A System for Classification and Control of Information in the Computer Aided Cooperative Work Place

M. Carl Drott
College of Information Studies
Drexel University

The computer aided cooperative work place (CACW) is at present an evolving entity that provides unique challenges for those who produce and study documentation. In my work as a consultant-researcher who designs, implements and studies industrial information systems, I have worked extensively with developing such systems and the systems of documentation which they of necessity must incorporate. The basic premise of CACW is the recognition that much work is the product of a number of individuals contributing their individual efforts to the completion of a project or task and that, by providing organizational coordination and appropriate tools, the effectiveness of each individual's effort can be increased. In fact, it is not inappropriate that the previous definition is broad enough to be applicable to the factory system which emerged in the nineteenth century.

This broad a definition could suggest that we are somehow re-inventing the factory system which emerged in the nineteenth century and in a sense this is true. The twentieth century culmination of the factory system, the assembly line, has become a significant organizing principal in tasks where the processing of information rather than physical objects has become the dominant task. In this sense, the move to CACW could be viewed as a reform of the "information assembly line" in the direction of task modularity in much the same spirit of experiments with autonomous work groups originating in Sweden to reform the physical production line.

The reform of the paper assembly line embodied by the CACW is characterized by the extensive use of computers to permit alternate work flows, partial completion, dynamic task reassignment, simple creation of effective work groups, and especially the timely sharing of information and evolving knowledge as the task progresses. At the same time, CACW can provide progress tracking and resource control necessary for effective management.

In the ideal computer aided cooperative work place we might imagine a integrated unified package of computer programs. Each program is equipped with "pipelines" for importing and exporting data, with "hooks" for sharing code and functions. Programs have the intelligence to track user actions and learn from them, and provide decision support that includes automatic inclusion of corporate goals and objectives. Unfortunately, the ideal of a single integrated software system for CACW is, in most cases, not yet technically or economically feasible. The present norm is the use of disjoint software, incompatible hardware and a mixture of machine and non machine tasks. The widespread use of such ad-hoc systems as CACW stations means that documentation is even more of a thorny issue than it would be if the CACW had from the beginning be realized in pure form.

After a brief overview of the role of documentation in CACW this paper will discuss CACW documentation which is automatically generated and then follow with an examination of several forms of user supplied documentation, some of which were expected in the design of the system, and some of which arose spontaneously.

The Role of Documentation in CACW

There have not yet been any comprehensive attempts to provide documentation which is specifically designed for the CACW environment. This is not surprising since in many situations the systems are experimental and subject to nearly continuous modification. Another complication in the design of documentation is that, by their nature, CACW systems are dynamically evolving and re-evolving to meet the developing needs of the work group. In this environment documentation that aims at a single task or system is still a necessity but there is an increasing need for a meta-documentation which responds to the growth of information inherent in the tasks themselves. This complexity can be seen in an example.

One of our experimental settings has been the product research and development division of a large manufacturer of electronic components. In this context every project has its own unique set of procedures and yet in each case the development process has a consistent underlying series of steps, milestones, and requirements necessary to bring any product from laboratory research to manufacturing. The development team assembled for each project is unique, yet as an audience for documentation they represent a repeating mix of skills and interests. They consist of Clerks, secretaries, and technicians; research engineers both beginning and advanced, specialists in testing, prototyping, and manufacturability; technical managers, administrative managers, and financial managers. In this setting, we have identified a number of quite different documentation needs.

Structuring and Indexing Knowledge

In our engineering research example there is a data base system which holds a library of technical information. Some of this information is in the form of titles and abstracts (and now some full text) that is purchased from an outside vendor. The necessary documentation includes the procedural steps of running conversion programs and loading the vendor data into the data base. This information system also includes internally generated research papers and technical reports. For these materials the documentation needs are more complex. There are procedural materials for the clerical preparation and formatting of reports and for their insertion into the data base. But there also needs to be documentation of the process for describing the content and structural requirements of the original report. Content and structural requirements include the informativeness of the title and abstract, consistency in the use of terms identifying measurements and test results, provision of fields which disclose and date stamp possibly patentable ideas, and indexing information which links this report to other reports. In our system design the responsibility for these aspects of the report are shared by engineers who are members of the research team, clerical support staff, and the research project manager. In addition, the computer participates in the description of the report by analyzing document content in comparison to other documents and to a concept network. Further, the completion of a report calls for updating project progress information, communication with other work groups to plan future directions, and perhaps reassignment of personnel and resources.

