Bridging the Gulf between Large-Scale Information Resources and Communities of Practice

Abstract

Community memory can provide the crucial bridge between large-scale information bases like digital libraries and the day-to-day activities of a community's members. Just as a digital library is based on a general structure and conventional means of access to diverse collections of materials, a community memory culls and shapes the structure of this collection to meet more particular needs; it provides a unique perspective on a larger, more general collection. Useful and usable community memories require support for: (1) the acquisition and evolution of content and structure; (2) the identification of relevant materials and community members; and (3) the maintenance of mutually intelligible organizations. In this paper, we explore issues related to these three requirements based on our experiences with the development and use of shared hypermedia information resources.

Keywords: community memory, hypermedia, digital libraries, collaboration, shared understanding, emerging structure

1. Introduction

In principle, digital libraries and large-scale information bases will provide physically distributed electronic communities access to a broad spectrum of archival materials, including those that we currently find in public, community, and work group repositories. But how will these communities bring these ever-increasing electronic resources to bear on their work? How will people use digital libraries in their day-to-day activities? How will they apply these emerging collections to information-intensive intellectual tasks -- research, design, education, analysis -- work that requires information to be gathered, understood, and communicated to others? We must bridge the gulf between the wealth of on-line resources and communities of practice in which these resources are used.

Community memories will form this vital link between large-scale collections and information-intensive work. Just as a digital library will provide a general structure and means of access to a collection of materials, a community memory will cull and shape the structure of this collection to meet particular needs; it will provide a unique perspective on the collection.

1.1 Community Memory and Electronic Communities

When people work together -- whether in designing a product, or creating training materials from video-based documentation, or writing a coherent analysis of a complex situation in the world -- they require, and put effort into constructing and maintaining, shared understandings of what they are doing: the task, the pertinent body of material, preliminary findings, progress, and methods.

We refer to the open-ended set of collective knowledge and shared understandings developed and maintained by the group as *community memory*.

Electronic communities are different than physical communities. They are ephemeral, forming and reforming according to interests, particular tasks, or issues. A person may be a member of many electronic communities, shifting his or her virtual presence from one locus of activity to another with ease. Our experience with electronic communities has drawn us to investigate ways of supporting these communities in the same medium in which they have formed.

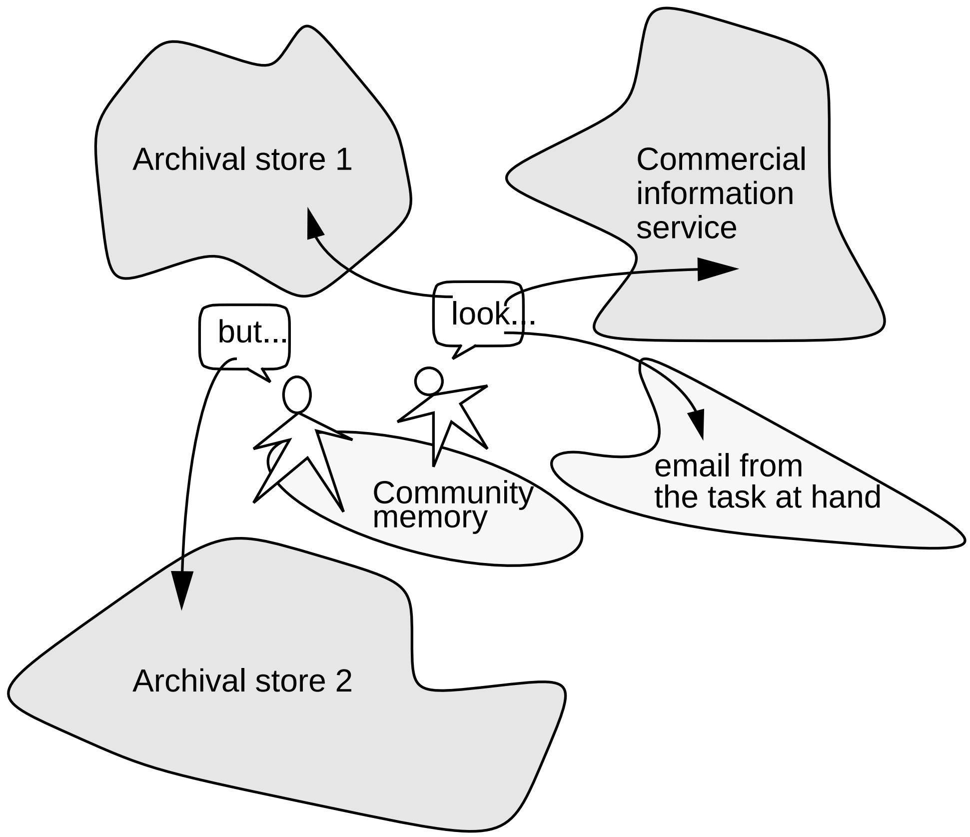

Figure 1. Digital libraries in use, forming community memory.

Through continuing discourse within a community, the materials in distributed digital library collections first become useful for performing the task at hand, then are recorded and made available for reuse through an evolving community memory.

Figure 1 illustrates our conceptualization of community memory in a distributed electronic community. Materials are drawn from many repositories, some archival, some transient (like news wires, email, and other semi-ephemeral forms) and, through discourse and interaction, combined with artifacts relating to the task at hand to form shared understandings that are community memory. Shared understandings in turn become the basis for communication and further work. Thus community memory may include discourse, collected materials, evaluations of these materials ("This is an important article") or sources ("This newsgroup has valuable information; this other one is a waste of time"), as well as marginalia and annotations, alternative organizations of materials, filters, and well-tuned queries.

1.2 Electronic Community Memory at Work: Early Examples

We can already see community memory at work in on-line communities on the Internet. Beyond providing access to distributed computing resources and remote information, the Internet has proven to be a particularly effective vehicle for human-human communication. It is the means by which electronic communities form and transient collections of materials grow in association with a task or topic. These collaborations, and their associated community memory, have the capacity to greatly extend the reach of the individual. This is not a new idea; Engelbart saw such a potential with Augment [4]. Instead this phenomenon is an important side effect of having the infrastructure in place to provide people with extensive digital resources and improved connectivity.

