Parsimony Analysis – a Novel Method of Structured Knowledge Representation for Teaching Computer Science

S P Maj (Corresponding author)

School of Computer Science and Security Science, Edith Cowan University

2 Bradford Street, Mount Lawley, Western Australia, 6050, Australia

Tel: 61-8-9370-6277   E-mail: p.maj@ecu.edu.au

D Veal

School of Computer Science and Security Science, Edith Cowan University

2 Bradford Street, Mount Lawley, Western Australia, 6050, Australia

Tel: 61-8-9370-6295   E-mail: p.maj@ecu.edu.au

Abstract

Knowledge may be structured in a manner that may either handicap or improve learning outcomes. Knowledge construction is therefore an important aspect of the learning and assessment process. This paper is primarily concerned with structural knowledge (knowing how) of which there are numerous representations and mapping methods each with their associated strengths and weaknesses. This paper presents the novel application of parsimony analysis for organizing and defining structural knowledge within the field of computer science. Work to date has clearly demonstrated there are considerable potential pedagogical advantages to this approach.

Keywords: Computer science education, Parsimony analysis, Cladistics, Cognitive load

1. Introduction

According to schema theory knowledge is stored as mental constructs called schemas (Rumelhart 1980). Schemas are arranged into interrelated networks called schemata. Memory can be considered therefore as schemata and learning is acquisition and the organization of schemata. There are different types of knowledge. Declarative knowledge (knowing that) is the ability to define or describe objects but does not imply understanding (Ryle 2009). Declarative knowledge is often defined using schemas (Rumelhart 1980). Procedural knowledge (knowing how) is the application of declarative knowledge and entails the application of schemas. Schemas must be selectively interpreted and employed to undertake a given task. Structural knowledge (knowing why) provides the conceptual basis for understanding and hence links declarative and procedural knowledge (Diekhoff 1983). Structural knowledge is also referred to as cognitive or internal knowledge. Meaning and hence understanding is implied by the pattern of schema relationships. All objects are identified, characterized and to various degrees related. Learning consists of building these new knowledge structures. Ideally knowledge should be related to existing knowledge hence reinforcing earlier learning. Hence learning outcomes may be improved if knowledge is structured. However the acquisition of deep conceptual knowledge is considered a complex scientific problem (Ohlsson 2009).

2. Structural Knowledge

A wide variety of different theories and methods have been employed to classify, represent, assess and convey structural knowledge each with their associated advantages and disadvantages (Jonassen 1993). All of these methods attempt to control complexity and explicitly define relationships in order to encourage higher order learning. Typically these methods are based on similarity data also referred to as semantic proximities (Nagy, 1984) or quantifiable measure of closeness (Keppens 2008). Concepts may be defined by characters (attributes) of which three types have been defined – intrinsic, functional and relational (Klausmeirer 1992). An intrinsic character is an invariant property of a class of things. The systematic grouping of concepts for instructional purposes may be based on either successive or coordinate concepts (Merrill 1977). Successive concepts share a common, super-ordinate parent concept with the associated defining characters. Coordinate concepts share characteristics with other concept groups.

Content structures are designed to convey structural knowledge and hence improve learning outcomes by organizing text into hierarchically structured relationships (Meyer 1985). Similarly the 'frames' and 'slots'

theory organizes information into data structures (frames) that can be filled with specific information (slots) (Minsky 1975). By contrast elaboration theory provides specific guidelines for content selection, sequencing and relationship identification (Reigeluth 1979). Spatial-semantic methods are often used because displays can concurrently display relationships (Kulhavy 1996), (Winn 1991). Spatial-semantic maps have been used as the basis of study and learning strategies (Larkin 1987). One technique employed is concepts maps (Novak 1984). Concept maps have both advocates and critics. It has been demonstrated that concept maps are relatively easy to use (Novak 2005). However other studies suggest concept maps do not accurately assess knowledge (Kinchin 2005), (Kinchin 2000). Recent studies indicate that computer-based instruction and learning, if based on cognitive learning theory, may improve not only instructional methods but also learning outcomes (Kiewra 2005), (Kiewra 2009).

