Exploitation of Digital Artifacts and Interactions to Enable Peer-to-Peer (P2P) Knowledge Management (KM)

Mark T. Maybury
The MITRE Corporation
202 Burlington Road
Bedford, MA 01730
maybury@mitre.org

Abstract

This paper surveys our efforts in the automated analysis of human created digital artifacts and human computer interactions to enable peer-to-peer (P2P) knowledge management (KM). We begin outlining the relationships among knowledge management, peer-to-peer computing, collaboration, and human language technology. We first discuss tools to support peer and group knowledge discovery, exemplifying these in the domains of global infectious disease management (TIDES) and global social indicator analysis (SIAM). Next we describe automated tools for profiling individual and collective expertise (Expert Finder) as well as organizational knowledge interactions within a distributed enterprise to detect expert communities (Expert Locator). We then describe tools that facilitate group knowledge annotation (KEAN), group learning (OWL) and group search (SCOUT). Finally, we discuss our efforts to create and deploy tools for peer-to-peer knowledge communication/exchange (CVW and TrIM). We describe the efficacy of these tools and illustrate how they collectively enable peer-to-peer knowledge management. We conclude summarizing some remaining challenges.

Keywords: peer-to-peer computing, knowledge management, expertise management, knowledge discovery, collaboration.

Peer-to-Peer (P2P) Knowledge Management (KM)

Figure 1 illustrates some of the interdisciplinary relationships among peer-to-peer computing, knowledge management, human language technology, and collaboration. For example, whereas *peer-to-peer computing* provides services such as efficient indexing and caching of distributed content to enable rapid and efficient content discovery, *human language technology* provides for the automated processing of speech and language in text, audio, and video to perform the extraction, summarization, generation, and interactive dialogue on particular content. In contrast, collaboration technology provides services such as awareness of virtual participants and materials, text, audio, and video conferencing, workflow, shared applications (including whiteboarding), and persistence of content across sessions. *Knowledge management* is the strategy, policy, and technology that empowers learning organizations. Knowledge management systems often seek to exploit services provided by peer-to-peer computing, collaboration, and even human language technology to enable knowledge discovery, expertise management, knowledge sharing, and peer-

Invited Talk. In Zaihrayeu, I. (ed.). 2004. Workshop on *Peer-to-Peer Knowledge Management*, August 22, 2004. Royal Sonesta Hotel, Cambridge, MA.

to-peer knowledge community support. Staab et al. (2001) detail the role of human language technologies for knowledge management.

The dramatic growth and success of peer-to-peer networking for file sharing and collaboration illustrates the value of placing power to the edge. Peer-to-peer file sharing services such as Napster, Morpheus and KaZaa illustrate the benefits of distribution efficiency and birds of a feather information discovery in dynamic, autonomous peer-to-peer networks. Peer-to-peer collaboration conversely illustrates the benefits of fault tolerance, efficiency, and scalability for group work. While peer-to-peer computing and communications illustrates the power of self organizing networks, it raises new challenges (e.g., discovery, management) as well as novel opportunities for solving these challenges.

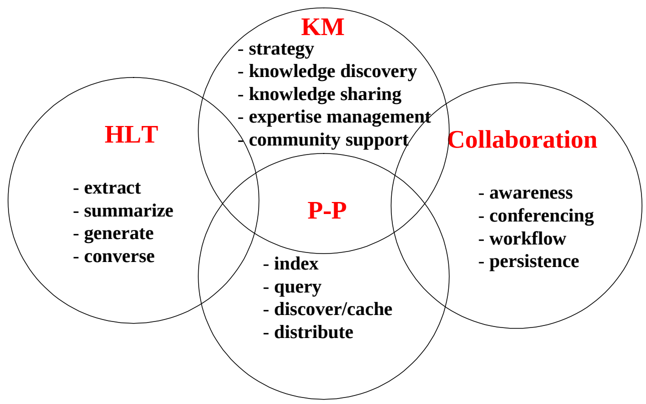

Figure 1. Interdisciplinary Relationships

As Figure 2 illustrates, our research and development aims to provide full knowledge life-cycle management, to include Web-based expertise and knowledge discovery, project and partnership creation (leader identification, team formulation, team facilitation/collaboration), support for knowledge creation, and knowledge delivery. All of this should be supported by a knowledge information infrastructure which includes tools for staff and project discovery, information sharing (e.g., transfer folders), capture/reuse of knowledge and lessons learned, and virtual place-based collaboration tools. We also are experimenting with new virtual organizational models to enable distributed teaming of expert talent (see nrrc.mitre.org). We document our knowledge management experiences at MITRE in Maybury (2003) and collect best practices in KM strategy, process, and benchmarking in Morey, Maybury, and Thuraisingham (2000).