For example, in recent years researchers at different sites have met using this infrastructure as a means of organizing their conferences; they have discussed individual papers, the program, and decisions about conference content while referring to the body of submissions, and implicitly, to the body of literature in the field. In effect, through their conversations, they have formed a shared understanding about the current state of the field. In other cases, a topic (rather than a task) helps maintain a community: for example, high-energy physicists exchange preprints through the World-Wide Web as a way of shortening the review and publication cycle. NSF Collaboratory projects acknowledge the feasibility of distributed communities in information-rich domains collaborating on line.

1.3 A Spectrum of Resources

Physical collections like libraries are a good example of how information resources serve communities of differing scopes and levels of diversity. An individual's well-stocked bookshelves may be thought of as his or her library; others close by may avail themselves of the collection and its owner's curatorial abilities. Groups maintain reading rooms that stock all the latest journals and the body of reference works that they regard as central to their work. Institutions like universities and large corporations have extensive libraries that reflect their goals and purposes. Physical communities have libraries that can be thought of as including not only resources but also services and spaces that are aligned with the needs of their constituencies. Broader still are cultural and archival resources like the Library of Congress. Each of these sorts of information resources relies on the existence of roles like publishers, libraries, readers, and writers [20].

How, then, are community memories different?

Community memories are an important artifact of library use at each level of diversity. For example, in the context of an individual's bookshelves, community memory may take the form of marked-up copies of materials filed in an idiosyncratically arranged file cabinet. A group might cull broader materials to build up a project notebook. An institution might maintain a database of intellectual property documents and internal communications about them. Cross-community interactions might result in something like the World-Wide Web. Notice that in each of these cases, the derived forms refer to materials in the larger-scale information resources, but they come into being and are maintained through an unpredictable blurring of roles: readers, writers, librarians, and publishers may all be the same people.

We have already seen people try to bridge the gap between tomorrow's electronic resources and the distributed communities that use them in their work. Proposed digital library functions include: guided tours and automatically recorded reader paths; the ability to self-publish and move material in and out of the library at will; sharing of annotations; and voting schemes [23]. Digital media make this bridge much easier to construct.

1.4 Challenges

We thus find community memory to be a linchpin to the effective performance of intellectual work. But our past experiences building systems to support the elicitation and reuse of community memory and experiences with network-wide collaborations have shown us that there are significant barriers to realizing a fully articulated, well organized, usable electronic community memory.

Building useful and usable memories for distributed communities presents fundamental challenges [30, 2]. Although it appears to be easy to amass the kinds of materials that are part of a community memory -- for example, electronic mail, culled, annotated library materials, "war stories" about how prototypical problems were solved in past situations, software that embodies a particular way of processing digital library information, or videotapes documenting an activity -- it is still problematic how to put these materials to productive use over time.

Community memories need to be seeded, maintained, and generalized; they need to reflect the evolution of shared understanding. Members of the community must be mutually aware of each other's contributions, and the contributions must be mutually intelligible. Effective community memories cannot exist in isolation either from the tasks at hand or the information resources they refer to. Finally, and most crucially, they have to be useful to the members of the community: they must contribute directly to the work activities.

2. Community Memory: Issues of Acquisition, Comprehension, and Location

We are looking at community memory as a shared interpretive layer on top of sifted subcollections that refer to materials taken from both within and outside of digital libraries. To provide technological support for community memory, we must examine the situations from which it arises, and the challenges associated with our collective set of experiences designing systems to support community memory.

How do people use community memory as a resource for performing intellectual work? First, they find the materials they need for their work (many times by consulting colleagues, assistants, librarians, experts, and other human resources); they read or otherwise apprehend portions of materials they've gathered; finally, they modify these materials to suit the purposes at hand, where modification may include synthesis of diverse sources, paraphrasing, quoting outright, or using the gathered information as a taking-off point. Thus, to perform information-intensive intellectual work, a member of a community will act in many different roles: as searcher, as reader, as contributor.

We will take each of these roles as a separate vantage point, and examine the issues and challenges raised by each. Because the materials must be in place before they may be used, we first take the contributor's perspective. Once these material exist as a community-maintained electronic resource, we can begin to examine how people locate the relevant portions. Finally, we take a reader's view -- how the materials may be understood outside of a prescribed, pre-defined structure. We ground our discussion in smaller-scale experiences with hypermedia systems, since hypermedia is a good representational medium for creating community memory. We use these experiences to anticipate how the issues revealed by these systems in use scale to much more extensive electronic resources.

2.1 Acquiring Information: Contributor's Perspective

Although it is easy to amass materials for community memory, it is difficult to provide the incentive to add the requisite organization that will make the shared resource useful to others [18, 3, 5]. In general, this difficulty is intrinsic to certain types of groupware: contributors' efforts may far outweigh the benefits they derive from the work [9]. Many existing efforts to provide group memory or support long-term community-wide discussion have found that without an individual's single-minded devotion to starting them, keeping them going, and maintaining them, the information space slowly dies and becomes irrelevant, even to its originators. It is difficult to ensure real, continuing participation as well as the casual browsing we might encounter today on the Web.

The difficulties of acquiring community memory are exacerbated by both technological and social factors. First, contributors often don't derive benefits commensurate with the amount of effort they expend: there is a large gap between the collected materials they've used in their work (their files, for example) and materials that have been organized so that others may profitably use them. Not only does the structure of these materials arise over time and in conjunction with particular tasks, but any additional structure brings with it a considerable amount of overhead [22]. Second, as a changing, evolving form, community memory requires continuing thoughtful maintenance to weed out growing inconsistencies and redundant contributions. Finally, community memory arises out of tasks that take place in a distributed, heterogeneous environment, one that involves paper as well as digital media (see for example the description of analytic work in [11]), multiple authoring tools, and many different collections of source material, retrieved from a multiplicity of information services, each with its own formats, access methods, and protocols; this blend of materials, media, and technology presents significant obstacles to the construction of a side-product like community memory.

