Annotate: A Web-based Knowledge Management Support System for Document Collections

Mark Ginsburg
Fisher Center
Haas School of Business
University of California, Berkeley
Berkeley, CA 94115
ginsburg@haas.berkeley.edu

Ajit Kambil
Information Systems Department
Stern School of Business
New York University
New York, NY 10012
akambil@stern.nyu.edu

Abstract

Knowledge management is an increasingly important source of competitive advantage for organizations. Knowledge is a renewable, re-usable and accumulating asset of value to firms that increases in value with employee experience and organizational life. Knowledge embedded in the organization's business processes or the employee's skills are assets are generally hard to discern, accumulate and replicate by competitors. It provides the firm with unique capabilities or "resources" to deliver customers with a product or service. In contrast, as we undertake electronic commerce, customer interfaces and business strategies generally become more visible to competitors. Thus, the organizations capacity to effectively accumulate and leverage knowledge assets better than its competitors becomes a key source of competitive differentiation.

As firms become more knowledge intensive, more effort is being expended on knowledge management (KM). While much progress has been made on designing IS to support decision making, the art and design of KM systems to preserve, index, formalize and leverage knowledge in organizations is still new (see O'Leary [1] for a review of best practices). Knowledge is fundamentally more complex than information or data, and systems supporting knowledge management have a broader range of design issues.

This paper reviews approaches to knowledge management support systems (KMSS) and proposes the need to design systems that carefully map their features to target organizations and user groups. We illustrate Annotate *as a specific KMSS designed to support the knowledge management of document collections in federated organizations which lack common vocabularies and central authority.*

This work has been supported in part by the Fisher Center for Management of Information Technology, Prof. Arie Segev, Director, and by NSF Grant No. 9319331, Prof. Ajit Kambil, co-principal investigator.

1. Knowledge Management Support Systems

Knowledge management support systems require new design principles because knowledge fundamentally differs from information and data in organizations. Knowledge is an organizational member's experience and values combined with and shaped by the information contained in various systems and data provided to the person [2] [3]. It is intrinsic to organizational members and focuses on the information recipient. In contrast, data refers to a set of discrete, objective facts about events recorded in an organization and information provides organizational members with contextual meaning to the data.

Knowledge can be tacit or explicit [3]. Tacit knowledge is the beliefs and values that are hard to express but inferred from the behaviors of organizational members. Explicit knowledge is easily expressible such as the formalization of an organizational routine or process through a flow diagram. Organizational and individual knowledge is created through a continuous dialogue between the tacit and explicit knowledge of individuals. Ideas are formed in the minds of individuals, but interaction between individuals typically plays a critical role in developing these ideas. Nonaka [3] identifies four knowledge transformation processes: socialization (tacit to tacit), internalization (explicit to tacit), externalization, (tacit to explicit), and combination (explicit to explicit) [3].

While new knowledge is developed by individuals, organizations play a critical role in articulating and amplifying that knowledge [2]. This requires organizations to provide a working infrastructure, composed of a set of knowledge management support systems (KMSS), and meaningful policies for knowledge sharing. Such an infrastructure would allow users to easily share information, with policies that provide incentives to organizational members to participate in knowledge sharing and refinement activities. The information shared among members should reflect their values and beliefs about the information stored and exchanged

to support KM.

As KMSS are embedded within an organizational system they must also be designed to fit within the cultural values, authority structures and other design features of the organization. Thus knowledge management consists of both the implementation of information systems and organizational systems with incentives, processes, and tasks to collectively generate, refine and manage organizational knowledge. As IS systems increasingly support KM we denote systems supporting KM as KMSS to note that an information system is only a support tool in an overall organizational KM system.

The ideal knowledge network as conceptualized by Nonaka assumes efficient search and retrieval of an abstract knowledge base; however, he does not indicate design approaches which would bring about this efficiency. This paper introduces the *Annotate* system to address the problem of design of an enhanced retrieval software tool for retrieval on un- or semi-structured document archives. The *Annotate* system captures user histories in a typical search session and, through the annotation facility, leverages the collective intelligence of the search community.

The remainder of the paper reviews critical issues for the design of KMSS in Section 2 and moves onto KMSS challenges, both technical and organizational, in Section 3 . Section 4 presents the *Annotate* system, our software tool to explore KM in the domain of Web-based document archive structure and retrieval. Section 5 presents first a technical review of the two fundamental data structures underlying *Annotate*, the *discussion* and the *session* data and discusses the organizational implications of this architecture. Section 6 discusses an ongoing field experiment with the *Annotate* system with special emphasis on incentive considerations. In Section 7 we discuss lessons learned from this project and provide concluding remarks.

2. Critical Issues in the Design of KMSS

Despite the widespread interest on KM in general, there has been surprisingly little work on what might constitute an effective KMSS and the tradeoffs an organization might face in achieving its KMSS goals. For example, KMSS systems often have some or all of the following components [1]:

- A Data or Knowledge warehouse — as the organization ages and continues to store transaction data in the warehouse, the costs to ensure efficient retrieval on the data store may increase sharply.
- Knowledge search and discovery mechanisms which are a particularly difficult problem in the case of multimedia, for example streaming audio and video.
- Knowledge representation via an ontology which presents a significant tradeoff. If an organization imposes an ontology on a series of document collections, there is the possibility of vocabulary conflict across business units. As Pejtersen notes [4], there is a significant cost associated with forming classification schemes which cover the organization's various work domains. For example, GrapeVine (a vendor product) uses a 'knowledge chart' to list topics of interest to end-users. This chart is manually set up by domain expert(s) beforehand, and in addition there is another setup cost: experts are designated for each topic which sort documents in legacy databases such as Lotus Notes and classify them according to importance. Users set up profiles on the topic and importance dimension and thereafter documents are broadcast to matched users via e-mail.
- Knowledge quality control is an important organizational goal; for example, establishing a minimum level of credibility for entries in a given knowledge base.
- Knowledge visualization techniques are a burgeoning field. For example, Phelps and Wilensky [5] have been researching Java applets at the client side to improve the presentation of documents (separating them into text, scanned OCR pages, and other layers).

