Semantic categorization of content and process concepts relevant to computer science education

Zendler, Andreas ✉
University of Education Ludwigsburg, Germany (zendler@ph-ludwigsburg.de)

Klaudt, Dieter
University of Education Ludwigsburg, Germany (klaudt@ph-ludwigsburg.de)

Spannagel, Christian
University of Education Heidelberg, Germany (spannagel@ph-heidelberg.de)

Reuter, Thomas
University of Education Ludwigsburg, Germany (tms.reuter@yahoo.de)

Abstract

A major task in planning the computer science curriculum is the specification of teaching and learning contents. This work needs to be based on knowledge of the content and process concepts central to the discipline of computer science. These central concepts are applicable or observable in multiple domains of computer science, can be taught on every intellectual level, will be relevant in the longer term, and are related to everyday language and/or thinking. Two empirically based catalogues of central content concepts (e.g., *algorithm, system, process*) and central process concepts (e.g., *problem solving and problem posing, analyzing, classifying*) for computer science education have recently been proposed. This article uses discriminant analysis techniques to provide a semantic categorization of both the content and the process concepts. On this basis, conclusions can be drawn about how individual groups of content and process concepts differ semantically.

Keywords: computer science education; central content concepts; central process concepts

Semantic categorization of content and process concepts relevant to computer science education

1. Introduction

Knowledge of the central contents and processes of computer science plays a key role in the planning of computer science instruction. To date, the contents of computer science education have primarily been discussed in the context of *fundamental ideas*. According to Schwill (1994), fundamental ideas are applicable or observable in multiple domains of a discipline (*horizontal criterion*), can be demonstrated and taught on every intellectual level (*vertical criterion*), will remain relevant in the longer term (*criterion of time*), and are related to everyday language and/or thinking (*criterion of sense*). Various catalogues of fundamental ideas or basic concepts for computer science education have been proposed (Nievergelt, 1980, 1990; Knöß, 1989; Schwill, 1994; Baumann, 1996; Denning, 2003; Modrow, 2003; Wursthorn, 2005; Armoni and Ginat, 2008). However, most of these catalogues are based on the subjective judgments of a single author or small group of authors. In recent empirical studies (Zendler & Spannagel, 2008; Zendler, Spannagel, & Klaudt, 2008; Zendler, Spannagel, & Klaudt, 2011; Zendler, McClung, & Klaudt, 2012) central content and process concepts for computer science education have been identified by asking experts to rate a set of concepts with respect to Schwill's four criteria. The advantage of this approach is that expert opinions can be averaged, and single subjective impressions are eliminated.

The specification of central content concepts can be assigned to the *structure of the discipline* approach to curriculum planning (Bruner, 1960), in which learning contents are derived from the basic concepts of a discipline and their interrelations. Another approach focuses not on contents, but on processes, which play a key role in science. The *process as content* approach is based on the work of Parker and Rubin (1966), which has more recently been revisited by Costa and Liebmann (1997a, 1997b, 1997c). The two approaches – *structure of the discipline* and *process as content* – are not mutually exclusive, but complementary: Both content and process concepts must be taken into account in the planning of the computer science curriculum. The semantics of the content and process concepts identified in the studies by Zendler and Spannagel (2008) as well as Zendler, Spannagel, and Klaudt (2008) have not yet been described in detail. The purpose of this article is therefore to categorize this content and process concepts according to semantic criteria.

In Section 2, we outline the findings of the two studies on which the present analyses are based. In Section 3, we present our discriminant analysis of the data and discuss the results. Finally, in Section 4, we summarize our findings and outline prospects for future research.

2. Background: Two empirical studies

The findings of the empirical studies by Zendler and Spannagel (2008) as well as Zendler, Spannagel, and Klaudt (2008) provide the basis for our semantic categorization of the content and process concepts relevant to computer science education.