It is important not only to consider how to teach a given subject but also how to assess student knowledge. Ideally assessment must be valid, reliable and efficient. A key element of assessment is to accurately measure not simply memory but also assimilation and understanding. There are a range of different assessment methods that include: multiple choice tests (Cordes 1994); (Parrish 1995); self-assessment (Whiddett 2000) and mastery-bridge tasks (Urban-Lurain 1999).

However regardless of the instructional and assessment methods employed the problem they all must address is to both systematically define classification characteristics and a quantitative method of defining and testing relationships – in a simple manner. Cladistic analysis has been employed with considerable success in the field of taxonomy in order to analyze and structure extremely complex systems. This paper demonstrates how the well established cladistic principles may also be used to structure knowledge within the field of computer science and hence potentially improve learning outcomes.

3. Cladistic Analysis – an overview

Biological systems are extremely diverse and complex. Taxonomy is the science and practice of classification. Classifying things into a hierarchical relationship of sub-type and super-type categories is potentially useful for helping to understand diverse and complex systems. Taxonomic principles, originally developed for the classification of living organisms, have successfully been used in other diverse disciplines such as manufacturing, biogeography and linguistics (McCarthy 2000), (Ganapathy 2006), (Rexova 2003).

Cladistics is a method of classification used in taxonomy to group taxa (named groups with common characteristics) into hierarchical sets and subsets and is sometimes referred to as phylogentic systematic (Henning 1965). The aim of cladistics is to hypothesis relationships that can be expressed in a branching diagram (cladogram). Using the principle of parsimony information may be ordered and represented in an optimal relationship. Information is defined by characters which may be plesiomorphic (primitive) or apomorphic (advanced or derived). Plesiomorphic and apomorphic are relative terms.

Sister groups are categorized by identifying apomorphic characters called synapomorphies and plesiomorphic characters called symplesiomorphies. Automorphic characters uniquely define terminal taxa and hence are not employed in the analysis. The possible relationships between characters, which define taxon groupings, are: consistent, congruent and homoplastic. A consistent character provides a unique definition that may be reinforced by another congruent character. A homoplastic character is inconsistent and conflicts with other characters. Finally there are three different character groups: monophyletic, paraphyletic and polyphyletic. A monophyletic group is defined by synapomorphies; a paraphyletic group is defined by symplesiomorphes or as a partially complete monophyletic group. A polyphyletic group is based on homoplastic characters.

Cladistic analysis orders synapomorphies into a hierarchy with the greatest number of characters arranged in the simplest and the most parsimonious way. Character distribution is considered as the number of steps in a cladogram. Fewer steps represent a more efficient hierarchical classification. Parsimony is the criterion for choosing between different hierarchical classification structures. Parsimony analysis becomes rapidly computationally intensive as the number of characters employed increases. However, software is available such as Phylogenetic Analysis Using Parsimony (PAUP) that runs on a standard PC. Consider four different taxa (A, B, C, D) with six different characters (u, v, w, x, y, z) (table 1).

Using PAUP the most parsimonious construction can be calculated (figure 1). The parsimonious topology, sometimes called a tree or cladogram, represents:

- Explicit overall structure and relationships

- Simplest hierarchical structure

- Simplest logical sequence between taxa

- Minimum number of character changes

Other, less parsimonious topologies are more complex sub-optimal solutions. Parsimonious analysis has potential pedagogical applications in the context of cognitive load. Cognitive load is the amount of mental activity imposed on limited working memory. Cognitive load theory is concerned with the instructional implications of information structures and the associated cognitive schemata (Sweller 1988), (Paas 2003). Understanding and measuring cognitive load is a fundamental challenge for cognitive science (DeLeeuw 2008). According to some authors there are three kinds of cognitive processing during learning that may contribute to cognitive load – extraneous, intrinsic and generative (Sweller 1999), (Sweller 2005). Poorly designed instructional material leads to extraneous processing in which cognitive processes do not support learning objectives. The quality of intrinsic processing depends upon the complexity of the material used i.e. the number of interacting elements. This paper suggests that using parsimonious analysis it is possible to sequence instructional material that minimizes cognitive load by logically sequencing and hence reducing the complexity of the material.