These components have to be integrated into a system that provides the functionality in the previous sections and maps to organizational requirements. The integration is done through the human organization and processes that overlay a KMSS. The integration works best when the KMSS features fit well with the organization structure, processes and values.

Without effective retrieval, information islands in a federated organization do not diffuse well across intra-organizational boundaries. Hence, knowledge transfer is limited in any structure with sub-optimal retrieval facilities. For example, if an organization relies on Lotus Domino to search its Web document archives, only Notes databases will be consulted: other Web servers (Unix, NT, Mac) will be not be included[1].

3. Challenges of Designing a Web-based Document KMSS

There are both technical and organizational factors which impact the design of a Web-based document KMSS. In this section, we review the key properties of documents in a Web development environment and discuss key features of the organizational document publishing process that we must keep in mind when designing the KMSS.

3.1. Documents as Web Knowledge Bases

In contrast to well-structured fielded database, unstructured or semi-structured (template-based) documents represent an increasingly important part of organizational knowledge bases. Documents have the potential to be highly expressive, with embedded multimedia objects. While ex-

[1] The *Annotate* system had to solve the important subproblem of integrated search across Unix and Domino servers in the field test described in Section 6 . Federalist firms are especially likely to require an integrated search mechanism across multiple server platforms; for technical details see [6].

pressive and strong in presentational markup (rendering) they are often poor in semantic markup making knowledge search and discovery difficult.

The Web facilitates distributed document publishing by virtue of its open HTTP protocol [7], however professional document work products typically incur a high cost of creation in time and effort.

Another problem of documents in a Web environment is inadequate support of the HTTP protocol for shared editing. There is active research studying a tighter integration of the Web with groupware systems. For example, Bentley et al.'s BSCW system [8] helps users manage projects and perform document workflow, much as Lotus Notes does. In addition, since the inception of the Web there has been interest to help users share documents. In a recounting of the origins of the Web, Robert Caillau tells us in an interview that: "The first [CERN web editor/browser made] no distinction between editing mode and browser mode. We lost all that along the way" [9]. In 1994, Frivold et al. modified the NCSA Mosaic client to support shared edits [10]. Other early annotation facilities were described by Röscheisen et al. [11]; more recently Salcedo and Decouchant [12] delineated an architecture to to enable workgroups to edit documents collaboratively at the sub-document (paragraph, sentence) level. The *Annotate* system does not make use of these advanced in-place edit or workflow features; rather, its lightweight annotation facility leaves the core document untouched.

3.2. Document 'Marketing' on the Web

Document repositories which span multiple intranet Web servers pose a marketing problem. With the advent of low-cost WWW publishing, it is quite easy to place a document on a given intranet server. It is quite another matter altogether to to let other business units know that a new document repository exists, or that interesting new documents exist on a server that another business unit may not consult very often.

The Web moves the firm to a peer information model, where clients can easily access servers throughout the intranet. Intranets in federalist organizations (those with semi-autonomous business units) [13] face practical difficulties. If each business unit maintains its own intranet server, a given business unit may become used to searching only its own server. How to increase the scope of the search so that functional overlaps between business units might be exploited? Note that the increased scope means that there is greater information throughput (and consequently, greater potential for knowledge gain) in the aggregate.

3.3. Pre-Coordinate Ontology vs. Post-Coordinate Full Text Search

Document indexing and search can be implemented through pre-coordination or post coordination. In pre-coordination, the documents are associated with subject headers by a collection administrator. The subject headers follow a standard order, for example *Mexico — Economy — Inflation*. Post-coordination is so named because the keywords are combined at search time; there is no subject term taxonomy specified *a priori*.

Pre-coordination implements a centralized ontology, but the effort to set up an ontology and classify documents is manually intensive. As a knowledge base grows, it becomes difficult and expensive to create ontologies and reconcile classifications to suit the interests of many different users. This problem is compounded as the interests of the knowledge/information seekers increase and diverge. Any KMSS system which implement static ontologies for classification and selection of control vocabularies must face this issue.

If an organization decides to map documents in heterogeneous databases to knowledge structure, as described in the Andersen consulting case [1], the maps themselves are susceptible to political processes, often hiding controversial areas and thus limiting the total amount of information available [2].

Organizations usually resort to post-coordination or full-text search and impose no vocabulary control. In standard Web-based full text search, we encounter problems such as homonymy, where words mean different things in different contexts, lowering precision and synonymy: search engines that lack a good thesaurus will artificially deflate the confidence scores of documents containing synonyms to the keywords.

3.4. Organizational KMSS design challenges

In addition to technical challenges organizations often lack adequate incentives for knowledge sharing and management. These difficulties are often exacerbated in emerging federalist organizations which are dynamic, team based problem solving structures with distributed authority. The first decision business units make is the choice of specific groupware products, such as Notes (Domino) or intranet product suites [14]; the broader issue is how to organize the documents accessed by the groupware product to facilitate knowledge transfer.

