Methodological Issues in Studying and Supporting Awareness on the World Wide Web

Lee Li-Jen Chen and Brian R. Gaines
Knowledge Science Institute
University of Calgary
Calgary, Alberta, Canada T2N-1N4
{lchen, gaines}@cpsc.ucalgary.ca
http://ksi.cpsc.ucalgary.ca/KSI/

Abstract: The World Wide Web has grown very rapidly to become a major resource supporting collaborative activities in a wide range of groups, disciplines and communities. However, the growth of the web creates problems of information overload and of maintaining awareness of activities at other sites relevant to one's own tasks. This article develops a methodological framework for studying and supporting awareness on the web; describes Chrono, a tool for supporting awareness of changes at another site; and uses the framework to classify Chrono and related tools, and to clarify the human factors design issues involved.

INTRODUCTION

The exponential growth of the World Wide Web (WWW) and the growing availability of collaborative tools and services on the Internet have facilitated innovated knowledge creation/dissemination infrastructures, such as: electronic libraries, digital journals, resource discovery environments, distributed co-authoring systems and virtual scientific communities [Schatz91]. Collectively, these Internet infrastructures have become integral parts of an emerging information system supporting collaboration in distributed scientific/research communities.

World Wide Web was originally conceived and developed at CERN for the purpose of assisting and facilitating collaborative interactions among high energy physicists, working at various institutions in different countries, to conduct joint research projects [Berners-Lee94]. Since 1993, it has diffused at a phenomenal rate and gradually has subsumed various popular Internet services such as: USENET Newsgroup, Electronic Mail, etc. The original charter of the web is summarized in the following quote: *The World Wide Web was developed to be a pool of human knowledge, which would allow collaborators in remote sites to share their ideas and all aspects of a common project* [Berners-Lee94].

However, the growth of the web, while creating a rich new resource, also creates problems of information overload. The management of the diffuse communities collaborating through the web raises human factors issues going beyond those of the coordination of smaller, goal-directed groups with well-defined roles and tasks. For example, what are the responsibilities of information providers in supporting users of whom they are unaware, and who may be using the information in very different ways from those originally envisioned? The web supports the collaborative activities of small work groups, but it also supports those of well-defined scholarly subdisciplines, and those of the much less defined community at large. To study and support collaborative activities on the web, we need a methodological framework that identifies the major distinctions between 'work groups' of widely differing sizes and structures, and between the various roles that originators, retrievers, and intermediaries can play.

This article identifies one of the major problems of collaboration on the web as that of maintaining *awareness* between remote research partners when changes occur in one location that affect activities in another. Such *chronological awareness* is an important issue for supporting task-oriented collaborative projects of research

groups or organizations. At the other end of spectrum, the issue of locating where specific information resource is on the web, i.e., *resource awareness* has become important for supporting the research community at large. In the subsequent sections, awareness issues in collaborative group are discussed, an awareness maintenance framework is developed which identifies the various dimensions for evaluating and designing awareness maintenance systems. CHRONO, a tool developed to support awareness of change, is described, and the awareness maintenance framework is used to analyze and classify this and other chronological awareness systems on the web.

COGNITIVE ARTIFACTS AND SITUATIONAL AWARENESS

The use of social interaction and cognitive artifacts as means to enhance human abilities has been analyzed by Norman [Norman91]. A cognitive artifact is defined as *"an artificial device designed to maintain, display, or operate upon information in order to serve a representational function."* The power of a cognitive artifact comes from its function as a representational device changing the nature of a task to match human abilities and hence to give users *greater capabilities based on those abilities*. When the informational and processing structures of the artifact are combined with those of the users, the result is to expand and enhance the cognitive capabilities of the total system of human, artifact, and task.