2.1.1 The emerging structure of shared resources: Incremental formalization.

Through our work with tools like Aquanet, VIKI, and HOS, we have shown that the groupware cost/benefit paradox [9] may be amenable to solutions like tools to support the gradual emergence and evolution of structure and techniques to support incremental formalization [16, 15, 27]. These tools and techniques emphasize low-cost means of adding the kinds of structure that may organize information from a digital library into a community memory.

Aquanet [14] is a good example of a group tool that suffered from the cost/benefit paradox. One of Aquanet's principle roles was to act as a collaborative front-end for the exploratory manipulation and organization of large collections of documents relevant to a particular task; in particular, we had hoped people would work together to create large, tightly interlinked structures of argumentation and evidence in the course of performing long-term analyses. These structures would encourage people to develop multiple interpretations of large collections of always-changing, possibly conflicting materials and would form a shared interpretive layer over institutional databases and commercial information services.

Aquanet provided specific support for users to create and manipulate complex graphical knowledge structures. In our original conception of the knowledge structuring task, users would define graphical representations of the elements in their problem domain and specify the ways in which these elements could be interconnected. Users could then apply and change these structuring schemes over the course of their tasks. Thus Aquanet provided a flexible way for people to record the abstractions they use to interpret information, to reflect and critique their analytic frameworks, and to explicitly negotiate about how information is structured (all crucial elements to a successful community memory).

But we did not anticipate the degree to which people found the definition of such domain descriptions (schemas) difficult. Not only were we requiring users to categorize materials in their domain according to a schema when they brought these materials into the tool; we were also requiring that they define the structuring schema itself before they embarked on their tasks (although the schemas could subsequently be modified).

From observing Aquanet in use, we found that informal representations are crucial to coaxing out partially formed, emerging interpretations. One of Aquanet's unexpected strengths was the ability it gave people to express interpretations -- interpretations that were less than fully formed -- in terms of visual appearance or position in the shared space. Extra-linguistic means of expression proved to be vital, allowing categories to be created without labels and relationships between documents to be expressed visually. The kinesthetic process of "trying things out" (as one might do wiggling molecular models in space or moving a jigsaw puzzle piece into different orientations) was not eliminated because a person was using a computer instead of manipulable paper objects in the world.

Figure 2 shows a portion of a shared space that was created in Aquanet, one constructed during the course of an evaluation of foreign language translation software. Each distinct visual symbol refers to an article from the trade press, notes about a software package, or contact information for a company.

But constructing such a (still small-scale) community memory proved to be difficult. How could we make it easier?

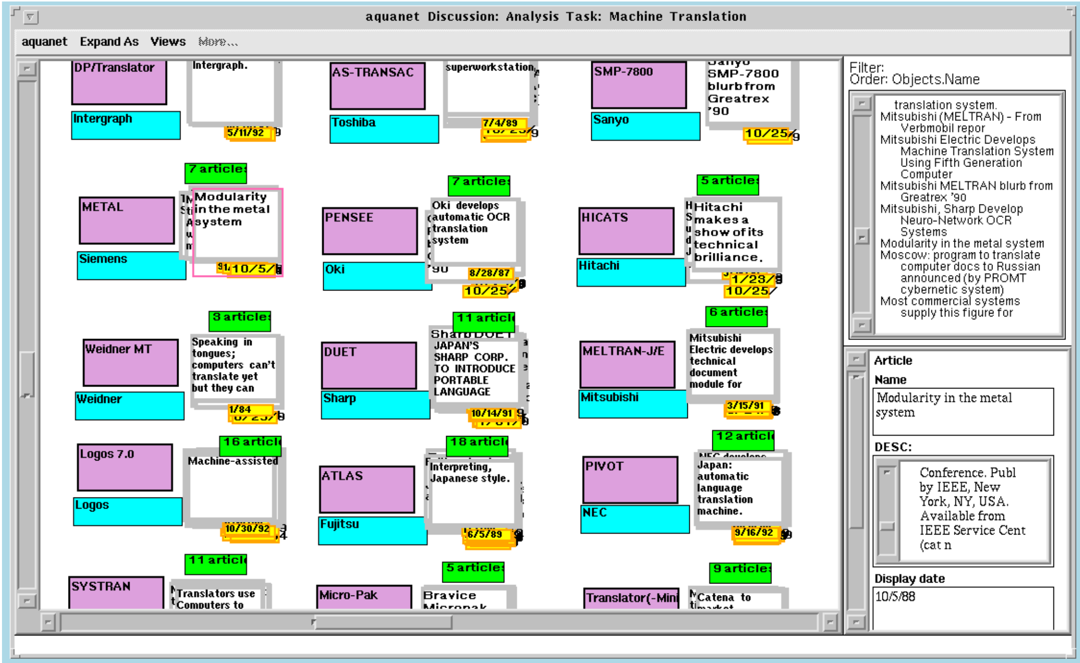

Figure 2. Aquanet containing information on foreign language translation software.

Out of our experiences with Aquanet, we designed VIKI, a tool to support emergent, dynamic, exploratory interpretation [15]. VIKI supports the ad hoc use of a visual symbol language so people can see and express structure as it becomes apparent to them. In contrast with Aquanet, developing this language is well-integrated with the task at hand. Because interpretation -- along with the concomitant act of organizing materials -- is opportunistic, users are not confined to a particular working style; they may work from gathered examples to develop structure, they can work schematically (the mode Aquanet enforced), or they may leave structure and meaning largely implicit. Thus, in VIKI, we complement the ability to develop abstraction and reflect on and critique interpretive frameworks with the flexibility offered by ad hoc, visually salient representations. We see support for emergent structure as a partial solution to the cost/benefit paradox inherent in computer support for community memory.

Figure 3 shows a shared space in VIKI. Note that the structures that people have built are similar to those shown in our Aquanet example -- visually salient, regular, but not wholly conforming to a pattern, a mixture of typed and untyped entities. VIKI includes facilities for recognizing, using, and declaring this kind of implicit structure. Visual structure is built up and becomes the basis for sharing knowledge.