In addition to the documetation materials provided by the engineer-users of the CACW, part of the documentation is generated automatically by a series of programs. The computer performs its analysis through the development and updating of a "concept network." It first processes the document to identify key words and phrases. Parsing approaches and syntactic analysis were not used both because of their high implementation cost and because of their demands on machine time. The selection of terms draws upon a knowledge structure to which all previous reports in the information system have contributed. The concept network establishes weighted links between report words and and concepts. This linkage in turn links to additional concepts and to reports linked to them. The material from the outside vendor is also processed, although since most of the information is in the form of abstracts only, the procedure is somewhat different

Documenting Dynamic Information

In our test environment many of the information linkage tasks are handled by computer (although at present some subsystems exist only as simulations.) One of our first goals was the automatic linking of items in the technical information library. Previous research had shown us that to be able to produce such linkage, we needed information about the topics and contents of the articles and reports from the engineers who were actually creating and using them. On the other hand, we found that traditional indexing structures were too complex for the engineers to learn and master given the time available for this aspect of the system. In response to this we developed a classification display system that gave the users choices in a tree structured display. Two of our goals here were to have a shallow tree structure so that the number of choices which had to be made with respect to any particular concept was small and to permit the user create a new entry into the tree at any point. Thus the tree display is a representation of the concept network in which the choices of each user can be individualized and are not required to be consistent with the current concept network. This approach was intended to overcome user resistance to being forced to learn and operate within a traditional indexing structure. It also preserves the cooperative nature of the system since a set of relationships among ideas may represent a single individual, a specific project group, or the entire work group.

An important force for encouraging individual effort in the classification of reports proved to be the project supervisors. They viewed the technical data base as a communications tool. One supervisor described it as an "ultimate routing system." "It puts in what people will ask for before they ask it." Another engineer said that the selection of terms was like describing an elephant. "I don't know all the right words to put in but but the machine makes it a whole picture rather than than just a blind man who knows only part of the animal."

The original documentation for the information retrieval system was organized around the online display of the hierarchies of terms. This is the same display that formed the search and retrieval system which included basis for the description of new reports and for the selection of search terms for retrieval. The user had the choice of seeing the tree augmented by his own terms or in standard form. Searching consisted of browsing the tree and selecting one or more points to form a query. Boolean searching was also available. Contrary to some reports in the literature that users are confused by boolean searching, the engineers using the system seemed to have no problems. Anecdotal evidence suggests that users tended to browse briefly in the display and then formulate a boolean search. This may in part stem from the limited number of terms that can be displayed on any one screen. It may have seemed faster to type more terms rather than to page through a lengthy display. Overall the search documentation seemed to be little used at least in part because the system was easily learned and straightforward.

Documenting Definitions

One of the surprises to emerge from our information system for research engineers was the need to provide definitions of terms. As noted above, when reports were put into the data base they included special descriptors describing specific tests or measurements that they reported. Upon implementation this aspect drew criticism that "other engineers" were using the terms inappropriately. As we investigated the situation, we found that there was a lack of agreement as to what many terms meant.

This problem with nomenclature is not a new phenomenon to electronics. Even references to published standards such as "RS 232 interface" are widely understood by engineers to mean that only part of the specific interface standard may have been followed. In many engineering situations the universities which train the engineers are great forces for standardization, both through widely used texts and through the professional contacts of the faculty who do research in the fields. For our present case, some of the technology was so new that this standardization had not yet had a chance to emerge.