As a result it is not surprising that most systems in the past have covered limited domains (see the O'Leary examples [1].) As document publishing is simplified, and Intranets link individuals in organizations to rapidly expanding Web document bases, the previous problems in the design and maintenance of KMSS become more pronounced.

4. *Annotate* : A Web-based Document KMSS

Typical Web Full Text Search (WFTS) engines which provide post-coordinate search have deficiencies which translate into inadequate support for KM. For example, there is no way to share resource discovery made during the course of an ad-hoc search session for one's future use

or between users. There are also extremely limited data and metadata clues to assist the user as he or she traverses the system from the front-end (the Query Layer) to the intermediate layer, which is an array of hyperlinks to the core documents (the Retrieval Layer) and on to the bottom layer, the Document Layer. In a typical implementation, the user has no knowledge of others' prior searches or results at the Query front-end and has very few clues of what the most interesting documents might be at the Retrieval layer.

To address some of these problems we have developed *Annotate* which provides a flexible KMSS to support federated document search and retrieval. Motivation for *Annotate* also comes from the CSCW field. *Annotate* is first and foremost a collaborative tool to help groups work together. As such, it stands rooted in the CSCW community. As Bentley et al. and Bowers note, we need to move out of the laboratory when studying technologies with enable collaboration [8] [15] and into the field. Thus we evaluate *Annotate* in a field experiment, realizing that work that appears to promise great gains may run into severe problems in their situations and use [15].

Increasing the data and metadata clues at each layer of the search navigation chain will aid in search efficiency and aid in the process of knowledge discovery, thereby realizing more of the knowledge potential of document archives. wherever they may be found in the corporate intranet. This is similar to the metadata focus in recent electronic document management work. For example, in the document management system described by Balasubramanian and Bashian [16], document metadata (attributes bestowed by a legacy document management application) are captured and used by the web search as a attribute-value search alternate mechanism to full text search.

One of the driving factors behind the *Annotate* design is to enable Nonaka's knowledge management processes of socialization, internalization, externalization and combination by:

- capturing individual and aggregate document appraisals (a means to aggregate individuals' externalization, or use of metaphor to express others' tacit knowledge),
- using individual appraisals of documents to augment document content (to support readers' internalization on an ongoing basis)
- using individual appraisals of documents to support a recommender system (which improves the efficiency of the search by filtering out unwanted documents, for example those from an untrusted domain),
- using individuals' free text annotations to support combination or the reshaping of information and data from one information system to another,
- and using the free text annotations as well to weakly support socialization or the transfer of tacit knowledge from one individual to another.

Annotate is predicated on the principle that the users and creators of knowledge best know the information relevant to their knowledge management task and that they can more effectively filter, discover and signal useful knowledge to their peers than an automatic system.

As annotations accrue in the system, so do the reasons the annotator had for making the note. Both the annotation text and its rationale (metadata) are logically bound to the core documents, thus increasing the semantic content of the document repositories. *Annotate* implements a *star* structure: "for each document, there is only one level of annotations — annotations of annotations are not possible. Stars are simpler for users in some ways because one can read through all unread annotations in a sequence. Since new annotations are always appended to the end of the list, one knows that readers are seeing the same thing, and thus the conversational style of communication is well modeled" [17]. The alternative, a *tree* structure where annotations can be made to annotations, diminishes the distinction between the (lengthier) core document and the brief secondary annotation — we wish to make the annotation process simple and limit the length of the annotation entity while keeping attention on the core document which had a greater social cost to produce. The tree structure works well where the core document is also a brief note, such as LaLiberte's HyperNews system (hosted at *http://www.hypernews.org/*) for Web-based threaded Usenet-style discussions. Note that the star structure does not preclude annotation post-processing to create more than one level of annotation. For example, one dynamic extension to *Annotate* is to scan annotations in batch for common hyperlinks or keywords and create inter-annotation links at regular intervals.

4.1. A Tour of the *Annotate* System

In a typical user session, the Query interface as shown in Figure 1 resembles that of a standard Web Full Text Search Engine, for example, Alta Vista, or Excite. The user enters keyword(s) to reach the Retrieval interface, a set of hyperlinks to base documents. There is an enhancement to standard full text search shown in this figure: the first is the ability to filter the result set by annotation domain or to set a minimum aggregate quality rating. Shown is a cutoff of 6.0 (out of 7.0) as a Quality Filter considering only the `FUND ADMIN` group annotations.

The result of the query is an array of hyperlinks — the Retrieval Interface. An excerpt of this screen is shown in Figure 2.

Reading from left to right in the first row (see Figure 3 for the icons and their meanings[2]), the leftmost four icons represent the most recent annotation for the Document *Interesting Servers on Related Subjects*. A user from D (domain) of `EQ` (Equities) assigns a Q (quality) rating of 3 (out of 7) and a F (factual accuracy) rating of 6 (out of 7). The

[2]The Likert 1—7 scale is mapped to a spectrum of facial expressions similar to Koda and Maes [18] in their interface agent usability study.