As we discuss later on in Section 2.3, structure helps keep community memories intelligible to the members of the community. But more importantly, formal structure is also computationally tractable, raising the possibility of computer support for a community's activities. With the Hyper-Object Substrate (HOS), we have investigated the process of incremental formalization to support the emergence of structure. To this end, HOS integrates hypermedia and knowledge-based representations. Hypermedia eliminates many of the cognitive costs of formalization that

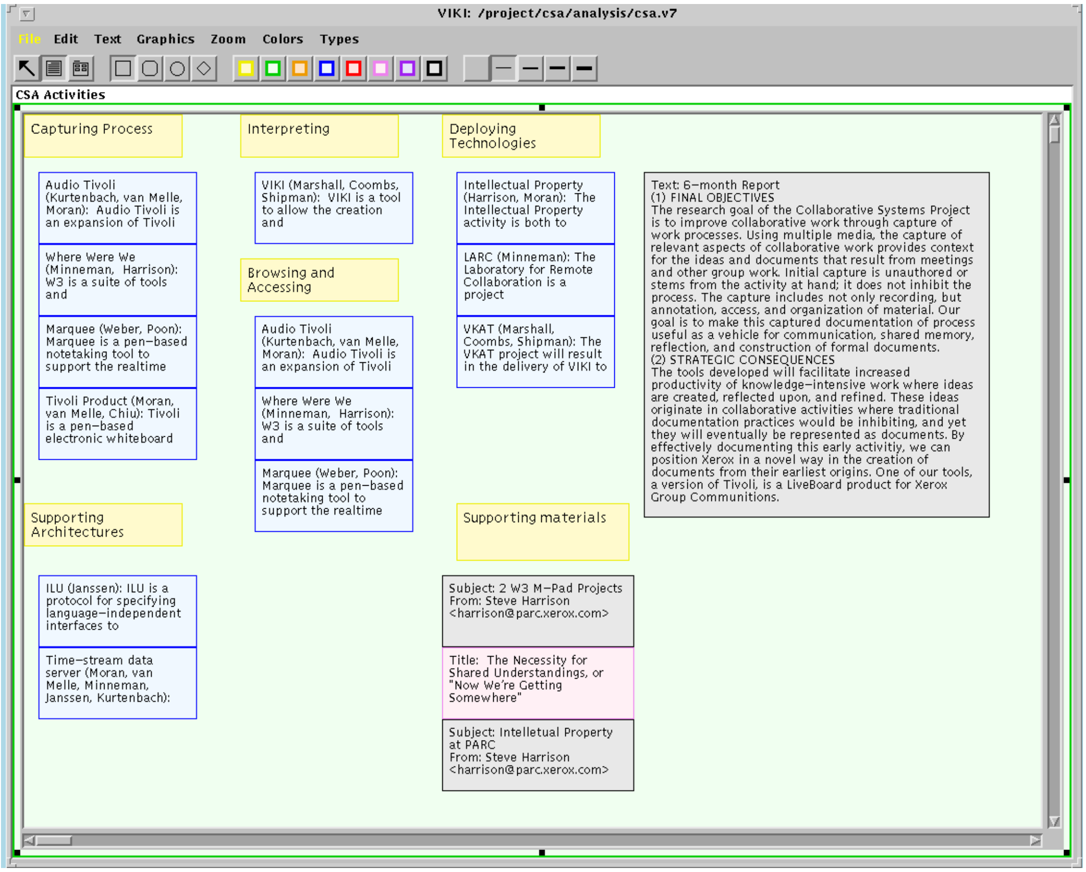

Figure 3. Objects in VIKI arranged into visual structures representing an evolving interpretation of the object's content.

inhibit user input. Integration with a formal knowledge representation reduces the burden of formalization by allowing it to be distributed and making it demand-driven [27].

To further lower the cost of formalizing information, HOS actively supports incremental formalization with mechanisms to recognize emergent structure implicit in the community memory and suggest formalizations based on this structure (a content-based analog to VIKI's visual/spatial structure perception). Experience with the use of HOS indicates some success and a greater potential for investigation of both methods for producing and interfaces to suggesting possible formalizations.

2.1.2 Maintenance of memory: Seeding, evolutionary growth, and reseeding.

We have observed three major types of processes -- and stages -- in the life cycle of community memories: seeding, evolutionary growth and reseeding. Seeding is the creation of the initial body of information in community memory. When this initial set of information reaches a certain size and level of relevance to the community, it starts to grow and evolve spontaneously as the result of additions made by its users (as has happened in the World-Wide Web). Seeding ends with the start of this evolutionary growth. After this growth proceeds for some time, the memory starts to become less and less useful; as a consequence, both use and growth may diminish. This happens for a number of reasons, such as growing disorder in the memory and the “needle in the haystack” problem -- i.e., the increasing difficulty of finding useful information in the growing information collection. At this point, the community memory must be revised -- i.e., reseeded. Its contents must be organized, winnowed, prioritized and generalized. The methods for locating things in memory may themselves need to be altered. If this reseeding is done successfully, the system can start another stage of evolutionary growth, after which it will in turn need to be reseeded if it is to continue to serve its users.

We have repeatedly experienced this three-fold process in our attempts to build community memories, for example, with large Issue-Based Information Systems (IBIS) structures [18]. Very few IBISs for groups have gotten started without the dedication of a single person or small core group of people who were willing to create the seed: i.e., the initial set of issues, positions and arguments. We have found that attempting to get the IBIS users themselves to invent -- out of the blue -- relevant issues, answers and arguments is a frustrating and generally unproductive experience for all concerned. Once there is some argumentative discussion for users to react to, the situation changes dramatically. It is easy to get people to react to what others have said, and the difficulty changes from trying to elicit information to trying to keep up with the information elicited. In our experience, this change makes it quite clear when the evolutionary growth stage of an IBIS has begun.

We found that as an issue base grows, its maintenance becomes increasingly difficult and error prone; growth leads to increasing disorder in the issue base. We also found that it became increasingly difficult to locate relevant information. These two problems have devastating synergies. For example, a given issue would often be raised and stored repeatedly, typically with slightly different wording. These redundancies were very difficult to detect, in part because of the difference in wording. Thus group discussion became fragmented into parallel discussions. As time went on the fragmentations grew in number and even compounded themselves -- with branches of the fragmented discussions in turn becoming fragmented. As a result, the IBIS increasingly ceased to function as a vehicle for group communication. To restore it to functionality, it was necessary to reseed the IBIS through a comprehensive edit of the issue base.