In a first study, Zendler and Spannagel (2008) identified the content concepts of central relevance to computer science education. To this end, computer science professors ($N = 37$) were asked to rate 49 concepts retrieved from the ACM Computing Classification System[1] with respect to Schwill's four criteria. The study identified three clusters ("winner" clusters or W clusters) comprising a total of 18 content concepts: cluster CW1 contains the concepts *problem, data, computer, test,* and *algorithm*; cluster CW2 the concepts *process, system, information, language, communication, software, program, computation, structure,* and *model*; and cluster CW3

---

[1] Available from http://www.acm.org/class/1998/ (last retrieved on January 6, 2013).

just the three concepts of *network, error*, and *hardware*.

In a second study, Zendler, Spannagel, and Klaudt (2008) used an analogous procedure to identify the process concepts of central relevance to computer science education. This study was based on a list of 44 processes that Costa and Liebmann (1997d) proposed as relevant to general education, which were again rated by computer science professors ($N = 24$) with respect to Schwill's four criteria. Again, three winner clusters were identified: cluster PW1 contains the processes *finding relationships, classifying, problem solving and problem posing, investigating, analyzing,* and *generalizing*; cluster PW2 the processes *finding cause-and-effect relationships, creating and inventing, categorizing,* and *transferring*; and cluster PW3 the processes *communicating, presenting, collaborating, questioning, ordering,* and *comparing*.

Figure 1 presents the "winner" clusters of the 18 content and 16 process concepts identified in the two studies. In addition, the figure reports the mean scores for each concept on the horizontal criterion, the vertical criterion, the criterion of time, and the criterion of sense. Although the previous analyses allow the content and process concepts to be assigned to clusters, the *characteristics* of the clusters identified have not yet been investigated. The present study aims to close this knowledge gap by means of discriminant analysis.

		horizontal criterion	vertical criterion	criterion of time	criterion of sense			horizontal criterion	vertical criterion	criterion of time	criterion of sense	
	Content concepts						**Process concepts**					
CW1	problem	3.70	3.86	4.30	4.08		finding relationships	4.13	3.38	4.04	3.96	PW1
	data	3.84	3.89	4.41	3.95		classifying	3.96	3.42	4.25	3.88	
	computer	3.57	4.05	4.46	4.14		problem solving and ...	4.17	3.50	4.38	4.13	
	test	3.49	3.78	3.73	4.05		investigating	3.83	3.58	4.21	3.67	
	algorithm	4.24	4.08	4.62	3.54		analyzing	4.50	3.17	4.50	3.83	
							generalizing	4.29	3.04	4.33	3.88	
CW2	process	3.69	3.11	4.32	3.65		finding cause-and-effect ...	3.58	3.04	3.88	3.83	PW2
	system	3.97	3.32	4.35	3.78		creating and inventing	3.50	3.13	3.88	3.63	
	information	3.86	3.46	4.38	4.19		categorizing	3.79	3.17	3.79	3.75	
	language	3.76	3.19	4.35	4.03		transferring	3.83	2.96	3.50	3.58	
	communication	3.46	3.35	4.16	3.89							
	software	3.70	3.62	4.35	3.22		communicating	3.42	3.54	3.46	4.29	PW3
	program	3.65	3.54	4.32	3.59		presenting	3.58	3.50	3.17	4.42	
	computation	3.51	3.24	4.30	3.27		collaborating	3.42	3.50	3.38	3.96	
	structure	3.70	3.32	3.76	3.24		questioning	3.25	3.42	3.33	3.83	
	model	4.05	3.00	4.19	3.14		ordering	3.42	3.63	3.79	4.08	
CW3	network	2.84	3.16	4.14	4.05		comparing	3.04	3.50	3.71	4.04	
	error	3.00	3.54	3.89	3.78							
	hardware	2.78	3.59	4.30	3.43							

Figure 1. The "Winner" clusters of content and process concepts

3. Semantic categorization of content and process concepts

The content and process concepts were categorized according to semantic criteria as follows: We first conducted group comparisons of the clusters of content concepts and of the clusters of process concepts. We then performed discriminant analyses on the content and process concepts using the methods of descriptive discriminant analysis discussed, for example, by Timm (2002) and Huberty and Olejnik (2006). All analyses were conducted with SPSS 17.0.