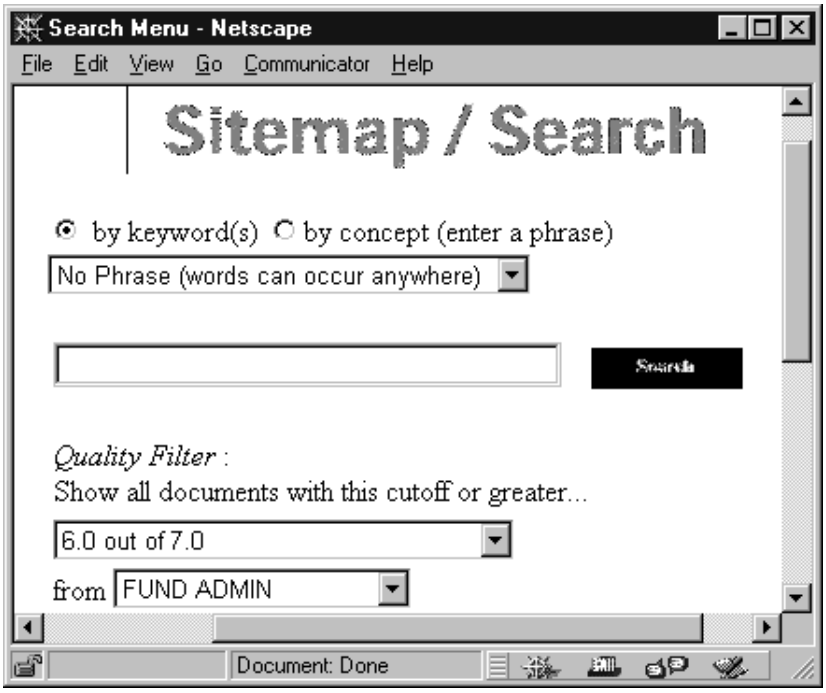

Fig. 1. *Annotate* Query Interface with a Social Filter Selected

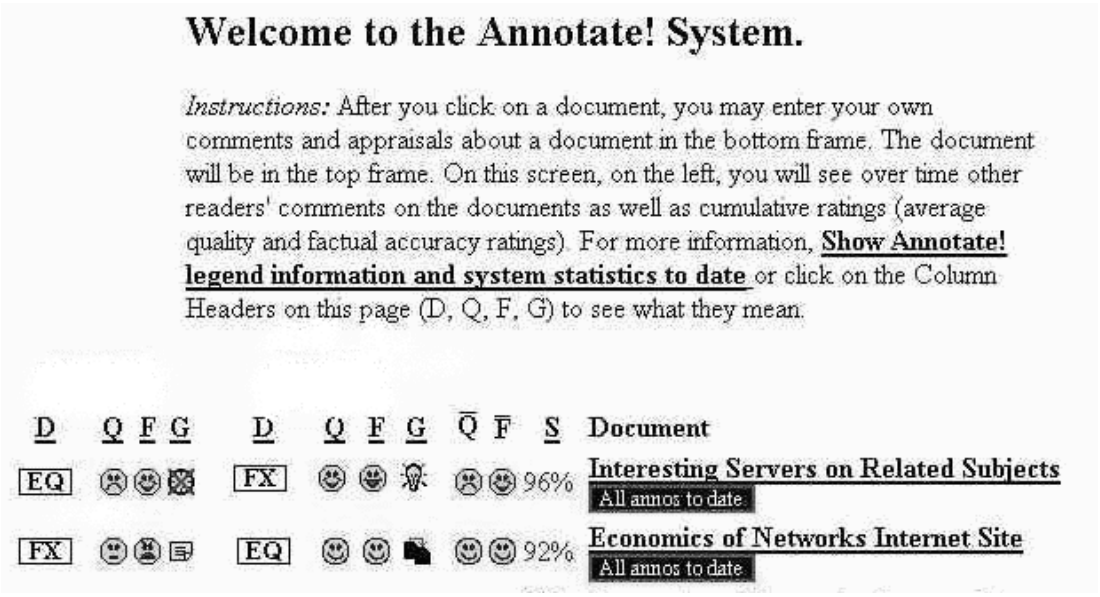

Fig. 2. Retrieval Layer Alterations in the *Annotate* System

reason for annotating is that the document was perceived to be out of date (the icon is a clock with an X through it). The next set of D, Q, F, G icons represent the next-most recent annotation. In this case, the FX group (Foreign Exchange) assigns $Q = 6$, $F = 7$, with a reason of "a more general lesson can be drawn" (the light bulb icon).

Figure 3 shows the complete range of possible icons that can be attached to the Retrieval Layer.

We then present aggregate statistics about the document annotations to date: $\bar{Q} = 3$ is the aggregate quality rating and $\bar{F} = 6$ is the aggregate factual accuracy rating. On the right we have the conventional Excite confidence score followed by a hyperlink to the core document. The confidence score is color-coded according to levels of annotation activity (blue — cold, green — medium, red — hot, with the exact cutoffs an implementation decision) to provide an extra metadata signal.

Thus *Annotate*, with its timely modifications of the retrieval interface, contributes to the organization's memory store which "concerns itself with usually recent events and outcomes within an organizational context and for organizational purposes" [19].

From the retrieval layer, users can also create a report on the most commonly queries keyword(s) to date, or the most heavily annotated documents to date, or the highest rated documents to date along several dimensions using the Session Data (cf. Section 5 .3).

The user selects a document hyperlink in Figure 2 to reach the Document Layer.

Figure 4 shows the user in the process of creating a new annotation; the form is kept simple in order to encourage participation in the system (i.e. lower the opportunity cost of annotating). Annotations grow the discussion data store and make add value to the document recommender system. As users contribute annotations over time, filtering can become a powerful mechanism to limit spurious results. This is depicted schematically in Figure 5.