We have also observed this three-fold process in the creation and development of a number of large software systems, such as Symbolic’s Genera, Unix and the X-Windows system. In such systems, after the creation of the initial versions of the systems (seeding), users developed ad hoc additions to system functionality and often shared these as a community (evolutionary growth). These additions were often winnowed, refined, combined and included in later official versions of the software (reseeding), after which they entered another stage of ad hoc additions to functionality (evolutionary growth) [6].

2.1.3 Non-monolithic resources: Connecting community memories to external information.

Monolithic attempts to create community memory are difficult and soon fail because they are disconnected from the tools and materials people use to perform the task at hand as well as from outside reference materials, including the outside information they obtain from other people [19,13].

The assumption implicit in the design of many computer-based tools is that communication is a separate process from the user's main task. An analysis of computer network designers showed how the logical map, a representation of the design which shows network device interconnections, acted as the central artifact around which most communication occurred [21]. In response, XNetwork, an environment for supporting network design, provides designers with an integrated view of the design and the discussions about the design in conjunction with methods for importing electronic mail and bulletin board discussions into the design space. Figure 4 shows how the discussion and the design can be created and viewed together in XNetwork. The need to integrate discussion and artifact signals a more general need to integrate source information and produced information.

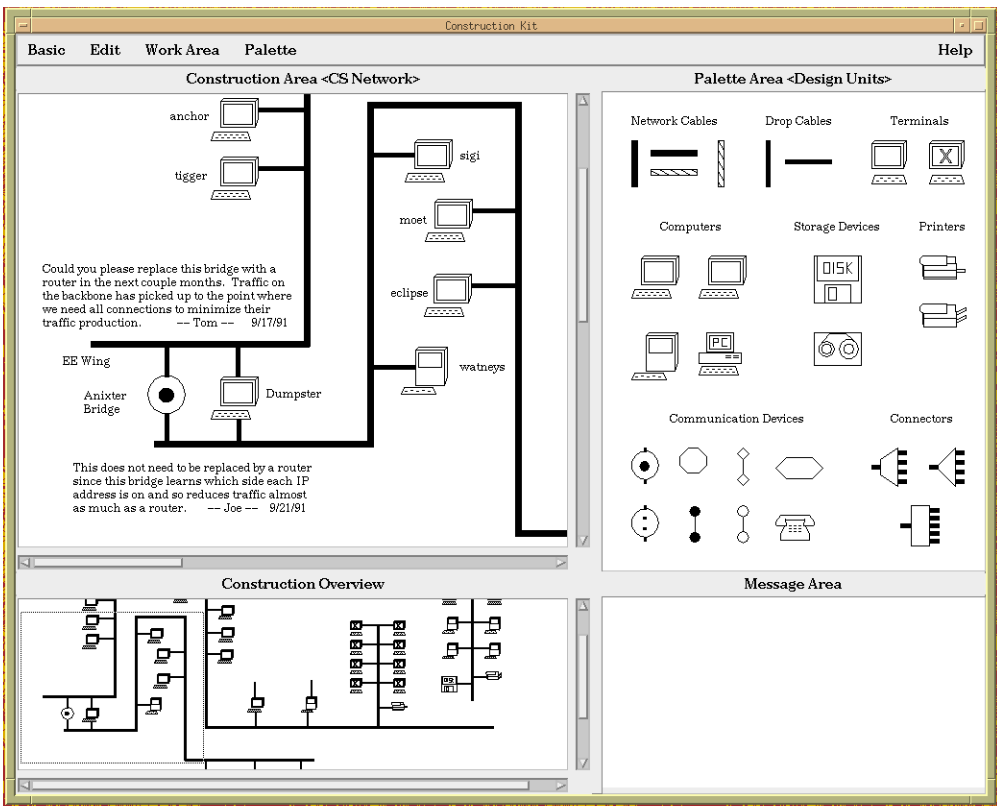

Figure 4. XNetwork showing part of a network design along with the discussion about an issue that occurred during the creation of the design.

Just as community memories must be connected with the means of communication about their content, so must they be connected with the universal collections from which they arise. For example, prototypes of the Virtual Notebook System (VNS) [24, 8] used generic hypermedia to overcome the difficulty of integrating various sources of biomedical research information. The VNS was intended as an electronic analog to a researcher's notebook that could also act as a shared repository of information gathered from early digital libraries and other on-line sources. Such external information resources included the National Library of Medicine's Medline database containing bibliographic and abstract information on articles from medical journals.

Users could connect to the Medline database through a graphical interface and could easily "paste" interesting information into their hypertext for later use.

Experience with these early prototypes of the VNS shows the difficulty of providing the needed connections to a variety of information sources and media. Besides the connection to Medline, the VNS included interfaces to the researchers' organizational information resources, i.e. hospital and departmental information systems, as well as to their research information resources, such as genome and experimental data databases. As these examples show, the specific information resources used by a particular community can differ greatly in scope. The experimental databases were used by only one research group; the genome database was shared by a number of groups; and the hospital and departmental information systems were used by most of the staff within the institutions.

To summarize, our experiences have pointed out how the problem of acquiring usefully structured information for the community memory involves supporting the emergent qualities of structure, the maintenance of the content and structure, and the connection to external sources of content and artifacts for discussion. Now we will examine issues concerning how information in the community memory is recovered.

2.2 Locating Information in a Community Memory: Searcher's Perspective

At first it might seem that the problem of locating information in community memory is exactly analogous to locating material in a larger digital library -- a classic problem of information retrieval, and thus amenable to treatment by a body of existing information retrieval techniques. While this is certainly true in part, there are a number of respects in which the retrieval issue is largely and fundamentally different for electronic communities. The first has to do with the continually evolving nature of community memory. The other has to do with the way in which the community itself mediates retrieval, even the retrieval of information from community memory.

