Knowledge Acquisition Processes in Internet Communities

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

Abstract

With the growth of usage of List Servers and the World Wide Web the Internet has become a major resource for the acquisition of knowledge, and it has given new prominence to human discourse as a continuing source of knowledge. The society of distributed intelligent agents that is the Internet community at large provides an 'expert system' with a scope and scale well beyond that yet conceivable with computer-based systems alone. It is important to model and support the processes by which knowledge is acquired through the net. In developing new support tools is one asks "what is the starting point for the person seeking information, the existing information that is the basis for their search." A support tool is then one that takes that existing information and uses it to present further information that is likely to be relevant. Such information may include relevant concepts, text, existing documents, people, sites, list servers, news groups, and so on. The support system may provide links to further examples of all of these based on content, categorization or linguistic or logical inference. The outcome of the search may be access to a document but it may also be email to a person, a list or a news group. This articles develops a model of services and knowledge processes on the Internet, describes various forms of support tool, and categorizes them in terms of the model.

1. INTRODUCTION

The growth of the Internet has provided major new channels for the dissemination of knowledge. Increasing international connectivity has made the net accessible to special-interest communities world wide, and electronic mail and list servers now provide a major communications medium supporting discourse in these communities. Until recent years, limitations on the presentation quality of on-line file formats restricted the publication capabilities of the net to rapid dissemination of files printable in paper form. However, advances in on-line presentation capabilities now allow high-quality typographic documents with embedded figures and hyper-links to be created, distributed and read on-line. Moreover, it has become possible to issue *active documents* containing animation, simulations, and supporting user interaction with computer services through the document interface. The major part of this functionality has become accessible through the protocols of the World Wide Web, and the web itself is seen as a precursor to an *information highway* subsuming all existing communications media.

The development of the net has been very rapid with little central planning, and, despite its widespread use, there is little information as yet on the social dynamics of net technologies. Many systems have been developed cope with the information overload generated by direct access to the net. The wide variety of indexing and search tools now available have in common that they support

selective attention and awareness in the communities using the net. It would be useful to be able to analyze the design issues and principles involved in these tools in terms of the knowledge and discourse processes in the communities using these tools.

This article provides a model of the Internet in terms of discourse and awareness and uses it to classify the types of support tools existing and required.

2. COMPUTER-MEDIATED COMMUNICATION (CMC)

It is tempting to consider the Internet as a new publication medium in which electronic documents emulate paper ones, and where the basic human factors issues are those of indexing and information retrieval. This makes the vast existing literature on information retrieval, its techniques and human factors, relevant to the net. However, this addresses only one aspect of computer-mediated communication, neglecting its function of supporting discourse within communities. Much of the information retrieved from the net is generated as needed through discourse on list servers—the Internet is a mixed community of publications and intelligent human agents that both stores knowledge and generates it on demand. When the information needed cannot be found through retrieval then it may be requested through discourse, a phenomenon prophesied in the early days of timeshared computing:

> *"No company offering time-shared computer services has yet taken advantage of the communion possible between all users of the machine...If fifty percent of the world's population are connected through terminals, then questions from one location may be answered not by access to an internal data-base but by routing them to users elsewhere—who better to answer a question on abstruse Chinese history than an abstruse Chinese historian."* (Gaines, 1971)

The society of distributed intelligent agents that is the Internet community at large provides an 'expert system' with a scope and scale well beyond that yet conceivable with computer-based systems alone. Computer-based discovery, indexing and retrieval systems have a major role to play in that community, but are only one aspect of Internet information systems.

Krol (1993) captures the essence of these consideration in Internet RFC1462 which replies to the question "What is the Internet" with three definitions:

1 a network of networks based on the TCP/IP protocols,
2 a community of people who use and develop those networks,
3 a collection of resources that can be reached from those networks.

These are complementary perspectives on the net in terms of its technological infrastructure, its communities of users, and their access to resources, respectively. Models of computer-mediated communication must taken into account all three perspectives: how agents interface to the network; how discourse occurs within communities; and how resources are discovered and accessed.

3. DIMENSIONS OF THE COMPUTER-MEDIATED COMMUNICATION

In examining the utility of the net and web it is useful to classify all the major services in terms of the significant distinctions that determine their relative utilities which characterizes the major net services in terms of their utility for computer-mediated communication, access to services or search.