Figure 2: Knowledge Management Processes

The remainder of this article outlines the field of research at the intersection of peer-to-peer computing and knowledge management. We first briefly describe peer-to-peer group knowledge discovery. We then discuss the analysis of peer-to-peer computing to support expert peer and expert network discovery. We then consider the facilitation of group knowledge creation and peer-to-peer knowledge communication/exchange. We finish discussing the facilitation of peer-to-peer knowledge communication/exchange.

Peer and Group Knowledge Discovery

Knowledge discovery ranges from search engines applied to the web or corporate holdings (e.g., Google) to advanced question answering systems which interpret natural language questions and extract answers from sources (See e.g., Language Computer Corp.). Our own research has focused on peer and group knowledge discovery. For example, in support of Translingual Information Detection, Extraction, and Summarization (TIDES), Figure 3 illustrates the MITRE Text and Audio Processing System (MITAP) created by Laurie Damianos and colleagues which takes input from over 90 sources ranging from formal content from the medical literature to content from the Center for Disease Control and World Health Organization to more informal email from the ProMed medical network. This content is categorized by disease, source, region, person, and organization using natural language information extraction. This automatically extracted semantic content enables content-based access to hundreds of users via a standard news reader. Messages are cross posted to relevant news groups and message subjects are automatically generated stating the disease type, location, and number of deaths (e.g., "Ebola Gabon 147 people"). This enables the user to rapidly find answers to questions such as "What is the status of the current Ebola outbreak?" to yield "The epidemic is contained; as of 12/22/00, there were 421 cases with 162 deaths." In summary, analysis of traditional peer-to-peer messaging, such as in the ProMED news group, enables situational awareness of global infectious diseases.

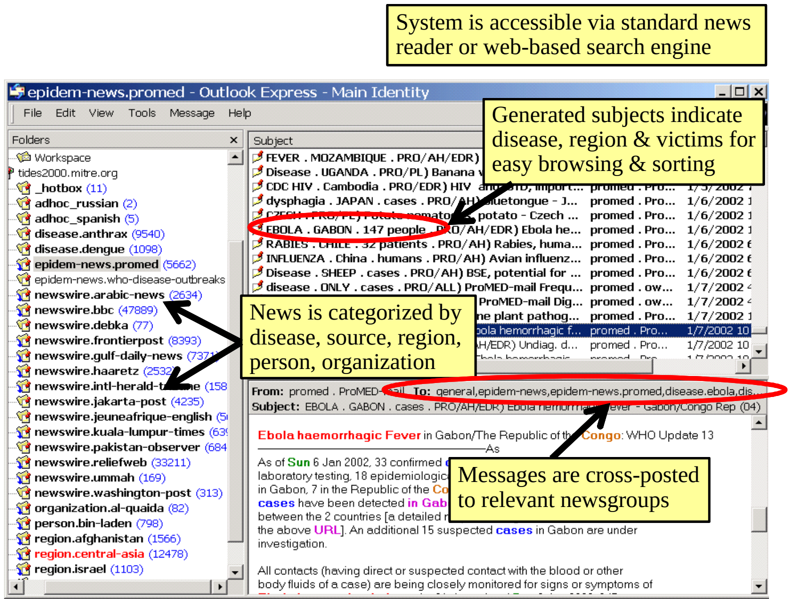

Figure 3. MITRE Text and Audio Processing System (MITAP)

Content analysis of peer-to-peer interactions associated with geospatial extent can analogously be used to enhance situational awareness. Figure 4 illustrates the Social Indicator Analysis Method (SIAM) investigated by Ray D'Amore and colleagues, in which a query is performed across multiple search engines, analyzed over time, location, and source to assess the social interest of sources in a topic or issues. Performing longitudinal analysis of the outcome can provide insight into the concerns of a particular geographic region (e.g., town, country, state, country). This automated method was successfully used to assess risks associated with Y2K much faster and with less cost than equivalent manual analyses. In summary, by performing geospatial and temporal analysis over peer-to-peer communications, an accurate situational picture could be rapidly and inexpensively formulated.