2.2.1 "You don't know what you don't know": Mechanisms for volunteering information

One of the biggest problems with shared, collaboratively-constructed resources like community memory is that members of the community are often unaware of when there is some critical piece of information (either represented in electronic form or in another community member's purview) that is critical to their task. With many people contributing to a community memory, knowledge of the overall contents is necessarily distributed. It is apparent from looking at how community memories grow that a central obstacle to their success is "not knowing what you don't know." This means that no matter how capable a retrieval mechanism is, the user has to know to ask for information to receive it. To facilitate the location of information relevant to the task at hand we have developed a variety of mechanisms to volunteer the user with information without their explicitly asking for it.

The JANUS system supporting design uses the relationships between domain-oriented construction kits and a domain-oriented issue base to integrate argumentative information into the task of constructing solution form [5]. JANUS employs knowledge-based critics that "look over the designer's shoulder" and critique partially constructed solutions, pointing out potential inadequacies and providing relevant rationale from a domain-oriented issue base.

Generalizing from JANUS's critics, XNetwork includes agents to support the recovery of relevant information from a community memory. In a manner similar to JANUS's critics, XNetwork's agents volunteer information or take some action based on the user's current actions. XNetwork agents can be created by designers to act as proponents of certain information and opinions. As part of this creation, the designer can select from a variety of methods for informing the recipient of relevant information, enabling different levels of intrusiveness depending on the nature of the content. In this case, agents act as surrogates for users, advertising the existence of information deemed important. In this way the agents support communication among the members of a community.

2.2.2 Finding good information is more than IR: Support for community-mediated location.

There are, of course, many instances where community members feel a need for information and set out to retrieve it. Perhaps the central point about retrieval in electronic communities is that informed people are the best sources of information [7]. Community memory can serve two crucial functions in helping people to find information. First of all, it can serve as a cache for that information and evaluations of its worth, thus reducing the difficulty of search and increasing its effectiveness. Secondly, it can serve as a means for identifying community members who either know the information or can help in locating it.

Community memory might consist in large part of explicit records of the knowledge of the individuals in the community. This knowledge can be stored in a number of ways, perhaps the most basic of which is frequently asked questions (FAQs). In fact, an IBIS on recurring issues can be seen as nothing more than a souped-up FAQ collection.

As research on IBIS hypermedia has shown, the problem of retrieving issues is by no means merely a conventional information retrieval problem. Above all, it requires more than retrieval by content or bibliographic reference. Retrieval of relevant information in complex question-based discussions is decisively aided by associative indexing -- i.e., indexing by the relationships among questions [19]. For one thing, answering a query (question/issue) might be aided by the answers given to similar queries. The answer might also depend on the answers given to other queries. Such similarity and dependency relationships are also valuable information that can aid retrieval.

Most of the knowledge of community members is not and cannot be stored in community memory. Even so, a community memory can still be a decisive aid to retrieval of such knowledge if it can guide the question-asker to the community member who has the knowledge. There are at least two ways in which community memory can be of help in this situation. One is by storing the questions that its members want answered, so that other members can become aware of these information needs. The other is by storing information about the types of knowledge possessed by its various members--i.e., who knows what types of things. Community members may themselves be the best guide for finding other knowledgeable community members.

From these experiences, the location of information in a community memory differs from many other information location problems because of the continually evolving nature of the information and the distributed responsibility for the information's structure. In response, we have described the use of information volunteering mechanisms and human-mediated location.

2.3 Comprehension of Community Memory: Reader's Perspective

Once located, information must be applied to the community member's task at hand to be useful. If we examine shared resources from a reader's perspective, two main challenges emerge. First we must be concerned with the community memory's intelligibility -- are the materials organized and represented in such a way that they may be understood not only by the person who contributed them, but also by other members of the community? Can a community reach the high ground of shared understanding?

Second, we must be concerned with reusability -- will readers be able to apply the collected materials to their tasks? Will they be able to reformulate and generalize materials specific to one task such that the materials are once again useful from a new perspective? Will members of the community be able to re-apply schematic structures to organize new material that they've brought to the task?

2.3.1 Achieving a shared understanding: Enabling meta-communication.

Community memory critically rests on idea that any one community member's contribution to such a shared resource is intelligible to other members of the community. But how do we ensure intelligibility of material that results from a task that's not necessarily accessible in time (community memory is usually an asynchronous form of communication) or place (as we have seen, electronic communities are distributed groups)? Our past efforts have focused on two different tactics to make shared spaces mutually intelligible: meta-discussions within a space [10, 12] to discuss the materials it contains and shared representations that structure and organize the materials [17, 14, 25]. Yet the problem becomes much harder to solve as the community memory grows in size; rationale for the content and structure of the shared resource becomes opaque and inaccessible over time.

Realistically, some portion of emerging structure (and structure is continually emerging) will always be implicit. In systems to support collaborative intellectual work like NoteCards and VIKI, the strategy to achieve mutual intelligibility has been to encourage contributors to explicitly record discussions about the work.

NoteCards is a hypertext-based information-organizing tool originally intended for individual use, but once a user community emerged, it became apparent that many tasks people were performing using the tool -- writing papers, managing projects, collecting and analyzing information -- were in fact group activities [29]. As a result of this observation, NoteCards developers added facilities to support collaborative work [10]. Three of the more important facilities were: History Cards, tailored event-centered record keeping that could be annotated by collaborators; Guided Tours, a technique that allowed a presentation structure to be overlaid on a hypertext network; and TableTops, a means of contextualizing work by allowing a number of cards to be grouped as a visual composite [31].

Each of these mechanisms involved a semi-automatic way of recording changes or state (for example, TableTops recorded which cards were together on the screen, including scrolling). What we learned is that these recordings of paths, process, or state need to be supplemented by human annotation [12]. This need for human annotation (or communication *about* a group information resource) has been confirmed by our experiences with subsequent systems. VIKI provided no explicit mechanisms for recording change history, so collaborators developed conventions (electronic post-its) for communicating their changes to each other.

