Computing Classification System 1998: Current Status and Future Maintenance

Report of the CCS Update Committee

Neal Coulter (Chair), James French, Ephraim Glinert, Thomas Horton, Nancy Mead, Roy Rada, Anthony Ralston, Craig Rodkin, Bernard Rous, Allen Tucker, Peter Wegner, Eric Weiss, Carol Wierzbicki

January 21, 1998

The ACM Computing Classification System (CCS), which for 20 years has served as the primary and most generally used system for the classification and indexing of the published literature of computing, has been substantially revised to reflect the greatly changed nature of computing. In this article, the revision committee reports on its work, describes the changes in the CCS, proposes processes for making annual changes, and recommends a future total revision.

Until 1995, the CCS was named the *Computing Reviews* Classification System, or CRCS. After a major overhaul of its basic structure in 1982 [1], the CCS was revised and updated in 1983 [2], in 1987 [3], and in 1991 [4]. The CCS is the basis for classifying all documents in the *ACM Guide to Computing Literature* (*Guide*), an annual bibliography that lists more than 20,000 new items, and has indexed more than 250,000 items since 1982. Various other organizations use it officially and unofficially to classify literature, reviewers, and so forth. A complete description of CCS '98 follows these discussions.

The Computing Classification System Update Committee was funded by the ACM SIG Discretionary Fund and by the ACM Publications Board. Its charter was:

- Review the structure, content, and utility of the Computing Classification System.
- Recommend changes in the CCS as required to keep the taxonomy current, while still preserving the historical value of the CCS in searching previously classified literature.
- Investigate the feasibility of having CCS supporting materials. An online dictionary of CCS terms might be especially useful.
- Investigate the utility of electronic access to the CCS and its supporting materials.
- Determine a mechanism for more streamlined and rapid CCS revisions.

This update commenced in early 1996 with the appointment of the Committee. The Committee met twice. The first meeting took place in June, 1996, in Cambridge, Massachusetts; the second meeting took place on March 1, 1997, in San Jose, California. It conducted extensive electronic discussions between the initial and final meetings.

The Committee sought wide input from the computing profession through various mechanisms, including direct mailing to or other direct contact with:

- ACM *Computing Reviews* category editors
- ACM journal editors
- IEEE-CS journal editors
- ACM SIG chairs and editors
- Chairs of ACM boards
- ACM *CR*/*Guide* professional staff
- ACM *Guide* indexers

In addition, the general professional community was invited to participate by ACM *MemberNet* and *Computing Reviews* announcements; and by a sustained announcement and update suggestion facility on the World Wide Web at: http://www.acm.org/class/update (also referenced on ACM's home page).

Table of Contents:

Table 1: Historical Overview of CCS Changes

Year	Level	Items Added	Items Retired	Items Renamed	Items Cross-Referenced
1983					
	2	2	0	0	2
	3	6	0	0	1
	4	16	0	1	NA
1987					
	2	0	0	0	0
	3	2	0	0	0
	4	16	0	0	NA
1991					
	2	1	0	1	0
	3	17	0	2	3
	4	103	4	8	NA
1998					
	2	1	0	3	0
	3	23	13	9	19
	4	246	171	32	NA

ACM professional staff provided complete CCS indexing counts for all index terms for the *Guide* years 1995, 1996, and the first half of 1997. These counts reveal how many times each CCS term was used to index documents included in *Guide* during these years. The counts helped identify candidates for retirement and ones for possible expansion (i.e., subnodes).

In addition to revising the CCS, the Update Committee considered methods for more consistent maintenance of the taxonomy and considered the advantages and disadvantages of a major redesign of the taxonomy. Later sections detail the Committee's recommendations on these various issues to the ACM Publications Board.

First, a summary of CCS changes helps put the maintenance and redesign issues in perspective.