CSCW (computer-supported collaborative work) research focuses on assisting people to work collaboratively as a cohesive team and providing them with a sense of common purposes (e.g., completion of the group task). For example, Landow's *In Memoriam* project [Landow90] utilizes hypertext's freedom of navigation and linking ability to alleviate the effects of physical separation and the univocal voice of textual conversation. In so doing it creates a new awareness of the processes of collaborative learning and collaborative work for group members in literary studies. Olson and Atkins' NSF EXPRES Project [Olson90]uses intelligent, multimedia email to facilitate cooperation within scientific and engineering community by increasing researchers' awareness about each other's work.

Norman notes that a critical requirement in shared tasks is maintaining *situational awareness* [Norman91] by keeping everyone adequately informed. In an environment where each member has a well defined role, the need to have face-to-face communication in order to perform a cooperative task becomes less necessary if mechanisms for situational awareness have been well established between members. For example, Hutchins notes that the navigation of a large ship requires effective coordination of various people with differing roles [Hutchins90]. Many key members of the navigation team are geographically separated and communicate with each other by a common telephone circuit. Those arrangements provide opportunities for the navigation crews to observe each other's work, contributing to partial redundancy in their joint knowledge. They also support maintenance of the group over time to provide fault-tolerance if some group members fail to perform their roles.

Thus, one of the important criteria for achieving group cohesiveness is the situational awareness of what other group members are doing. Together the functional specificity of the crews and the cognitive artifacts that facilitate situational awareness (e.g., the single telephone circuit, the pilot house with high visibility among navigational team members) can create an effective collaborative system. As already noted, for the web these awareness issues are made more complex by the diffuseness and lack of overall goal and task definitions for many of the 'work groups' involved.

FRAMEWORK FOR HUMAN FACTORS OF AWARENESS MAINTENANCE

This section examines various human factor issues relating to awareness of changes in a web-based working environment. The main focus here is on *chronological awareness* as a special case of situational awareness, that is, the awareness of *when* something (an event or an artifact) has changed.

Web pages, FTP archives, Listservers, and other common Internet infrastructures are the primary means for information dissemination on the web, and these infrastructures are being constantly updated to reflect members' current states of knowledge on their portions of collaborating tasks. In a dynamic environment where large amounts of information are created and updated frequently, the need to keep up with the most up-to-date and relevant information has become more important as the Internet community expands.

DIMENSIONS OF AWARENESS MAINTENANCE ARTIFACT

There are four main dimensions of design considerations for awareness maintenance artifacts for web users:

1. Locus of Responsibility: *Server-Side, Client-Side*, or *Centralized Dispatcher*
2. Level of 'Work Group' Hierarchy: *Group, Organization*, or Community.
3. Method of Locating Changes: *Browsing* vs. *Targeting*
4. Complexity of User Interaction: *Simplicity* vs. *Customization*

The first dimension, the *locus of responsibility*, differentiates who is responsible for maintaining the record-keeping mechanisms for supporting awareness maintenance.

- A *Server-side* approach ensures that only the users who are current visiting the web site would need to know what information has been changed. Hence, it reduces network traffic by avoiding needless broadcasting of chronological information to some users who might not be concerned with it. The main disadvantage of this strategy is that in order to know whether or not any particular page has been changed, a user would need to check out the specific web site periodically.
- A *Client-side* approach periodically monitors specific pages at various web sites and report whether or not they have been changed recently to ensure the user would be aware of any changes. The main disadvantage is that the users need to remember to run such a system, or it must be set up to run periodically, in turn causing it to consume higher network bandwidth.
- A *Centralized-dispatcher* approach put the monitoring responsibility at some specific central registry locations that automatically monitor the registered pages for the users. Its main disadvantage is the high network traffic involved in such a centralized broadcasting scheme.

The first dimension can also be considered from a information resource provider/user perspective. Various information systems, such as the web, use the client-server model to partition the computational division of labor. Similarly the *locus of responsibility* of awareness maintenance at every level can be divided into *originators* (i.e., providers) of information, *retrievers* of information resource, and *intermediaries* of information retrieval exchange. Therefore:

- An *originator* is a source/server of information dissemination.
- A *retriever* is an user/client of information resource.
- An *intermediary* is a meta-information resource mediating between originators and retrievers.