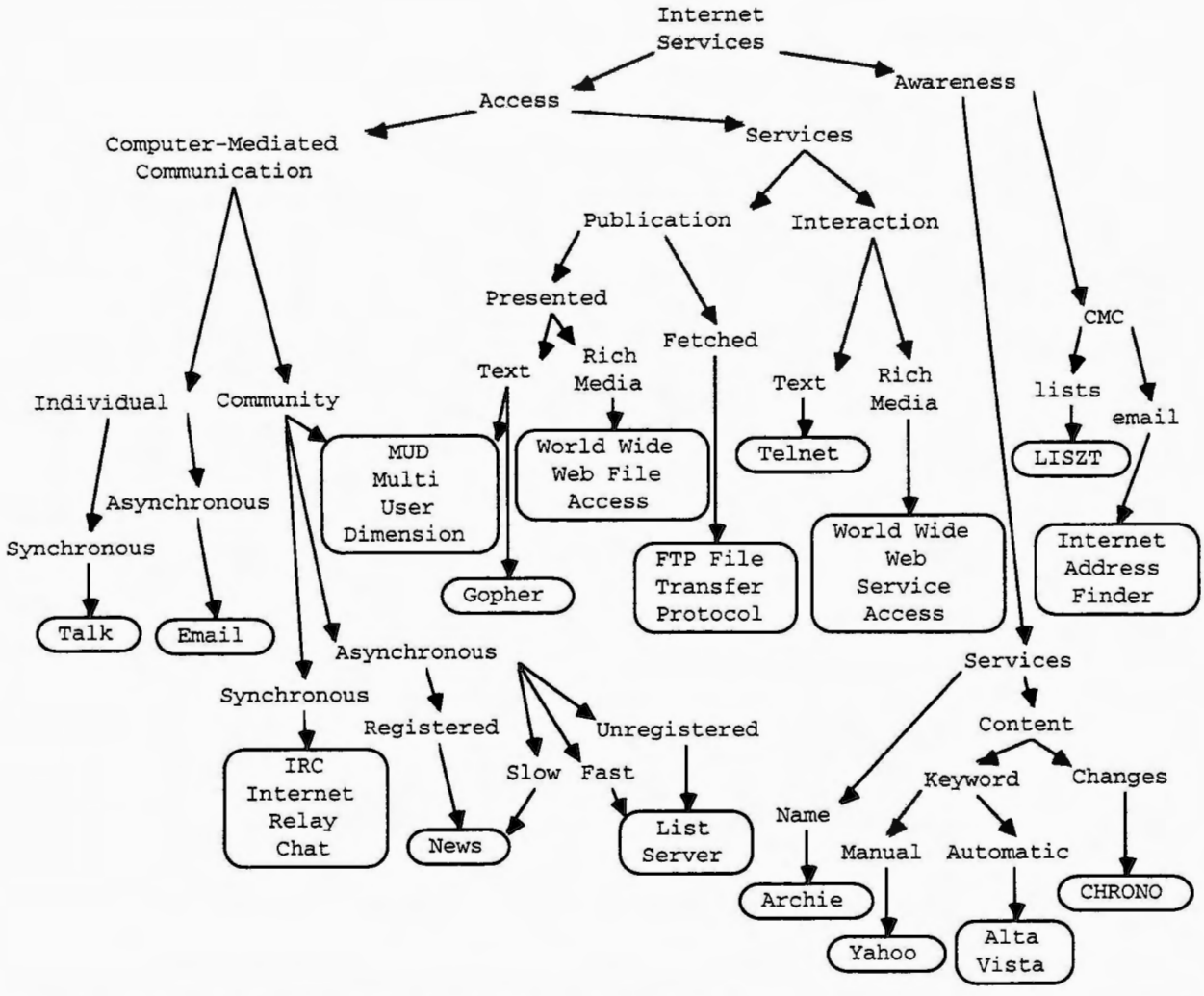


Figure 1 Internet services in terms of dimensions of computer-mediated communication

Figure 1 is a concept map presenting the major services on the net in terms of a small set of fundamental distinctions:-

- At the top level the major net services are characterized in terms of their utility for access to resources or awareness of resources.
- Access is sub-classified as to discourse, publications or services.
- Discourse is sub-classified by whether it is:-
  - — agent-to-agent discourse or community discourse;
  - — synchronous with the agents conversing in real time or asynchronous with substantial time delays in responses.
- Asynchronous community discourse is sub-classified by whether the channel is slow or fast, and whether the community is centrally registered or not.
- Publications are sub-classified by whether they are:-
  - — just fetched or presented when fetched;
  - — text or rich media.
- Services are sub-classified by whether they are text or rich media.

- Resource awareness is sub-classified by whether it is:
  — by resource name or content;
  — by keywords or by change in contents;
  — by keywords generated manually or automatically.

The less well-know systems classified are: *Internet Address Finder* which provides an index of email addresses; LISZT which assists users to search for a list server by its name; and *CHRONO* (Chen, 1995) which indexes a web site in reverse chronological order to provide automatic "what's new" pages. MUDs, multi-user dungeons/dimensions, are interesting in providing a mix of services supporting both discourse and resource access. Web browsers such as Netscape are interesting in providing a single tool accessing nearly all the services shown except talk and chat.

4. COOPERATIVE INTERACTIONS AND COMMUNITIES ON THE INTERNET

The exponential growth of the web 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. Collectively, the web/net can be considered as a large scale groupware for supporting special interest communities (e.g., researchers in high-energy physics).

Large scale groupware differs not only in the quantity, but also in the quality of cooperative interaction (Dennis, Valacich & Nunamaker, 1990). The fundamental nature of interaction on the web can be characterized as **virtual cooperative interaction**. The word "virtual" has two senses here: first, it denotes the notion of *virtual space,* i.e., the cooperative interaction occurs in a non-physical space which allows participants to be situated in geographically separate locations; second, it denotes that the *intention* to engage in cooperative interaction itself may not necessary pre-exist or be conscious. Traditional notions of groupware focus on the first sense (tele-presence in virtual space), but there is a need to extend the notion of cooperative interaction to encompass the latter sense of virtual cooperative interaction also.

Frequently, information resource contribution and exchange on the web involve cooperative interaction without pre-planned coordination. In fact, participants on the web may have no intention to cooperate in the first place. Quite often, a resource provider and a resource user are unaware of each other's existence until their first interaction. Nevertheless, the interactive process between them is still loosely cooperative in nature. It differs from the traditional team-oriented cooperation where group tasks, goals, and purposes are usually well-defined.

A classical *social exchange* model like Interactional Matrix model (Kelly & Thibaut, 1978; Cook, 1987) cannot readily account for this unusual form of cooperation where a resource provider might never know the identity of her resource users, nevertheless still continues to contribute anyway. On the web the only feedback she may receive might be the frequency of accesses to her information resources. What does she gain in return in such a seemingly one-way cooperative interaction? Is it simply an expression of altruism? What are some possible motivations for her to contribute to the web? In general, how would one ensure the continual contribution of an information provider? These questions can be answered more clearly in the context of *socioware*.

The present article defines **socioware** as: *computer-mediated environments for supporting community-wide processes which expedite virtual cooperative interactions*. Information inquiry and

response, dissemination of ideas, and social networking are examples of virtual cooperative interactions. USENET news groups and list servers are two prototypes of socioware that support dialogues within well defined special-interest communities on the Internet.

The proliferation of personal home pages with cross-linkage of web pages by people who share common interests has made the exploration process on the web (i.e., net surfing) a *social experience*. Such a seemingly intrinsic rewarding experience can often be characterized as serendipitous and not necessarily task-oriented (as in traditional groupware).

Through home pages, individuals create their own *virtual persona* on the web without any awareness of whom their eventual audience might actually be (i.e. without *extensional awareness* of particular recipients). However they often have a sense of who the potential audience might be (i.e. with *intensional awareness* of the type of recipient). Sometimes individuals provide information resource to the web as a by-product during some self organization processes of their own knowledge. As observed earlier, this form of apparently cooperative behavior is prevalent on the web.

In essence, the goal of socioware is to facilitate emergent pro-social behaviors for self-organized, virtual collaborative communities.

5. CONCEPTUAL MODEL OF VIRTUAL COOPERATIVE INTERACTION

This section describes a detailed model for that encompasses collaborative activities supported by traditional groupware and by emergent socioware. The model analyzes the following five basic elements for virtual cooperative interactions in CMEs:

1. *discourse patterns*
2. *time-dimension of virtual interactions*
3. *awareness hierarchy*
4. *motivations for cooperative behaviors*
5. *emergence and maintenance of virtual cooperative interaction*

Together they present three aspects (what, why, and how) of the conceptual model: (i) the descriptive aspect comprised of the first three elements which characterize and classify virtual cooperative interactions; (ii) the prescriptive aspect that provide motivational reasons for individuals to participate in virtual cooperative interactions; and (iii) the operational aspect of how virtual cooperative interactions initiate and function.

5.1 Some Definitions

Before describing the conceptual model in detail, the definitions of some frequently used terms in the model are introduced in this subsection.