4. Parsimony Analysis in Computer and Engineering Science Education

The Cisco Network Academy Program (CNAP) is an international program for teaching network technology and defines the global standard for curriculum within this field. This program is supported by recommended text books, on-line material and assessments. The learning sequence for teaching routing protocols in this material was analyzed according to page occurrence (table 2) (Cisco 2003). Routing protocols such as EIGRP and OSPF are complex and cannot be understood without the explanation of a lot of preliminary concepts such as: Administrative Distance, Metric, vector etc. Despite this these complex protocols are introduced in the introductory chapters.

A more detailed analysis indicated that routing protocols are categorized according to the characters: metric, type (distance vector, balanced hybrid or link state), class and loop prevention (table 3). The metric for RIP is explained but no details are provided for either EIGRP or OSPF metrics. Loop prevention is defined as a characteristic of OSPF. However all three routing protocols are automatic and hence have loop prevention mechanisms. The structural knowledge provided by this chapter is arguably incomplete, inconsistent and not logically sequenced.

4.1 Cladistic analysis

The routing protocols above were analyzed for Cladistic based characters and taxa. Six taxa (table 4) were selected along with six associated defining characters (table 5).

All these routing protocols are designed to provide information about remote networks (character 1). Static and default router configuration are both manual. All other protocols obtain information about remote networks automatically (character 2). It is possible to represent a character as a binary value (present/absent) or a as multi-state value (0, 1, 2). Metric is a tri-state character; 0 no metric; 1 simple metric and 2 composite metric. Other characteristics, such as partial routing table updates (character 6), represent improvements in a routing protocol. Taxa and characters were entered into a matrix (table 6).

Even with a relatively small data set this represents a large number of different ways the taxa may be ordered using these characters. This ordering is pedagogically important because this represents the structured knowledge with the associated teaching sequence. Using PAUP v4 these taxa and characters were analyzed. The parsimonious analysis, based on the taxa/character matrix (table 6), resulted in 105 trees being evaluated with the score of the worst tree being 14 and the best tree with a length of 7 (table 7). In the most parsimonious tree (tree length 7) there were no homoplastic characters.

The parsimonious cladogram (Figure 2) represents a structured and logical teaching sequence with a minimum cognitive load. This represents a minimum cognitive load because each the sequence is hierarchically structured. The static routing protocol is the simplest protocol and represents the plesiomorphic (most primitive) character set. This plesiomorphic set is inherited by other more advanced apomorphic (advanced or derived) character sets. In effect students are able to systematically scaffold knowledge. In this context new knowledge reinforces existing knowledge. The limitations of both static and default routing are addressed by the characters: metric, update and automatic. These are common characters of all the more advanced routing protocols. However some characters, such as metric (composite) uniquely define the advanced routing protocol EIGRP.

Furthermore there are various phylogeny programs, such as McClade, that can selectively provide alternative tree diagrams via a Graphic User Interface. Using McClade software the parsimonious classification tree can be selectively displayed for individual characters.

5. Parsimony Analysis compared to other methods

A similar analysis of routing protocols was conducted using a range of different techniques: semantic maps (Fisher 1990); causal interaction and concept maps (Jonassen 1993); cross classification tables (Frase 1969); semantic feature analysis (Anders 1986); pattern and spider maps (Hanf 1971). All the techniques and methods could be used to represent relationships and hence define explicit structured knowledge for routing protocols. Most were easy to learn and use with supporting software for some methods. However none of the methods analyzed provided rigorous methods for defining concepts and characters. Furthermore none of the methods provided a rigorous, quantifiable mechanism for determining and calculating structured relationships (table 8). By contrast using parsimony analysis it is possible to define the optimum learning sequence. Namely static and default addressing represent the plesiomorphic character set and hence the initial topics to be taught. The limitations of these manual protocols are addressed by the more advanced apomorphic character sets of automatic updates with an associated simple metric of the RIP v1 routing protocol. RIP v2 inherits all these more primitive characters but incorporates an additional more advanced feature i.e. classless addressing. Finally the problems associated with RIP v2 are solved by advanced features such as partial updates, hierarchical topologies and composite metrics. Because each apomorphic character set inherits characters from the plesiomorphic characters leaning is scaffold i.e. new learning reinforces earlier learning. Significantly it can be proven that this learning sequence represents the optimum cognitive load.