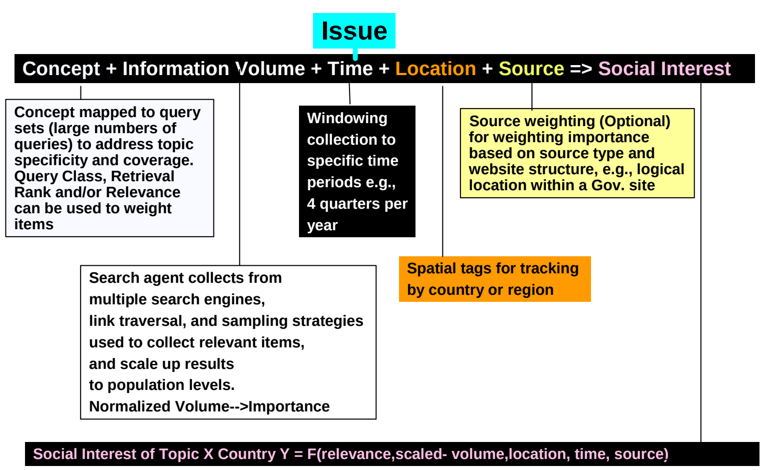

Figure 4. Social Indicator Analysis Method (SIAM)

Expert and Expert Community Discovery

Distribution of staff, decreasing project size, and cost/time pressure are driving a need to leverage enterprise expertise by quickly discovering who knows what and forming expert teams. Those in need typically have little or no means of finding experts other than by recommendation. Unfortunately, busy experts do not have time to maintain adequate descriptions of their continuously changing specialized skills. Also, our experience with "skills" databases indicates that they are difficult to maintain and quickly outdated.

Analysis of peer-to-peer interactions and artifacts created by individuals is an obvious method of expertise discovery and assessment. A number of researchers and commercial enterprises have explored automating the process of expertise discovery and assessment. These include manual expert self-nomination and discovery (e.g., Dataware II Knowledge Directory[1]), analysis of users' search and publication histories (Autonomy Agentware Knowledge Server[2]), analysis of email content (Yenta by Foner, 1997; Tacit KnowledgeMail[3]), WWW browsing patterns (Cohen et al. 1998), software library usage (Vivacqua 1999), bibliographic reference patterns (AT&T's Referral Web (Kautz et al. 1997)), and Latent Semantic Indexing (LSI) (U.S. West's Expert-Expert Locator by Streeter & Lochbaum 1988).

In contrast, Abuzz's Beehive[4] is one of many systems that provide an on-line community environment to support question-answer dialogues between users and registered "experts". Users can learn from other user's question-answer dialogues posted under specific topics. This is similar to The Answer Garden (Ackerman and Malone, 1990) which categorized questions into an ontology which could be browsed by users to find questions-answers similar to their own question. If users did not find a related question they were referred to an expert. Emerging on-line commercial systems attempt to also track each experts' performance; and the general trend is to use user ratings and experts response times as a basis for measuring competence. Essentially, social filtering is used to qualify the level of expertise of registered experts. As such systems often suffer from the cold-start problem where there is a mismatch between the number of experts and users. In some cases experts outnumber users; discouraging experts' participation or affecting revenue. In other cases, there is a dearth of experts (or qualified experts) and users become frustrated because of poor response times or low quality answers. While these systems (e.g., XperSite.Com[5]) present interesting expertise management paradigms, a number of core problems remain, including representing and measuring an expert's qualifications, as well as matching questions to the

---

[1] Dataware Knowledge Management Systems White Paper *(http://www1.dataware.com/forum/kms/kmsfull.htm)*

[2] Autonomy Technology White Paper *(http://www.autonomy.com/tech/wp.html)*

[3] Tacit KnowledgeMail (*http://www.tacit.com/products/knowledgemail.html*)

[4] Abuzz "Ask Anything" (http://www.abuzz.com/)

[5] XperSite.com (www.xpersite.com)

appropriate experts.

Expertise location can be broken down into *identification* and *selection* phases (Ackerman et al, 1999). We extend this to include expertise *measurement* and *representation* as fundamental to the overall problem of capturing and exploiting expertise. In Maybury, D'Amore, and House (2002) we describe a component-based architecture that supports a wide-range of expertise management services, defining major services encompassing brokering, registration, finding, qualifying, selection, and Q&A. We next describe our efforts to create tools to support individual expert and expert community management.

Expert Finder

Expert Finder (Mattox, Smith, and Seligman 1998) is an expert skill finder that exploits intellectual products to support automated expertise classification. Expert Finder mines information and activities on MITRE's corporate intranet and provides this in an intuitive fashion to end users. Figure 5 illustrates the system in action. In this example, a user is trying to find machine translation experts in the corporation. When the user searches using the term "machine translation," the system ranks employees by the number of mentions of a term or phrase and its statistical association with the employee name either in corporate communications (e.g., newsletters) or based on what they have published in their resume or document folder (a shared, indexed information space). Integrated with MITRE's corporate employee database, employees are ranked by frequency of mentions, pointing to sources in which they appear.