Overview of 1998 and Previous CCS Changes

The CCS remains a four-level, hierarchical taxonomy with 11 unchanged top-level nodes. At each of the three lower levels, index terms were added, retired, or revised, with increasing frequency through levels 2, 3, and 4. "Retired" means that a term will no longer be used for future indexing, but is still a descriptor for pertinent previously indexed items. Items at levels 2 and 3 are sometimes cross-referenced to indicate close relationships. As intended in the original design of the CCS, lower-level nodes (and their associated terms) allow the tree to expand and—occasionally—contract most easily to accommodate computing's rapidly changing nature.

The numbers of nodes at the first, second, third, and fourth levels of the CCS (1991 version) were approximately 11, 40, 250, and 850, respectively. All first- and second-level nodes have General and Miscellaneous subnodes (at levels 2 and 3, respectively); these subnodes were not candidates for retirement because of their essential utility.

Table 1 summarizes CCS changes for the current and past updates.

The number of changes in this update is far greater than in any previous revision. These changes demonstrate the rapid evolution of computing, while at the same time raising concerns about how to keep the CCS current and whether its underlying structure remains adequate after nearly 20 years.

Computing Classification System (CCS) Maintenance Plan

We propose that a CCS Maintenance Committee of ACM volunteers be permanently constituted to review on an annual basis proposed changes to the CCS, and to recommend an annual update to the CCS. The recommended changes will be sent to the Publications Board. Annually, an updated CCS will appear on the ACM Web site http://www.acm.org/class, and every second year, a paper version of the updated CCS will be published.

The proposed new terms for the CCS Maintenance Committee will come from four general sources: **Indexers**; a list of **free-text words** (i.e., words derived directly from the texts themselves, as opposed to index terms selected from a prescribed taxonomy by professional indexers) and their frequency of appearance in relevant bibliographic databases; **CR editors**; and the **general public**.

Indexers will be given "update forms" to use each time they cannot find in the CCS a concept needed to index an article. After a non-CCS concept has been needed five times to index the literature, that update form will be forwarded to the *Computing Reviews* (*CR*) database manager. At the end of the year, the update forms will be provided to the CCS Maintenance Committee. The benefits of this kind of activity include systematic, unbiased, relevant suggestions on how to improve the CCS, and the institutionalization of a kind of quality control in indexing. The cost is the effort on the part of the indexers to make this part of their workflow and the central collection activity by the database manager.

ACM will obtain a **list of words from all bibliographic sources** (other than existing CCS index terms) along with their frequency of occurrence in the literature databases. This list will be provided online to the Maintenance Committee, sorted both alphabetically and by frequency of occurrence. The heuristic is that a word that appears frequently but is not an index term might be considered as a new concept for the CCS. The annual cost of obtaining this list should be no more than a few days' effort from the *CR* database manager. We might consider in subsequent years applying more sophisticated analytic methods to the raw text of the database, such as determining the frequency of co-occurring words, morphosemantic analysis, and phrase processing.

CR editors will be asked to submit requests for changes to the CCS to the *CR* Editor-in-Chief on an ongoing basis. This is a volunteer activity at no monetary cost to ACM.

The **general public** will have an open invitation to complete a form on the Web (now at http://www.acm.org/class/update) or to write directly to the *CR* database manager with any recommended changes to the CCS. While this ideally would be the most common source of valuable input, in practice relatively little input is received in this way. Collection of these volunteer submissions is managed by the *CR* database manager at minimal cost.

Retirement of terms from the CCS will be facilitated by a count of frequency of usage of CCS terms over the past three years. If any term has been used less than five times in each of the past three years to index documents, it will be automatically deemed appropriate to be retired unless the Maintenance Committee sees some reason not to retire the term.

The CCS Maintenance Committee will then take the responsibility of considering broader shifts in the structure of the CCS to accommodate various additions or retirements of terms.

The Need For a New Computing Classification System

The Computing Classification System Update Committee developed a set of recommendations for revising the current CCS. This was a significant effort, and the Committee worked hard to develop a reasonable update. However, the Committee observed that while the current system has served the community well in the past, the fundamental structure of the CCS categories, as well as the methodologies and supporting processes (including software) for maintaining the CCS, have serious problems that the Committee was not able to address under the constraints of its present charge.