3.1 Database

An 18×4 data matrix with the three clusters of content concepts – CW1 ($n_1 = 5$), CW2 ($n_2 = 10$), and CW3 ($n_3 = 3$) – and a 16×4 data matrix with the three clusters of process concepts – PW1 ($n_1 = 6$), PW2 ($n_2 = 4$), and PW3 ($n_2 = 6$) – served as the database (see Figure 1). The dependent variables were the horizontal criterion,

vertical criterion, criterion of time, and criterion of sense.

3.2 Group Comparisons

The group comparisons involved two one-factor 4-variate variance analyses. Wilks' Λ was transformed to an F statistic and used to establish significance. The F value for the three clusters of content concepts was F = 9.73; that for the three clusters of process concepts was F = 13.59. Both F values were greater than the critical F value at α = .01: $F_{(8, 24)} < 3.37$ and $F_{(8, 20)} < 3.57$, respectively. In other words, the three clusters of content concepts and the three clusters of process concepts differed significantly.

3.3 Discriminant functions and dimensionality

To further examine the differences identified, we calculated linear discriminant functions. Table 1 reports the following statistics for the discriminant functions: the relative discriminating power λ, Wilks' Λ criterion, the χ^2 statistic (transformed from Wilks' Λ), the number of degrees of freedom df, and the probabilities p for significance testing.

Table 1

Determining Dimensionality and Degree of Discrimination

Content concepts

Discriminant function	λ	Wilks Λ	$\chi2$	df	p
# 1 - #2	0.66	0.06	39.03	8	< 0.01
# 2	0.34	0.30	16.12	3	< 0.01

Process concepts

Discriminant function	λ	Wilks Λ	$\chi2$	df	p
# 1 - #2	0.83	0.02	42.82	8	< 0.01
# 2	0.17	0.30	13.78	3	< 0.01

The tables show that the discriminant functions determined for the content and process concepts display different discriminating power. For the content concepts, for example, the discriminating power of discriminant function #1 is 66%,[2] whereas that of discriminant function #2 is 34%. The distribution for the process concepts is more one-sided: the discriminating power of discriminant function #1 for the process concepts is 83%; that of discriminant function #2 is just 17%. The discriminant functions for both the content and the process concepts are significant at the α = .01 level.

3.4 Discriminant coefficients

To categorize the discriminant functions according to semantic criteria, we next calculated standardized discriminant coefficients with respect to the four dependent variables (horizontal criterion, vertical criterion, criterion of time, and criterion of sense). Table 2 shows the discriminant coefficients of discriminant functions #1 and #2 for the content and process concepts.

Semantics of the discriminant functions for the content concepts

The horizontal and the vertical criterion are of great relevance for discriminant function #1 for the content concepts, as is apparent from the discriminant coefficients (0.85 and 0.83, respectively). The criteria of time and of sense are also relevant for this discriminant function (-0.48 and 0.47, respectively), although it should be taken into account that the discriminant coefficient for the time criterion is negative. In contrast, the discriminant coefficients indicate that only the horizontal and vertical criterion are of relevance for discriminant function #2

---

[2] The discriminating power is calculated using the following formula:

$$s = 100\% \frac{\lambda_s}{\lambda_1 + \lambda_2 + ... + \lambda_s + ... + \lambda_r}$$

(-0.74 and 0.64, respectively); the high negative value of the horizontal criterion is notable here.