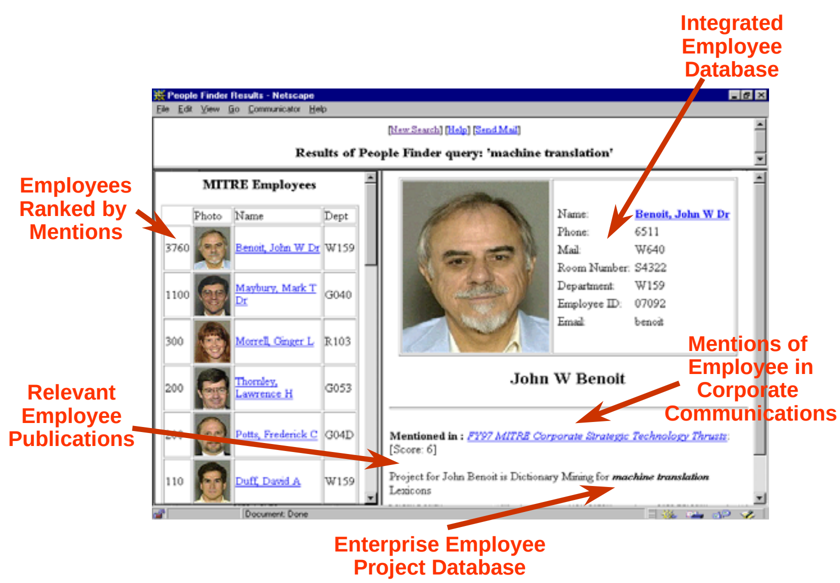

Figure 5: MII Expert Finder "Machine Translation" Example

In and of itself, each source of employee information mentioned above is generally not sufficient to determine if an employee is an "expert" in a particular topic. Expert Finder relies on the combination of evidence from many sources, and considers someone an expert in a particular topic if they are linked to a wide range of documents and/or a large number of documents about that topic. For documents about a topic that are published by an employee, Expert Finder relies on the number of documents published by an employee about a given topic to provide an "expert score" for that employee. The only exception is that of an employee's resume, which is given additional weight as a self-definition of an individual's expertise. For documents that mention employees and topics, Expert Finder first locates the proper names within the text using a commercial product that tags names within a document[6], and subsequently analyzes frequency and location of query topics based on document type for evidence of expertise.

The performance of ten technical human resource managers, professionals at finding experts, was compared with that of Expert Finder for the task of identifying the top five corporate experts in five diverse specialty areas. Expert Finder was able to find approximately 30% of the experts humans could find (recall) and of those reported by Expert Finder, approximately 40% were considered expert by human experts (precision). We found this impressive given that human experts agreed at best about 60% of the time on who was an expert. Users of Expert Finder have had generally positive comments about the system, in particular because the system tends to find an expert, a list of experts, or someone who is one phone call away from an expert, for a large variety of queries.

Expert Locator

Whereas ExpertFinder exploits individually created artifacts, Expert Locator (D'Amore et al. 2003) addresses the problem of detecting extant or emerging areas of expertise and associated communities without a priori knowledge of their existence by analyzing peer-to-peer interactions. In large, dynamic organizations expertise networks are often emergent and difficult to discern from the formal organizational structure. Expert Locator is designed to extract expertise networks and integrate them into the overall expertise management system.

The core model associates activities with workplace semantics and social context. In the baseline design, activities are represented by an activity identifier and associated with a membership list (individuals involved in the activity), and a semantic context (a set of terms or other descriptors that describe key themes or topics associated with the activity). For example, a technical exchange meeting (TEM) addressing *mobile computing* consists of the TEM identifier (title, corporate activity number, date and location), a list of participants or contributors, an "owning" organization and an activity description. Similarly, corporate share folders assigned to each staff member can be represented as a dissemination activity, with membership (the share folder's owner) and semantics extracted from the associated items (e.g., briefing paper). Note that each staff member has specific attributes such as job title or position, organizational home, and technical level (seven technical levels cover the technical staff). These attributes provide another basis for identifying relationships between staff members.

---

[6] NameTag from IsoQuest Corporation.