The term **social entrainment** refers to some endogenous biological and behavioral processes that are captured, and modified in their phase and periodicity, by powerful (internal or external) cycles or pacer signals The notion of entrainment contains two kind of synchrony: (i) The *mutual entrainment* of endogenous rhythms to one another; (ii) the *external entrainment* of such a rhythm by powerful external *signals* or *pacers* (McGrath, 1990).

When individuals participate in virtual cooperative interactions, depending the nature of their present focus (e.g., discuss an idea, co-reviewing a book), there is a natural cognitive processing time involved in each activity. This processing time generates an endogenous rhythm within individual participants. This natural rhythm of interactions consequently creates *mutual entrainment*

in sustaining continuation of virtual cooperative interactions. The processes of social entrainment are important in the time-dimension of virtual cooperative interaction.

One can regard a collaborative community as a set of individuals that provide resources to one other with the most significant dimension relating to the coordination of the community being that of the **awareness** of who is providing a particular resource and who is using it (Gaines, Shaw & Chen, 1996).

5.2 A Punctuated Discourse Model of Computer-Mediated Communications

Figure 1 presents a conventional model of Internet services in terms of their utility, but it does not provide an integrative model of the way in which they support communities. Such a model can be developed by noting that what distinguishes discourse from publication is that in discourse it is expected that the recipient responds to the originator, whereas publication is generally a one-way communication. However, on list servers some material is published in that the originator expects no specific response, and material published in electronic journals or archives often evokes a response. Computer-mediated communication offers a very flexible medium that breaks down the conventions of other media. The following diagrams show the different characteristics of the main Internet services in terms of these issues.

Figure 2 shows email discourse as a cycle of origination and response between a pair of agents communicating through a computer-mediated channel.

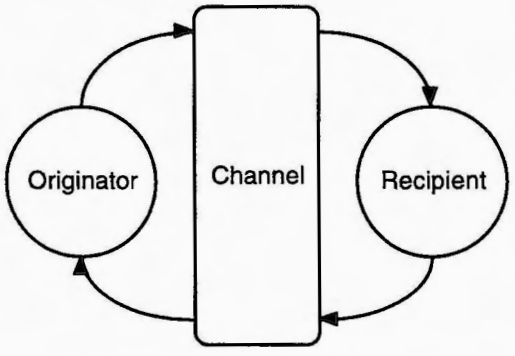


Figure 2 Email discourse

Figure 3 extends Figure 2 to show list server discourse as a cycle of origination and response between agents that is shared with a community through a computer-mediated channel. The community involvement leads to more complex discourse patterns in that: the originator may not direct the message to a particular recipient; there may be multiple responses to a message; and the response from the recipient may itself trigger responses from others who did not originate the discourse. For a particular discourse sequence this leads to a natural division of the community into active participants who respond and passive participants who do not.

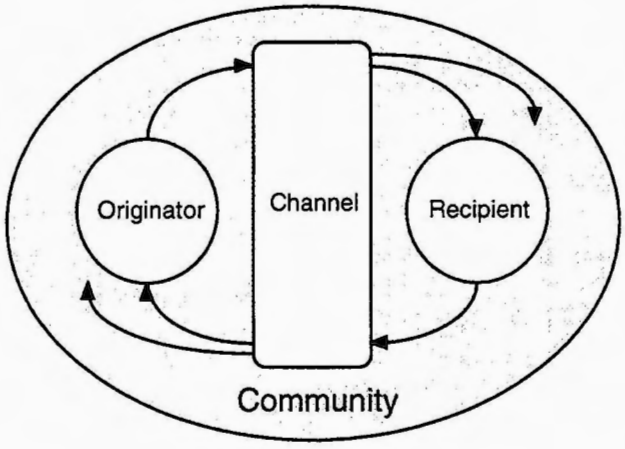


Figure 3 List server discourse

Figure 4 modifies Figure 3 to show web publication as an activity in which the channel is buffered to act as a store also. The material published is available to a community and the originator is unlikely to target it on a particular recipient. Recipients are not expected to respond direct to the originator, but responses may occur through email, list servers or through the publication of material linked to the original. Because the published material is not automatically distributed to a list, recipients have to actively search for and discover the material.

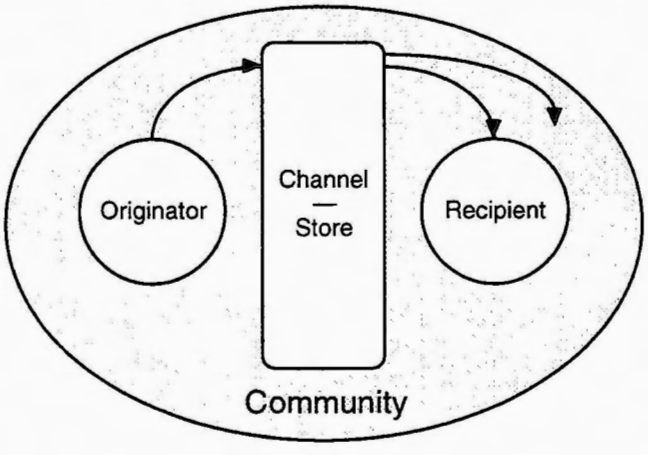


Figure 4 World Wide Web publication

The common structure adopted for the diagrams is intended to draw attention to the commonalties between the services. List server discourse is usually archived and often converted to hyper-mail on the web. Web publications do trigger responses through other services or through links on the web. A search on the web may not discover a specific item but rather a related item on a news group, list or by an author, and result in an request for information to the news group, list or author. Individuals and communities use many of the available Internet services in an integrated way to support their knowledge processes.

Figure 5 subsumes Figure 2 through Figure 4 to provide an integrated model of Internet knowledge processes that captures all the issues discussed. It models the processes as discourse punctuated by the intervention of a store allowing an indefinite time delay between the emission of a message and

its receipt. It introduces two major dimensions of analysis: the *times* for each step in a discourse cycle; and the *awareness* by originators of recipients and vice versa.