Within the research organization the electronic communication system provided an excellent vehicle for working toward standardized definition and there were soon files on the computer representing proposed definitions and commentaries on them. Interest in these definitions spread to the sales department who wanted to be sure that when they gave potential customers the technical specifications of a product there would be no confusion about what was being reported about the product. The definitions were added to the technical data base and could be accessed by a simple command.

Because the definitions, like other data base items are linked by the concept map it has been suggested that the appropriate display for them would be a form of hypertext. This would allow the reader to switch to the definition of related terms or to reports which amplified the definition. For example the definition of a measurement might be linked to a description of the laboratory setup for making that measurement. We have not yet attempted to implement such a system.

Other User Supplied Documentation

Our original design envisioned that when it came to describing the information in the technical data base it would be necessary to collect information from the project engineers. In this sense, the CACW project began with the concept that some of the documentation would be provided by the users. We discovered, however, that user supplied documentation began appearing in other situations.

The most obvious examples we have of user contributed documentation come from the electronic mail system. (Note: the examination of electronic mail messages was done with the knowledge and consent of the users. At no time did the investigators reveal the contents of any specific message without the permission of the sender.)

Almost from the first, "How to do it" messages started to appear about the operation of the system. Some of these were responses to specific questions which were themselves sent as e-mail. Other messages were anticipations that some computer processes might offer difficulty. Other messages reported in detail how a program or set of programs had been set up to perform a specific analysis. We requested that the authors of these messages send them to the technical data base so that they could be more widely retrieved. To facilitate we set up two more categories "hints" and "instructions" in the same way that "definitions" had been established earlier. One of the popular ways of retrieving "instructions" was the linkage to the report of the outcome. We characterized this type of query as "How did he do that?"

The appearance of documentation on the electronic mail gave rise to a concern for the orderly collection and timely dissemination of this information. Electronic mail is frequently treated like the paper memos which it replaces. Most messages are deleted as soon as they are received. The remainder are electronically filed by each individual. We had asked individuals to submit their instruction and tip messages to the data base but this left the capture of the information in the hands of the author. We have considered trying to automatically transfer this information to the data base. But the volume of mail large and most e-mail is ephemeral. Further, the monitoring of e-mail messages makes some users uncomfortable. On the other hand, depending on the authors will not capture all of the available information.

Special Documentation

A small number of "documentation" messages took the form of an amplification of management directives. A management message would appear and it would be followed by one or more messages, usually from more senior workers which amplified the original message. The amplification might indicate what resources were needed for a task, what precursor work had to be done before a specific task could be done. Sometimes the messages were in fact corrections to the original management message, suggesting a change in the persons assigned, or proposing an alternate schedule. Most of these documents were messages to the manager. The manager readdressed these messages as general mail. But in some instances the original sender simply addressed the message to everyone involved.

At this stage of our research the people originating these messages seem to be fulfilling a role much like the "technological gatekeepers" identified by Thomas Allen. Technological gatekeepers are people within organizations who take it upon themselves to be information sources. They attend conferences, read technical literature, and talk to others. Within the organization they are recognized as the person to asks when one has an information problem. We believe that we are seeing what we might call "managerial gatekeepers."

If we are seeing managerial gatekeepers because of the monitoring power of electronic mail we may wish to consider whether the role of explaining and amplifying managerial directives is a new form of gatekeeping made possible through the electronic work place or a long standing role made visible through the the exchange of e-mail. We did not find any managerial annoyance over these messages, and tentatively conclude that the amplification of a manager's directives is an accepted activity only now made visible.

In another case of examining e-mail we found a mail item entitled "How to use the word processor to compose e-mail." This document gave detailed steps for preparing a word processing document, spell checking it, and then using the communications software to put it on the e-mail system. This document seemed a straightforward example of user supplied documentation. We interviewed the originator, a senior secretary, and found that one of her principal reasons for writing the instructions was her dissatisfaction with the poor formatting and misspellings which appeared in the e-mail. On the one hand it seems perfectly reasonable for an individual to give suggestions for the correction of the work of others, on the other hand we are not used to the idea of documentation existing as a way of propagating one individual's personal preferences throughout a group.