The second dimension, the *level of 'work group' hierarchy*, reflects the need for maintaining mutual awareness among members exist in various collaborative arrangements. There are three main levels of awareness arrangements which constitute the awareness maintenance hierarchy.

- The *Group* level usually consists of closely coordinated members in a relatively small or medium size project. There is usually a need to be aware of short-term changes in the data being managed by different group members as part of their task activities, and the nature of this need is relatively well-defined.
- The *Organization* level frequently involves loosely coordination of different group projects within an organizational structure such as a scholarly sub-discipline. There is usually a need to be aware of significant changes in the knowledge available in the organization, and satisfying this need involves mechanisms ranging from well-defined channels, such as electronic journals or reprint archives, to organizational/cultural/ethical norms on how information is to be disseminated.
- The *Community* level often involves mechanisms for providing its members at large some form of *resource awareness*, for example in locating where specific information resource is on the web.

Viewed from Miller's *living systems theory* [Miller78] perspective, *situational awareness* is essential at every level above the individual level in the system hierarchy and that originators and retrievers of information are situated at the opposite ends of the *channel* and *net* information subsystems [Chen96]. The various awareness maintenance mechanisms (such as the chronological awareness support systems for groups) serve as (or provide the functionalities of) the *timer*, *associator*, and *memory* subsystems in various system levels.

The third dimension, the *method of locating changes*, involves two different ways of locating documents that have been changed: browsing and targeting.

- A *Browsing* approach facilitates the chronological browsing characteristic: visitors of a site may find relevant information via browsing the concurrently created/modified web pages, because closely related documents are sometimes created (or modified) around the same time. This browsing approach allows the *at glance* attribute for accidental discovery of relevant information without prior awareness of their existence.
- A *Targeting* approach focuses on specific pages or information that users have previously specified. Therefore this method of locating changes is more direct and efficient. However the main disadvantage of such approach is that the users cannot be made aware of any new information which have been created recently. They are limited only to changes made to prior knowledge.

Finally, the fourth dimension, the *complexity of user interaction*, denotes system usability in terms of simplicity vs. customization.

- A *Simplicity* approach tries to make its user interface simple and familiar to web users. The goal is gearing toward ease of use and a shallow learning curve. On the web this is typified by systems that present awareness information through the generation of familiar web documents.
- A *Customization* approach, in contrast, tries to allow elaborate customization of features, but it also demands more efforts by its users to learn and utilize its functionalities. On the web this is typified by systems that use separate tools with their own user interfaces to support awareness.

CHRONOLOGICAL AWARENESS SUPPORT SYSTEMS

This section presents a concrete example of a class of awareness support systems specially designed for supporting *chronological awareness* maintenance on the web. One such system is described in detail, followed by brief overviews of three other systems.

By definition, a *chronological awareness support system* for the web provides each individual an appropriate awareness of relevant activities of other individuals. Hence it allows team members to synchronize their activities in a more coherent way by keeping them informed and aware of any changes made to each other's web pages or other information resource that might be relevant to their current tasks. The information systems that support such awareness correspond closely to the timer critical subsystems in a living system [Miller78, Chen96].

THE CHRONO CHRONOLOGICAL AWARENESS SYSTEM

CHRONO [1] is an HTTPD server-side system which generates chronological listings of web pages that have been changed recently at specific sites. It provides a basic awareness-support that lets visitors of a web site (e.g., members of a group, or other Netsurfers) see which web pages have been modified since their last visit. Currently, the CHRONO system is implemented on a UNIX platform. CHRONO presents visitors with an HTML document that lists the titles of web pages at the site in reverse chronological order. This chronological listing also functions as a collection of hyperlinks to the listed web pages.