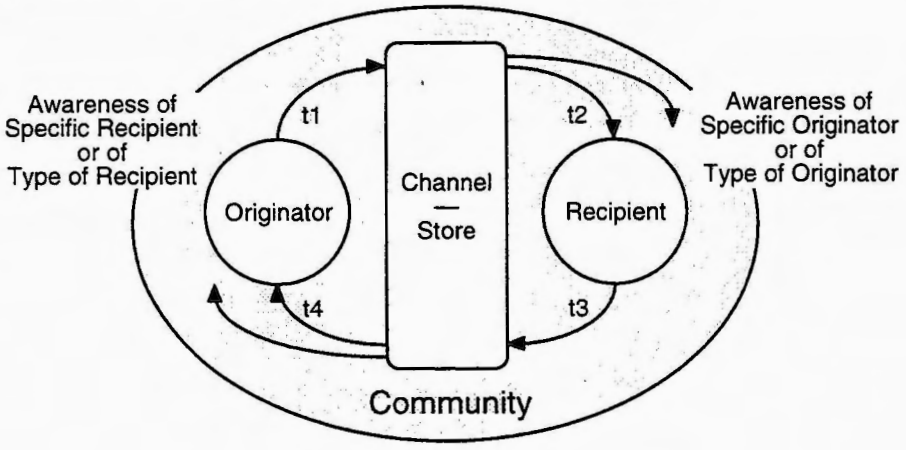


Figure 5 Punctuated discourse

5.3 Time Structure of Punctuated Discourse

The four times shown in Figure 5 are:

t1: the origination time—the time from a concept to its expression and availability
t2: the discovery time—the time from availability to receipt
t3: the response time—the time from receipt to expression and availability of a response
t4: the response discovery time—the time from response availability to receipt

Note that agent processing times and channel delays have been lumped. A study focusing on the impact of communication delays would want to consider them separately, otherwise there is no significant distinction—a general principle might be that communication delays should not be greater than agent processing times. Note also that the diagram is to a large extent symmetrical—the recipient becomes an originator when responding.

An important overall parameter is **time cycle**: the round-trip discourse time, t1+t2+t3+t4. If this is small, a few seconds or less, we talk in terms of synchronous communication. If its is large, a few hours or more, we talk in terms of asynchronous communication. If it is infinite, so that there is no response, we talk in terms of publication. However, this analysis shows that there is a continuous spectrum from synchronous through asynchronous to publication.

The discovery times, t2 and t4, are very significant to publication-mode discourse, and attempts to reduce them have lead to a wide range of awareness-support tools that aid potential recipients to discover relevant material and originators to make material easier to discover.

5.4 The Time Dimension in Virtual Cooperative Interaction

Awareness and coordination of cooperative interaction involve the processes of *social entrainment.* This subsection examines the relationship between the time cycle of virtual cooperative interactions and the relative strength of extensional and intensional awareness.

When two or more individuals participate in virtual cooperative interaction, they often take on dual roles of originator and recipient in punctuated discourse. Gradually they become locked into *social entrainment* processes. Computer-mediated environments, such as news groups (Resnick et al, 1994)

and shared drawing systems (Ishii & Kobayashi, 1992), provide specific external signals which set the pace of virtual cooperative interactions for participants. For example, the average *time cycle* for posting to a news group and receiving a feedback is about one to few days. Whereas the partial time cycle (t1+t2) for moving a mouse cursor in a real-time shared drawing system is around one to ten seconds. *External entrainment* occurs when the *actual* time cycle of a virtual interaction fall into the range of the *expected* time cycle anticipated by individual participants. When there is a wide discrepancy between the expected and the actual time cycle of interaction, participants often feel frustrated and decrease their desire to interact. For example, if cursor movements in a shared drawing system begin to take more than a few seconds to complete, the participants will tend to stop their interaction.

Continuation of virtual cooperative interaction can also break down when mutual and external entrainment processes are not synchronized with one another. When co-reviewing a book, the natural time cycle for mutual entrainment is in days and weeks, since it often take that a mount of time to read a book and absorb the material properly. It is unlikely that co-reviewers will want to use Internet Relay Chat (Reid, 1991) to disseminate and exchange their reviews. Such a fast time cycle of interaction is not well suited for activities involving deep, reflective cognitive processes.

The relationship between the time cycle and the relative strength of *extensional/intensional awareness* in virtual cooperative interactions can be illustrated in a time-dimension diagram (Figure 6). If the time cycle is relative short, say in few seconds or minutes, we have an interaction that can be characterized as synchronous (real-time). If it is longer, we have interaction that is often described as asynchronous (delay-time). The key notion here is that the types of virtual cooperative interactions are differentiated on a temporal continuum rather than by discrete categories.

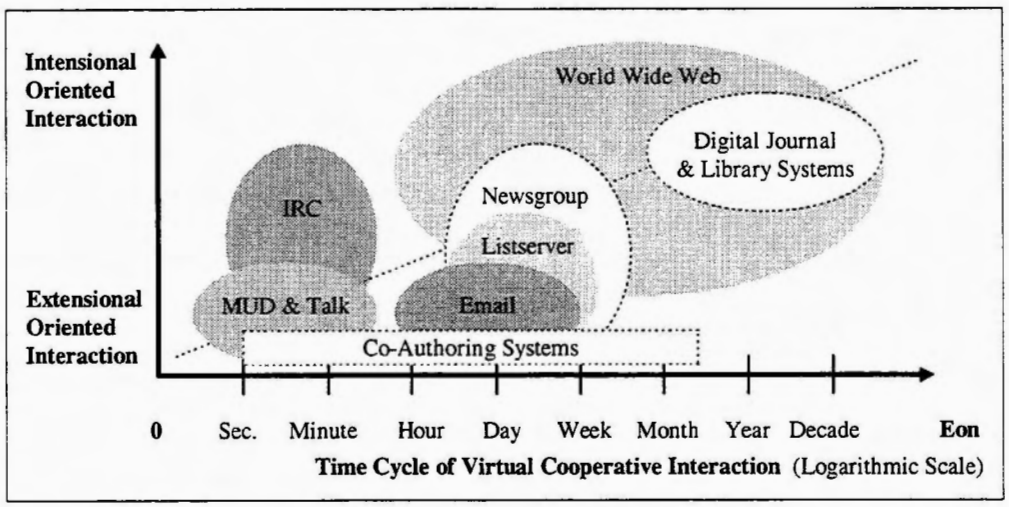


Figure 6 Time Dimension of Virtual Cooperative Interaction

In intensional oriented interactions, the level of intensional awareness is relatively high compared to extensional awareness; whereas in extensional oriented interactions, extensional awareness predominates. Many groupware systems (e.g., co-authoring systems, shared workspace systems) have been designed to support collaborative teams in which interactions are between known group members. Therefore, in these CMEs, the virtual cooperative interactions focus on extensional awareness. In contrast, interactions in USENET news groups involve both extensional and intensional awareness of targeted audience. For example, one can respond to a question from a

specific individual (an extensional oriented interaction) but do so publicly with the intention to address others who may have a similar question in mind (intensional oriented interaction).

The time cycle in virtual cooperative interaction often varies according to cognitive processes involved in any given moment of an activity. For example, during a collaborative writing session (Neuwirth et al, 1994): when co-authors' focus is on correcting sentences or paragraphs, the time cycle involved is usually around few minutes; and when they focus on reviewing chapters, the time cycle involved shifts to hours. Therefore co-authoring systems are classified in the range of time cycles from second to day, in addition to be extensional oriented.