6. Conclusions

Learning outcomes have variously been classified and defined (Bloom, Engelhart et al. 1956), (Biggs and Collis 1989). In order to achieve these outcomes a wide variety of different theories and instructional methods have been developed and employed – with various degrees of success. Similarly there are many different methods for displaying and representing structured knowledge each with relative strengths and weaknesses. However none of the methods analyzed provided systematic, quantifiable and easy to use procedures for defining and representing structured knowledge. This is important because quality learning depends upon cognitive load. This paper suggests that it is possible to quantifiably minimize the complexity of the material being taught and hence minimize the cognitive load using parsimony analysis. Using parsimony analysis it is possible to define the optimum learning sequence. Advantages of parsimony analysis include:

- Defines logical learning sequence with minimum cognitive distance
- Quantitative model
- Automated tool available
- Automated tool allows selective reordering according to different criteria

However, the parsimony software tools are complex and require the user to have some knowledge of classification. In particular the user must be able to be able to distinguish between taxa (concepts) and characters (attributes). Further work is being undertaken to evaluate student learning outcomes for instructional material based on parsimonious analysis.

References

Anders, P. L., Bos, C. S. (1986). Semantic feature analysis: An interactive strategy for vocabulary development and text comprehension. *Journal of Reading, 29*, 610-616.

Biggs, J., & Collis, K. (1989). Towards a Model of School-based Curriculum Development and Assessment Using the SOLO Taxonomy. *Australian Journal of Education, 33*(2), 151-163.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives - The Classification of Educational Goals*. New York: David McKay Company, Inc.

Cisco. (2003). *CCNA 1 and 2, Companion Guide*. Indianapolis: Cisco Press.

Cordes, D., Parrish, A., Vrbsky, S. (1994). *Breadth-oriented outcomes assessment in computer science*. Paper presented at the National Educational Computing Conference.

DeLeeuw, K. E., Mayer, R. E. (2008). A Comparision of Three Measures of Cognitive Load: Evidence for Seperable Measures of Intrinsic, Extraneous, and Germane Load. *Journal of Educational Psychology, 100*(1), 223-234.

Diekhoff, G. M. (1983). Relationship judgements in the evaluation of structural understanding. *Journal of Educational Psychology, 75*, 227-233.

Fisher, K. M. (1990). Semantic Networking: The new kid on the block. *Journal of Research in Science Teaching, 27*(11), 1001-1018.

Frase, L. T. (1969). Paragraph organisation of written materials. The influence of conceptual clustering upon level of organisation. . *Journal of Educational Psychology, 60*, 394-401.

Ganapathy, G., Goodson, B., Jansen, R., Le, H., Ramachandra, V., Warnow, T. (2006). Pattern Identification in Biogeography. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 34*(4), 334-346.

Hanf, M. B. (1971). Paragraph organisation of written materials. The influence of conceptual clustering upon level of organisation. *Journal of Educational Psychology, 60*, 394-401.

Henning, W. (1965). Phylogenetic Systematics. *Annual Review of Entemology, 10*, 97-116.

Jonassen, D., Beissner, K., Yacci, M. (1993). *Structural Knowledge: Techniques for Representing, Conveying and Acquiring Structural Knowledge*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Keppens, J., Hay, D. (2008). Concept map assessment for teaching computer science programming. *Computer Science Education, 18*(1), 31-42.

Kiewra, K. A. (2005). *Learn how to study and SOAR to success*. Upper Saddle River, NJ: Pearson, Prentice Hall.