It is difficult to decide whether this is an isolated incident or an expected behavior, and it is even more puzzling to determine whether the introduction of personal biases is benign or threatens the integrity of documentation. In this particular case it seems hard to argue that the writer's goal, better spelt and formatted messages, was in any way at odds with corporate goals. But it sensitized us to an important question of documentation integrity. How would we deal with a message telling how to cut corners on a complex test? If it was suggested that the parameters in a decision algorithm be set in a certain way, who was to decide if this was a policy matter or simply a technical issue of how to run the program?

A Range of Information

From the analysis above it should be clear that the CACW poses unique challenges for documentation. In particular, the issue of information management or documentation management and coordination may in fact call for a shift away from the notion of an "invisible" computer system to one which makes its presence felt in a systematic and memorable way, both to give the user a mental model of the system and the system a model of the needs and interests of the user. Working from an information point of view we consider documentation on a conceptual basis rather a procedural one, From this conceptual standpoint the following is a partial list of what needed to be known:

• Training information including training by procedure and by example.

• Tracking information which creates a growing history of the project and a linking of the information produced.

• Organizational and sequencing information for tracking progress and controlling resources.

• Communications information, including the identification of changes in the overall knowledge of the problem which require re-examination of previous decisions, plans or findings.

• Structuring and indexing information for the growing database of technical knowledge.

• Definitional information for providing consistency of nomenclature for the project staff.

This list includes information aspects that are traditionally clearly associated with the documentation, while other aspects such as scheduling and indexing are frequently not included in documentation packages. Our finding is that in a complex system these aspects require documentation.

A Documentation Problem

In a section above we listed the types of information that had to be tracked by the CACW system and noted that some of them were matters of project scheduling and management which are not normally considered a part of documentation. The integrated nature of the CACW system makes it less obvious where the line between documentation and management information should be drawn. This is complicated in the research and development environment because the way in which the computer system is being used is changing and shifting.

One of the problems was that there was no one who was charged with the responsibility for sorting out and tagging which of the various messages, files, and data base entries were which kind of information. Some of this tagging was done by the authors. The artificial intelligence and decision support subsystems added to the organization of information but their main function was control and tracking of the processes rather than the information.

When the original system was designed, management was very clear that they did not want to hire a staff person to run the system. The normal actions of the managers should be enough to keep the information organized. This seemed a sensible objective. After all, if the computer provides more of the coordinating function in the work place isn't it counterproductive to have a system that needs another human coordinator?

The problem here is not unlike the situation in which an organization has an ongoing need for updating and correcting documentation but there is not sufficient work to justify a full time documentation specialist. At present one of the technical supervisors has "adopted" the system and keeps the entries in the data base weeded. It is clear however that as cooperative systems grow, there will have to be new formal roles defined to see to the orderly maintenance of documentation.

Making the System Visible

The freedom to proceed in parallel proliferates individual versions of shared files. In many simple computer systems it is reasonable to keep many kinds of system information invisible to the user. This information includes details of what software version created a file and dates of creation and modification. In addition there is ephemeral information that even the system does not keep including the dates on which copies were made, the names of the input files when the present file was created and others. However, in the multiple file environment of the CACW system it is mandatory that this information not only be retained but that it be displayed for the user.

Questions arise about whether a person has seen a particular revision, how different two (or many) versions of a file are, whether information from a particular data set has been merged into a resulting table. CACW systems must have greater file tracking capabilities than previous systems. Along with the requirement that file history documentation be constantly available is the need for a powerful query and comparison package to support this documentation.

Some experimental CACW systems use artificial intelligence methods for version control and to alert workers of mismatches. But these do not reduce the users' need for information, quite the contrary, they increase it. In manual systems it may be possible to impose some sort of file discipline on the work group but this is not possible in cooperative systems. The user must not only have access to this information but must be prompted to use it. The question what files do I have is replaced by what files do I have access to and which ones should I use.