The time dimension diagram of virtual cooperative interactions allows us to visualize CMEs in terms of an *interaction area* they encompass as shown in Figure 6. The area denotes the range of time cycle and the degree of intensional vs. extensional awareness.

5.5 Awareness Structure of Punctuated Discourse

One can regard a community as a set of agents that provide resources to one other with the most significant dimension relating to the coordination of the community being that of the *awareness* of who is providing a particular resource and who is using it. In the tightly-coupled team, each person is usually aware of who will provide a particular resource and often of when they will provide it. In logical terms, this can be termed *extensional awareness* because the specific resource and provider are known, as contrasted to *intensional awareness* in which only the characteristics of suitable resources or providers are known.

In a special interest community resource providers usually do not have such extensional awareness of the resource users, and, if they do, can be regarded as forming teams operating within the community. Instead, resource providers usually have an *intensional awareness* of the resource users in terms of their characteristics as *types* of user within the community. The classification of users into types usually corresponds to social norms within the community, such as the ethical responsibilities in a professional community to communicate certain forms of information to appropriate members of the community. Resource users in a special interest community may have an extensional awareness of particular resources or resource providers, or an intensional awareness of the types of resource provider likely to provide the resources they require. This asymmetry between providers and users characterizes a special interest community and also leads to differentiation of the community in terms of core members of whom many users are extensionally aware, and sub-communities specializing in particular forms of resource.

In the community of Internet users at large, there is little awareness of particular resources or providers and only a general awareness of the rich set of resources is available. Awareness of the characteristics of resources and providers is vague, corresponding to *weak intensional awareness.*

These distinctions are summarized in Figure 7 and it is clear that the classification of awareness can lead to a richer taxonomy of communities than the 3-way division defined. Analysis of awareness in these terms allows the structure of a community to be specified in operational terms, and in complex communities there will be complex structures of awareness. The coarse divisions into sub-teams and sub-special interest communities provides a way of reducing this complexity in modeling the community.

Locus of responsibility	Team	Special-Interest Community	Community at Large
Originator	*Extensional awareness of actual recipients.* Use email to notify. Use CHRONO to index.	*Intensional awareness of types of recipient.* Broadcast to list server. Establish HTML links. Use CHRONO to index.	*No awareness of recipients, or only weak intensional awareness of types of recipients.* Broadcast to news groups. Register in Yahoo. Initialize Alta Vista.
Recipient	*Extensional awareness of actual resources and originators.* Use email to inquire. Check CHRONO index.	*Extensional awareness of actual resources and originators, or intensional awareness of types of resources and originators.* Subscribe to list server. Follow HTML links. Check CHRONO index. Use WebWatch, Katipo or URL-Minder	*No awareness of resources or originators, or only weak intensional awareness of types of resources and originators.* Read news groups. Browse Yahoo. Search with Alta Vista. Search with MetaCrawler.

Figure 7 Communities and tools distinguished in terms of awareness

The differentiation of communities in terms of awareness draws attention to the significance of supporting various aspects of awareness in a CME system. *Resource awareness*, the awareness that specific resources or resources with specified characteristics exists, may be supported by various indexing and search procedures. However, there is also a need to support *chronological awareness*, the awareness that a resource has changed or come into existence (Chen & Gaines, 1996b). Figure 7 also shows the way in which current tools for awareness support are classified within this framework.

5.5.1 CHRONO: Chronological Awareness Support Tools

CHRONO 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 let of a Web site visitors (e.g., members of a group, an organization, or other net surfers) see which Web pages have been modified since their last visit. Currently, the CHRONO system is implemented on a UNIX platform and has been made widely available for use at other sites.[1] As shown in Figure 8, CHRONO presents to the visitors an HTML document that lists the titles of Web pages at the site in reverse chronological order. This chronological listing of Web pages also functions as a collection of hyper-links to the listed pages.

[1] To obtain a copy of CHRONO, see http://ksi.cpsc.ucalgary.ca:8008/cgi-bin/release?6ka