There is some perception that support for strong typing of materials will naturally bring about coherence; we have not found this to be the case. Aquanet relied on domain schemas to make contributions self-organizing and self-documenting, thereby rendering them intelligible to other group members; if one contributor creates, for example, a claim as part of an argument, the contribution's type (along with the role it plays in a community-defined structure) would allow other group members to interpret it. This strategy is based on two important assumptions: (1) people understand the meaning of the abstract schematic description and use it in a uniform way and (2) people fully use the schematic structures, and leave little implicit. In practice, neither of these assumptions has been found to hold. Collaborators still found themselves discussing the abstractions and how they ought to be applied. They also left a great deal implicit (including why a particular element should occupy a particular position in the shared space), thereby introducing a great deal of ambiguity and inconsistency.

Structure recognition and incremental formalization techniques may be useful in finding implicit structure and making it available for discussion within a community of practice; the implicit structure does not need to be declared, only located. Specific support for conversations about recognizable implicit structures may help members of a community keep their own contributions to the shared resource intelligible.

2.3.2 Situatedness and task specificity: Representational fluidity.

From our earlier discussion of acquisition, it is clear we are assuming that contributions to community memory must be well-connected to the task at hand. Thus we must call into question the applicability of the material, representations, and structure that is constructed in service of one intellectual activity to another. There are two sides to reuse of materials in community memory: the ability to reuse the materials themselves (through generalization and reapplication) and the ability to reuse the abstractions that structure these materials. We look first at techniques for generalizing the materials themselves.

Generalization is a process in which details are removed and the resulting information is, in part, abstracted from its original context so that it may be applied to other situations. Generalizations are created with an expectation of future use. Different generalizations will be appropriate for different future situations. For example, in our experience with network design, the same design can be used as an example in situations with similar budgetary considerations and in situations using similar technology [6]. Providing fluid representations, where information can flow in both structure and use, can facilitate such generalizations. For example, XNetwork allowed designers to continually add and remove structure from the representation of the design and to make copies of the design available as more general examples within the community memory.

Since contributors cannot completely predict the situation of their audience, it is difficult to know how much background to provide to make their interpretations and knowledge useful at a later date [1]. As a result, not only does the knowledge itself need to be generalized; it is also important to record the context in which the materials constituting community memory were created.

We now turn our attention to the abstractions used to structure the materials -- the meta-schematic descriptions of domains of interest. One of the original motivations for providing this kind of abstraction is the ability to reapply it to interpret related materials. We found this kind of reuse may be difficult to support with tools that do not acknowledge the fluidity of abstraction, since the structures people define are based on an idealization of the task and of the materials and may not fit well with the contingencies of the actual situation [28, 26].

For example, in our experiences performing a long term analysis task that involved assessing machine translation systems (see [16]), we found that the abstract types that highlighted certain appropriate aspects of the systems (like the approach they took to translation of natural language) were not entirely appropriate for a seemingly similar task of identifying candidate Spanish-English translation software for purchase. The new task required that aspects like cost and hardware platform be made perspicuous. In general, fixed representations of domain structure tend to cause material that doesn't quite fit into the abstractions to get lost, to drop from sight. This problem with the application of abstractions would surely be amplified as a community memory grew and encompassed more materials and more related tasks.

We addressed this problem in VIKI by assuming that representations are fluid, lightweight, and locally-defined for the task. In the case of a community memory, interpretive abstractions would need to part of a *view* of the underlying materials rather than a property of the materials themselves.

3. Conclusions

We see community memory as the set of shared materials, understandings, and conventions that emerge within a community, a way of culling and bringing context and perspective to large scale information resources. It is a critical element in the evolving picture of what it means to make these resources effective within communities of practice.

In this paper, we have discussed issues from three different perspectives:

How community memory is *acquired* from contributors;

How people *locate* the information they need in a community memory;

How community memory is *understood* by its diverse members.

Each of these perspectives raises its own attendant set of issues. Acquisition raises three issues that in many ways stem from the blurring of roles among readers, writers, librarians, and publishers. First, structure is rarely in place from the start. Instead it emerges over time and through use of materials in a task setting. We pose incremental formalization and representational flexibility to address these characteristics. Second, community-directed resources must be maintained. Support for the process of seeding, evolutionary growth, and reseeding helps address maintenance issues. Finally, distributed communities use distributed, heterogeneous resources. Any community memory must provide connectivity and an open architecture.

Location evokes two sets of issues. The first arises from the simple observation that practitioners often times don't know what they don't know. This issue may be addressed by techniques for the automated volunteering of material from community memory. The second issue, finding high-quality information (beyond immediate exigencies met by information retrieval), may find a solution in explicit support for community-mediated location. In effect, librarianship is distributed among members of the community.

Finally, problems of comprehension, mutual intelligibility, and applying materials collected in service of one task to another can be addressed through better support for human-annotated records of process, state, or path and by expected fluid, high-situated representational forms.

By applying this collection of techniques, the emerging wealth of large-scale information resources can be put to work by groups and communities, and truly make a difference in the way we conduct our day-to-day lives.

References

[1] Ackerman, M.S. "Definitional and Contextual Issues in Organizational and Group Memories," *Proceedings of the Twenty-seventh Hawaii International Conference on System Sciences (HICSS'94)*, 1994.

[2] Berlin, L., Jeffries, R., O'Day, V.L., Paepcke, A., and Wharton, C. "Where Did You Put It? Issues in the Design and Use of a Group Memory," *Proceedings of InterCHI `93*, (Amsterdam, The Netherlands, Apr. 24-29). 1993, pp. 23-30.

[3] Conklin, E.J., and Yakemovic, K.C. "A Process-Oriented Approach to Design Rationale," *Human Computer Interaction* 6, 3-4 (1991), pp. 357-391.

[4] Engelbart, D.C., and English, W.K. "A Research Center for Augmenting Human Intellect," *Proceedings of the Fall Joint Computer Conference*, 1968, pp. 395-410.

[5] Fischer, G., Lemke, A., McCall, R., and Morch, A. "Making Argumentation Serve Design," *Human Computer Interaction* 6, 3-4 (1991), pp. 393-419.

[6] Fischer, G., McCall, R., Ostwald J., Reeves, B., and Shipman, F. "Seeding, Evolutionary Growth, and Reseeding: Supporting the Incremental Development of Design Environments," *Proceedings of CHI `94*, (Boston, Mass.) 1994, pp. 292-298.