The fact that decisions within the system may be wholly of partially under machine control creates another documentation problem. Jens Rasmussen has noted that users of computerized process control systems are frequently confused because the system operates on a logic built by the programmers rather than on a logic dictated by the specifics of the task. He notes that in cases where users do not understand the basis for machine actions or decisions they are unable to participate in the operation of the system. It is important in CACW that system actions be self explaining.

Some of this explanation resides in providing each user with a conceptual overview of what stage a system is at and what immediate goals it is programmed to achieve. It is also necessary to be able to communicate the specific steps that the program used in achieving the result at hand. Fortunately, decision support systems are frequently programmed so that they can recapitulate the evidence used and the steps taking in reaching a decision.

Thus the user needs to be kept in synchronization with the system in two ways: First he needs to have a mental model of what process the system is currently involved in and why the process is occurring. Second he must have a way of relating his observations of the state of the system to the actions taken by the system. Providing this information requires the selection of the appropriate conceptual description from the data base and augmenting it with a specially generated explanation of how the present status was reached.

Context Sensitive Conceptual Document

The idea of context sensitive documentation is not new. In computer systems it is fairly straightforward to provide a help command that directly retrieves and displays information on the last command used or attempted. This type of context sensitivity is procedural, trying to anticipate the next step which the user wants to perform. The greater complexity of the CACW environment calls for a more sophisticated approach, the availability of conceptual help.

Providing conceptual help in an appropriate context is still in early experimental stages. We need not only to understand how to recognize what conceptual help is appropriate but how to present that information. This problem is related to an information retrieval called selective dissemination of information (SDI.) In some SDI systems the documents that a user has recently retrieved are analyzed to construct a search request for similar material. This approach might give access to a more general conceptual explanation based on a continuing observation of a user's actions. In all attempts to provide conceptual explanations we need to be concerned that a conceptual explanation seems more like a human response that does a procedural one and so the wrong conceptual response is more jarring.

As a preliminary approach we propose a system that both tracks user actions and queries the user about his actions on a regular basis. Why are you opening this file? How does this fit in with what you did before? The questions must be infrequent and easily answered. This information can be combined with information in the database which ties a file into the concept network. Programs are also represented in the data base and have links to the concept network, although at present all of these links are built by human intervention.

User responses to the questions themselves form a weighted network linked to the concept network so that a set of responses together with the specific documents or programs being used represent a query. At present these are matched only to specifically designated concept statements and not to the entire data base. It is however for the user to examine the portions of the tree display which the query has activated and to construct a search of the data base from that.

At present the technological feasibility of the process is under review and the development of appropriate questions to the user is being studied. Simulations of providing context sensitive conceptual descriptions have been promising.

Summary and Conclusions

The computer aided cooperative work place multiplies the quantity and kinds of information which must be controlled. The information needed includes technical knowledge, vocabulary coordination, organization and planning, project history and tracking, individual and group communications, and documentation. The variety of information needed requires many forms of documentation; documentation of the processes, documentation of the system, and documentation of the conceptual framework of the overall task.

To control this information the system must provide its own organization of the information. All information is linked by a concept network. This approach integrates the many forms of information while allowing each individual piece of information to have special links to documents of its own type.

The capture and control of this information is shared by the computer system and the users. A tree structured display of key terms is available to assist users in entering information items into the data base or retrieving the. This tree can reflect either the conceptual network or special vocabularies unique to the user.

From a documentation viewpoint it is important that the system collect and organize both user generated documentation and system generated documentation. The latter, information necessary to understand both the current state of the system and its files and why the system is in that state, becomes especially important in CACW systems because of file proliferation and the capacity of the system for independent action.

The expansion of traditional notions of documentation and the increased dependence on both the users and the system for part of that documentation raise important problems of documentation completeness, quality , and appropriateness. These issues will require much attention in the future. However the methods of documentation control used in this experiment are promising. In this model the documentation is integrated into the information resources of the system. Formally designated documentation be retrieved. More important, other types of documents which serve to inform and direct users are available.

This unified information model of a CACW system involves both users and the system in the collection and retrieval of information resources. It reflects more closely the actual working of groups of people, while providing an integration unavailable in more traditional systems.