Due to these constraints, we were only able to recommend changes at the second, third, and fourth levels in the taxonomy. Other changes, especially consolidation and renaming of nodes at all levels, would have been desirable because of the dramatic ways in which the structure of the discipline has changed in the last two decades. However, no significant reorganization of the node structure and content of the taxonomy was possible, even though some of this structure does not match well the literature and major subject areas of the discipline today.

For example, some of the nine major subject areas identified in the 1989 ACM report "Computing as a Discipline" [5] and further refined in such documents as the ACM/IEEE-CS report "Computing Curricula 1991" [6] and the 1997 *CRC Handbook of Computer Science and Engineering* [7] are not coherently identified in the higher levels of the CCS. That is, some of the newer major subfields of computing, such as Computational Science and Human-Computer Interaction, are not clearly represented in the CCS. Moreover, some major categories of the CCS (e.g., category E, "Data") have become increasingly irrelevant in the modern literature (witness their low indexing rate and lack of richness at the second and third levels), and ought to be redesigned or combined with other categories (e.g., perhaps as "Data and Databases") to reflect a more contemporary and enriched major subject category. There are numerous other examples of this sort, which together suggest that the structural problems with the current CCS at all four levels are widespread. It will be interesting to see if the forthcoming (1998) fourth edition of the *Encyclopedia of Computer Science* [8] modifies its classification system, which is closely related to the current CCS.

We appreciate the fact that making these sorts of changes using the present indexing and search system would raise havoc with the principle of archival integrity (historical continuity) in searching, and would also be impossible with the

current software and staffing level at ACM. When users perform a search, they want all the literature, not just that which is classified under a "revised" taxonomy. So any new system would need to incorporate the mappings that would preserve archival integrity, while remaining transparent to the average user doing a search. Such a system would also need to automate or at least support the indexers in the process of reclassifying the existing literature into a new system.

We also appreciate that any new system would have to support an ongoing process of evolution for the entire classification system: the CCS for the year 2018 will hopefully reflect reality as it is then, not as it is in 1998. From our experience with this round of revisions, we believe that evolution of the classification system over time will continue to rely heavily on expert advice from across the broad field of computing (suggesting new and retired terms, as well as structural changes, at all four levels of the taxonomy), but it should also rely on a more sophisticated set of automatic techniques (e.g., statistical content analysis of the existing recent literature) for validating all suggested changes.

Finally, we appreciate that other organizations, such as the National Library of Medicine, have developed their own methodologies, some of which are similar and some of which are quite different from those in use for the CCS, for classifying and retrieving information in the published literature of other fields, such as medical research. It would be fruitful if a revised CCS system were well informed by a study of the classification conventions and software techniques used by these other organizations [9].

We therefore recommend that a new process be initiated that will address these problems, with the aim of putting in place a greatly revised CCS that 1) better reflects the modern taxonomy of the discipline of computing; 2) better serves the needs of those who use, update, and maintain the system; and 3) is more receptive to structural change as the discipline evolves in the future, than is the current CCS. To this end, the process should include at least the following tasks:

1. Perform a more careful and conclusive analysis of the current CCS's strengths and weaknesses. (What are its goals, and how does it meet these goals?) This analysis would include an evaluation of the range of documents that are now indexed in the *ACM Guide to Computing Literature* to ensure that it is widely representative of the field of computing. Also, this analysis should be done by a team of experts representing all subfields of computing, including—in particular—representatives from the IEEE Computer Society.

2. Study other bibliographic database systems, including their indexing schemes, methodologies for keeping their taxonomies current, software support, and other tools (dictionaries) for database maintenance and searching.

3. Consider the development of an online dictionary or thesaurus for the CCS that would interact with the new taxonomy and assist users and indexers who work with the database. Such a dictionary would provide not only definitions of terms but also important information about synonymy (related terms) and semantic hierarchy (broader terms and narrower terms).