On a daily basis, Expert Locator collection agents mine various MITRE workspaces to identify activities such as projects, technical exchange meetings, and public share folders that may be new or may have changed since the last update. Expert Locator works without user queries to identify expertise areas; a distinction between it and other expertise locator tools. Figure 6 illustrates Expert Locator in use illustrating the network of experts generated by typing in the search term “peer-to-peer”. The network shows individuals such as Paul Silvey connected to others such as Amy Kazura via an organization link, department G036. Individual experts such as Amy Kazura are further related to other experts via organization links (G036), projects (e.g., Semantic Web Services), and mailing lists (ontology-list). This interactive graph can be further explored (navigated, expanded, collapsed) to perform further discovery of MITRE’s knowledge network. Computational details of Expert Locator and its performance are detailed in (Maybury, D’Amore, and House 2002). Current results suggest automatic extraction of expertise networks is feasible.

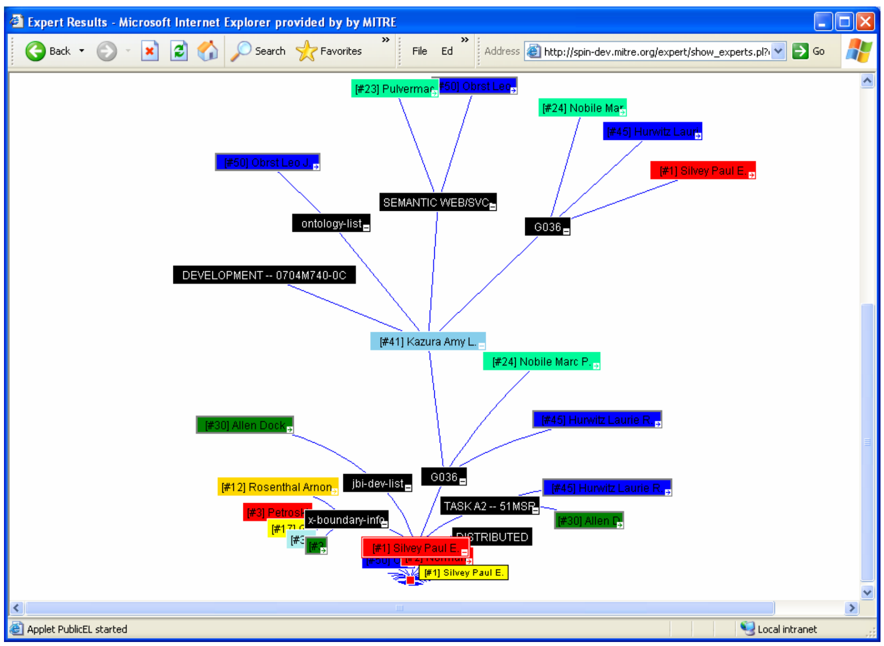

Figure 6. Expert Locator Example

Facilitating Group Knowledge Creation

In addition to discovering experts from artifacts, benefiting from their artifacts and activities is also important. For example, Figure 7 illustrates the Knowledge Exchange and Annotation eNgine (KEAN) prototype which enables experts to rapidly assess and annotate content for subsequent retrieval. While searching the web, users of KEAN can easily assign a topic classification and rating to any web page. These are subsequently aggregated across users and provided to all for search. Thus if a user knows or finds an expert, they can search for the content considered most valuable by that expert on any topic. Using KEAN to support web search experiments exploring the correlation between time and utility, we discovered that 66% of all URLs “viewed” for greater than 78 seconds were classified as high utility (6-10).

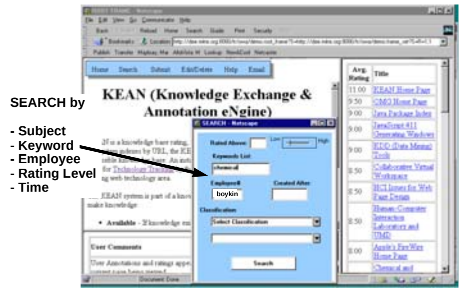

Figure 7. Knowledge Exchange and Annotation eNgine (KEAN)

In related research, we explored the digital instrumentation of tools used by experts to determine their level of utilization and, by implication, expertise. Organization Wide Learning (OWL), created by Frank Linton, is an agent that sits in the background of an application (currently Microsoft's Word), observes a user and a user's peers work, and provides each individual with unique and timely answers to the question: What should I learn next (about this application)? OWL bases its answer on an analysis of the functionality that the user's peers have already found useful and that the user is not using or is under-using. The approach taken in OWL will work with any application used in a networked environment where multiple users perform sets of similar tasks. OWL could also recommend URLs on an intranet, classes in an object-oriented language such as Java, and so on. In the case of Microsoft Word, in a long term study of over 20 users, three commands (delete, save, and open) accounted for over 50% of command usage and 10 commands accounted for 81% of usage. However, an individual user oftentimes does not use common and useful actions. Figure 8 illustrates the number of users of a command and overall command usage for the top 50 most used Word commands. This data allows the system to compare what it expects a user to exhibit and contrast this with their actual command usage. This then is used to make recommendations for learning assuming that if peers have found a command useful so will the user.