CHRONO USER INTERFACE AND ITS FUNCTIONALITY

The user interface of the CHRONO system is straightforward and intuitive for web users. It looks like an automated *what's new* page to the users. From the list, the visitors are able to tell at a glance what documents have been modified or created recently. They can also scroll down the list to check those older documents at the site. Because the titles of the listed pages also act as hyperlinks to the actual web pages, visitors can simply click and jump to the relevant pages of interests.

Chronological hyperlinks presented in the CHRONO listings provide the visitors the means to access the newly modified or created pages. This *time-line* (or history) dimensionality complements the functionality of the associative memory characteristic found in typical hyperlinks which join related information.

The *time-line* dimension allows frequent visitors of a web site an immediate awareness on what have been changed since their latest visit. The changes may reflect some web pages in which they have been previously interested or they may show some pages that the visitors have never seen before but now appeal to them. Hence

[1] CHRONO and related systems can be accessed via: http://ksi.cpsc.ucalgary.ca:8008/cgi-bin/release?1wn

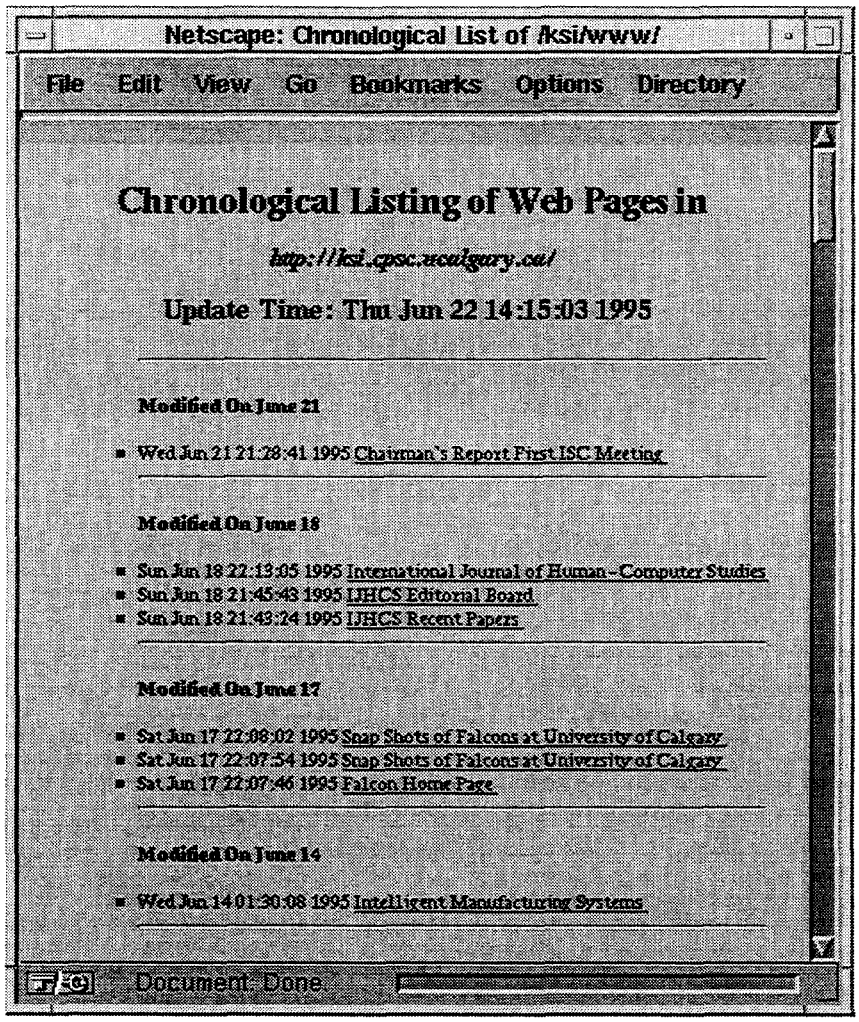

Figure 1 Screen Snapshot of CHRONO

this *chronological* browsing characteristic is analog to *spatial* (subject-category) browsing characteristic that library patrons have often experienced when looking for books on open book-shelves (i.e., accidentally finding other relevant books near the books that they were looking for originally). Unlike a manually updated *What's New* page in which the users have to rely on timely updates made by a webmaster (or by the document authors), CHRONO provides the time-line dimension to the users automatically, in a reliable and periodic fashion.