4. Design a new classification system, using a variety of contemporary sources (including the current CCS and the documents mentioned above) that might more accurately reflect the modern structure of the field and its published literature.

5. Use statistical content analysis methods to corroborate/refine this new taxonomy, using a significantly large corpus of text from the current and recently published literature.

6. Suggest a mapping function that preserves the historical integrity of the current database and CCS taxonomy in future searching and indexing activities. It should connect the terms in the new taxonomy both with the terms in the current taxonomy and with the terms in the online dictionary.

7. Estimate the cost and timeline for accomplishing a conversion of the current CCS to a new system that would have the above characteristics.

ACKNOWLEDGMENTS

We thank Peter Denning, Chair of the ACM Publications Board, and Doris Lidtke, Chair of the ACM SIG Board, for their assistance in securing funding for this update and for their continuing support throughout this project.

The Committee appreciates the many suggestions provided by the professional community. In addition, we would like to thank the following individual participants: Walter Allweil, Rajeev Pandey, Ian Gladwell, Michael Loui, Vicki Rosenzweig, Valerie Greco, Charles Kacmar, Dianne Martin, Al Lederer, Rangachar Kasturi, Joe Halpren, D. A. Griewank, Ron Fagin, Oscar Firschein, Robert France, Roy Levow, Robert Cooper, Richard Austing, B. L. Trippett, Don Herbison-Evans, Maxine Cohen, Giovanni DePoli (and his group), Jon Crowcroft, William Newman, Nina Wishbow, David Wise, and Andrew Hanson.

REFERENCES

[1] Sammet, J. and Ralston, A. The proposed new *Computing Reviews* Classification Scheme. *Commun. ACM* **24**, 7 (July 1981), 419–433.

[2] Sammet, J. Summary of changes from 1982 to 1983 version of *CR* Classification System. *Comput. Rev.* **24**, 1 (Jan. 1983), 7–8.

[3] Sammet, J. Summary of additions from 1983 to 1987 version of *CR* Classification System. *Comput. Rev.* **28**, 1 (Jan. 1987), 5–6.

[4] Coulter, N. Changes to the *CR* Classification System. *Comput. Rev.* **32**, 1 (Jan. 1991), 7–10.

[5] Denning, P.; Gries, D.; Mulder, M.; Tucker, A.; Turner, J.; and Young, P. Computing as a discipline. *Commun. ACM* **32**, 1 (Jan 1989), 2–23.

[6] Tucker, A.; Barnes, B.; Aiken, R.; Barker, K.; Bruce, K.; Cain, T.; Conry, S.; Engel, G.; Epstein, R.; Lidtke, D.; Mulder, M.; Rogers, J.; Spafford, E.; and Turner, J. Computing Curricula 1991. *Commun. ACM* **34**, 6 (June 1991), 69–84.

[7] Tucker, A. (Ed.) *The computer science and engineering handbook.* CRC Press, Boca Raton, FL, 1997.

[8] Hemmendinger, D.; Ralston, A.; and Reilly, E. (Eds.). *Encyclopedia of computer science (4th ed.).* International Thomson, Boston, to appear 1998.

[9] Rada, R.; Blum, B.; Calhoun, E.; Hafedh, M.; Helmuth, O.; and Singer, S. A vocabulary for medical informatics. *Comput. Biomed. Res.* **20** (1987), 244–263.

Authors' affiliations: Neal Coulter, Florida Atlantic University; James French, University of Virginia; Ephraim Glinert, Rensselaer Polytechnic Institute; Thomas Horton, Florida Atlantic University; Nancy Mead, Software Engineering Institute; Roy Rada, Washington State University; Anthony Ralston, SUNY-Buffalo; Craig Rodkin, ACM; Bernard Rous, ACM; Allen Tucker, Bowdoin College; Peter Wegner, Brown University; Eric Weiss, ACM; Carol Wierzbicki, ACM.

Contact Information: Neal Coulter, Department of Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL 33431-0991. Email: coultern@acm.org or neal@cse.fau.edu