Table 2

Discriminant Coefficients of the Dependent Variables

Content concepts

Discriminant function	Horizontal Criterion	Vertical Criterion	Criterion of Time	Criterion of Sense
#1	0.85	0.83	-0.48	0.47
#2	-0.74	0.64	-0.01	0.15

Process concepts

Dscriminant function	Horizontal Criterion	Vertical Criterion	Criterion of Time	Criterion of Sense
#1	1.04	0.42	0.67	-0.53
#2	0.28	0.94	0.01	0.38

Semantics of the discriminant functions for the process concepts

The horizontal criterion and the criterion of time are of particular relevance for discriminant function #1 for the process concepts, as is reflected by the discriminant coefficients (1.04 and 0.67, respectively). The discriminant coefficient of the vertical criterion is positive (0.42) for this function, whereas that of the criterion of sense is negative (-0.53). For discriminant function #2, only the vertical criterion is of great relevance (0.94). However, this result is qualified by the fact that the discriminating power of discriminant function #2 for the process concepts is just 17% (see Table 1).

3.5 Discriminant spaces and territories

Based on the findings presented in the previous sections, it is possible to plot the content and process concepts within discriminant spaces defined by the respective discriminant functions. Figure 2 presents these discriminant spaces. Within these discriminant spaces, we have marked out "territories" for the content and process concepts in accordance with the results of the discriminant analysis. The territories are defined by boundaries, within which the content/process concepts are plotted.

Figure 2 underlines the pattern of findings to emerge from the data presented in Table 1: Both discriminant functions #1 and #2 are required to distinguish among the content concepts.[3] In contrast, only discriminant function #1 is needed to distinguish among the process concepts,[4] with the horizontal criterion and the criterion of time having particular relevance.

3.6 Semantics of the clusters

The findings yielded by the discriminant analysis of the data make it possible to categorize both the content and process concepts according to semantic criteria.

Categorization of the content concepts

Cluster CW1(+H+V+S). The content concepts (*problem, data, computer, test, algorithm*) in this cluster are

---

[3] Because of overlaps among the content concepts.

[4] No overlaps among the process concepts with respect to discriminant function #1.

characterized by high scores on the horizontal and the vertical criterion in combination with particularly high scores on the criterion of sense. Their specific relevance for computer science education thus derives from their close relation to students' everyday life experience.
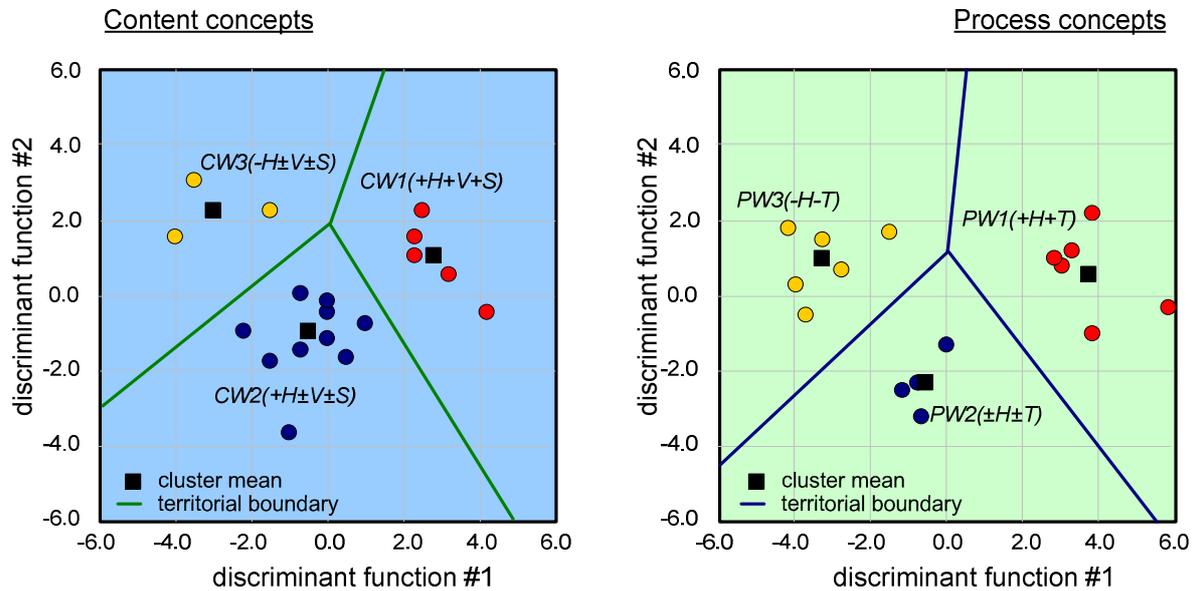