CHRONO CURRENT SYSTEM USAGE

The CHRONO system has been running at the Department of Computer Science at the University of Calgary since March, 1995. It currently services seven web locations: two research units and five individuals. People associated with the two research units: Knowledge Science Institute and GroupLab periodically have utilized the system to check on new developments of each other (both within group and between groups).

From a preliminary examination of the HTTPD access_log of the CHRONO system site and from talking with individual group members, we have found that the chronological listings of the five personal sites have offered other group members more focused chronological awareness about these five individuals' working patterns. Occasionally, some people had discovered new projects the targeted individuals were working on that they were not previously aware. This positive usage experience of personal chronological listings suggests to us that there is a

need to further examine the effectiveness of providing different sub-groupings of chronological listings to group members.

RELATED CHRONOLOGICAL AWARENESS SYSTEMS

There are three other systems that provide support for chronological awareness based on different designs and implementations. They are briefly examined in this subsection, then followed by a comparative evaluation.

WebWatch is a client-side chronological awareness system for keeping track of changes in selected web documents. Given an HTML document referencing URLs on the web, it produces a filtered list, containing only those URLs that have been modified since a given time. The criteria used for filtering can be given as a global setting that applies to all URLs, or can be derived automatically, using the time of user's last visit to the document, as recorded by the web browser in the user's local HTML (e.g., bookmark) file. In contrast with the simple timeline listing strategy used in CHRONO, WebWatch stores its arguments in a parameter file. Once the users have customized the program to their needs, using its graphical front-end, they can have it run periodically in unattended mode.

Katipo is another client-side chronological awareness system built for Macintosh that shares many similar concepts with WebWatch. It reads through the Global History file maintained by some web browsers checking for documents that have changed since the last time a user viewed them. The basic difference between it and WebWatch is that it uses the Global History file as its reference for checking URLs, whereas WebWatch uses the Bookmark file.

URL-minder is a centralized system that keeps track of resources on the Net and sends registered users e-mail whenever their personally registered resources change. Users can have URL-minder keeps track of any web resource accessible via HTTP. It keeps track of one web page, image file, or other Internet resource at a time. It tracks the actual HTML markup, binary contents, or ASCII contents of the URL users have submitted. If an HTML page includes a GIF or JPEG graphic, the URL-minder will inform them via E-mails when the reference to the graphic changes. The URL-minder currently checks on users' registered URL's at least once per week, and will inform users if it fails to retrieve their registered URL after trying twice.

SYSTEMS EVALUATION WITHIN THE METHODOLOGICAL FRAMEWORK

This section presents a comparative evaluation of the four chronological awareness support systems discussed earlier: CHRONO, WebWatch, Katipo, and URL-minder. Each system has its unique approaches for achieving chronological awareness for web users and complement each other along four main dimensions.

The first dimension, the *locus of responsibility*, differentiates who is responsible for maintaining the record-keeping mechanisms for chronological awareness. For example, CHRONO is a server-side system in which chronological listings are being updated and kept at the web server-side. Thus, CHRONO can be thought of as offering "*chronological awareness on demand*". WebWatch and Katipo, however, put the responsibility of maintaining chronological awareness on the client-side. Both client side systems periodically monitor specific pages at various web sites and report whether or not they have been changed recently. Finally, URL-minder requires its users to register at a centralized site.

The second dimension, the *level of 'work group' hierarchy,* signifies that all of the current chronological awareness support systems are mainly focused at the group level (and to some extend at the organizational level). Collectively, they are supporting chronological awareness of information resource typified by closely-coupled collaboration at the group level.