Kiewra, K. A. (2009). *Teaching how to learn: The teacher's guide to student success*. Thousand Oaks, CA: Corwin Press.

Kinchin, I., DeLeji, F., Hay, D. (2005). The evolution of a collaborative concept mapping activity for undergraduate microbiology students. *Journal of   Further and Higher Education, 29*(1), 1-14.

Kinchin, I., Hay, D. (2000). How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational Research, 42*(1), 43-57.

Klausmeirer, H. J. (1992). Concept learning and concept teaching. *Educational Psychologist, 27*(3), 267-289.

Kulhavy, R., Stock, W. A. (1996). How cognitive maps are learned and remembered. *Annals of the Association of American Geographers, 86*, 123-145.

Larkin, J., H., Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science 11*, 65-99.

McCarthy, I., Ridgway, K. (2000). Cladistics: a taxonomy for manufacturing organisations. *Integrated Manufacturing Systems, 11*(1), 16-29.

Merrill, M. D., Tennyson, R. D. (1977). *Teaching concepts: An instructional design guide*. Englewood Cliffs: NJ: Educational Technology Publications.

Meyer, B. J. F. (1985). Signalling the structure of text. In D. H. Jonassen (Ed.), *Technology of Text* (Vol. 2). Englewood Cliffs: NJ: Educational Technology Publications.

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill.

Novak, J. (2005). Results and implications of a 12-year longitudinal study of science concept learning. *Research in Science Education, 35*, 23-40.

Novak, J., Gowin, D. (1984). *Learning how to learn*. New York: Cambridge University Press.

Ohlsson, S. (2009). Meaning Change, Multiple Routes, and the Role of Differentiation in Conceptual Change: Alternatives to Resubumption? *Educational Psychologist Review, 44*(1), 64-71.

Paas, F., Renkl, A., Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist, 38*(1), 1-4.

Parrish, A., Vrbsky, S., Cordes, D., Fortner, C. (1995). *Assessing Computer Knowledge amoung college students*. Paper presented at the National Educational Computing Conference.

Reigeluth, C. M. (1979). In search of a better way of organising instruction: The elaboration theory. *Journal of Instructional Development, 2*, 8-15.

Rexova, K., Frynta, D. Zrzavy, J. . (2003). Cladistic analysis of languages: Indo-European classification based on lexicostatistical data. *Cladistics, 19*, 120-127.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, Bruce, B. C., Brewer, W. F. (Ed.), *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics,*

artificial intelligence, and education. Hillsdale: NJ: Lawrence Erlbaum.

Ryle, G. (2009). *Collected Papers* (Vol. II). New York, NY: Routledge.

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science, 12*, 257-285.

Sweller, J. (1999). *Instructional design in technical areas*. Camberwell, Victoria, Australia: ACER Press.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 19-30). New York: Cambridge University Press.

Urban-Lurain, M., Weinshank, D. J. (1999). *Matering computing technology: A new approach for non-computer science majors*. Paper presented at the Annual Meeting of the American Educational Research Association.

Whiddett, R. J., Jackson, B. X., Handy, J. A. (2000). Teaching Information Systems Management Skills: Using Integrated Projects and Case Studies. *Computer Science Education, 10*(2), 165-177.

Winn, W. D. (1991). Learning from maps and diagrams. *Educational Psychologist Review, 3*, 211-247.

Table 1. Taxa-character matrix

Character	u	v	w	x	y	z
Taxa						
A						
B	•	•				•
C	•		•	•	•	•
D	•		•		•	

Table 2. Cisco book chapter page occurrences

	Taxa	Chapter page occurrences
1	Static	1, 3, 4, 5, 6, 7, 8, 15
2	Default	4, 6, 7, 8, 9, 15
3	RIP	3, 5, 8, 14
4	EIGRP	3, 4, 5, 9, 14
5	OSPF	3, 5, 9, 12, 14