Figure 2. Discriminant Spaces for the Content and Process Concepts

Cluster CW2(+H±V±S). This cluster, which contains the content concepts *process, system, information, language, communication, software, program, computation, structure,* and *model*, is characterized by particularly high scores on the horizontal criterion and mid-range scores on the vertical criterion and the criterion of sense. It is their relevance in many different domains of computer science that makes them particularly interesting for computer science education.

Cluster CW3(-H±V±S). This cluster consists of the content concepts *network, error,* and *hardware*, which are distinguished from the other content concepts by their relatively low scores on the horizontal criterion. These concepts are thus suitable for special topics within computer science education.

Categorization of the process concepts

Cluster PW1(+H+T). The process concepts in this cluster (finding relationships, classifying, problem solving and problem posing, investigating, analyzing, generalizing) are distinguished from other process concepts by their high scores on the horizontal criterion and the criterion of time. As such, they are especially suited for instruction that covers process concepts in several domains of computer science and that focuses on concepts that will remain relevant in the longer term.

Cluster PW2(±H±T). This cluster comprises process concepts (finding cause-and-effect relationships, creating and inventing, categorizing, transferring) whose scores on the horizontal criterion and on the criterion of time are, on average, much lower than those given to the concepts in cluster PW1(+H+T). They are of interest for computer science instruction that can cover only a relatively limited number of domains and that selects process concepts likely to remain relevant at least in the mid-term.

Cluster PW3(-H-T). This cluster contains the process concepts communicating, presenting, collaborating, questioning, ordering, and comparing, which are distinguished from the other process concepts primarily by their lower scores on the horizontal criterion and the criterion of time. They are thus interesting for computer science education focusing on selected areas of the discipline. Relative to the process concepts in the other winner clusters, their relevance in terms of the criterion of time is rather low.

4. Summary and outlook

The results of the discriminant analysis show that distinguishing among content and process concepts requires different discriminant functions, whose semantics are defined by the horizontal criterion, the vertical criterion, the criterion of time, and the criterion of sense.The main finding is that the clusters of process concepts are much easier to interpret semantically than are the clusters of content concepts. Whereas two discriminant functions employing the horizontal criterion, the vertical criterion, the criterion of time, and the criterion of sense are required to interpret the three clusters of content concepts, a single discriminant function employing the horizontal criterion and the criterion of time is, in principle, sufficient to interpret the three clusters of process concepts.

The semantic categorization of individual process concepts is thus also more straightforward than is that of individual content concepts: In principle, only one discriminant factor is needed to categorize the process concepts, whereas both discriminant factors are required for the categorization of the content concepts. In future studies, the content concepts should be specified in more detail by reference to the process concepts identified as central to computer science education. Topic maps (Maicher, Sigel, & Garshol, 2007) – in which the relationships between content and process concepts are more closely defined in terms of equivalence, associative, and hierarchical relations – may inform this specification. Moreover, future research needs to address concrete questions of teaching practice. For example, experts can be consulted to determine which teaching methods and groupings of students are appropriate for teaching specific content and process concepts and in which grades of which school types these concepts play a role.

5. Acknowledgments

6. References:

Armoni, M., & Ginat, D. (2008). Reversing: A fundamental idea in computer science. *Computer Science Education, 18*(3), 213–230. http://dx.doi.org/10.1080/08993400802332670

Baumann, R. (1996). *Didaktik der Informatik* (2nd, fully revised edition). Stuttgart, Germany: Klett.

Bruner, J. S. (1960). *The process of education*. Cambridge: Harvard University Press.