[7] Goldberg, D., Nichols, D., Oki, B.M., and Terry, D. “Using Collaborative Filtering to Weave an Information Tapestry,” *Communications of the ACM*, 35, 12 (December 1992), pp. 61-70.

[8] Gorry, G.A., Burger, A., Chaney, J., Long, K., and Tausk, C. “Computer Support for Biomedical Research Groups,” *Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW’88)* (Portland, Oregon, Sept. 26-28). ACM, New York, 1988, pp. 39-51.

[9] Grudin, J. “Why CSCW Applications Fail: Problems in the Design and Evaluation of Organizational Interfaces,” *Proceedings of the Conference on Computer-Supported Cooperative Work (CSCW’88)* (Portland, Oregon, Sept. 26-28). ACM, New York, 1988, pp. 85-93.

[10] Irish, P.M., and Trigg, R.H. “Supporting Collaboration in Hypermedia: Issues and Experiences,” *Journal of the American Society for Information Science*, March, 1989.

[11] Levy, D. and Marshall C.C., “What Color was George Washington’s White Horse? A Look at the Assumptions Underlying Digital Libraries,” *Proceedings of Digital Libraries `94*, (College Station, Texas, June 19-21). 1994.

[12] Marshall C.C. and Irish P., “Guided Tours and On-Line Presentations: How Authors Make Existing Hypertext Intelligible for Readers,” *Hypertext `89 Proceedings*, (Pittsburgh, Penn., Nov. 5-8). 1989.

[13] Marshall, C.C. “Work Practice Study: Analysts and Notetaking,” Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304.

[14] Marshall, C.C., Halasz, F., Rogers, R., and Janssen, W. “Aquanet: a hypertext tool to hold your knowledge in place,” *Proceedings of Hypertext `91* (San Antonio, Texas, Dec. 15-18). ACM, New York, 1991, pp. 261-275.

[15] Marshall, C.C., Shipman, F.M., and Coombs, J.H., “VIKI: Spatial Hypertext Supporting Emergent Structure,” *Proceedings of the 1994 European Conference on Hypermedia Technologies* (Edinburgh, Scotland, Sept. 18-23). 1994, pp. 13-23.

[16] Marshall, C.C., and Rogers, R.A. “Two Years before the Mist: Experiences with Aquanet,” *Proceedings of European Conference on Hypertext (ECHT `92)* (Milano, Italy, Dec. 1992). pp. 53-62.

[17] McCall, R. “On the structure and use of issue systems in design,” Doctoral Dissertation (1978) University of California, Berkeley, University Microfilms (1979).

[18] McCall, R., Schaab, B., and Schuler, W. “An Information Station for the Problem Solver: System Concepts,” *Applications of Mini- and Microcomputers in Information, Documentation and Libraries*, C. Keren, L. Perlmutter, Eds. New York: Elsevier, 1983.

[19] McCall, R., Bennett, P., d'Oronzio, P., Ostwald, J., Shipman, F., and Wallace, N. "PHIDIAS: Integrating CAD Graphics into Dynamic Hypertext," *Proceedings of the European Conference on Hypertext (ECHT'90)* (Paris, France, Nov.). 1990, pp. 152-165.

[20] McKnight, C., Meadows, J., Pullinger, D., and Rowland, F. "ELVYN -- Publisher and Library Working Towards the Electronic Distribution and Use of Journals," *Proceedings of Digital Libraries '94*, (College Station, Texas, June 19-21, 1994), pp. 6-11.

[21] Reeves, B.N., and Shipman, F.M. "Supporting Communication between Designers with Artifact-Centered Evolving Information Spaces," *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW `92)* (Toronto, Canada, Oct. 31-Nov. 4). ACM, New York, 1992, pp. 394-401.

[22] Russell, D.M., Stefik, M.J., Pirollli, P., and Card, S.K. "The Cost Structure of Sensemaking," *Proceedings of InterCHI `93* (Amsterdam, Netherlands). 1993, pp. 269-276.

[23] Schnase, J.L., Leggett, J.J., Furuta, R.K., and Metcalfe, T. (Editors) *Proceedings of Digital Libraries '94*, (College Station, Texas, June 19-21, 1994).

[24] Shipman, F.M., Chaney, R.J., and Gorry, G.A. "Distributed Hypertext for Collaborative Research: The Virtual Notebook System," *Proceedings of Hypertext `89* (Pittsburgh, Penn., Nov. 5-8). ACM, New York, 1989, pp. 129-135.

[25] Shipman, F.M. "Supporting Knowledge-Base Evolution with Incremental Formalization," Doctoral Dissertation, Department of Computer Science, University of Colorado, Boulder, 1993.

[26] Shipman, F.M., and Marshall, C.C. "Formality Considered Harmful: Experiences, Emerging Themes, and Directions," Xerox PARC Technical Report ISTL-CSA-94-08-02, 1994.

[27] Shipman, F.M., and McCall, R. "Supporting Knowledge-Base Evolution with Incremental Formalization," *Proceedings of CHI'94* (Boston, Mass., Apr. 24-28). 1994, pp. 285-291.

[28] Suchman, L.A. *Plans and Situated Actions: The problem of human-machine communication*. Cambridge University Press, Cambridge, UK, 1987.

[29] Suchman, L., R. Trigg, and F. Halasz. "Supporting Collaboration in NoteCards," In D. Marca and G. Bock (Eds.) *Groupware: Software for Computer-Supported Cooperative Work*, Los Alamitos, CA: IEEE Computer Society Press, 1992; pp. 394-403.

[30] Terveen, L., Selfridge, P.G., Long, M. D. "From `Folklore' to `Living Design Memory'," *Proceedings of InterCHI `93*, Amsterdam, The Netherlands, (24-29 April 1993), pp. 15-22.

[31] Trigg, R. "Guided Tours and Tabletops: Tools for Communicating in a Hypertext Environment," *CSCW `88 Proceedings*, Portland, Oregon, (September 26-28, 1988), pp. 216-226.