Table 3. Cisco routing protocol characteristics

Taxa	Chapter defined characteristics
RIP	Simple metric (hop count)
	Distance vector
EIGRP	Classless
	Balanced hybrid
OSPF	Classless
	Loop prevention
	Link state

Table 4. Routing taxa

	Taxa	Definition
1	Static	All routing information must be entered manually.
2	Default	Manual method for entering a default route.
3	RIP v1	Routing Information Protocol version 1
4	RIP v2	Routing Information Protocol version 2
5	EIGRP	Enhanced Interior Gateway Routing Protocol
6	OSPF	Open Shortest Path First routing protocol

Table 5. Routing characters

	Character	Definition
1	Remote	Networks can either be directly connected or remotely connected Remotely connected network must be discovered using a routing protocol. This is a binary character – remote or not remote.
2	Automatic	Routing protocols must be configured manually, others automatically discover remote networks. This is a binary character.
3	Classless	Some routing protocols only allow variable subnet boundaries. This is a binary character.
4	Metric	Routing protocols may employ a metric to determine path selection to remote network. Metric derivation may be simple or composition. This is a tri-state character: 0 no metric 1 simple metric 2 composite metric.
5	Hierarchical	Some routing protocols allow networks to be divided into smaller systems. This is a binary character
6	Updates	Routing protocols distribute routing table information between routers. This is a tri-state character: 0 no routing update 1 full routing table exchange 2 partial routing table exchange

Table 6. Taxa/Character matrix

	Character					
	1	2	3	4	5	6
Taxa	Remote 0 directly 1 remote	Automatic 0 manual 1 automatic	Classless 0 classfull 1 classless	Metric 0 absent 1 simple 2 composite	Hierarchical 0 flat 1 hierarchical	Update 0 absent 1 full 2 partial
Static	1	0	0	0	0	0
Default	1	0	0	0	0	0
RIP v1	1	1	0	1	0	1
RIP v2	1	1	1	1	0	1
EIGRP	1	1	1	2	1	2
OSPF	1	1	1	1	1	2

Table 7. Parsimonious analysis

Tree length	Number of tree occurrences
7	1
8	2
9	4
10	12
11	14
12	10
13	42
14	20
Mean	12.18
Standard deviation	1.61

Table 8. Structured knowledge modeling techniques

|	Advantage	Disadvantage
Semantic maps	Define hierarchical relationships Easy to learn and use	No rigorous method or technique for defining hierarchical relationships
Causal interaction maps	Define causal relationships between variable. Easy to learn and use	No rigorous method or technique for defining causal relationships
Concept maps	Defines linked hierarchal relationships Easy to learn and use	Maps may become complex No rigorous method or technique for defining causal relationships
Cross classification tables	Tabulates similarities and differences concepts using characters Easy to learn and use	No rigorous method for defining concepts or characters
Semantic feature analysis	Tabulates similarities and differences between concepts using character states	No rigorous method for defining concepts or characters
Pattern maps	Defines linked hierarchical relationships	No rigorous method for defining hierarchical relationships
Spider maps	Defines highly structured hierarchical relationships – concepts, sub concepts etc	More difficult to learn and use. No rigorous method for defining hierarchical relationships
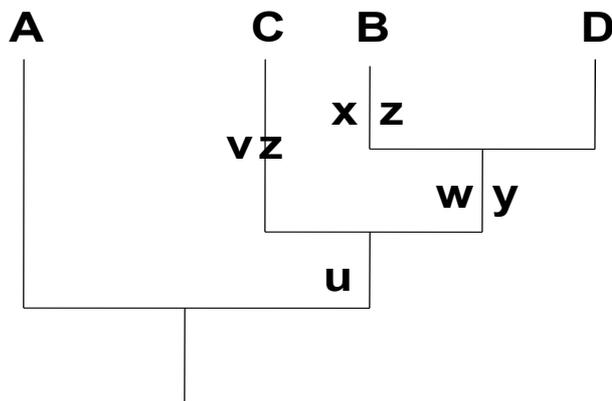
Figure 1. Parsimonious topology

Figure 2. Parsimonious cladogram