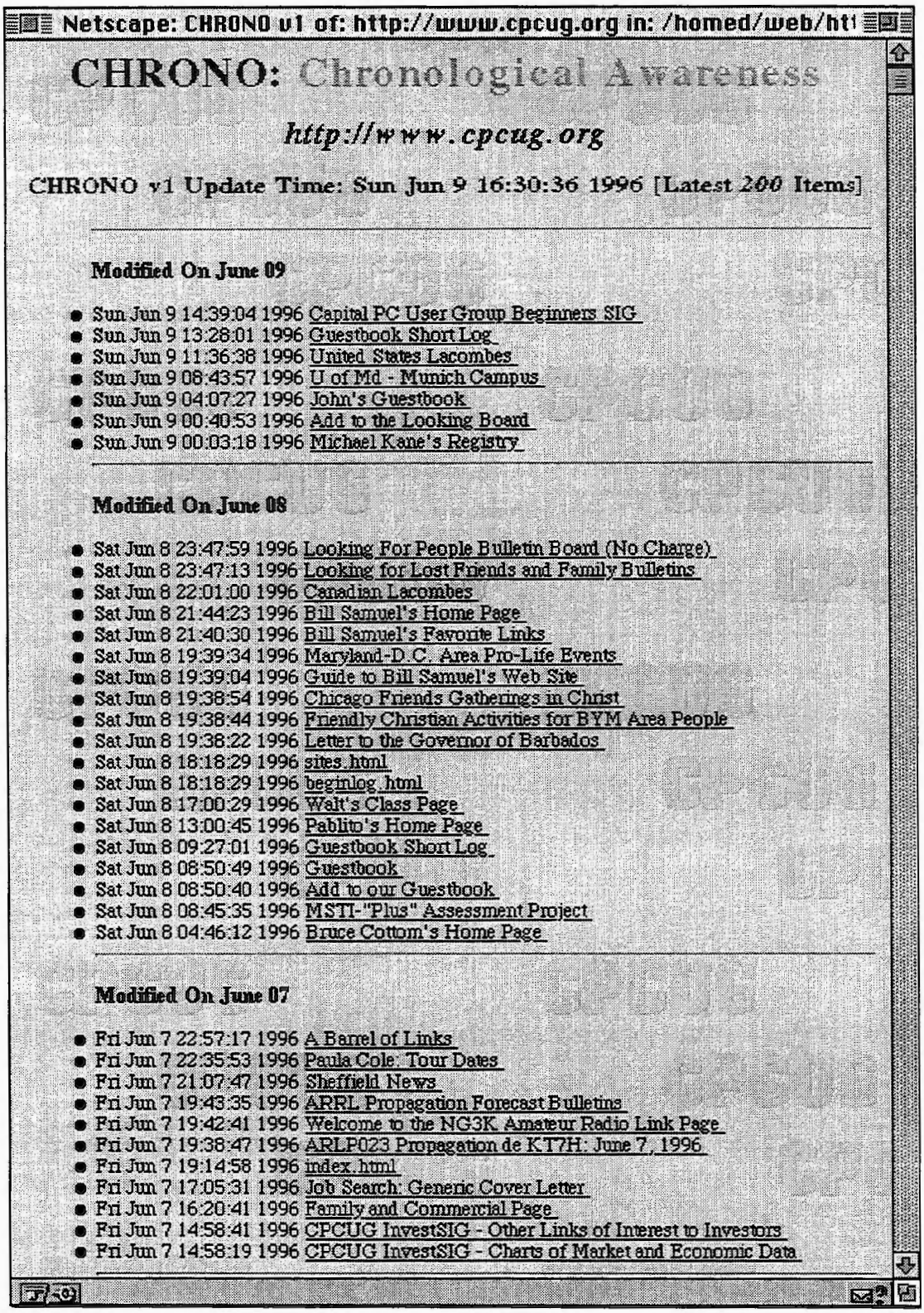


Figure 8 CHRONO in use at a PC user group site

This time-line dimension allows frequent visitors of a Web site an immediate awareness on what have been changed since their latest visit. The changes they see may be some Web pages in which they have particular prior-interests of or may be some pages that they have never seen before but now appeal to them. Hence this *chronological browsing* characteristic is analogue to *spatial (subject-category) browsing* characteristic that library patrons have often experienced when looking for books on open book-shelves (i.e., accidentally finding (more) relevant books near by the books that they are looking for originally).

What is different here is that instead of finding relevant information via browsing the near by subject-categories, now the users may find relevant information via browsing the concurrently modified/created web pages. Sometimes, conceptually related documents are created (or modified) around the same time, however their author(s) may not remember to update the HTML links to them. Unlike a manually updated *what's new* page in which the users have to rely on the timely updates made by a Web-master (or by the document authors), CHRONO provides the time-line dimension to the users automatically, in a reliable, periodic fashion.

WebWatch (Specter, 1995), *Katipo* (Newberry, 1995), and *URL-Minder* (NetMind, 1995) are other chronological awareness tools that track changes in specified documents. 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. Katipo is another client-side chronological awareness system built for Macintosh that shares many similar concepts as 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. It writes a report file (in HTML format) listing all such documents in a format that allows you to easily visit the updated documents. URL-Minder is a centralized system that keeps track of resources on the Web, and sends registered users e-mail whenever their personally registered resources change. Web users can have the URL-Minder keep track of any Web resource accessible via HTTP. It can be anything, not just Web pages that users personally maintain.

5.5.2 Further Developments

The CHRONO research program is now at a stage of measuring the structures and time constants of discourse on the Internet from empirical data. Studies are being carried out of the rates of diffusion of information and the various knowledge acquisition paths and processes whereby individuals become aware of information on the net. List server archives are being analyzed to determine the fine structure of discourse and to track the trajectories of ideas. CHRONO was issued in May 1996 and is now being used at a number of sites. A new program META-CHRONO is under development which will collect and collate information from multiple sites running CHRONO and provide awareness of activities being carried out on a distributed basis.

5.6 Motivations for Participation in Virtual Communities

This subsection examines the motivational dimension of virtual cooperative interaction. Here, a theory of *collective social exchange* attempts to explain the behaviors of participants in terms of exchange theory, effects of norms in virtual community, capacity of power and social influence. These motivational explanations together with *social learning theory* (in the next subsection) examine the fairness and reinforcement issues involved in virtual cooperative interactions.

When many individuals participate in a multitude of punctuated discourses (Figure 5), a chain reaction occurs. The accumulative effect generated by this chain of inquiry-response-reaction-response-reaction (and so on) is an evolving topical thread that can become a part of *shared knowledge* among community members. Through automatic archival services such as Hyper-mail or some individual efforts such as FAQs (Frequently Asked Questioned) and web pages, the shared knowledge persists and grows. An interesting question is: why should individuals contribute to this pro-social process? Correspondingly, how does the virtual community ensure its participants to contribute to the growth of the knowledge pool?

First, why would individuals want to participate in virtual cooperative interactions? Generally, interpersonal behavior can be characterized as a *social exchange* between people, and these social exchanges typically involved both rewards and costs to participants. On a balance, an individual will perform those actions which produce the greatest rewards at the least cost (Shaver, 1987). Therefore according to this cost-benefit calculus, a perceived potential for rewards must exist for individuals to participate and contribute in a cooperative relationship.

In contrast with *classical social exchange theories* (Cook, 1987) (e.g., Kelly and Thibaut's (1978) *Interactional Matrix* model) which emphasize dyadic interactions between individuals, **collective social exchange theory** focuses on interactions between individuals and their community. Conceptually, the Internet community is viewed from a *collective stance* (Gaines, 1994) as an entity to 'whom' individual participants exchange information resource with. This collective entity offers participants a valuable informational service (namely, as *a pool of human knowledge* (Berners-Lee et al, 1994) in exchange for their contributions.

The *norm of reciprocity* is fundamental to social exchange and leads to contributing behavior. The reciprocity norm creates an obligation for repayment that must be satisfied if the interaction is to continue (Shaver, 1987). However, the way reciprocity operates in collective social exchange is more subtle than in conventional social exchange between individuals. Why should one reciprocate (through contribution) in a situation where social responsibility is relatively *diffused* among community members?

One motivation for contributing to the net is for individual to gain positive *self-image* (Jones & Pittman, 1982). In this case, an individual has internalized the norm of reciprocity and acts according to the principle of **equity theory**: that is, a person will seek to maintain his ratio of rewards to costs as the same as that of relevant comparison persons (Wąlster, Walster & Berscheid, 1978). A sense of *guilt* would occur if the individual perceives he has not contributed enough to the community. Hence, he would want to reciprocate fairly.

Another more subtle motivation is that of contribution as an investment in **social power**, that is, the *capacity* of a person or group to affect the behavior of another person or group (Schopler, 1965). Contributions made by an individual may not only help others but may also help her to gain name recognition from peers. The more one contributes publicly and receives recognition for one's contributions, the more one gains the capacity of power to influence others or the community as a whole. The added weight in recognizing who is first to contribute relevant information also motivates individuals to volunteer information resource more readily. The competition for priority in contribution has been well documented in Merton's studies on the reward system in scientific discovery (Merton, 1973).