Costa, A. L., & Liebmann, R. M. (1997d). Toward a renaissance curriculum. In L.A. Costa & R.M. Liebmann (Eds.), *Envisioning process as content. Toward a renaissance curriculum* (pp. 1–20). Thousand Oaks, CA: Corwin Press.

Costa, A. L., & Liebmann, R. M. (Eds.). (1997a). *Envisioning process as content. Toward a renaissance curriculum*. Thousand Oaks, CA: Corwin Press.

Costa, A. L., & Liebmann, R. M. (Eds.). (1997b). *The process-centered school. Sustaining a renaissance community.* Thousand Oaks, CA: Corwin Press.

Costa, A. L., & Liebmann, R. M. (Eds.). (1997c). *Supporting the spirit of learning. When process is content.* Thousand Oaks, CA: Corwin Press.

Denning, P. J. (2003). Great principles of computing. *Communications of the ACM, 46*(11), 15–20. http://dx.doi.org/10.1145/948383.948400

Huberty, C.J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis.* New York: Wiley. http://dx.doi.org/10.1002/047178947X

Knöß, P. (1989). Fundamentale Ideen der Informatik im Mathematikunterricht: grundsätzliche Überlegungen und Beispiele für die Primarstufe. Wiesbaden, Germany: Deutscher Universitäts-Verlag.

Maicher, L., Sigel, A., & Garshol, L. M. (2007). *Leveraging the semantics of topic maps.* Berlin, Germany: Springer. http://dx.doi.org/10.1007/978-3-540-71945-8

Modrow, E. (2003). Pragmatischer Konstruktivismus und fundamentale Ideen als Leitlinien der Curriculumsentwicklung. Dissertation. Halle (Saale): Martin Luther University of Halle-Wittenberg.

Nievergelt, J. (1980). Computer science education: An emerging consensus on basic concepts. In S. H. Lavington (Ed.), *Information processing 80* (pp. 927–933). Amsterdam, Netherlands: North Holland.

Nievergelt, J. (1990). Computer science for teachers: A quest for classics and how to present them. In D.H. Norrie & H.W. Six (Eds.), *Computer Assisted Learning, Lecture Notes in Computer Science 438* (pp. 2–15). Berlin, Germany: Springer. http://dx.doi.org/10.1007/BFb0020866

Parker, J. C., & Rubin, L. J. (1966). Process as content. Curriculum design and the application of knowledge. Chicago, IL: Rand McNally.

Schwill, A. (1994). Fundamental ideas of computer science. *EATCS Bulletin, 53*, 274–295.

Timm, N. H. (2002). *Applied multivariate analysis*. Berlin, Germany: Springer.

Wursthorn, B. (2005). Fundamental concepts of computer science in a Logo-environment. In G. Gregorczyk, A. Walat, W. Kranas, & M. Borowiecki (Eds.), *Digital tools for lifelong learning. Proceedings of the 10th European Logo Conference* (pp. 219–227). Warsaw: Centre for Informatics and Technology in Education.

Zendler, A., & Spannagel, C. (2008). Empirical foundation of central concepts for computer science education. *ACM Journal on Educational Resources in Computing, 8*(2), Article No. 6. http://dx.doi.org/10.1145/1362787.1362790

Zendler, A., McClung, O. W., & Klaudt, D. (2012). Content and process concepts relevant to computer science education: A cross-cultural study. *International Journal of Research Studies in Computing, 1*(2), 27–47. http://dx.doi.org/10.5861/ijrsc.2012.144

Zendler, A., Spannagel, C., & Klaudt, D. (2008). Process as content in computer science education: Empirical determination of central processes. *Computer Science Education, 18*(4), 231–245.

Zendler, A., Spannagel, C., & Klaudt, D. (2011). Marrying content and process in computer science education. *IEEE Transactions on Education, 54*(3), 387–397. http://dx.doi.org/10.1109/TE.2010.2062184