To accomplish knowledge search and discovery mechanisms, documents and annotations are indexed using the Excite search engine. The user can search using keywords, and refine the search filtering on annotation variables. Knowledge quality control is a subjective process which is completely dependent on the user. Annotations can provide readers with rich data and opinions to aid belief development about the documents. Furthermore, readers who frequently contribute high-quality annotations can become opinion leaders and in a fully authenticated system will gain a sort of 'brand-name' recognition causing their notes to gain readership.

To aid in knowledge visualization, we use various small icons to denote appraisals and comments on the documents and convey them quickly to users.

5. *Annotate* System Architecture

Two key data stores, *discussion* data and *session* data, underly the *Annotate* system. In this section we describe these data structures and show how they relate to the interface layers a user encounters in a search session.

5.1. The Discussion Data Store

The discussion data store stores the appraisal ratings, the free-text annotations of specific documents, and the reason for annotating (document out of date, a 'see' reference to another document, and so on). The discussion store is a hybrid of the originally published ('core') document with zero or more annotations. The annotation Document Type Definition (DTD) is declared in a e**X**tensible **M**arkup **L**anguage (XML) Document Type Definition[3].

Figure 5 presents the high-level view of the relationship between the discussion instances and the information retrieval interface layers (Query, Retrieval, and Document). The Document layer is a receptacle to collect annotations. When an annotation event occurs, the discussion data store grows. This growth in turn may alter the look and feel of the Retrieval layer depending on simple trigger rules (refer to Section 4 .1 for a full description).

Annotations add value to existing data: a legacy HTML or ASCII document is now coupled with XML annotations adding value to the existing document base by increasing its overall semantics. This conforms to Maurer's recommendation that metadata be increased in legacy document stores to facilitate customized information access [20].

The annotations help users to refine their search by filtering on the annotation categories and enables collaborative (social) filtering as discussed theoretically by Avery and Zeckhauser [21]. The annotations help users to socialize through allowing asynchronous collaboration, internalize or combine knowledge through looking at user defined annotations that guide to other sources, and support externalization by providing a mechanism for expressing annotations. Anonymous but authenticated annotations identifying the authors' workgroup can increase trust within an organizational setting.

Finally, the discussion instances *leverage the weaknesses of conventional full text information retrieval*. The *Annotate* system anticipates that the results of full-text queries lack precision and are often spurious in the context of the original query. Annotations allow us to capture individuals' associative trails, and note interesting documents even if they did not match the original query; this has the potential to create new knowledge with subsequent system use.

[3] The DTD is available at *http://edgar.stern.nyu.edu/xml/discussion.html.* The DTD can be published on the network and revised for site-specific customization.