The third dimension, the *method of locating changes*, involves two different ways of locating documents that have been changed: browsing and targeting. CHRONO uses the browsing approach in order to facilitate the chronological browsing characteristic. Conversely, WebWatch, Katipo, and URL-minder employ a targeting approach in which they are targeted on specific pages or information that users have previously specified.

Finally, the fourth dimension, the *complexity of user interaction*, denotes system usability in terms of simplicity vs. customization. CHRONO and URL-minder are in the simplicity category; their user interfaces are simple and familiar to web users (i.e., scrolling list of hyperlinks and fill-in form of URLs and e-mail address). They are

geared toward ease of use and a shallow learning curve. Both systems, however, have no capability for individual customization. In contrast WebWatch and Katipo allow elaborate customization of features, but they also demand more efforts by the users to learn and utilize their functionalities.

Therefore, each chronological awareness support system examined so far have various degrees of advantages and disadvantages along the four dimensions. CHRONO has the advantages of: (i) simplicity of user interface; (ii) supporting accidental discovery via its browsing characteristic; and (iii) server-side chronological awareness information on demand. It is nicely complimented by WebWatch and Katipo for their strength in the efficiency of targeting approach and customization capabilities. And finally URL-minder offers another unique service: it uses e-mail as its notification channel. This approach is useful for users who use their e-mail systems more frequently than web browsers. Together as a whole, these chronological awareness support systems have covered a wide range of approaches in respect to four major dimensions of chronological awareness support.

CLASSIFICATION OF EXISTING WEB SERVICES

The methodological framework developed in this article can be used to study a wide range of general awareness systems for the web. There are various awareness maintenance artifacts on the Internet that address different system levels. The following analysis examines and categorizes them along the two major dimensions: *level of 'work group' hierarchy and locus of responsibility*. Currently, *intermediary* mechanisms are still in a nascent stage of evolution (e.g. Universal Resource Agents), hence the present taxonomy does not yet reflect the intermediary locus of responsibility (Table 1).

At the *group* level, originators of the information resource can organize and implement work flow models of the group activities, use server-side chronological awareness support systems such as CHRONO, or/and send e-mail notification to users of information. Retrievers of the group level information resource can use client-side chronological awareness tools such as WebWatch and Katipo or register in centralized dispatcher service like URL-minder. Alternatively, they can send e-mail to inquire to information originators to see if any new things have come up.

At the *organization* level, originators of the organizational resource can broadcast to concerned individuals, groups, or organizations via specific Listservers, register in NCSA's *What's New* service, or announce in organization-maintained MOO or MUD. They can also establish what's new HTML links in organization news while retrievers can participate in HyperNews or MOO, and follow the new HTML links in organizational web pages.

At the *community* level, originators can register the information resource in hierarchical subject services like Yahoo, or initialize their pages in searching and navigational services like Web Crawler while retrievers can browse Yahoo, search LYCOS or Web Crawler, or read and post to the USENET newsgroups.

One interesting observation can be made about the level of awareness in relation to the level of coordination. As the level of awareness moves from the group level to the community level, the need for closely-coupled coordination decreases among members. What happens in practice is the awareness maintenance becomes asymmetrical, rather than mutual, at the higher system level. However major awareness requirement continues. Resource providers may not need to be aware of who their users are, but the users' activities may be critically dependent on the status of the resources.

