepts and expressions in language. They are not verifiable by logical, formal, or empirical methods but are accepted as true by a set of values in language (or the deep grammar of language) or by intuition.

Theoretical-hypothetical-empirical knowledge involves general statements about natural phenomena. These statements are testable against such phenomena. The prefix "by definition" is not meaningful for the statements in this category, but the prefixes "according to the theory" and "according to the hypothesis" are meaningful. Generalizations and specializations are applicable to theoretical statements.

Historical knowledge includes the records of past experience. They are verified or falsified by the methods of historical study. *Factual knowledge*, however, reflects the current state of affairs about the world. Factual statements are verified by observation. They describe specific facts or events; therefore, specialization does not apply to them.

Applications in Modeling Scientific Discovery

In chapter 5, I describe four systems of scientific reasoning and discovery. The first one is QR-1, which builds new concepts from a set of basic concepts in astronomy and formulates relationships between them. The others—AR-4, BR-3 (Kocabas 1991), and CR-3—simulate reasoning and discoveries by theory formation and revision in the field of particle physics. In these systems, knowledge is organized in accordance with the principles of categorization introduced in chapter 4. All four systems use exactly the same knowledge representa-

tion and can be joined together. In fact, the third and the fourth systems, BR-3 and CR-3, were developed as an integral part of the second one, AR-4. Such different systems were developed independently to demonstrate the applicability of the proposed knowledge organization formalism.

Toward a Comprehensive Model of Scientific Research

Chapter 6 describes CER, a theory-driven system that is being developed as a more detailed model of scientific research and discovery than the earlier systems. It is designed to incorporate most of the basic operations (or elements) of scientific research. CER partially simulates the research and discoveries in high-temperature superconductivity through a series of activities such as formulating and choosing research goals; choosing research strategies, methods, and processes; and proposing experiments.

CER's descriptive and definitive knowledge is organized in accordance with the categorization principles introduced in chapter 4. However, the system's prescriptive knowledge (for example, its rules of action) is organized into a set of operators (or homunculi), whereby each operator carries out a task through a series of activities.

Summary

In complex systems, descriptive and definitive knowledge can be organized into functional categories; this categorization provides clarity and efficiency in representation and facilitates the integrated use of various methods of learning and discovery. Knowledge represented in categorized predicate statements can automatically be translated into frames and frames into predicate statements.

As a result of its bringing clarity to the representation of descriptive and definitive knowledge, categorization also facilitates the organization of prescriptive knowledge that must interact with the descriptive and definitive knowledge. Because of the uniformity that is achieved through propositional representation, descriptive and prescriptive knowledge are more easily integrated in this organization. This ease of integration is expected to facilitate the construction of more complex systems.

Acknowledgments

This work benefited from the general

comments of M. B. Sandler and H. A. Fatmi and from the detailed comments of Pat Langley on the programs described in chapters 5 and 6 of the dissertation.

References

Buchanan, B. G., and Feigenbaum, E. A. 1978. Dendral and Meta-Dendral: Their Application Dimension. *Artificial Intelligence* 11:5–24.

Friedland, P. 1979. Knowledge-Based Experiment Design in Molecular Genetics. In Proceedings of the Sixth International Joint Conference on Artificial Intelligence, 285–287. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence.

Kocabas, S. 1991. Conflict Resolution as Discovery in Particle Physics. *Machine Learning* 6:277–309.

Kocabas, S. 1989. Functional Categorization of Knowledge: Applications in Modeling Scientific Research and Discovery. Ph.D. diss., Dept. of Electronic and Electrical Engineering, King's College London, Univ. of London.

Kulkarni, D., and Simon, H. A. 1988. The Processes of Scientific Discovery: The Strategy of Experimentation. *Cognitive Science* 12:139–175.

Langley, P.; Simon, H. A.; Bradshaw, G. L.; and Zytkow, J. M. 1987. *Scientific Discovery: Computational Explorations of the Creative Processes*. Cambridge, Mass.: The MIT Press.

Lenat, D. B. 1983. EURISKO: A Program That Learns New Heuristics and Domain Concepts. *Artificial Intelligence* 21(1–2): 61–98.

Lenat, D. B. 1979. On Automated Scientific Theory Formation: A Case Study Using the AM Program. In *Machine Intelligence, volume 9*, eds. J. Hayes, D. Michie, and L. I. Mikulich, 251–283. New York: Halstead.

Piaget, J. 1971. *The Language and Thought of the Child*. London: Routledge & Kegan Paul.

Rose, D., and Langley, P. 1986. Chemical Discovery as Belief Revision. *Machine Learning* 1:423–452.

Thagard, P., and Holyoak, K. 1985. Discovering the Wave Theory of Sound: Inductive Inference in the Context of Problem Solving. In Proceedings of the Ninth International Joint Conference on Artificial Intelligence, 610–612. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence.

Wittgenstein, L. 1974. *Wittgenstein's Lectures on the Foundations of Mathematics, Cambridge, 1939*. Sussex: Harvester.

Zytkow, J. M., and Simon, H. A. 1986. A Theory of Historical Discovery: The Construction of Componential Models. *Machine Learning* 1:107–137.

Sakir Kocabas is a research scientist with Marmara Research Center, Gebze, Turkey. His research interests include machine learning, scientific discovery, and the philosophy of science.