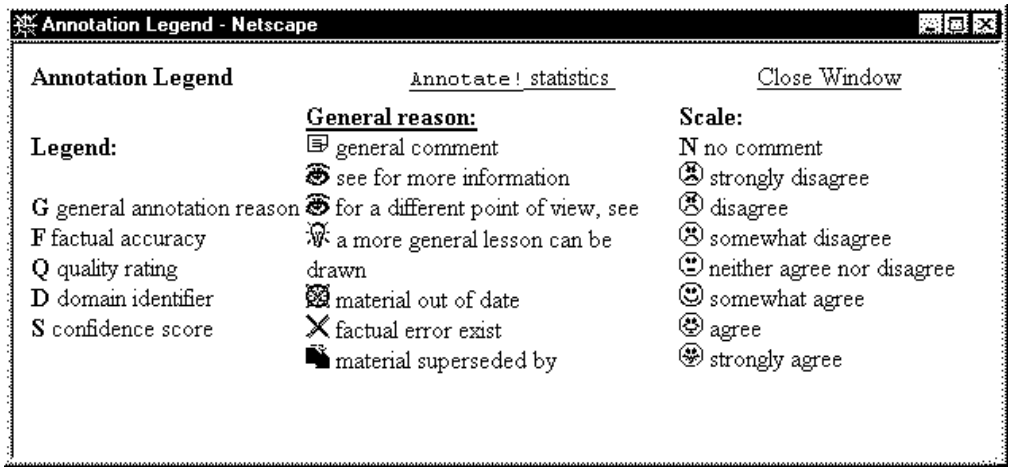

Fig. 3. The Icon Legend

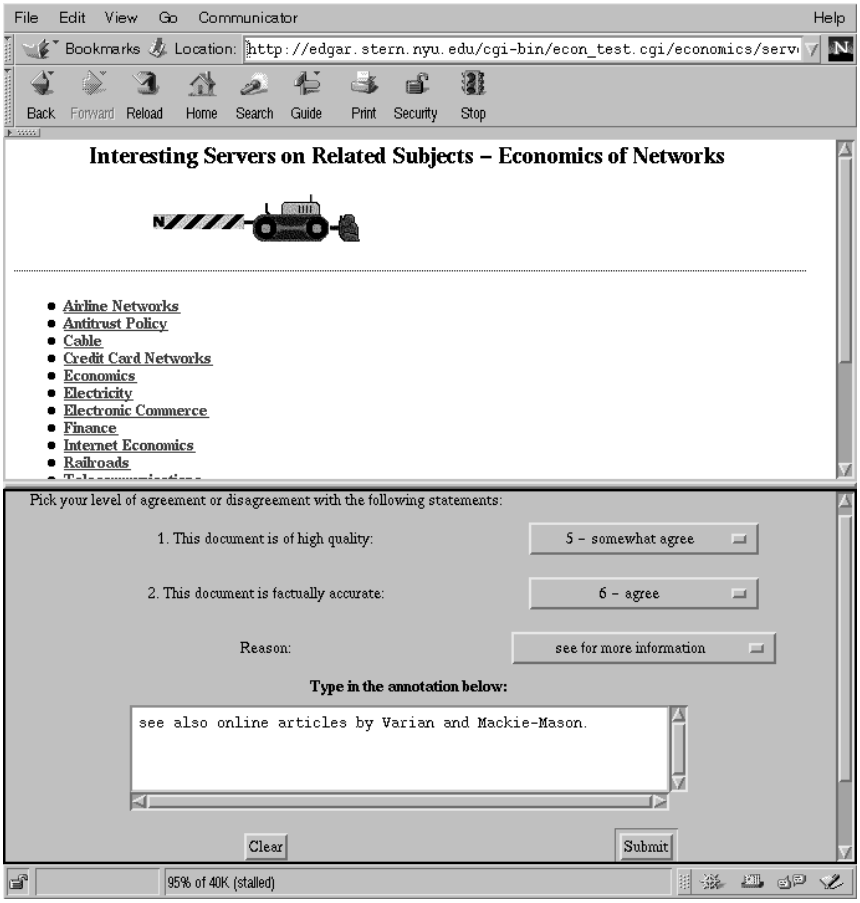

Fig. 4. Document Layer: Creating a New Annotation

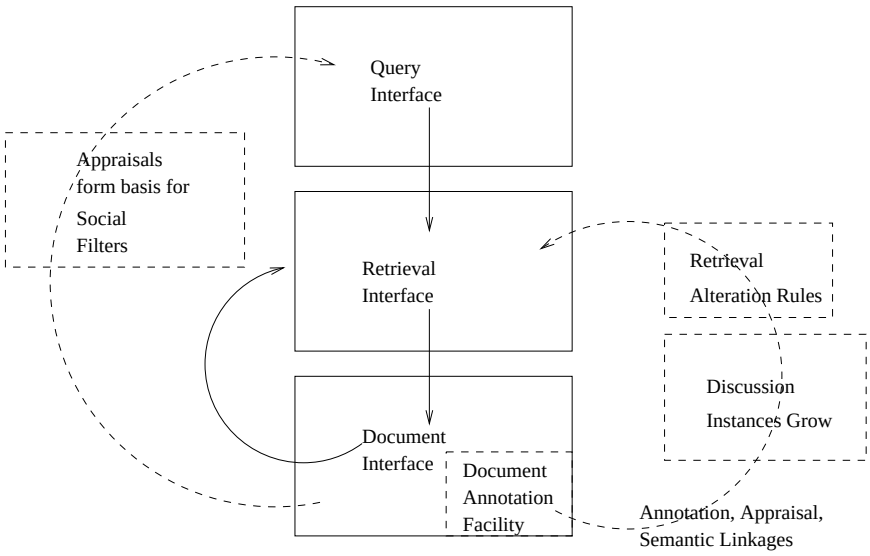

Fig. 5. The Discussion Data Store influences, or is influenced by, all of the *Annotate* Interface layers

5.2. Policies to Manage the Discussion Instances

To realize the benefits of discussion data, organizations need to have supportive policies for knowledge management. Three policy decisions regard:

- Incentives and rewards for adding annotations and conversely, sanctions for non-participation, may be chosen. Without an explicit incentive scheme, Orlikowski [22] demonstrated that Lotus Notes groupware was not well utilized at a management consulting company because its workers had little incentive to share information. The trade-off to supplying annotation is the cost (time and effort) of constructing the notes versus the value of becoming an opinion leader and/or distinguishing oneself from one's peer group.
- The level of anonymity of the annotator can be specified: anonymous, semi-anonymous (only the workgroup identified), or non-anonymous. Prior research highlights the importance of anonymity but says little about group identification. For example, the social issues of anonymity and annotation have been explored in the Group Support System (GSS) setting by Connolly, Jessup and Valacich [23] with the result that anonymous readers are more likely to offer critical remarks.
- Annotation controls may be imposed. It is possible to limit the annotator population to designated experts in a given subject. For example, the *Annotate* system can be extended to form a scholarly peer review system whereby domain experts annotate a draft manuscript.

5.3. The Session Data Store

The session data store keeps track of user queries, keywords, retrieval lists and the timings of the users' navigation through the document base. The relationship of the session instances to the search interface layers is shown in Figure 6.

As Figure 6 shows, the experimenter can use session data to evaluate *Annotate* in a field setting. We have written custom data analysis modules to perform more sophisticated tests of system usage; for example, in-depth analyses of document readership demographics and the times spent at the Query, Retrieval, and Document interface screens[4].

6. Evaluating the *Annotate* System

We are presently evaluating *Annotate* in an ongoing field trial at a federalist financial services firm. We are collecting system variables, such as user navigation timings, document readership demographics, and annotation statistics. We also collect qualitative data such as general user satisfaction measures and suitability of the system to the task at hand.

Thus far, few readers ($\leq$ 8 per week, data collected weekly, 8 data points collected thus far) are choosing to annotate in this setting. In each week's sample, there are between 5,000 and 7,000 Intranet accesses and between 35

[4] If the user is inactive at the document layer for a period of time, then returns to the same search session, the browsing time for that document will be inflated. More than one distribution can be expected in this data. Technical details are available in [24].