Table 1 Taxonomy of Mechanisms Supporting Awareness Maintenance

Locus of Responsibility	**Level of 'Work Group' Hierarchy**		
	Group	**Organization**	**Community**
Originator	Work flow Model **CHRONO** E-mail to Notify	Broadcast to Listserver Register in **What's New** Announce in MOO Establish HTML Links	Register in **Yahoo** Initialize **Web Crawler**
Retriever	**WebWatch/Katipo** **URL-minder** E-mail to Inquire	Participate in **HyperNews** Participate in MOO Follow HTML links	Browse **Yahoo** **LYCOS** Search Read/Post to USENET **Web Crawler** Search

DISCUSSION AND CONCLUSION

The discussions in the previous sections have utilized the concept of *system levels* in *living systems theory* [Miller78, Chen96] and described an awareness maintenance framework for the web that focuses on the dynamic interactions between individuals, groups, organizations and communities. Within this framework, individuals constitute the fundamental level of analysis in the overall system. Self awareness provides a sense of identity, purpose, and consciousness to each individual. When perceiving all levels collectively, the notion of self awareness can be extended to different hierarchical partitions in the system for the purpose of analyzing the interrelationship between the overall system and its parts in the *collective stance* [Gaines94].

Within in this collective stance perspective, individuals, groups, organizations, and communities together create a multi-leveled *collective intelligence* [Smith94] that becomes a pool of human knowledge for individual collaborators in remote sites to share their ideas and all aspects of collaboration. In order for the collective intelligence to function in a cohesive and integrative manner, coordination at various system levels must be maintained. However the means for maintaining a system-wide cohesion and integrity for the web are still emerging and evolving.

One practical issue inspired from the locus of responsibility for awareness maintenance at the community level is that of awareness of individual member's profile of specialties as a information resource. For example in terms of scientific communities, one common question among many scientists/scholars is: who and how should one contact when specific research questions, needs, or opportunities have arisen? The challenge is how can we create an environment inductive for informal social networking among community members on the web? What is the proper balance among originators, retrievers, and intermediary mechanisms for such an information resource?

Also in term of chronological awareness, there is a need to examine further the effectiveness of providing different sub-groupings of chronological listings to group members, such as chronological listings for particular subdirectories of a web site, for particular projects of a group, or for chronological access patterns to particular documents which may prove to be useful on different occasions and for varying user needs. In addition, the issues of how to establish mutual awareness among individuals, groups, organizations need to be explored further.

REFERENCES

[Berners-Lee94] Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. F. & Secret, A. (1994). The World-Wide Web. Communications *of the ACM*, August, Vol. 37, No. 8, pp. 76-83.

[Chen96] Chen, L. L.-J. & Gaines, B. R. (1996). A CyberOrganism Model for Awareness in Collaborative Communities on the Internet. *International Journal of Intelligent Systems*, to appear.

[Gaines94] Gaines, B. R. (1994). The Collective Stance in Modeling Expertise in Individuals and Organizations. *International Journal of Expert Systems.* 7(1), pp. 22-51.

[Galegher90] Galegher, J., Kraut, R. E. & Egido, C., (ed.) (1990). *Intellectual Teamwork: Social and Technological Foundations of Cooperative Work.* Erlbaum, Hillsdale, New Jersey.

[Hutchins90] Hutchins, E. (1990) The Technology of Team Navigation. In [Galegher90], pp. 407-428.

[Landow90] Landow, G. P. (1990) Hypertext and Collaborative Work: the example of Intermedia. In [Galegher90], pp. 407-428.

[Miller78] Miller, J. G. (1978) *Living Systems.* McGraw Hill, New York, NY.

[Norman91] Norman, D. A. (1991) Cognitive Artifacts. Carroll, J. M., (ed.) *Designing Interaction: Psychology at the Human-Computer Interface.* pp. 17-38. Cambridge University Press, Cambridge, UK.

[Olson90] Olson, G. M. & Atkins, D. E. (1990) Supporting Collaboration With Advanced Multimedia Electronic Mail: the NSF EXPRES project. In [Galegher90], 429-451.

[Schatz91] Schatz, B. R. (1991). Building an Electronic Scientific Community. *Proceedings of 24th HICSS-- 1991,* Vol. 3.

[Smith94] Smith, J. B. (1994). *Collective Intelligence in Computer-Based* Collaboration. Lawrence Erlbaum Associates, Hillsdale, NJ.