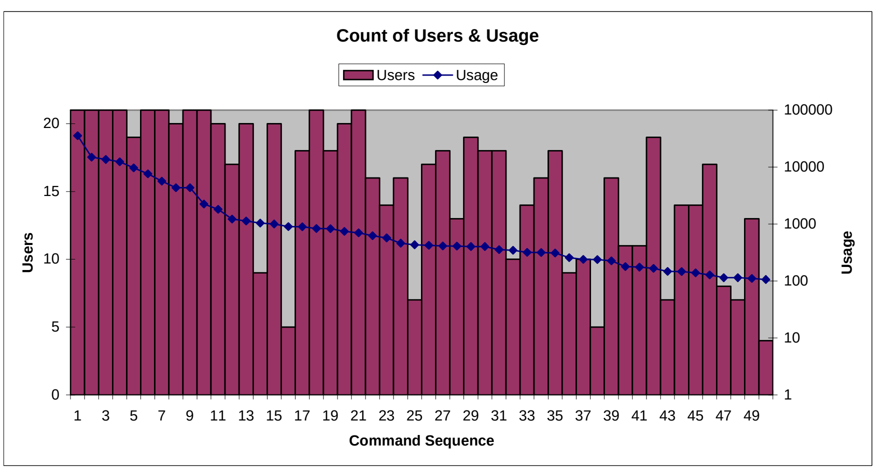

Figure 8. Organization Wide Learning (OWL)

Peers can even more directly benefit from one another's activities by performing shared tasks. Scout is a multi-user collaborative retrieval tool. Its basic hypothesis is that group (coordinated) searching can be more effective than multiple (independent) searchers working autonomously. This "next generation" information retrieval system addresses multi-user, coordinated searching, shared analysis, and has a built-in recommender system. The system tracks topics, users, and provides a persistent knowledge store. Peers start by generating a shared task folder. Retrieved information is organized by domain. Search engine statistics are provided to the users and results are clustered and categorized offline. Ratings, annotations and actions associated with particular content are stored and made available to the work group. Group search is more efficient and comprehensive as users continuously benefit from each others cumulative knowledge.

Facilitating Peer-to-Peer Knowledge Communication/Exchange

Enabling peer-to-peer shared work and knowledge communication exchange are valuable P2P KM functions. The Collaborative Virtual Workspace (CVW), shown in Figure 9, is a pioneering collaborative software environment that provides a "virtual building" where teams can communicate, collaborate, and share information, regardless of their geographic location. CVW takes virtual meetings one step further and enables virtual co-location through persistent virtual rooms, each incorporating people, information, and tools appropriate to a task, operation, or service. MITRE has ceded CVW to the public domain (http://cvw.sourceforge.net) in order to encourage the widest possible dissemination and use of CVW to help communities further understand the deployment and development of synchronous computer mediated collaboration. We have learned peer-to-peer collaboration is enhanced when groups have a clear joint mission and are trained together and is mitigated by security, privacy, and lack of incentives for group work. P2P collaboration solutions (e.g., Groove) illustrate the benefits of peer-to-peer communication including survivability and scalability.