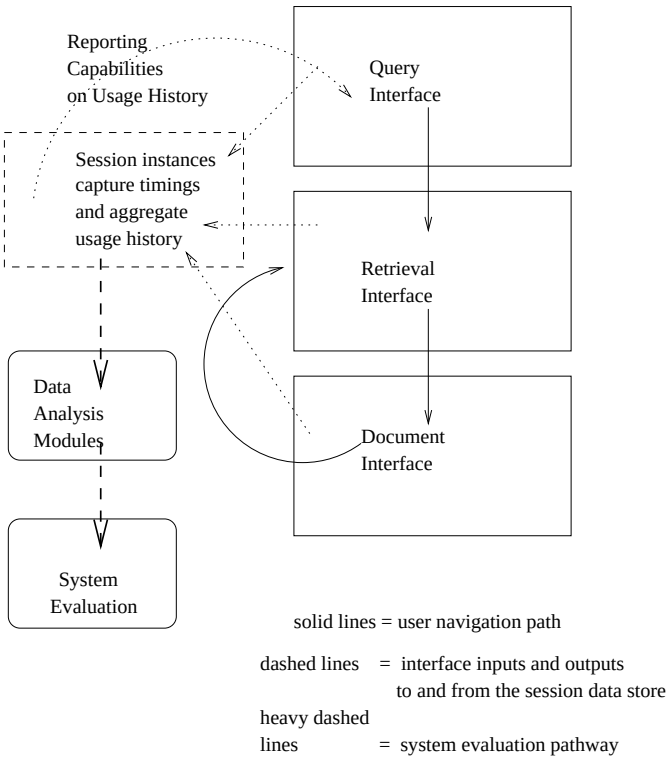

Fig. 6. Session Instances Provide Interface Timings and Usage History

and 60 web searches. This result can be thought of graphically by the Schelling Diagram shown in Figure 7.

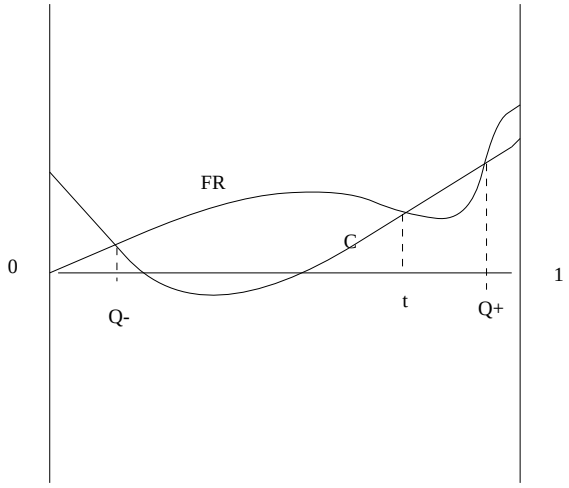

Fig. 7. Incentives and Social Considerations Alter the Free Ridership and Cooperation Curves

In this figure, the incentives facing the marginal chooser are shown contingent on all others' choice. Going from the left (0, or no cooperation) to the right (1, or full cooperation) we consider the marginal choice. The vertical distance from the horizontal axis measures V, the value to the individual[25] for choosing Free Ridership (FR) or Cooperation (C).

There exist two equilibria in Figure 7. The inferior one, $Q-$, is reached when the cooperation level is shifted to the left of the *tipping point*, t. The superior equilibrium, $Q+$, is reached when one moves to the right of t.

The low levels of annotation activity are consistent with an inferior equilibrium point of $Q-$. Management has offered no explicit positive incentives to participate, and the promotional e-mail announcing the system merely solicited usage in general terms, which cannot be construed as a negative incentive (risk to non-participants).

Still, some users may attach social value to the act of annotating. For example, a user might want to help his or peers improve the quality of a work product and offer constructive feedback using the *Annotate* mechanism. Others may have private, self-interested reasons for annotating. Suppose an author wants to attract attention to his or her document. He or she might 'bootstrap' the system by adding some annotations, to make the document appear 'busy' (changing the Retrieval Interface to add annotation icons).

The social challenge therefore is how to move most effectively from $Q-$ to $Q+$. Keeping in mind that the organization must be able to *afford* the incentives (the cost must be lower than the gain expected) one possibility is

to recognize the most active secondary authors publicly in a broadcast e-mail. Another way to increase positive social perceptions of annotating is to broadcast the annotation event to the original author(s), thus accelerating information throughput. Again, this extension must be carefully mapped to organizational policies and norms.

7. Conclusions and Further Research

The WWW and Internet technologies enable new ways of implementing KMSS. *Annotate* provides one mechanism to support knowledge management in federated organizations focusing on documents as repositories of relevant information for knowledge creation and use. Federated organizational forms are becoming more prevalent in a knowledge economy. The WWW and Intranet facilitate distributed document publishing necessitating effective storage,retrieval and KM mechanisms. KMSS should be designed to fit the organization form and enable organizations to implement policies for effective knowledge sharing.

Annotations improve the overall semantics of Web document (ASCII or HTML) by declaring user values and beliefs formally about documents. *Annotate* in an Intranet increases *knowledge throughput* by increasing the flow of relevant information across business units. Even when users pursue documents irrelevant to the original query, the possibility of capturing subjective reactions will help in this regard. *Annotate* begins to instantiate Nonaka's ideal of the *knowledge network* through provision of recommendations and navigation assistance. Furthermore, by helping to increase the knowledge value of document repositories which span many business groups, *Annotate* is designed to increase the interoperability of federated document collections [26].