The motivational dimension of the model illustrates the importance of *feedback loops* (Losada, Sanchez & Noble, 1990) in the reinforcement of virtual cooperative interactions. It provides a coherent explanation for the apparent altruistic behavior of information providers on the web/net.

5.7 Reinforcement of Virtual Cooperative Interaction

One question raised earlier in the chapter is that: why people publish information resource to the web in the first place? Usually a resource provider might never know the identity of her resource user, nevertheless she contributes even without any potential and apparent playbacks for her effort. Two possible motivations described earlier for providing information resource on the web are gaining positive self-image and name recognition. How does such a pro-social behavior initiate and continue?

The concern here is with the relationship between the *effect* of an individual's behavior in a *virtual cooperative community* and its impact on the individual's later behavior. This is the basic to **operant conditioning**, the learning process by which behavior is modified by it the consequences of previous similar behavior (Ritzer, 1992). An individual emits some behavior. The community in which the behavior occurs in tern "acts" back in various ways. The reaction—positive, negative, or neutral—affects the individual's later behavior.

Social learning theory suggests that novel social behavior is first learned through imitation of actions taken by others who act as (social) models (Bandura & Walters, 1963). The reinforcement received by a model serves as information to the person about which behaviors are acceptable and appropriate for the circumstances. Once a novel action has been acquired through imitation, its probability of continuation is depended on the reinforcement it receives. *Vicarious reinforcement*, as well as direct reward or punishment, can play a part in social learning (Shaver, 1987).

On the web, an individual's first successful encounter with a home page full of relevant information resource provides a positive role model for imitation. Her subsequent positive net-surfing experience will further increase her exposure to other positive models. Once an individual internalizes the web culture which encourages construction of personal home page (which coincidentally also provides virtual persona for self-image), she will come to view that contribution to the web as a pro-social behavior and act accordingly. The dynamic of *social exchange* then comes into play here, if the costs of putting up information resources (e.g., research papers, hyper-links to relevant web pages) are relatively low to her (e.g., she has necessary skills and resources), she would contribute to the web. In addition, an original intention to contribute to the web community does not need to exist, she may coincidentally using her home page to organize her knowledge resource and contribute to the web community as an after thought (or as a by-product). In this situation, the extensional audience is herself together with a vague sense of intensional awareness of other potential resource users.

How does reinforcement come into the picture? Frequently, one would encounter some home pages that had been constructed months or years ago without any revisions or new contributions. Their authors have neglected them and ceased to contribute. Once a novel behavior has been acquired, it needs to have *intermittent, positive reinforcements* to sustain the behavior (Bandura & Walters, 1963). In order for reinforcement to take place, there must exist a feedback loop. The round-trip cycle of virtual cooperative interaction provides an individual the necessary awareness of the *effectiveness of investment in social power* which is crucial to reinforcing the behavior and leading to similar future actions.

An observable measurement of the effectiveness of social power on the web is the relative popularity of a web site. The popularity of a home page can be inferred from recognition earned by its visitor frequency counter, commentaries in its public guest-book, awards given by reviewers of popular web sites, and the number of other web pages linked to the page, etc. These gauges of popularity (which measure the relative power for social influence) provide direct reinforcements

(can be either positive or negative) to an information provider. They also offer indirect, vicarious reinforcements to other information providers by providing social models for comparisons.

6. SUMMARY OF THE CONCEPTUAL MODEL

The 1990s have seen the emergence of large scale collaborative activities on the Internet using email, list servers, news groups and the World Wide Web. There have also been developments of systems using some of these technologies to support smaller closely-coupled teams. In terms of the standard time/space taxonomy for CSCW, these uses of the Internet are generally virtual in space and range from highly synchronous to highly asynchronous interactions. However, many of the major applications of the Internet raise new issues that are not adequately addressed by existing models and taxonomies of CSCW.

Small groups of individuals working together generally have well-defined roles and mutual awareness of roles, tasks and activities. However on an Internet list server, a discussion may be initiated with only a vague concept of other potential participants but with strong expectations that a collaborative activity will result. On the World Wide Web, material may be published without only a vague conception of potential users yet that material may play an essential role in a collaborative active in some community, possibility not involving the originator, and perhaps a community of which the originator is not part. These phenomena are common in various collaborative scientific communities conducting interdisciplinary research. Those loosely collaborative communities are moving their knowledge acquisition processes to the Internet. It is interesting to know whether they can be modeled and supported using some extended CSCW frameworks (Chen & Gaines, 1996b).

The article presents a conceptual model for virtual cooperative interaction which encompasses the collaborative knowledge acquisition activities of closely-coupled teams and those of the very diffuse communities. It analyzes these activities in terms of the punctuated discourse processes, breaking down the cycles of action and response involved into a continuous temporal dimension. It analyzes them also in terms of awareness by originators of recipients and vice versa. The temporal dimension and awareness hierarchy enable the existing taxonomies and models of CSCW to be extended to encompass a very wide range of systems operating in both the short- and long-term and ranging from small teams to large communities. The model analyzes motivational aspects of virtual cooperative interactions. It gives rise to natural structural analyses of the activities which allows the types of communities involved to be identified from their observed activities. It can also be used to categorize computer-mediated environments roughly into groupware and socioware (Figure 9).

The conceptual model presented in the article implies that for successful maintenance of continual virtual cooperative interactions, the following criteria must exist:

- establishment of resource awareness for initial encounter
- establishment of mutual awareness as a feedback loop for continual virtual cooperative interactions
- compatibility between the expected and the actual time cycles of virtual cooperative interactions
- properly situated expectation of fairness in terms of collective social exchange
- accumulation of positive feedback for reinforcements in virtual cooperative interactions

	Groupware	Socioware
Awareness	strong mutual extensional	weak mutual intensional
Time cycle of Interaction	short to medium (seconds to days)	medium to long (days to years)
Motivation for Cooperation	individual social exchange	collective social exchange
Power Relations	well-defined roles as part of team definition	emergent roles from investment in social power capacity

Figure 9 Groupware and socioware comparisons

The current model also identifies the types of socioware systems that are needed to expedite collaborative activities, and provides a framework for classifying existing tools in use on the Internet. It focuses on participants' motivations and power relationship which determine their social roles, goals, expectations in virtual cooperative interactions. They are generally implicitly defined in groupware by the nature of group tasks (Mandviwalla & Olfman, 1994) and organizational structures (Kling, 1980). The article contributes to knowledge acquisition research by drawing attention to significance in large-scale socioware like the web where social and organizational structures are fluid and less well-defined. The conceptual model for virtual cooperative interaction expands the scope of groupware research. It provides a framework encompassing all forms of virtual knowledge acquisition processes from teams through organizational work groups to diffused, evolving communities (Chen, 1996b). Modeling and supporting virtual cooperative interactions on the Internet are important new challenges for knowledge acquisition research.