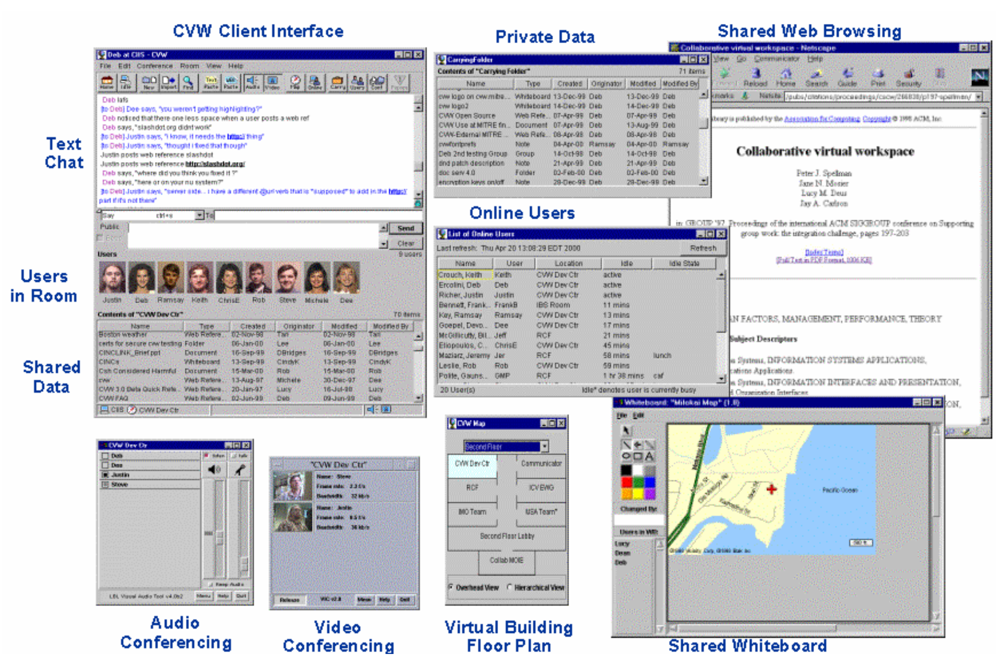

Figure 9. Collaborative Virtual Workspace (CVW)

To further enable collaboration with more heterogeneous groups of peers, the Translingual Instant Messaging (TrIM) system was developed. TrIM uses MITRE's open architecture secure "simple instant messaging and presence" protocol (SIMP) integrated with standard machine translation tools. SIMP provides standard notions of buddy lists to identify and maintain awareness of peers. Figure 10 illustrates the use of TrIM between an English and Spanish speaker. TrIM currently supports languages including English, Spanish, Portuguese, German, Italian, French, Korean, Thai, Japanese, Chinese (simplified), Russian, Polish, Ukrainian, Arabic, and Hebrew.

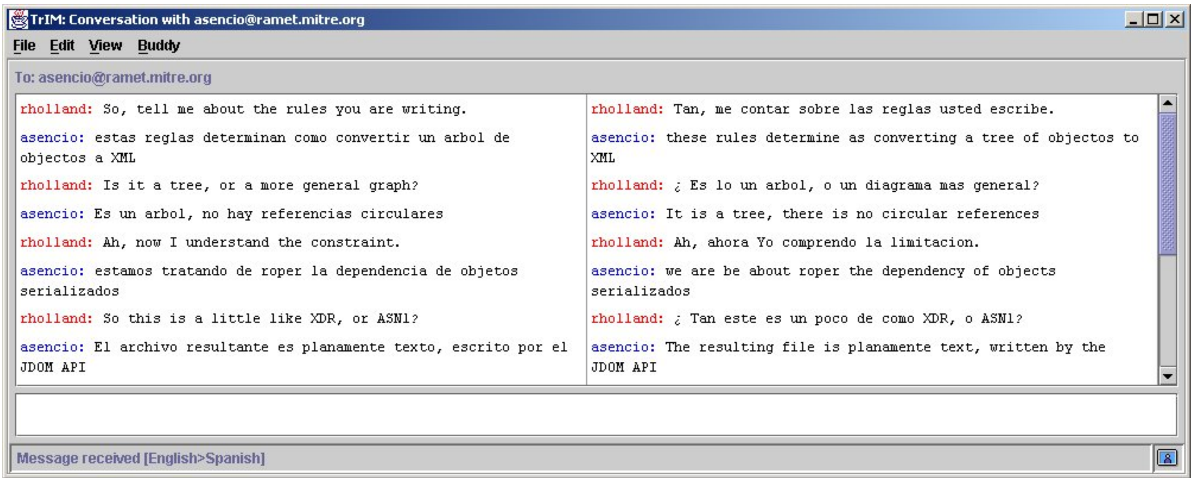

Figure 10. Translingual Instant Messaging (TrIM)

Lessons Learned and Future Research

Our investigations into peer-to-peer knowledge management illustrate the diversity of capabilities necessary to support effective peer synergy across an enterprise. In this article, we have illustrated the roles human language technology and collaboration play in support of peer-to-peer knowledge management. In our P2P KM experiences we have learned a number of key lessons (Morey et al. 2000). These include:

- <u>People</u>, and the <u>cultures</u> that influence their behaviors, are the single most critical resource for successful knowledge creation, dissemination, and application. *Understand and influence them*.

- <u>Cognitive</u>, <u>social</u>, and <u>organizational learning processes</u> are essential to the success of a knowledge management strategy. *Focus your strategy on enhancing these processes*.

- <u>Measurement</u>, <u>benchmarking</u>, and <u>incentives</u> are essential to accelerate the learning process and to drive cultural change. *Create a tailored balanced scorecard to target what you want to improve*.