Ultimately, such a system's effective use will be predicated on organizational policies and choices users make to define their own ontology. As we apply this tool in organizational settings, our current research examines incentives, authentication, anonymity and the impact of other policy choices on system use and effectiveness. Specifically, we are modeling knowledge as a "collective organizational good", examining different levels or authentication, anonymity and policy choices on system use and effectiveness. Tools like *Annotate* enable us to easily collect data and study the diffusion and sharing of know-how in organizations through electronic means.

References

[1] Daniel E. O'Leary, "Enterprise knowledge management," *IEEE Computer*, vol. 31, no. 3, pp. 54–61, March 1998.

[2] Thomas H. Davenport and Laurence Prusak, *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, Boston, MA, 1998.

[3] Ikujiro Nonaka, "A dynamic theory of organizational knowledge creation," *Organization Science*, vol. 5, no. 1, pp. 14–37, 1994.

[4] Annelise Mark Pejtersen, "Semantic information retrieval," *Communications of the ACM*, vol. 41, no. 4, pp. 90–92, April 1998.

[5] Thomas A. Phelps and Robert Wilensky, "Toward active, extensible, networked documents: Multivalent architecture and applications," in *Proceedings of the 1st ACM International Conference on Digital Libraries*, Bethesda, MD, March 1996, pp. 100–108, ACM.

[6] Mark Ginsburg, *Annotate: A Web-based Knowledge Management Support System for Document Collections*, Ph.D. thesis, New York University, 1998.

[7] Carliss Y. Baldwin and Kim B. Clark, "Managing in an age of modularity," *Harvard Business Review*, pp. 84–93, September-October 1997.

[8] Richard Bentley, Thilo Horstmann, and Jonathan Trevor, "The World Wide Web as enabling technology for CSCW: The case of BSCW," *International Journal of CSCW Special Issue on CSCW and the Web*, vol. 6, no. 2-3, pp. 111–134, Spring 1997.

[9] Robert Caillau, "How it really happened," Tech. Rep., CERN, April 1998, Internet Computing Online.

[10] Thane J. Frivold, Ruth E. Lang, and Martin W. Fong, "Extending WWW for synchronous collaboration," in *Proceedings, 2th World Wide Web Conference: Mosaic and the Web*, Chicago, IL, 1994, NCSA, NCSA.

[11] Martin Röscheisen, Christian Mogensen, and Terry Winograd, "Interaction design for shared world-Wide Web annotations," in *CHI 95: Human Factors in Computing Systems*, New York, May 1995, pp. 328–329, ACM.

[12] Manuel Romero Salcedo and Dominique Decouchant, "Structured cooperative authoring for the World Wide Web," *International Journal of CSCW Special Issue on CSCW and the Web*, vol. 6, no. 2-3, pp. 157–174, Spring 1997.

[13] Jeanne W. Ross and John F. Rockart, "Enabling new organizational forms: A changing perspective on infrastructure," *Proceedings of the International Conference on Information Systems (ICIS)*, December 1996.

[14] Mark Ginsburg and Katherine Duliba, "Enterprise-level groupware choices: Evaluating lotus notes and intranet-based solutions," *CSCW: The Journal of Collaborative Computing*, vol. 6, pp. 201–225, 1997.

[15] John Bowers, "The work to make a network work: Studying CSCW in action," *Proceedings of the Conference on Computer-Supported Cooperative Work*, pp. 287–298, 1994.

[16] V. Balasubramanian and Alf Bashian, "Document management and Web technologies: Alice marries the mad hatter," *Communications of the ACM*, vol. 41, no. 7, pp. 107–115, July 1998.

[17] Daniel LaLiberte, "WWW collaboration projects," Tech. Rep., UIUC, http://www.hypernews.org/HyperNews/get/www/annotations.html, 1998.

[18] Tomoko Koda and Pattie Maes, "Agents with faces: The effects of personification of agents," in *Proceedings of HCI'96*, London, 1996.

[19] Mark Ackerman, *Answer Garden: A Tool for Growing Organizational Memory*, Ph.D. thesis, Massachusetts Institute of Technology, 1994.

[20] Hermann Maurer, "Web-based knowledge management," *IEEE Computer*, pp. 122–123, March 1998.

[21] Christopher Avery and Richard Zeckhauser, "Recommender systems for evaluating computer messages," *Communications of the ACM*, vol. 40, no. 3, pp. 88–89, March 1997.

[22] Wanda Orlikowski, "Learning from Notes: Organizational issues in groupware implementation," *Proceedings of CSCW 1992*, 1992.

[23] Terry Connolly, Leonard M. Jessup, and Joseph S. Valacich, "Effects of anonymity and evaluative tone on idea generation in computer-mediated groups," *Management Science*, vol. 36, no. 6, pp. 689–703, 1990.

[24] Mark Ginsburg, "Annotate! a tool for collaborative information retrieval," in *Proceedings, WETICE98 Conference*, Palo Alto, CA, June 1998, vol. forthcoming, IEEE.

[25] Howard Margolis, *Strategy and Choice*, chapter Free Riding versus Cooperation, MIT Press, 1991.

[26] Andreas Paepcke, Chen-Chuan K. Chang, Héctor García-Molina, and Terry Winograd, "Interoperability for digital libraries worldwide," *Communications of the ACM*, vol. 41, no. 4, pp. 33–43, April 1998.