7. CONCLUSIONS

The purpose of the research reported in this article has been to develop a finer-grained conceptual model of the knowledge acquisition processes that occur in Internet communities in order to support and improve those processes through new and better services. The model developed suggests three levels of analysis of services:

- *Message quality*—the improvement of the multimedia capabilities of the basic message channel—there are been continuous improvement from simple text to typography, images, movies, sounds, animations, simulations, and so on.
- *Relationship modeling*—the incorporation of linkage information preserving discourse relationships—the hypertext links of the original web technology introduced this capability and clickable maps extended it—there is scope for further extension based on greater understanding of the roles the links play in enable people to grasp the argument forms of information on the web.
- *Awareness support*—the systematic reduction of the time (t2 and t4 in Figure 5) for a potential recipient to become aware of relevant information—manual and automatic indexing and various forms of search engine have made massive advances in coping with the information overload

resulting from the growth of the web—however, there is scope for many different tools supporting the various ways in which people manage their awareness.

The key question to ask in developing new awareness support tools is "what is the starting point for the person seeking information, the existing information that is the basis for their search." A support tool is then one that takes that existing information and uses it to present further information that is likely to be relevant. Such information may include relevant concepts, text, existing documents, people, sites, list servers, news groups, and so on. The support system may provide links to further examples of all of these based on content, categorization or linguistic or logical inference. The outcome of the search may be access to a document but it may also be email to a person, a list or a news group.

The net is a vehicle for discourse in which the goals of individual agents are supported through social knowledge processes, and support tool design needs to be based on increasingly refined models of those processes. Much of our current research is concerned with the empirical studies of discourse processes on the net through analysis of information diffusion, list server archives, and so on. We conjecture that tools that develop models of such processes and make them available to the participants may themselves result in improved usage of net resources.

Acknowledgments

Financial assistance for this work has been made available by the Natural Sciences and Engineering Research Council of Canada.

8. REFERENCES

Bandura, A. and Walters, R. (1963). *Social Learning and Personality Development.* Holt, Rinehart & Winston, New York, NY.

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

Chen, L. L.-J. (1995). CHRONO: *A Chronological Awareness Tool.* Knowledge Science Institute, University of Calgary. http://ksi.cpsc.ucalgary.ca:8008/cgi-bin/release?6ka

Chen, L. L.-J. and Gaines, B. R. (1996). Methodological Issues in Studying and Supporting Awareness on the World Wide Web. *Proceedings of WebNet96.* Association for the Advancement of Computing in Education, Charlottesville, VA.

Chen, L. L.-J. and Gaines, B. R. (1996b). A CyberOrganism Model for Awareness in Collaborative Communities on the Internet. *International Journal of Intelligent Systems*, (recently accepted).

Cook, K. S., ed. (1987). *Social Exchange Theory*, Sage, Newbury Park, CA.

Dennis, A. R., Valacich, J. S. and Nunamaker, J. F. Jr. (1990). An Experimental Investigation of the Effects of Group Size in an Electronic Meeting Environment. *IEEE Transactions on Systems, Man, and Cybernetics*, 20, pp. 1049-1059.

Gaines, B. R. (1971). Through a Teleprinter darkly. *Behavioural Technology* 1(2) 15-16.

Gaines, B. R. (1994). The collective stance in modeling expertise in individuals and organizations. *International Journal of Expert Systems,* 7 (1) 21-51.

Gaines, B. R., Shaw, L. G., and Chen, L. L.-J. (1996). Utility, Usability, and Likeability: dimensions of the net and web. *Proceedings of WebNet96.* Association for the Advancement of Computing in Education.

Ishii, H. and Kobayashi, M. (1992) ClearBoard: a seamless medium for shared drawing and conversion with eye contact. *Proceedings, CHI '92*. ACM, New York.

Jones, E. E. and Pittman, T. S. (1982). Toward a General Theory of Strategic Self-Presentation. In Suls, J. (ed.), *Psychological Perspectives on the Self*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Kelly, H. H. and Thibaut, J. W. (1978). *Interpersonal Relations: a theory of interdependence*. Wiley, New York, NY.

Kling, R. (1980). Social Analysis of Computing: theoretical perspectives in recent empirical research. *Computing Surveys*, Vol. 12, No. 1. ACM Press, pp. 61-110.

Krol, E. (1993). FYI on "What is the Internet?". Internet. RFC 1462.

Losada, M., Sanchez, P. and Noble, E. E. (1990). Collaborative Technology and Group Process Feedback: their impact on interactive sequence in meetings. *Proceedings of CSCW '90*, pp. 53-64.

Mandviwalla, M. and Olfman, L. (1994). What Do Groups Need? A Proposed Set of Generic Groupware Requirements. *ACM Transactions on Computer-Human Interaction*, pp. 245-268.

McGrath, J. E. (1990). Time Matters in Groups. Galegher, J., Kraut, R. E. and Egido, C., (Eds.). *Intellectual Teamwork: social and technological foundations of cooperative work*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Merton, R. K. (1973). *The Sociology of Science: theoretical and empirical investigations*. University of Chicago Press, Chicago, IL.

NetMind (1995). *The URL-Minder: Your Own Personal Web Robot*. NetMind. http://www.netmind.com/URL-minder/URL-minder.html

Neuwirth, C. M., Kaufer, D. S., Chandhok, R. and Morris, J. H. (1994). Computer Support for Distributed Collaborative Writing: defining parameters of interaction. *Proceedings of CSCW 94*, ACM Press, pp. 145-152.

Newberry, M. (1995). *Katipo--a Web Lurker*. Victoria University of Wellington, New Zealand. http://www.vuw.ac.nz/~newbery/Katipo.html

Reid, E. M. (1991). *Electropolis: communication and community on Internet relay chat*. Honour thesis. Dept. of History, University of Melbourne, Australia.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, and Riedl, J. (1994). GroupLens: an open architecture for collaborative filtering of netnews. *Proceedings of CSCW 94*, ACM Press, pp. 175-186.

Ritzer, G. (1992). *Sociological Theory*, 3/E. McGraw-Hill, New York, NY.

Schopler, J. (1965). Social Power. In: Berkowitz, L. (ed.) *Advances in Experimental Social Psychology*, Vol. 2. Academic Press, NY.

Shaver, K. G. (1987). *Principles of Social Psychology, 3/E*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Specter (1995). *WebWatch*. Specter Communications. http://www.specter.com/

Walster, E. H., Walster, G. W. and Berscheid, E. (1978). *Equity: theory and research*. Allyn & Bacon, Boston, MA.