- Knowledge management programs can yield impressive benefits to individuals and organizations if they are <u>purposeful</u>, <u>concrete</u>, and <u>action-oriented</u>. *Make yours so*.

Exploiting peer-to-peer interactions raises a number of technical challenges and issues that require further exploration. The first issue is *security and privacy*. By exposing information about individual activities and interests, users hope to benefit from other peers' knowledge. However, they potentially sacrifice confidentiality although aggregation and anonymization can be used to ensure privacy. A second issue is one of *timeliness and performance*. For example, expensive tagging and analysis of source documents requires time. This can effect a range of P2P KM applications including knowledge discovery, shared search, and expert discovery. Finally, the *dynamicity and autonomy* provided by the self organizing properties of peer-to-peer computing require further exploration to assess their benefit for both collaboration and knowledge management services.

Acknowledgements

The TIDES/MITAP effort is led by Laurie Damianos and SIAM is led by Ray D'Amore. Daryl Morey created KEAN, Frank Linton developed OWL, and Ray D'Amore built SCOUT. David Mattox and Inderjeet Mani created MITRE's Expert Finder, the original idea for which came from Chris Elsaesser. Ray D'Amore and Manu Konchandy designed and developed the Expert Locator tools. Peter Spellman created CVW and Rod Holland led a team to create TrIM.

References

Ackerman, M. and Malone, T., (1990). The Answer Garden: A Tool for Growing Organizational Memory. Proceedings of the ACM Conference of Computer Supported Cooperative Work (CSCW'94).

Ackerman, M, McDonald, D., Lutters, W., and Muramatsu, J. (1999). "Recommenders for Expertise Management", ACM SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation University of California, Berkeley August 19, 1999.

Cohen, A. L., Maglio, P. P. and Barrett, R. (1998). "The Expertise Browser: How to Leverage Distributed Organizational Knowledge." Paper presented at the Workshop on Collaborative Information Seeking at CSCW 98, Seattle, WA.

D'Amore, R., Helm, D., Valley, G. and Damianos, L. (2003). Expertise Sharing and Community Detection European Conference on Knowledge Management (ECKM 2003), Oxford, U.K., September, 2003.

Foner, L. (1997). Yenta: A Multi-Agent Referral System for Matchmaking System. Proceedings of The First International Conference on Autonomous Agents, Marina Del Ray, CA, 1997.

Mattox, D., Maybury, M. and Morey, D. (1999). Enterprise Expert and Knowledge Discovery. International Conference on Human Computer International (HCI 99). 23-27 August 1999. Munich, Germany, 303-307.

Mattox, D., Smith, K., and Seligman, L. (1998). Software Agents for Data Management. In Thuraisingham, B. Handbook of Data Management, CRC Press: New York. 703-722.

Maybury, M., D'Amore, R. and House, D. (2002). Automated Discovery and Mapping of Expertise. In Ackerman, M., Cohen, A., Pipek, V. and Wulf, V. (eds.). Beyond Knowledge Management: Sharing Expertise. Cambridge: MIT Press.

Maybury, M. (2003). Knowledge Management at MITRE. In Rao, M (editor) Case Studies in Knowledge Management Practices. Leading With Knowledge: KM Practices in Global InfoTech Companies. New Delhi: Tata McGraw-Hill. http://www.mitre.org/support/papers/tech_papers_02/maybury_knowledge/index.html

Morey, D.; Maybury, M. and Thuraisingham, B. editors. (2000). *Knowledge Management: Classic and Contemporary Works*. Cambridge: MIT Press. (www.amazon.com/exec/obidos/ASIN/0262133849/ref=lm_lb_11/104-8249298-4021519, http://mitpress.mit.edu/book-home.tcl?isbn=0262133849)

Staab, S., Ciravegna, F., Maybury, M., Moldovan, D., Niklfeld, G. Nov/Dec. (2001). Human Language Technologies for Knowledge Management. *IEEE. Intelligent Systems, Trends & Controversies*, 16(6): 84 –94.

Streeter, L. & Lochbaum, K. (1988). An Expert/Expert-Locating System Based on Automatic Representation of Semantic Structure, *Proceedings of the Fourth Conference on Artificial Intelligence Applications*, San Diego, CA, March 1988, 345-350.

Vivacqua, A. (1999). Agents for Expertise Location. In Proceedings 1999 AAAI Spring Symposium on Intelligent Agents in Cyberspace. Technical Report SS-99-03. Stanford, CA, USA, March 1999.